SUPPLEMENTAL MATERIALS

Table of Contents

1. SUPPLEMENTARY METHODS

Prairie Rattlesnake Genome Sequencing and Assembly

A male Prairie Rattlesnake (*Crotalus viridis viridis*) collected from a wild population in Colorado was used to generate the genome sequence. This specimen was collected and humanely euthanized according to University of Northern Colorado Institutional Animal Care and Use Committee protocols 0901C-SM-MLChick-12 and 1302D-SM-S-16. Colorado Parks and Wildlife scientific collecting license 12HP974 issued to S.P. Mackessy authorized collection of the animal. Genomic DNA was extracted using a standard Phenol-Chloroform-Isoamyl alcohol extraction from liver tissue that was snap frozen in liquid nitrogen. Multiple short-read sequencing libraries were prepared and sequenced on various platforms, including 50bp single-end and 150bp paired-end reads on an Illumina GAII, 100bp paired-end reads on an Illumina HiSeq, and 300bp paired-end reads on an Illumina MiSeq. Long insert libraries were also constructed by and sequenced on the PacBio platform. Finally, we constructed two sets of mate-pair libraries using an Illumina Nextera Mate Pair kit, with insert sizes of 3-5 kb and 6-8 kb, respectively. These were sequenced on two Illumina HiSeq lanes with 150bp paired-end sequencing reads. Short and long read data were used to assemble the previous genome assembly version CroVir2.0 (NCBI accession SAMN07738522). Details of these sequencing libraries are in Supplemental Table S1. Prior to assembly, reads were adapter trimmed using BBmap (Bushnell 2014) and we quality trimmed all reads using Trimmomatic v0.32 (Bolger et al. 2014). We used Meraculous (Chapman et al. 2011) and all short-read Illumina data to generate a contig assembly of the Prairie Rattlesnake. We then performed a series of scaffolding and gap-filling steps. First, we used L_RNA_scaffolder (Xue et al. 2013) to scaffold contigs using the complete transcriptome assembly (see below), SSPACE Standard (Boetzer et al. 2010) to scaffold contigs using mate-pair reads, and SSPACE Longread to scaffold using long PacBio reads. We then used GapFiller (Nadalin et al. 2012) to extend contigs and fill gaps using all short-read data cross five iterations. We merged the scaffolded assembly with a contig assembly generated using the *de novo* assembly tool in CLC Genomics Workbench (Qiagen Bioinformatics, Redwood City, CA, USA).

We improved the CroVir2.0 assembly using the Dovetail Genomics HiRise assembly v2.1.3-59a1db48d61f method (Putnam et al. 2016), leveraging both Chicago and Hi-C sequencing. This assembly method has been used to improve numerous draft genome assemblies (e.g., Jiao et al. 2017; Rice et al. 2017). Chicago assembly requires large amounts of high molecular weight DNA from a very fresh tissue sample. We thus extracted high molecular weight genomic DNA from a liver of a closely related male to the CroVir2.0 animal (i.e., from the same den site). This animal was collected and humanely euthanized according to the Colorado Parks and Wildlife collecting license and UNC IACUC

protocols detailed above. Hi-C sequencing data were derived from the venom gland of the same animal (see details below on venom gland Hi-C and RNA-seq experimental design). The assembly was carried out using the existing CroVir2.0 draft genome assembly, short read data used in the previous assembly, Chicago, and Hi-C datasets. The HiRise assembly method then mapped Chicago and Hi-C datasets to the draft assembly and generated a model fit of the data based on insert size distributions (Supplemental Fig. S1; Supplemental Material 2). Models were generated with read pairs that mapped within the same scaffold and were used in successive join, break, and final join phases of the pipeline to perform final scaffolding. Dovetail Genomics HiRise assembly resulted in a highly contiguous genome assembly (CroVir3.0) with a physical coverage of greater than 1,000× (Supplemental Table S2).

We estimated the size of the genome using k-mer frequency distributions (19, 23, and 27mers) quantified using Jellyfish (Marçais and Kingsford 2011). Raw Illumina 100bp paired-end reads (Supplemental Table S1) were quality trimmed using Trimmomatic (Bolger et al. 2014) using the settings LEADING:10, TRAILING:10, SLIDINGWINDOW:4:15, and MINLEN:36. The total number of output sequences and bases were 400,983,222 and 38,471,185,282, respectively. Quality trimmed reads were then used for Jellyfish *k*-mer counting, and the Jellyfish *k*-mer table output per *k*-mer was used to estimate genome size with GCE (Liu et al. 2013).

We generated transcriptomic libraries from RNA sequenced from 16 different tissues: two venom gland tissues; 1 day and 3 days post-venom extraction (see Hi-C and RNA sequencing of Venom Gland section below), one from pancreas, and one from tongue were taken from the Hi-C sequenced genome animal. Additional samples from other individuals included a third venom gland sample from which venom had not been extracted (‘unextracted venom gland’), three liver, three kidney, two pancreas, and one each of skin, lung, testis, accessory venom gland, shaker muscle, brain, stomach, ovaries, rictal gland, spleen, and blood tissues. Total RNA was extracted using Trizol, and we prepared RNAseq libraries using an NEB RNA-seq kit for each tissue, which were uniquely indexed and run on multiple HiSeq 2500 lanes using 100bp paired-end reads (Supplemental Table S3). We used Trinity v. 20140717 (Grabherr et al. 2011) with default settings and the ‘--trimmomatic' setting to assemble transcriptome reads from all tissues. The resulting assembly contained 801,342 transcripts comprising 677,921 Trinity-annotated genes, with an average length of 559 bp and an N50 length of 718 bp.

Repeat Element Analysis

Annotation of repeat elements was performed using homology-based and *de novo* prediction approaches. Homology-based methods of transposable element identification (e.g., *RepeatMasker*) cannot recognize

elements that are not in a reference database, and have low power to identify fragments of repeat elements belonging to even moderately diverged repeat families (Platt et al. 2016). Since the current release of the Tetrapoda RepBase library (v.20.11, August 2015; Bao et al. 2015) is unsuitable for detailed repeat element analyses of most squamate reptile genomes, we performed *de novo* identification of repeat elements on 6 snake genomes (*Crotalus viridis*, *Crotalus mitchellii*, *Thamnophis sirtalis*, *Boa constrictor*, *Deinagkistrodon acutus*, and *Pantherophis guttatus*) in RepeatModeler v.1.0.9 (Smit and Hubley 2015) using default parameters. Consensus repeat sequences from multiple species were combined into a large joint snake repeat library that also includes previously identified elements from an additional 12 snake species (Castoe et al. 2013). All genomes were annotated with the same library with the exception of the green anole lizard, for which we used a lizard specific library that includes *de novo* repeat identification for *Pogona vitticeps*, *Ophisaurus gracilis*, and *Gekko japonicus*. To verify that only repeat elements were included in the custom reference library, all sequences were used as input in a BLASTx search against the SwissProt database (UniProt 2017), and those clearly annotated as protein domains were removed. Finally, redundancy and possible chimeric artifacts were removed through clustering methods in CD-HIT (Li and Godzik 2006) using a threshold of 0.85.

Homology-based repeat element annotation was performed in RepeatMasker v.4.0.6 (Smit et al. 2015) using a PCR-validated BovB/CR1 LINE retrotransposon consensus library (Castoe et al. 2013), the Tetrapoda RepBase library, and our custom library as references. Output files were post-processed using a modified implementation of the ProcessRepeat script (RepeatMasker package).

Gene Annotation

We used MAKER v. 2.31.8 (Cantarel et al. 2008) to annotate protein-coding genes in an iterative fashion. Several sources of empirical evidence of protein-coding genes were used, including the full *de novo C. viridis* transcriptome assembly and protein datasets consisting of all annotated proteins from NCBI for *Anolis carolinensis* (Alfoldi et al. 2011), *Python molurus bivittatus* (Castoe et al. 2013), *Thamnophis sirtalis* (Perry et al. 2018), and *Ophiophagus hannah* (Vonk et al. 2013), and from GigaDB for *Deinagkistrodon acutus* (Yin et al. 2016). We also included 422 protein sequences for 24 known venom gene families that were used to infer *Python* venom gene homologs in a previous study (Reyes-Velasco et al. 2015). Prior to running MAKER, we used BUSCO v. 2.0.1 (Simão et al. 2015) and the full *C. viridis* genome assembly to iterative train AUGUSTUS v. 3.2.3 (Stanke and Morgenstern 2005) HMM models based on 3,950 tetrapod vertebrate benchmarking universal single-copy orthologs (BUSCOs). We also ran this analysis on the previous genome assembly (CroVir2.0) as a comparison, and provide the details of these analyses in Supplemental Table S4. We ran BUSCO in the 'genome' mode and specified the '--

long' option to have BUSCO perform internal AUGUSTUS training. We ran MAKER with the 'est2genome=0' and 'protein2genome=0' options set to produce gene models using the AUGUSTUS gene predictions with hints supplied from the empirical transcript and protein sequence evidence. We provided the coordinates for all interspersed, complex repetitive elements for MAKER to perform hard masking before evidence mapping and prediction, and we set the 'model_org' option to 'simple' to have MAKER soft mask simple repetitive elements. We used default settings for all other options, except 'max_dna_len' (set to 300,000) and 'split_hit' (set to 20,000). We iterated this approach an additional time and we manually compared the MAKER gene models with the transcript and protein evidence. We found very little difference between the two gene annotations and based on a slightly better annotation edit distance (AED) distribution in the first round of MAKER, we used our initial round as the final gene annotation. The resulting annotation consisted of 17,486 genes and we ascribed gene IDs based on homology using reciprocal best-BLAST (with e-value thresholds of 1×10^{-5}) and stringent one-way BLAST (with an e-value threshold of 1×10^{-8}) searches against protein sequences from NCBI for *Anolis*, *Python*, and *Thamnophis*.

Hi-C and RNA Sequencing of the Venom Gland

We dissected the venom glands from the Hi-C *Crotalus viridis viridis* 1 day and 3 days after venom was initially extracted in order to track a time-series of venom production. A subsample of the 1-day venom gland was sent to Dovetail Genomics where DNA was extracted and replicate Hi-C sequencing libraries were prepared according to their protocol (see above). We also extracted total RNA from both 1-day and 3-day venom gland samples, along with tongue and pancreas tissue from the Hi-C genome animal (see Sequencing and Assembly and Annotation sections above). mRNA-seq libraries were generated and sequenced at Novogene on two separate lanes of the Illumina HiSeq 4000 platform using 150 bp paired-end reads (Supplemental Table S3).

Chromosome Identification and Synteny Analyses

Genome assembly resulted in several large, highly-contiguous scaffolds with a relative size distribution consistent with the karyotype of *C. viridis* (Baker et al. 1972), representing nearly-complete chromosome sequences. We determined the identity of chromosomes using a BLAST search of the chromosome-specific markers linked to snake chromosomes from (Matsubara et al. 2006), downloaded from NCBI (accessions SAMN00177542 and SAMN00152474). We kept the best alignment per cDNA marker as its genomic location in the Prairie Rattlesnake genome, except when a marker hit two high-similarity matches on different chromosomes. The vast majority of markers linked to a specific macrochromosome

(i.e., Chromosomes 1-7; Supplemental Table S6) in *Elaphe quadrivirgata* mapped to a single genomic scaffold; only six of 104 markers did not map to the predicted chromosome from *E. quadrivirgata*. Possible reasons for unmatched chromosomal locations for these markers in *Elaphe* and the Prairie Rattlesnake include 1) original localizations in *Elaphe* that are unique to the species or were localized in error, 2) translocations have occurred, leading to divergent locations in each genome, or 3) misassembly errors in the rattlesnake genome assembly. To distinguish between these possibilities, we first identified the chromosomal location of each marker in the Anole Lizard (*Anolis carolinensis*) genome (Alfoldi et al. 2011) to determine if their locations are expected based on *Elaphe-Anolis* synteny. Three markers mapped to unexpected chromosomes in *Anolis* (*NOSIP*, *GNAI2*, and *P4HB*), which instead matched syntenic locations in the rattlesnake (Supplemental Table S7). *Anolis* synteny for a fourth marker (*UCHL1*) suggested correct assembly in the rattlesnake, but was unclear because it mapped to *Anolis* Chromosome 5, which is syntenic with both snake Chromosomes 6 and 7 (Fig. 1). To determine if the two remaining markers (*ZNF326* and *KLF6*) were placed on unexpected chromosomes due to misassembly, and to identify further evidence that the other markers were assembled correctly, we leveraged our intrachromosomal Hi-C data to deeply investigate contact patterns around these markers. Specifically, we plotted heatmaps of $\log_{10}$ normalized contact frequencies in 10 kb bins using R (R Core Team 2017). Regional dropout in intrachromosomal contact frequencies in the focal regions would be expected if mismatched chromosome locations were due to misassembly in the rattlesnake. We focused our searches on genomic intervals around each of the six focal genes and the nearest confirmed marker from Supplemental Table S6. The genomic region around each gene showed intrachromosomal contact frequencies consistent with correct assembly for five of six markers (Supplemental Fig. S2). Only *ZNF326* was adjacent to a region with intrachromosomal contact dropout that could have resulted from misassembly. All snake microchromosome markers mapped to a single 139Mb scaffold, which was later broken into 10 microchromosome scaffolds (scaffold-mi1-10; see below).

We identified a single 114Mb scaffold corresponding to the Z Chromosome, as 10 of 11 Z-linked markers mapped to this scaffold. To further vet this as the Z-linked region of the genome, we mapped reads from male and female *C. viridis* (Supplemental Table S9) to the genome using BWA (Li and Durbin 2009) using program defaults. Male and female resequencing libraries were prepared using an Illumina Nextera prep kit and sequenced on an Illumina HiSeq 2500 using 250bp paired-end reads. Adapters were trimmed and low-quality reads were filtered using Trimmomatic (Bolger et al. 2014). After mapping, we filtered reads with low mapping scores and quantified per-base read depths using SAMtools (Li et al. 2009). We then totaled read depths for consecutive 100 kb windows and normalized windowed totals for female and male by dividing the value for each window by the median autosomal 100 kb window value for each sex,

then determined the normalized ratio of female to male coverage by calculating $\log_2$(female normalized coverage/male normalized coverage) per window. Here, the expectation is that a hemizygous locus will show roughly half the normalized coverage, which we observe for females over the majority of the Z Chromosome scaffold length, and not elsewhere in the genome. To demonstrate Z Chromosome conservation among pit vipers and to further determine the identity of this scaffold, we mapped male and female Pygmy Rattlesnake (*Sistrurus catenatus*) reads from Vicoso et al. (2013) and female and male Five Pace Viper (*Deinagkistrodon acutus*) reads from Yin et al. (2016) to the genome using the same parameters detailed above (Supplemental Fig. S7). *Anolis* Chromosome 6 is homologous with snake sex chromsomes (Srikulnath et al. 2009), thus we aligned *Anolis* Chromosome 6 (Alfoldi et al. 2011) to the Prairie Rattlesnake genome using a chromosome painting technique described below. As expected, we found a large quantity of high-similarity hits to the rattlesnake Z Chromosome scaffold, specifically, which were organized in a sequential manner across the Z scaffold (Fig. 1B).

We used multiple sources of information to identify the best candidate breakpoints between microchromosomes within the 139Mb fused microchromosome scaffold in the initial Hi-C assembly. First, because Chicago scaffolds must be assembled from fragments that are physically linked (Rice et al. 2017), we used breakpoints between adjacent Chicago scaffolds on the microchromosome scaffold as candidate misjoins between microchromosomes, which identified 305 candidate break points. Second, intrachromosomal contact frequencies have been shown to be exponentially higher than contacts between chromosomes (Lieberman-Aiden et al. 2009), and we used shifts in intrachromosomal Hi-C data to further identify the nine most biologically plausible candidate break points among microchromosomes (Supplemental Fig. S16). Here, we stress two things relevant to using Hi-C contact data for this purpose: 1) intrachromosomal contacts within candidate microchromosomes were far more frequent than contacts between candidate microchromosomes, as expected (Supplemental Fig. S16), and 2) the nine Hi-C derived breakpoints overlapped consistently with breaks between Chicago scaffolds. Because reptile microchromosomes are highly syntenic (Alfoldi et al. 2011), we also aligned the microchromosome scaffold to microchromosome scaffolds from chicken (Hillier et al. 2004) and *Anolis* using LASTZ (Harris 2007) to determine if likely chromosomal breakpoints also had shifts in synteny. To retain only highly similar alignments per comparison, we set the 'hspthresh' option equal to 10,000 (default is 3,000). We also set a step size equal to 20 to reduce computational time per comparison. We further validated candidate break points using genomic features that consistently vary at the ends of chromosomes. Here, we specifically evaluated if candidate breakpoints exhibited regional shifts in GC content and repeat content, similar to the ends of macrochromosomes (Fig. 1). Finally, if no annotated genes spanned this junction, we considered it biologically plausible. There were nine candidate breakpoints that met each of

these criteria, equaling the number of boundaries expected given ten microchromosomes (Supplemental Fig. S16).

To explore broad-scale structural evolution across reptiles, we used the rattlesnake genome to perform in silico painting of the chicken (*Gallus gallus* version 5) and green anole *Anolis carolinensis* (version 2) genomes. Briefly, we divided the rattlesnake genome into 2.02 million potential 100 bp markers. For each of these markers, we used BLAST to record the single best hit in the target genome requiring an alignment length of at least 50 bp. This resulted in 41,644 potential markers in *Gallus* and 103,801 potential markers in *Anolis*. We then processed markers on each chromosome by requiring at least five consecutive markers supporting homology to the same rattlesnake chromosome. We consolidated each group of five consecutive potential markers as one confirmed marker. In *Gallus*, we rejected 12.4% of potential markers and identified 7,291 confirmed merged markers. In *Anolis*, we rejected 39.7% of potential markers and identified 12,511 confirmed merged markers.

This approach demonstrates considerable stability at the chromosomal level despite 158 million years of divergence between *Anolis* and *Crotalus* (Fig. 1B, Supplemental Fig. S5), and between squamates and birds, despite 280 million years of divergence between *Gallus* and *Crotalus* (and between *Gallus* and *Anolis*). This stability is evident not only in the macrochromosomes but also in the microchromosomes. In fact, 7 of 10 *Crotalus* microchromosomes had greater than 80% of confirmed markers associated with a single chromosome in the chicken genome (Fig. 1B, microchromosome inset). Comparisons among the three genomes suggest that the *Crotalus* genome has not experienced some of the fusions found in *Anolis*. Specifically, we infer that *Anolis* Chromosome 3 is a fusion of *Crotalus* Chromosomes 4 and 5. Likewise, *Anolis* Chromosome 4 is a fusion of *Crotalus* Chromosome 6 and 7. Divergence time estimates discussed above and shown in Fig. 1B were taken from the median of estimates for divergence between *Crotalus* and *Gallus* and between *Crotalus* and *Anolis* from Timetree (www.timetree.org; Kumar et al. 2017).

To validate the genome-wide *k*-mer based approach used to identify homology among reptile chromosomes, we also performed a more traditional analysis using only protein-coding genes. We first identified 2,190 three-way reciprocal best BLAST hits among rattlesnake, anolis, and chicken protein-coding genes that we used as markers. Both the chicken and *Anolis* genomes contain genes that have not been placed on chromosomes and remain in unmapped scaffold or contigs, which reduced the number of markers available to 2,105 in chicken and 2,135 in *Anolis*. Results from this approach indicate that the *k*-mer approach is consistent with this more traditional approach but provides approximately three times the density of markers (Fig. 1B, Supplemental Fig. S5).

Genomic Patterns of GC Content

We quantified GC content in sliding windows of 100 kb and 1Mb across the genome using a custom Python script (https://github.com/drewschield/Comparative-Genomics-Tools/blob/master/slidingwindow_gc_content.py). GC content in 100 kb windows is presented in Fig. 1 in the Main Text.

To determine if there is regional variation in nucleotide composition consistent with isochore structures across the rattlesnake genome, we quantified GC content and its variance within 5, 10, 20, 40, 80, 160, 240, and 320 kb windows. The variation (standard deviation) in GC content is expected to decrease by half as window size increases four-fold if the genome is homogeneous (i.e., lacks isochore structures; (Venter et al. 2001). By comparing the observed variances of GC content across spatial window scales to those from 11 other squamate genomes, including lizards (*Anolis* has been shown to lack isochore structure (Alfoldi et al. 2011), henophidian snakes, and colubroid snakes, we were able to determine the relative heterogeneity of nucleotide composition in the rattlesnake (Supplemental Table S8). To reduce potential biases from estimates from small scaffold sizes, we filtered to only retain scaffolds greater than the size of the window analyzed (e.g., only scaffolds longer than 10 kb when looking at the standard deviation in GC content over 10 kb windows) and for which there was less than 20% of missing data.

