

SUPPLEMENTARY INFORMATION

1. SUPPLEMENTARY METHODS

1.1 Garter snake genome sequencing

A single adult female *Thamnophis sirtalis* was obtained from a wild population at Coffin Butte Road, Benton County, Oregon, USA, 44° 41' 57.02'' N, 123° 13' 18.06'' W, 229 feet elevation (collected by Brian Gall; field tag: EDBJR-23777). The individual was euthanized and tissues were preserved under Utah State University approved IACUC protocols (#1008). The specimen was deposited in the University of Texas at Arlington Amphibian and Reptile Diversity Research Center collection (voucher accession: UTA R 62823). This individual was used for all complete genome sequencing. All reads were sequenced on the HiSeq2000 (Software version HCS1.5.15.0 and RTA version 1.13.48.0)

Kmer-based estimates of genome size were generated using the following reads from the NCBI Sequence Read Archive (SRA): SRR1029946, SRR1029947, SRR2017627, SRR770194, SRR770253, SRR770448, SRR770501, SRR786677, SRR786678, SRR786680. Reads were quality trimmed using Trimmomatic v0.33 (Bolger, et al. 2014) with the settings LEADING:10 TRAILING:10 SLIDINGWINDOW:4:15 MINLEN:36. Kmers were extracted using jellyfish v2.2.4 (Marçais and Kingsford 2011), and GCE v1.0.0 (Liu, et al. 2013) was used to estimate genome size using both 19mers and 23mers.

1.2 Annotation and Characterization of Transposable Elements

Repeat elements were annotated by homology-based and by *de novo* prediction approaches. *De novo* identification was performed on the garter snake and the five-pacer pitviper (*Deinagkistrodon acutus*; (Yin, et al. 2016)) genome assemblies in *RepeatModeler* v.1.0.8 (Smit, et al. 2014), followed by further classification in CENSOR (Kohany, et al. 2006). The resulting libraries were then combined with a 12-species snake specific repeat library (Castoe, et al. 2013) to generate a reference library of repeat elements for use in comparative analyses. For *Anolis carolinensis*, we generated a lizard specific library that includes *de novo* repeat identification for

the *Anolis carolinensis* (Alföldi, et al. 2011), *Gekko japonicas* (Liu, et al. 2015), and *Pogona vitticeps* (Georges, et al. 2015) genomes. 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 (v.20.11, 07 August 2015), and our snake specific library as references. Output files were post-processed using a modified implementation of the ProcessRepeat script (RepeatMasker package) to incorporate summary statistics of TEs prevalent in squamate reptile genomes (i.e. CR1/CR1-L3, L2).

1.3 Analysis of gene, exon, and intron number and length

To quantify the number and length of genes, exons, and introns in the *Thamnophis sirtalis* genome, we compared distributions of gene features (i.e., genes, exons, and introns) from the garter snake with those from other highly complete vertebrate genomes. We downloaded the genome feature files (GFF) for the human (GCA_000001405.25), chicken (GCA_000002315.3), and anole lizard (GCA_000090745.1) from Ensembl. We then parsed each GFF using a Python script (gff_gene_stat.py; https://github.com/drewschield/Comparative-Genomics-Tools/blob/master/gff_gene_stat.py) to determine the length of each gene, exon, and intron, and the numbers of exons and introns per gene across the genome of each species.

1.4 Analyses of positive selection

To infer patterns of positive selection in protein coding genes, we generated orthologous coding sequence alignments for 27 species from available genomes and transcriptomes, including 18 reptiles and nine mammals (see Supplementary Table S2). Of the 18 reptiles, one species, *Thamnophis elegans*, was represented twice for two independent assemblies of two distinct ecomorphs (meadow and lake; See Supplementary Table S2 for details). For simplicity, hereafter we refer to these 28 genome/transcriptomes as 28 species. Methods for creating alignments followed a previous established protocol (McGaugh, et al. 2015), and are based on OrthoMCL clustering analyses followed by a series of filtering and annotation steps (Li, et al. 2003) and alignment with MSAProbs (Liu, et al. 2010).

Since alignments were processed in a high-throughput manner for all genomic data available, some alignments were missing in some species out of the 28 total that were included in the OrthoMCL analyses. This is due to either extremely divergent genes or incomplete sequences in the available transcriptomes and/or genomes. To address this, we required 15 or greater species to be in a gene's alignment to include that gene in subsequent analyses. We also required that the alignments included *Lamprophis fuliginosus*, *Thamnophis sirtalis*, *Thamnophis couchii*, and at least one of the *Thamnophis elegans* ecomorphs. In total, we identified 4,639 alignments that met this inclusion criterion.

Alignments were used to identify evidence for positive selection in the codeml program in PAML version 4.7 (Yang 2007). We used a likelihood-ratio test (LRT) and three separate branch-site models to test for evidence of positive selection. The three branch-site models varied in which branch was placed in the foreground and included: 1) the branch to the colubrid family *Colubridae*, 2) the branch to the *Thamnophis* genus, and 3) the branch to *Thamnophis sirtalis*. For each LRT, the test statistic was compared to a 1:1 mixture of χ^2 distributions with one and zero degrees of freedom (Goldman and Whelan 2002). Genes with significant LRT identified in codeml were corrected for multiple testing with sequential Bonferroni (Holm 1979) or false discovery rate calculations (Benjamini and Hochberg 1995). Results of both approaches resulted in a nearly identical set of genes. For all genes that were significant with initial branch-site tests, we hand-corrected alignments (mostly realigning indels) and reran the branch-site tests to ensure that the signal was not driven by poor alignments. Only results from these hand-corrected alignments were included in the summary. For additional details on the orthologs, their annotation, and results of the branch-site tests, see Supplementary Table S3 and Supplementary File 2.

For each gene with a significant PAML result, the Ensembl gene identifier for the human ortholog (or if human not present, mouse) was included in Ingenuity Pathway Analysis (IPA) Comparison Analysis (Krämer, et al. 2013) as the knowledge base for IPA contains only human, mouse, rat, and zebrafish gene identifiers. One gene without an identifier from the 10 genes significant in the *Thamnophis* genus branch-site test could not be included in the IPA analyses.

For all genes, best BLAST hits to the Uniprot database are reported, regardless of whether the alignment exhibited Ensembl gene identifiers that were annotated in the IPA knowledge base.

1.5 Sex Chromosomes

We chose fragments of six genes that have been shown to have Z and W alleles in caenophidian snakes: WAC (WW domain containing adaptor with coiled-coil; 1522 bp), CTNNB1 (beta-catenin; 691 bp), COMMD3 (COMM domain containing 3; 359 bp), MSL1 (male specific lethal 1 homolog; 454 bp), LSM12 (LSM12 homolog; 4342 bp), and WDR48 (WD repeat domain 48; 349bp) (Laopichienpong, et al. 2017; Matsubara, et al. 2006; Vicoso, et al. 2013). We used BLAST (Altschul, et al. 1990) to identify *T. sirtalis* Z and W scaffolds for each of the six Z-linked genes. Exons and flanking introns were targeted for phylogenetic analysis although difficulty aligning non-coding regions limited the length of intronic sequences that we could analyze. Two assembled female snake genomes on GenBank also included Z and W alleles, *Pantherophis guttatus* (GCA_001185365.1) (Ullate-Agote, et al. 2015) and *Crotalus horridus* (GCA_001625485.1), and were included in the phylogenetic analyses. We used sequences from two snake species with XX/XY sex chromosomes, *Boa constrictor* and *Python molurus bivittatus* (Gamble, et al. 2017), and three lizards, *Anolis carolinensis*, *Pogona vitticeps*, and *Gekko japonicus*, as outgroups (Alföldi, et al. 2011; Bradnam, et al. 2013; Castoe, et al. 2013; Georges, et al. 2015; Liu, et al. 2015). Orthologous sequences were not identified for *G. japonicus* in three of the genes and *P. vitticeps* for one of the genes and therefore excluded in those phylogenetic analyses.

Sequence alignment was conducted using MUSCLE implemented in Geneious 11.0.3 (Edgar 2004; Kearse, et al. 2012). We estimated maximum likelihood trees using RAxML-HPC BlackBox, version 8.2.10 (Stamatakis, et al. 2008) on the CIPRES (Cyberinfrastructure for Phylogenetic Research) Science Gateway (Miller, et al. 2010). Nodal support was estimated using rapid bootstrapping with automatic stopping criterion (Stamatakis, et al. 2008).

W-specific PCR primers were designed based on the alignment of *T. sirtalis* Z and W scaffolds in Geneious 11.0.3 (Edgar 2004; Kearse, et al. 2012). Primers were designed to amplify in the W allele but not the Z allele following Gamble et al. (In Press). PCR was performed using the

following PCR profile: an initial 5-min denaturation at 94°C followed by 32 cycles of denaturation (30 s at 94°C), annealing (45 s at 52–57°C) and extension (1 min at 72°C), followed by a final extension of 5 min at 72 °C. See Supplementary Table S2 for locus-specific annealing temperatures and primer sequences.

Chromosome spreads from a female *T. sirtalis* were prepared from fibroblasts established from a tail clip (Ezaz, et al. 2008; Gamble, et al. 2014). Fibroblasts were grown at 28–31° C in media containing: DMEM 1X (Invitrogen, Carlsbad, CA) with 4.5 g/L glucose and L-glutamine without sodium pyruvate; 20% fetal bovine serum; and anti-anti (Invitrogen, Carlsbad, CA), which contains penicillin, streptomycin, and amphotericin. Cells were arrested in metaphase using vinblastine sulphate (1 mg/ml), collected after trypsin digestion, and incubated in a 0.07 M KCl hypotonic solution for 20 min in a 37°C water bath. Cells were fixed and washed in methanol:acetic acid (3:1). Cell suspensions were then dropped onto clean glass slides, allowed to air dry, and dehydrated in an ethanol series (70%, 95%, 100%). We assessed the accumulation of repetitive DNA sequences on the putative sex chromosome by hybridizing a fluorescently labeled (GATA)_n satellite repeat onto metaphase spreads. The GATA satellite repeat, also called the Bkm satellite repeat, has been shown to accumulate onto the sex chromosomes of multiple animal species and is a good candidate marker identifying sex chromosomes (Gamble, et al. 2014; Nanda, et al. 1990; O’Meally, et al. 2010; Singh, et al. 1980). (GATA)_n probes were generated by PCR in the absence of template DNA (Ijdo, et al. 1991) using (GATA)₇ and (TATC)₇ primers. Probes were labeled via nick translation with Chromatide/Alexa Fluor fluorescently labeled dUTP 488-5 (Life Technologies). We checked the sizes of the nick translated fragments by electrophoresis on a 1% TBE gel. Labeled DNA was ethanol precipitated and resuspended in 100 ul hybridization buffer (Ezaz, et al. 2005), denatured at 72°C for 10 minutes and snap-cooled on ice for five minutes. We added 20 ul of probe to each slide, affixed a cover slip using rubber cement, heated slides again to 72°C for 5 minutes and incubated overnight at 37°C. Slides were washed once at 60°C in 0.4%SSC, 0.3% Igepal CA-630 (Sigma Aldrich) for two minute followed by a second two minute wash in 2%SSC, 0.1% Igepal CA-630 at room temperature. Slides were dehydrated in an ethanol series (70%, 95%, 100%) and air dried. Slides were stained with 4,6-diamidino-2-phenylindole (DAPI) and mounted with a cover

slip using Permafluor (Lab Vision). All slides were photographed on a Zeiss Imager Z1 microscope using a Zeiss MRm camera. Images were captured using Zeiss Axiovision software.

1.6 Identification and Analysis of microRNA and Associated mRNA Targets

The sequence-specific microRNA prediction (SMIRP) tool was used for the identification of microRNA that are specific to *T. sirtalis* (Peace, et al. 2015). The species-specific positive training datasets are built using of the miRBase (v.21) database (<http://www.mirbase.org>). Clusters were generated using a threshold of 80% sequence similarity, using default CD-hit parameters. Using the miRBase sequence dataset and the generated sequence clusters, we developed positive training datasets for microRNA classification, which are targeted toward arbitrary species. These datasets were developed as follows: For a given positive integer n , the largest n clusters are chosen from the CD-hit clustering results, then a representative sequence is chosen from each of these clusters. Each representative sequence is the sequence whose species is nearest our target species in terms of phylogenetic classification. In the event of multiple sequences within a cluster whose species are equally close to our target species in terms of phylogenetic classification, the sequence among these candidate representative sequences whose length is closest to the mean length of sequences within the cluster is chosen as the representative sequence for the cluster. Negative training datasets are generated from sequences that resemble microRNA within the annotated coding regions of the *T. sirtalis* genomes.

HeterMirPred was used for feature set generation for the classification of *T. sirtalis* microRNA using the targeted positive and negative training sets. The feature set consists of 20 sequence and structural features. We then used a support vector machine (SVM) as our classifier as opposed to the ensemble classifier, due to the ability of a SVM to provide clear probability estimates along with its determinations of real or pseudo-microRNA.

