Molecular Biology and Evolution, 2025, **42**, 1–16
https://doi.org/10.1093/molbev/msaf058
Advance access publication 18 March 2025
Discoveries

MBE OXFORD

A Segregating Structural Variant Defines Novel Venom Phenotypes in the Eastern Diamondback Rattlesnake

Pedro G. Nachtigall [1,2] Gunnar S. Nystrom [1] Emilie M. Broussard [1] Kenneth P. Wray [3] Inácio L. M. Junqueira-de-Azevedo [2] Christopher L. Parkinson [4] Mark J. Margres [5] Darin R. Rokyta [1,*]

[1]Department of Biological Science, Florida State University, Tallahassee, FL, USA
[2]Laboratório de Toxinologia Aplicada, CeTICS, Instituto Butantan, São Paulo, SP, Brazil
[3]Biodiversity Center, University of Texas at Austin, Austin, TX, USA
[4]Department of Biological Sciences, Clemson University, Clemson, SC, USA
[5]Department of Integrative Biology, University of South Florida, Tampa, FL, USA

***Corresponding author**: E-mail: drokyta@bio.fsu.edu.
Associate editor: Sandro Bonatto

Abstract

Of all mutational mechanisms contributing to phenotypic variation, structural variants are both among the most capable of causing major effects as well as the most technically challenging to identify. Intraspecific variation in snake venoms is widely reported, and one of the most dramatic patterns described is the parallel evolution of streamlined neurotoxic rattlesnake venoms from hemorrhagic ancestors by means of deletion of snake venom metalloproteinase (SVMP) toxins and recruitment of neurotoxic dimeric phospholipase A2 (PLA2) toxins. While generating a haplotype-resolved, chromosome-level genome assembly for the eastern diamondback rattlesnake (*Crotalus adamanteus*), we discovered that our genome animal was heterozygous for a ∼225 Kb deletion containing six SVMP genes, paralleling one of the two steps involved in the origin of neurotoxic rattlesnake venoms. Range-wide population-genomic analysis revealed that, although this deletion is rare overall, it is the dominant homozygous genotype near the northwestern periphery of the species' range, where this species is vulnerable to extirpation. Although major SVMP deletions have been described in at least five other rattlesnake species, *C. adamanteus* is unique in not additionally gaining neurotoxic PLA2s. Previous work established a superficially complementary north–south gradient in myotoxin (MYO) expression based on copy number variation with high expression in the north and low in the south, yet we found that the SVMP and MYO genotypes vary independently, giving rise to an array of diverse, novel venom phenotypes across the range. Structural variation, therefore, forms the basis for the major axes of geographic venom variation for *C. adamanteus*.

Keywords: venom, rattlesnake, structural variation, conservation, genomics

Introduction

Structural variants (SVs) are a class of mutation affecting large (typically defined as >50 nucleotides) regions of a genome (Weischenfeldt et al. 2013) and include changes in chromosomal organization (e.g. translocations and inversions) or content (e.g. deletions and duplications). SVs can impact gene-expression patterns through duplication or deletion of genes, thereby increasing or decreasing mRNA and protein levels, and alter linkage patterns in affected regions (Mérot et al. 2020). SVs appear to be ubiquitous and are known to affect phenotypes critical to adaptation and speciation (Zhang et al. 2021b). For example, SVs are involved in high-altitude adaptation in humans (Shi et al. 2023), and a large chromosomal inversion was found to control tail length in deer mice, which is relevant for adaptation to forest or prairie habitats (Hager et al. 2022). Additionally, a 2.25 Kb retrotransposon insertion contributes to differences in plumage patterns involved in premating isolation in two subspecies of European crow (Weissensteiner et al. 2020). Despite such canonical examples, most SVs and their roles in evolutionary processes remain uncharacterized because technical challenges preclude their detection and study (Mérot et al. 2020; Xu et al. 2021). Read length and accuracy, in particular, limit the size and nature of SVs that can be detected with high confidence (Ho et al. 2020); many SVs are not detectable without long-read sequencing data (Chaisson et al. 2019).

Recent innovations in DNA sequencing and new bioinformatic approaches facilitate the generation and assembly of genomes at chromosome-level with haplotype resolution (Giani et al. 2020; Cheng et al. 2022). Specifically, the combination of PacBio HiFi sequencing, which results in reads >15 Kb with error rates of ∼1%, Hi-C, which produces short reads capturing 3-dimensional chromatin organization, and novel computational pipelines allows the generation of high-resolution genomes (Garg 2021; Rhie et al. 2021). Haplotype-resolved assemblies enable precise identification of all classes of genetic variation, from single-nucleotide polymorphisms (SNPs) to large SVs, including those in the heterozygous state in the source organism. The resolution of whole-genome haplotypes of species and populations has, in particular, revealed SVs affecting adaptive traits in diverse lineages (Chaisson et al. 2019; Low et al. 2020; Ebert et al. 2021; Hämälä et al. 2021; Garg 2023; Nakandala et al. 2023; Li et al. 2024a), suggesting that such approaches are critical for understanding phenotypic evolution.

Snake venoms are variable in composition at all phylogenetic scales (Casewell et al. 2020; Holding et al. 2021), yet the most phenotypically consequential known variants tend to involve

Received: November 4, 2024. **Revised:** March 3, 2025. **Accepted:** March 6, 2025

large SVs segregating within species (Dowell et al. 2018). At the family level, major differences exist between the dominant types of toxins present. For example, elapids have venoms comprised largely of phospholipases A2 (PLA2s) and three-finger toxins (3FTxs), whereas the dominant components in viperid venoms tend to be snake venom metalloproteinases (SVMPs), snake venom serine proteases (SVSPs), and an independently recruited class of PLA2s (Oliveira et al. 2022). Among closely related species, SVs affecting numbers and identities of toxin-family paralogs commonly account for major venom differences (Dowell et al. 2016, 2018; Almeida et al. 2021; Margres et al. 2021; Nachtigall et al. 2022). Several rattlesnake species are polymorphic for vastly different venom phenotypes, with some populations expressing predominantly neurotoxic venoms and others expressing hemorrhagic venoms. Where examined, the underlying genetics involve distinct haplotypes for two venom-gene tandem arrays, PLA2s and SVMPs, that differ by major SVs (Dowell et al. 2018). The haplotypes are maintained such that the potent PLA2 haplotype, which encodes a dimeric PLA2 neurotoxin (Whittington et al. 2018), is nearly always associated with an SVMP haplotype with major portions of the SVMP array deleted. The maintenance of this polymorphism within species has been described in *Crotalus scutulatus*, *C. horridus*, and *C. helleri* (Dowell et al. 2018) and suggests that the loss of SVMP paralogs is beneficial only in the presence of a complementary neurotoxic PLA2 haplotype.

The eastern diamondback rattlesnake (*Crotalus adamanteus*) has one of the most thoroughly studied venoms of any animal (Rokyta et al. 2011, 2012; Margres et al. 2014, 2015a, 2015b; Rokyta et al. 2015; Wray et al. 2015; Margres et al. 2016a, 2016b, 2017a, 2017b; Rokyta et al. 2017; Margres et al. 2019; Schonour et al. 2020; Hogan et al. 2021; Harrison et al. 2022; Hogan et al. 2024), particularly at the genomic level, but still continues to yield novel insights into the mechanisms of venom evolution. *C. adamanteus* is the largest species of rattlesnake and is endemic to the southeastern United States, where it feeds primarily on mammals such as mice, rats, and rabbits (Means 2017). Its venom was among the first to be fully characterized by means of next-generation sequencing and proteomic approaches (Rokyta et al. 2011, 2012; Margres et al. 2014; Rokyta et al. 2015), and the species has been used to reveal patterns of geographic (Margres et al. 2015a, 2016a, 2016b, 2017b, 2019) and ontogenetic (Wray et al. 2015; Rokyta et al. 2017; Schonour et al. 2020) venom variation, the effects of hybridization on venom composition (Harrison et al. 2022), and the relationships between venom and other trophic adaptations (Margres et al. 2015b; Hogan et al. 2021). The major geographic pattern in venom variation previously described for this species involves a north–south gradient in copy number of the Myotoxin A (MYO) gene that correlates with a dramatic variation in the abundance of the encoded toxin in the venom, ranging from complete absence to a majority of the protein content (Margres et al. 2017a). Most recently, a chromosome-level genome assembly was described for *C. adamanteus* and used to characterize the gene-expression and epigenetic bases for its ontogenetic venom change (Hogan et al. 2024). Despite these substantial efforts, we are far from a comprehensive understanding of the patterns and causes of venom and genetic variation in this, or any, species.

The examination of patterns of genetic variation in *C. adamanteus* is of particular interest given growing concerns related to the conservation and management of this species, which is experiencing a rapid population decline and is considered vulnerable throughout much of its range (Waldron et al. 2013). *C. adamanteus* is state-endangered in North Carolina, a species of special concern in South Carolina, likely extirpated from Louisiana, and is currently under review for federal protection under the Endangered Species Act (Fish and Service 2012). This species is also included in the Department of Defense at-risk herpetofaunal species priority list. The rapid decline is primarily caused by anthropogenic forces (Means 2017), such as habitat loss and degradation, human persecution (Means 2009), and road mortality. Characterization of patterns of genetic variation, particularly fitness-related genetic variation (i.e. functional diversity, reviewed in Mable 2019), could, and perhaps should, provide the basis for targeted conservation and management efforts for this species.

We sequenced and assembled a chromosome-level, haplotype-resolved genome of *C. adamanteus* that fortuitously revealed a large SV spanning six SVMP genes between haplotypes. This SV substantially alters the venom proteomes of homozygotes for the deletion, suggesting a large phenotypic effect. Homozygotes for the deletion are rare range-wide but are the predominant genotype along the northwestern edge of the species' range. The distribution of this deletion complements the previously described pattern in this species (Margres et al. 2017a) of copy-number variation for myotoxins to generate overlapping gradients of venom phenotypes across the species' range, necessitating a reexamination of conservation strategies based on these patterns of fitness-related venom phenotypes.

Results

Chromosome-level Genome Assembly

Our *C. adamanteus* genome assembly, derived from a heterogametic female, comprised 19 chromosomes, including seven macrochromosomes (ma1–7), 10 microchromosomes (mi1–10), and both sex chromosomes (Z and W; Fig. 1a and b; Table 1), as expected for *Crotalus* species (Baker et al. 1972; Schield et al. 2019; Hogan et al. 2021; Margres et al. 2021; Hogan et al. 2024). We estimated a haploid genome size of 1.69 Gb and achieved a scaffold *N*50 of 208.8 Mb. The assembly was both highly accurate (QV score = 44.7) and complete (BUSCO = 95.7% complete, 94.6% single-copy, 1.1% duplicated, 0.9% fragmented, and 3.4% missing using the tetrapoda BUSCO gene set, which contains a total of 5,310 genes). Moreover, we identified telomeric repeats at both ends of 11 chromosomes and on one end of an additional five chromosomes, further attesting to the completeness of the assembly. Our primary assembly was similar to that previously published (Hogan et al. 2024) for *C. adamanteus* (Table 1), which was based on the same PacBio HiFi data but lower-coverage Hi-C data (~5×) from a different individual. For our new assembly, we used a greater depth of Hi-C data (~49×) from the same individual that was used for the HiFi data, allowing haplotype resolution. Our assembly statistics revealed that our genome assembly was of higher quality than any previously published snake genomes (Schield et al. 2019; Peng et al. 2020; Suryamohan et al. 2020; Li et al. 2021; Zhang et al. 2022; Peng et al. 2023).