To study patterns of molecular evolution across squamate evolution, we generated whole genome alignments of 12 squamates including the Green Anole (*Anolis carolinensis* v. anoCar2.0; Alfoldi et al. 2011), Australian Bearded Dragon (*Pogona vitticeps* v. pvi1.1; Georges et al. 2015), Crocodile Lizard (*Shinisaurus crocodilurus* GigaDb version; Gao et al. 2017), Glass Lizard (*Ophisaurus gracilis* v. O.gracilis.final; Song et al. 2015), Schlegel's Japanese Gecko (*Gekko japonicus* v. 1.1 ; Liu et al. 2015), Leopard Gecko (*Eublepharis macularius* v. 1.0; Xiong et al. 2016), Prairie Rattlesnake (*Crotalus viridis* v. CroVir3.0; current study), Five-pacer Viper (*Deinagkistrodon acutus* GigaDb version; Yin et al. 2016), Burmese Python (*Python bivittatus* v. Python_molurus_bivittatus-5.0.2; Castoe et al. 2013), Boa Constrictor (*Boa constrictor* v. 7C; Bradnam et al. 2013), Garter Snake (*Thamnophis sirtalis* NCBI version; Perry et al. 2018), and King Cobra (*Ophiophagus Hannah* v. OphHan1.0; Vonk et al. 2013). We obtained the repeat libraries for each species and softmasked each assembly. The repeat library was not available for *Deinagkistrodon*, so we annotated repeats in that assembly using RepeatMasker v4.0.5 (Smit et al. 2015) with the vertebrate library from RepBase (Jurka et al. 2005). First, we generated pairwise syntenic alignments of each species as a query to the green anole genome (anoCar2.0) as a target using LASTZ v1.02 (Harris 2007) with the HoxD55 scoring matrix, followed by chaining to form gapless blocks and netting to rank the highest scoring chains (Kent et al. 2003) . The pairwise alignments were

used to construct a multiple sequence alignment with MULTIZ v11.2 (Blanchette et al. 2004) with Green Anole as the reference species. We then filtered the multi-species whole genome alignment to retain only blocks for which information for all 12 species was available, and concatenated blocks according to their organization in the anole lizard genome. We then calculated GC content within consecutive 50 kb windows of this concatenated alignment using the 'slidingwindow_gc_content.py' script detailed above.

Comparative Microchromosome Genomics

To understand evolutionary shifts in microchromosome composition among amniotes, we compared measures of gene density, GC content, and repeat content of macro- and microchromosomes between the rattlesnake, anole (Alfoldi et al. 2011), bearded dragon (Georges et al. 2015; Deakin et al. 2016), chicken (Hillier et al. 2004), and zebra finch (Warren et al. 2010) genomes. These species were chosen because their scaffolds are ordered into chromosomes and because their karyotypes contain microchromosomes. For each species, we downloaded relevant data from Ensembl and quantified the total number of genes per chromosome, total number of G+C bases, and total bases masked as repeats in RepeatMasker. We then normalized each measure by the total length of macrochromosome and microchromosome sequences in each genome, then calculated the ratio of microchromosome:macrochromosome proportions. We then used Fisher's exact tests determine if one chromosome set possessed a significantly greater proportion of each measure. We generated a phylogenetic tree (Supplemental Fig. S4) for the five species based on divergence time estimates from TimeTree (Kumar et al. 2017), and plotted the ratio values calculated above onto the tree tips for between-species comparisons.

Hi-C analysis

Raw Illumina paired-end reads were mapped and processed using the Juicer pipeline (Durand et al. 2016) to produce Hi-C maps binned at multiple resolutions, as low as 5 kb resolution, and for the annotation of contact domains. These data were aligned against the CroVir3.0 assembly. All contact matrices used for further analysis were KR-normalized in Juicer. TAD domains were called using Juicer's Arrowhead algorithm for finding contact domains at various resolutions (5 kb, 10 kb, 25 kb, 50 kb and 100 kb) using the default settings (Durand et al. 2016). 175 TADs were identified at 5 kb resolution, 16 at 10 kb, 53 at 25 kb, 175 at 50 kb, and 126 at 100 kb. Additionally, TADs were annotated at 20kb resolution using the HiCExplorer software (Ramírez et al. 2018). Raw reads were mapped and processed separately through HiCExplorer and 1,262 TADs were called at 20 kb resolution using the default settings with the p-value set to 0.05. We further identified TADs by eye at finer scale (i.e., 5 kb) resolution.

We compared intra and interchromosomal contact frequencies between the rattlesnake venom gland and various tissues from mammals. To do this we quantified the total intra- and interchromosomal contacts between chromosome positions from the rattlesnake and the following Hi-C datasets: human lymphoblastoma cells (Rao et al. 2014) and human retinal epithelial cells, mouse kidney, and rhesus macaque tissue (Darrow et al. 2016). To investigate patterns of intra- and interchromosome contact frequency, we normalized contact frequencies by chromosome length. In the case of the mouse, we removed the Y chromosome due to its small size and relative lack of interchromosomal contacts. We then performed linear regressions of chromosome length and normalized intra- and interchromosomal contact frequencies (i.e., contact frequency/chromosome length). In all cases we observed a positive relationship between normalized intrachromosomal contacts and chromosome size and a negative relationship between normalized interchromosomal contacts and chromosome size (Fig. 3B). We also tested for significant differences in intra- and interchromosomal contact between the rattlesnake and mammals using *t*-tests.

Sex Chromosome Analysis

We identified the Prairie Rattlesnake Z Chromosome using methods described in the 'Chromosome Identification and Synteny Analyses' section above. We localized the candidate pseudoautosomal region (PAR) based on normalized female/male coverage (Fig. 2A; the PAR is the only region of the Z consistent with equal female and male coverage. We quantified gene content, GC content, and repeat content across the Z Chromosome and PAR (Supplemental Figs. S8, S9, and S10), and tested for gene enrichment in the PAR using a Fisher's exact test, where we compared the number of genes within each region to the total length of the region.

To compare nucleotide diversity (π) across the genome between male and female *C. viridis*, we called variants (i.e., heterozygous sites) from the male and female reads used in coverage analysis detailed above. With the mappings from coverage analysis, we used SAMtools (Li et al. 2009) to compile all mappings into pileup format, from which we called variant sites using BCFtools. We filtered sites to retain only biallelic variants using VCFtools (Danecek et al. 2011) and calculated the proportion of heterozygous sites using a custom pipeline of scripts. First, calcHet (https://github.com/darencard/RADpipe) outputs details of heterozygous site and window_heterozygosity.py (https://github.com/drewschield/Comparative-Genomics-Tools/blob/master/window_heterozygosity.py) uses this output in conjunction with a windowed bed file generated using BEDtools 'make_windows' tool to calculate π within a given window size. We then

normalized π for each genomic window in the female and male by the median value of π for female and male autosomes, respectively.

Evolutionary patterns of the Z Chromosome were also analyzed by examining transposable element age and composition along the whole chromosome, and between the PAR and the Z, specifically (see Main Text). Since the length of the PAR is significantly smaller than the length of the Z, to rule out potential biases due to unequal sample size we also independently analyzed fragments of the Z with lengths equal to the PAR (total of 15 7.18 Mbp fragments). Each region was analyzed in RepeatMasker using a single reference library that included the squamate fraction of the RepBase Tetrapoda library, and the snake specific library clustered at a threshold of 0.75. The age distribution of TE families was estimated by mean of the Kimura 2-parameter distance from the consensus sequence per element (CpG corrected) calculated from PostProcessed.align outputs (see 'Repeat Analysis' section above), and using a modified Perl script from Kapusta et al. (2017). We then merged estimates of repeat content from each of these regions for comparison to the PAR region, specifically.

To quantify gene expression on the rattlesnake Z Chromosome and across the genome, we prepared RNA-seq libraries from liver and kidney tissue from two males and females and sequenced them on an Illumina HiSeq using 100bp paired-end reads (Supplemental Table S9). Samples and libraries were prepared following the previously described methods of (Andrew et al. 2017). After filtering and adapter trimming using Trimmomatic v. 0.32 (Bolger et al. 2014), we mapped RNA-seq reads to the *C. viridis* genome using STAR v. 2.5.2b (Dobin et al. 2013) and counts were determined using featureCounts (Liao et al. 2013). To be comparable to anole and chicken RNA-seq data described below, we analyzed the rattlesnake RNA-seq reads as single-end data by ignoring the second read of each read pair. We normalized read counts across tissues and samples using TMM normalization in edgeR (Robinson and Oshlack 2010) to generate both counts per million (CPM) for use in pairwise comparisons between males and females, and reads per kilobase million (RPKM) normalized counts for comparisons of chromosome-wide expression within samples. All subsequent analyses of gene expression included only genes with expression information in both the male and female (>1 average RPKM in each sex; average overall for female and male were roughly equal). Mann-Whitney *U* tests in R (R Core Team 2017) were used to compare median expression level between chromosomes and/or chromosomal regions (i.e., the PAR) within males and females. Per gene female-to-male ratios of expression in the Z Chromosome were normalized by taking the $\log_2$ of the female and male Z expression values scaled to the median expression level of autosomal genes in female and male, respectively:

$$\text{Current female/male Z} = log_2 \left[\left(\frac{female\ Z}{median\ female\ Auto} \right) / \left(\frac{male\ Z}{median\ male\ Auto} \right) \right]$$

To explore regional variation in the current female-to-male (F/M) gene expression ratio across the Z Chromosome, we performed a sliding window analysis of the $\log_2$ F/M expression ratio with a window size of 30 genes and a step size of 1 gene.

To further investigate patterns of gene expression in females and males across the Z Chromosome, we compared current levels of female and male expression for Z-linked genes to inferred ancestral levels of expression using autosomal 1:1 orthologs in the anole lizard and the chicken. Comparisons of sex chromosome-linked genes to autosomal orthologs in outgroup species have been shown to provide robust information about global ancestral expression patterns in the 'proto-sex' chromosomes of the focal species (Julien et al. 2012; Marin et al. 2017), and can be used to determine if patterns of gene expression between sexes are consistent with each other and with the evolution of dosage compensation mechanisms. We first filtered to retain only the 1,343 non-PAR genes on the rattlesnake Z Chromosome for comparison, and used reciprocal best BLAST searches to find putative 1:1 orthologs in the Ensembl anole (version 2) and chicken (version 5) cDNA datasets, respectively. This resulted in 682 1:1 orthologs between the rattlesnake and the anole, and 291 between the rattlesnake and the chicken, and 260 shared orthologs among the three species (i.e., 'proto-Z' genes). All putative orthologs are located on autosomes in both the anole and chicken. We also identified 3,059 1:1 orthologs that are autosomal in all three species (i.e., 'proto-autosomal' genes). We then obtained RNA-seq data from Marin et al. (2017) for female and male kidney and liver tissue for the chicken and anole (at least two replicates per tissue per sex) and performed filtering, mapping, and normalization of counts using the methods described above for the rattlesnake.

We used female and male expression levels from rattlesnake Z autosomal orthologs in the anole and chicken to infer ancestral (i.e., proto-Z) female and male expression levels. To do this, we first calculated the average expression value per proto-autosomal gene between the anole and chicken for each sex, and then calculated the median expression value from each of these distributions. We used these median values to normalize female and male expression in the anole and chicken 1:1 rattlesnake Z orthologs (proto-Z genes) to a common scale (these values are analogous to the median female or male autosomal denominators in the equations above for current female/male expression).

$$\text{Proto-Z female} = \frac{female\ proto{-}Z\ gene}{median\ female\ proto{-}autosomal}$$

$$\text{Proto-Z male} = \frac{male\ proto{-}Z\ gene}{median\ male\ proto{-}autosomal}$$

We then calculated a weighted average of female and male proto-Z expression per gene between the anole and chicken designed to account for the more recent divergence between the anole and rattlesnake, which was equal to the reciprocal of the sum of branch lengths based on the divergence times in millions of years between rattlesnake and anole and between rattlesnake and chicken:

$$\text{Branch length weight} = \frac{branch\ length\ (rattlesnake\ to\ anole)=158}{branch\ length\ (rattlesnake\ to\ chicken)=402} = 0.393$$

$$\text{Weighted Proto-Z female} = log2\left[\frac{[(Proto{-}Z\ female\ anole*1)+(Proto{-}Z\ female\ chicken*0.393)]}{1.393}\right]$$

$$\text{Weighted Proto-Z male} = log2\left[\frac{[(Proto{-}Z\ male\ anole*1)+(Proto{-}Z\ male\ chicken*0.393]}{1.393}\right]$$

To further compare current and ancestral Z expression to the female and male distributions of proto-autosomal expression, we calculated the average expression between the anole and chicken per proto-autosomal gene, then normalized the averaged expression by the median of proto-autosomal expression detailed above:

$$\text{Proto-autosomal female} = log2\left[\frac{female\ proto{-}autosomal\ gene}{median\ female\ proto{-}autosomal}\right]$$

$$\text{Proto-autosomal male} = log2\left[\frac{male\ proto{-}autosomal\ gene}{median\ female\ proto{-}autosomal}\right]$$

We also calculated the distribution of current autosomal expression in the rattlesnake by normalizing the current female and male expression of rattlesnake autosomal genes by the median of female and male expression of all autosomal genes, respectively:

$$\text{Current autosomal female} = log_2\left[\frac{female\ autosomal\ gene}{median\ female\ autosomal}\right]$$

$$\text{Current autosomal male} = log_2\left[\frac{male\ autosomal\ gene}{median\ female\ autosomal}\right]$$

We tested for enrichment of male and female-biased gene expression on chromosomes by first characterizing genes as male or female biased if their current $\log_2$(female/male) expression ratio was less than -0.5 or greater than 0.5, respectively. We then compared proportions of male-biased, female-biased, and unbiased between the Z Chromosome and autosomes using Fisher's exact tests to determine if the Z Chromosome is enriched or depleted for sex-biased gene expression.

A potential mechanism for upregulation of Z-linked genes in females is positive regulation through estrogen response elements (EREs), which can enable binding of enhancers and promote transcription of genes over long distances (Lin et al. 2007). Rice et al. (2017) identified that the binding domain of *ESR1* is completely conserved among humans, chickens, and alligators, thus we obtained a position weight matrix for the *ESR1* binding motif (ERE) of humans (Lin et al. 2007) from the CisBP database, and performed binding site prediction using PoSSuM Search (Beckstette et al. 2006). For more details on PoSSuM Search parameters, see the 'Transcription Factor Binding Site Prediction' section below. We quantified the number of predicted EREs and the average current female/male gene expression ratio (see above) along the Z Chromosome in 100 kb windows, and tested for a relationship between these variables using a Pearson's correlation coefficient in R.

We also quantified the number of predicted EREs in the entire genome, as well as the entire *Anolis* genome. We then compared the density of EREs (i.e., number of EREs divided by total scaffold length) between the rattlesnake *Anolis* genomes, and between the rattlesnake Z Chromosome and *Anolis* Chromosome 6, specifically. We tested for ERE enrichment on the Z Chromosome compared to *Anolis 6* using a Fisher's exact test in R. To test more broadly for an expansion of EREs in snakes, we repeated this analysis using Z-linked and autosomal scaffolds from the five pace viper (*Deinagkistron acutus*; Yin et al. 2016).

Transcription Factor Binding Site Prediction

To identify putative transcription factor binding sites throughout the rattlesnake genome, we obtained the TRANSFAC position weight matrix (PSSM) for transcription factors of interest from the CIS-BP database (Weirauch et al. 2014). The focal transcription factors (e.g., *CTCF*, *NFI, GRHL1*, *ESR1*, and the remaining transcription factors on Supplemental Table S12) have conserved DNA binding domains among vertebrates, and where possible we obtained the chicken binding PSSM. In some cases there was no curated PSSM for chicken, and we used the PSSMs for human, and in the case of *NCOA2* (Supplemental Table S12), there was no available PSSM for a close relative. We searched for putative binding sites throughout the genome using PoSSuM Search (Beckstette et al. 2006). Because each PSSM has a different probability distribution based on the relative frequencies of observed binding and the length of the element, we pre-calculated the complete probability distribution for each PSSM using PoSSuMdist. We then used the resulting distribution in conjunction with relative base frequencies for the genome calculated using PoSSuMfreqs to identify putative binding sites exceeding a significance threshold. This threshold necessarily varied for different PSSMs, but was never higher than $p < 1 \times 10^{-5}$.

Venom Gene Annotation and Analysis

We took a multi-step approach toward identifying venom gene homologs in the rattlesnake genome. We first obtained representative gene sequences for 38 venom gene families from GenBank (Supplemental Table S10), comprising known enzymatic and toxin components of snake venoms. We then searched our transcript set using the venom gene family query set using a tBLASTx search, defining a similarity cutoff e-value of 1×10^{-5}. For each candidate venom gene transcript identified in this way, we then performed a secondary tBLASTx search against the NCBI database to confirm its identity as a venom gene. In the case of several venom gene families, such as those known only from elapid snake venom, we did not find any candidate genes. Three venom gene families that are especially abundant, both in terms of presence in the venom proteome (Fig. 4a) and in copy number, in the venom of *C. viridis* are phospholipases A2 (*PLA2s*), snake venom metalloproteinases (*SVMPs*), and snake venom serine proteases (*SVSPs*). Rattlesnakes possess multiple members of each of these gene families (Mackessy 2008; Casewell et al. 2011; Dowell et al. 2016), and the steps taken above appeared to underestimate the total number of copies in the *C. viridis* genome. Therefore, for each of these families, we performed an empirical annotation using the Fgenesh++ (Solovyev et al. 2006) protein similarity search. We first extracted the genomic region annotated for each of these families above plus and minus a 100 kb flanking region. We used protein sequences from Uniprot (*PLA2*: APD70899.1; *SVMP*: Q90282.1; and *SVSP*: F8S114.1) to query the region and confirm the total number of copies per family. Each gene annotated in this way was again searched against NCBI to confirm its identity and manual searches of aligned protein sequences (see phylogenetic analyses below) further confirmed their homology to each respective venom gene family. Genomic locations and details of annotated venom genes in the rattlesnake genome are provided in Table S9. We tested for venom gene enrichment on microchromosomes versus macrochromosomes using a Fisher's exact test, where numerator for each category was the number of venom genes located on each chromosome type, and the denominator in each category was the background number of genes, which allowed us to account for different levels of gene density on microchromosomes and macrochromosomes.

We used LASTZ (Harris 2007) to align the genomic regions containing *PLA2*, *SVMP*, and *SVSP* genes to themselves. We used program defaults, with the exception of the 'hspthresh' command, which we set to 8,000. This was done to only return very high similarity matches between compared sequences. Here the expectation is that when alignments are plotted against one another, we will observe a diagonal line demonstrating perfect matches between each stretch of sequence and itself. In the case of segmental duplications, we also expect to see parallel and perpendicular (if in reverse orientation) segments adjacent to the diagonal 'self' axis. We plotted LASTZ results for each of the regions using the base plotting function in R (R Core Team 2017).

We then performed Bayesian phylogenetic analyses to further evaluate evidence of tandem duplication and monophyly among members of the *PLA2*, *SVMP*, and *SVSP* venom gene families. We generated protein alignments of venom genes with their closest homologs, which we identified using tBLASTx searches between venom genes and our whole gene set) using MUSCLE (Edgar 2004) with default parameters, with minor manual edits to the alignment to remove any poorly aligned regions. We analyzed the protein alignments using BEAST2 (Bouckaert et al. 2014), setting the site model to 'WAG' for each analysis. We ran each analysis for a minimum of 1×10^8 generations, and evaluated whether runs had reached stationarity using Tracer (Drummond and Rambaut 2007). After discarding the first 10% of samples as burnin, we generated consensus maximum clade credibility trees using TreeAnnotator (distributed with BEAST2).

Analyses of Venom Gland Gene Expression

To explore venom gland gene expression in comparison to other body tissues, raw Illumina RNA-seq reads from all tissues (Supplemental Table S3) were quality trimmed using Trimmomatic v. 0.36 (Bolger et al. 2014) with default settings. We used STAR (Dobin et al. 2013) to align reads to the genome. Raw expression counts were estimated by counting the number of reads that mapped uniquely to a particular annotated transcript using HTSeq-count (Anders et al. 2013). These raw counts were then normalized and filtered in edgeR using TMM normalization (Oshlack et al. 2010; Robinson et al. 2010), and all subsequent analyses were done using these normalized data. To test for significant expression differences between venom gland and body tissues, we performed pairwise comparisons between combined venom gland (i.e., 1 day venom gland, 3 day venom gland, and unextracted venom gland) and body (all other tissues, except for accessory venom gland) tissue sets using an exact test of the binomial distribution estimated in edgeR, integrating tagwise dispersion (Robinson and Oshlack 2010). Genes with differential expression at an FDR value ≤ 0.05 were considered significant. Heatmaps were generated in R using the heatmap function from the R Stats package (R Core Team 2017).