Mature microRNAs were searched against the *T. sirtalis* coding sequences using miRanda (v.3.3a) (Enright, et al. 2003). To search the predicted microRNA target genes within *T. sirtalis* the miRanda software was used within the following parameters and conditions: a gap opening penalty of -9 , a gap extension penalty of -4 ; match with minimum score threshold 160, target duplex with maximum threshold free energy -25 kCal/mol, scaling parameter 4 for

complementary nucleotide match score, counting from the microRNA 5' end, and demand strict 5' seed pairing on between 2 and 9 nucleotides.

1.7 Analyses of Visual Gene Loss and Opsin Expression Localization

A set of 119 visual genes were targeted for analysis (Schott et al. 2017). These were extracted from the *Thamnophis sirtalis* genome assembly and CDS annotation using BLAST searches (blastn, discontinuous megablast, megablast, tblastn, and tblastx) using *Python* and *Anolis* mRNA, exon, and protein queries. CDS and BLAST results were manually edited to ensure proper exon boundaries and corrected with direct parsing of the genomes, as necessary. Visual genes were additionally obtained from mammals, crocodylians (*Alligator mississippiensis*, *A. sinensis*, *Crocodylus porosus*, *Gavialis gangeticus*), and a gecko (*Gekko japonicus*) from NCBI GenBank and through BLAST searches of genome assemblies as necessary to confirm loss. Gene loss was assumed if a sequence could not be found in any of the available genomes. For mammals, if a sequence was found in monotremes and/or marsupials, but was absent in placental genomes, it was considered present in mammals, but lost in placentals.

Fixation of garter snake eyes was conducted as previously described (Bhattacharyya et al, 2017). Briefly, after enucleating *T. sirtalis* eyes in the light, they were rinsed in PBS (0.8% NaCl, 0.02% KCl, 0.144% NaHPO₄, and 0.024% KH₂PO₄, pH 7.4), fixed overnight at 4°C in 4% paraformaldehyde, infiltrated with increasing concentrations of sucrose (5%, 13%, 18%, 22%, 30%) in PBS, and embedded in a 2:1 solution of 30% sucrose and O.C.T compound (Tissue-Tek, Burlington, NC, USA) at -20°. The eyes were cryosectioned transversely at -25°C in 20 µm sections using a Leica CM3050 (Wetzlar, Germany) cryostat, placed onto positively charged microscope slides, and stored at -80°C until use.

Slides were first rehydrated in PBS and then air-dried to ensure adhesion. Sections were rinsed in PBS with 0.1% Tween-20 (PBT) and fixed in 4% paraformaldehyde PBS. After rinsing in PBT and PDT (PBT with 0.1% DMSO), the slides were incubated with blocking solution (1% BSA in PDT with 2% normal goat serum), and then incubated with primary antibody diluted in blocking solution overnight at 4° in a humidity chamber. Antibodies used were the K20 antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA, sc-389, lot#:C1909, dilution: 1:500) and RET-P1

anti-rhodopsin antibody (Sigma-Aldrich, St. Louis, MO, USA, O-4886, lot#: 19H4839, dilution: 1:100).

After extensive rinsing and soaking in PDT, secondary antibody was added to the samples and incubated at 37°C. Secondary antibodies used for the fluorescent staining were the AlexaFluor-488 anti-rabbit antibody (Life Technologies, Waltham, MA, USA, A11034, lot#: 1298480, dilution: 1:1000) and the Cy-3 anti-mouse antibody (Jackson ImmunoResearch, West Grove, PA, USA, 115-165-003, dilution: 1:800). After rinsing with PBS, followed by PDT, sections were stained with 10 µg/mL Hoechst. The sections were then mounted with ProLong® Gold antifade reagent (Life technologies, Waltham, MA, USA) and cover-slipped. Sections were visualized via a Leica SP-8 confocal laser microscope (Wetzlar, Germany).

1.8 Phylogenetic Analysis of Olfactory Receptors

Completely intact OR amino acid sequences were preliminarily grouped into known subfamilies by using a database of known human ORs from the HORDE database (Olender, et al. 2013). We used BLASTP to group intact ORs into putative subfamilies defined by Glusman, et al. (2000) and Hayden, et al. (2010). All intact OR sequences were aligned using MUSCLE (Edgar 2004) and a phylogenetic tree was constructed using a neighbor joining method in MEGA6 (Tamura, et al. 2013). Node support was based on 1,000 bootstrap replicates. The resulting phylogenetic tree was then used to correct preliminary subfamily annotations.

1.9 Venom Gene and Protein Analysis

We used tBLASTx to identify potential venom homologs in the *T. sirtalis* genome, based on a query set of known venom genes from diverse snakes. We inferred orthologous venom genes using phylogenetic analyses. Putative venom homologs were aligned with additional sequences, sourced from Genbank and used in previous studies (Junqueira-de-Azevedo, et al. 2016; Reyes-Velasco, et al. 2014; Vonk, et al. 2013), using MUSCLE v3.5 (Edgar 2004) before best-fit models of molecular evolution were fit to each gene using PartitionFinder v1.1.1 (Lanfear, et al. 2012). We inferred phylogenies for each gene using MrBayes v3.2.1 (Huelsenbeck and Ronquist 2001) implemented on the CIPRES science gateway server (Miller, et al. 2010), with two

simultaneous chains run for a total of 10^7 generations (sampled every 10^3 generations). We confirmed mixing and convergence using Tracer v1.5 (Drummond and Rambaut 2007) and discarded the first 25% of each run as burn-in.

To examine expression patterns of venom gene orthologs in the Duvernoy's gland and other tissues, RNAseq data for eight organs of *T. sirtalis* (Duvernoy's gland, brain, kidney, liver, lung, ovary, upper digestive tract, and lower digestive tract) were downloaded from NCBI: SRA SRP018850. STAR v2.4.4a (Dobin, et al. 2013) was used to map RNAseq data to the garter snake genome and generate read counts. TMM normalization in edgeR (Robinson, et al. 2010) was used to convert raw counts to normalized counts per million (CPM). Heatmaps of gene expression across tissues were generated in R.

Analysis of venom protein composition. Snakes were collected from Rio Blanco Co. (Colorado: *Thamnophis elegans vagrans*) and Benton Co. (supplied by E. D. Brodie Jr.; Oregon: *T. sirtalis parietalis*) and housed at the University of Northern Colorado Animal Resource Facility under approved animal protocols (#9204). Venom extractions were performed as described previously (Hill and Mackessy 1997), using doses of 30 μ g ketamine-HCL per g body weight and 6.0 μ g/g pilocarpine-HCL. Venom was collected into capillaries, frozen, and lyophilized. One mg of crude venom was subjected to fractionation on a Jupiter C18 column (250 x 4.6 mm, 5 μ m, 300 Å pore size) as described previously (Smith and Mackessy 2016). One-minute fractions were collected, frozen, and lyophilized for analysis by SDS-PAGE and MALDI-TOF mass spectrometry. Samples were resuspended in 10 μ L MilliQ-filtered H₂O; 5 μ L was added to 5 μ L 2X LDS/DTT buffer and electrophoresed on 12% acrylamide NuPage gels (Smith and Mackessy 2016); 20 μ g of crude venom (*T. s. parietalis* and *T. e. vagrans*) were also analyzed. Bands were assigned to major protein families based on mass, comparison with transcriptomic data, and previously published data. Samples for MALDI (1 μ L) were mixed with 1 μ L sinapinic acid (10 mg/mL in 50% acetonitrile, 0.1% TFA) and spotted onto MALDI target plates. Samples were analysed on a Bruker Microflex LRF MALDI-TOF mass spectrometer operated in linear mode (Smith and Mackessy 2016).

2. SUPPLEMENTARY RESULTS

2.1 Location of *SCNA* Genes

Although the genome has not yet been assembled into chromosomes, evidence suggests that patterns of synteny observed in other species (Zakon, et al. 2010) are conserved in *T. sirtalis*. Scaffold 360 contains the paralogs *SCN9A* (forward strand), *SCN1A* (forward strand), and *SCN2A* (reverse strand) in that order, with the genes *TTC21B*, *GALNT3*, and *CSRNP3* located between *SCN1A* and *SCN2A* as in other species. Sequence from a fourth paralog that typically occurs adjacent to these three, *SCN3A*, was found across two different scaffolds with multiple assembly gaps. Another group of typically syntenic paralogs, *SCN11A*, *SCN10A*, and *SCN5A*, were found in their predicted order on scaffold 243. However, the second half of *SCN5A* was placed on a separate scaffold. As predicted, *SCN4A* and *SCN8A* were not found on scaffolds with other paralogs. We describe our annotations of each of the genes in turn below, comparing the sequences found in the current assembly to previous results.

The coding sequence of *SCN1A* was found beginning at position 367,500 of the forward strand of scaffold 360 and was annotated by Gnomon as gene LOC106541394. The coding sequence was identical to that previously described (McGlothlin, et al. 2014) except for 2 bp in exon 5a, 2 bp in exon 9, 1 bp in exon 24, and 2 bp in exon 26. The discrepancies in exon 9 led to amino acids differences and appear to represent a sequencing error in the previous study, as this segment of exon 9 was not found in the BAC library and was derived from cDNA. In addition, one of the differences in exon 5a led to an amino acid difference, which may represent a segregating polymorphism. The other nucleotide differences were synonymous.

The coding sequence of *SCN2A* was found beginning at position 141,945 of the reverse strand of scaffold 360 and was annotated by Gnomon as gene LOC106541399. The coding sequence was identical to that previously described except for a 34 bp assembly gap in exon 9, a large assembly gap that affects 29 bp of exon 14 and all of exon 15 (357 bp, including the DII P-loop), one nonsynonymous difference in exon 19, and 8 bp (2 nonsynonymous, 6 synonymous). As above, the differences in exons 19 and 26 likely represent sequencing errors in the previous

study. One of these differences affected the predicted amino acid sequence of the DIV P-loop (NFETFGNMLCLFQITTSAGWDGLL), but did not occur at a site known to influence TTX binding (Feldman, et al. 2012). The previous study was also unable to sequence the last 274 bp of *SCN2A*, which are found in the genome.

Exons 1-3 and 15-26 of *SCN3A* were found beginning at position 57,110 on the reverse strand of scaffold 1732 and were annotated by Gnomon as gene LOC106551485. A large assembly gap occurred after exon 3. Exons 6-13 were found beginning at position 69,150 on the reverse strand of scaffold 2787 and were annotated by Gnomon as part of gene LOC106554673. Exons 4, 5a, 5b, and 14 could not be located in the genome assembly. Within the exons that were present, coding sequence was identical to that found in the previous study except for a 1 bp (nonsynonymous) difference in exon 11. In addition, a large assembly gap was located with exon 26, causing 456 bp (including the DIV P-loop) to be missing from the genome assembly.

The complete coding sequence of *SCN4A* was found, but was spread across three scaffolds: exons 1-13 were annotated as gene LOC106557226 beginning at position 1,122,874 of the reverse strand of scaffold 101, exons 14-25 were annotated as part of gene LOC106551420 beginning at position 137,946 of the forward strand of scaffold 1719, and exon 26 was annotated as gene LOC106554426 beginning at position 583 on the forward strand of scaffold 583. The coding sequence was identical to that described in the previous study.

The first 16 exons of *SCN5A* were found in scaffold 243, with the coding sequence of exon 1 beginning at position 711,957 of the reverse strand. A large (63,130 bp) fragment located from positions 243,186 to 306,315 on the forward strand of scaffold 243 represented an apparent assembly error. This fragment contained *SCN5A* exons 5-8 beginning at position 250,855 on the positive strand and is located between two large assembly gaps. Our annotation of *SCN5A* corrects this apparent error by reversing this segment. Exons 5-8 were automatically annotated by Gnomon as part of gene LOC106539742, which also contained 4 other misidentified exons. Exons 1-4 and 9-16 were not annotated by Gnomon. Exons 17-26 are found on the forward strand of scaffold 774 beginning at position 107,110 and were annotated by Gnomon as part of gene LOC106545622.

The complete coding sequence of *SCN8A* was found beginning at position 323,784 of the forward strand of scaffold 893 and correctly annotated by Gnomon as *SCN8A*. There were four

synonymous differences when compared to the sequence from the previous study, one each in exons 8, 14, 16, and 25.

Most of the coding sequence of *SCN9A* was found beginning at position 226,516 of the forward strand of scaffold 360 and annotated by Gnomon as gene LOC106541401. Exons 17-19 were missing due to an assembly gap; however, exon 18 was placed within exon 26 in an apparent assembly error. As a consequence, exon 26 was split into two parts including a large gap, and 161 bp were missing, including the DIV P-loop. The rest of the coding sequence was identical to that from the previous study.

Most of the coding sequence of *SCN10A* was found beginning at position 519,213 of the reverse strand of scaffold 243. Many assembly gaps were present, and exons 5, 10, and 14 were missing. Exon 4 was found out of place before exon 2, representing an apparent assembly error. Much of this gene was annotated by Gnomon as gene LOC106539743.

Some of the coding sequence of *SCN11A* was found beginning at position 396,886 of the reverse strand of scaffold 243, although exons 7, 10-12, and 21-23 (DIII P-loop) were missing due to large assembly gaps. Exons 17 and 18 and a fragment of exon 9 were located out of place and are reinstated to their proper place in our annotation. Some of this gene was annotated by Gnomon as gene LOC106539744. We located exons 22-23 beginning at position 33,473 of the positive strand of scaffold 674.