Repeats accounted for 52.61% of the genome, including 7.09% tandem repeats and 44.17% transposable elements (TEs). Among the TEs, 22.11%, 11.98%, and 7.39% were

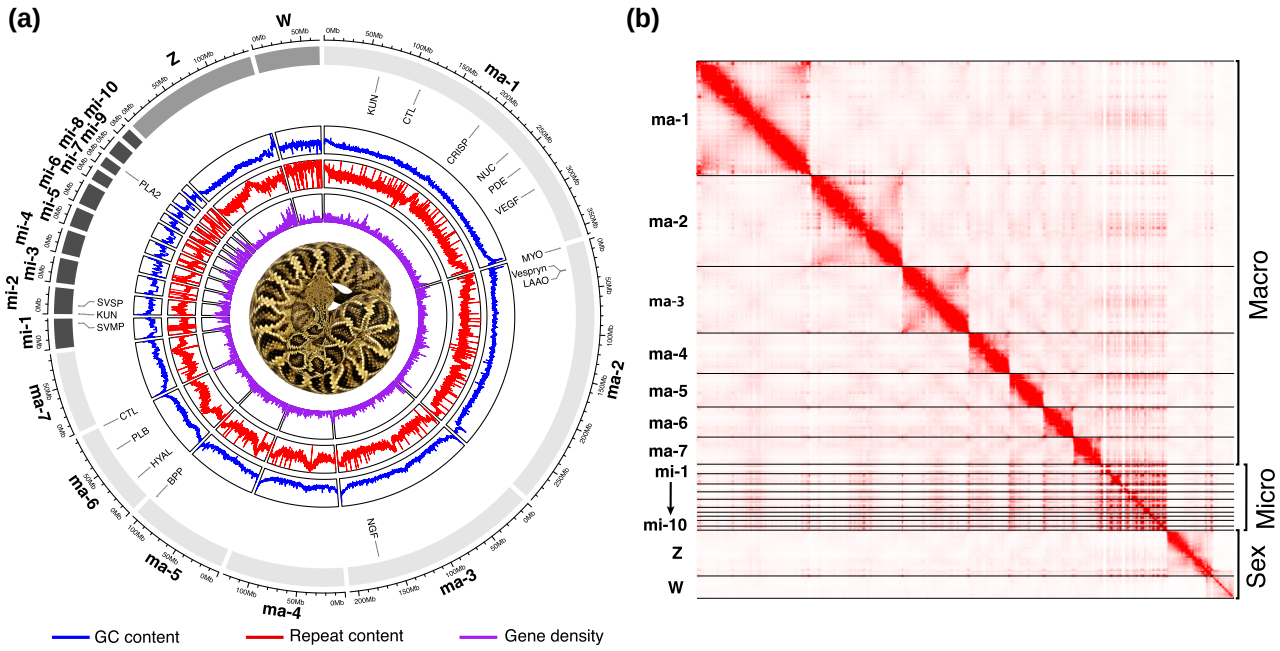

Fig. 1. Overview of the genome assembly for *C. adamanteus*. a) A circos plot shows the distribution of venom genes across the genome. Circular rings from inner to outer display the gene density (purple), repeat content (red), and GC content (blue) within 100-Kb windows. The macro-, micro-, and sex chromosomes are colored in light gray, dark gray, and gray, respectively. b) An Hi-C contact map for all 19 assembled chromosomes, with darker colors indicating stronger interactions, indicates well-defined chromosomes consistent in structure and number with previous rattlesnake genome assemblies. Snake image credit: Michael P. Hogan.

Table 1 Statistics for the primary and haplotype assemblies of *C. adamanteus* compared with the previously published assembly (Hogan et al. 2024)

	Primary	Hap1	Hap2	Previous
Total size (Gb)	1.69	1.52	1.62	1.69
Scaffold *N*50 (Mb)	208.8	207.8	208.2	208.9
Contig *N*50 (Mb)	57.8	45.3	29.4	67.5
No. chromosomes	19	18	18	19
Sex chromosome	ZW	W	Z	ZW
GC content (%)	39.9	39.9	39.9	39.9
BUSCO (%)[a]	95.7	89.9	95.7	95.7
Quality value (QV) score	44.4	40.4	43.4	42.8
Genes[b]	34,471 (20,156)	32,461 (17,861)	34,255 (19,470)	21,841 (17,810)
Venom genes	77	63	73	134

[a]Percentage of complete BUSCO genes using the tetrapoda gene set (odb10; total of 5,310 genes).
[b]Distinct gene-annotation approaches were applied in these studies. The number of genes matching the ENSEMBL database are provided in parentheses.

LINE, LTR, and DNA (i.e. DNA and MITE) TE families, respectively. To check for evidence of recent TE activity, we estimated the repeat landscape by calculating Kimura substitution levels (supplementary fig. S1, Supplementary Material online). We found evidence for a recent burst of TE activity from the LINE family, suggesting that these TEs are relevant for the genome dynamics and evolution of *C. adamanteus*. The high abundance and recent bursts of LINE families are in accordance with a previous comparative analysis of repeat sequences in squamates (Pasquesi et al. 2018), which showed that specific LINE elements may be responsible for rearrangement events in snake genomes.

We used funannotate (Palmer and Stajich 2017) to annotate 34,471 protein-coding genes, of which 15,037 were attributed to meaningful functional annotation beyond "hypothetical protein." A similarity search revealed that 20,156 of the 34,471 (58.47%) predicted protein-coding genes matched with high identity and coverage to entries in the ENSEMBL database. Using ToxCodAn-Genome (Nachtigall et al. 2024), we identified 77 venom protein-coding genes (supplementary table S1, Supplementary Material online). These numbers differ from the previous assembly (Hogan et al. 2024, supplementary tables S2 and S3, Supplementary Material online), because we only included genes with medium to high expression in the venom-gland transcriptomes relative to transcriptomes from other tissues and that have been confirmed proteomically in vipers (Oliveira et al. 2022). The venom-gland transcriptome from the genome individual showed that the major components of the venom were MYOs, SVSPs, PLA2s, C-type lectins (CTLs), and SVMPs (supplementary fig. S2, Supplementary Material online), consistent with previous studies (Rokyta et al. 2012; Margres et al. 2014, 2015a, 2016a; Rokyta et al. 2017). The SVMPs, SVSPs, and PLA2s were each organized in single tandem arrays and located on microchromosomes, whereas MYO and CTLs were each organized in single tandem arrays and located

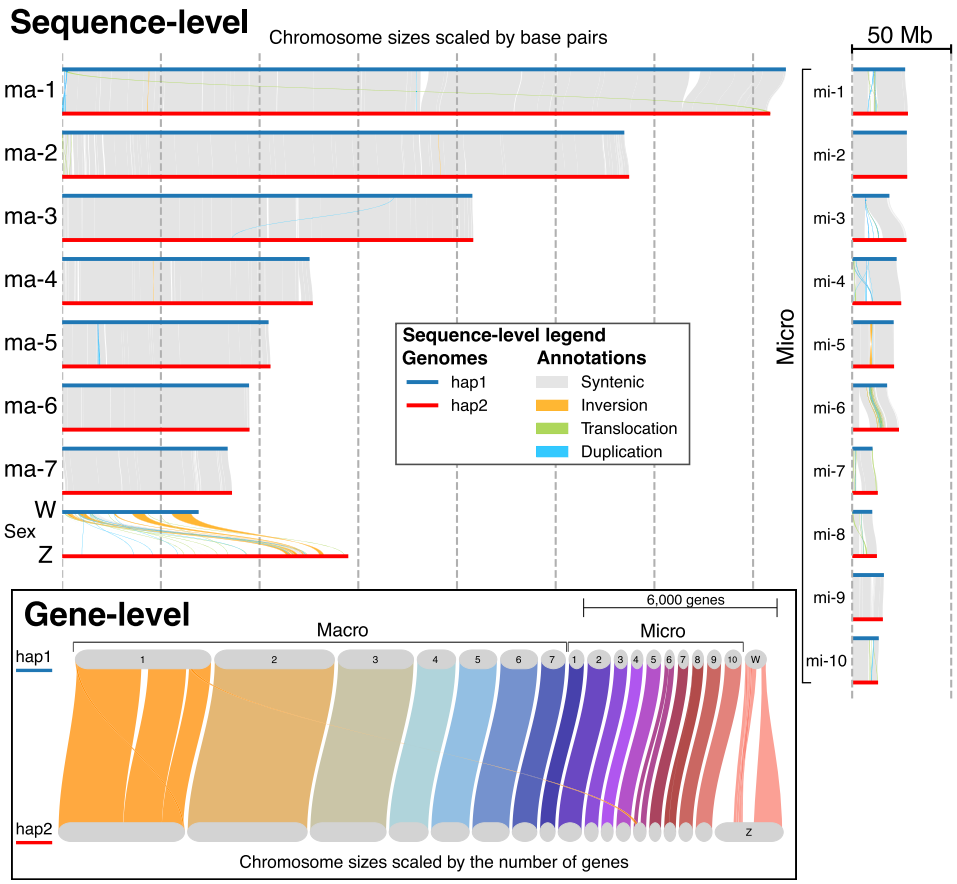

Fig. 2. Sequence-level and gene-level synteny analyses between haplotypes of the *C. adamanteus* genome. For the sequence-level analysis, the chromosome sizes are scaled in base pairs, where the distances between the vertical gray dashed lines are 50 Mb. For the gene-level analysis, the chromosome sizes are scaled by the number of genes. Both sequence-level and gene-level analyses reveal that most rearrangements are occurring in intergenic regions, which may be a result of recent bursts of TE activity. The large numbers of rearrangements in the sex chromosomes are in agreement with previous cytogenetic studies in snakes (Matsubara et al. 2006; Viana et al. 2019).

on macrochromosomes, as previously reported in other *Crotalus* genomes (Schield et al. 2019; Hogan et al. 2021, 2024; Margres et al. 2021).

Rearrangements Primarily Affected Intergenic Regions

The two haplotype assemblies returned different statistics relative to the primary assembly due to distinct rearrangements and associated sex chromosomes (Table 1). Their QV scores (>40) and Hi-C interaction maps indicated robust assemblies for both haplotypes (supplementary fig. S3, Supplementary Material online). Haplotype 1 (hap1) included the W chromosome, and haplotype 2 (hap2) included the Z chromosome. A synteny analysis revealed that the haplotypes were predominantly syntenic, with specific genomic regions showing rearrangements (Fig. 2). Rearrangements were primarily detected at the sequence-level (i.e. alignment of sequences between haplotypes, independent of gene presence/absence), and almost no rearrangements were observed affecting genes (i.e. when comparing gene positions between haplotypes) for macro- and microchromosomes. Together, these results indicate that most rearrangements were detected in intergenic regions, and these were enriched in microchromosomes and sex chromosomes (supplementary fig. S4A, Supplementary Material online). For all chromosomes, the regions containing rearrangements also showed a higher content of repetitive elements (supplementary fig. S4B, Supplementary Material online), suggesting that TE elements may play a role in generating these rearrangements.

The Z and W sex chromosomes presented high proportions of rearrangements at the sequence-level, which resulted in some rearrangements in gene locations along the chromosomes. Such rearrangements were previously reported by cytogenetic studies in snakes (Matsubara et al. 2006). The W chromosome showed the highest proportion of repeat-element-derived sequences (~85%; supplementary fig. S5, Supplementary Material online), far exceeding the corresponding value for the Z chromosome (~55%). Other snakes (Viana et al. 2019; Schield et al. 2022) show similar enrichment of repeat elements on the W chromosome. These results suggest that the W chromosome exhibits dynamic evolution and that the Z chromosome is more stable. However, further studies assembling telomere-to-telomere sex chromosomes of snake species may help confirm such features and reveal the evolutionary history of the sex-determination system in caenophidian snakes.

Venom-gene Haplotypes Revealed a Large SV in the SVMP Array

To determine whether venom genes were affected by SVs in the heterozygous state for our genome animal, we compared these regions in the two haplotype assemblies (hap1 and hap2), and

Table 2 Numbers of paralogs within each toxin gene family in primary and haplotype genome assemblies of *C. adamanteus*

	Primary	Haplotype 1	Haplotype 2
BPP	1	1	1
CRISP	1	1	1
CTL	10	10	10
HYAL	1	1	1
KUN	2	2	2
LAAO	2	2	2
MYO	4	4	4
NGF	1	1	1
NUC	1	1	1
PDE	1	1	1
PLA2	3	3	3
PLB	1	1	1
SVMP	23	17	23
SVSP	24	16	20
VEGF	1	1	1

Abbreviations: BPP, bradykinin-potentiating peptide; CRISP, cysteine-rich secretory protein; CTL, C-type lectin; HYAL, hyaluronidase; KUN, Kunitz-type protease inhibitor; LAAO, L-amino acid oxidase; MYO, myotoxin/crotamine; NGF, nerve growth factor; NUC, nucleotidase; PDE, phosphodiesterase; PLA2, phospholipase A_2; PLB, phospholipase B; SVMP, snake venom metalloproteinase; SVSP, snake venom serine protease.

in the primary assembly (Table 2; supplementary table S1, Supplementary Material online). We detected no differences for single-paralog toxin genes between haplotypes (Table 2). Among the multiparalog toxin-gene arrays, the PLA2, CTL, and MYO arrays were consistent across haplotypes (Table 2; supplementary fig. S6, Supplementary Material online), whereas the SVSP and SVMP arrays showed apparent gene-content differences. To assess whether the latter two putative SVs were real or assembly artifacts, we evaluated read alignments and examined relationships among paralogs.

The differences in paralog content between assemblies in the SVSP gene array (Table 2) were likely due to assembly artifacts. The primary assembly had 23 SVSP paralogs, whereas hap1 and hap2 had 14 and 19 paralogs, respectively. Further inspection revealed that this region contained signal for being error-prone according to VerityMap and showed collapsed reads in the breakpoints defining the differences (supplementary figs. S10 and S11, Supplementary Material online). Moreover, the paralog phylogeny showed that variable SVSP paralogs were identical to paralogs shared by all assemblies, indicating either recent duplications or assembly artifacts (supplementary fig. S12, Supplementary Material online). Because we could not reject assembly artifacts as the basis for the assembly differences in this region, we did not pursue this potential SV further.

