



# The transcriptomic and proteomic basis for the evolution of a novel venom phenotype within the Timber Rattlesnake (*Crotalus horridus*)



Darin R. Rokyta<sup>\*</sup>, Kenneth P. Wray, James J. McGivern, Mark J. Margres

Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA

## ARTICLE INFO

### Article history:

Received 15 December 2014

Received in revised form

6 February 2015

Accepted 25 February 2015

Available online 26 February 2015

### Keywords:

Gene expression

Genetics of adaptation

Venom evolution

Snake venom

Hybridization

## ABSTRACT

The genetics underlying adaptive trait evolution describes the intersection between the probability that particular types of mutation are beneficial and the rates they arise. Snake venoms can vary in a directly meaningful manner through coding mutations and regulatory mutations. The amounts of different components determine venom efficacy, but point mutations in coding sequences can also change efficacy and function. The Timber Rattlesnake (*Crotalus horridus*) has populations that have evolved neurotoxic venom from the typical hemorrhagic rattlesnake venom present throughout most of its range. We identified only a handful of nonsynonymous differences in just five loci between animals with each venom type, and these differences affected lower-abundance toxins. Expression of at least 18 loci encoding hemorrhagic toxins was severely reduced in the production of neurotoxic venom. The entire phospholipase A<sub>2</sub> toxin family was completely replaced in the neurotoxic venom, possibly through intergeneric hybridization. Venom paedomorphosis could, at best, explain only some of the loss of expression of hemorrhagic toxins. The number of potential mechanisms for altering venom composition and the patterns observed for *C. horridus* suggest that rapid venom evolution should occur primarily through changes in venom composition, rather than point mutations affecting coding sequences.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Adaptation should proceed along more likely mutational pathways, provided mutations on those pathways are beneficial (Rokyta et al., 2005; Weinreich et al., 2006). The genetics underlying adaptive trait evolution can therefore describe the intersection between the probability that a given class of mutation is beneficial and the rate at which that class of mutation arises. The form of this intersection may vary among traits, with some, such as morphological traits, perhaps favoring regulatory mutations because of intense pleiotropy for the genes involved in development, although the existence of such pleiotropy is not certain (Hoekstra and Coyne, 2007), and individual mutations in some proteins can benefit multiple traits (McGee et al., 2014). The evolution of functional aspects of individual proteins must involve changes to their primary amino-acid sequences, but few traits in complex organisms are determined by single proteins. Disregarding the probability of generating beneficial variation, more mutational mechanisms exist

for altering the amounts of proteins produced than for altering their functions through their primary sequences. The latter must occur through nonsynonymous point mutations affecting specific regions of the encoded protein, whereas the former can occur through, for example, mutations in *cis*-regulatory regions (Carroll, 2008), copy-number variation (Stranger et al., 2007), post-transcriptional small RNA regulation (Flynt and Lai, 2008), and translational efficiency (Man and Pilpel, 2007). We therefore might expect that, without any innate effect-based bias, traits should preferentially evolve through mutations affecting regulatory changes.

Snake venoms are an emerging system for the study of the genetics of adaptation because of their genetic tractability, direct contributions to fitness, and high evolutionary rates. Venoms combine strong positive selective pressures through antagonistic coevolutionary interactions (Biardi et al., 2005; Biardi and Coss, 2011), defensive (Jansa and Voss, 2011) and feeding functions that contribute directly to survival and reproduction, and, being proteinaceous secretions from specialized glands, simple methods for characterizing their genomic bases. High-throughput, high-resolution venom-gland transcriptomics have been used to characterize the toxin-encoding genes for the venoms of several venomous snake species, including the Eastern Coral Snake (*Micrurus fulvius*;

<sup>\*</sup> Corresponding author. Florida State University, Department of Biological Science, 319 Stadium Dr. Tallahassee, FL 32306-4295, USA.

E-mail address: [drokyta@bio.fsu.edu](mailto:drokyta@bio.fsu.edu) (D.R. Rokyta).

Margres et al., 2013), the King Cobra (*Ophiophagus hannah*; Vonk et al., 2013), the Eastern Diamondback Rattlesnake (*Crotalus adamanteus*; Rokyta et al., 2011, 2012), and the Timber Rattlesnake (*Crotalus horridus*; Rokyta et al., 2013). As reference transcriptomes accumulate, we can shift from questions regarding the composition of venoms (Margres et al., 2014; McGivern et al., 2014) to questions about how venoms evolve within species (Margres et al., 2015) or between closely related species (Rokyta et al., 2013). Because venom genes are not known to be expressed in tissues other than venom glands (but see Casewell (2012)), we expect no significant pleiotropic constraints on mutations affecting primary protein sequences of toxins. Because venoms are secretions, changes in toxin-gene expression levels directly alter protein amounts in the venom and thereby directly influence venom efficacy, particularly given the synergism between some venom components (Casewell et al., 2011). These unique characteristics of venom allow the straightforward ascertainment of the relative roles of expression versus coding-sequence changes in adaptive evolution for a trait with little inherent bias toward either class of mutation.

Rattlesnake venoms are typically hemorrhagic, but potent neurotoxic venoms have arisen in several lineages (Mackessy, 2008), including intraspecific evolution in the Mojave Rattlesnake (*Crotalus scutulatus*), the South American Rattlesnake (*Crotalus durissus*), and the Timber Rattlesnake (*C. horridus*). In these three species, populations with the typical, probably ancestral, hemorrhagic venoms, referred to as type B or type I venoms, coexist with populations that show neurotoxic venom, referred to as type A or type II venoms (Glenn et al., 1994; Wooldridge et al., 2001; Mackessy, 2008; Calvete et al., 2010). These two venom types show an inverse relationship between toxicity and metalloproteinase activity. Snake-venom metalloproteinases (SVMPs) break down components of the capillary basement membrane, resulting in local and systemic hemorrhage, and are otherwise known to disrupt hemostasis and cause inflammation and apoptosis (Fox and Serrano, 2005). Type I or type B venoms have high metalloproteinase activity and low toxicity ( $LD_{50} > 1.0 \mu\text{g/g}$  mouse body weight), whereas type II or type A venoms have low metalloproteinase activity and high toxicity ( $LD_{50} < 1.0 \mu\text{g/g}$  mouse body weight). All type A venoms that have been investigated contain potent heterodimeric phospholipase  $A_2$  (PLA2) toxins homologous to crotoxin from *C. durissus*, Mojave toxin from *C. scutulatus*, and canebrake toxin from *C. horridus* (Hendon and Fraenkel-Conrat, 1971; Straight and Glenn, 1989; Wooldridge et al., 2001) that are the primary venom components responsible for their high lethality. Calvete et al. (2010) hypothesized that the type A venom of *C. durissus* may have arisen through paedomorphism, because its close relative *Crotalus simus* undergoes an ontogenetic shift from type A to type B venom (Durban et al., 2013). This hypothesis, however, was made on the basis of superficial similarity at the proteomic level. This similarity may represent paedomorphosis or simply convergence; finer-scale data might reveal different molecular mechanisms underlying these phenotypes.

We provided the first full transcriptomic and proteomic characterization of the evolutionary pathway to neurotoxic venom within a snake species and the first full-venom characterization of intraspecific evolution. The Timber Rattlesnake (*C. horridus*) occurs from New England and extreme southern Ontario, southward to northern Florida, and westward to eastern Texas and extreme southeastern Minnesota (McDiarmid et al., 1999). This species is of conservation concern because of habitat loss and human persecution (Stechert, 1982; Tyning, 1990; Brown, 1993). *C. horridus* has been extirpated from many areas, particularly in the northern part of its range, and is classified as endangered in six states and threatened in five others. The diet of *C. horridus* consists primarily

of rabbits, squirrels, rats, mice, and occasionally birds, frogs, and snakes (Klauber, 1997). Rokyta et al. (2013) described the transcriptomic basis for the evolution of neurotoxic venom in *C. horridus* in comparison with its congener, the Eastern Diamondback Rattlesnake (*C. adamanteus*). We focused instead on identifying the genetic pathway to the rapid evolution of neurotoxic venom within *C. horridus* by studying the venom-gland transcriptomes and venom proteomes of individual snakes of this species with each venom type. We ascertained the relative roles of expression versus fixed coding-sequence changes in intraspecific venom evolution and tested whether neurotoxic venom in adult *C. horridus* arose through retention of juvenile gene-expression patterns.