To identify candidate transcription factors regulating venom gene expression, we searched the genome annotation for all genes included on the UniProt (http://www.uniprot.org) reviewed human transcription factor database, by specifying species = 'Homo sapiens' and reviewed = 'yes' in the advanced search terms. Using this list, we parsed our significant venom gland expressed gene results detailed above for candidate venom gland transcription factors, which showed a pattern of overall low body-wide expression and statistically significant evidence of higher expression in the venom gland, specifically. We identified 12 candidates using this approach, including four members of the *CTF*/*NFI* family of RNA polymerase II core promoter-binding transcription factors (*NFIA,* two isoforms of *NFIB*, and *NFIX*). *NFI* binding sites

have been identified upstream of venom genes in several venomous snake taxa, including viperids, elapids, and colubrids (e.g., crotamine/myotoxin in *Crotalus durissus* (Rádis-Baptista et al. 2003) and three finger toxins in *Naja sputatrix* (Lachumanan et al. 1998) and *Boiga dendrophila* (Pawlak and Kini 2008). *NFI* family members were also found to be expressed in the venom glands of several species in a previous study exploring putative venom gland transcription factors (Hargreaves et al. 2014), but information about whether they showed venom gland-specific expression was not provided. This set also included the grainyhead-like homolog 1 (*GRHL1*) transcription factor Other significantly up-regulated transcription factors in the venom gland appear to be involved in the unfolded protein stress response of the endoplasmic reticulum and in glandular epithelium development and maintenance (Fig. 4B; Supplemental Table S12). We quantified the distance between predicted binding sites of all transcription factors upregulated in the venom gland (Supplemental Table S12) from 1) venom genes and 2) non-venom genes and compared these distance distributions using *t*-tests.

Because four transcription factors of the *NFI* family each showed evidence of venom gland-specificity, we tested the hypothesis that their binding motifs are also upstream of venom genes by quantifying the number of predicted *NFI* binding sites from PSSM analyses detailed above in the 1 kb upstream region of each venom gene. We also searched for proximity of *GRHL1* binding sites to venom gene regions, as well as all nonvenom genes, using BEDtools (Quinlan and Hall 2010) to calculate the number of predicted binding sites within 100 kb, 50 kb, 10 kb, and 5 kb intervals up and downstream of each gene. Here, we did not confine our search only to promoter regions. To test for enrichment of *NFI* binding sites in the upstream regions of venom genes, we divided the number of predicted binding sites upstream of venom genes by the total length of upstream regions and compared this value to the analogous proportion for upstream regions of all nonvenom genes using a Fisher's exact test (Supplemental Table S13). We performed a similar analysis for *GRHL1* at each interval size, again comparing the density of predicted *GRHL1* binding sites within intervals of venom genes to nonvenom genes (Supplemental Table S13). We also used the Bedtools 'closest' function (Quinlan and Hall 2010) to calculate the distribution of distances between genes and predicted *GRHL1* binding sites.

Supplementary References

Alfoldi J, Di Palma F, Grabherr M, Williams C, Kong L, Mauceli E, Russell P, Lowe CB, Glor RE, Jaffe JD et al. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**(7366): 587-591.

Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD. 2013. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat Protoc* **8**(9): 1765.

Andrew AL, Perry BW, Card DC, Schield DR, Ruggiero RP, McGaugh SE, Choudhary A, Secor SM, Castoe TA. 2017. Growth and stress response mechanisms underlying post-feeding regenerative organ growth in the Burmese python. *BMC Genomics* **18**(1): 338.

Baker RJ, Bull JJ, Mengden GA. 1972. Karyotypic studies of 38 species of North-American snakes. *Copeia* (2): 257-&.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**(1): 11.

Beckstette M, Homann R, Giegerich R, Kurtz S. 2006. Fast index based algorithms and software for matching position specific scoring matrices. *BMC Bioinformatics* **7**(1): 389.

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AFA, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res* **14**(4): 708-715.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2010. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**(4): 578-579.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15): 2114-2120.

Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comp Bio* **10**(4): e1003537.

Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S, Chapman JA, Chapuis G, Chikhi R et al. 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience* **2**(1): 10.

Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US).

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell M. 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* **18**(1): 188-196.

Casewell NR, Wagstaff SC, Harrison RA, Renjifo C, Wüster W. 2011. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. *Mol Biol Evol* **28**(9): 2637-2649.

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

Chapman JA, Ho I, Sunkara S, Luo S, Schroth GP, Rokhsar DS. 2011. Meraculous: de novo genome assembly with short paired-end reads. *PLoS One* **6**(8): e23501.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**(15): 2156-2158.

Darrow EM, Huntley MH, Dudchenko O, Stamenova EK, Durand NC, Sun Z, Huang S-C, Sanborn AL, Machol I, Shamim M. 2016. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc Natl Acad Sci USA*: E4504-E4512.

Deakin JE, Edwards MJ, Patel H, O'Meally D, Lian J, Stenhouse R, Ryan S, Livernois AM, Azad B, Holleley CE. 2016. Anchoring genome sequence to chromosomes of the central bearded dragon (*Pogona vitticeps*) enables reconstruction of ancestral squamate macrochromosomes and identifies sequence content of the Z chromosome. *BMC Genomics* **17**.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**(1): 15-21.

Dowell NL, Giorgianni MW, Kassner VA, Selegue JE, Sanchez EE, Carroll SB. 2016. The deep origin and recent loss of venom toxin genes in rattlesnakes. *Curr Biol* **26**(18): 2434-2445.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**.

586 Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides
587 a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* **3**(1): 95-98.
588 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
589 *Nucleic Acids Res* **32**(5): 1792-1797.
590 Gao J, Li Q, Wang Z, Zhou Y, Martelli P, Li F, Xiong Z, Wang J, Yang H, Zhang G. 2017. Sequencing,
591 de novo assembling, and annotating the genome of the endangered Chinese crocodile lizard
592 *Shinisaurus crocodilurus*. *GigaScience* **6**(7): 1-6.
593 Georges A, Li Q, Lian J, O'Meally D, Deakin J, Wang Z, Zhang P, Fujita M, Patel HR, Holleley CE.
594 2015. High-coverage sequencing and annotated assembly of the genome of the Australian dragon
595 lizard *Pogona vitticeps*. *GigaScience* **4**(1): 45.
596 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury
597 R, Zeng Q. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference
598 genome. *Nat Biotechnol* **29**(7): 644.
599 Hargreaves AD, Swain MT, Hegarty MJ, Logan DW, Mulley JF. 2014. Genomic and transcriptomic
600 insights into the regulation of snake venom production. *bioRxiv*: 008474.
601 Harris RS. 2007. *Improved pairwise alignment of genomic DNA*. The Pennsylvania State University.
602 Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen
603 MAM, Delany ME. 2004. Sequence and comparative analysis of the chicken genome provide
604 unique perspectives on vertebrate evolution. *Nature* **432**(7018): 695-716.
605 Jiao W, Garcia Accinelli G, Hartwig B, Kiefer C, Baker D, Severing E, Willing E, Piednoel M, Woetzel
606 S, Madrid-Herrero E et al. 2017. Improving and correcting the contiguity of long-read genome
607 assemblies of three plant species using optical mapping and chromosome conformation capture
608 data. *Genome Res* **27**: 778-786.
609 Julien P, Brawand D, Soumillon M, Necsulea A, Liechti A, Schütz F, Daish T, Grützner F, Kaessmann H.
610 2012. Mechanisms and evolutionary patterns of mammalian and avian dosage compensation.
611 *PLoS Biol* **10**(5): e1001328.
612 Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a
613 database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**(1-4): 462-467.
614 Kapusta A, Suh A, Feschotte C. 2017 Dynamics of genome size evolution in birds and mammals. *Proc*
615 *Natl Acad Sci USA* **114**(8): 1460-1469.
616 Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution's cauldron: duplication,
617 deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci USA* **100**(20):
618 11484-11489.
619 Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and
620 divergence times. *Mol Biol Evol* **34**(7): 1812-1819.
621 Lachumanan R, Armugam A, Tan C-H. 1998. Structure and organization of the cardiotoxin genes in *Naja*
622 *naja sputatrix*. *FEBS letters* **433**(1-2): 119-124.
623 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.
624 *Bioinformatics* **25**(14): 1754-1760.
625 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome
626 Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools.
627 *Bioinformatics* **25**(16): 2078-2079.
628 Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or
629 nucleotide sequences. *Bioinformatics* **22**(13): 1658-1659.
630 Liao Y, Smyth GK, Shi W. 2013. featureCounts: an efficient general purpose program for assigning
631 sequence reads to genomic features. *Bioinformatics* **30**(7): 923-930.
632 Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR,
633 Sabo PJ, Dorschner MO et al. 2009. Comprehensive mapping of long-range interactions reveals
634 folding principles of the human genome. *Science* **326**(5950): 289-293.
635 Lin C-Y, Vega VB, Thomsen JS, Zhang T, Kong SL, Xie M, Chiu KP, Lipovich L, Barnett DH, Stossi F.
636 2007. Whole-genome cartography of estrogen receptor α binding sites. *PLoS Genet* **3**(6): e87.

637 Liu B, Shi Y, Yuan J, Hu X, Zhang H, Li N, Li Z, Chen Y, Mu D, Fan W. 2013. Estimation of genomic
638 characteristics by analyzing k-mer frequency in *de novo* genome projects. *arXiv:13082012*.
639 Liu Y, Zhou Q, Wang Y, Luo L, Yang J, Yang L, Liu M, Li Y, Qian T, Zheng Y. 2015. *Gekko japonicus*
640 genome reveals evolution of adhesive toe pads and tail regeneration. *Nat Commun* **6**: 10033.
641 Mackessy SP. 2008. Venom composition in rattlesnakes: trends and biological significance. In *The*
642 *Biology of Rattlesnakes*, (ed. WK Hayes, KR Beaman, MD Cardwell, SP Bush). Loma Linda
643 University Press, Loma Linda, CA.
644 Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of
645 k-mers. *Bioinformatics* **27**(6): 764-770.
646 Marin R, Cortez D, Lamanna F, Pradeepa MM, Leushkin E, Julien P, Liechti A, Halbert J, Brüning T,
647 Mössinger K. 2017. Convergent origination of a *Drosophila*-like dosage compensation
648 mechanism in a reptile lineage. *Genome Res* **27**(12): 1974-1987.
649 Matsubara K, Tarui H, Toriba M, Yamada K, Nishida-Umehara C, Agata K, Matsuda Y. 2006. Evidence
650 for different origin of sex chromosomes in snakes, birds, and mammals and step-wise
651 differentiation of snake sex chromosomes. *Proc Natl Acad Sci USA* **103**(48): 18190-18195.
652 Nadalin F, Vezzi F, Policriti A. 2012. GapFiller: a de novo assembly approach to fill the gap within
653 paired reads. *BMC Bioinformatics* **13**(14): S8.
654 Oshlack A, Robinson MD, Young MD. 2010. From RNA-seq reads to differential expression results.
655 *Genome Biol* **11**(12): 220.
656 Pawlak J, Kini RM. 2008. Unique gene organization of colubrid three-finger toxins: complete cDNA and
657 gene sequences of denmotoxin, a bird-specific toxin from colubrid snake *Boiga dendrophila*
658 (Mangrove Catsnake). *Biochimie* **90**(6): 868-877.
659 Perry BW, Card DC, McGlothlin JW, Pasquesi GI, Hales NR, Corbin AB, Adams RH, Schield DR, Fujita
660 MK, Demuth JP et al. In Review. Molecular adaptations for sensing and securing prey, and
661 insight into amniote genome diversity, revealed by the garter snake genome.
662 Platt RN, Blanco-Berdugo L, Ray DA. 2016. Accurate transposable element annotation is vital when
663 analyzing new genome assemblies. *Genome Biol Evol* **8**(2): 403-410.
664 Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD,
665 Sugnet CW et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-
666 range linkage. *Genome Res* **26**: 1-9.
667 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
668 *Bioinformatics* **26**(6): 841-842.
669 R Core Team. 2017. R: A language and environment for statistical computing.
670 Rádis-Baptista G, Kubo T, Oguiura N, Svartman M, Almeida TMB, Batistic RF, Oliveira EB, Vianna-
671 Morgante ÂM, Yamane T. 2003. Structure and chromosomal localization of the gene for
672 crotamine, a toxin from the South American rattlesnake, *Crotalus durissus terrificus*. *Toxicon*
673 **42**(7): 747-752.
674 Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A,
675 Manke T. 2018. High-resolution TADs reveal DNA sequences underlying genome organization
676 in flies. *Nat Commun* **9**(1): 189.
677 Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I,
678 Omer AD, Lander ES. 2014. A 3D map of the human genome at kilobase resolution reveals
679 principles of chromatin looping. *Cell* **159**(7): 1665-1680.
680 Reyes-Velasco J, Card DC, Andrew AL, Shaney KJ, Adams RH, Schield DR, Casewell NR, Mackessy
681 SP, Castoe TA. 2015. Expression of venom gene homologs in diverse python tissues suggests a
682 new model for the evolution of snake venom. *Mol Biol Evol* **32**(1): 173-183.
683 Rice ES, Kohno S, John JS, Pham S, Howard J, Lareau LF, O'Connell BL, Hickey G, Armstrong J, Deran
684 A et al. 2017. Improved genome assembly of American alligator genome reveals conserved
685 architecture of estrogen signaling. *Genome Res* **27**(5): 686-696.
686 Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential
687 expression analysis of digital gene expression data. *Bioinformatics* **26**(1): 139-140.

688 Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of
689 RNA-seq data. *Genome Biol* **11**(3): R25.
690 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing
691 genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*
692 **31**(19): 3210-3212.
693 Smit AFA, Hubley R. 2015. RepeatModeler Open 1.0.
694 Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0. 2013–2015. *Institute for Systems Biology*
695 *http://repeatmasker org*.
696 Solovyev V, Kosarev P, Seledsov I, Vorobyev D. 2006. Automatic annotation of eukaryotic genes,
697 pseudogenes and promoters. *Genome Biol* **7**(1): S10.
698 Song B, Cheng S, Sun Y, Zhong X, Jin J, Guan R, Murphy RW, Che J, Zhang Y, Liu X. 2015. A genome
699 draft of the legless anguid lizard, *Ophisaurus gracilis*. *GigaScience* **4**(1): 17.
700 Srikulnath K, Nishida C, Matsubara K, Uno Y, Thongpan A, Suputtitada S, Apisitwanich S, Matsuda Y.
701 2009. Karyotypic evolution in squamate reptiles: comparative gene mapping revealed highly
702 conserved linkage homology between the butterfly lizard (*Leiolepis reevesii rubritaeniata*,
703 Agamidae, Lacertilia) and the Japanese four-striped rat snake (*Elaphe quadrivirgata*, Colubridae,
704 Serpentes). *Chromosome Res* **17**(8): 975-986.
705 Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows
706 user-defined constraints. *Nucleic Acids Res* **33**(suppl_2): W465-W467.
707 UniProt C. 2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**(D1): D158-D169.
708 Vicoso B, Emerson JJ, Zektser Y, Mahajan S, Bachtrog D. 2013. Comparative sex chromosome genomics
709 in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Biol*
710 **11**(8): e1001643.
711 Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans MA,
712 Holt RA et al. 2001. The sequence of the human genome. *Science* **291**: 1304-1351.
713 Vonk FJ, Casewell NR, Henkel CV, Heimberg AM, Jansen HJ, McCleary RJR, Kerkkamp HME, Vos
714 RA, Guerreiro I, Calvete JJ et al. 2013. The king cobra genome reveals dynamic gene evolution
715 and adaptation in the snake venom system. *Proc Natl Acad Sci USA* **110**(51): 20651-20656.
716 Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Kunstner A, Searle S, White S, Vilella AJ,
717 Fairley S et al. 2010. The genome of a songbird. *Nature* **464**(7289): 757-762.
718 Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert
719 SA, Mann I, Cook K. 2014. Determination and inference of eukaryotic transcription factor
720 sequence specificity. *Cell* **158**(6): 1431-1443.
721 Xiong Z, Li F, Li Q, Zhou L, Gamble T, Zheng J, Kui L, Li C, Li S, Yang H. 2016. Draft genome of the
722 leopard gecko, *Eublepharis macularius*. *GigaScience* **5**(1): 47.
723 Xue W, Li J-T, Zhu Y-P, Hou G-Y, Kong X-F, Kuang Y-Y, Sun X-W. 2013. L_RNA_scaffolder:
724 scaffolding genomes with transcripts. *BMC genomics* **14**(1): 604.
725 Yin W, Wang ZJ, Li QY, Lian JM, Zhou Y, Lu BZ, Jin LJ, Qiu PX, Zhang P, Zhu WB et al. 2016.
726 Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-
727 pacer viper. *Nat Commun* **7**: 13107.

728

729

2. SUPPLEMENTAL TABLES

Supplemental Table S1. Sequencing libraries used in the Prairie Rattlesnake genome assembly. Where noted, various libraries were used in the previous assembly (CroVir2.0). Data from Chicago and Hi-C libraries are available under NCBI BioProject accession PRJNA413201.

Library	Read Type	Number of Reads	Assembly Version
50bp short read	single end	9,536,384	CroVir2.0
100bp short read	paired end	449775645	CroVir2.0, CroVir3.0
150bp short read	paired end	41,211,014	CroVir2.0
150bp long insert mate pair (3-5 kb)	paired end	188,532,564	CroVir2.0
150bp long insert mate pair (6-8 kb)	paired end	189,928,342	CroVir2.0
PacBio long reads	-	1,027,365	CroVir2.0
Chicago library 1 (150 bp)	paired end	251,689,106	CroVir3.0
Chicago library 2 (150 bp)	paired end	206,176,028	CroVir3.0
Hi-C library 1 (150 bp)	paired end	230,083,402	CroVir3.0
Hi-C library 2 (150 bp)	paired end	160,673,944	CroVir3.0

Supplemental Table S2. Basic information about assembly versions for the Prairie Rattlesnake genome.

Assembly	CroVir2.0	Chicago Assembly	Chicago + Hi-C Assembly
Longest Scaffold (bp)	1,184,546	11,576,738	311,712,589
Number of Scaffolds	47,782	8,183	7,034
Number of Scaffolds > 1 kb	47,658	8,059	6,910
Contig N50 (kb)	15.81	14.91	14.96
Scaffold N50 (kb)	139	2,472	179,898
Number of Gaps	112,369	158,269	159,024
Percent of Genome in Gaps	5.84%	6.15%	6.16%

Supplemental Table S3. RNA-seq libraries used for transcriptome assembly. Raw reads for each library are available on the NCBI Short Read Archive, accession PRJNA477004.

Sample ID	Tissue	Raw Reads	Quality Trimmed Reads
CroVirPan	pancreas	28,126,703	27,073,946
CroVirTon	tongue	24,451,116	23,561,349
CroVirVG1	venom gland	41,744,110	40,147,306
CroVirVG3	venom gland	29,216,664	28,035,353
Cvv01	liver	7,833,506	7,365,740
Cvv02	liver	7,451,792	7,064,234
Cvv11	liver	9,218,939	8,441,587
Cvv20	kidney	6,958,120	6,580,387
Cvv22	kidney	8,116,679	7,601,517
Cvv23	kidney	7,193,762	6,785,947
Cvv25	skin	7,849,895	7,303,441
Cvv26	pancreas	8,886,612	8,160,214
Cvv27	venom gland	3,098,151	2,928,974
Cvv28	lung	6,613,196	6,024,613
Cvv29	testes	5,055,189	4,745,375
Cvv30	accessory venom gland	3,261,326	3,053,142
Cvv31	shaker muscle	4,290,989	3,996,274
Cvv32	pancreas	4,836,715	4,566,165
Cvv33	brain	3,815,570	3,569,113
Cvv34	stomach	5,297,110	4,993,142
Cvv35	ovaries	3,737,870	3,528,104
Cvv36	rictal gland	6,654,626	6,070,883
Cvv37	spleen	7,776,020	6,975,210
Cvv38	blood	2,550,433	2,364,162

Supplemental Table S4. BUSCO results for assembly versions of the prairie rattlesnake genome. Proportions of each category are in parentheses.

BUSCO category	CroVir2.0	CroVir3.0 (current)
Complete	3,277 (83.0 %)	3,372 (85.3 %)
Complete and single-copy	3,253 (82.4%)	3,347 (84.7%)
Complete and duplicated	24 (0.6%)	25 (0.6%)
Fragmented	364 (9.2%)	298 (7.5%)
Missing	309 (7.8%)	280 (7.2%)
Total searched	3,950	3,950

745 **Supplemental Table S5.** Genome-wide annotated repeat proportions identified using RepeatMasker.