We identified a large deletion in the SVMP array in hap1 relative to hap2 and the primary assembly (Fig. 3a), which comprised ~225 Kb and six SVMP paralogs (*SVMP-11-mdc*, *SVMP-12-mdc*, *SVMP-13-mdc*, *SVMP-14-mad*, *SVMP-15-mad*, and *SVMP-16-mdc*). VerityMap confirmed that the SVMP loci in all assemblies did not occur in error-prone regions (supplementary fig. S7, Supplementary Material online). Read coverage for all assemblies corroborated the VerityMap output by presenting a similar pattern along the SVMP region in the primary assembly and both haplotype assemblies. We found no evidence for collapsed regions in the SVMP array for all assemblies (supplementary fig. S8, Supplementary Material online). Moreover, the SVMP paralog phylogeny revealed that none of the SVMP paralogs were identical copies (supplementary fig. S9, Supplementary Material online), providing additional evidence in support of the heterozygous SV in this region. Our data therefore provided a robust characterization of a six-paralog deletion affecting the SVMP toxin array for one of the haplotypes of our genome individual.

We identified conserved LINE elements at both boundaries of the SVMP deletion, as well as one LTR/Gypsy element surrounded by these LINE elements at one end (Fig. 3b). The recent genome-wide burst of LINE and LTR element activity in *C. adamanteus* (supplementary fig. S1, Supplementary Material online) suggests an active role for these elements in shaping the genome of this species in addition to a specific role in the evolution of the SVMP array. TEs have been hypothesized to be responsible for rearrangements, such as duplications, deletions, and gene fusions of SVMP genes in other *Crotalus* species (Giorgianni et al. 2020).

The SVMP Deletion is Rare but Regionally Abundant

To assess the geographic distribution of the SVMP deletion, we used anchored hybrid enrichment sequencing data designed to capture the exons of toxin and nontoxin genes (Margres et al. 2019) from 139 *C. adamanteus* individuals from across the species' range (Fig. 3b and Fig. 4). On the basis of mapping coverage, we were able to categorize each individual as homozygous for the full SVMP array (HomoRef), heterozygous (Het), or homozygous for the deletion (HomoDel). We detected 123 HomoRef, 5 Het, and 11 HomoDel individuals, indicating that the SVMP deletion haplotype was rare across the entire range. HomoDel and Het individuals were primarily restricted to the northwestern edge of the species' distribution and were the dominant genotypes in Mississippi (Fig. 4b).

We observed anomalously high coverage in the last exons of the *SVMP-12-mdc* gene in HomoDel individuals despite this gene being within the deletion (Fig. 3b). These peaks resulted from multimapped reads due to a high similarity between *SVMP-12-mdc* and nondeleted paralogs *SVMP-17-mdc* and *SVMP-19-mdc* (supplementary fig. S9A, Supplementary Material online). *SVMP-12-mdc* shows 92.9% overall identity with *SVMP-17-mdc* and 91.5% identity with *SVMP-19-mdc*. In particular, the region of *SVMP-12-mdc* with reads mapping (the last five exons) shows the highest identity with *SVMP-17-mdc* (97.5%; supplementary fig. S9B, Supplementary Material online). Higher-than-average-coverage peaks in this region of the *SVMP-12-mdc* gene are also present in Het and HomoRef individuals. The peaks observed in last exons of *SVMP-12-mdc* of HomoDel individuals comprised ~95% multimapped reads, whereas the average of multimapped reads for this gene in HomoRef and Het individuals was ~80%. These anomalous peaks therefore represent read-mapping artifacts due to recently derived SVMP paralogs and conserved exons.

The SVMP Deletion Affects the Venom Phenotype

To assess the potential for phenotypic effects of the deletion, we analyzed the venom-gland transcriptomes of 18 individuals (Fig. 3c) to estimate typical expression levels for the deleted genes in wild-type individuals. On the basis of their sampling localities (supplementary table S4, Supplementary Material online) and their high expression of the deleted paralogs, these 18 individuals should all be HomoRef or Het genotypes. Two SVMP paralogs within the deletion (*SVMP-11-mdc* and

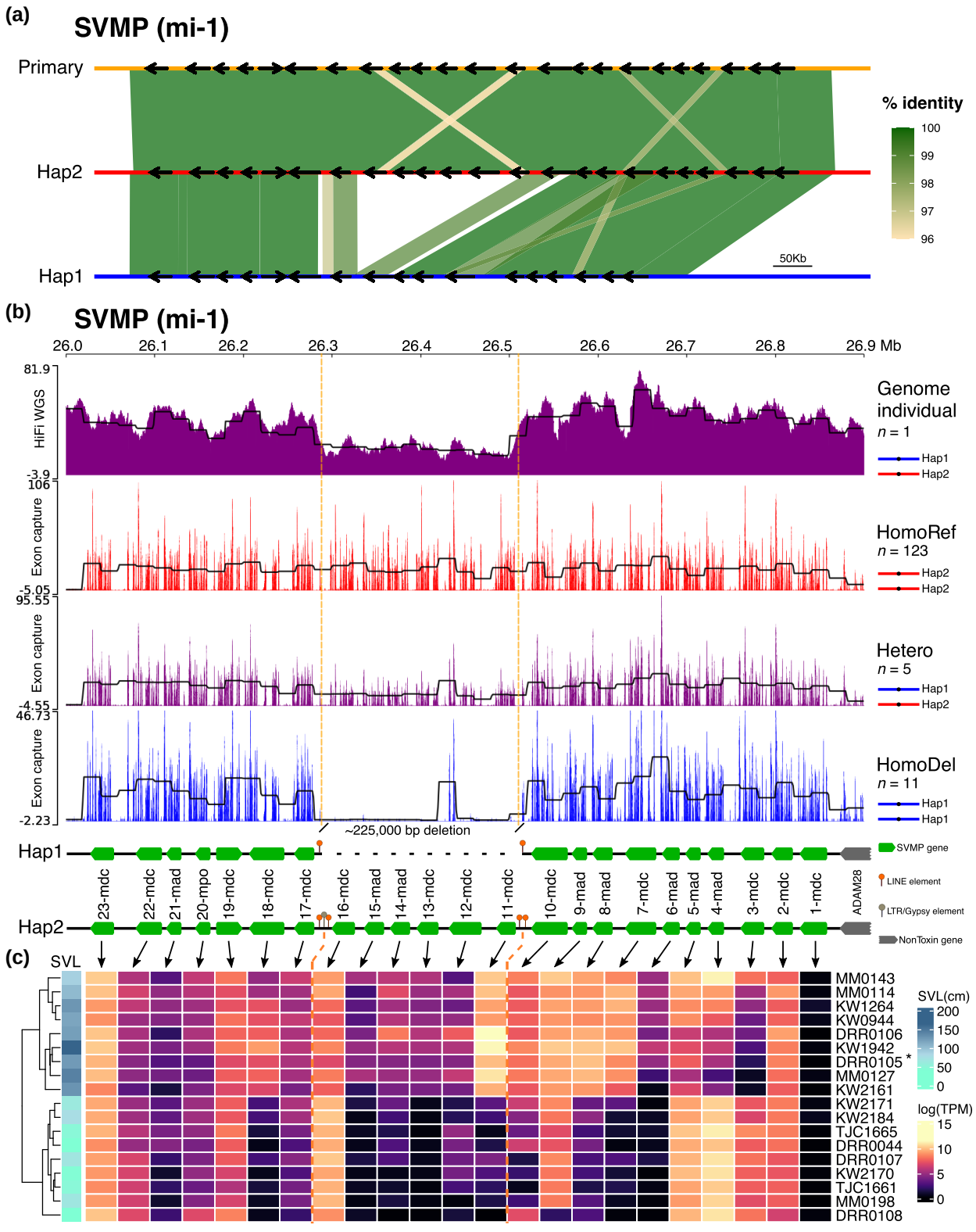

Fig. 3. Genomic alignment, genomic read coverages, and expression levels of SVMPs. a) Riparian plot showing the genomic alignment of SVMP loci between the primary assembly (orange) and each haplotype (red and blue). Black arrows represent each SVMP gene in that specific assembly. The shaded areas represent the percentage identity of alignments obtained through BLAST; alignments were filtered to sizes >10 Kb and percentage identity >95%. b) SVMP coverage in the genome individual (heterozygote), using the HiFi whole-genome data, and the other 139 individuals, using exon-capture sequencing data. Example coverage tracks of five individuals each of HomoRef (homozygotes for the entire SVMP array), Heterozygotes, and HomoDel (homozygotes for the SVMP deletion) are displayed in red, purple, and blue, respectively. Black lines represent average coverage across all individuals in that class in 20-Kb sliding windows. The vertical, dashed orange lines indicate the SVMP deletion region. At the bottom, the alignment of both haplotypes shows the ~225 Kb deletion in hap1 and the transposable elements (TE) flanking the deleted region. Only TEs at the deletion boundaries are shown. c) SVMP expression levels in the venom-gland transcriptomes of nine adults (snout-to-vent length >100 cm) and nine juveniles (snout-to-vent length <100 cm) show that the deleted paralogs include ontogenetically regulated genes. The genome individual (DRR0105) is indicated with an asterisk. Abbreviations: SVL, snout-to-vent length; TPM, transcripts per million.

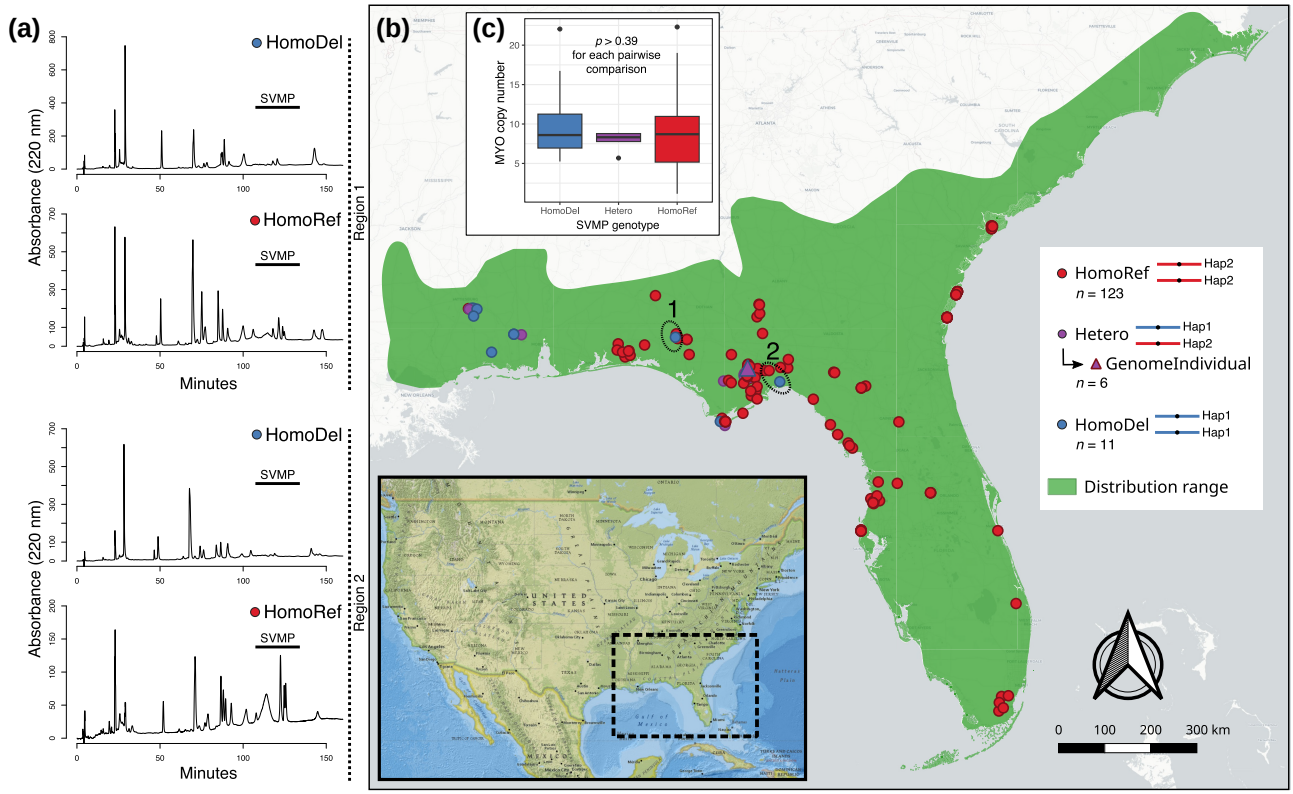

Fig. 4. The geographic distribution of the SVMP deletion and the phenotypic differences between homozygotes of each genotype. a) Reversed-phase high-performance liquid chromatography (RP-HPLC) of individual venoms from animals confirmed to be homozygous for each SVMP genotype from two different regions demonstrate clear differences in SVMP content. Region 1 was Eglin Air Force Base, and region 2 was the Apalachicola National Forest. b) Sampling distribution of specimens relative to the overall species range. Red dots represent individuals that were confirmed as homozygous for the wild-type SVMP region (HomoRef) on the basis of exon-capture data. Purple circles represent heterozygous (Hetero) individuals. Blue circles represent individuals that were homozygous for the SVMP deletion (HomoDel). The purple triangle represents the genome individual, which was a heterozygote for the deletion. c) Estimated copy number of MYO genes in individuals genotyped for the SVMP deletion. We found no statistically significant differences between groups using the Wilcoxon rank sum test.