3. SUPPLEMENTARY TABLES

Supplementary Table S1. Characterization of the number and length of genes, introns, and exons in the human, garter snake, anole, and chicken genomes.

	Human	Garter snake	Anole	Chicken
Number of genes	21,407	20,186	18,595	18,346
Mean gene length (bp)	62,634.77	24,532.96	35,331.78	24,354.34
Median gene length (bp)	23,247.50	13,384	16,110	8,287
Number of exons	1,199,851	235,641	206,290	323,585
Mean number exons per gene	8.85	9.03	9.94	10.092
Median number exons per gene	6	7	7	8
Mean exon length (bp)	249.22	280.12	260.3	211.62
Median exon length (bp)	128	133	128	128
Number of introns	748,295	202,369	171,351	275,075
Mean number introns per gene	7.85	8.03	8.94	9.092
Median number introns per gene	5	6	6	7
Mean intron length (bp)	6,449.16	3,553	3,675	2,695.47
Median intron length (bp)	1,638	1,439	1,351	777

Supplementary Table S2. List of species included in ortholog clustering and analyses of positive selection.

Species	Common name	Order
<i>Ornithorhynchus anatinus</i>	Platypus	Monotreme
<i>Monodelphis domestica</i>	Opossum	Marsupial
<i>Bos taurus</i>	Cow	Artiodactyl
<i>Canis lupus familiaris</i>	Dog	Carnivore
<i>Mus musculus</i>	Mouse	Rodent
<i>Otolemur garnettii</i>	Bushbaby	Primate
<i>Pongo abelii</i>	Orangutan	Primate
<i>Homo sapien</i>	Human	Primate
<i>Pan troglodytes</i>	Chimp	Primate
<i>Chrysemys picta</i>	Painted Turtle	Turtle
<i>Pelodiscus sinensis</i>	Chinese Softshell Turtle	Turtle
<i>Alligator mississippiensis</i>	Alligator	Crocodile
<i>Anas platyrhynchos</i>	Duck	Bird
<i>Gallus gallus</i>	Chicken	Bird
<i>Meleagris gallopavo</i>	Turkey	Bird
<i>Ficedula albicollis</i>	Collared Flycatcher	Bird
<i>Taeniopygia guttata</i>	Zebra Finch	Bird
<i>Pogona vitticeps</i>	Bearded Dragon	Squamate (Iguania)
<i>Anolis carolinensis</i>	Green Anole	Squamate (Iguania)
<i>Ophisaurus gracilis</i>	Glass Lizard	Squamate (Anguimorph)
<i>Python vivittatus</i>	Python	Squamate (Snake)
<i>Ophiophagus hannah</i>	Cobra	Squamate (Snake)
<i>Agkistrodon piscivorus</i>	Cottonmouth	Squamate (Snake)
<i>Lamprophis fuliginosus</i>	African House Snake	Squamate (Snake, Colubrid)
<i>Thamnophis couchii</i>	W. Aquatic Garter Snake	Squamate (Snake, Colubrid)
¹ <i>Thamnophis elegans</i>	W. Terrestrial Garter Snake	Squamate (Snake, Colubrid)
<i>Thamnophis sirtalis</i>	Common Garter Snake	Squamate (Snake, Colubrid)

¹Two distinct assemblies from two distinct ectomorphs

Supplementary Table S3. Genes with evidence of positive selection with the branch leading to colubrids, *Thamnophis*, and *Thamnophis sirtalis* placed in the foreground in codeml in PAML. See Supplementary File 2 for additional information on these orthologs and tests.

Foreground Branch	UNIPROT ID	Entrez Gene Name	X ² branch-sites test	FDR q-value
Colubrids	<i>EMC7</i>	ER membrane protein complex subunit 7	19.40887	0.004896
	<i>WNT9A</i>	Wnt family member 9A	46.897616	4.34E-09
	<i>PICK1</i>	protein interacting with PRKCA 1	49.796718	1.32E-09
	<i>STK40</i>	serine/threonine kinase 40	104.224818	2.10E-21
<i>Thamnophis</i>	<i>RBPM5</i>	RNA binding protein with multiple splicing	13.577426	0.048282
	<i>DYHC1</i>	dynein cytoplasmic 1 heavy chain 1	14.79828	0.027758
	<i>GBF1</i>	golgi brefeldin A resistant guanine nucleotide exchange factor 1	14.978168	0.027758
	<i>TBCD</i>	tubulin folding cofactor D	16.97549	0.010981
	<i>URB2</i>	NA	17.813104	0.008077
	<i>RPAB3</i>	RNA polymerase II subunit H	25.046968	0.000216
	<i>SULF2</i>	sulfatase 2	34.134766	2.39E-06
	<i>EXT2</i>	exostosin glycosyltransferase 2	99.703774	1.03E-20
	<i>K1161</i>	myogenesis regulating glycosidase (putative)	105.038932	9.26E-22
	<i>P5CS</i>	aldehyde dehydrogenase 18 family member A1	137.724304	9.71E-29
<i>T. sirtalis</i>	<i>MLTK</i>	mitogen-activated protein kinase kinase kinase 20	14.599276	0.034276
	<i>A16L1</i>	autophagy related 16 like 1	22.55583	0.000592
	<i>TMCO4</i>	transmembrane and coiled-coil domains 4	28.889262	2.54E-05
	<i>ANO5</i>	anoctamin 5	32.078458	5.73E-06
	<i>PCSK7</i>	proprotein convertase subtilisin/kexin type 7	33.467668	3.36E-06
	<i>PR40A</i>	pre-mRNA processing factor 40 homolog A	39.019608	2.43E-07
	<i>PK3CD</i>	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta	62.293054	2.29E-12
	<i>MED24</i>	mediator complex subunit 24	65.653098	6.24E-13

Supplementary Table S4. Table showing the prior calibration settings for the 9 interior nodes used for divergence dating and four-fold degenerate substitution rate estimation. The mean and standard deviations were designed using the estimates and confidence intervals available on the TimeTree database (Kumar, et al. 2017).

Clade	CI lower	CI upper	Distribution	Mean	Sigma
Human-Chimp	6.23	7.07	Normal	6.4	1
Cobra-Thamnophis	44	59	Normal	51	4
Human-Chimp-Mouse	85	94	Normal	88	2
Pogona-Anole	138	177	Normal	148	6
Root	297	326	Normal	321.3	4
Snakes	72	92	Normal	86	5
Squamates	155	179	Normal	158	4
Pogona-Anole-GlassLizard	152	178	Normal	157	4
Cobra-CottonMouth-Thamno	49	74	Normal	62	5

Supplementary Table S5. PCR primers that amplify W alleles for six *Thamnophis sirtalis* genes. Primers are written 5' to 3'.

Gene	Z scaffold	W scaffold	Primer names	Forward primer sequence	Reverse primer sequence	PCR size (bp)	annealling temp (°C)
COMMD3	NW_013657999	NW_013658453	COMMD-W-F/R	GCTGCTCTCCATTTCCCTGT	TTAAGGGGCAGACGTTCTCG	434	56
CTNNB1	NW_013659367	NW_013658577	CTNNB1-W-F2/R2	ACTGTGTTTTAAAAACATCCCAGA	CTCAAAGCAAGGCCAATYCC	309	52
WDR48	NW_013657923	NW_013658244	WDR48-W-F/R	TCCAGATGTGAGGAAGGGGT	CCTCCAGCCTCCCCAAAAAT	266	57
WAC	NW_013657751	NW_013657718	WAC-W-F/R	GTGAACCTATGGCATGCGTG	GCACACAAGGTTGCAGTCAG	394	56
LSM12	NW_013661471	NW_102658438	LSM12-tsW - F/R	GTTGGTGGTGTGAATTCGCC	GGGCGTACCACACCCATTAT	499	56
MSL1	NW_013657752	NW_013658736	MSL1-WF2/WR2	ACAGCAGGGTTTCAAAGGCT	TCAAAAGGACCTGTGTGCCA	339	56

Supplementary Table S6. Number of annotated olfactory receptors in each genome.

Species	Intact	Pseudogenes	Partial
Green anole*	108	33	1
Bearded dragon	180	65	43
Boa constrictor	441	137	242
Burmese python*	481	319	598
King cobra	368	189	46
Garter snake	306	130	54

*Annotations from Vandewege et al. (2016)

Supplementary Table S7: Amino acid sequences of P-loops in Na_v1.5, Na_v1.8, and Na_v1.9 across sampled taxa. TTX-resistance substitutions are color-coded by their predicted resistance, with warmer colors signifying higher TTX resistance. (Green = 2-10×; yellow = 10-100×; orange = 100-1000×; red = > 1000×. See references for experimentally verified substitutions; the predicted contributions of untested substitutions are based on the effects of other tested substitutions at the same sites). Protein names are similarly color-coded, with warmer colors representing channels with higher predicted TTX resistance.