	# elements	length masked (bp)	% of sequence	% element masked
Total masked	2,966,274	489,373,735	38.91	100.00
Total interspersed repeats	2,348,232	463,237,605	36.83	79.16
Retroelements	1,139,213	295,244,109	22.81	38.41
SINEs	173,332	22,894,322	1.82	5.84
Squam1/Sauria	19,230	3,376,458	0.27	0.65
Other SINEs	126,898	15,602,678	1.24	4.28
LINEs	621,859	170,275,973	13.54	20.96
CR1-Like	359,387	91,177,000	7.25	12.12
CR1/L3	288,888	74,285,822	5.91	9.74
L2	53,219	12,036,490	0.96	1.79
Rex	19,032	5,339,363	0.42	0.64
R1/LOA/Jockey	3,272	854,611	0.07	0.11
R2/R4/NeSL	35,256	9,045,775	0.72	1.19
RTE/Bov-B	101,958	32,795,496	2.61	3.44
L1/CIN4	78,926	28,358,227	2.25	2.66
Other LINEs	154,019	16,472,232	0.64	5.19
Other nonLTR	10,119	1,572,442	0.13	0.34
DIRS	28,657	13,553,057	1.08	0.97
PLEs	120,162	19,278,497	1.53	4.05
LTR elements	156,427	54,116,761	4.30	5.27
BEL/Pao	4,007	1,927,682	0.15	0.14
Ty1/Copia	9,160	3,340,874	0.27	0.31
Gypsy	77,793	35,080,772	2.79	2.62
Retroviral	16,727	5,393,228	0.43	0.56
Other LTR	48,740	8,374,205	0.67	1.64
DNA transposons	850,487	125,287,793	9.96	28.67
hobo-Activator	428,247	60,243,144	4.79	14.44
Tc1-IS630-Pogo	283,367	48,888,185	3.89	9.55
En-Spm	12,485	1,964,905	0.16	0.42
MuDR-IS905	1,300	383,077	0.03	0.04
PiggyBac	131	22,504	0.00	0.00
Tourist/Harbinger	80,904	7,193,605	0.57	2.73
P elements	155	45,074	0.00	0.01
Rolling-circles	3,736	635,885	0.05	0.13
SPIN	253	26,640	0.00	0.01
Other DNA	39,909	5,884,774	0.47	1.35
Unclassified	358,532	48,493,199	3.86	12.09
Total interspersed repeats	2,348,232	463,237,605	36.83	79.16
Small RNA	2,054	174,940	0.01	0.07
Satellites	4,952	1,104,344	0.09	0.17
Simple repeats	540,288	28,572,170	2.27	18.21
Low complexity	70,748	4,755,565	0.38	2.39

746

Supplemental Table S6. Mapping of cDNA markers from Matsubara et al. 2006 to the Prairie Rattlesnake genome. Locations of best BLAST hits of each cDNA marker to the genome are reported. Markers that mapped with exceptional similarity to multiple locations in the genome are denoted with a '#', and markers that did not map to the chromosome as predicted by Matsubara et al. (2006) are denoted with a '*'. Details for these markers are provided in Supplemental Table S7 and Supplemental Fig. S2.

Marker	Accession	Chromosome	Scaffold	e-value	bit-score	Start Position	End Position
OMG	BW999947	1p	scaffold-ma1	6.00E-115	398	309337082	309336564
XAB1	AU312353	1p	scaffold-ma1	2.00E-46	122	297437298	297437486
MGC15407	AU312344	1p	scaffold-ma1	2.00E-65	92.3	288097081	288097206
XPO1	AU312325	1p	scaffold-ma1	2.00E-113	153	289547707	289547901
DEGS	AU312341	1p	scaffold-ma1	5.00E-106	356	269312409	269311948
KIAA0007	AU312332	1p	scaffold-ma1	5.00E-50	120	265943692	265943841
EPRS	AU312324	1p	scaffold-ma1	2.00E-91	174	270708945	270709160
ARID4B	AU312346	1p	scaffold-ma1	1.00E-129	333	252059286	252059699
QKI	AU312356	1p	scaffold-ma1	5.00E-112	124	246094729	246094887
MDN1	AU312339	1p	scaffold-ma1	7.00E-60	109	211517498	211517349
AFTIPHILIN	AU312311	1p	scaffold-ma1	5.00E-75	112	170752748	170752888
SF3B1	AU312337	1q	scaffold-ma1	7.00E-95	215	150078848	150078576
CACNB4	BW999948	1q	scaffold-ma1	1.00E-47	102	127283965	127283819
ZFHX1B	BW999949	1q	scaffold-ma1	6.00E-93	204	123301385	123301101
UMPS	AU312331	1q	scaffold-ma1	8.00E-95	198	113761458	113761724
TCIRG1	BW999950	1q	scaffold-ma1	2.00E-72	164	102088882	102089094
TSG101	AU312316	1q	scaffold-ma1	4.00E-76	113	88358887	88359054
M11S1	AU312350	1q	scaffold-ma1	4.00E-31	94.5	70777673	70777560
GPHN	AU312327	1q	scaffold-ma1	5.00E-68	116	60249829	60249644
DNCH1	AU312310	1q	scaffold-ma1	1.00E-71	145	25060055	25059885
HSPCA	BW999951	1q	scaffold-ma1	2.00E-123	149	25029984	25030184
ISYNA1	AU312338	1q	scaffold-ma1	2.00E-89	178	7770987	7771196
TUBGCP2	AU312343	1q	scaffold-ma1	4.00E-74	136	9697568	9697377
ZFR	AU312309	2p	scaffold-ma2	8.00E-110	208	222653709	222653461
PHAX	AU312322	2p	scaffold-ma2	3.00E-99	224	189308026	189307715
VPS13A	BW999952	2p	scaffold-ma2	9.00E-70	109	179725513	179725656
UBQLN1	BW999953	2p	scaffold-ma2	2.00E-87	132	182156077	182156238
C9orf72	AU312326	2p	scaffold-ma2	5.00E-91	203	164760033	164760347
KIAA0368	BW999954	2p	scaffold-ma2	1.00E-56	116	161287251	161287397
TOPORS	BW999955	2p	scaffold-ma2	8.00E-118	410	162258381	162257809
FAM48A	BW999956	2cen	scaffold-ma2	1.00E-45	102	157286823	157286680
UNQ501	AU312305	2cen	scaffold-ma2	6.00E-118	284	142895238	142895636
DCTN2	AU312317	2q	scaffold-ma2	4.00E-80	122	122527271	122527110
EXOC7	BW999957	2q	scaffold-ma2	3.00E-93	121	92952368	92952526
DDX5	BW999958	2q	scaffold-ma2	7.00E-112	144	108253948	108253775
CCNG1	AU312308	2q	scaffold-ma2	6.00E-70	173	80553964	80553731

CPEB4	AU312333	2q	scaffold-ma2	3.00E-119	250	72297563	72297874
FLJ22318	AU312329	2q	scaffold-ma2	2.00E-105	194	51908839	51908582
DCTN4	AU312349	2q	scaffold-ma2	4.00E-50	99.6	58962806	58962928
C5orf14	AU312304	2q	scaffold-ma2	4.00E-120	329	64853582	64853127
NOSIP[*]	AU312303	2q	scaffold-Z	1.00E-51	93.6	92988551	92988661
RBM5[#]	BW999960	2q	scaffold-mi8	6.00E-78	90.4	9620291	9620181
RBM5[#]	BW999960	2q	scaffold-ma2	7.00E-13	76.1	130725514	130725606
ITPR1	BW999961	2q	scaffold-ma2	9.00E-53	135	23858424	23858585
ENPP2	BW999962	3p	scaffold-ma3	6.00E-90	121	9756367	9756209
YWHAZ	BW999963	3p	scaffold-ma3	2.00E-99	180	16759896	16760114
LRRCC1	BW999964	3p	scaffold-ma3	4.00E-83	150	21993774	21993565
LYPLA1	BW999965	3p	scaffold-ma3	3.00E-107	149	31673258	31673440
SS18	AU312302	3p	scaffold-ma3	1.00E-83	126	36811554	36811724
MBP	AU312318	3p	scaffold-ma3	7.00E-111	179	49049170	49049382
EPB41L3	BW999966	3p	scaffold-ma3	3.00E-84	141	40222999	40222808
TUBB2A	BW999967	3p	scaffold-ma3	8.00E-91	155	59187732	59187532
LRRC16	BW999968	3p	scaffold-ma3	2.00E-100	144	51025171	51025350
SERPINB6[#]	BW999969	3p	scaffold-ma5	5.00E-99	130	36540937	36540755
SERPINB6[#]	BW999969	3p	scaffold-ma3	2.00E-76	113	60484038	60483865
BPHL	BW999970	3p	scaffold-ma3	1.00E-87	118	59199779	59199621
KIF13A	BW999971	3p	scaffold-ma3	3.00E-78	139	53681516	53681349
TPR	BW999972	3q	scaffold-ma3	6.00E-83	122	93408800	93408636
AKR1A1	BW999973	3q	scaffold-ma3	9.00E-75	153	133869419	133869619
ZNF326[*]	BW999974	3q	scaffold-ma2	2.00E-77	120	224940437	224940586
YIPF1	BW999975	3q	scaffold-ma3	6.00E-52	112	127724189	127724353
BCAS2	AU312354	3q	scaffold-ma3	3.00E-51	141	151621402	151621229
KIAA1219	BW999976	3q	scaffold-ma3	4.00E-101	158	155122635	155122844
STAU1	BW999977	3q	scaffold-ma3	2.00E-116	169	165663812	165663594
RBM12	BW999978	3q	scaffold-ma3	2.00E-152	406	154706304	154705780
TPT1	BW999979	4p	scaffold-ma4	2.00E-68	148	1006155	1006349
EIF2S3	AU312306	4p	scaffold-ma4	1.00E-111	126	49115724	49115885
SYAP1	AU312328	4p	scaffold-ma4	3.00E-96	121	46147275	46147135
DSCR3	AU312319	4q	scaffold-ma4	1.00E-74	119	60873037	60872873
DCAMKL1	BW999980	4q	scaffold-ma4	8.00E-49	110	86291138	86291302
ELMOD1	BW999981	4q	scaffold-ma4	1.00E-56	147	93207704	93207522
BCCIP	AU312307	5q	scaffold-ma5	1.00E-46	148	32597249	32597061
SH3MD1	AU312347	5q	scaffold-ma5	2.00E-119	378	45831798	45832379
PPP1R7	BW999982	5q	scaffold-ma5	2.00E-92	228	56956062	56955736
PDCD10	AU312342	5q	scaffold-ma5	4.00E-61	143	74805371	74805547
TLOC1	AU312335	5q	scaffold-ma5	2.00E-45	101	76109988	76110125
UCHL1[*]	BW999983	6p	scaffold-ma7	4.00E-89	210	33298090	33298407
GNAI2[*]	BW999984	6p	scaffold-ma2	2.00E-106	126	49893686	49893841
P4HB[*]	BW999985	6p	scaffold-ma2	2.00E-69	100	97717890	97718012

FLJ12571	AU312352	6q	scaffold-ma6	2.00E-46	117	46698606	46698752
RANGAP1	AU312313	6q	scaffold-ma6	7.00E-71	95	47795604	47795500
LDHB	BW999986	6q	scaffold-ma6	2.00E-60	117	69268248	69268418
SEC3L1	AU312345	7p	scaffold-ma7	3.00E-58	125	55644074	55643916
KIAA1109	AU312348	7q	scaffold-ma7	2.00E-60	124	30398905	30398711
RAP1GDS1	AU312351	7q	scaffold-ma7	2.00E-91	112	12141068	12140931
GAD2	BW999991	Zp	scaffold-Z	1.00E-109	136	17484512	17484336
WAC	AU312355	Zp	scaffold-Z	3.00E-93	209	16303681	16303947
KLF6[*]	BW999992	Zp	scaffold-ma2	1.00E-99	366	47130305	47130796
LOC90693[#]	BW999993	Zp	scaffold-ma7	4.00E-127	301	34444161	34444577
LOC90693[#]	BW999993	Zp	scaffold-Z	1.00E-107	291	34827559	34827182
TAX1BP1	AU312320	Zp	scaffold-Z	1.00E-86	141	36989995	36990174
RAB5A	BW999994	Zp	scaffold-Z	9.00E-94	166	40227424	40227215
CTNNB1	BW999995	Zcen	scaffold-Z	3.00E-129	275	49548885	49549226
AMPH	BW999996	Zcen	scaffold-Z	1.00E-66	101	55612836	55612955
TUBG1	BW999997	Zq	scaffold-Z	5.00E-89	116	17359265	17359113
GH1	BW999998	Zq	scaffold-Z	2.00E-115	179	77397011	77396727
MYST2	BW999999	Zq	scaffold-Z	6.00E-122	293	90785118	90784714
NEF3	BW999987	micro	scaffold-mi1	1.00E-102	352	13833430	13832942
ASB6	AU312340	micro	scaffold-mi7	1.00E-95	161	6270589	6270353
RPL12	BW999988	micro	scaffold-mi7	6.00E-67	95.5	7974658	7974542
FLJ25530	AU312336	micro	scaffold-mi1	4.00E-98	255	8157147	8156806
HSPA8[#]	BW999989	micro	scaffold-ma1	2.00E-124	236	20422342	20422662
HSPA8[#]	BW999989	micro	scaffold-mi1	3.00E-123	259	2089357	2089025
GLCE	AU312330	micro	scaffold-mi10	1.00E-79	234	24861	24577
POLG	AU312315	micro	scaffold-mi3	4.00E-97	116	10042696	10042845
LOC283820	AU312323	micro	scaffold-mi5	8.00E-71	116	3659851	3659708
PARN	AU312312	micro	scaffold-mi7	1.00E-66	73.9	12029447	12029361
ATRX	BW999990	micro	scaffold-mi4	3.00E-63	102	1268001	1268126

753

Supplemental Table S7. Details of mismatched cDNA markers from *Elaphe quadrivirgata* (Matsubara et al. 2006), their locations in *Crotalus* and *Anolis*, and notes on likelihood of misassembly based on synteny and intrachromosomal Hi-C.

Marker	***Elaphe* Chromosome**	***Crotalus* Scaffold**	***Anolis* Scaffold**	**Notes**
NOSIP	2q	scaffold-Z	6	Unique or erroneous original cDNA placement; *Anolis* and *Crotalus* synteny suggest correct placement in *Crotalus*; Hi-C data inconsistent with misassembly
ZNF326	3q	scaffold-ma2	4	Possible misassembly error; *Anolis* and *Elaphe* synteny suggest incorrect placement in Crotalus; Hi-C data consistent with regional misassembly
UCHL1	6p	scaffold-ma7	5	Hit to *Anolis* 5 is inconclusive because it is syntenic with snake chromosomes 6 & 7; Hi-C data inconsistent with misassembly
GNAI2	6p	scaffold-ma2	2	Unique or erroneous original cDNA placement; *Anolis* and *Crotalus* synteny suggest correct placement in *Crotalus*; Hi-C data inconsistent with misassembly
P4HB	6p	scaffold-ma2	2	Unique or erroneous original cDNA placement; *Anolis* and *Crotalus* synteny suggest correct placement in *Crotalus*; Hi-C data inconsistent with misassembly
KLF6	Zp	scaffold-ma2	6	Possible misassembly error; *Anolis* and *Elaphe* synteny suggest incorrect placement in *Crotalus*; Hi-C data inconsistent with misassembly

Supplemental Table S8. GC variation in windows of various sizes for 12 squamate species. Values for each species are measured as the standard deviation (SD) of GC content in all sampled windows of a given size. Information for 5, 20, and 80 kb windows are also presented in Fig. 1c. Missing data (i.e., window sizes that were too large and contained greater than the threshold allowed missing data) are denoted with '-'.

Window Size (bp)	*Gekko japonicus*	*Eublepharis macularius*	*Ophisaurus gracilis*	*Shinisaurus crocodilurus*	*Pogona vitticeps*	*Anolis carolinensis*
5,000	0.039295606	0.037140406	0.037038224	0.03488877	0.03681681	0.032312269
20,000	0.028980944	0.027338004	0.029217483	0.027425317	0.030930264	0.021209
40,000	0.025219459	0.024838347	0.027141528	0.025322106	0.029367252	0.017608402
80,000	0.021385708	0.023326607	0.025558162	0.023843432	0.028238318	0.015121097
160,000	0.01811246	0.022646783	0.024536212	0.022632678	0.027330318	0.013089382
240,000	-	0.022203903	0.023356372	0.021943776	0.026943855	0.012088733
320,000	-	0.022121291	0.022899173	0.021312719	0.026617904	0.011287772

Window Size (bp)	*Boa constrictor*	*Python molurus*	*Ophiophagus hannah*	*Thamnophis sirtalis*	*Deinagkistrodon acutus*	*Crotalus viridis*
5,000	0.043942864	0.042024505	0.040098669	0.047076022	0.047062019	0.041210929
20,000	0.034934365	0.035837726	0.031894398	0.037865804	0.03882085	0.032232558
40,000	0.030576918	0.033337717	0.028952912	0.03429097	0.036517713	0.029884634
80,000	0.023292703	0.030197592	0.026685436	0.031202717	0.034964163	0.0281043
160,000	0.014736549	0.02736241	0.024597185	0.02894796	0.033486765	0.026806291
240,000	-	0.024725646	0.023968494	0.026250057	0.032562166	0.02616041
320,000	-	0.023707617	0.023468328	0.024606171	0.031784231	0.025840409

Supplemental Table S9. Details of Illumina Nextera resequencing and RNAseq libraries used for comparative female/male read coverage across the rattlesnake genome and sex-specific gene expression analyses. Raw read data are available on NCBI under accession PRJNA476794.

Library Type	**Read Length**	**Sample ID**	**Tissue**	**Sex**	**Number of Mapped Reads**
Illumina Nextera	150 bp paired end	CV0007	Liver	Male	20,279,801
Illumina Nextera	150 bp paired end	CV0011	Liver	Female	4,975,491
RNAseq	100 bp paired end	Cv3	Liver	Female	3,774,322
RNAseq	100 bp paired end	Cv8	Liver	Female	3,680,195
RNAseq	100 bp paired end	Cv3	Kidney	Female	3,256,208
RNAseq	100 bp paired end	Cv3	Kidney	Female	4,565,008
RNAseq	100 bp paired end	Cv5	Liver	Male	3,330,125
RNAseq	100 bp paired end	Cv6	Liver	Male	3,837,264
RNAseq	100 bp paired end	Cv5	Kidney	Male	3,729,811
RNAseq	100 bp paired end	Cv6	Kidney	Male	4,673,928

Supplemental Table S10. Representative sequences for known snake venom gene families used to annotate venom genes in the rattlesnake genome.

Gene Family	Accession	Sequence Type	Species
5'Nucleotidase	AK291667.1	mRNA	*Homo sapiens*
Acetylcholinesterase	U54591.1	mRNA	*Bungarus fasciatus*
AVItoxin	EU195459.1	mRNA	*Varanus komodoensis*
C-type Lectin	JF895761.1	mRNA	*Crotalus oreganus helleri*
Cobra Venom Factor	U09969.2	mRNA	*Naja kaouthia*
CRISp (cysteine-rich secretory protein)	HQ414088.1	mRNA	*Crotalus adamanteus*
Cystatin	FJ411289.1	mRNA	*Naja kaouthia*
Extendin	EU790960.1	mRNA	*Heloderma suspectum*
Exonuclease	XM_015826835.1	mRNA	*Protobothrops mucrosquamatus*
Hyaluronidase	HQ414098.1	mRNA	*Crotalus adamanteus*
LAAO (L-amino acid oxidase)	HQ414099.1	mRNA	*Crotalus adamanteus*
SVMP I (class I snake venom metalloproteinase)	HM443635.1	mRNA	*Bothrops neuwiedi*
SVMP II (class II snake venom metalloproteinase)	HM443637.1	mRNA	*Bothrops neuwiedi*
SVMP III (class III snake venom metalloproteinase)	HM443632.1	mRNA	*Bothrops neuwiedi*
Nerve growth factor	AF306533.1	mRNA	*Crotalus durissus terrificus*
Phosphodiesterase	HQ414102.1	mRNA	*Crotalus adamanteus*
PLA2-I (vipers)	AF403134.1	mRNA	*Crotalus viridis viridis*
PLA2-II (elapids)	GU190815.1	mRNA	*Bungarus flaviceps*
Sarafotoxin	L07528.1	mRNA	*Atractaspis engaddensis*
Serine Proteinase	HQ414121.1	mRNA	*Crotalus adamanteus*
3FTX (Three-finger Toxin)	DQ273582.1	mRNA	*Ophiophagus hannah*
Veficolin	GU065323.1	mRNA	*Cerberus rynchops*
VEGF (Vascular Endothelial Growth Factor)	AB848141.1	mRNA	*Protobothrops mucrosquamatus*
Vespryn	EU401840.1	mRNA	*Oxyuranus scutellatus*
Waprin	EU401843.1	mRNA	*Oxyuranus scutellatus*
Kunitz (serine peptidase inhibitor, Kunitz type)	JU173666.1	mRNA	*Crotalus adamanteus*
Thrombin-like (thrombin-like venom gland enzyme)	AJ001209.1	mRNA	*Deinagkistrodon acutus*
Ficolin	GBUG01000048.1	mRNA	*Echis coloratus*
Disintegrin	AJ131345.1	mRNA	*Deinagkistrodon acutus*
FactorV (venom coagulation factor V)	XM_015815922.1	mRNA	*Protobothrops mucrosquamatus*
FactorX	XM_015819885.1	mRNA	*Protobothrops mucrosquamatus*
Prokineticin	XM_015822870.1	mRNA	*Protobothrops mucrosquamatus*
Ohanin (ohanin-like)	XM_015818414.1	mRNA	*Protobothrops mucrosquamatus*
Complement C3 (Cadam VF)	JU173742.1	mRNA	*Crotalus adamanteus*

Crotasin	AF250212.1	mRNA	*Crotalus durissus terrificus*
Endothelin	XM_015810852.1	mRNA	*Protobothrops mucrosquamatus*
Kallikrein	GALC01000005.1	mRNA	*Crotalus oreganus helleri*
Lynx1 (Ly6/neurotoxin 1)	XM_014066791.1	mRNA	*Thamnophis sirtalis*
Natriuretic Peptide (bradykinin potentiating peptide and C-type natriuretic peptide precursor isoform 2)	AF308594.2	mRNA	*Crotalus durissus terrificus*
sPla/ryanodine receptor	XM_015823102.1	mRNA	*Protobothrops mucrosquamatus*
WAP four-disulfide core domain protein 5 (Whey Acidic Protein/secretory leuki proteinase inhibitor)	XM_015822353.1	mRNA	*Protobothrops mucrosquamatus*
Myotoxin	HQ414100.1	mRNA	*Crotalus adamanteus*
PLA2	APD70899.1	protein	*Crotalus atrox*
SVMP	Q90282.1	protein	*Crotalus atrox*
Serine Proteinase	F8S114.1	protein	*Crotalus adamanteus*

772

773 **Supplemental Table S11.** Annotated venom gene homologs in the Prairie rattlesnake Genome. Genes
774 were annotated using materials detailed in Supplemental Table 8.