SVMP-12-mdc) were highly expressed in adults (snout-to-vent length >100 cm; Waldron et al. 2013), and one paralog (*SVMP-16-mdc*) was highly expressed in juveniles (snout-to-vent length <100 cm; Waldron et al. 2013). Previous work showed that *SVMP-11-mdc* and *SVMP-12-mdc* were up-regulated in adults and that *SVMP-16-mdc* was up-regulated in juveniles (Hogan et al. 2024). Not only are some of the SVMP paralogs highly expressed, but some are involved in the fine-tuning of venom composition across life history. Individuals heterozygous or homozygous for the deletion therefore potentially show unique venom phenotypes for both adults and juveniles.

To directly measure the effects of the SVMP deletion on venom composition, we performed reverse-phase high-performance liquid chromatography (RP-HPLC) and mass spectrometry (MS) on venoms from genotyped individuals. We first visually compared representative homozygotes of each genotype from the same populations by means of RP-HPLC (Fig. 4a) to show that HomoRef and HomoDel individuals have striking differences in the SVMP peak region (i.e. peaks eluting at ~120 min; Margres et al. 2014). We then performed MS analyses on adult venoms from five HomoRef, one Het, and four HomoDel individuals. We detected unique proteomic signal for 17 SVMP paralogs, including four paralogs within the deletion (Fig. 5). For many of these paralogs, however, the signal was at a low background level (i.e. ≤2 Exclusive Unique Spectra Counts), leaving their presence in the venom unconfirmed. The two most highly expressed deletion paralogs identified in the transcriptomic data above (*SVMP-11-mdc* and *SVMP-16-mdc*) were, as expected, abundant in the venoms of HomoRef and Het individuals, but not HomoDel individuals; we found no proteomic evidence for any of the four detected deletion paralogs in the venoms from HomoDel individuals (supplementary fig. S13, Supplementary Material online).

Independence between the SVMP Deletion and MYO Copy Number

Previous work in *C. adamanteus* (Margres et al. 2017a) described substantial variation in copy number for the MYO gene with corresponding dramatic effects on venom composition. The general trend was that MYO was at low copy number or absent from the genomes of individuals in the southern portion of the range and at high copy number in the north of the range, superficially complementing our pattern for the SVMP gene array (Fig. 4). To test for a correlation between SVMP and MYO genotypes, we estimated MYO copy number from the hybrid-enrichment data for the same individuals genotyped for the SVMP deletion. We found no statistically significant relationship between MYO copy number and SVMP genotype (Fig. 4c and supplementary S14, Supplementary Material online). The comparison between homozygotes returned a *P*-value of 0.39, whereas comparisons between the heterozygotes and homozygotes for the presence and deletion of the SVMP paralogs resulted in *P*-values of 0.70 and 0.59,

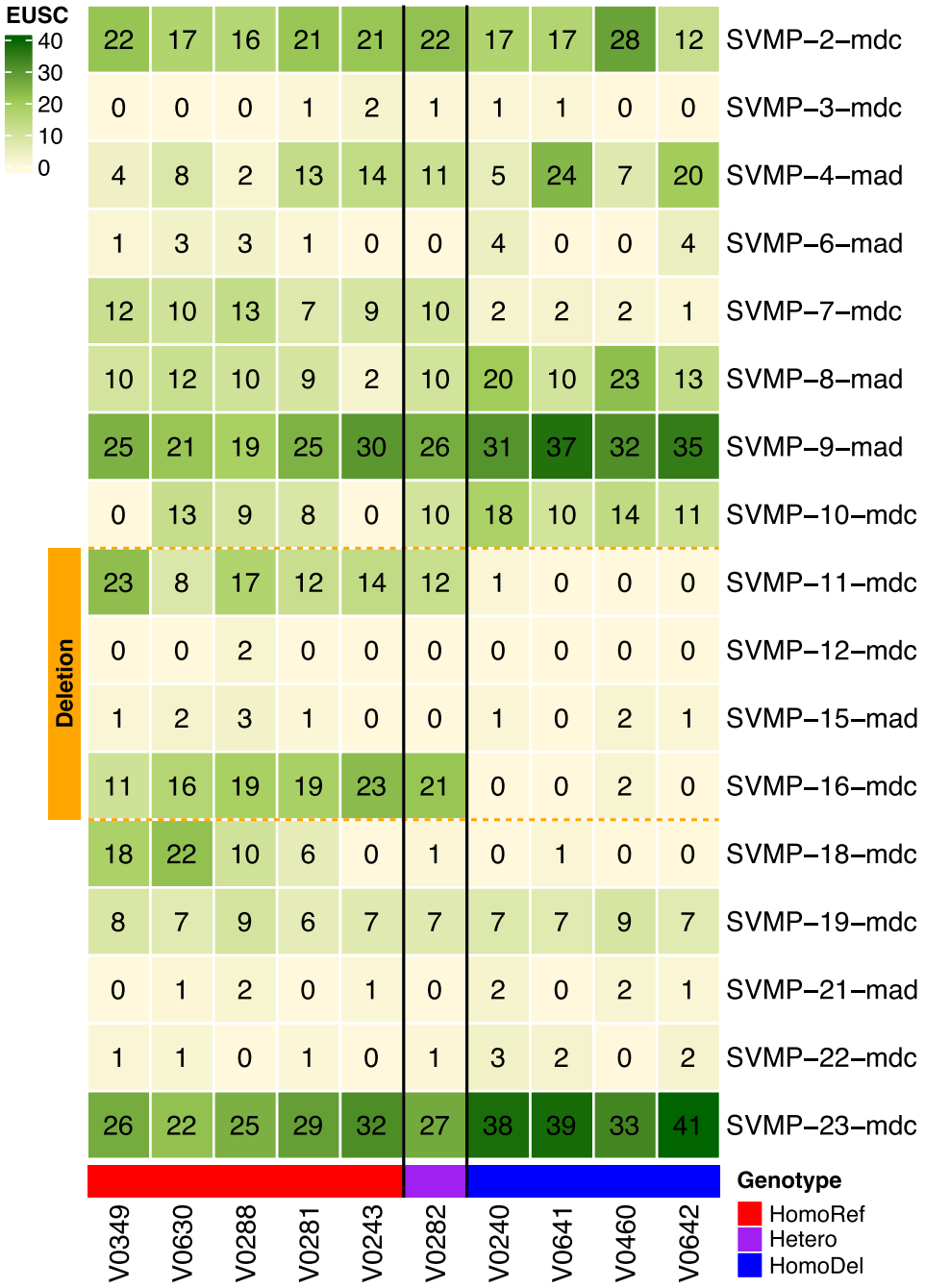

Fig. 5. SVMP proteomic abundances from venoms of genotyped individuals. The heatmap shows the normalized Exclusive Unique Spectra Count (EUSC) for SVMPs with at least one unique peptide matching in at least one sample. Deleted SVMPs are represented with orange colored squares on the left. The genotype of each sample is at the bottom, where red indicates homozygotes for the entire wild-type SVMP array (HomoRef), purple indicates heteorozygotes (Hetero), and blue indicates homozygotes for the deletion (HomoDel). At least two of the six deleted SVMPs have major effects on venom proteomic composition.

respectively. The SVMP deletion distribution showed a clear increase in frequency toward the western periphery of the range, with no detected occurrences east of the Suwannee River (supplementary fig. S14, Supplementary Material online), and we confirmed the complementary pattern of MYO copy-number increasing from south to north (supplementary fig. S14, Supplementary Material online). However, rather than finding evidence for linkage disequilibrium between the regions, we find apparent independence, suggesting that the prevalence of the SVMP deletion in the northwestern region of the range (Mississippi) is not contingent on the presence of high-copy-number MYO haplotypes. Furthermore, this lack of association indicates that a large component of geographic variation in the venom phenotype of *C. adamanteus* was generated by two independent gradients in the frequencies of SV-based haplotypes on two different chromosomes.

Discussion

Biases in mutational processes can be as determinative of evolutionary outcomes as the selective pressure acting on the resulting mutations (Rokyta et al. 2005; Sackman et al. 2017;

Stoltzfus and McCandlish 2017; Cano et al. 2023). Observations of the genetic bases of beneficial phenotypes can provide critical information about whether certain types of mutations are more likely to contribute to adaptation by virtue of some combination of their rates of occurrence and phenotypic consequences. Because the genes encoding venom components are readily identifiable, we know that positive selection is rampant within these coding sequences (Lynch 2007; Gibbs and Rossiter 2008; Rokyta et al. 2011). Expression differences, however, are also widespread in venoms and may be the quickest characteristics to evolve (Margres et al. 2016a, 2017a), possibly because they can be accomplished by means of multiple mutational mechanisms, including gene duplication and deletion and changes to multiple *cis*-regulatory regions. Only as high-contiguity genome assemblies become available for venomous species are we able to begin to comprehensively assess the full spectrum of mutational types, and SVs are emerging as a major component of intraspecific variation for many species, including *C. adamanteus*. Structural variants represent a large proportion of genetic variants affecting phenotypes in eukaryotes (Ho et al. 2020), but their relative contributions to trait evolution are largely unknown. Starting with a single genome assembly of sufficient quality to identify large SVs in the heterozygous state, we uncovered a new SV with a major impact on venom variation for *C. adamanteus*. Although the majority of SVs are likely to be deleterious due to their impacts on gene expression and recombination rates, they are also known to contribute to adaptation (Tigano et al. 2018; Vickrey et al. 2018; Catanach et al. 2019; Faria et al. 2019; Todesco et al. 2020; Weissensteiner et al. 2020; Hämälä et al. 2021; Hager et al. 2022; Saitou et al. 2022; Shi et al. 2023; Li et al. 2024b). We showed that a large SV affecting six toxin genes is locally dominant in a population of *C. adamanteus*, which, in conjunction with parallel occurrences of similar deletions in other rattlesnake species, suggests it confers a local fitness advantage. We also demonstrated that multimapped short reads can potentially confound coverage-based genotyping for SVs involving genes within tandem duplicate arrays (like those containing many, if not most, venom genes). Shared, conserved exons among paralogs can result in spurious signal for the presence of deleted paralogs. For our data, the spurious signal was limited to a small minority of exons in a single-deleted paralog and was therefore straightforward to exclude as an alignment artifact.

Parallel evolution (Ventura et al. 2011; Pearse et al. 2014; Bohutínská et al. 2021) occurs when similar phenotypes evolve independently in two distinct lineages in response to similar selective pressures. The large SVMP deletion we identified in *C. adamanteus* parallels similar deletions in this same genomic region in multiple other *Crotalus* species (supplementary figs. S15 and S16, Supplementary Material online; Giorgianni et al. 2020; Margres et al. 2021). The SVMP deletion of *C. adamanteus*, however, was not correlated with the acquisition of neurotoxic PLA2s as in every other reported case (Dowell et al. 2016, 2018; Margres et al. 2021). A dichotomy in venom types for rattlesnakes has long been hypothesized, with type I venoms showing high metalloproteinase activity and low neurotoxic activity and type II venoms showing the inverse (Mackessy 2010). Individuals of *C. adamanteus* with the SVMP deletion appear to depart from this pattern (Fig. 4a).

MYO expression also shows parallel patterns of extreme expression variation within other *Crotalus* species, including *C. scutulatus* (Strickland et al. 2018) and *C. viridis* (Smith et al. 2023). Remarkably, for *C. viridis*, MYO expression is almost perfectly inversely correlated with SVMP expression levels (Smith et al. 2023), although the genetic basis for this phenotypic variation is not known. This pattern suggests that MYO activity could replace neurotoxic PLA2 activity in some scenarios (Smith et al. 2023). In contrast to the results from *C. viridis*, however, we found no statistical association between the genotypes underlying high MYO expression and low SVMP expression. This suggests a more complex pattern of geographic variation in selective pressures as well as a remarkable mosaic of SV-based phenotypes in these venoms that does not neatly partition into a distinct dichotomy.