<i>Xenopus</i>	SFDTFGWAFSLSLFRLMTQDYWENL	WHMHDFHSHSLIVFRVLCGEWIETMWDCME	NFDNVGAGYLALLQVATFKGWMDIMYA	NFTFGNSMICLFQITTSAGWDGLL
<i>Homo</i> Na _v 1.5	SFDSFAWAFSLSLFRLMTQDCWE ^{RL}	WHMMDFFHAFSLIIFRILCGEWIETMWDCME	NFDNVGAGYLALLQVATFKGWMDIMYA	NFQTFANSMCLCFQITTSAGWDGL
<i>Chrysemys</i> Na _v 1.5	SFDTFGWAFSLSLFRLMTQDCWE ^{RL}	WHMNDFFHSHSLIIFRILCGEWIETMWDCME	NFDNVGAGYLALLQVATFKGWMDIMYA	NFQTFANSMCLCFQITTSAGWDGL
<i>Gallus</i> Na _v 1.5	SFDTFGWAFSLSLFRLMTQDYWE ^{RL}	WHMNDFFHSHSLIIFRILCGEWIETMWDCME	NFDNVGAGYLALLQVATFKGWMDIMYA	NFQTFANSMCLCFQITTSAGWDGL
<i>Anolis</i> Na _v 1.5	SFDTFGWAFSLSLFRLMTQDYWE ^{RL}	WHMKDFFHSHSLIIFRILCGEWIETMWDCMR	NFDNVGAGYLALLQVATFKGWMEIMYA	NFQTFANSMCLCFQITTSAGWDGL
<i>Ophisaurus</i> Na _v 1.5	SFDSFGWAFSLSLFRLMTQDCWE ^{RL}	WHMHDFHSHSLIIFRILCGEWIETMWDCML	NFDNVGAGYLALLQVATFKGW ^{TE} IMYA	NFQTFANSMCLCFQITTSAGWDGL
<i>Boa</i> Na _v 1.5	SFDTFGWAFSLSLFRLMTQDCWE ^{RL}	WHMKDFFHSHSLIIFRILCGEWIETMWDCMK	NFDNVGAGYLALLQVATFKGWMEIMYA	NFQTFANSMCLCFQITTSAGW ^{NAL}
<i>Python</i> Na _v 1.5	SFDTFGWAFSLSLFRLMTQDCWE ^{RL}	WHMKDFFHSHSLIIFRILCGEWIETMWDCMK	NFDNVGAGYLALLQVATFKGWMQIMYA	NFETFANSMCLCFQITTSAGWDGL
<i>Cobra</i> Na _v 1.5	SFDTFGWAFSLSLFRLMTQDCWE ^{RL}	WHMKDFFHSHSLIIFRILCGEWIETMWDCMK	NFDNVGAGYLALLQVATFKGWMSIMYA	NFETFANSMCLCFQITTSAGWDGL
<i>Thamnophis</i> Na _v 1.5	SFDTFGWAFSLSLFRLMTQDCWE ^{RL}	WHMKDFFHSHSLIIFRILCGEWIETMWDCMK	NFDNVGAGYLALLQVATFKGWMSIMYA	NFETFANSMCLCFQITTSAGWDGL
<i>Homo</i> Na _v 1.8	SFDSFAWAFSLSLFRLMTQDSWE ^{RL}	WHMHDFHSHSLIVFRILCGEWIENMWACME	NFDNVAMGYLALLQVATFKGWMDIMYA	NFQTFANSMCLCFQITTSAGWDGL
<i>Chrysemys</i> Na _v 1.8	SFDTFGWAFSLSLFRLMTQDYWE ^{RL}	WHFMDFFHSHSLIIFRILCGEWIETMWGCMV	NFDNVGSGYLALLQVATFKGWMDIMYA	NFQTFANSMCLCFQITTSAGWDGL
<i>Gallus</i> Na _v 1.8	SFDTFGWAFSLSLFRLMTQDYWE ^{RL}	WHMMDFFHSHSLIVFRILCGEWIETMWDCMV	NFDNVGSGYLALLQVATFKGWMEIMYA	NFQTFANSMCLCFEITTSAGWDGL
<i>Anolis</i> Na _v 1.8	SFDTFGWAFSLSLFRLMTQDYWE ^{RL}	WHMEDFFHSHSLIVFRILCGEWIETMWGCMV	NFDNVGSGYLSLLQVATFKGWMDIMYA	NFKTFGNSMLCLCFQITTSAGW ^{AEL}
<i>Ophisaurus</i> Na _v 1.8	SFDTFGWAFSLSLFRLMTQD ^N WE ^{RL}	WHMQDFFHSHSLIIFRILCGEWIETMWGCMV	NFDNVGSGYLALLQVATFKGWMDIMYA	NFETFGNSMLCLFEITTSAGWDGL
<i>Boa</i> Na _v 1.8	SFDTFGWAFSLSLFRLMTQDYWE ^{RL}	WHMEDFFHSHSLIIFRILCGEWIETMWQCMV	NFDNVGSGYLALLQVATFKGWMDIMYA	NFETFGNSMLCLCFQITTSAGWDGL
<i>Python</i> Na _v 1.8	SFDTFGWAFSLSLFRLMTQDYWE ^{RL}	WHMEDFFHSHSLIIFRILCGEWIETMWHCMV	NFDNVGLGYLALLQVATFKGWMDIMYA	NFETFGNSMLCLCFQITTSAGWDGL
<i>Cobra</i> Na _v 1.8	SFDTFGWAFSLSLFRLMTQDYWE ^{RL}	WHMEDFFHSHSLIIFRILCGEWIETMWHCMV	NFDNVGLGYLALLQVATFKGWMDIMYA	NFETFGNSMLCLCFQITTSAGWDGL
<i>Thamnophis</i> Na _v 1.8	SFDTFGWAFSLSLFRLMTQDYWE ^{RL}	WHMEDFFHSHSLIIFRILCGEWIETMWHCMV	NFDNVGLGYLALLQVATFKGW ^T DIMYA	NFETFGNSMLCLCFQITTSAGWDGL
<i>Homo</i> Na _v 1.9	NFDNFGWSFLAMFRLMTQDSWE ^{KL}	WHMGDFWHSFLVFRILCGEWIENMWECMQ	NFDNVGNAYLALLQVATFKGWMDI IYA	NFKTFASSMLCLCFQISTAGWD ^{SL}
<i>Chrysemys</i> Na _v 1.9	SFDHFGWAFSLSVFRLMTQD ^N WE ^{RL}	WHMIDFFHSHSLIIFRILCGEWIETLWGCME	TFDHVGMGYLALLQVATFKGWMDIMYA	NFQTFGGSMLCLCFQITTSAGWDGL
<i>Gallus</i> Na _v 1.9	SFDHFGWAFSLSLFRLMTQDSWE ^{RL}	WHMKDFYHSFLVIFRILCGEWIETMWECME	NFDHVGMYLALLQVATFKGWMDIMYA	NFQTFDSSILCLCFQITTSAGWD ^{SL}
<i>Anolis</i> Na _v 1.9	XXXXXXXXXXXXXXXXXXXXXXXXXX	WHMEDFFHSHSLIIFRILCGEWIETMWGCMV	NFDNIMKGYLALLQVATFKGWMEIMYA	NFQTFGGSMLCLFEITTSAGW ^{HGF}
<i>Ophisaurus</i> Na _v 1.9	SFDHFGWAFSLSLFRLMTQDSWE ^{RL}	WHMKDFLHSHSLIIFRILCGEWI ^D TMMGCMQ	NFDNVWKGYLALLQVATFKGWMQIMYA	NFQTFGGSMLCLCFQITTSAGWD ^{EL}
<i>Boa</i> Na _v 1.9	SFDHLGWAFSLSLFRLMTQDSWE ^{RL}	WHMKDFLHSHSLIIFRILCGEWI ^{STM} WECME	NFDHVGKGYLALLQVATFKGW ^{LD} DIMYA	NFQTFGNSMLCLCFQIMTSAGW ^{NEL}
<i>Python</i> Na _v 1.9	SFDHLGWAFSLSLFRLMTQDSWE ^{CL}	WHMKDFLHSHSLIIFRILCGEWI ^{STM} WECMQ	NFDHVIGYLYLALLQVATFKGW ^T DIMYA	NFQTFGNSMLCLCFQITTSAGW ^{NEL}
<i>Cobra</i> Na _v 1.9	SFDHLGWAFSLSLFRLMTQDSWE ^{GL}	WHMKDFLHSHSLIIFRILCGEWI ^{STM} WECMQ	NFDHVGNGYLYLALLQVATFKGWMDIMYA	NFQTFGNSMLCLCFQITTSAGWD ^{EL}
<i>Thamnophis</i> Na _v 1.9	SFDHLGWAFSLSLFRLMTQDSWE ^{RL}	WHMKDFLHSHSLIIFRILCGEWI ^{STM} WECMQ	XXXXXXXXXXXXXXXXATFKGWMDIMYA	NFQTFGNSMLCLCFQ ^M TTTSAGWD ^{EL}

References: (Backx, et al. 1992; Choudhary, et al. 2003; Geffeney, et al. 2005; Jost, et al. 2008; Leffler, et al. 2005; Penzotti, et al. 1998; Terlau, et al. 1991)

4. SUPPLEMENTARY FIGURES

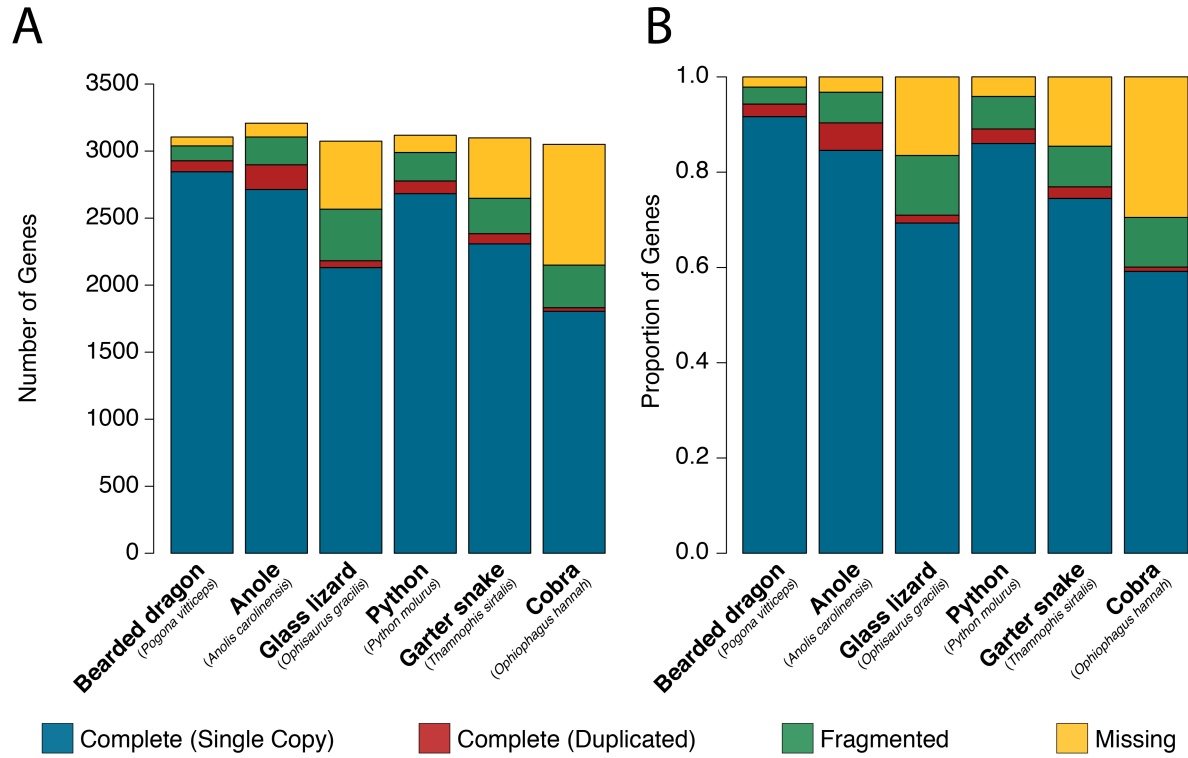


Figure S1. The number (**A**) and proportion (**B**) of BUSCOs identified in the *Thamnophis* genome and 5 other squamate reptile genomes.

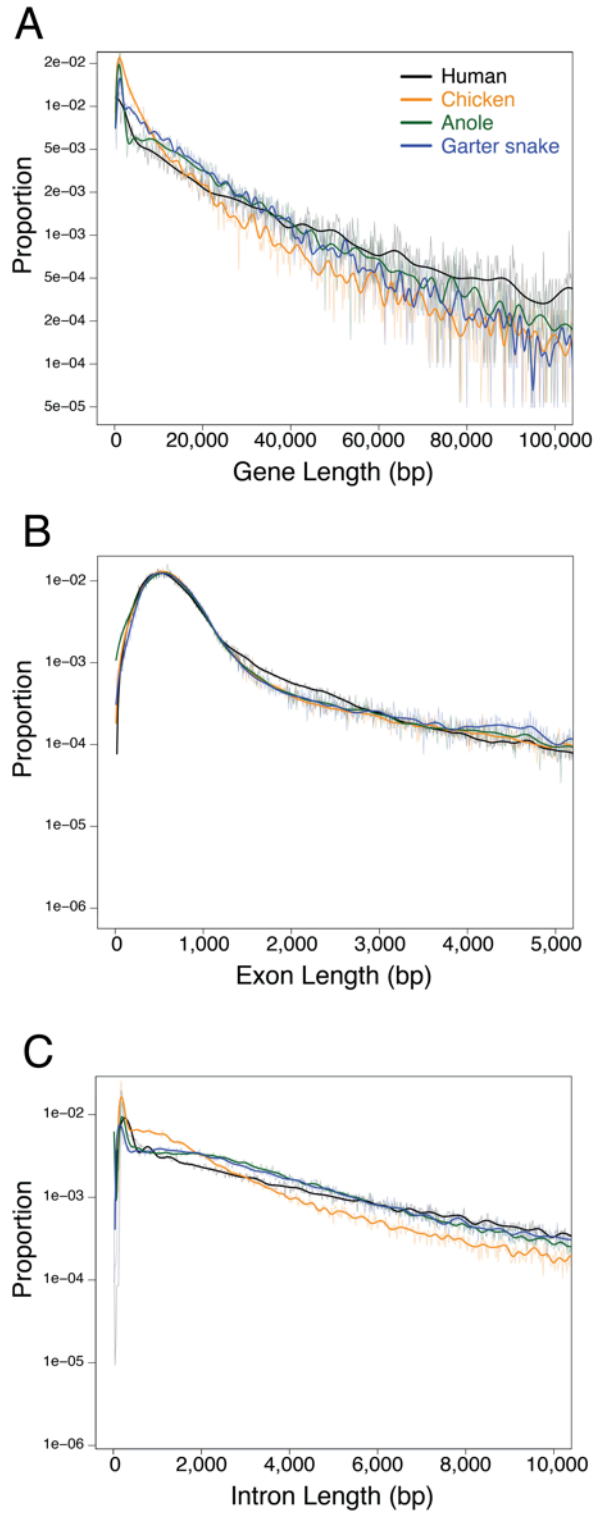


Figure S2. Smoothed spline distributions of A) gene, B) exon, and C) intron length and relative frequency in the human, chicken, anole, and garter snake genomes.