Venom Gene Family	Scaffold	Start Position (bp)	End Position (bp)
3-Finger toxin	scaffold-ma1	103004868	103021927
3-Finger toxin	scaffold-ma1	102999393	103000958
5' Nucleotidase	scaffold-ma5	46133017	46179118
5' Nucleotidase	scaffold-ma6	55711914	55732365
5' Nucleotidase	scaffold-mi1	18004217	18021456
5' Nucleotidase	scaffold-ma2	45090212	45121335
5' Nucleotidase	scaffold-ma2	134237148	134264183
Acetylcholinesterase	scaffold-ma2	4047955	4053281
Acetylcholinesterase	scaffold-ma2	3948506	3952373
Acetylcholinesterase	scaffold-ma2	4016363	4018146
Acetylcholinesterase	scaffold-ma2	4026170	4045822
Acetylcholinesterase	scaffold-ma5	73971094	73976212
Acetylcholinesterase	scaffold-ma5	74015346	74036663
Acetylcholinesterase	scaffold-un210	16032	17552
Bradykinin potentiating and natriuretic peptide	scaffold-un187	22386	23524
C-type lectin	scaffold-mi5	3276042	3284747
C-type lectin	scaffold-mi5	11650747	11653723
C-type lectin	scaffold-Z	21883578	21895509
C-type lectin	scaffold-Z	21706900	21776775
C-type lectin	scaffold-Z	21786524	21797211
C-type lectin	scaffold-Z	108214710	108236532
Cysteine-rich secretory protein	scaffold-ma1	169434958	169437996
Cysteine-rich secretory protein	scaffold-ma1	169423774	169434684
Cysteine-rich secretory protein	scaffold-ma3	25391938	25416947
Cysteine-rich secretory protein	scaffold-mi6	1021447	1040191
Exonuclease	scaffold-mi7	8097114	8103411
Exonuclease	scaffold-ma1	5804894	5842638
Exonuclease	scaffold-mi3	10271502	10274220
Exonuclease	scaffold-ma6	12590208	12591465
Factor V	scaffold-mi4	8493826	8518402
Factor V	scaffold-mi4	8479637	8493564
Factor V	scaffold-ma4	81074882	81113119
Glutaminyl cyclase	scaffold-ma1	256551622	256564040
Glutaminyl cyclase	scaffold-mi7	5091107	5094268
Hyaluronidase	scaffold-ma6	14952252	14955850
Hyaluronidase	scaffold-ma2	45901201	45920587
Hyaluronidase	scaffold-ma2	49137409	49145188
Hyaluronidase	scaffold-ma2	49106981	49118469
Kunitz peptide	scaffold-mi7	3590975	3597607
Kunitz peptide	scaffold-mi8	4992795	5002390

L-amino acid oxidase	scaffold-ma4	56914906	56948498
L-amino acid oxidase	scaffold-ma4	85461961	85468906
L-amino acid oxidase	scaffold-ma2	4658599	4661642
L-amino acid oxidase	scaffold-ma2	4654769	4658293
Myotoxin/crotamine	scaffold-ma1	289328153	289328605
Nerve growth factor	scaffold-Z	93342025	93347811
Nerve growth factor	scaffold-ma1	76711308	76727703
Phospholipase A2	scaffold-mi7	3019970	3021876
Phospholipase A2	scaffold-mi7	3027607	3029199
Phospholipase A2	scaffold-mi7	3031464	3033348
Phospholipase A2	scaffold-mi7	3037103	3038488
Phospholipase A2	scaffold-mi7	3042118	3043697
Serine Proteinase	scaffold-mi2	8569773	8575182
Serine Proteinase	scaffold-mi2	8588278	8593660
Serine Proteinase	scaffold-mi2	8628274	8636651
Serine Proteinase	scaffold-mi2	8664603	8670797
Serine Proteinase	scaffold-mi2	8739986	8745649
Serine Proteinase	scaffold-mi2	8752578	8759324
Serine Proteinase	scaffold-mi2	8864675	8879153
Serine Proteinase	scaffold-mi2	8937526	8947481
Serine Proteinase	scaffold-mi2	8960028	8980478
Snake venom metalloproteinase	scaffold-mi1	13901629	14014239
Snake venom metalloproteinase	scaffold-mi1	14022082	14075370
Snake venom metalloproteinase	scaffold-mi1	14091987	14112667
Snake venom metalloproteinase	scaffold-mi1	14147865	14170405
Snake venom metalloproteinase	scaffold-mi1	14174872	14190142
Snake venom metalloproteinase	scaffold-mi1	14211673	14242249
Snake venom metalloproteinase	scaffold-mi1	14248933	14272689
Snake venom metalloproteinase	scaffold-mi1	14281564	14300774
Snake venom metalloproteinase	scaffold-mi1	14368422	14393313
Snake venom metalloproteinase	scaffold-mi1	14401627	14424637
Snake venom metalloproteinase	scaffold-mi1	14310844	14338336
Veficolin/Ficolin	scaffold-mi7	5271880	5282014
Veficolin/Ficolin	scaffold-ma3	179788950	179790745
Veficolin/Ficolin	scaffold-ma1	232337083	232340714
Veficolin/Ficolin	scaffold-ma1	232312034	232335439
Vascular endothelial growth factor	scaffold-ma7	40288572	40327884
Vascular endothelial growth factor	scaffold-ma1	40733075	40747358
Vascular endothelial growth factor	scaffold-ma1	260248287	260272500
Venom Factor	scaffold-Z	79798672	79803249
Venom Factor	scaffold-Z	79749464	79761456
Venom Factor	scaffold-ma2	1573588	1616446
Venom Factor	scaffold-ma2	137559964	137560374

Venom Factor	scaffold-ma2	137553669	137558461
Venom Factor	scaffold-ma2	137623562	137648584
Venom Factor	scaffold-ma2	137651285	137653877
Venom Factor	scaffold-ma2	137710627	137728987
Venom Factor	scaffold-ma2	137753804	137775039
Venom Factor	scaffold-ma2	137735629	137741352
Vespryn/Ohanin	scaffold-ma2	4377779	4385668
Vespryn/Ohanin	scaffold-ma2	109834300	109838076
Waprin	scaffold-ma1	204655764	204666466

775

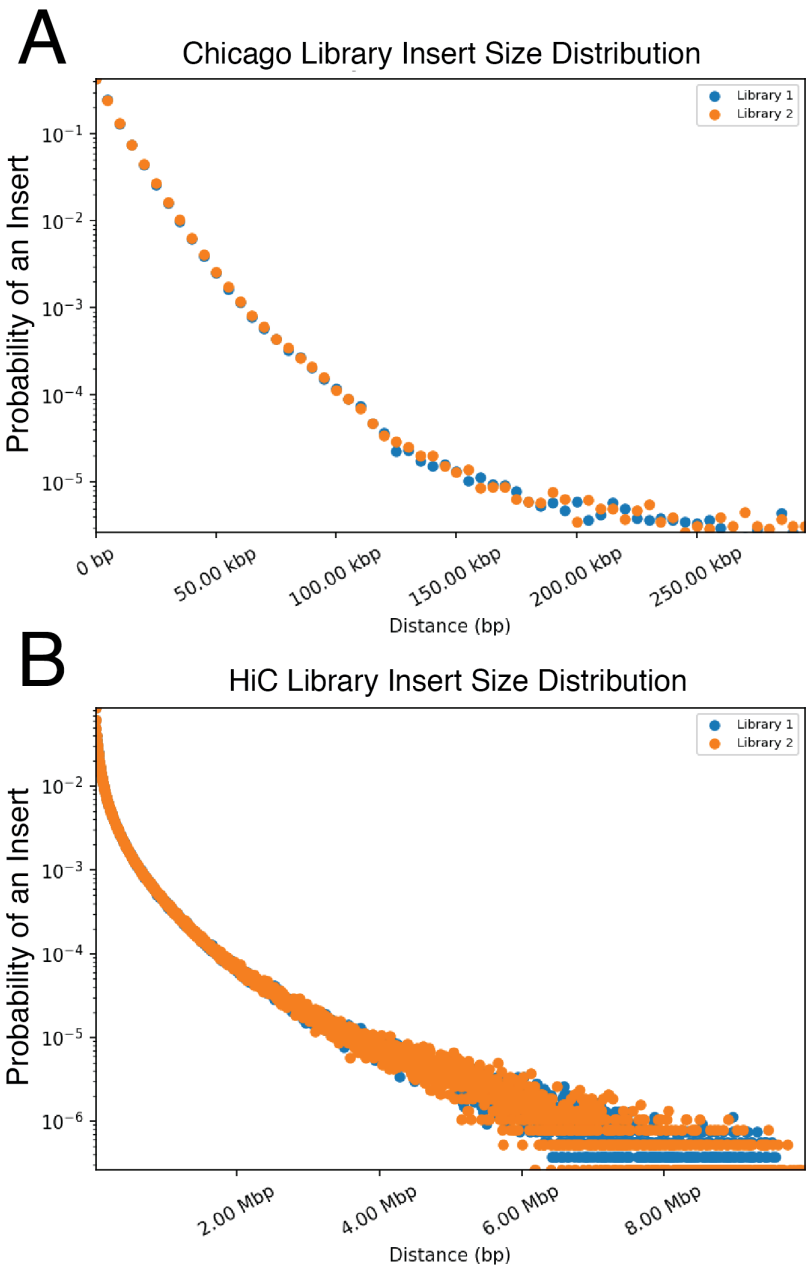
Supplemental Table S12. Transcription factors significantly upregulated in the venom gland. Mean distances summarize the distribution of distances between gene venom genes and non-venom genes and the nearest predicted binding site of each transcription factor. No position weight matrix for *NCOA2* was available for a close relative to the rattlesnake, and the *NFI* family transcription factors have a conserved binding motif, and are summarized together under *NFIA*. P-values are from *t*-test comparisons of distance distributions.

Gene ID	Rattlesnake Gene Detail	Mean Distance to Venom Gene (bp)	Mean Distance to Non-venom Gene (bp)	p-value
ATF6	augustus_masked-scaffold-ma3-processed-gene-300.3	421,305.1	595,006.2	0.002793
ELF5	maker-scaffold-ma1-augustus-gene-235.5	1,121.3	1,203.9	0.7953
FOXC2	augustus_masked-scaffold-mi6-processed-gene-2.1	202,416.2	251,898.5	0.02967
CREB3L2	maker-scaffold-ma6-augustus-gene-195.2	32,227.9	29,708.3	0.5558
GRHL1	maker-scaffold-ma1-augustus-gene-601.8	78,954.0	86,147.0	0.4343
NCOA2	maker-scaffold-ma3-augustus-gene-89.6	-	-	-
NFIA	maker-scaffold-ma3-augustus-gene-414.2	336,765.8	328,556.3	0.7968
NFIB	maker-scaffold-ma2-augustus-gene-569.3	-	-	-
NFIB	maker-scaffold-ma2-augustus-gene-569.2	-	-	-
NFIX	maker-scaffold-ma2-augustus-gene-473.3	-	-	-
NR4A2	maker-scaffold-ma1-augustus-gene-428.4	100,375.5	92,292.3	0.492
SREBF2	maker-scaffold-ma6-augustus-gene-158.15	306,901.1	328,081.4	0.4302

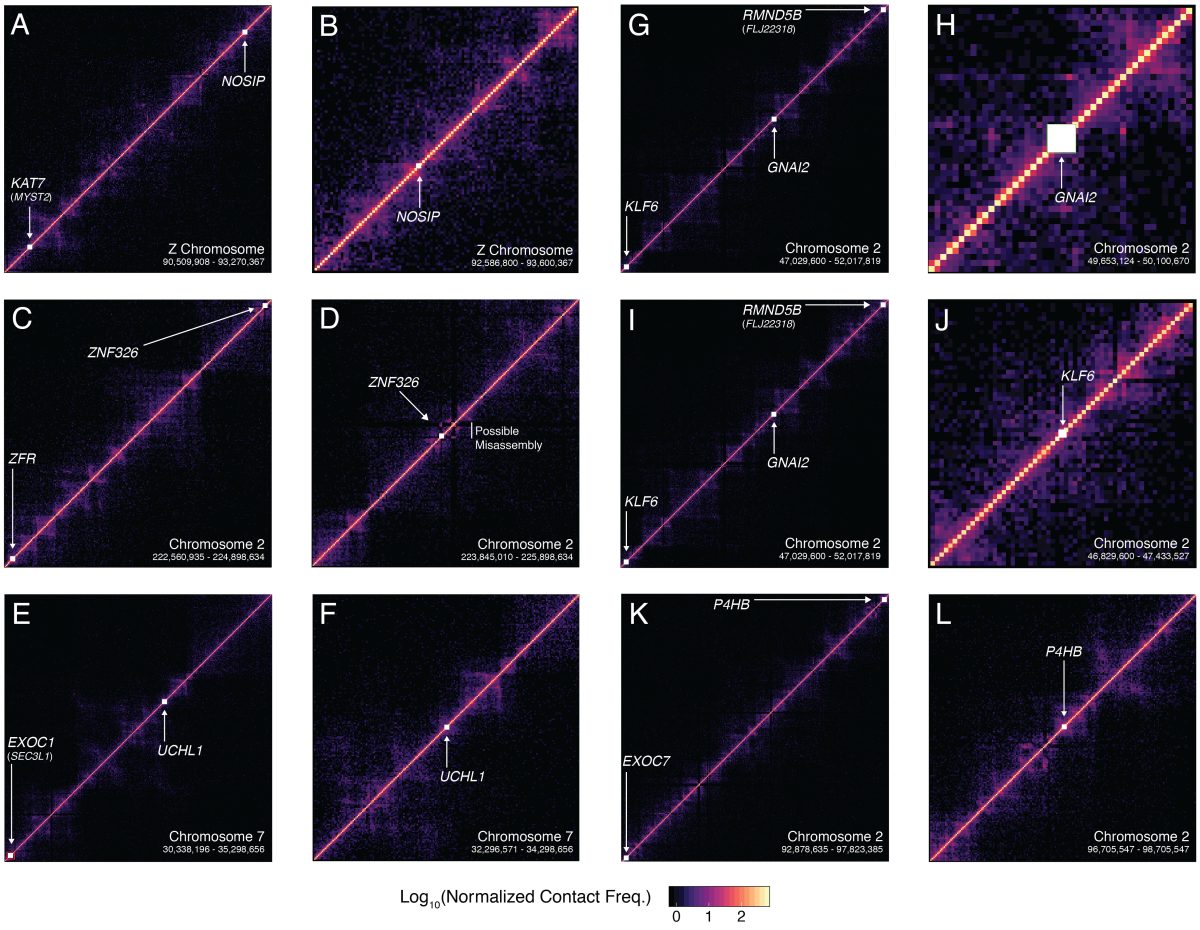
Supplemental Table S13. Density of predicted *GRHL1* and *NFI* binding sites within given intervals of venom genes and all nonvenom genes. P-values are reported from Fisher's exact tests, which compared the number of predicted binding sites by the total length of sequenced searched between venom and nonvenom gene sets.

Transcription Factor	Interval (kb)	Venom Gene Density	Nonvenom Gene Density	p-value
GRHL1	100 kb	7.44E-06	8.03E-06	0.4022
GRHL1	50 kb	1.49E-05	1.74E-05	0.2127
GRHL1	10 kb	3.00E-05	2.78E-05	0.6875
GRHL1	5 kb	4.44E-05	3.97E-05	0.554
NFI	Promoter (1kb)	1.80E-03	1.43E-03	0.1305

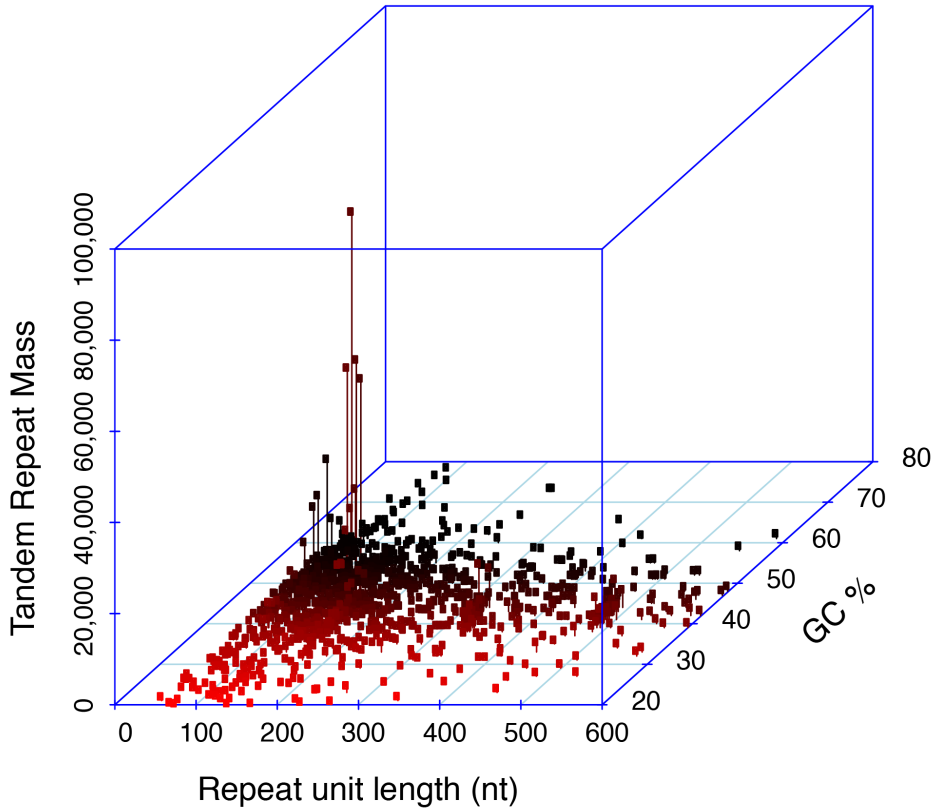
3. SUPPLEMENTAL FIGURES



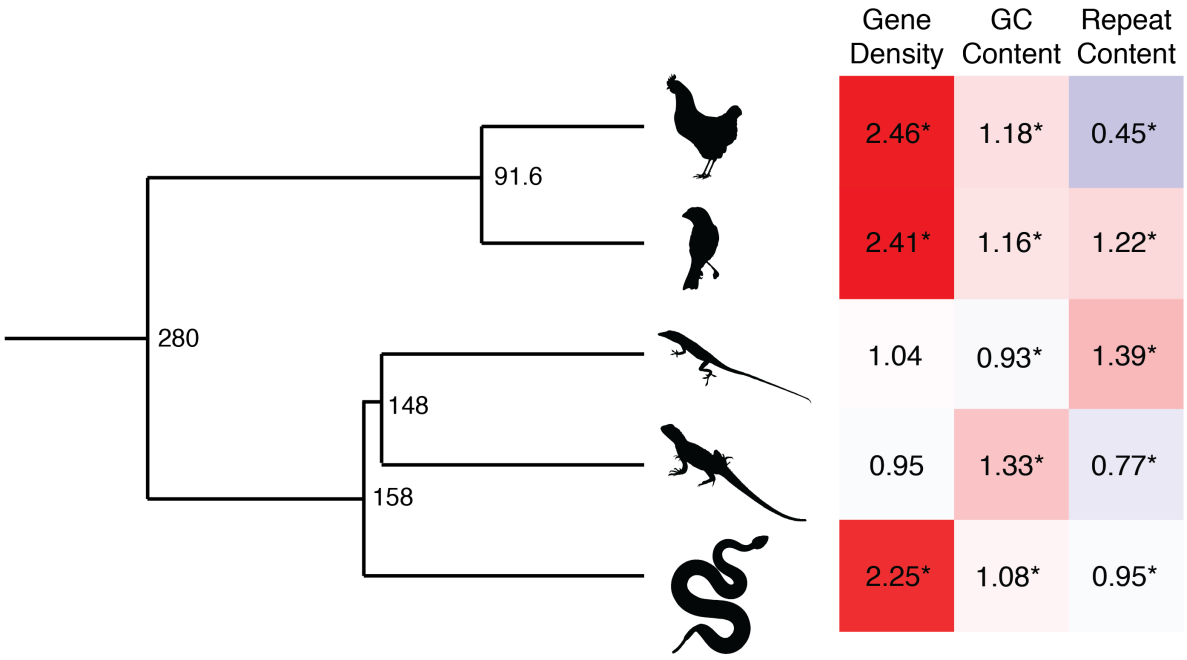
Supplemental Figure S1. Insert size probability distributions used in the Dovetail Genomics HiRise assembly method from paired Chicago (A) and Hi-C (B) datasets.