## 2. Materials and methods

### 2.1. Snakes, venoms, and venom glands

The adult venom and transcriptome used in the present work for type A *C. horridus* were described previously (Rokyta et al., 2013). This type A *C. horridus* was an adult female from Bradford County, Florida, with a snout-to-vent length of 108.0 cm and a total length of 116.0 cm (Rokyta et al., 2013). The type B adult was a male from Baker County, Georgia, with a snout-to-vent length of 107.0 cm and a total length of 117.0 cm. The juvenile type B animal was a female from Mitchell County, Georgia with a snout-to-vent length of 39.5 cm and a total length of 41.5 cm. The type A juvenile was from Columbia County, Florida, with a snout-to-vent length of 44.0 cm and a total length of 47.5 cm.

We stimulated venom-gland transcription by means of venom extraction under anesthesia (McCleary and Heard, 2010). Snakes were anesthetized with a propofol injection (10 mg/kg) and exposure to isoflurane gas, and venom expulsion was initiated by means of electrostimulation. After allowing four days for transcription to be maximized (Rotenberg et al., 1971), animals were euthanized by injection of sodium pentobarbital (100 mg/kg), and their venom glands were removed and transferred into RNAlater. The above techniques were approved by the Florida State University Institutional Animal Care and Use Committee (IACUC) under protocol #0924.

### 2.2. Transcriptome sequencing

Venom-gland tissue was diced, placed in TRIzol (Invitrogen 15596-018), homogenized by mortar, and aspirated through a 20 gauge needle. The RNA was isolated from the lysate using a chloroform extraction in conjunction with Heavy Phase Lock Gel tubes (5 PRIME 2302810) and further purified by ethanol precipitation.

Sequencing and nonnormalized cDNA library preparation for our two adults were performed by the HudsonAlpha Institute for Biotechnology Genomic Services Laboratory (<http://www.hudsonalpha.org/gsl/>). Transcriptome sequencing was performed essentially as described by Mortazavi et al. (2008) in a modification of the standard Illumina methods described in detail in Bentley et al. (2008). Total RNA was reduced to poly-A+ RNA with oligo-dT beads. Two rounds of poly-A+ selection were performed. The purified mRNA was then subjected to a mild heat fragmentation followed by random priming for first-strand synthesis. Standard second-strand synthesis was followed by standard library preparation with the double-stranded cDNA as input material. This approach is similar to that of Illumina's TruSeq RNA-seq library preparation kit. Sequencing was performed on an Illumina HiSeq 2000 with 100-base-pair paired-end reads.

Library preparation for the MiSeq was performed on the selected mRNA using the Illumina TruSeq RNA Sample Preparation

v2 kit and multiplex oligos for Illumina sequencing. Quality of the isolated RNA was assessed by Experion StdSens RNA Analysis Kit (Bio Rad). The mRNA was isolated using NEBNext Poly(A) mRNA Magnetic Isolation Module using 500 ng of total RNA. Incubation and PCR steps were carried out in a Veriti Thermocycler (Applied Biosystems). During and after the protocol, DNA was purified using Agencourt AMPure XP PCR Purification Beads. Final PCR amplification of library consisted of 15 cycles. Libraries were quantified and assessed for quality with an Agilent 2100 Bioanalyzer. Samples were pooled and sequenced with MiSeq Version 2 Reagent Kits on the Illumina MiSeq platform. Samples were prepared according to manufacturer's protocol (revision C). The pooled samples were sequenced with two kits with 151-nucleotide paired-end reads.

The original, unmerged sequencing reads for all runs were submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive. The previously published type A adult *C. horridus* HiSeq data (Rokyta et al., 2013) is under the accession SRR575168, and the type B adult HiSeq data is under accession SRR1554232. The two MiSeq runs were submitted separately with accession numbers as follows: type A adult—SRR1554276 and SRR1554277, type B adult—SRR1554280 and SRR1554283, and type B juvenile—SRR1554285 and SRR1554286.

### 2.3. Transcriptome assembly and analysis

The raw 100-nt paired-end reads passing the Illumina quality filter were merged if their 3' ends overlapped as described previously (Rokyta et al., 2012, 2013; Margres et al., 2013). This step also removed adapter sequences present because of fragment read-through. To eliminate reads corresponding to extremely high-abundance transcripts, we used the Extender program (Rokyta et al., 2012) with 1000 merged reads as seeds to attempt to generate complete transcripts using only the merged reads. Extension of seeds required an overlap of 100 nucleotides, phred scores of at least 30 at each position in the extending read, and an exact match in the overlapping region. We performed a reference-based assembly against the 3031 nontoxins previously annotated for type A *C. horridus* (Rokyta et al., 2013) with NGen version 11 using both the merged and unmerged reads and a minimum match percentage of 95. Consensus sequences were retained if they had at least 10× coverage over the entire coding sequence. Regions outside the coding sequence with less than 10× coverage were removed. Toxin sequences were clustered into groups with <1% nucleotide divergence in their coding sequences, and duplicate nontoxin sequences were eliminated following alignment of the final transcripts with NGen. We used one representative from each toxin cluster and all of the unique nontoxins to filter the corresponding reads in a reference-based transcriptome assembly in NGen with a minimum match percentage of 98, using only the merged reads. The unfiltered reads were then used in a *de novo* transcriptome assembly in NGen with the default minimum match percentage of 93, retaining only contigs comprised of at least 100 reads.

To increase our chances of identifying all toxin sequences, we performed four additional *de novo* assemblies for each species. We ignored sequences without homology with known toxins for all four assemblies. Three assemblies were performed with NGen with a minimum match percentage of 98, using 1, 5, and 10 million reads. We opted for the high stringency for these assemblies to attempt to differentiate closely related paralogs and varied the number of reads because we found that some extremely high-abundance transcripts were difficult to assemble with too many reads, apparently because of low levels of unspliced transcripts. The fourth additional *de novo* assembly used the Extender program as above

on 1000 new random reads, and we only used this assembly to identify SVMPS, which were difficult for other methods to assemble.

After combining the results of all of the above assemblies and eliminating duplicates as described above, we performed one final round of read filtering of the merged reads, followed by a *de novo* assembly of the unfiltered reads as above, keeping only those contigs comprising ≥1000 reads. We ignored sequences without homology with known toxins families and added any resulting unique toxins to our database. This step was included to ensure that we missed no toxin sequences with appreciable expression.

To screen for and eliminate potentially chimeric sequences in our toxin databases for both species, we first screened for evidence of recombination within each toxin family with GARD (Kosakovsky Pond et al., 2006). We used the general reversible model of sequence evolution and gamma-beta rates. If we found a signal for recombination resulting in significantly different tree topologies for different regions of the alignments based on KH tests, we performed a reference-based assembly with NGen version 11 with a minimum match percentage of 98 and the autotrim parameter set to false, using the toxin coding sequences as references and 10 million merged reads. Such high-stringency alignments facilitate the identification of chimeric sequences by producing either multimodal or extremely uneven coverage distributions, particularly in combination with our long, merged reads. Suspect sequences were confirmed to be chimeras of other sequences in our toxin database before removal.

Sequences were identified by means of blastx searches against the NCBI non-redundant (nr) protein database with a minimum E-value of  $10^{-4}$  and retaining only 10 hits. *De novo* assembled transcripts were only retained and annotated if they had complete protein-coding sequences. Putative toxins were identified by searching their blastx match descriptions for toxin-related key words as described previously (Rokyta et al., 2012, 2013; Margres et al., 2013). The final set of unique transcripts for each individual was generated by combining the results from all assemblies and eliminating duplicates by means of an NGen assembly and a second, more stringent assembly in SeqMan Pro. Final transcript abundances were estimated by means of a reference-based transcriptome assembly with NGen with a minimum match percentage of 95, using only the coding sequences of transcripts. Signal peptides for toxins were identified by means of SignalP analyses (Bendtsen et al., 2004). Putative toxins were named with a toxin-class abbreviation, a number indicating cluster identity, and a lower-case letter indicating the particular member of a cluster.