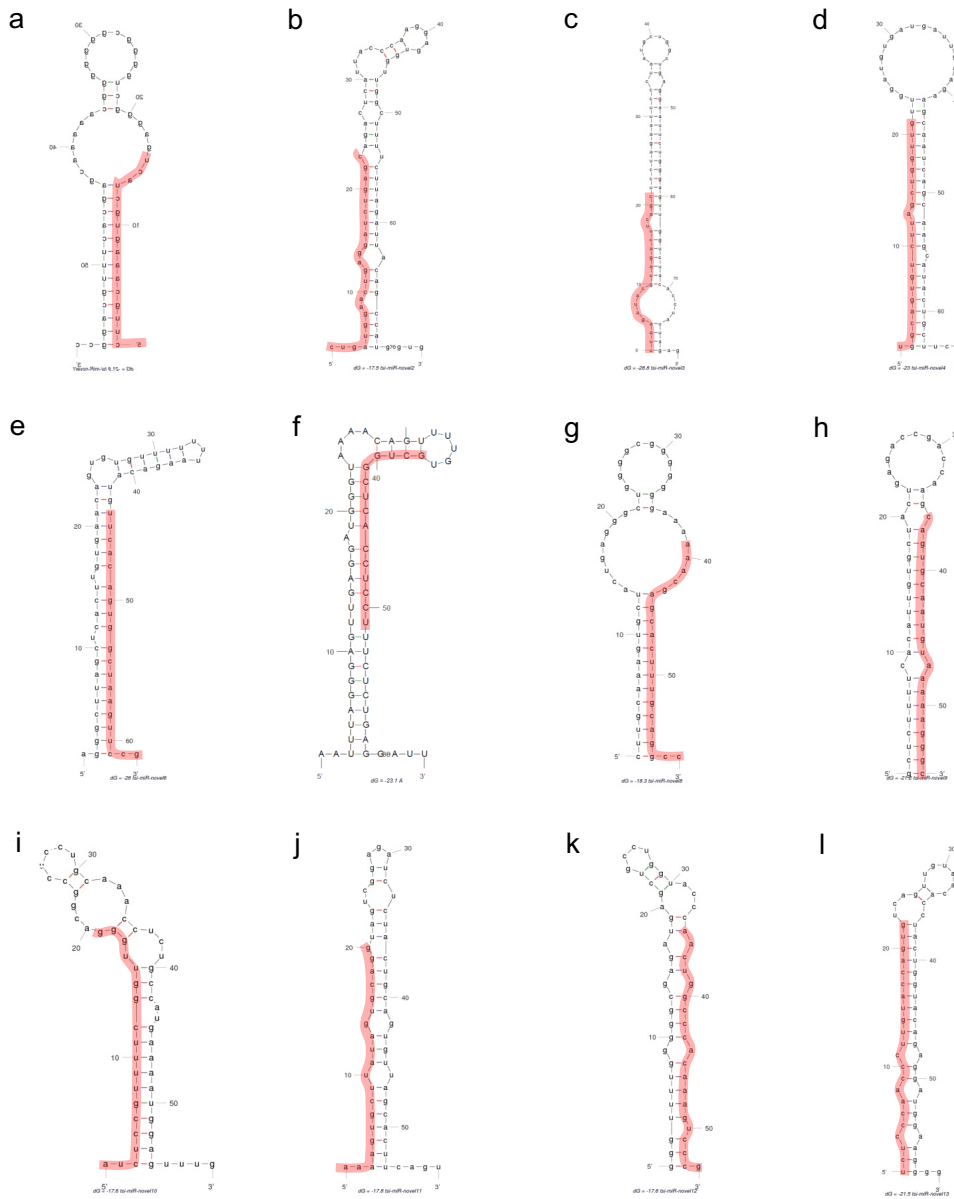


Figure S3. Species-specific microRNA precursor structure. (a) *tsi-miR-novel1*, (b) *tsi-miR-novel2*, (c) *tsi-miR-novel3*, (d) *tsi-miR-novel4-1* and *tsi-miR-novel4-2*, (e) *tsi-miR-novel5*, (f) *tsi-miR-novel6*, (g) *tsi-miR-novel7*, (h) *tsi-miR-novel8*, (i) *tsi-miR-novel9*, (j) *tsi-miR-novel10*, (k) *tsi-miR-novel11*, and (l) *tsi-miR-novel12-1* and *tsi-miR-novel12-2*.

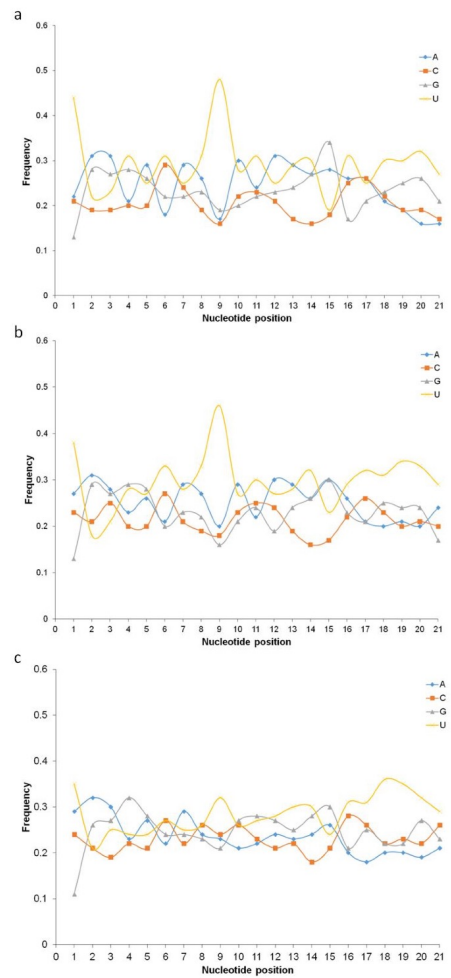


Figure S4. Nucleotide frequency between the (A) garter snake, (B) king cobra, and (C) human mature microRNA.

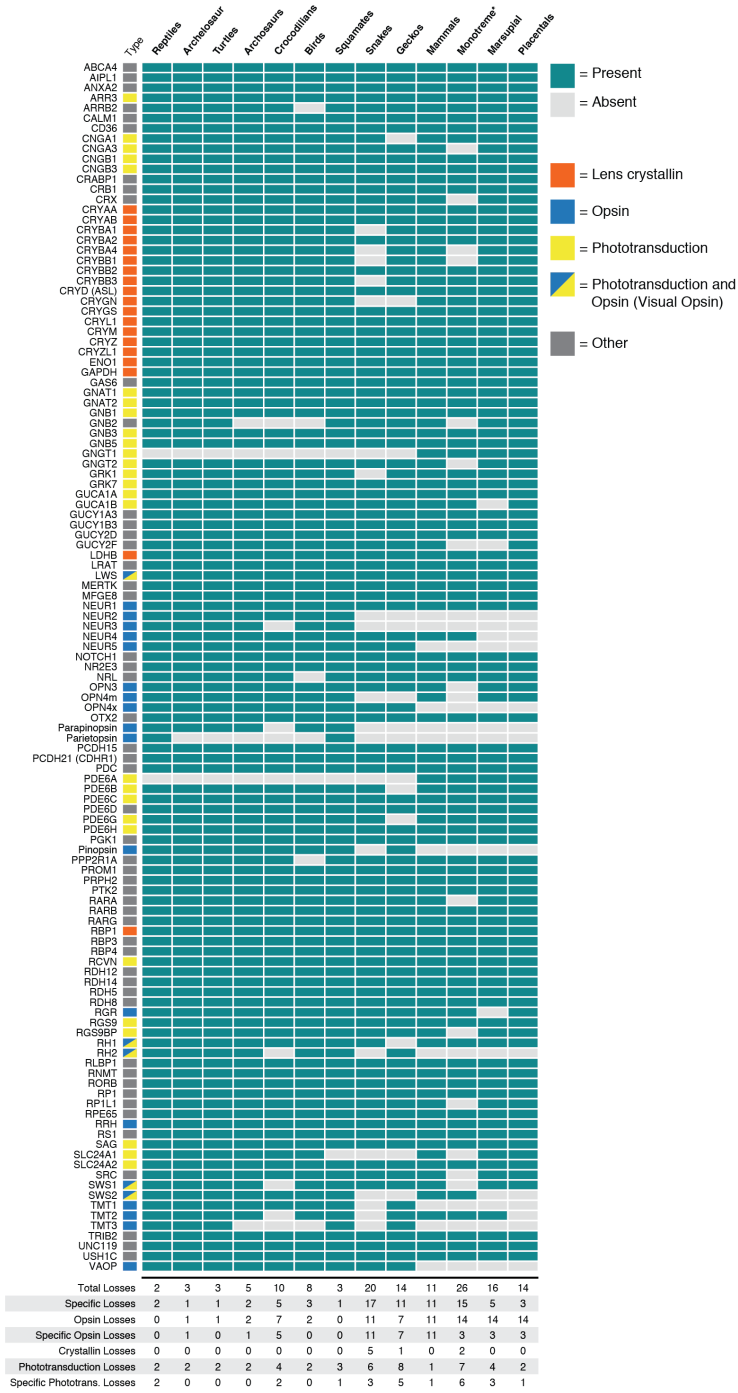


Figure S5. Visual gene loss in groups with nocturnal ancestry, with total numbers of gene losses given below. *Note: Losses inferred for monotremes are less reliable as they are based on a single genome and thus may be artefacts of incomplete genome sequencing and assembly rather than true losses.

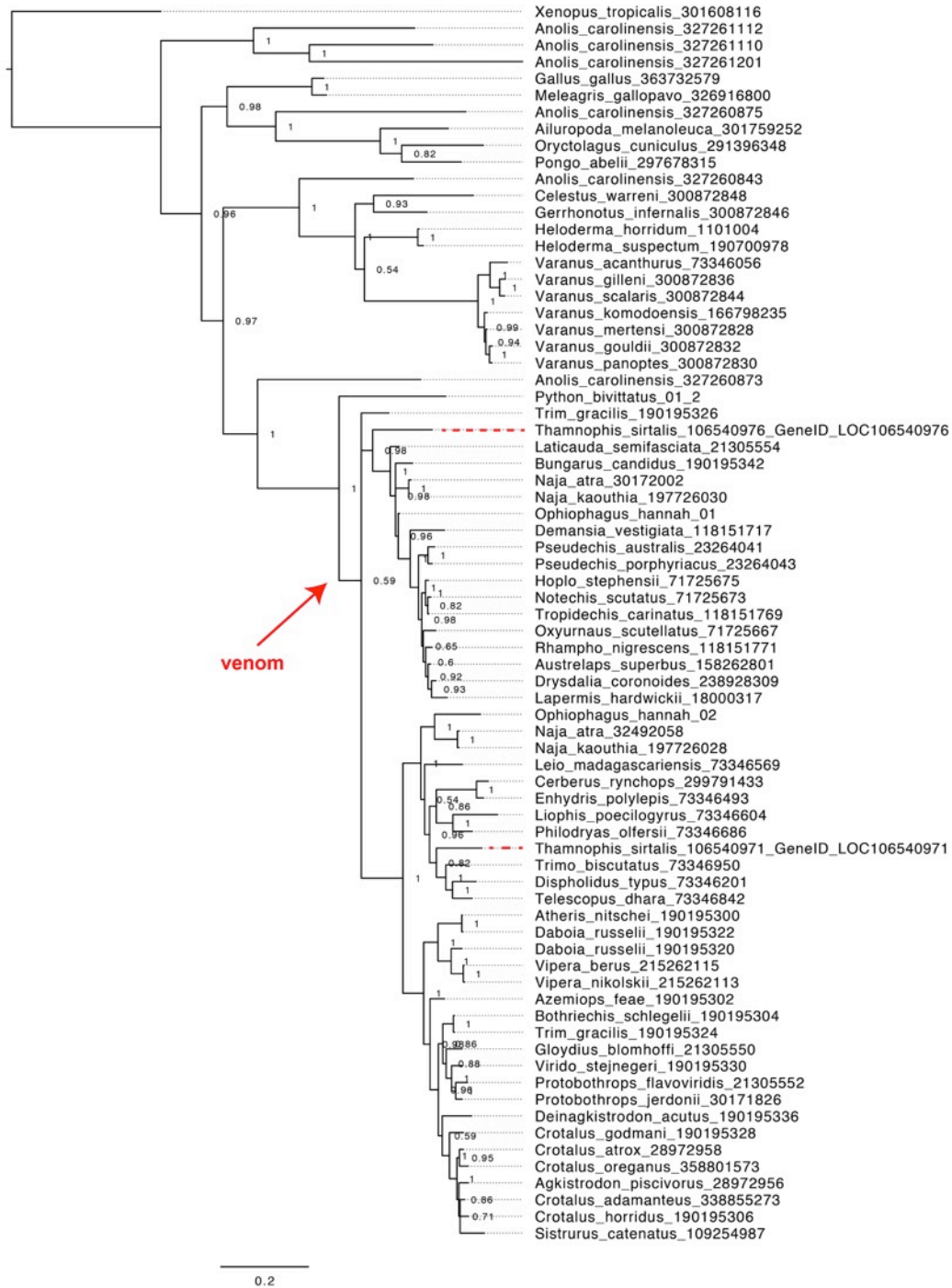


Figure S6. The cysteine-rich secretory protein (CRISP) gene family. The venom clade is indicated with a red arrow, and *Thamnophis sirtalis* venom gene orthologs are indicated with red dotted lines.