The geographic regions with the highest density of the SVMP deletion (the MS population in supplementary fig. S14A, Supplementary Material online) also showed among the highest average MYO copy numbers, yet we found no evidence for an association between these SVs. These loci appear to be evolving independently (Fig. 4c). The SVMP deletion was primarily detected in the western periphery of the species' range and gradually declined in frequency eastward, becoming undetectable east of the Aucilla River in our sampling. MYO SVs had no well-defined west-to-east gradient, but instead decreased toward the southern edge of the range (supplementary fig. S14, Supplementary Material online). Previous work in *C. adamanteus* using neutral data from these same individuals (Fig. 4b; Margres et al. 2019) identified three genetically-distinct populations: one predominantly east and south of the Suwannee River, one predominantly west of the Suwannee River, and a distinct island population. MYO SVs were present in both western and eastern populations, with deletions being more frequent in the eastern population south of the Suwannee River (Margres et al. 2015b, 2017a). Given that the SVMP deletion was unique to the western population, our current sampling indicated that the Suwannee River may also be a phylogeographic barrier for the SVMP deletion; we do note that our current sampling does not indicate the presence of the SVMP deletion near the river (~100 km west), and dense sampling along the contact zone would be needed to determine if the Suwannee River is a barrier to both neutral and putatively adaptive alleles. Nevertheless, the Suwannee River is thought to be a suture zone for numerous Florida lineages (Bert 1986), including *C. adamanteus* (Margres et al. 2015b, 2019). The SV in the SVMP array is a single large deletion that likely originated a single time in the western population, whereas the SVs involving the MYO genes comprise several deletion and/or duplication events that may have originated multiple times independently across the range, predate the phylogeographic split at the Suwannee River, and/or be shared through gene flow across the river (i.e. leaky barrier). The distinct genomic architectures of these loci may also partially explain their independent segregation and different geographic patterns. In contrast to SVMP loci (Fig. 2), the MYO loci are located in a genomic region of ma-2 that appears to be predisposed to intergenic rearrangements (Fig. 2) with a high density of TEs, which may promote copy-number variation. This region also harbors other gene families known to be highly duplicated, such as chemosensory and immunoglobulin genes (Hogan et al. 2021). Where examined in other rattlesnake species, MYO expression levels have nearly always been found to be highly variable within species (Bober et al. 1988; Oguiura et al. 2009; Gopalan et al. 2022), providing further evidence that this genomic region may experience

higher rates of SV generation than others. We found that, in general, heterozygous SVs in the *C. adamanteus* genome occurred primarily in intergenic regions and were associated with enrichment of repetitive sequences and TEs. We also detected TEs flanking the region deleted in the SVMP tandem array, corroborating previous findings suggesting that TEs can directly affect snake venom variation (Dowell et al. 2016; Giorgianni et al. 2020; Perry et al. 2022). Structural variants differentiating haplotypes have been observed in numerous genomes from a diversity of eukaryotic species (Armstrong et al. 2022; Toh et al. 2022; Barros et al. 2023; Chang et al. 2023; Han et al. 2023; Qi et al. 2023; Zhao et al. 2023a), and these SVs were also found to be enriched in intergenic and repetitive regions (Casacuberta and González 2013; Serrato-Capuchina and Matute 2018).

Our detection of the SVMP deletion was contingent on fortuitously selecting a rare genotype in the region from which we sampled our genome animal. Given that this discovery resulted from a sample size of one (or two if counting haplotypes), additional genome sequencing from throughout the range is likely to reveal substantial genetic novelty that may affect management decisions for this species. *C. adamanteus* is a charismatic and emblematic species of the southeastern US coastal plain with a range substantially diminished from its historical extent as a result of human persecution and habitat loss (Means 2017). Given the numerous examples of large-scale venom-related expression differences within and between snake species with unknown genetic origins, including for *C. adamanteus* (Margres et al. 2016b, 2017b; Smith et al. 2023), we expect SVs to be a major source of this phenotypic variation and a critical component of functional genetic variation within species. These patterns of functional genetic variation should be incorporated into conservation and species-management decision-making processes, such as those currently ongoing for *C. adamanteus*. Previous work in *C. adamanteus* using neutral data from these same individuals (Fig. 4b; Margres et al. 2019) failed to detect any neutral population structure west of the Suwannee River. We found a major segregating SV in this region with a pronounced longitudinal frequency gradient, highlighting how the inclusion of functional genetic variation can refine our view of optimal conservation strategies. Not only does the western periphery of the range harbor a genetically and phenotypically unique population of *C. adamanteus*, the presence of this unique phenotype may indicate a novel ecology for the species in this region. High-contiguity genomes from multiple individuals from throughout the range will be necessary for a comprehensive accounting of all forms of genetic variation. Genomic approaches have long been used to identify cryptic species (Hinojosa et al. 2019; Christmas et al. 2021); high-accuracy, long-read sequencing technologies are now facilitating the identification of previously cryptic forms of genetic variation within species with implications for both species management as well as fundamental evolutionary processes.

Materials and Methods

Genome Sequencing

The PacBio HiFi data were described previously (Hogan et al. 2024). Briefly, data were generated using genomic DNA from an adult female (DRR0105). The resulting data comprised 3,910,111 reads with an average read length of 15.0 Kb and a total of 58,702,723,931 bp (>36× coverage). We used cutadapt version 4.1 (Martin 2011) to remove reads containing adapters and kraken2 version 2.1.2 (Wood et al. 2019) to remove human or bacterial contaminants.

A blood sample from the same individual (DRR0105) was used to construct an Hi-C library, following the protocol for nucleated blood cells for the Arima High Coverage Hi-C Kit (Arima Genomics) to crosslink DNA and generate the proximity-ligated DNA. We then constructed the final Hi-C library using the proximity-ligated DNA and the Arima High Coverage Hi-C Library Preparation Kit (Arima Genomics) following the manufacturer's instructions. The Hi-C library was sequenced using the NovaSeq 6,000 platform (Illumina) with paired-end reads layout (2×150 bp) at the Florida State University College of Medicine Translational Science Laboratory. Sequencing yielded 269,208,645 paired-end reads. We used trim_galore! to trim adapters and to remove low-quality reads (–q 25) and reads shorter than 75 bp (–length 75), which returned a final dataset consisting of 265,684,197 paired-end reads (total of 79,218,703,363 bp; >49× coverage).

Genome Assembly

The HiFi long reads and paired-end Hi-C short reads were provided to hifiasm version 0.16.1 (Cheng et al. 2021, 2022) to generate the primary and paired haplotype-resolved assembly contig graphs with default parameters. The contigs of the primary assembly were used as reference to map the Hi-C reads using Chromap v0.2.3 (Zhang et al. 2021a) and to scaffold using YaHS version 1.2a.2 (Zhou et al. 2023b). We then used Juicer version 1.6 (Durand et al. 2016) to manually review the scaffolded genome following the standard DNA Genome Assembly Cookbook instructions (https://aidenlab.org/assembly/manual_180322.pdf). The haplotype-resolved assemblies were generated using RagTag version 2.1.0 (Alonge et al. 2022) with each haplotype contig as a query and the primary chromosome-level assembly as a reference. Genome assembly statistics, for primary and both haplotypes, were obtained using Inspector version 1.0.1 (Chen et al. 2021), which also calculates a QV score to measure putative errors in the assembly, and the completeness was assessed using BUSCO version 5.2.2 (Waterhouse et al. 2018) with the Tetrapoda gene set (odb10; total of 5,310 genes). We also verified assembly quality using VerityMap version 1.0 (Bzikadze et al. 2022), which allows accurate mapping of long-reads to determine details about heterozygous and error-prone assembled regions. We characterized the identity of chromosomes performing BLAST searches against a set of chromosome-specific markers (NCBI accessions SAMN00177542 and SAMN00152474) of snakes (Matsubara et al. 2006) and the chromosome-level assembly of *C. viridis* (Schield et al. 2019). We also confirmed identities of sex chromosomes based on male–female read coverage ratio mapping whole-genome sequencing data of male and female individuals as previously described (Hogan et al. 2021). The mitochondrial genome was assembled using the long-read mode of MITGARD version 1.1 (Nachtigall et al. 2021a) with the HiFi reads and the *C. adamanteus* mitogenome as reference (NC_041524.1). The assembled mitogenome was annotated using MitoZ version 3.6 (Meng et al. 2019), followed by manual verification.

Genome Annotation

We performed genome annotation on the primary chromosome-level assembly. We annotated repetitive regions

and TEs using RepeatModeler2 and RepeatMasker. We used the RepeatModeler2 version 2.0.1 (Flynn et al. 2020) to generate a *de novo* species-specific repetitive-sequence and TE library. We split the library into "known" and "unknown" sets as output by RepeatModeler2. The "unknown" set was classified using DeepTE version 1.0 (Yan et al. 2020) with the model designed for metazoans. To remove false-positive repetitive elements, we removed any sequence classified as "NonTE" using TERL version 1.0 (da Cruz et al. 2021). Then, the species-specific TE library (i.e. the "known" set and the "unknown" re-classified set) was merged to a curated TE library designed for snakes (Castoe et al. 2013), and the final TE library was used to perform the repetitive annotation using RepeatMasker version 4.1.1 (https://www.repeatmasker.org/). The divergence between the individual TE copies versus their consensus sequences based on CpG-adjusted Kimura distance was estimated using RepeatMasker built-in scripts. We searched for telomeric sequences at chromosome terminals using tidk-search version 0.2.0 (Brown et al. 2023) using the conserved vertebrate telomeric repeat sequence TTAGGG.

Gene annotation was performed using the funannotate pipeline (Palmer and Stajich 2017), which integrates several *ab initio* gene predictors (i.e. AUGUSTUS, SNAP, and GeneMark-ES) to build gene models and uses transcript and protein evidence to generate the final annotation set. We used available transcriptomic data from several tissues from males and females of *C. adamanteus* (supplementary table S4, Supplementary Material online) as transcript evidence and the protein sequences available for the Tetrapoda clade in Uniprot and NCBI databases as protein evidence. We also performed the functional annotation step using InterProScan5 version 5.54 (Jones et al. 2014). Due to the high number of proteins predicted by funannotate, we compared the predicted proteins to the proteins annotated in the genomes of mouse, chicken, green anole, Central bearded dragon, Komodo dragon, Common wall lizard, Mainland tiger snake, and Eastern brown snake available in the ENSEMBL database (downloaded August 2023) using DIAMOND version 2.1.9 (Buchfink et al. 2021) with high stringency (parameters set to be more sensitive, minimum coverage of 50%, and *e*-value <0.001) to confirm conserved and confident predicted proteins. To annotate toxins, we used ToxCodAn-Genome version 1.0 (Nachtigall et al. 2024) with default parameters and followed their guide to ensure a confident toxin annotation set (Nachtigall 2023). Briefly, the genome individual venom-gland transcriptomic data were assembled and annotated using ToxCodAn version 1.0 (Nachtigall et al. 2021b) with default parameters to generate a species-specific toxin database. The species-specific and the Viperidae toxin databases were used as database sources to annotate the toxins in the genome using ToxCodAn-Genome version 1.0 (Nachtigall et al. 2024). We then merged the toxin and nontoxin annotations to generate a final annotation set. The final annotation set obtained in the primary assembly was lifted to the haplotype-resolved assemblies using liftoff version 1.6.3 (Shumate and Salzberg 2021).

Comparative Analysis

We compared haplotypes at the sequence level using syri version 1.6.3 (Goel et al. 2019) and at the gene level using genespace version 1.3.0 (Lovell et al. 2022). We also compared the assemblies on a smaller scale by aligning the genomic region containing the multiloci toxin families (i.e. SVMP, SVSP, PLA2, MYO, and CTL) to check for differences in these regions between haplotypes and primary assemblies. Genomic alignments were performed using BLAST (blastn) with an identity percentage threshold set to 95%. We then filtered results to keep alignments >10 Kb and plotted the alignments using ggplot2 in R.

To ensure that putative SVs detected between haplotypes were not assembly artifacts, we mapped the HiFi reads against each of the assemblies (i.e. primary and both haplotypes) using VerityMap, Inspector, and minimap2. We then analyzed read coverage and the VerityMap output in the toxin genomic regions of interest to check for error-prone regions. VerityMap maps long reads using a k-mer method and enables checking for possible errors and heterozygous sites in the assembly on the basis of the proportion of rare k-mers. To check whether multimapped reads could be influencing detection of assembly artifacts in those regions, we filtered the minimap2 output to only keep uniquely mapped and high-quality alignments by removing reads with MAPQ <30 using samtools. We also checked for collapsed regions in the highly duplicated toxin genomic regions using NucFreq version 0.1 (Vollger et al. 2019) and the mapping files output by VerityMap, Inspector, and minimap2. NucFreq checks for collapsed regions that may indicate assembly artifacts occurring in error-prone regions. We performed this additional step because analyzing the mapping status of the original reads along a genome assembly allows assessment of the overall assembly quality and can reveal putative assembly artifacts and error-prone regions (Li et al. 2023).

We performed a phylogenetic inference for the SVSP and SVMP toxin genes to better understand their relationships. We aligned their coding sequences using MAFFT version 7.450 (Rozewicki et al. 2019) with default parameters and searched for the maximum likelihood tree using IQTree version 1.6.12 (Nguyen et al. 2015) with parameters `– m TEST –bb 1000 –alrt 1000`. The final tree was adjusted using FigTree version 1.4.4 (https://github.com/rambaut/figtree/).