RNA-seq data represents a type of compositional data (Aitchison, 1986), and comparisons of expression levels are complicated by values confined to a simplex. We measured expression levels by aligning 10 million merged reads against the coding sequences identified through *de novo* assembly and analysis using the templated transcriptome settings in NGen11 with a minimum match percentage of 95. To compare expression levels among classes or individual transcripts, we normalized the values by the total number of reads among all transcripts considered (i.e., converted to proportions or a composition), then performed a centered logratio transformation. If the normalized data are  $\mathbf{x} = (x_1, \dots, x_n)$  such that  $\sum_{i=1}^n x_i = 1$ , then

$$\text{clr}(\mathbf{x}) = \left( \ln \frac{x_1}{g(\mathbf{x})}, \dots, \ln \frac{x_n}{g(\mathbf{x})} \right) \quad (1)$$

where  $g(\mathbf{x}) = \sqrt[n]{x_1 \cdots x_n}$  is the geometric mean. This transformation takes the data from the simplex to real space.

To identify variable sites, we aligned 10 million merged

transcriptome reads of each genotype against the final toxin transcripts of each genotype and identified SNPs at frequencies of at least 25% with  $\geq 10\times$  coverage. For the alignment and SNP calling, we used NGen version 11 with a minimum match percentage of 95.

The assembled transcripts for both the adult transcriptomes were deposited at DDBJ/EMBL/GenBank in the Transcriptome Shotgun Assembly database under the following accession numbers: type A adult—GBKC000000000 and type B adult—GBKD000000000.

#### 2.4. Hybridization tests

All alignments were constructed with ClustalW (Thompson et al., 1994). Evolutionary models were selected using MrModelTest version 2 with Akaike Information Criterion values. To estimate Bayes factors (Kass and Raftery, 1995) under different topological constraints, we ran four independent analyses under each constraint in MrBayes version 3.1.2 (Huelsenbeck and Ronquist, 2001) for 10 million generations each, sampling every 1000 generations. Markov chain Monte Carlo searches were run for 10 million generations with four chains, the temperature parameter set to 0.2, and samples taken every 1000 generations. We discarded samples from the first one million generations for each analysis as burn-in and used the harmonic mean estimator for marginal likelihoods as implemented in MrBayes. We used the samples from all four runs for each estimate.

#### 2.5. Proteomics

Reversed-phase high-performance liquid chromatography was performed on a Beckman System Gold HPLC (Beckman Coulter, Fullerton, CA). Dried venom samples were re-suspended in water and centrifuged to remove insoluble material. Approximately 100  $\mu$ g of total protein were injected onto a Jupiter C18 column, 250  $\times$  4.6 mm (Phenomenex, Torrance, CA) using the standard solvent system of A = 0.1% trifluoroacetic acid (TFA) in water and B = 0.075% TFA in acetonitrile. After five minutes at 5% B, a 1% per minute linear gradient of A and B was run to 25% B, followed by a 0.25% per minute gradient from 25 to 65% B at a flow rate of 1 mL per minute. Column effluent was monitored at 220 and 280 nm and peak fractions were collected manually.

Reversed-phase high-performance liquid chromatography fraction digestion was performed using the Calbiochem ProteoExtract All-in-One Trypsin Digestion Kit (Merck, Darmstadt, Germany) according to the manufacturer's instructions, using LC/MS grade solvents. The digestion supernatants were stored at  $-80^\circ\text{C}$  prior to analysis. Fraction digests were chromatographically separated prior to online analysis by tandem mass spectrometry. Using a nanoLC 1D system (Eksigent, Dublin, CA), samples were passed through a vented column system beginning with a 300  $\mu$ m ID  $\times$  5 mm C18 trap column (Agilent, Santa Clara, CA) for online desalting and sample clean-up. Samples were then loaded onto a 10 cm bed of C18 reversed phase chromatography resin packed into a 360  $\mu$ m OD  $\times$  75  $\mu$ m ID fused silica emitter tip (PicoFrit column, New Objective, Woburn, MA). Mobile phase A was 0.1% formic acid in water and mobile phase B was 0.1% formic acid in acetonitrile. Approximately 500 ng of total protein were loaded onto the column. To maintain optimal column pressure, a flow rate of 600 nL per minute was used to run a linear gradient of 0–40% B for 45 min.

An externally calibrated Thermo LTQ Orbitrap Velos nLC-ESI-LIT-Orbitrap was used to perform high-resolution tandem mass spectrometry on the peaks isolated from the venoms of each proteome animal. The nLC-MS/MS was run in triplicate. A 2 cm, 100  $\mu$ m i.d. trap column (SC001 Easy Column, Thermo Scientific, Waltham, MA) was followed by a 10 cm analytical column of 75  $\mu$ m i.d. (SC200

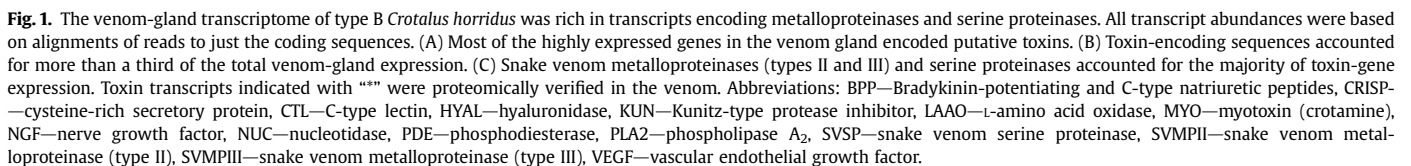
Easy Column, Thermo Scientific, Waltham, MA). Both trap column and analytical column contained C18-AQ packaging. Separation was carried out using Easy nanoLC II (Thermo Scientific, Waltham, MA) with a continuous, vented column configuration. A 2  $\mu$ L (200 ng) sample was aspirated into a 20  $\mu$ L loop and loaded onto the trap. The flow rate was set to 300 nL/min for separation on the analytical column. Mobile phase A was composed of 99.9% water (EMD Omni Solvent) and 0.1% formic acid. Mobile phase B was composed of 99.9% acetonitrile and 0.1% formic acid. A one-hour linear gradient from 0 to 45% B was performed. The LC effluent was directly nanosprayed into a LTQ Orbitrap Velos mass spectrometer (Thermo Scientific, Waltham, MA). During chromatographic separation, the LTQ Orbitrap Velos was operated in a data-dependent mode and under direct control of the Xcalibur software (Thermo Scientific, Waltham, MA). The MS data were acquired using 10 data-dependent collisional-induced-dissociation (CID) MS/MS scans per full scan. All measurements were performed at room temperature and each sample was run in triplicate.

Tandem mass spectra were extracted by Proteome Discoverer version 1.4.0.288. Charge state deconvolution and deisotoping were not performed. All MS/MS samples were analyzed using Sequest (Thermo Fisher Scientific, San Jose, CA, USA; version 1.4.0.288) and X! Tandem (The GPM, thegpm.org; version CYCLONE (2010.12.01.1)). Sequest and X! Tandem were set up to search the transcriptome databases (4432 entries for *C. horridus* type A and 2785 entries for type B) assuming trypsin was the digestion enzyme used. Sequest and X! Tandem were searched with a fragment ion mass tolerance of 0.50 Da and a parent ion tolerance of 10.0 PPM. Carbamidomethyl of cysteine was specified in Sequest and X! Tandem as a fixed modification. Conversion of glutamine to pyroglutamate of the N-terminus and oxidation of methionine were specified in Sequest and X! Tandem as variable modifications.

Scaffold (version Scaffold\_4.3.2, Proteome Software Inc., Portland, OR) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability. Peptide Probabilities from X! Tandem were assigned by the Peptide Prophet algorithm (Keller et al., 2002) with Scaffold delta-mass correction. Peptide Probabilities from Sequest were assigned by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least one unique peptide and two or more identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped together.