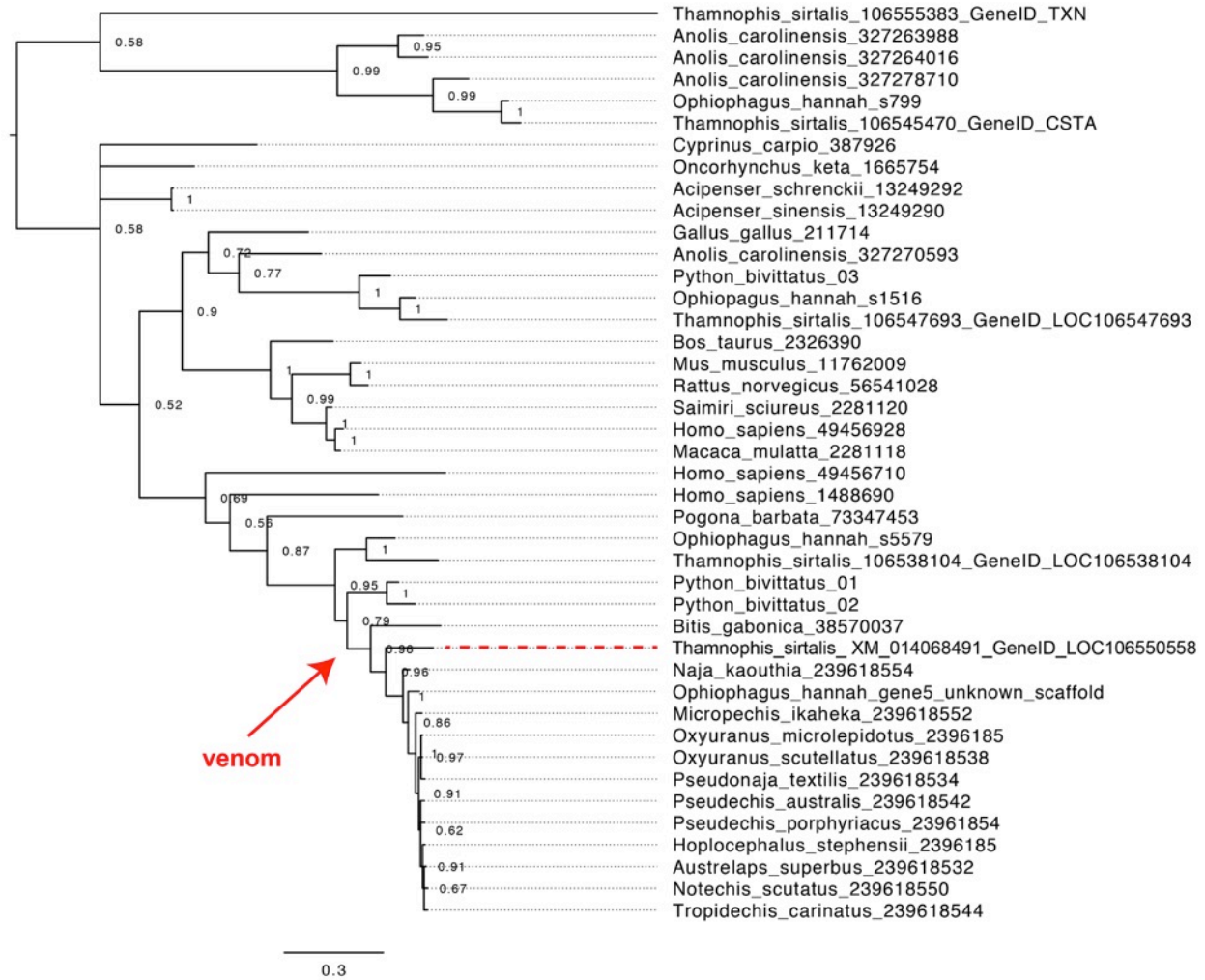


Figure S7. The cystatin gene family. The venom clade is indicated with a red arrow, and *Thamnophis sirtalis* venom gene orthologs are indicated with red dotted lines.

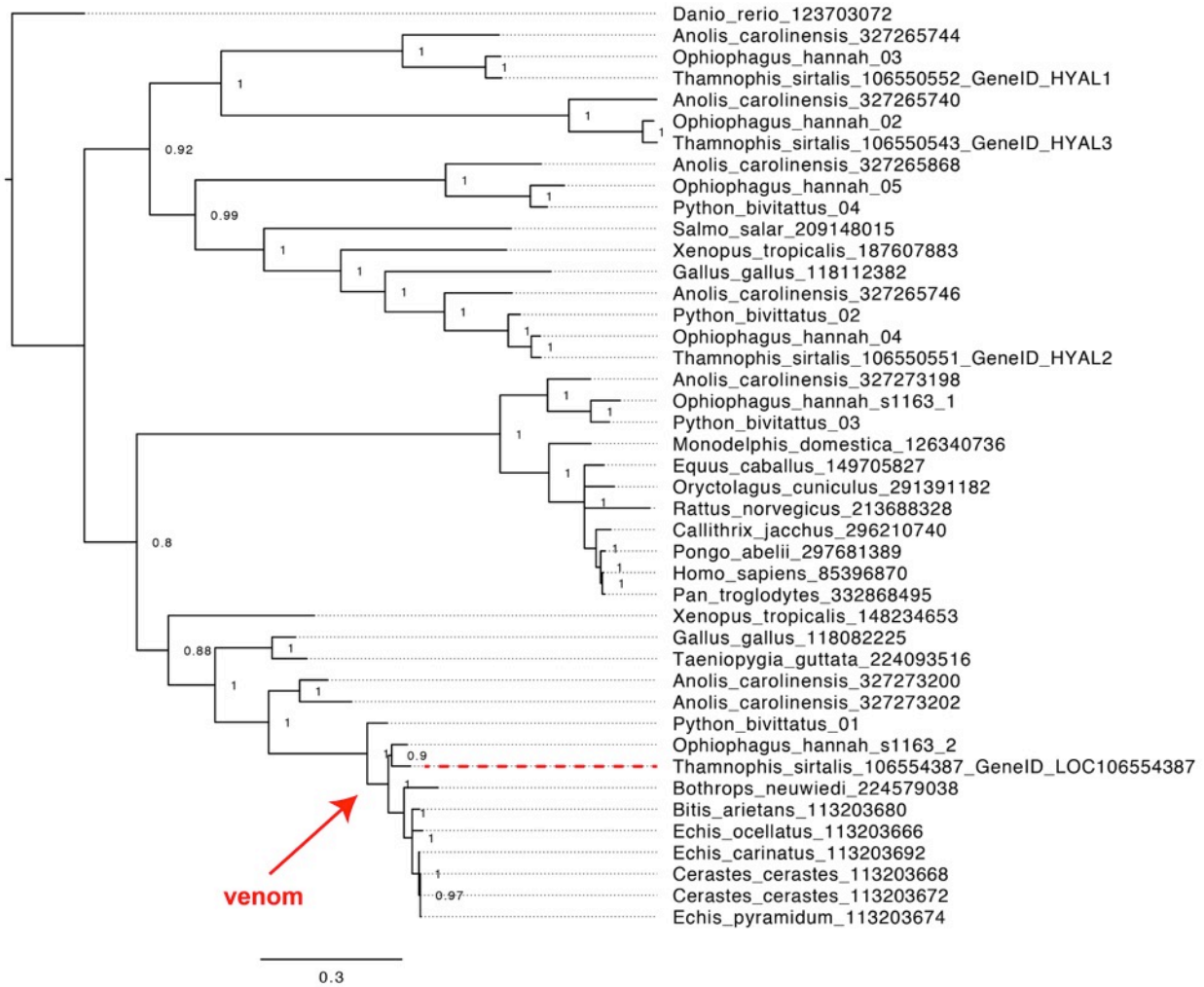


Figure S8. The hyaluronidase gene family. The venom clade is indicated with a red arrow, and *Thamnophis sirtalis* venom gene orthologs are indicated with red dotted lines.

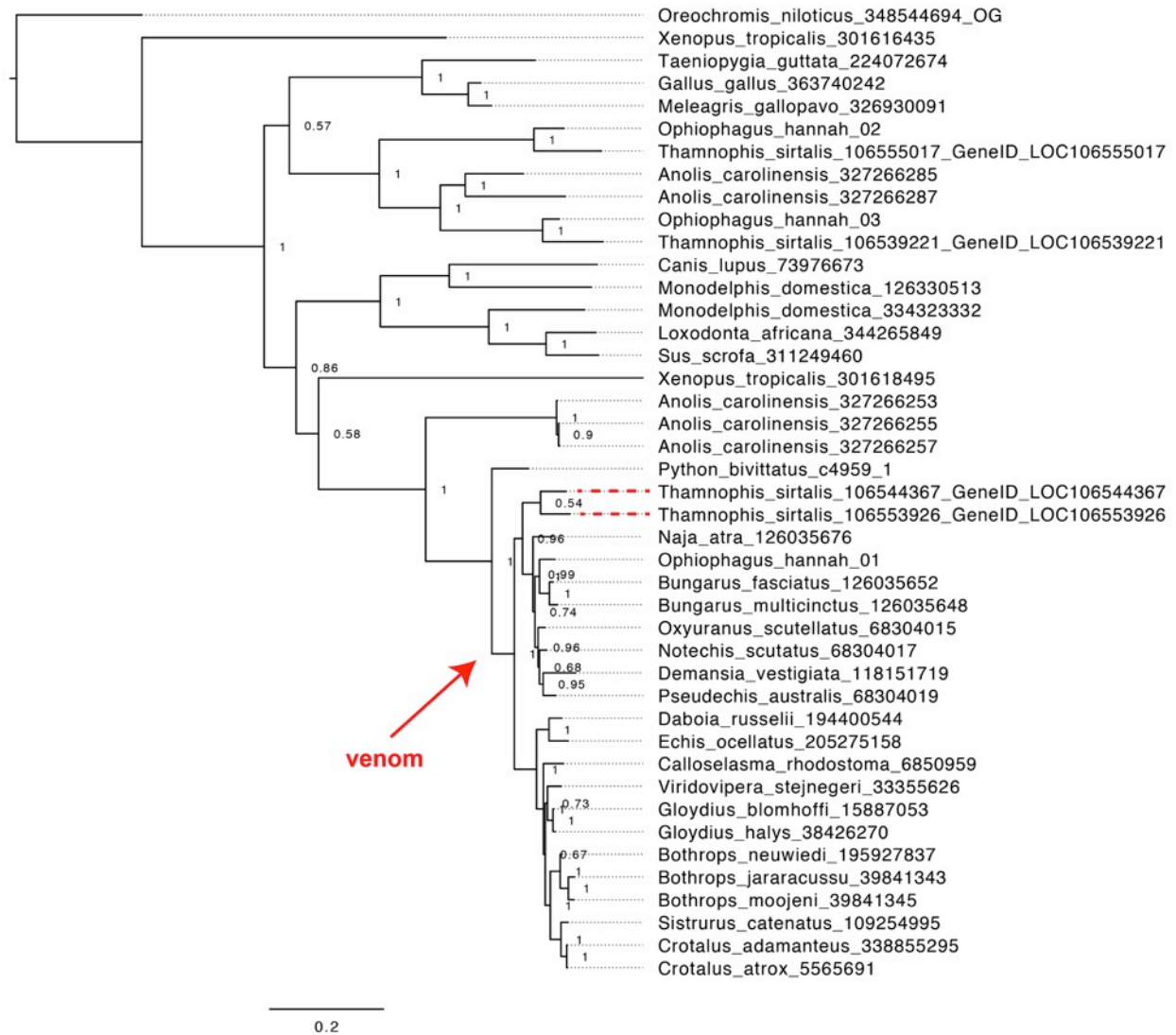


Figure S9. The L-amino acid oxidase (LAO) gene family. The venom clade is indicated with a red arrow, and *Thamnophis sirtalis* venom gene orthologs are indicated with red dotted lines.

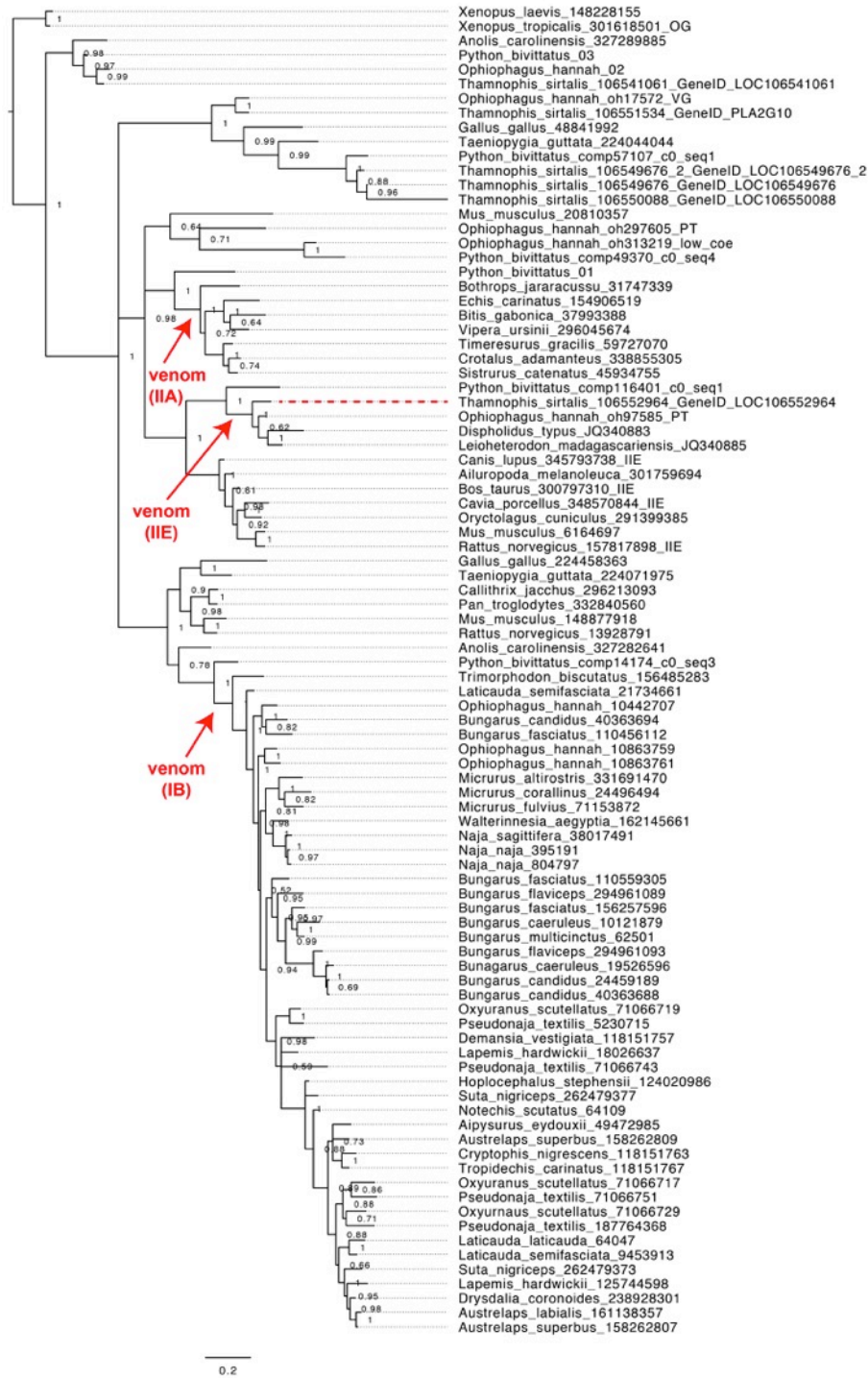


Figure S10. The phospholipase A₂ (PLA₂) gene family. The venom clades are indicated with a red arrow, and *Thamnophis sirtalis* venom gene orthologs are indicated with red dotted lines.