Venom-gland Transcriptomic Analysis

We used previously published venom-gland transcriptomic data (supplementary table S4, Supplementary Material online) for 18 individuals of *C. adamanteus* (Hogan et al. 2024). Adapters and low-quality reads were removed using trim_galore! as previously described. Expression levels of annotated coding sequences were estimated using RSEM version 1.3.1 (Li and Dewey 2011) using Bowtie2 version 2.4.2 (Langmead and Salzberg 2012) as the aligner with the mismatch rate parameter set to 0.02.

Exon-capture Data Analysis

We used a set of anchored data available for 139 individuals of *C. adamanteus* designed to sequence the exon of toxin genes and other probes as previously described (Margres et al. 2017a, 2019). These data comprise individuals sampled from throughout the species distribution and contain representatives of most *C. adamanteus* populations. Adapters and low-quality reads were trimmed using trim_galore! as previously described. The trimmed reads were mapped against the primary assembly using Bowtie2. We removed PCR duplicates and mapped reads with MAPQ <30 before calculating the average coverage of each SVMP paralog to genotype each individual as homozygous or heterozygous for the SVMP deletion observed in the haplotype-revolved

assemblies. Specifically, we calculated the average of coverage for each SVMP gene in the SVMP array. Then, the individuals were genotyped as follows: (i) homozygous for the entire SVMP array (HomoRef), when all genes were presenting a similar average of coverage; (ii) heterozygous for the SVMP deletion (Het), when genes in the SVMP deletion were presenting half of average of coverage when compared with the SVMP genes not located in the SVMP deletion; and (iii) homozygous for the SVMP deletion (HomoDel), when genes in the SVMP deletion presented an average of coverage <10% of the other genes in the SVMP array.

To estimate the copy number of MYO genes in each sample, we mapped reads as described above, but the multimapped reads were kept due to the high similarity of MYO genes (i.e. we did not remove mapped reads with MAPQ <30). We then used the coverage of exon 2 and exon 3 from MYO genes to calculate the average of coverage, as previously performed (Margres et al. 2017a), and compared it with the average coverage of 10 nontoxin genes available in the probe set and located on the macrochromosome 2 as well (i.e. ATPSynLipid-1, ATPase-lys70, CD63, Calreticulin, DAZ-2, GADD45, Glutaredoxin-1, Leptin-1, PDI, and Nexin-2).

Venom Reversed-Phase High-Performance Liquid Chromatography

To visualize the compositional effects of the SVMP deletion, we performed reversed-phase high-performance liquid chromatography (RP-HPLC) for two individuals genotyped as homozygotes for each haplotype (i.e. two individuals homozygous for the complete SVMP array and two individuals homozygous for the six-paralog SVMP deletion). We performed RP-HPLC analysis on pairs of individuals collected in close geographic proximity. RP-HPLC was performed and analyzed as previously described (Margres et al. 2015a).

Venom Mass Spectrometry

To generate a genotype–phenotype map and verify the toxin expression proteomically, we performed mass spectrometry on whole venom samples from genotyped individuals for the SVMP deletion. Proteomics data were generated and analyzed following (Hofmann et al. 2018). See Supplementary Material online for details.

Permits and Protocols

The specimen used for genome sequencing was collected under the Florida USA permits LSSC-13-00004A, LSSC-13-00004B, and LSSC-13-00004C. All animal procedures were performed under active IACUC protocols: Florida State University protocols 0924, 1230, 1333, 1529, and 1836.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Funding

This work was supported by the National Science Foundation (NSF DEB 1638902 to D.R.R. and NSF DEB 1638879 to C.L.P.) and Fundação de Amparo à Pesquisa no Estado de São Paulo (FAPESP; grant nos. 2018/26520-4 and 2022/04988-0 to P.G.N. and 2013/07467-1 and 2016/50127-5 to I.L.M.J.-d.-A.). Additional support for this work was provided by the Clemson University Genomics and Bioinformatics Facility, which receives support from the College of Science and two Institutional Development Awards (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant nos. P20GM146584 and P20GM139769.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability

A list with all datasets used in the present study are available in supplementary table S4, Supplementary Material online. The genome assembly and the Hi-C data generated in the present study are available in the NCBI under the project number PRJNA868880. In addition, the assembled genome and annotations are available in the figshare database (https://figshare.com/projects/Eastern_diamondback_rattlesnake_Crotalus_adamanteus_-_haplotype-resolved_genome_assembly/200614). All code and commands used in this study are available on GitHub (https://github.com/pedronachtigall/Cadamanteus_SV).

References

Almeida DD, Viala VL, Nachtigall PG, Broe M, Gibbs HL, Serrano SMdT, Moura-da Silva AM, Ho PL, Nishiyama-Jr MY, Junqueira-de Azevedo IL. Tracking the recruitment and evolution of snake toxins using the evolutionary context provided by the *Bothrops jararaca* genome. *Proc Natl Acad Sci U S A*. 2021:118(20):e2015159118. https://doi.org/10.1073/pnas.201515911.

Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol*. 2022:23(1):258. https://doi.org/10.1186/s13059-022-02823-7.

Armstrong EE, Perry BW, Huang Y, Garimella KV, Jansen HT, Robbins CT, Tucker NR, Kelley JL. A beary good genome: haplotype-resolved, chromosome-level assembly of the brown bear (*Ursus arctos*). *Genome Biol Evol*. 2022:14(9):evac125. https://doi.org/10.1093/gbe/evac125.

Baker RJ, Mengden GA, Bull JJ. Karyotypic studies of thirty-eight species of North American snakes. *Copeia*. 1972:1972(2):257–265. https://doi.org/10.2307/1442486.

Barros CP, Derks MF, Mohr J, Wood BJ, Crooijmans RP, Megens HJ, Bink MC, Groenen MA. A new haplotype-resolved turkey genome to enable turkey genetics and genomics research. *Gigascience*. 2023:12(1):giad051. https://doi.org/10.1093/gigascience/giad051.

Bert T. Speciation in western Atlantic stone crabs (genus *Menippe*): the role of geological processes and climatic events in the formation and distribution of species. *Mar Biol*. 1986:93(2):157–170. https://doi.org/10.1007/BF00508253.

Bober MA, Glenn JL, Straight RC, Ownby CL. Detection of myotoxin a-like proteins in various snake venoms. *Toxicon*. 1988:26(7):665–673. https://doi.org/10.1016/0041-0101(88)90248-6.

Bohutínská M, Vlček J, Yair S, Laenen B, Konečná V, Fracassetti M, Slotte T, Kolář F. Genomic basis of parallel adaptation varies with divergence in *Arabidopsis* and its relatives. *Proc Natl Acad Sci U S A*. 2021:118(21):e2022713118. https://doi.org/10.1073/pnas.2022713118.

Brown M, González De la Rosa PM, Mark B. A telomere identification toolkit. https://github.com/tolkit/telomeric-identifier. 2023.

Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 2021:18(4):366–368. https://doi.org/10.1038/s41592-021-01101-x.

Bzikadze AV, Mikheenko A, Pevzner PA. Fast and accurate mapping of long reads to complete genome assemblies with VerityMap. *Genome Res*. 2022:32(3):2107–2118. https://doi.org/10.1101/gr.276871.122.

Cano AV, Gitschlag BL, Rozhoňová H, Stoltzfus A, McCandlish DM, Payne JL. Mutation bias and the predictability of evolution. *Philos Trans R Soc B*. 2023:378(1877):20220055. https://doi.org/10.1098/rstb.2022.0055.

Casacuberta E, González J. The impact of transposable elements in environmental adaptation. *Mol Ecol*. 2013:22(6):1503–1517. https://doi.org/10.1111/mec.12170.

Casewell NR, Jackson TN, Laustsen AH, Sunagar K. Causes and consequences of snake venom variation. *Trends Pharmacol Sci*. 2020:41(8):570–581. https://doi.org/10.1016/j.tips.2020.05.006.

Castoe TA, de Koning APJ, Hall KT, Card DC, Schield DR, Fujita MK, Ruggiero RP, Degner JF, Daza JM, Gu W, *et al*. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc Natl Acad of Sci USA*. 2013:110(51):20645–20650. https://doi.org/10.1073/pnas.1314475110.

Catanach A, Crowhurst R, Deng C, David C, Bernatchez L, Wellenreuther M. The genomic pool of standing structural variation outnumbers single nucleotide polymorphism by threefold in the marine teleost *Chrysophrys auratus*. *Mol Ecol*. 2019:28(6):1210–1223. https://doi.org/10.1111/mec.15051.

Chaisson MJ, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, Gardner EJ, Rodriguez OL, Guo L, Collins RL, *et al*. Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat Commun*. 2019:10(1):1784. https://doi.org/10.1038/s41467-018-08148-z.

Chang Y, Zhang R, Ma Y, Sun W. A haplotype-resolved genome assembly of *Rhododendron vialii* based on PacBio HiFi reads and Hi-C data. *Sci Data*. 2023:10(1):451. https://doi.org/10.1038/s41597-023-02362-1.

Chen Y, Zhang Y, Wang AY, Gao M, Chong Z. Accurate long-read *de novo* assembly evaluation with inspector. *Genome Biol*. 2021:22(1):312. https://doi.org/10.1186/s13059-021-02527-4.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with Hifiasm. *Nat Methods*. 2021:18(2):170–175. https://doi.org/10.1038/s41592-020-01056-5.

Cheng H, Jarvis ED, Fedrigo O, Koepfli KP, Urban L, Gemmell NJ, Li H. Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol*. 2022:40(9):1332–1335. https://doi.org/10.1038/s41587-022-01261-x.

Christmas MJ, Jones JC, Olsson A, Wallerman O, Bunikis I, Kierczak M, Peona V, Whitley KM, Larva T, Suh A, *et al*. Genetic barriers to historical gene flow between cryptic species of alpine bumblebees revealed by comparative population genomics. *Mol Biol Evol*. 2021:38(8):3126–3143. https://doi.org/10.1093/molbev/msab086.

da Cruz MHP, Domingues DS, Saito PTM, Paschoal AR, Bugatti PH. TERL: classification of transposable elements by convolutional neural networks. *Brief Bioinform*. 2021:22(3):bbaa185. https://doi.org/10.1093/bib/bbaa185.

Dowell NL, Giorgianni MW, Griffin S, Kassner VA, Selegue JE, Sáanchez EE, Carroll SB. Extremely divergent haplotypes in two toxin gene complexes encode alternative venom types within rattlesnake species. *Curr Biol*. 2018:28(7):1016–1026. https://doi.org/10.1016/j.cub.2018.02.031.

Dowell NL, Giorgianni MW, Kassner VA, Selegue JE, Sanchez EE, Carroll SB. The deep origin and recent loss of venom toxin genes in rattlesnakes. *Curr Biol*. 2016:26(18):2434–2445. https://doi.org/10.1016/j.cub.2016.07.038.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. 2016:3(1):95–98. https://doi.org/10.1016/j.cels.2016.07.002.

Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, Sulovari A, Ebler J, Zhou W, Serra Mari R, *et al*. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science*. 2021:372(6537):eabf7117. https://doi.org/10.1126/science.abf7117.

Faria R, Chaube P, Morales HE, Larsson T, Lemmon AR, Lemmon EM, Rafajlović M, Panova M, Ravinet M, Johannesson K, *et al*. Multiple chromosomal rearrangements in a hybrid zone between *Littorina saxatilis* ecotypes. *Mol Ecol*. 2019:28(6):1375–1393. https://doi.org/10.1111/mec.14972.

Fish US, Service W. Endangered and threatened wildlife and plants: 90-day finding on a petition to list the eastern diamondback rattlesnake as threatened. *Fed Regist*. 2012:77:27403–27411.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*. 2020:117(4):9451–9457. https://doi.org/10.1073/pnas.1921046117.

Garg S. Computational methods for chromosome-scale haplotype reconstruction. *Genome Biol*. 2021:22(1):101. https://doi.org/10.1186/s13059-021-02328-9.

Garg S. Towards routine chromosome-scale haplotype-resolved reconstruction in cancer genomics. *Nat Commun*. 2023:14(1):1358. https://doi.org/10.1038/s41467-023-36689-5.

Giani AM, Gallo GR, Gianfranceschi L, Formenti G. Long walk to genomics: history and current approaches to genome sequencing and assembly. *Comput Struct Biotechnol J*. 2020:18(Suppl 2):9–19. https://doi.org/10.1016/j.csbj.2019.11.002.

Gibbs HL, Rossiter W. Rapid evolution by positive selection and gene gain and loss: PLA2 venom genes in closely related *Sistrurus* rattlesnakes with divergent diets. *J Mol Evol*. 2008:66(2):151–166. https://doi.org/10.1007/s00239-008-9067-7.