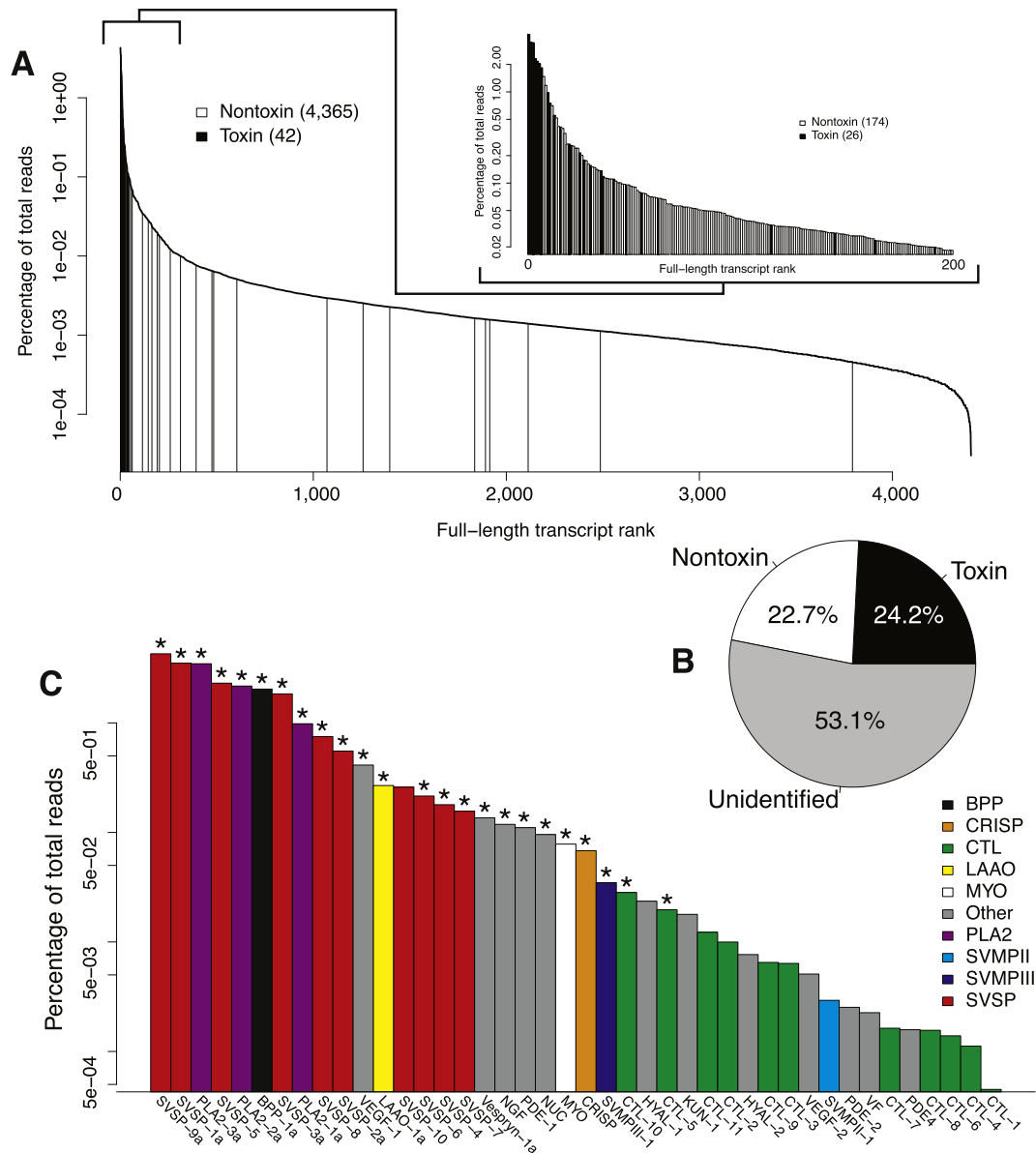
#### 2.6. Canebrake toxin PCR assay

Genomic DNA was extracted from blood samples drawn from caudal veins using the Omega bio-tek E.Z.N.A Tissue DNA Kit according to the manufacturer's protocol. The acidic and basic subunits of canebrake toxin were amplified from template DNA in 25  $\mu$ L PCR reactions using the MTA-sense (5' GGT ATT TCG TAC TAC AGC TCT TAC GGA 3'), MTA-antisense (5' AAT TGC CAG GGG GAA TCA 3'), MTB-sense (5' AAC GCT ATT CCC TTC TAT GCC TTT TAC 3'), and MTB-antisense (5' GGA ACA GAT TTG TGA GTG CGA CAG G 3') primers, respectively, under the following thermal cycling protocol: 95  $^\circ\text{C}$  for five minutes, 35 cycles of 95  $^\circ\text{C}$  for 30 s, 55  $^\circ\text{C}$  for 30 s, and 72  $^\circ\text{C}$  for five minutes, followed by 72  $^\circ\text{C}$  for ten minutes (Wooldridge et al., 2001). Evidence for amplification was visualized by means of a 0.7% agarose gel and ethidium bromide and imaged on a Bio Rad Gel Doc using Quantity One Version 4.1.1.



### 3.1. Comparative transcriptomics of *C. horridus* types A and B

selected on the basis of these geographic patterns, then confirmed to have venom profiles matching the two venom types described by [Glenn et al. \(1994\)](#). We generated 64,555,690 pairs of 100-nucleotide Illumina reads passing filter with an average phred quality score of 36 and merged 41,053,061 pairs on the basis of their 3' overlaps. The merged reads had an average length of 131 nucleotides and an average phred quality of 49. Assembly of the type B transcriptome gave 2691 full-length nontoxin coding sequences and 94 full-length toxin sequences, which we clustered into 49 groups with <1% nucleotide divergence ([Fig. 1](#)). The coding sequences of the identified transcripts accounted for 53.9% of the reads, and the full transcripts (with UTRs) accounted for 66.5% of the reads. Toxin coding sequences accounted for 38.1% of the reads



**Fig. 2.** The venom-gland transcriptome of type A *Crotalus horridus* was rich in transcripts encoding serine proteinases and phospholipases A<sub>2</sub>. The transcriptome described by Rokyta et al. (2013) was re-analyzed through a more stringent procedure. All transcript abundances were based on alignments of reads to just the coding sequences. (A) Most of the highly expressed genes in the venom gland encoded putative toxins. (B) Toxin-encoding sequences accounted for nearly a quarter of the total venom-gland expression. (C) Serine proteinases and phospholipases A<sub>2</sub> accounted for the majority of toxin-gene expression. Toxin transcripts indicated with "\*" were proteomically verified in the venom. Abbreviations: BPP—Bradykinin-potentiating and C-type natriuretic peptides, CRISP—cysteine-rich secretory protein, CTL—C-type lectin, HYAL—hyaluronidase, KUN—Kunitz-type protease inhibitor, LAAO—L-amino acid oxidase, MYO—myotoxin (crotonamine), NGF—nerve growth factor, NUC—nucleotidase, PDE—phosphodiesterase, PLA2—phospholipase A<sub>2</sub>, SVSP—snake venom serine proteinase, SVMPII—snake venom metalloproteinase (type II), SVMPIII—snake venom metalloproteinase (type III), VEGF—vascular endothelial growth factor, VF—venom factor.

(Fig. 1B). A re-assembly of the type A transcriptome (Rokyta et al., 2013) using the more stringent protocol described in the methods (and used for type B) yielded 4365 nontoxins and 67 toxins in 42 clusters (Fig. 2). The original data consisted of 113,344,311 pairs of 100-nucleotide Illumina reads passing filter with an average phred quality score of 30. We merged 64,169,665 pairs to give an average length of 133 nucleotides and an average quality of 47. The coding sequences of all of the identified transcripts accounted for 46.9% of the reads, and the full transcripts (with UTRs) accounted for 73.3% of the reads. Toxin coding sequences accounted for 24.2% of the reads (Fig. 2B).