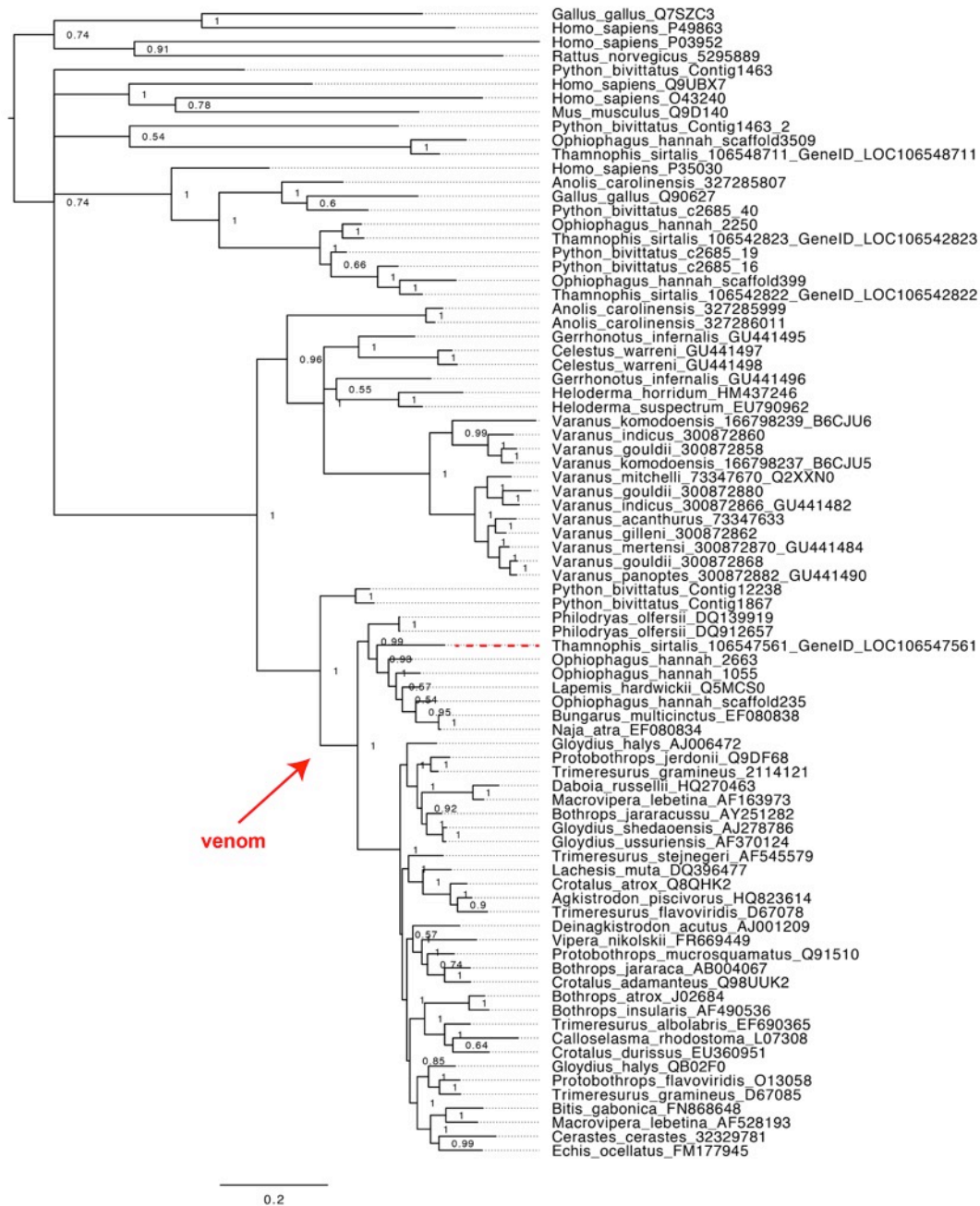


Figure S11. The snake venom serine protease (SVSP) gene family. The venom clade is indicated with a red arrow, and *Thamnophis sirtalis* venom gene orthologs are indicated with red dotted lines.

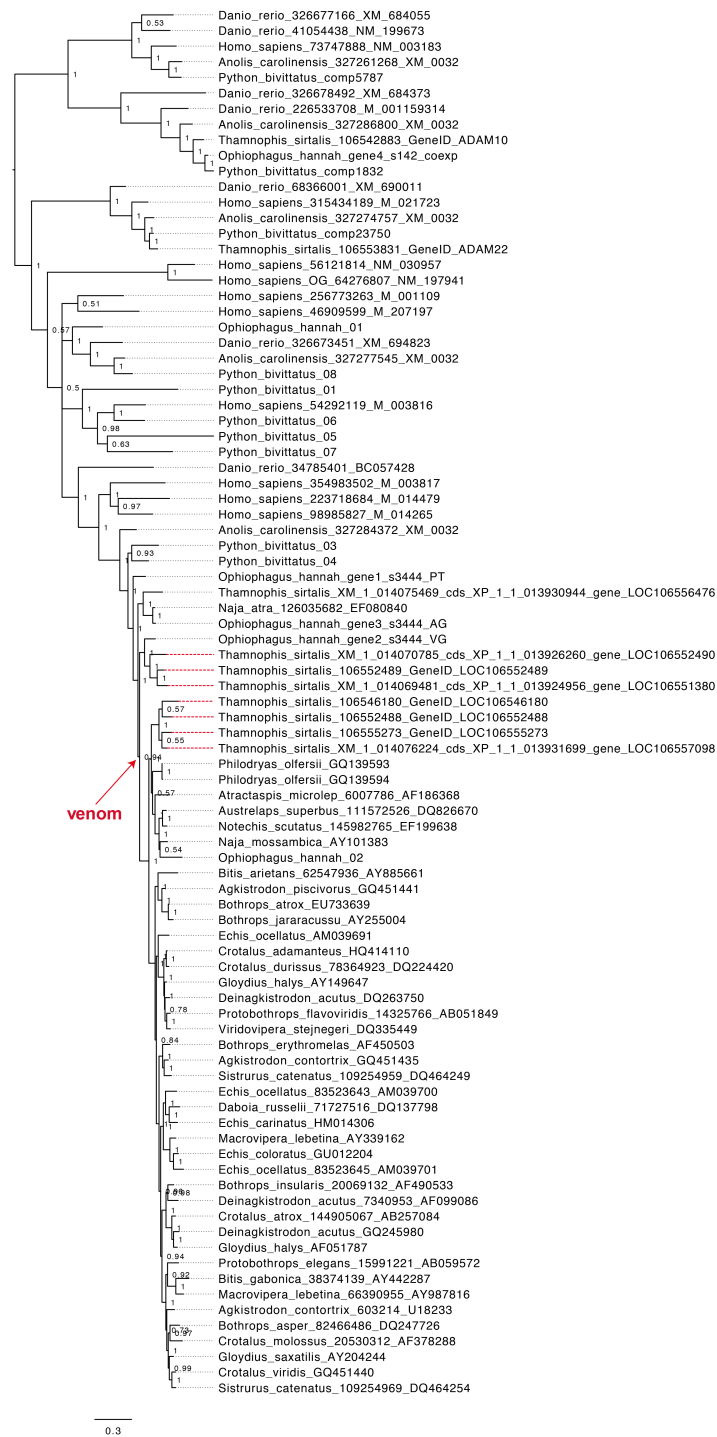


Figure S12. The snake venom metalloproteinase (SVMP) gene family. The venom clade is indicated with a red arrow, and *Thamnophis sirtalis* venom gene orthologs are indicated with red dotted lines.

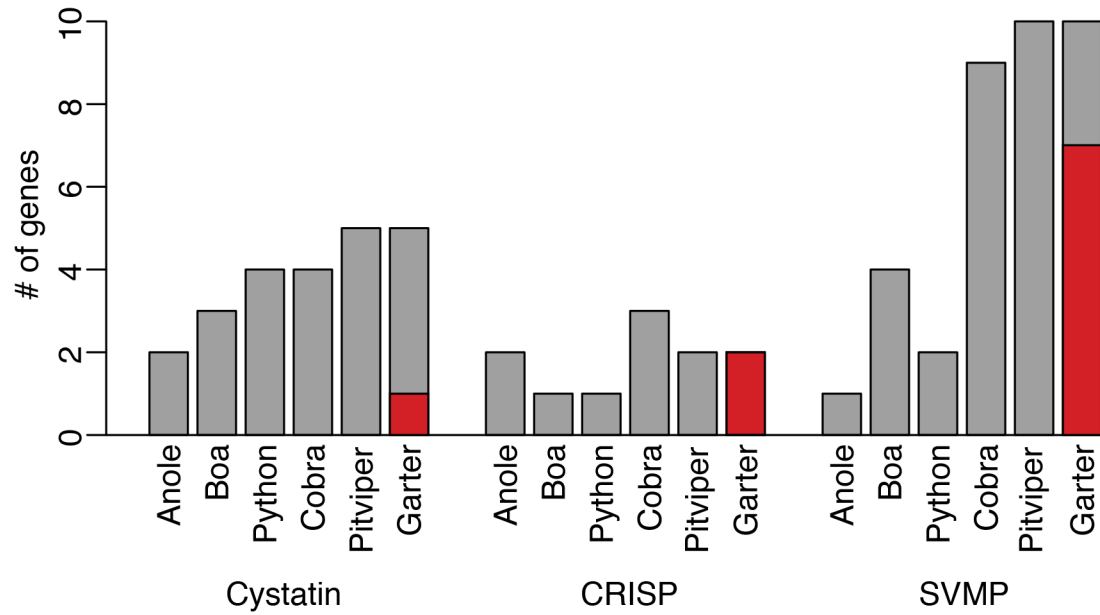


Figure S13. Total number of cystatin, CRISP, and SVMP toxin family genes identified in the genomes of the garter snake (*Thamnophis sirtalis*), two front-fanged venomous snakes, the five-pace viper (*Deinagistrodon acutus*) and king cobra (*Ophiophagus hannah*), two nonvenomous snakes, the Burmese python (*Python molurus*) and *Boa constrictor*, and one nonvenomous lizard, the green anole (*Anolis carolinensis*). Red bars indicate the number of genes in the garter snake genome that clustered with known snake venom toxins in phylogenetic reconstructions, indicative of a role in venom. Numbers for all species other than the garter snake are derived from (Yin, et al. 2016).