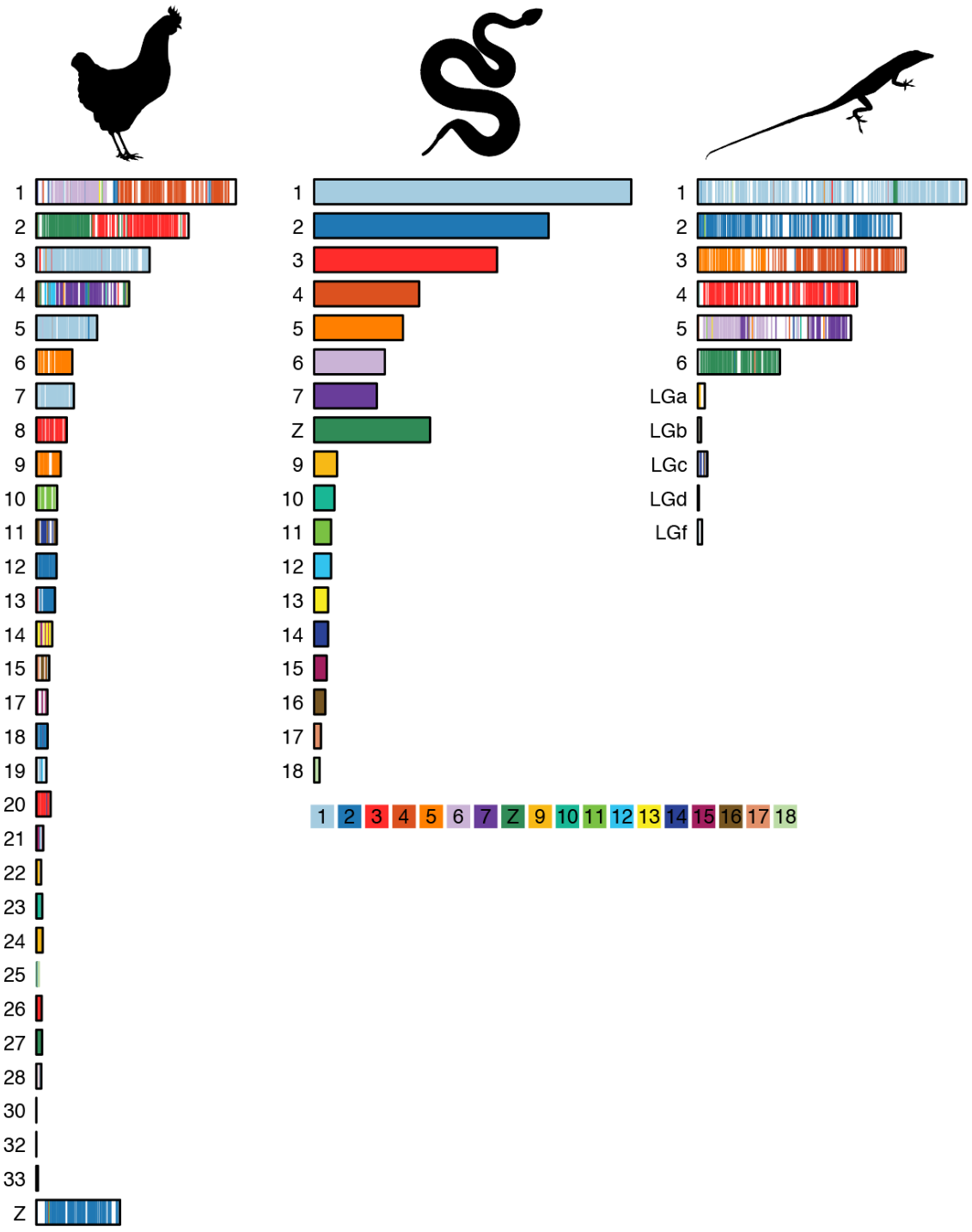
Supplemental Figure S2. Heatmaps of Log_{10} normalized intrachromosomal Hi-C contact frequencies around mapping locations for cDNA markers from *Elaphe quadrivirgata* (Matsubara et al. 2006) in the rattlesnake genome. For each of the six markers, panels showing the contact frequencies between the focal marker and its nearest confirmed marker (see Supplemental Methods), and panels zoomed to the region immediately around the focal marker are shown: *NOSIP* (A-B), *ZNF326* (C-D), *UCHL1* (E-F), *GNAI2* (G-H), *KLF6* (I-J), and *P4HB* (K-L). Marker locations are shown with white squares, and chromosomal coordinates for each panel are shown in the bottom right corner. The location of a potential misassembly error is shown in panel D.



Supplemental Figure S3. Centromeric tandem repeat motif characterized using tandem repeats finder. Analysis of high frequency tandem repeats identified a 164-mer with high relative GC to the genomic background. The y-axis, tandem repeat mass, represents the relative abundance of tandem repeats of a given unit length and GC content.

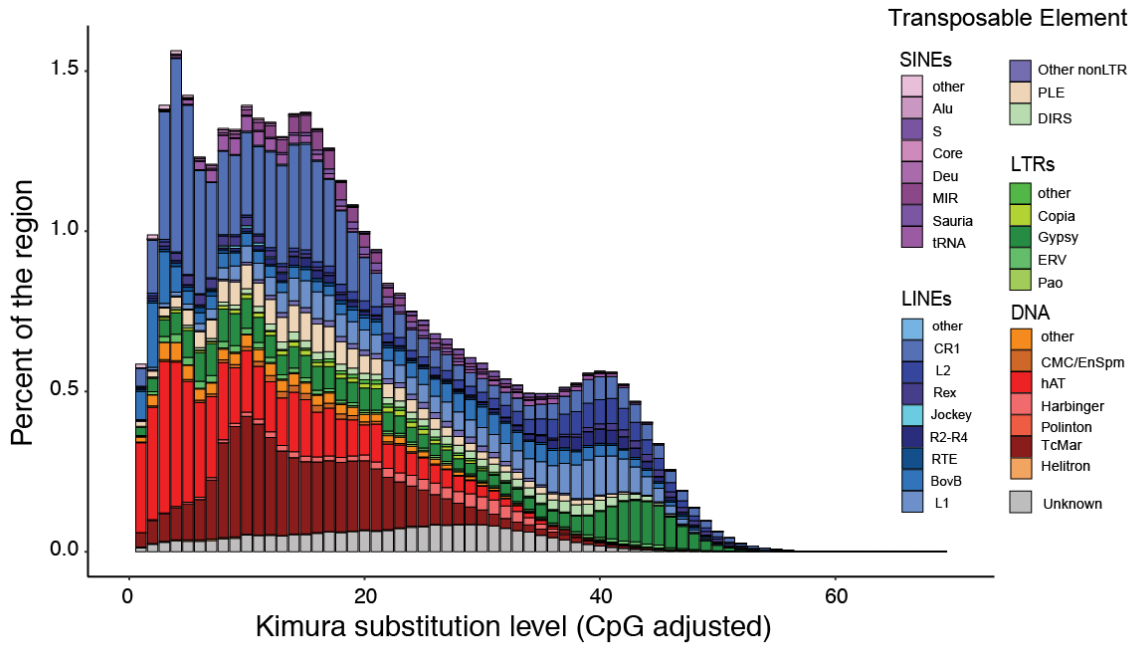


Supplemental Figure S4. Evolutionary patterns of genomic features of microchromosomes among reptiles. Values at nodes on the phylogenetic tree represent the node age in millions of years, and were obtained using median estimates from TimeTree. The heatmap to the right represents the relative abundance of a given measure on microchromosomes versus macrochromosomes within each species (blue values represent greater abundance on macrochromosomes and red values represent greater abundance on microchromosomes). Values in each heatmap cell equal the ratio of each measure on microchromosomes/macrochromosomes, and values with asterisks represent significant differences between microchromosomes and macrochromosomes.

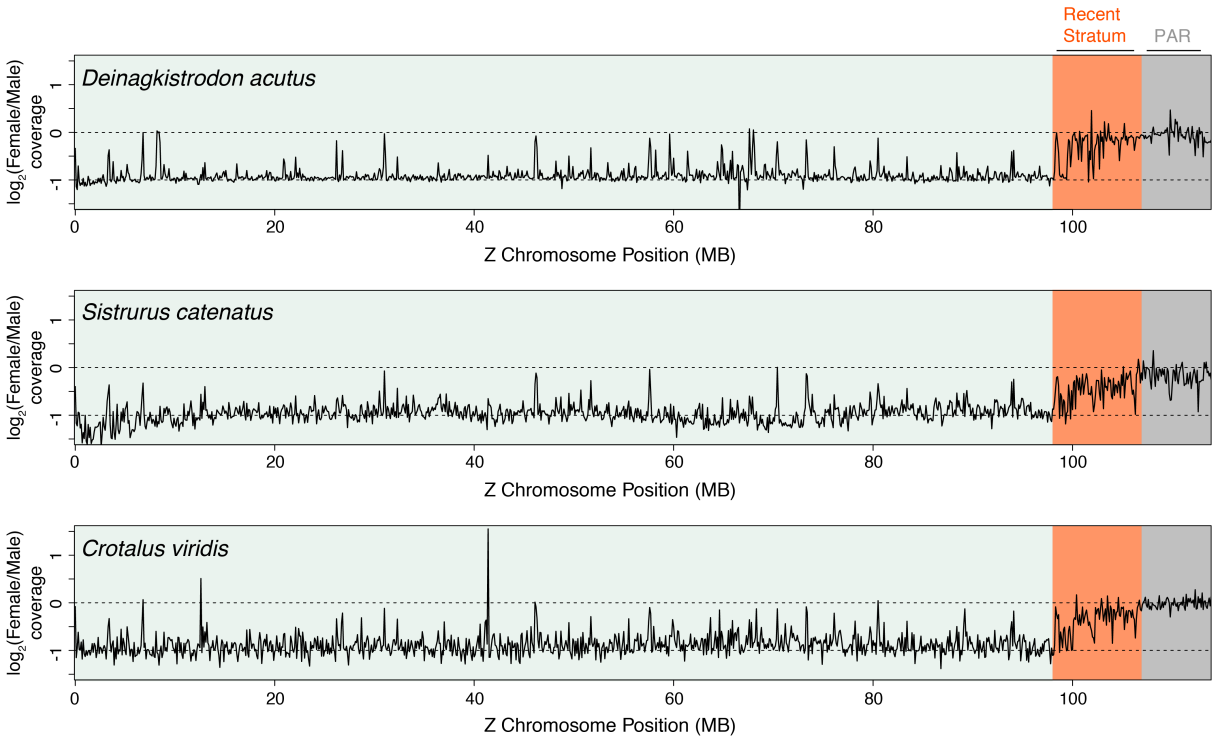


816

817 **Supplemental Figure S5.** Results of gene-based synteny analyses between the chicken (left), rattlesnake
818 (center), and anole lizard (right). Chromosome numbers for each species are shown to the left of the
819 chromosome ideograms, which are scaled by total length. Colors for chromosome paints are based on the
820 rattlesnake genome.



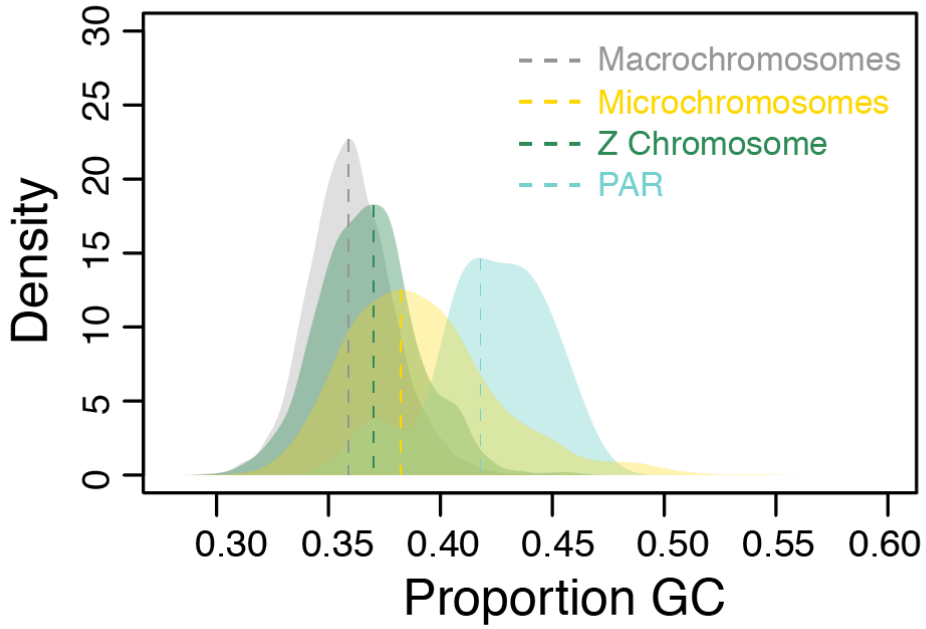
Supplemental Figure S6. Genomic repeat element abundance at a range of relative age values. Age is measured using the Kimura substitution level of transposable elements when compared to a consensus sequence. Transposable element types are colored according to the legend at the right.



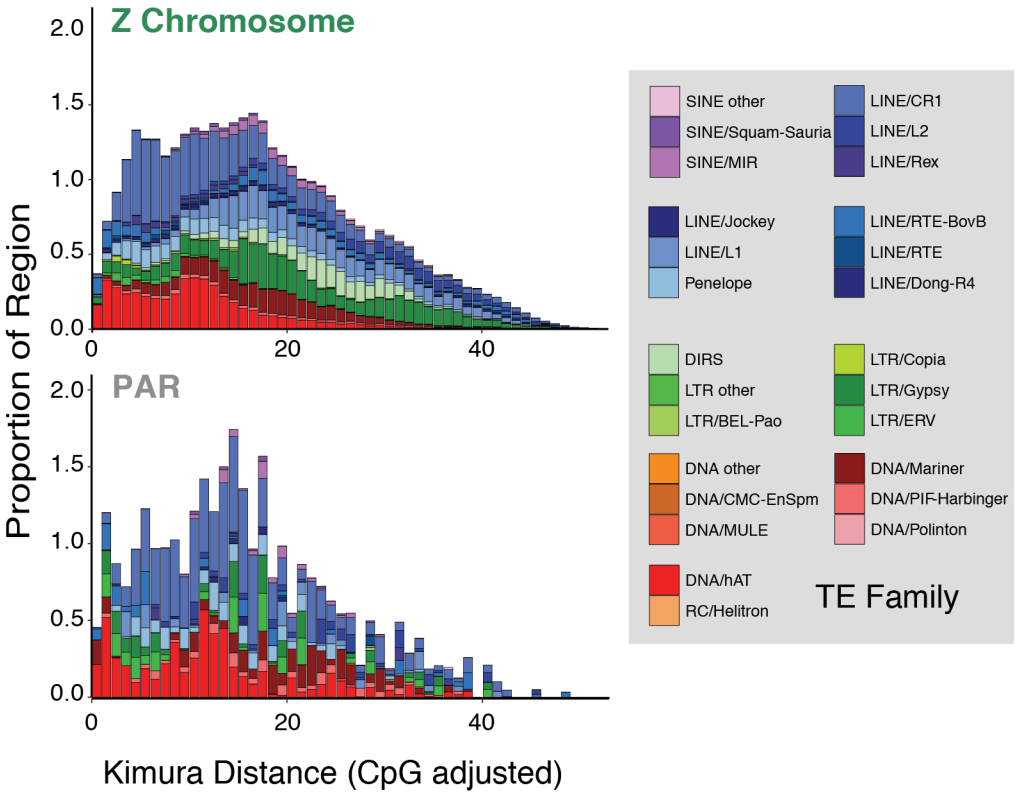
825

826 **Supplemental Figure S7.** Log_2 normalized female/male coverage ratio of pitviper species (Five Pace
827 viper (*Deinagkistrodon acutus*), Pygmy Rattlesnake (*Sistrurus catenatus*), and Prairie Rattlesnake
828 (*Crotalus viridis*), when mapped to the prairie rattlesnake reference genome. The dashed line at zero
829 represents the normalized coverage expectation for diploid loci, and the dashed line at -1 represents the
830 expectation of a hemizygous locus. Ratios are shown show values for each 100 kb window in a sliding
831 window analysis of coverage. Colored backgrounds depict the major regions discussed in the Main Text.

832

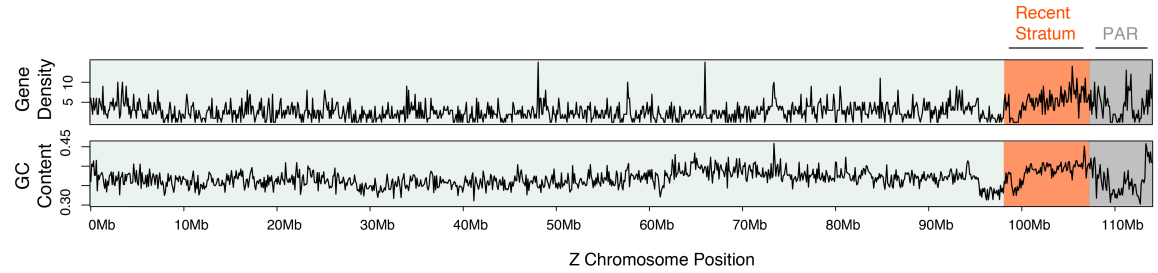


833 **Supplemental Figure S8.** Density distributions of GC content across Prairie Rattlesnake chromosomes,
834 showing specific distributions of macrochromosomes, microchromosomes, the Z Chromosome, and the
835 pseudoautosomal region (PAR) of the sex chromosomes, specifically.



836

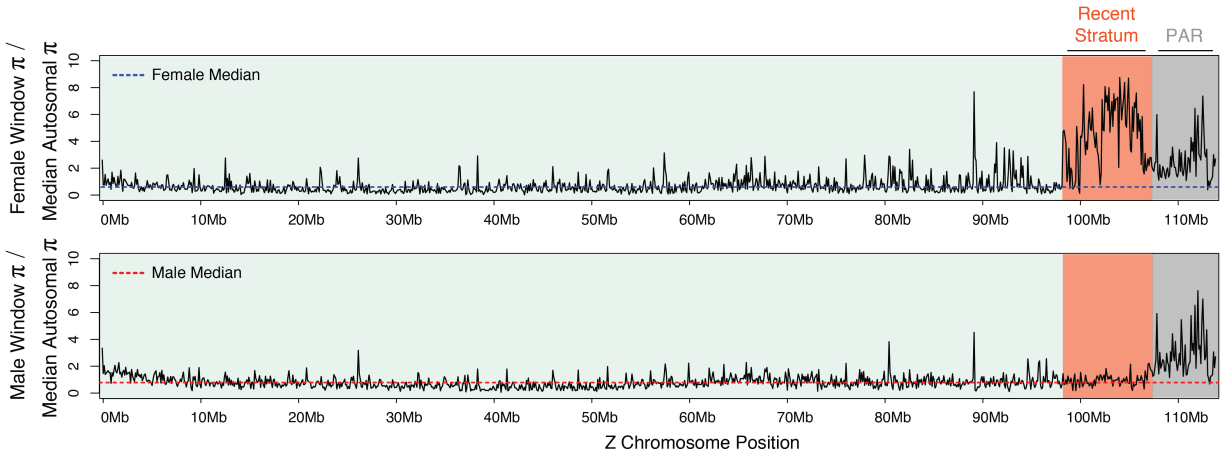
837 **Supplemental Figure S9.** Comparative age distributions of proportions of transposable elements (TEs)
838 across the Z Chromosome (upper) and the pseudoautosomal region (PAR; lower) of the rattlesnake Z
839 Chromosome. TE families contributing to proportions in each region at each age are shown at the right.