A reciprocal-blastp analysis of the 49 type B and 42 type A toxin clusters with a 5% cut-off on the nucleotide divergence of matches

yielded 35 matched pairs of putatively orthologous toxins (Table 1). The most striking discrepancy between the two transcriptomes was the absence of matches between the three clusters of PLA2s for type A and two clusters of PLA2s for type B. This entire toxin gene family appears to have been replaced in one of the two lineages. This difference includes the two subunits of canebrake toxin; the type A PLA2-3 transcript encodes the acidic subunit of canebrake toxin, and the type A PLA2-2 transcript encodes the basic subunit. The remaining type A clusters not accounted for in the type B transcriptome included venom factor (VF), a phosphodiesterase (PDE-2), and two C-type lectins (CTL-5 and CTL-10). These transcripts were all expressed at low levels in the type A transcriptome (Fig. 2). In addition to the two PLA2s, the type B toxin clusters not

**Table 1**

Toxin orthologs identified for *Crotalus horridus* types A and B by means of reciprocal blastp.

A transcript name	B transcript name	% nt difference	% AA difference
BPP-1a	BPP-1a	0.0	0.0
CRISP	CRISP	0.3	0.8
CTL-1	CTL-1	0.0	0.0
CTL-2	CTL-9a	0.0	0.0
CTL-3	CTL-5	0.4	0.0
CTL-4	CTL-3	1.4	3.5
CTL-6	CTL-4	0.2	0.0
CTL-7	CTL-6	0.2	0.6
CTL-8	CTL-10a	0.2	0.0
CTL-9	CTL-11a	0.4	0.0
CTL-11	CTL-7	0.4	0.0
HYAL-1	HYAL-1a	0.1	0.2
KUN-1	KUN	0.0	0.0
LAAO-1a	LAAO-1a	0.1	0.2
MYO	MYO	0.5	1.5
NGF	NGF-1a	0.1	0.0
NUC	NUC	0.1	0.0
PDE-1	PDE	0.3	0.0
PDE4	PDE4	0.2	0.7
SVMPII-1	SVMPII-2	1.3	1.7
SVMPIII-1	SVMPIII-5a	0.9	1.5
SVSP-1a	SVSP-2a	0.6	1.6
SVSP-2a	SVSP-11a	0.3	0.8
SVSP-3a	SVSP-12a	0.1	0.4
SVSP-4	SVSP-9	0.0	0.0
SVSP-5	SVSP-6	0.0	0.0
SVSP-6	SVSP-7	0.4	1.2
SVSP-7	SVSP-1	0.1	0.4
SVSP-8	SVSP-5	0.3	0.4
SVSP-9a	SVSP-13a	0.3	0.8
SVSP-10	SVSP-10	3.1	6.8
TrunchYAL	TrunchYAL	0.2	0.5
VEGF-1	VEGF-1	0.0	0.0
VEGF-2	VEGF-2	0.0	0.0
VESP-1a	VESP	0.2	0.0

Abbreviations: AA—amino acid, BPP—Bradykinin-potentiating and C-type natriuretic peptides, CRISP—cysteine-rich secretory protein, CTL—C-type lectin, HYAL—hyaluronidase, KUN—Kunitz-type protease inhibitor, LAAO—L-amino acid oxidase, MYO—myotoxin-A (crotoxin), NGF—nerve growth factor, nt—nucleotide, NUC—nucleotidase, PDE—phosphodiesterase, SVMPII—type II snake-venom metalloproteinase, SVMPIII—type III snake-venom metalloproteinase, SVSP—snake venom serine proteinase, TrunchYAL—truncated hyaluronidase, VEGF—vascular endothelial growth factor, VESP—vespryn.

accounted for in the type A transcriptome included CTL-2, CTL-8, three serine proteinases (SVSP-3, SVSP-4, and SVSP-8), three type II metalloproteinases (SVMPII-1, SVMPII-3, and SVMPII-4), and four type III metalloproteinases (SVMPIII-1, SVMPIII-2, SVMPIII-3, and SVMPIII-4). Some of these toxins, particularly the metalloproteinases, were expressed at high levels in the type B transcriptome (Fig. 1). The type B venom is most likely the ancestral form given its predominance among the rattlesnakes (Mackessy, 2008) and that most species with type A venom remain polymorphic for type B. Assuming that the type B venom phenotype was ancestral, the primary evolutionary changes related to the presence or absence of toxins in the origin of type A venom in *C. horridus* were the complete replacement of the PLA2 gene family and the loss (from the transcriptome) of most of the SVMPs.

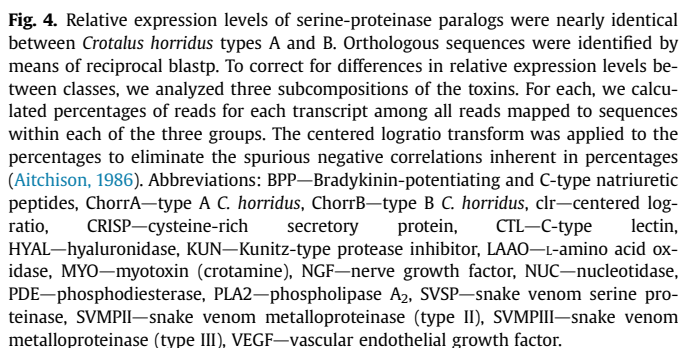
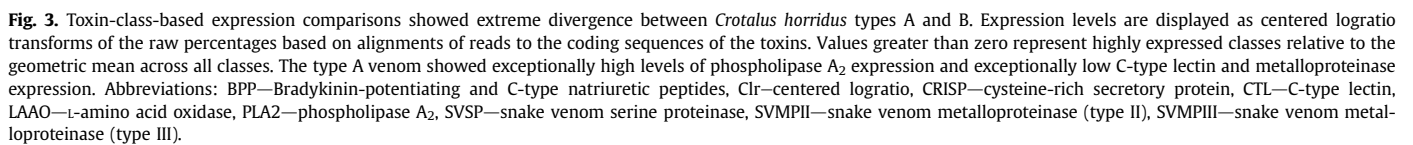
To determine whether our failure to find orthologs for some transcripts truly indicated an absence in the corresponding transcriptome, we aligned the type B reads against the type A transcript coding sequences and the type A reads against the type B transcript coding sequences. We aligned 10 million merged reads with a minimum match percentage of 95. Because of the presence of gene families of toxins, we expect some reads to align to some regions of transcripts in these families even if the particular paralog in question was absent from the transcriptome. We therefore assessed

presence by requiring complete coverage of the coding sequence. We found no evidence in the type A reads for any of the 14 type B transcripts lacking type A orthologs. For the seven type A transcripts without orthologs in type B, we found evidence for the presence of orthologs of ChorrA PDE-2 and ChorrA VF, but not for any of the other five type A transcripts. The PDE-2 and VF genes were expressed at low levels in both transcriptomes.

The transcriptome for type B *C. horridus* showed an overall higher percentage of reads mapping to toxin coding sequences, with 38.1% compared to 24.2% for type A (Figs. 1 and 2). Class-level expression patterns differed dramatically between the type A and B *C. horridus* venom-gland transcriptomes (Fig. 3). The major differences included much higher relative expression of SVSPs and PLA2s and much lower expression of CTLs and SVMPs in the type A transcriptome (Fig. 3). The larger total percentage of reads mapping to toxins for type B and the differences in relative expression levels of the toxin classes could mostly be accounted for simply by the loss of reads mapping to CTLs and SVMPs for type A. Removal of the CTL and SVP reads from the type B left 18.1% of the total reads; removing these reads and switching the read counts for the PLA2s between type A and type B (to account for the replacement of this gene family) gave 23.0% of the total reads, which is remarkably similar to the 24.2% observed for type A. The combination of relatively simpler venom and higher toxicity of the type A venom may allow for a lower cost for venom production for these animals and therefore may explain, in part, the evolution of this phenotype.