Giorgianni MW, Dowell NL, Griffin S, Kassner VA, Selegue JE, Carroll SB. The origin and diversification of a novel protein family in venomous snakes. *Proc Natl Acad Sci U S A*. 2020:117(20):10911–10920. https://doi.org/10.1073/pnas.1920011117.

Goel M, Sun H, Jiao WB, Schneeberger K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol*. 2019:20(1):277. https://doi.org/10.1186/s13059-019-1911-0.

Gopalan SS, Perry BW, Schield DR, Smith CF, Mackessy SP, Castoe TA. Origins, genomic structure and copy number variation of snake venom myotoxins. *Toxicon*. 2022:216(9):92–106. https://doi.org/10.1016/j.toxicon.2022.06.014.

Hager ER, Harringmeyer OS, Wooldridge TB, Theingi S, Gable JT, McFadden S, Neugeboren B, Turner KM, Jensen JD, Hoekstra HE. A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science*. 2022:377(6604):399–405. https://doi.org/10.1126/science.abg0718.

Hämälä T, Wafula EK, Guiltinan MJ, Ralph PE, Tiffin P. Genomic structural variants constrain and facilitate adaptation in natural populations of *Theobroma cacao*, the chocolate tree. *Proc Natl Acad Sci U S A*. 2021:118(35):e2102914118. https://doi.org/10.1073/pnas.2102914118.

Han X, Zhang Y, Zhang Q, Ma N, Liu X, Tao W, Lou Z, Zhong C, Deng XW, Li D, *et al*. Two haplotype-resolved, gap-free genome assemblies for *Actinidia latifolia* and *Actinidia chinensis* shed light on the regulatory mechanisms of vitamin C and sucrose metabolism in kiwifruit. *Mol Plant*. 2023:16(2):452–470. https://doi.org/10.1016/j.molp.2022.12.022.

Harrison CM, Colbert J, Richter CJ, McDonald PJ, Trumbull LM, Ellsworth SA, Hogan MP, Rokyta DR, Margres MJ. Using morphological, genetic, and venom analyses to present current and historic evidence of *Crotalus horridusxadamanteus* hybridization on Jekyll Island, Georgia. *Southeast Nat*. 2022:22(7):158–174. https://doi.org/10.1656/058.021.0209.

Hinojosa JC, Koubínová D, Szenteczki MA, Pitteloud C, Dincă V, Alvarez N, Vila R. A mirage of cryptic species: genomics uncover striking mitonuclear discordance in the butterfly *Thymelicus sylvestris*. *Mol Ecol*. 2019:28(17):3857–3868. https://doi.org/10.1111/mec.15153.

Ho SS, Urban AE, Mills RE. Structural variation in the sequencing era. *Nat Rev Genet*. 2020:21(3):171–189. https://doi.org/10.1038/s41576-019-0180-9.

Hofmann EP, Rautsaw RM, Strickland JL, Holding ML, Hogan MP, Mason AJ, Rokyta DR, Parkinson CL. Comparative venom-gland transcriptomics and venom proteomics of four sidewinder rattlesnake (*Crotalus cerastes*) lineages reveal little differential expression despite individual variation. *Sci Rep*. 2018:8(1):15534. https://doi.org/10.1038/s41598-018-33943-5.

Hogan MP, Holding ML, Nystrom GS, Colston TJ, Bartlett DA, Mason AJ, Ellsworth SA, Rautsaw RM, Lawrence KC, Strickland JL, *et al*. The genetic regulatory architecture and epigenomic basis for age-related changes in rattlesnake venom. *Proc Natl Acad Sci U S A*. 2024:121(16):e2313440121. https://doi.org/10.1073/pnas.2313440121.

Hogan MP, Whittington AC, Broe MB, Ward MJ, Gibbs HL, Rokyta DR. The chemosensory repertoire of the eastern diamondback rattlesnake (*Crotalus adamanteus*) reveals complementary genetics of olfactory and vomeronasal-type receptors. *J Mol Evol*. 2021:89(6):313–328. https://doi.org/10.1007/s00239-021-10007-3.

Holding ML, Strickland JL, Rautsaw RM, Hofmann EP, Mason AJ, Hogan MP, Nystrom GS, Ellsworth SA, Colston TJ, Borja M, *et al*. Phylogenetically diverse diets favor more complex venoms in North American pitvipers. *Proc Natl Acad Sci U S A*. 2021:118(17):e2015579118. https://doi.org/10.1073/pnas.2015579118.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014:30(9):1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012:9(4):357–359. https://doi.org/10.1038/nmeth.1923.

Li A, Wang J, Sun K, Wang S, Zhao X, Wang T, Xiong L, Xu W, Qiu L, Shang Y, *et al*. Two reference-quality sea snake genomes reveal their divergent evolution of adaptive traits and venom systems. *Mol Biol Evol*. 2021:38(11):4867–4883. https://doi.org/10.1093/molbev/msab212.

Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinform*. 2011:12(1):323. https://doi.org/10.1186/1471-2105-12-323.

Li K, Xu P, Wang J, Yi X, Jiao Y. Identification of errors in draft genome assemblies at single-nucleotide resolution for quality assessment and improvement. *Nat Commun*. 2023:14(1):6556. https://doi.org/10.1038/s41467-023-42336-w.

Li W, Chu C, Li H, Zhang H, Sun H, Wang S, Wang Z, Li Y, Foster TM, López-Girona E, *et al*. Near-gapless and haplotype-resolved apple genomes provide insights into the genetic basis of rootstock-induced dwarfing. *Nat Genet*. 2024a:56(3):505–516. https://doi.org/10.1038/s41588-024-01657-2.

Li X, Wang Y, Cai C, Ji J, Han F, Zhang L, Chen S, Zhang L, Yang Y, Tang Q, *et al*. Large-scale gene expression alterations introduced by structural variation drive morphotype diversification in *Brassica oleracea*. *Nat Genet*. 2024b:56(3):517–529. https://doi.org/10.1038/s41588-024-01655-4.

Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, Schmutz J. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *Elife*. 2022:11(9):e78526. https://doi.org/10.7554/eLife.78526.

Low WY, Tearle R, Liu R, Koren S, Rhie A, Bickhart DM, Rosen BD, Kronenberg ZN, Kingan SB, Tseng E, *et al*. Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and Brahman cattle. *Nat Commun*. 2020:11(4):2071. https://doi.org/10.1038/s41467-020-15848-y.

Lynch VJ. Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. *BMC Evol Biol*. 2007:7(1):2. https://doi.org/10.1186/1471-2148-7-2.

Mable BK. Conservation of adaptive potential and functional diversity: integrating old and new approaches. *Conserv Genet*. 2019:20(1):89–100. https://doi.org/10.1007/s10592-018-1129-9.

Mackessy SP. Evolutionary trends in venom composition in the Western Rattlesnakes (*Crotalus viridis* sensu lato): Toxicity vs. tenderizers. *Toxicon*. 2010:55(8):1463–1474. https://doi.org/10.1016/j.toxicon.2010.02.028.

Margres MJ, Bigelow AT, Lemmon EM, Lemmon AR, Rokyta DR. Selection to increase expression, not sequence diversity, precedes gene family origin and expansion in rattlesnake venom. *Genetics*. 2017a:206(3):1569–1580. https://doi.org/10.1534/genetics.117.202655.

Margres MJ, McGivern JJ, Seavy M, Wray KP, Facente J, Rokyta DR. Contrasting modes and tempos of venom expression evolution in two snake species. *Genetics*. 2015a:199(1):165–176. https://doi.org/10.1534/genetics.114.172437.

Margres MJ, McGivern JJ, Wray KP, Seavy M, Calvin K, Rokyta DR. Linking the transcriptome and proteome to characterize the venom of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *J Proteom*. 2014:96(6):145–158. https://doi.org/10.1016/j.jprot.2013.11.001.

Margres MJ, Patton A, Wray KP, Hassinger AT, Ward MJ, Lemmon EM, Lemmon AR, Rokyta DR. Tipping the scales: the migration–selection balance leans toward selection in snake venoms. *Mol Biol Evol*. 2019:36(2):271–282. https://doi.org/10.1093/molbev/msy207.

Margres MJ, Rautsaw RM, Strickland JL, Mason AJ, Schramer TD, Hofmann EP, Stiers E, Ellsworth SA, Nystrom GS, Hogan MP, *et al*. The tiger rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. *Proc Natl Acad Sci U S A*. 2021:118(4):e2014634118. https://doi.org/10.1073/pnas.2014634118.

Margres MJ, Walls R, Suntravat M, Lucena S, Sánchez EE, Rokyta DR. Functional characterizations of venom phenotypes in the eastern diamondback rattlesnake (*Crotalus adamanteus*) and evidence for expression-driven divergence in toxic activities among populations. *Toxicon*. 2016a:119(9):28–38. https://doi.org/10.1016/j.toxicon.2016.05.005.

Margres MJ, Wray KP, Hassinger ATB, Ward MJ, McGivern JJ, Lemmon EM, Lemmon AR, Rokyta DR. Quantity, not quality: rapid adaptation in a polygenic trait proceeded exclusively through expression differentiation. *Mol Biol Evol*. 2017b:34(12):3099–3110. https://doi.org/10.1093/molbev/msx231.

Margres MJ, Wray KP, Seavy M, McGivern JJ, Herrera ND, Rokyta DR. Expression differentiation is constrained to low-expression proteins over ecological timescales. *Genetics*. 2016b:202(1):273–283. https://doi.org/10.1534/genetics.115.180547.

Margres MJ, Wray KP, Seavy M, McGivern JJ, Sanader D, Rokyta DR. Phenotypic integration in the feeding system of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *Mol Ecol*. 2015b:24(13):3405–3420. https://doi.org/10.1111/mec.13240.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal*. 2011:17(1):10–12. https://doi.org/10.14806/ej.17.1.200.

Matsubara K, Tarui H, Toriba M, Yamada K, Nishida-Umehara C, Agata K, Matsuda Y. Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. *Proc Natl Acad Sci U S A*. 2006:103(48):18190–18195. https://doi.org/10.1073/pnas.0605274103.

Means DB. Effects of rattlesnake roundups on the eastern diamondback rattlesnake (*Crotalus adamanteus*). *Herpetol Conserv Biol*. 2009:4(2):132–141.

Means DB. *Diamonds in the rough: natural history of the eastern diamondback rattlesnake*. Tallahassee (FL): Tall Timbers Press; 2017.

Meng G, Li Y, Yang C, Liu S. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res*. 2019:47(11):e63. https://doi.org/10.1093/nar/gkz173.

Mérot C, Oomen RA, Tigano A, Wellenreuther M. A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends Ecol Evol*. 2020:35(7):561–572. https://doi.org/10.1016/j.tree.2020.03.002.

Nachtigall PG. Guide to annotate toxin genes in the genome of venomous lineages. https://github.com/pedronachtigall/ToxCodAn-Genome/tree/main/Guide. 2023.

Nachtigall PG, Durham AM, Rokyta DR, Junqueira-de Azevedo ILM. ToxCodAn-Genome: an automated pipeline for toxin-gene