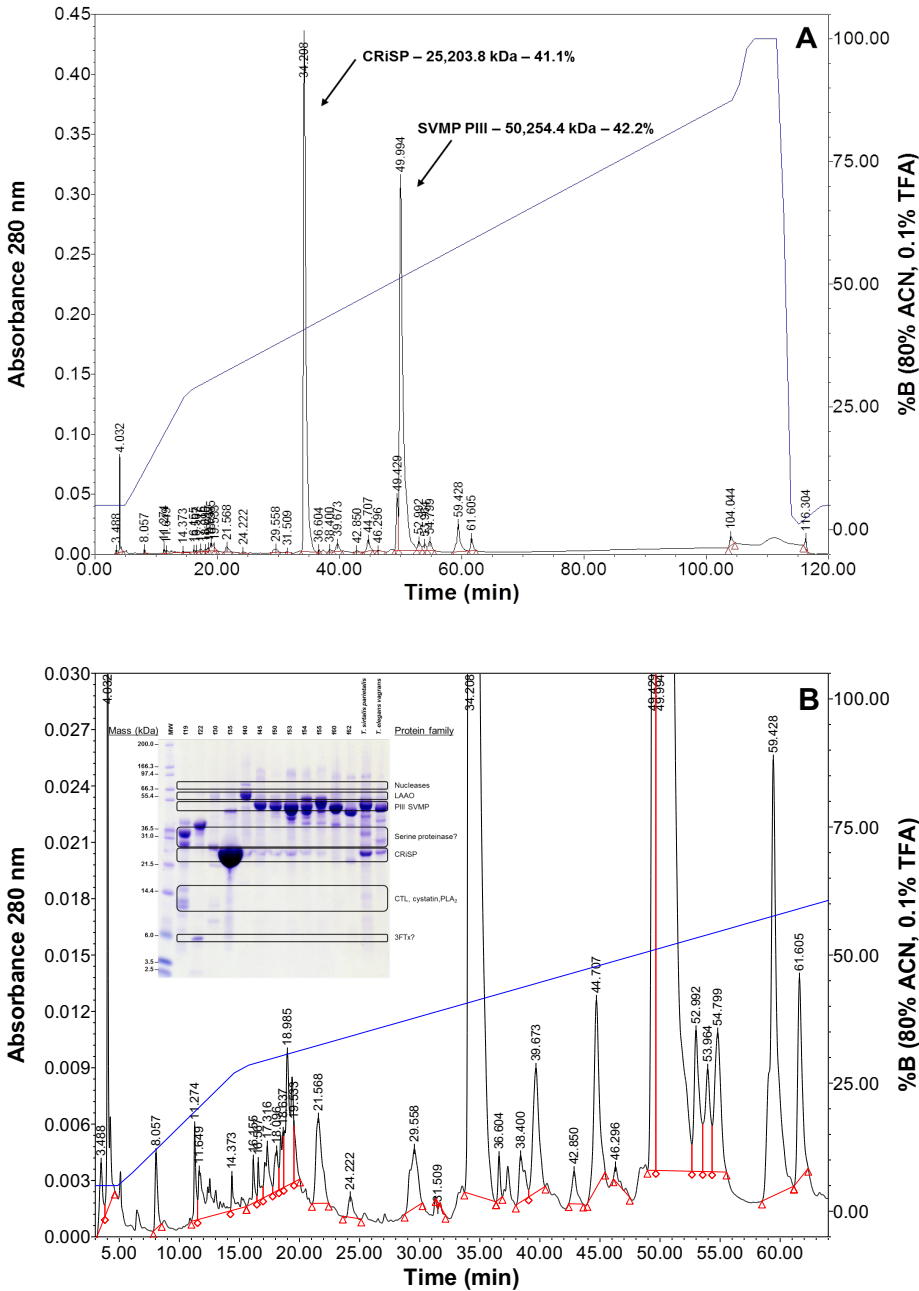


Figure S14. A) RP-HPLC fractionation of 1.0 mg *Thamnophis sirtalis parietalis* (Oregon - EB) on a Jupiter C18 column. 1.0 mL/min – 18 Jan 2016. Note that CriSP and SVMP P-III metalloproteinases dominate. B) adjusted *Y-axis*. Inset is a 12% acrylamide NuPAGE SDS-PAGE gel of major fractions, indicating major protein families represented.

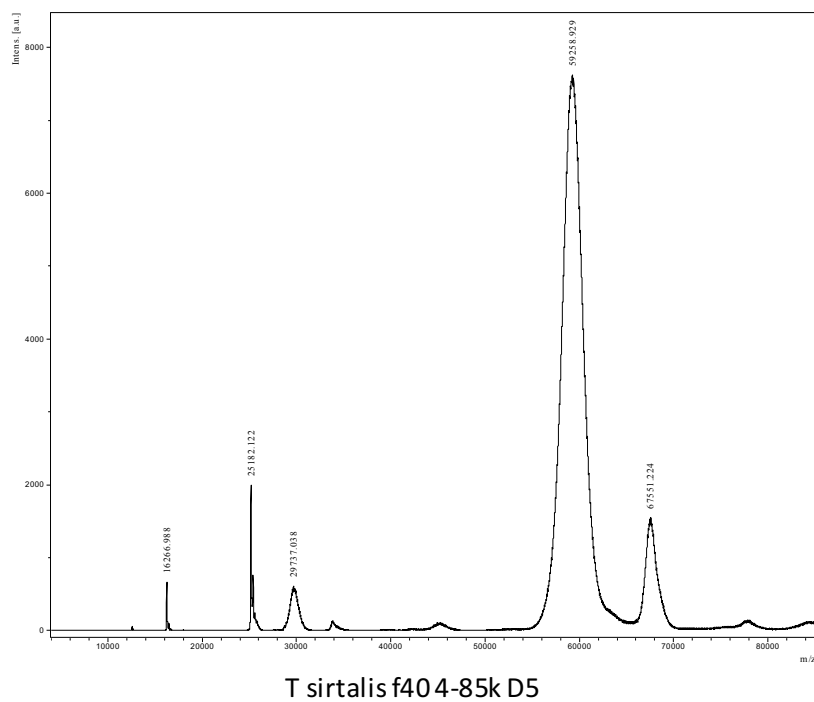


Figure S15. MALDI-TOF mass spectrum of RP-HPLC fraction 40 - LAAO monomer – 59,258.9 Da. 0.76% of total protein.

Works Cited

- Alföldi J, et al. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* 477. doi: 10.1038/nature10390
- Altschul SF, et al. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
- Backx PH, et al. 1992. Molecular localization of an ion-binding site within the pore of mammalian sodium channels. *Science* 257: 248-252.
- Benjamini Y, Hochberg Y 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Ser. B. (Stat. Method.)*: 289-300.
- Bolger AM, Lohse M, Usadel B 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114-2120.
- Bradnam KR, et al. 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience* 2: 10.
- Castoe TA, et al. 2013. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc. Natl. Acad. Sci. USA* 110: 20645-20650. doi: 10.1073/pnas.1314475110
- Choudhary G, et al. 2003. Interactions of the C-11 hydroxyl of tetrodotoxin with the sodium channel outer vestibule. *Biophys. J.* 84: 287-294.
- Dobin A, et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15-21.
- Drummond AJ, Rambaut A 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7: 214.
- Edgar RC 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32: 1792-1797.
- Enright AJ, et al. 2003. MicroRNA targets in *Drosophila*. *Genome Biol.* 5: R1.
- Ezaz T, et al. 2008. A simple non-invasive protocol to establish primary cell lines from tail and toe explants for cytogenetic studies in Australian dragon lizards (Squamata: Agamidae). *Cytotechnology* 58: 135-139.
- Ezaz T, et al. 2005. The dragon lizard *Pogona vitticeps* has ZZ/ZW micro-sex chromosomes. *Chromosome Res.* 13: 763-776.
- Feldman CR, Brodie Jr. ED, Brodie III ED, Pfrender ME 2012. Constraint shapes convergence in tetrodotoxin-resistant sodium channels of snakes. *Proc. Natl. Acad. Sci. USA* 109: 4556-4561.
- Gamble T, et al. 2017. The discovery of XY sex chromosomes in a boa and python. *Curr. Biol.* 27: 2148-2153. e2144.

- Gamble T, Geneva AJ, Glor RE, Zarkower D 2014. Anolis sex chromosomes are derived from a single ancestral pair. *Evolution* 68: 1027-1041.
- Geffeney SL, et al. 2005. Evolutionary diversification of TTX-resistant sodium channels in a predator-prey interaction. *Nature* 434: 759-763.
- Georges A, et al. 2015. High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard *Pogona vitticeps*. *Gigascience* 4: 45.
- Glusman G, et al. 2000. The olfactory receptor gene superfamily: data mining, classification, and nomenclature. *Mamm. Genome* 11: 1016-1023.
- Goldman N, Whelan S 2002. A novel use of equilibrium frequencies in models of sequence evolution. *Mol. Biol. Evol.* 19: 1821-1831.
- Hayden S, et al. 2010. Ecological adaptation determines functional mammalian olfactory subgenomes. *Genome Res.* 20: 1-9.
- Hill RE, Mackessy SP 1997. Venom yields from several species of colubrid snakes and differential effects of ketamine. *Toxicon* 35: 671-678.
- Holm S 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*: 65-70.
- Huelsenbeck JP, Ronquist F 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.
- Ijdo JW, Wells RA, Baldini A, Reeders ST 1991. Improved telomere detection using a telomere repeat probe (TTAGGG)_n generated by PCR. *Nucleic Acids Res.* 19: 4780.
- Jost MC, et al. 2008. Toxin-resistant sodium channels: parallel adaptive evolution across a complete gene family. *Mol. Biol. Evol.* 25: 1016-1024.
- Junqueira-de-Azevedo ILM, Campos PF, Ching ATC, Mackessy SP 2016. Colubrid venom composition: An-omics perspective. *Toxins* 8: 230.
- Kearse M, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647-1649.
- Kohany O, Gentles AJ, Hankus L, Jurka J 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 7: 474.
- Krämer A, Green J, Pollard Jr J, Tugendreich S 2013. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 30: 523-530.
- Kumar S, Stecher G, Suleski M, Hedges SB 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* 34: 1812-1819.

- Lanfear R, Calcott B, Ho SYW, Guindon S 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* 29: 1695-1701.
- Laopichienpong N, et al. 2017. Evolutionary Dynamics of the Gametologous CTNNB1 Gene on the Z and W Chromosomes of Snakes. *J. Hered.* 108: 142-151.
- Leffler A, et al. 2005. Pharmacological properties of neuronal TTX-resistant sodium channels and the role of a critical serine pore residue. *Pflügers Archiv* 451: 454-463.
- Li L, Stoeckert CJ, Roos DS 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13: 2178-2189.
- Liu B, et al. 2013. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *arXiv preprint arXiv:1308.2012*.
- Liu Y, Schmidt B, Maskell DL 2010. MSAProbs: multiple sequence alignment based on pair hidden Markov models and partition function posterior probabilities. *Bioinformatics* 26: 1958-1964.
- Liu Y, et al. 2015. Gekko japonicus genome reveals evolution of adhesive toe pads and tail regeneration. *Nat. Commun.* 6: 10033.
- Marçais G, Kingsford C 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27: 764-770.
- Matsubara K, et al. 2006. Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. *Proc. Natl. Acad. Sci. USA* 103: 18190-18195. doi: 0605274103 [pii] 10.1073/pnas.0605274103
- McGaugh SE, et al. 2015. Rapid molecular evolution across amniotes of the IIS/TOR network. *Proc. Natl. Acad. Sci. USA* 112: 7055-7060.
- McGlothlin JW, et al. 2014. Parallel evolution of tetrodotoxin resistance in three voltage-gated sodium channel genes in the garter snake *Thamnophis sirtalis*. *Mol. Biol. Evol.* 31: 2836-2846.
- Miller MA, Pfeiffer W, Schwartz T editors. Gateway Computing Environments Workshop (GCE), 2010. 2010.
- Nanda I, et al. 1990. Simple repetitive sequences are associated with differentiation of the sex chromosomes in the guppy fish. *J. Mol. Evol.* 30: 456-462.
- O'Meally D, et al. 2010. Non-homologous sex chromosomes of birds and snakes share repetitive sequences. *Chromosome Res.* 18: 787-800.
- Olender T, Nativ N, Lancet D 2013. HORDE: comprehensive resource for olfactory receptor genomics. *Olfactory Receptors: Methods and Protocols*: 23-38.

- Peace RJ, Biggar KK, Storey KB, Green JR 2015. A framework for improving microRNA prediction in non-human genomes. *Nucleic Acids Res.* 43: e138-e138.
- Penzotti JL, Fozzard HA, Lipkind GM, Dudley SC 1998. Differences in saxitoxin and tetrodotoxin binding revealed by mutagenesis of the Na⁺ channel outer vestibule. *Biophys. J.* 75: 2647-2657.
- Reyes-Velasco J, et al. 2014. Expression of venom gene homologs in diverse python tissues suggests a new model for the evolution of snake venom. *Mol. Biol. Evol.* 32: 173-183.
- Robinson MD, McCarthy DJ, Smyth GK 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139-140.
- Singh L, Purdom IF, Jones KW 1980. Sex chromosome associated satellite DNA: evolution and conservation. *Chromosoma* 79: 137-157.
- Smit AFA, Hubley R, Green P 2014. RepeatModeler Open-1.0. 2008-2010. Access date Dec.
- Smit AFA, Hubley R, Green P 2015. RepeatMasker Open-4.0. 2013–2015. Institute for Systems Biology. <http://reapeatmasker.org>.
- Smith CF, Mackessy SP 2016. The effects of hybridization on divergent venom phenotypes: characterization of venom from *Crotalus scutulatus scutulatus* × *Crotalus oreganus helleri* hybrids. *Toxicon* 120: 110-123.
- Stamatakis A, Hoover P, Rougemont J 2008. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* 57: 758-771.
- Tamura K, et al. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30: 2725-2729.
- Terlau H, et al. 1991. Mapping the site of block by tetrodotoxin and saxitoxin of sodium channel II. *FEBS Lett.* 293: 93-96.
- Ullate-Agote A, Milinkovitch MC, Tzika AC 2015. The genome sequence of the corn snake (*Pantherophis guttatus*), a valuable resource for EvoDevo studies in squamates. *Int. J. Dev. Biol.* 58: 881-888.
- Vicoso B, et al. 2013. Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Biol.* 11: e1001643.
- Vonk FJ, et al. 2013. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc. Natl. Acad. Sci. USA* 110: 20651-20656. doi: 10.1073/pnas.1314702110
- Yang Z 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24: 1586-1591.

Yin W, et al. 2016. Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. *Nat. Commun.* 7.

Zakon HH, Jost MC, Lu Y 2010. Expansion of voltage-dependent Na⁺ channel gene family in early tetrapods coincided with the emergence of terrestriality and increased brain complexity. *Mol. Biol. Evol.* 28: 1415-1424.