Serine proteinases were expressed in high levels in both transcriptomes (Fig. 3), and we identified 10 pairs of SVSP orthologs through reciprocal blastp (Table 1). The relative expression levels of these orthologs were nearly identical across transcriptomes (Fig. 4) despite different relative class-level expression (Fig. 3). The centered logratio values from the B transcripts nearly perfectly predicted those from A ( $F_{1,8} = 280.1$ ,  $P = 1.6 \times 10^{-7}$ ,  $R^2 = 0.97$ ; Fig. 4). Serine proteinase expression levels appeared to be unchanged; the increase in relative expression level for SVSPs in the type A transcriptome appeared to merely reflect the loss of expression in other classes. Serine proteinases therefore appear to be effective in both hemorrhagic and neurotoxic venoms. The relative expression levels of the nine reciprocal-blastp CTL pairs were highly discordant between the A and B transcriptomes ( $F_{1,7} = 1.0$ ,  $P = 0.35$ ,  $R^2 = 0.13$ ; Fig. 4). The global reduction in class-level CTL expression in the type A venom gland (Fig. 3) was accompanied by idiosyncratic changes in relative expression among the paralogs, suggesting that expression of these genes was reduced on a gene-by-gene basis. The remainder of the reciprocal-blastp pairs were in fairly close agreement ( $F_{1,14} = 7.76$ ,  $P = 0.015$ ,  $R^2 = 0.36$ ; Fig. 4). For each of these three analyses (Fig. 4), expression values were normalized by the total number of reads mapped to that particular group (SVSPs, CTLs, or everything else).

In addition to the gene-expression changes described above, toxin sequence differences may have contributed to the evolution of the two venom types. To look for fixed differences in toxins between our type A and type B *C. horridus*, we ran four SNP analyses. We aligned the transcriptome reads of each type against the toxin transcripts of each type and identified SNPs at frequencies of at least 25% with  $\geq 10\times$  coverage. We counted SNPs as fixed differences if they were present as fixed differences in both cross-type alignments for orthologs identified by means of the reciprocal-blastp analysis (Table 1) and were not also detected as SNPs in the type-specific alignments. Because we were interested in differences that could affect venom function, we only considered nonsynonymous substitutions. We identified eight fixed, nonsynonymous differences between our type A and type B *C. horridus* affecting five of the 35 loci analyzed (Table 2). For comparison, we identified 14 heterozygous sites resulting in amino-acid changes in



Fixed toxin amino-acid differences between *Crotalus horridus* types A and B.

A transcript name	B transcript name	AA position	Type A AA	Type B AA
CRISP	CRISP	36	K	E
CRISP	CRISP	43	F	S
CTL-7	CTL-6	97	V	I
HYAL-1	HYAL-1a	381	S	G
MYO	MYO	8	F	S
SVSP-6	SVSP-7	67	Y	H
SVSP-6	SVSP-7	79	T	M
SVSP-6	SVSP-7	204	G	S

Abbreviations: AA—amino acid, CRISP—cysteine-rich secretory protein, CTL—C-type lectin, HYAL—hyaluronidase, MYO—myotoxin-A (crotonamine), SVSP—snake venom serine proteinase.

Table 3

Heterozygous toxin amino-acid sites for type A *Crotalus horridus*.

Transcript	SNP %	AA position	AA 1	AA 2
LAAO-1a	46.00	147	R	Q
SVMP11-1	38.90	9	I	L
SVMP11-1	37.50	41	A	P
SVMP11-1	28.20	44	R	K
SVSP-10	87.60	51	G	S
SVSP-10	85.60	53	T	I
SVSP-1a	29.70	26	V	I
SVSP-1a	30.00	132	P	T
SVSP-3a	54.70	246	M	I
SVSP-4	35.10	253	A	V
SVSP-7	40.30	9	T	N
SVSP-8	52.40	230	E	G
SVSP-9a	60.40	164	K	N
SVSP-9a	58.70	179	H	Q

Abbreviations: AA—amino acid, LAAO—L-amino acid oxidase, SVMPII—snake venom metalloproteinase (type II), SVSP—snake venom serine proteinase.

differences. Given that we only compared a single individual of each venom type, these differences should represent an upper bound on the number of fixed nonsynonymous differences between these populations.

**Table 4**  
Heterozygous toxin amino-acid sites for type B *Crotalus horridus*.

Transcript	SNP %	AA position	AA 1	AA 2
CTL-10a	42.60	98	G	S
CTL-5	80.70	22	E	A
CTL-6	28.90	4	L	F
LAAO-1a	47.90	21	N	D
LAAO-1a	41.00	147	R	Q
LAAO-1a	34.30	420	E	Q
LAAO-1a	37.20	431	H	R
LAAO-1a	33.90	434	K	M
LAAO-1a	45.60	477	T	I
PDE4	55.60	213	Y	H
PDE4	38.50	316	V	I
PDE4	46.20	371	P	S
SVMPII-1	78.10	79	E	K
SVMPII-1	53.70	108	H	R
SVMPII-1	42.00	110	Q	E
SVMPII-1	85.50	142	K	E
SVMPII-1	40.00	424	R	G
SVMPII-1	40.00	424	R	T
SVMPIII-1	39.30	21	I	V
SVMPIII-1	37.00	69	L	H
SVMPIII-1	36.40	98	P	L
SVMPIII-1	25.40	477	Q	R
SVMPIII-1	26.30	495	K	N
SVMPIII-4a	54.20	231	T	A
SVMPIII-5a	25.60	35	H	Y
SVMPIII-5a	35.20	69	L	H
SVMPIII-5a	37.70	98	P	L
SVMPIII-5a	44.20	440	G	E
SVSP-10	32.00	141	D	V
SVSP-10	32.00	152	A	P
SVSP-10	41.70	153	T	S
SVSP-10	53.20	159	N	D
SVSP-10	25.60	168	I	L
SVSP-10	75.60	170	N	D
SVSP-11a	78.10	46	R	S
SVSP-11a	78.80	51	A	G
SVSP-11a	84.90	85	K	N
SVSP-11a	74.90	89	K	T
SVSP-13a	36.10	15	I	L
SVSP-13a	38.20	47	N	H
SVSP-13a	37.70	153	E	K
SVSP-1	25.10	9	N	T
SVSP-2a	25.70	176	A	T
SVSP-4	56.00	37	H	R
SVSP-4	54.30	182	Q	K
SVSP-4	54.70	212	V	L
SVSP-4	37.70	232	P	L
SVSP-5	25.00	230	E	G

Abbreviations: AA—amino acid, CTL—C-type lectin, LAAO—L-amino acid oxidase, PDE—phosphodiesterase, SVMPII—snake venom metalloproteinase (type II), SVMPIII—snake venom metalloproteinase (type III), SVSP—snake venom serine proteinase.

### 3.2. Comparative proteomics of *C. horridus* types A and B

We conducted proteomic analyses of venom samples from the same type A and type B individuals used in venom-gland transcriptome sequencing. We followed the approach of Margres et al. (2014) of separating the whole venom into a moderate number of fractions by means of reversed-phase high performance liquid chromatography (RP-HPLC) and then performing LC-MS/MS identification of the proteins present in each fraction by searching against the corresponding transcriptomic database. We separated the type A venom into nine fractions (Fig. 5 and Table 5) and the type B venom into 18 fractions (Fig. 5 and Table 6). For type A, we found proteomic evidence for the proteins encoded by 24 of the 42 putatively toxin-encoding transcript clusters identified in the venom-gland transcriptome, including 24 of the 26 most highly expressed transcripts (Fig. 2 and Tables 5 and 7). For type B, we