annotation in genome assembly of venomous lineages. *Gigascience*. 2024:13(17):giad116. https://doi.org/10.1093/gigascience/giad116.
Nachtigall PG, Freitas-de Sousa LA, Mason AJ, Moura-da Silva AM, Grazziotin FG, Junqueira-de Azevedo IL. Differences in PLA2 constitution distinguish the venom of two endemic Brazilian mountain lanceheads, *Bothrops cotiara* and *Bothrops fonsecai*. *Toxins (Basel)*. 2022:14(4):237. https://doi.org/10.3390/toxins14040237.
Nachtigall PG, Grazziotin FG, de Azevedo ILMJ. MITGARD: an automated pipeline for mitochondrial genome assembly in eukaryotic species using RNA-seq data. *Brief Bioinform*. 2021a:22(5): bbaa429. https://doi.org/10.1093/bib/bbaa429.
Nachtigall PG, Rautsaw RM, Ellsworth SA, Mason AJ, Rokyta DR, Parkinson CL, Junqueira-de Azevedo IL. Toxcodan: a new toxin annotator and guide to venom gland transcriptomics. *Brief Bioinform*. 2021b:22(5):bbab095. https://doi.org/10.1093/bib/bbab095.
Nakandala U, Masouleh AK, Smith MW, Furtado A, Mason P, Constantin L, Henry RJ. Haplotype resolved chromosome level genome assembly of *Citrus australis* reveals disease resistance and other citrus specific genes. *Hortic Res*. 2023:10(5):uhad058. https://doi.org/10.1093/hr/uhad058.
Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015:32(1):268–274. https://doi.org/10.1093/molbev/msu300.
Oguiura N, Collares MA, Furtado MFD, Ferrarezzi H, Suzuki H. Intraspecific variation of the crotamine and crotasin genes in *Crotalus durissus* rattlesnakes. *Gene*. 2009:446(1):35–40. https://doi.org/10.1016/j.gene.2009.05.015.
Oliveira AL, Viegas MF, Soares AM, Ramos MJ, Fernandes PA. The chemistry of snake venom and its medicinal potential. *Nat Rev Chem*. 2022:6(7):451–469. https://doi.org/10.1038/s41570-022-00393-7.
Palmer J, Stajich J. Funannotate: eukaryotic genome annotation. https://github.com/nextgenusfs/funannotate. 2017.
Pasquesi GI, Adams RH, Card DC, Schield DR, Corbin AB, Perry BW, Reyes-Velasco J, Ruggiero RP, Vandewege MW, Shortt JA, *et al.* Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nat Commun*. 2018:9(1): 2774. https://doi.org/10.1038/s41467-018-05279-1.
Pearse DE, Miller MR, Abadía-Cardoso A, Garza JC. Rapid parallel evolution of standing variation in a single, complex, genomic region is associated with life history in steelhead/rainbow trout. *Proc Biol Sci*. 2014:281(1783):20140012. https://doi.org/10.1098/rspb.2014.0012.
Peng C, Ren JL, Deng C, Jiang D, Wang J, Qu J, Chang J, Yan C, Jiang K, Murphy RW, *et al.* The genome of Shaw's sea snake (*Hydrophis curtus*) reveals secondary adaptation to its marine environment. *Mol Biol Evol*. 2020:37(6):1744–1760. https://doi.org/10.1093/molbev/msaa043.
Peng C, Wu DD, Ren JL, Peng ZL, Ma Z, Wu W, Lv Y, Wang Z, Deng C, Jiang K, *et al.* Large-scale snake genome analyses provide insights into vertebrate development. *Cell*. 2023:186(14):2959–2976. https://doi.org/10.1016/j.cell.2023.05.030.
Perry BW, Gopalan SS, Pasquesi GI, Schield DR, Westfall AK, Smith CF, Koludarov I, Chippindale PT, Pellegrino MW, Chuong EB, *et al.* Snake venom gene expression is coordinated by novel regulatory architecture and the integration of multiple co-opted vertebrate pathways. *Genome Res*. 2022:32(6):1058–1073. https://doi.org/10.1101/gr.276251.121.
Qi H, Cong R, Wang Y, Li L, Zhang G. Construction and analysis of the chromosome-level haplotype-resolved genomes of two *Crassostrea* oyster congeners: *Crassostrea angulata* and *Crassostrea gigas*. *Gigascience*. 2023:12(4):giad077. https://doi.org/10.1093/gigascience/giad077.
Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, *et al.* Towards complete and error-free genome assemblies of all vertebrate species. *Nature*. 2021:592(7856):737–746. https://doi.org/10.1038/s41586-021-03451-0.
Rokyta DR, Joyce P, Caudle SB, Wichman HA. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. *Nat Genet*. 2005:37(4):441–444. https://doi.org/10.1038/ng1535.
Rokyta DR, Lemmon AR, Margres MJ, Aronow K. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genom*. 2012:13(7):312. https://doi.org/10.1186/1471-2164-13-312.
Rokyta DR, Margres MJ, Calvin K. Post-transcriptional mechanisms contribute little to phenotypic variation in snake venoms. *G3: Genes Genomes Genet*. 2015:5(11):2375–2382. https://doi.org/10.1534/g3.115.020578.
Rokyta DR, Margres MJ, Ward MJ, Sanchez EE. The genetics of venom ontogeny in the eastern diamondback rattlesnake (*Crotalus adamanteus*). *PeerJ*. 2017:5(4):e3249. https://doi.org/10.7717/peerj.3249.
Rokyta DR, Wray KP, Lemmon AR, Lemmon EM, Caudle SB. A high-throughput venom-gland transcriptome for the eastern diamondback rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes. *Toxicon*. 2011:57(5): 657–0671. https://doi.org/10.1016/j.toxicon.2011.01.008.
Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res*. 2019:47(W1):W5–W10. https://doi.org/10.1093/nar/gkz342.
Sackman AM, McGee LW, Morrison AJ, Pierce J, Anisman J, Hamilton H, Sanderbeck S, Newman C, Rokyta DR. Mutation-driven parallel evolution during viral adaptation. *Mol Biol Evol*. 2017:34(12): 3243–3253. https://doi.org/10.1093/molbev/msx257.
Saitou M, Masuda N, Gokcumen O. Similarity-based analysis of allele frequency distribution among multiple populations identifies adaptive genomic structural variants. *Mol Biol Evol*. 2022:39(3): msab313. https://doi.org/10.1093/molbev/msab313.
Schield DR, Card DC, Hales NR, Perry BW, Pasquesi GM, Blackmon H, Adams RH, Corbin AB, Smith CF, Ramesh B, *et al.* The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. *Genome Res*. 2019:29(4): 590–601. https://doi.org/10.1101/gr.240952.118.
Schield DR, Perry BW, Card DC, Pasquesi GI, Westfall AK, Mackessy SP, Castoe TA. The rattlesnake W chromosome: a GC-rich retroelement refugium with retained gene function across ancient evolutionary strata. *Genome Biol Evol*. 2022:14(9):evac116. https://doi.org/10.1093/gbe/evac116.
Schonour RB, Huff EM, Holding ML, Claunch NM, Ellsworth SA, Hogan MP, Wray K, McGivern J, Margres MJ, Colston TJ, *et al.* Gradual and discrete ontogenetic shifts in rattlesnake venom composition and assessment of hormonal and ecological correlates. *Toxins (Basel)*. 2020:12(10):659. https://doi.org/10.3390/toxins12100659.
Serrato-Capuchina A, Matute DR. The role of transposable elements in speciation. *Genes (Basel)*. 2018:9(5):254. https://doi.org/10.3390/genes9050254.
Shi J, Jia Z, Sun J, Wang X, Zhao X, Zhao C, Liang F, Song X, Guan J, Jia X, *et al.* Structural variants involved in high-altitude adaptation detected using single-molecule long-read sequencing. *Nat Commun*. 2023:14(12):8282. https://doi.org/10.1038/s41467-023-44034-z.
Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics*. 2021:37(12):1639–1643. https://doi.org/10.1093/bioinformatics/btaa1016.
Smith CF, Nikolakis ZL, Ivey K, Perry BW, Schield DR, Balchan NR, Parker J, Hansen KC, Saviola AJ, Castoe TA, *et al.* Snakes on a plain: biotic and abiotic factors determine venom compositional variation in a wide-ranging generalist rattlesnake. *BMC Biol*. 2023:21(6):136. https://doi.org/10.1186/s12915-023-01626-x.
Stoltzfus A, McCandlish DM. Mutational biases influence parallel adaptation. *Mol Biol Evol*. 2017:34(9):2163–2173. https://doi.org/10.1093/molbev/msx180.
Strickland JL, Mason AJ, Rokyta DR, Parkinson CL. Phenotypic variation in Mojave rattlesnake (*Crotalus scutulatus*) venom is driven by four toxin families. *Toxins (Basel)*. 2018:10(4):135. https://doi.org/10.3390/toxins10040135.

Suryamohan K, Krishnankutty SP, Guillory J, Jevit M, Schröder MS, Wu M, Kuriakose B, Mathew OK, Perumal RC, Koludarov I, *et al.* The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. *Nat Genet*. 2020:52(1):106–117. https://doi.org/10.1038/s41588-019-0559-8.

Tigano A, Reiertsen TK, Walters JR, Friesen VL. A complex copy number variant underlies differences in both colour plumage and cold adaptation in a dimorphic seabird. bioRxiv 507384. https://doi.org/10.1101/507384, 2018, preprint: not peer reviewed.

Todesco M, Owens GL, Bercovich N, Légaré JS, Soudi S, Burge DO, Huang K, Ostevik KL, Drummond EB, Imerovski I, *et al.* Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*. 2020:584(7822):602–607. https://doi.org/10.1038/s41586-020-2467-6.

Toh H, Yang C, Formenti G, Raja K, Yan L, Tracey A, Chow W, Howe K, Bergeron LA, Zhang G, *et al.* A haplotype-resolved genome assembly of the Nile rat facilitates exploration of the genetic basis of diabetes. *BMC Biol*. 2022:20(11):245. https://doi.org/10.1186/s12915-022-01427-8.

Ventura M, Catacchio CR, Alkan C, Marques-Bonet T, Sajjadian S, Graves TA, Hormozdiari F, Navarro A, Malig M, Baker C, *et al.* Gorilla genome structural variation reveals evolutionary parallelisms with chimpanzee. *Genome Res*. 2011:21(10):1640–1649. https://doi.org/10.1101/gr.124461.111.

Viana PF, Ezaz T, de Bello Cioffi M, Jackson Almeida B, Feldberg E. Evolutionary insights of the ZW sex chromosomes in snakes: a new chapter added by the Amazonian puffing snakes of the genus *Spilotes*. *Genes (Basel)*. 2019:10(4):288. https://doi.org/10.3390/genes10040288.

Vickrey AI, Bruders R, Kronenberg Z, Mackey E, Bohlender RJ, Maclary ET, Maynez R, Osborne EJ, Johnson KP, Huff CD, *et al.* Introgression of regulatory alleles and a missense coding mutation drive plumage pattern diversity in the rock pigeon. *Elife*. 2018:7:e34803. https://doi.org/10.7554/eLife.34803.

Vollger MR, Dishuck PC, Sorensen M, Welch AE, Dang V, Dougherty ML, Graves-Lindsay TA, Wilson RK, Chaisson MJ, Eichler EE. Long-read sequence and assembly of segmental duplications. *Nat Methods*. 2019:16(12):88–94. https://doi.org/10.1038/s41592-018-0236-3.

Waldron JL, Welch SM, Bennett SH, Kalinowsky WG, Mousseau TA. Life history constraints contribute to the vulnerability of a declining North American rattlesnake. *Biol Conserv*. 2013:159(3):530–538. https://doi.org/10.1016/j.biocon.2012.11.021.

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*. 2018:35(3):msx319. https://doi.org/10.1093/molbev/msx319.

Weischenfeldt J, Symmons O, Spitz F, Korbel JO. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet*. 2013:14(2):125–138. https://doi.org/10.1038/nrg3373.

Weissensteiner MH, Bunikis I, Catalán A, Francoijs KJ, Knief U, Heim W, Peona V, Pophaly SD, Sedlazeck FJ, Suh A, *et al.* Discovery and population genomics of structural variation in a songbird genus. *Nat Commun*. 2020:11(7):3403. https://doi.org/10.1038/s41467-020-17195-4.

Whittington AC, Mason AJ, Rokyta DR. A single mutation unlocks cascading exaptations in the origin of a potent pitviper neurotoxin. *Mol Biol Evol*. 2018:35(4):887–898. https://doi.org/10.1093/molbev/msx334.

Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 2019:20(11):257. https://doi.org/10.1186/s13059-019-1891-0.

Wray KP, Margres MJ, Seavy M, Rokyta DR. Early significant ontogenetic changes in snake venoms. *Toxicon*. 2015:96(3):74–81. https://doi.org/10.1016/j.toxicon.2015.01.010.

Xu M, Guo L, Du X, Li L, Peters BA, Deng L, Wang O, Chen F, Wang J, Jiang Z, *et al.* Accurate haplotype-resolved assembly reveals the origin of structural variants for human trios. *Bioinformatics*. 2021:37(15):2095–2102. https://doi.org/10.1093/bioinformatics/btab068.

Yan H, Bombarely A, Li S. DeepTE: a computational method for de novo classification of transposons with convolutional neural network. *Bioinformatics*. 2020:36(15):4269–4275. https://doi.org/10.1093/bioinformatics/btaa519.

Zhang H, Song L, Wang X, Cheng H, Wang C, Meyer CA, Liu T, Tang M, Aluru S, Yue F, *et al.* Fast alignment and preprocessing of chromatin profiles with Chromap. *Nat Commun*. 2021a:12(11):6566. https://doi.org/10.1038/s41467-021-26865-w.

Zhang L, Reifová R, Halenková Z, Gompert Z. How important are structural variants for speciation? *Genes (Basel)*. 2021b:12(7):1084. https://doi.org/10.3390/genes12071084.

Zhang ZY, Lv Y, Wu W, Yan C, Tang CY, Peng C, Li JT. The structural and functional divergence of a neglected three-finger toxin subfamily in lethal Elapids. *Cell Rep*. 2022:40(2):111079. https://doi.org/10.1016/j.celrep.2022.111079.

Zhao SW, Guo JF, Kong L, Nie S, Yan XM, Shi TL, Tian XC, Ma HY, Bao YT, Li ZC, *et al.* Haplotype-resolved genome assembly of *Coriaria nepalensis* a non-legume nitrogen-fixing shrub. *Sci Data*. 2023a:10(5):259. https://doi.org/10.1038/s41597-023-02171-6.

Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics*. 2023b:39(1):btac808. https://doi.org/10.1093/bioinformatics/btac808.