840

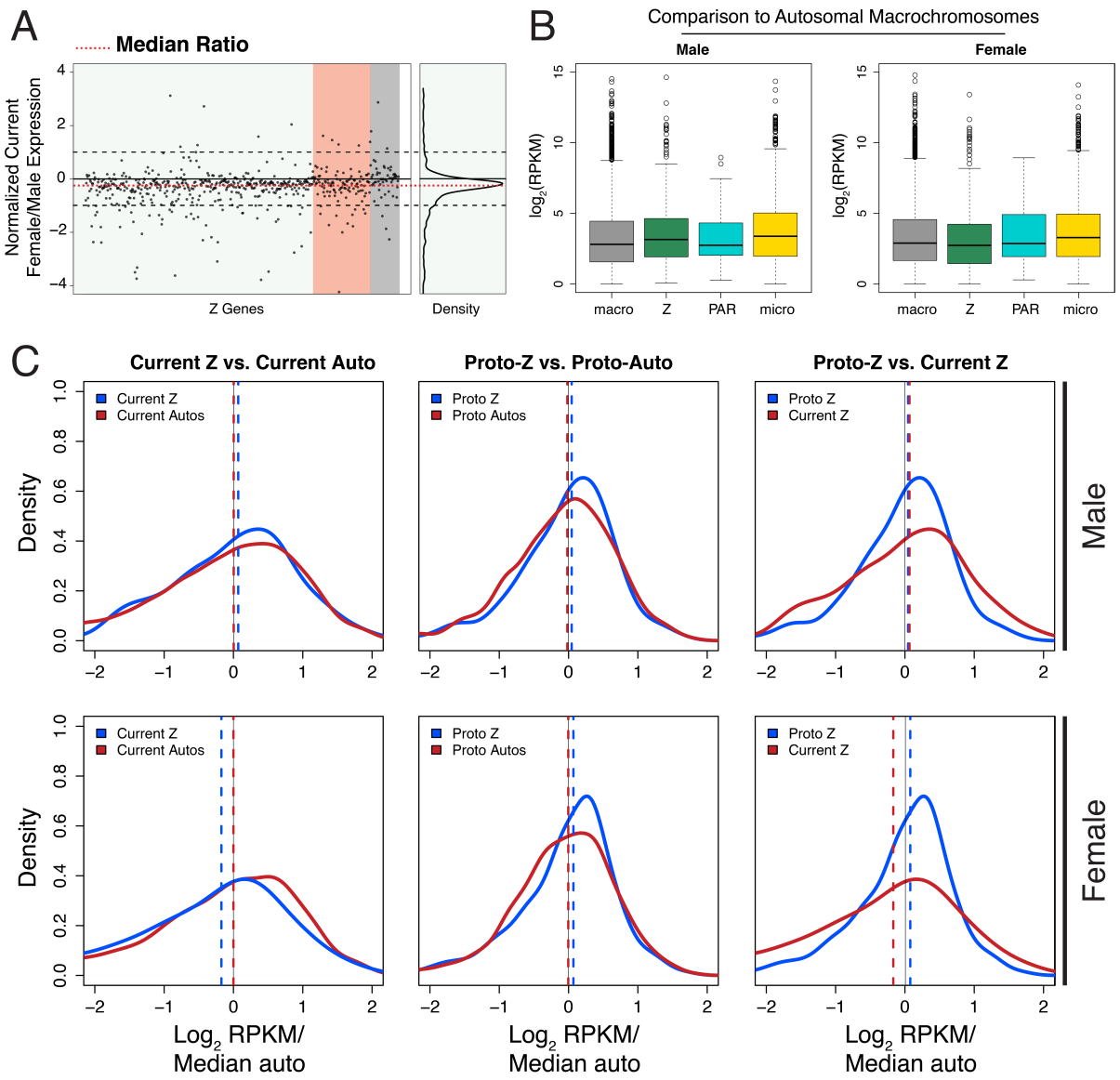
841 **Supplemental Figure S10.** 100 kb windowed scans of gene density (measured as number of genes per
842 window) and GC content (i.e., proportion of GC bases within each window) across the Z Chromosome of
843 the prairie rattlesnake. The regions on the Z correspond to those demarcated in Fig. 2 in the main text.

844

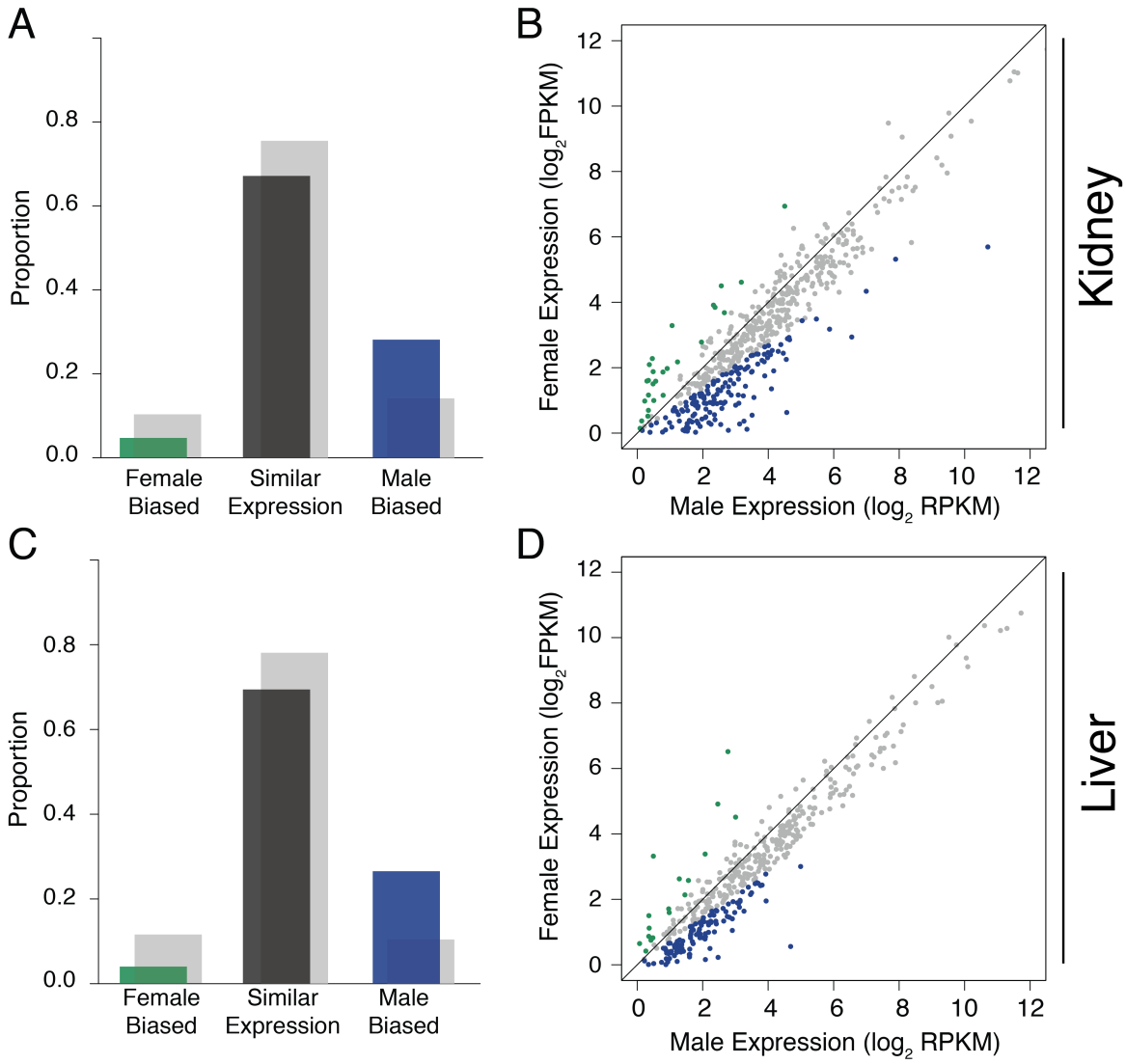


845

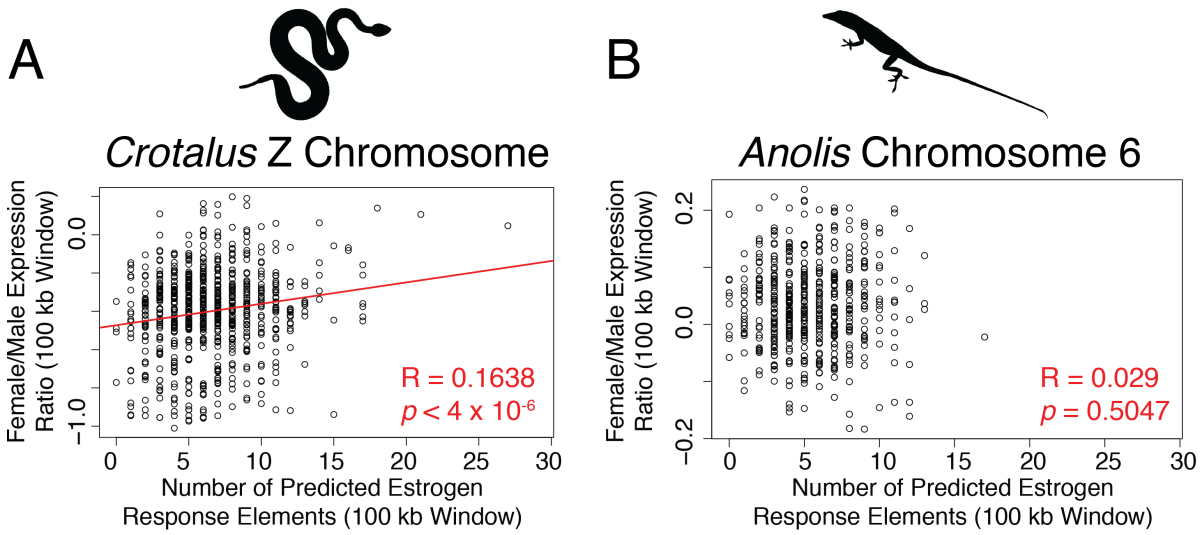
846 **Supplemental Figure S11.** 100 kb windowed scans of nucleotide diversity (π) for each sex across the Z
847 Chromosome of the Prairie Rattlesnake. The regions on the Z correspond to those demarcated in Fig. 2 in
848 the main text. Blue and red dashed lines correspond to median female and male values, respectively.



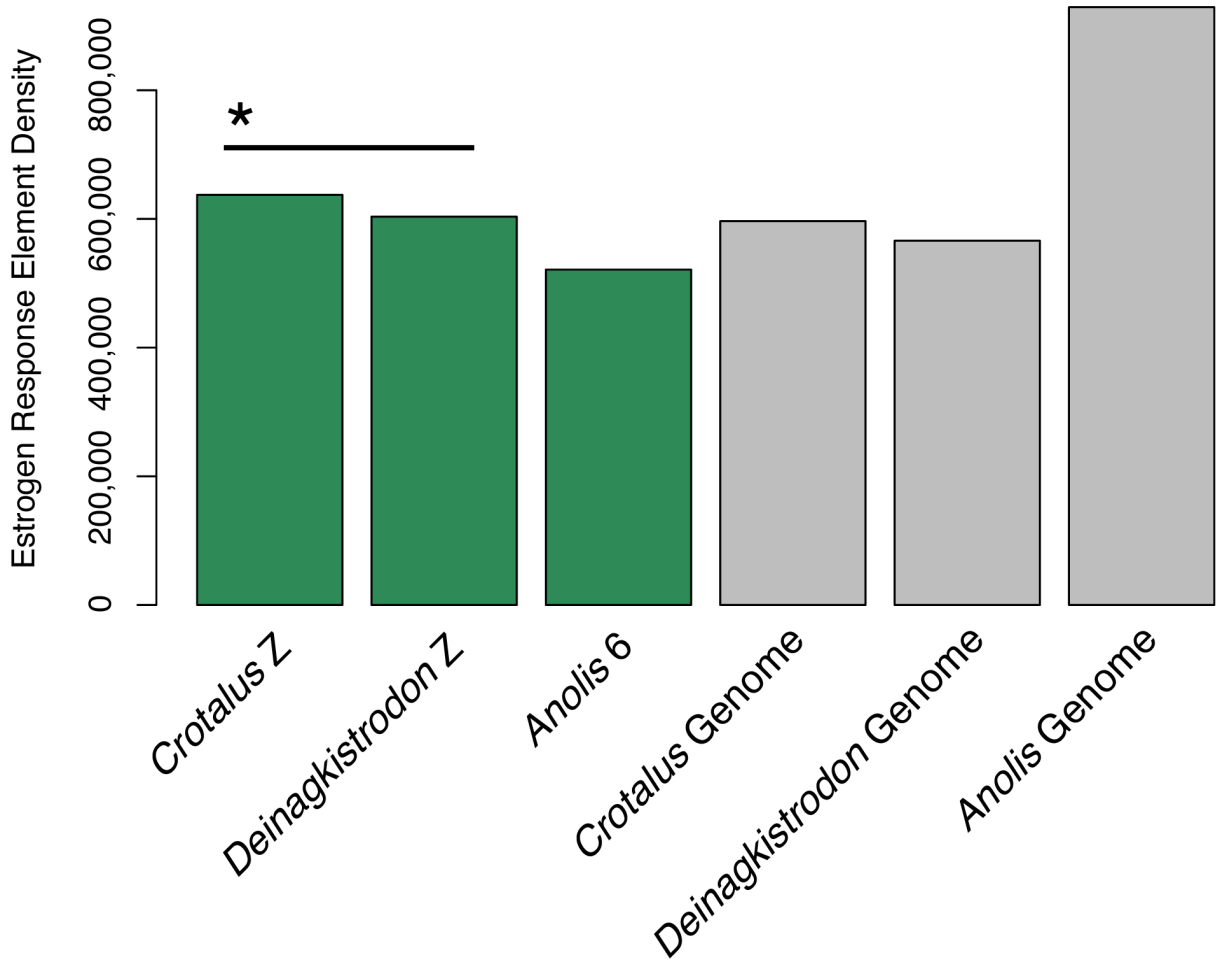
Supplemental Figure S12. Patterns of liver gene expression in females and males across the Z chromosome. (*A*) Log_2 normalized female/male gene expression per gene (black dots) across the Z. The red dashed line is the median ratio, and relative density is shown to the right. (*B*) Gene expression (log_2 RPKM) distributions for male and female across macrochromosomes, Z chromosome, the PAR, and microchromosomes. (*C*) Density plots of current and inferred ancestral patterns of gene expression (log_2 RPKM) in male and female, respectively. Dashed lines represent the median of each distribution.



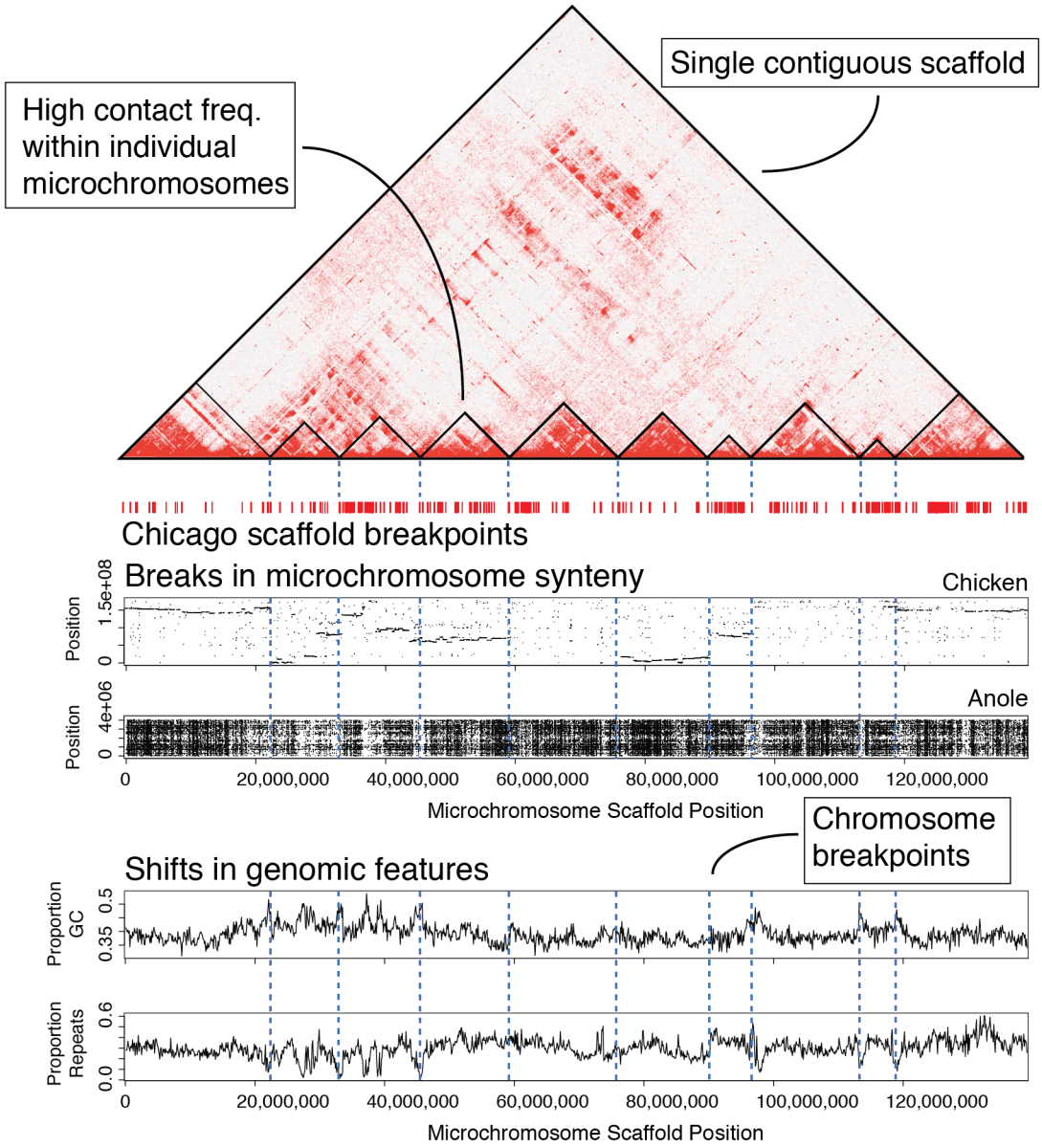
Supplemental Figure S13. Proportions of genes on the Z that exhibit female-biased (i.e., $\log_2$ female/male RPKM > 0.5; green bars), unbiased (dark grey bars), and male-biased (i.e., $\log_2$ female/male RPKM < -0.5; blue bars) expression in the kidney (*A*) and liver (*C*). Light grey bars in the background represent proportions of autosomal genes meeting the same criteria. Scatterplots of male versus female gene expression ($\log_2$ RPKM), with points showing expression of male-biased (blue), unbiased (grey), and female-biased (green) genes for kidney (*B*) and liver (*D*).