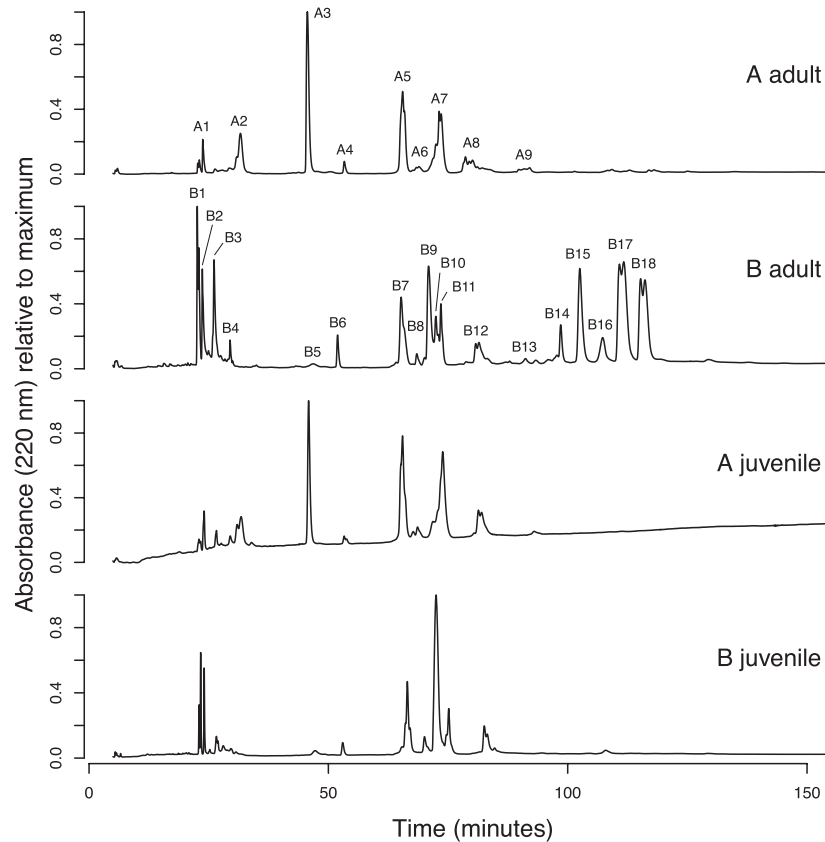
found proteomic evidence for 39 of the 49 putatively toxin-encoding transcripts, including 37 of the 39 most highly expressed transcripts (Fig. 1 and Tables 6 and 8). The major differences in peaks matched the results from the transcriptomes. The type A venom lacked peaks corresponding to the type B peaks 14–18 (Fig. 5), and the major proteins in these peaks were the type II and type III SVMPIs (Table 6). The A2 and A3 peaks containing the two PLA2 subunits of canebrake toxin (PLA2-2 and PLA2-3; Table 5) had no corresponding peaks in the type B venom (Fig. 5). The peaks that did match across venoms generally showed evidence for the presence of orthologous toxins as expected. For example, peaks A4 and B6 correspond and both had evidence for CRISP and NGF. Peaks A5 and B7 each showed evidence for two distinct clusters of SVSP, and these represented orthologous pairs (Tables 1, 5 and 6). Similar correspondence was present in the other matching peaks.

Our proteomic analysis showed a tendency to fail to detect proteins in the venom corresponding to the lower-abundance transcripts for both individuals (Figs. 1 and 2). Nearly all of the transcripts without proteomic support were in the lower end of the coverage distribution. This may simply reflect a detection threshold and the extremely large range of expression levels for toxin-encoding genes; the signal for low-expression proteins may be overwhelmed by that for high-expression proteins. These putative toxin transcripts, however, were identified through homology with known toxins and may not encode true toxic proteins that are secreted into the venom. C-type lectins appeared to be particularly prone to being missed in the proteome when expressed at low levels in the transcriptome. The phosphodiesterases (PDEs), venom factor (VF), Kunitz-type protease inhibitors (KUNs), hyaluronidases (HYALs), and vascular endothelial growth factors (VEGFs) were also among those transcripts without supporting proteomic data, suggesting that some of these may either not be true toxins or at least not significant components of the venoms.

### 3.3. Ontogeny and the origin of neurotoxic venom

Calvete et al. (2010) hypothesized that the evolution of neurotoxic venom in *C. durissus* represented a case of paedomorphism, because the venom proteome of neurotoxic *C. durissus* was similar to that of juvenile *C. simus*, which changes from neurotoxic to hemorrhagic during maturation. This mechanism is an attractive hypothesis because the co-option of a pre-programmed regulatory change could allow the evolution of major, large-scale expression changes with minimal genetic change. Juvenile type B *C. horridus* are known to have different venoms than adults (Glenn et al., 1994). Proteomic analysis of the venoms of juveniles from the same populations as our transcriptome animals showed no apparent ontogenetic change for the type A population, but did show an obvious change in the type B population (Fig. 5). The type B juvenile, much like the type A adult, lacked many of the later-eluting RP-HPLC peaks that correspond to the SVMPIs. The juvenile type B, however, also lacked the two peaks corresponding to the two subunits of canebrake toxin (peaks A2 and A3), clearly indicating that paedomorphosis is not the entire story behind the evolution of type A venom in *C. horridus*.

To determine whether gene expression levels of the adult type A animal matched those of the juvenile type B for genes other than the PLA2s, consistent with a partial paedomorphic origin of type A venom, we sequenced the juvenile venom-gland transcriptome and re-sequenced the two adult transcriptomes on an Illumina MiSeq to standardize the measurements across genotypes. For the adult type A, we generated 2,810,495 pairs of 151-nt reads and merged 2,088,342 on the basis of their 3' overlaps. For the adult type B, we generated 3,949,870 pairs of reads and merged 2,309,434, and we generated 3,839,186 pairs and merged 2,345,251 for the juvenile



**Fig. 5.** Proteomic analysis revealed loss and gain of venom components in type A *Crotalus horridus* relative to type B. The reversed-phase high-performance liquid chromatography profiles of the venoms showed an absence of peaks after approximately 100 min in the neurotoxic type A venoms. The juvenile type B venom similarly lacked these peaks but also lacked the peaks corresponding to the two canebrake toxin subunits (peaks A2 and A3), indicating that venom paedomorphosis cannot entirely account for the evolution of type A venom in *C. horridus*. The juvenile type A venom showed no apparent compositional difference from the adult type A. Labeled peaks were analyzed by mass spectrometry for identification (Tables 5 and 6). Venom samples were taken from the animals used for transcriptome sequencing. Chromatograms are scaled relative to the maximum detected absorbance for each sample.

type B. Despite using different RNA-seq prep kits and different sequencing platforms, the estimated expression levels of the toxins for the type A and type B animals were in close agreement (type A:  $F_{1,40} = 1577.0$ ,  $P < 10^{-16}$ ,  $R^2 = 0.98$ ; type B:  $F_{1,47} = 553.6$ ,  $P < 10^{-16}$ ,  $R^2 = 0.92$ ). To compare the three transcriptomes, we constructed a superset of toxin sequences by combining the 49 type B clusters with the seven type A clusters that did not have reciprocal blastp matches in the type B transcriptome. This set of toxins therefore included representatives of every toxin sequence identified across both animals. For each individual, we aligned all of the merged reads against the toxin superset. At the level of toxin classes, the type A adult was more similar to the type B juvenile than to the type B adult (Fig. 6). Relative to the type B adult, the type A adult and type B juvenile expressed much lower levels of SVMs, but expression of SVMs was higher in the type B juvenile than the type A adult, particularly for the type III class of SVMs (Fig. 6). C-type lectins were severely down-regulated in the type A adult, but appeared to be expressed at comparable levels in type B adult and juvenile. At the level of individual transcripts, expression levels were more similar between the type B adult and juvenile ( $F_{1,54} = 29.9$ ,  $P < 10^{-5}$ ,  $R^2 = 0.36$ ) than between the type A adult and type B juvenile ( $F_{1,54} = 16.3$ ,  $P < 10^{-3}$ ,  $R^2 = 0.23$ ). The absence of the canebrake toxin subunits, the high level of CTL expression, and the relatively higher level of SVM expression in the type B juvenile compared to the type A adult suggest that much of the evolutionary pathway to type A venom cannot be attributed to retention of the

juvenile venom expression pattern. The evolution of type A venom in *C. horridus* was therefore not accomplished only through a simple genetic change co-opting a pre-existing ontogenetically variable expression pattern.