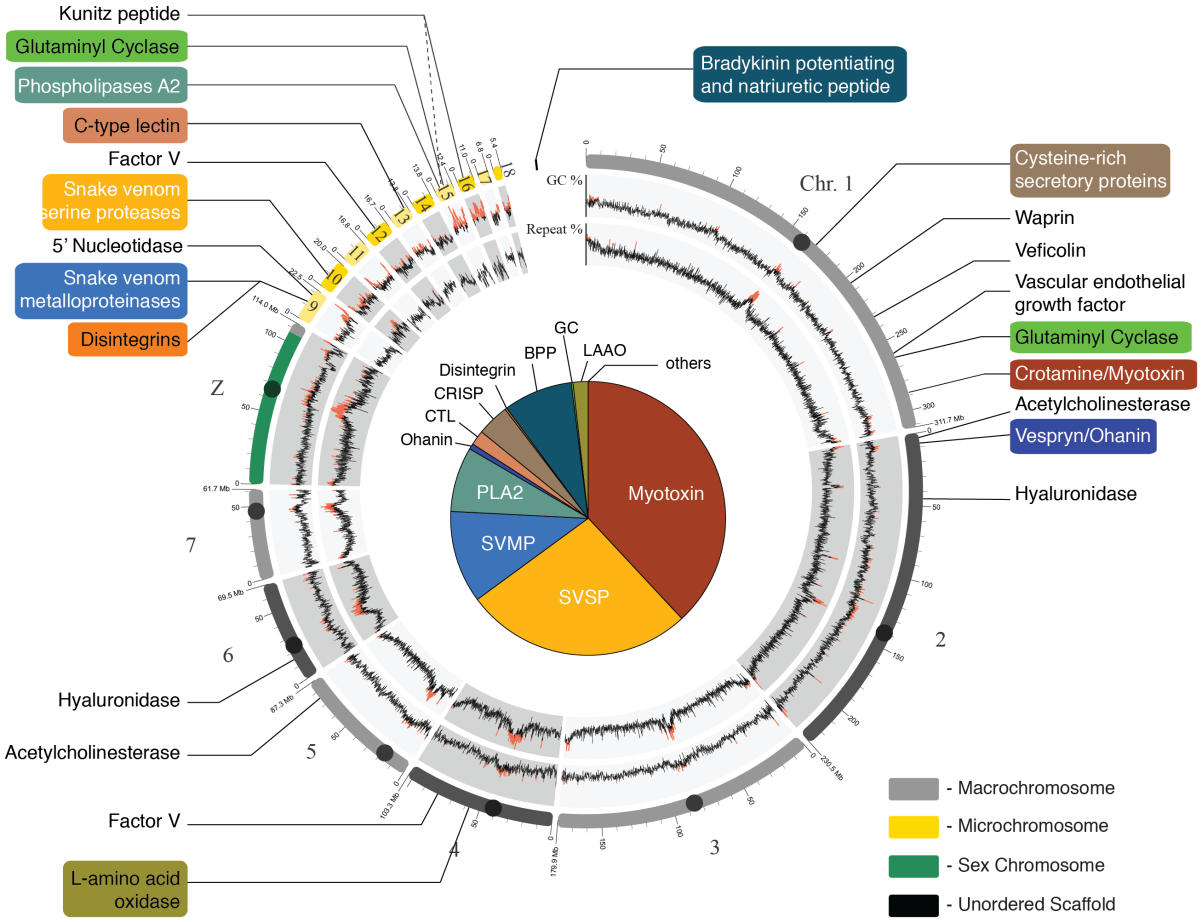
Supplemental Figure S14. Scatterplots of the number of predicted estrogen response elements versus the ratio of $\log_2$(female/male) gene expression in 100 kb windows across the rattlesnake Z Chromosome (*A*) and *Anolis* Chromosome 6 (*B*). The significant positive correlation between variables on the rattlesnake Z is shown by the red line.



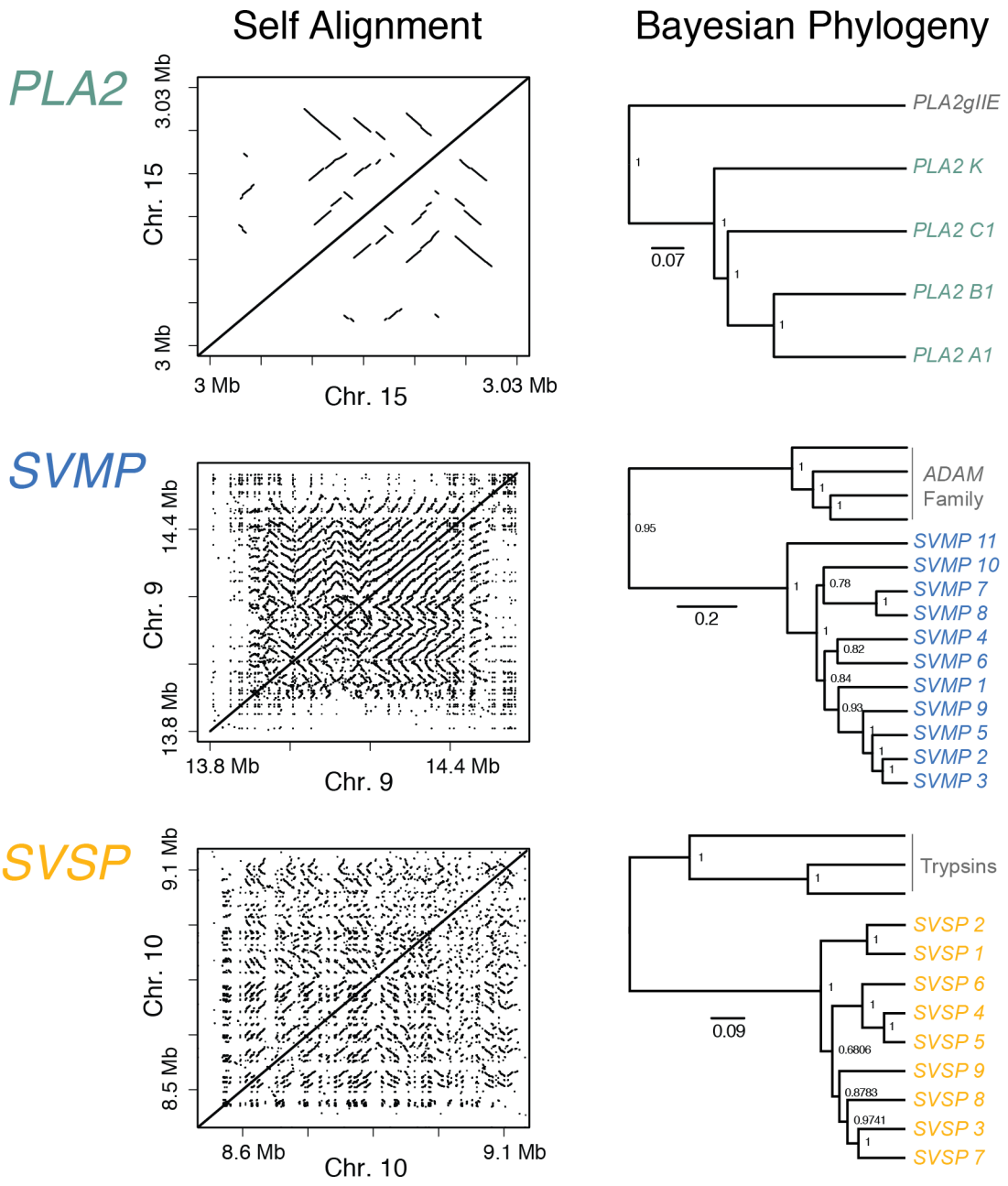
Supplemental Figure S15. Density of estrogen response elements (EREs) across the genomes of squamate species. Density in Z-linked regions of the Prairie Rattlesnake (*Crotalus*) and Five Pace Viper (*Deinagkistrodon*) and the syntenic Anole lizard (*Anolis*) Chromosome 6 regions are depicted in green, and the genomic background for each species is shown in grey bars. The black bar and asterisk depict that EREs are enriched on the pitviper Z Chromosome relative to the homologous autosome in *Anolis* (Chromosome 6).



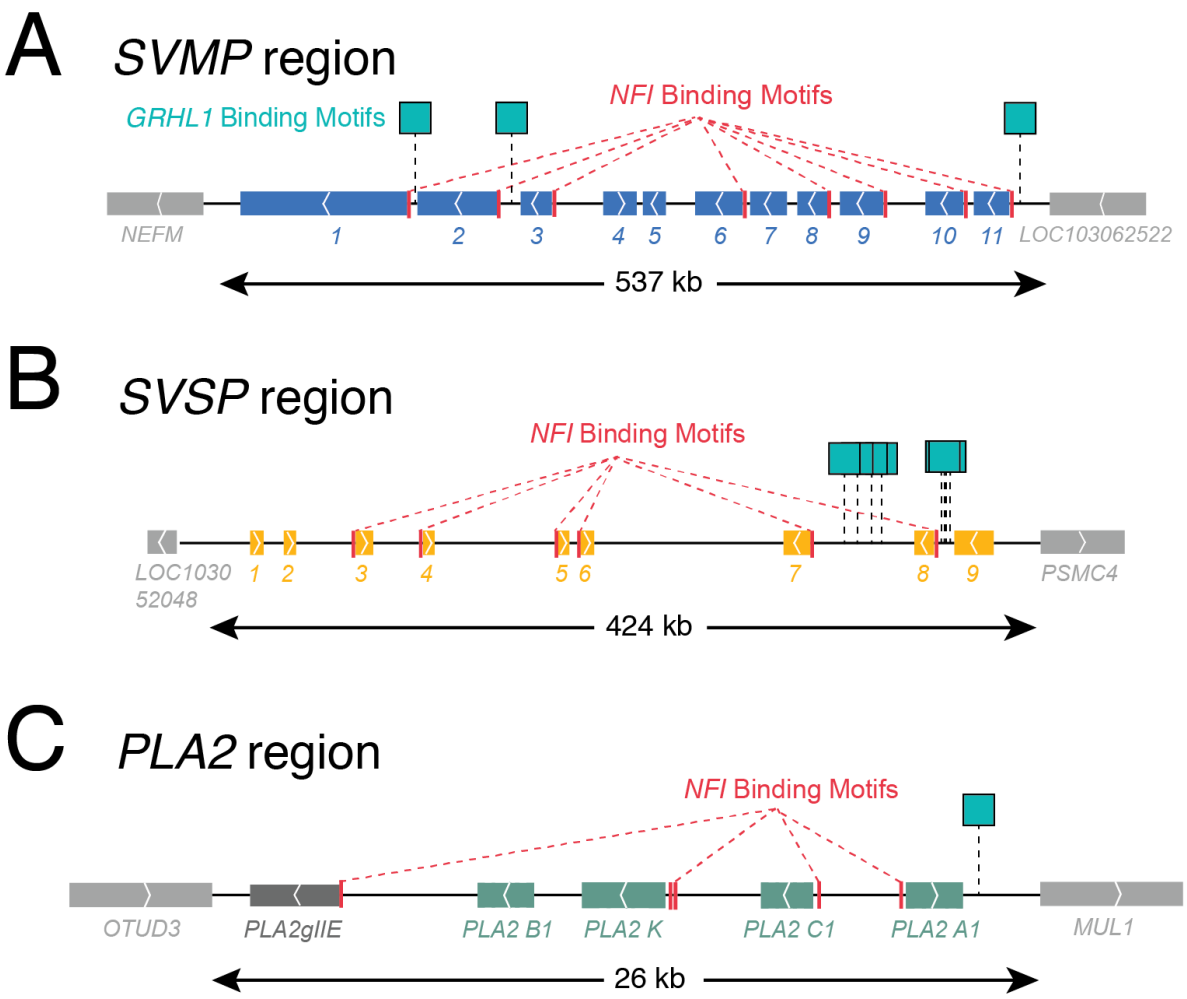
Supplemental Figure S16. Schematic of the initial misassembled microchromosome scaffold. The heatmap panel at the top depicts the high frequency intrachromosomal contacts within individuals microchromosomes, and black triangles depict boundaries between microchromosomes. Breakpoints between Chicago scaffolds used as initial microchromosome breakpoint hypotheses are shown as red dashes below the Hi-C heatmap. The middle two panels show synteny alignments between rattlesnake, chicken, and anole microchromosomes. The bottom two panels show windowed GC and repeat content across microchromosomes. Blue dashed lines in the lower panels show breakpoints between individual microchromosomes.



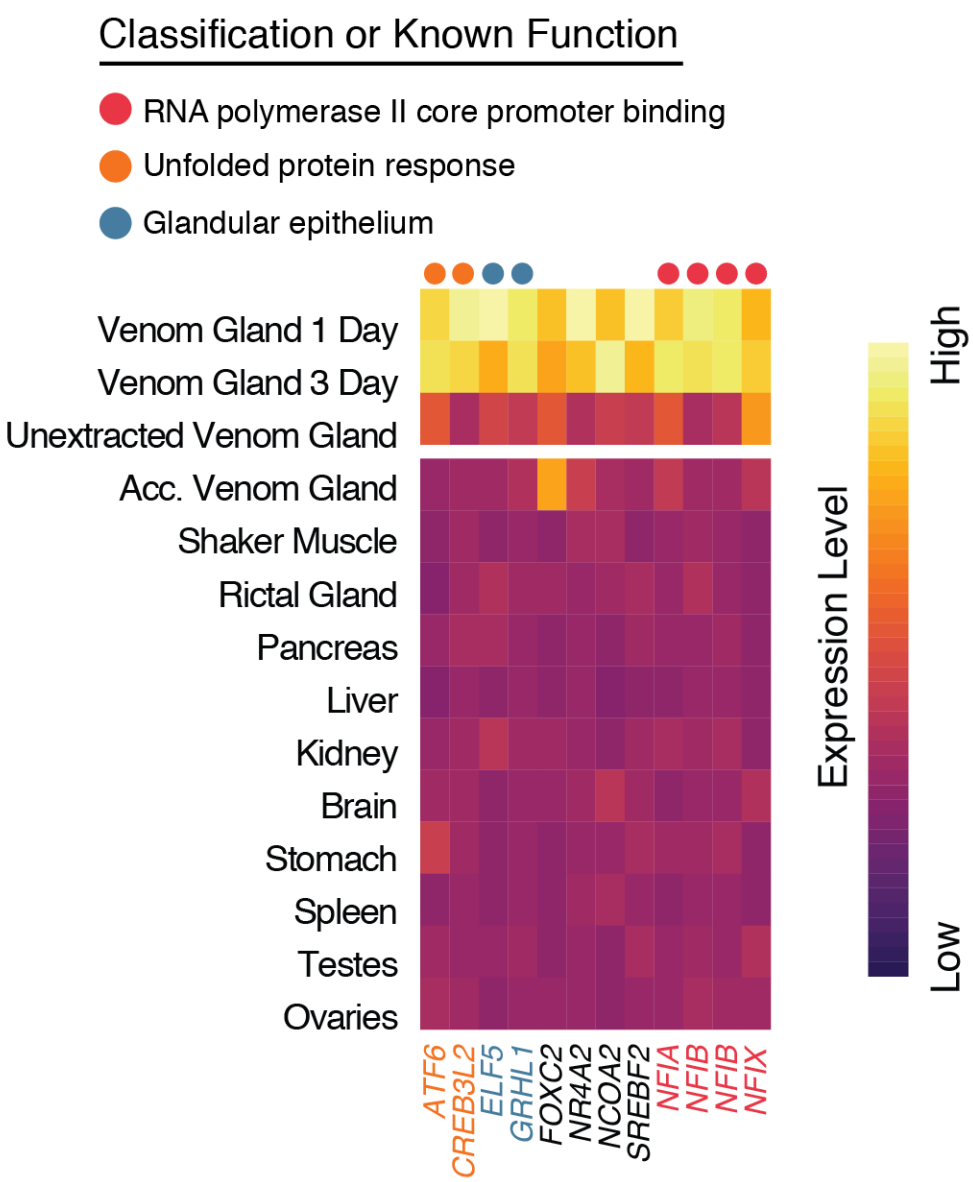
Supplemental Figure S17. Chromosomal locations of snake venom gene families in the prairie rattlesnake. The pie chart in the center depicts the relative abundance of venom families in the prairie rattlesnake proteome. Chromosomal ideograms and windowed scans of GC content (%) and repeat content (%) correspond to those described in Fig. 1 in the main text).



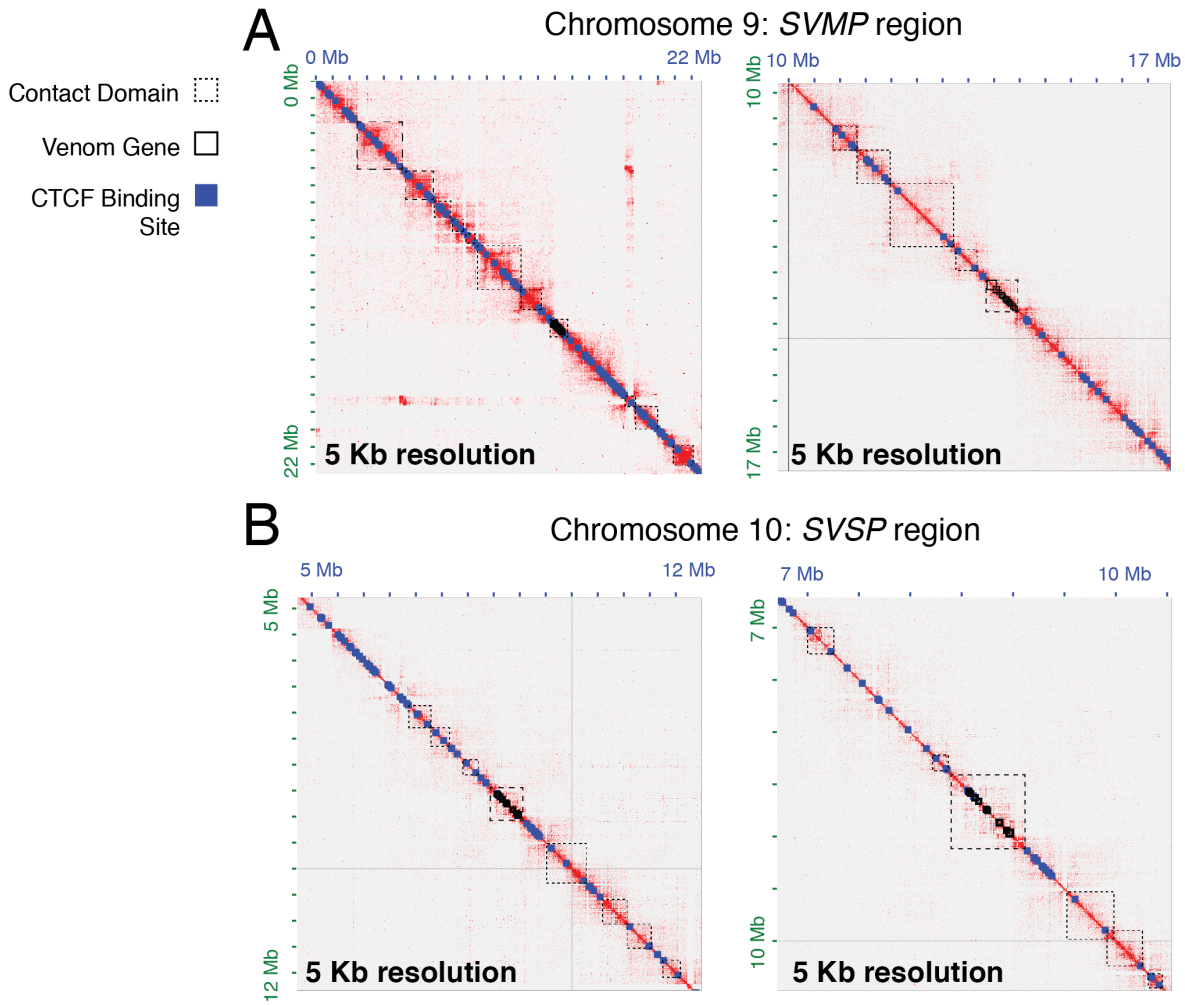
Supplemental Figure S18. Regional self alignment of phospholipase A2 (*PLA2*), snake venom metalloproteinase (*SVMP*), and serine proteinase (*SVSP*) venom gene clusters (left). Parallel and perpendicular lines off of the central diagonal line indicate segmental duplications. Bayesian phylogenetic tree estimates for each of the three gene families constructed based on protein alignments (right), with venom gene paralogs shown in color, and non-venom paralogs in grey. Values at nodes represent posterior probabilities.



Supplemental Figure S19. Structure of annotated *SVMP* (*A*), *SVSP* (*B*), and *PLA2* (*C*) venom gene clusters in the prairie rattlesnake genome. Strandedness (i.e., +/-) of genes is summarized by arrows in the center of each gene. The length of each cluster is shown at the bottom of each panel. Non-venom genes flanking each cluster are shown in grey. In the *PLA2* region, *PLA2gIIE* (non-toxin) is depicted in dark grey. Predicted *NFI* transcription factor binding sites within the 1 kb upstream region of venom genes are shown in red, and locations of predicted *GRHL1* binding sites between genes are shown as turquoise squares.



Supplemental Figure S20. Gene expression across tissues of 12 transcription factors (TFs) significantly upregulated in the venom gland. Broad classifications of known TF functions are annotated at the top of each gene, where applicable.



910

911 **Supplemental Figure S21.** Zoomed out Hi-C heatmaps of the SVMP (*A*) and SVSP (*B*) venom gene
912 regions at two scales (left and right) on microchromosomes, depicting chromatin contact domain
913 structure. Inferred contact domains are represented by dashed black boxes, venom genes in each venom
914 gene region are depicted by solid black boxes, and predicted *CTCF* binding sites are represented by blue
915 squares. Zoomed in versions of these schematics are presented in Fig. 4 in the main text.