### 3.4. The evolution of type A venom

Rokyta et al. (2013) examined the patterns of evolutionary divergence between the hemorrhagic venom of *C. adamanteus* and the type A, neurotoxic venom of *C. horridus*. They found complex genetic and expression patterns differentiating the venom-gland transcriptomes. Toxin classes and even paralogs within classes differed dramatically in expression levels across the two species. The nonsynonymous substitution rates in toxins were higher than in nontoxins, and gene duplication appeared to have played a substantial role in the divergence. The genetics of venom divergence within *C. horridus* was found to be far simpler, but was nonetheless striking for having occurred within a species and between populations in close geographic proximity. The animals used were collected within 200 km of each other. The major evolutionary changes differentiating type A from type B were regulatory. Two classes of toxin, CTLs and SVMs, were essentially lost from the venom-gland transcriptome of the type A animal. The CTLs were still expressed at low levels, but the SVMs were generally not detectable. Whether the loss of SVMs reflects genome deletion or transcriptional silencing remains to be determined. Fixed

**Table 5**Type A *Crotalus horridus* protein identifications by RP-HPLC fractions.

Fraction	Transcript	Exclusive	Exclusive	TSC	% TSC	% Seq. cov.
		UPC	USC			
A1	H33B/H3-2/H3-3	5	7	13	56.52	26.70
	HistoneH2A/H2AZ	2	2	4	17.39	12.50
	Tubulin- $\alpha$ 1B/ $\alpha$ 6	1	1	3	13.04	5.32
A2	PLA2-3a	3	4	198	58.06	43.40
	PLA2-1a	5	5	95	27.86	45.90
	PLA2-2a	4	5	27	7.92	32.80
	SVMP111-1	3	3	5	1.47	4.07
	MYO	2	2	2	0.59	18.60
A3	PLA2-2a	17	25	757	91.87	88.50
A4	CRISP	18	28	146	69.52	50.70
	SVSP-8	6	9	19	9.05	39.20
	NGF	6	8	16	7.62	25.60
	PLA2-2a	5	6	16	7.62	35.20
A5	SVSP-1f	1	1	196	28.24	40.80
	SVSP-1e	1	1	194	27.95	41.70
	SVSP-5	14	22	111	15.99	32.90
	PLA2-1a	2	3	41	5.91	33.60
A6	SVSP-8	11	21	72	19.25	35.00
	SVSP-1e	1	1	52	13.90	38.30
	SVSP-1f	1	1	49	13.10	32.90
	PLA2-3a	2	3	37	9.89	55.70
	PLA2-1a	5	5	31	8.29	50.00
	SVSP-5	8	11	26	6.95	27.50
	PLA2-2a	3	3	23	6.15	29.50
A7	VESP-1a	7	8	20	5.35	18.30
	PLA2-1a	7	11	231	37.93	53.30
	SVSP-9a	1	2	85	13.96	22.00
	PLA2-2a	3	3	78	12.81	30.30
	SVSP-9b/d	1	2	69	11.33	22.00
	SVSP-1e	1	1	32	5.25	32.10
	SVSP-4	8	9	29	4.76	30.50
A8	PLA2-3a	1	2	28	4.60	55.70
	SVSP-2a	1	1	70	14.00	32.10
	SVSP-1e	1	1	67	13.40	45.00
	SVSP-2b	2	2	64	12.80	32.10
	SVSP-1f	1	1	58	11.60	39.60
	SVSP-7	14	17	44	8.80	41.20
	SVSP-8	5	5	34	6.80	27.50
A9	SVSP-3a	2	3	34	6.80	36.30
	LAAO-1a/b	14	17	39	43.33	22.10
	H33B/H3-2/H3-3	11	13	23	25.56	48.10
	PDE-1	9	10	13	14.44	10.10
	SVSP-5	3	3	6	6.67	15.00

Toxin identification was based on peptide spectral evidence of isolated peaks from RP-HPLC. Proteins were first identified using transcriptome database of mRNA sequences identified from the venom gland. Isolated RP-HPLC peaks are separated by horizontal lines. Abbreviations: CRISP—cysteine-rich secretory protein, LAAO—L-amino acid oxidase, MYO—myotoxin-A, NGF—nerve growth factor, PDE—phosphodiesterase, PLA2—phospholipase A<sub>2</sub>, SVMP111—type III snake-venom metalloproteinase, SVSP—snake venom serine proteinase, TSC—total spectrum count, UPC—unique peptide count, USC—unique spectrum count, VESP—vespryn.

nonsynonymous changes in the venom genes appeared to be relatively unimportant. Although we detected eight differences, none of these affected toxins were expressed at high levels. None of the 20 most highly expressed toxin transcripts for type B or the 13 most highly expressed toxin transcripts for type A were affected. We found very few transcripts lacking orthologs across the two individuals, suggesting that gene duplication with subsequent differentiation of the paralogs was not important. Duplication as a mechanism for affecting expression levels (i.e., copy number variation) would have been undetectable from transcriptomic data, except perhaps as a difference in expression level. Because the high toxicity of type A venoms is attributable to the presence of heterodimeric PLA2s (canebrake toxin, Mojave toxin, crotoxin, and sistruxin), the lack of orthologs in the transcriptome of the type B *C. horridus* leaves one of the major aspects of the origin and evolution of type A venom completely unresolved. Although the

**Table 6**Type B *Crotalus horridus* protein identifications by RP-HPLC fractions.

Fraction	Transcript	Exclusive	Exclusive	TSC	% TSC	% Seq. cov.
		UPC	USC			
B1	SVMP11-3a	1	1	20	29.85	9.39
	SVMP11-4a	1	1	15	22.39	5.68
	SVSP-6	4	6	9	13.43	17.50
	VEGF-1	3	4	8	11.94	27.40
	BPP-1a/c	1	1	8	11.94	14.90
B2	HistoneH2AZ/H2A	2	2	4	5.97	12.50
	SVMP11-4b	1	3	1313	38.24	47.40
	SVMP11-4a	1	3	1280	37.27	47.40
	SVMP11-2	1	2	176	5.13	12.40
	SVMP11-4b	1	2	584	27.14	18.70
B3	SVMP11-4a	1	2	581	27.00	18.70
	SVMP11-3a	8	15	511	23.75	38.20
	LAAO-1e	2	4	164	7.62	25.70
	LAAO-1b/d	1	2	155	7.20	21.70
	MYO	3	4	40	27.40	48.80
B4	SVMP11-4a	5	5	27	18.49	17.00
	LAAO-1b/d	4	5	23	15.75	10.20
	LAAO-1e	4	5	22	15.07	12.20
	SVMP11-3a	2	2	13	8.90	17.20
	SVSP-6	4	4	7	4.79	13.80
B5	VEGF-1	7	9	19	79.17	51.30
	NGF-1a	2	2	5	20.83	11.20
	CRISP	14	25	153	85.96	61.50
	NGF-1a	6	10	16	8.99	26.90
	VEGF-1	5	6	9	5.06	28.20
B6	SVSP-2a	1	1	944	28.30	84.20
	SVSP-2b	1	1	848	25.42	84.20
	SVSP-6	24	56	475	14.24	82.90
	PLA2-1d	4	11	440	13.19	83.60
	SVSP-5	9	13	54	26.73	32.50
B7	SVSP-6	10	16	47	23.27	27.50
	SVSP-2b	1	1	34	16.83	38.30
	SVSP-2a	1	1	34	16.83	38.30
	NGF-1a	5	7	13	6.44	22.00
	VESP	4	4	10	4.95	16.10
B8	PLA2-1c	1	1	3325	30.60	88.50
	PLA2-1e	1	1	2827	26.02	86.90
	PLA2-1d	5	14	2459	22.63	100.00
	SVSP-13b	5	9	1914	26.10	93.80
	SVSP-13c	5	13	1871	25.52	94.60
B9	PLA2-1d	3	7	1634	22.29	87.70
	SVSP-2b	1	1	508	6.93	80.40
	SVSP-5	8	13	348	4.75	70.00
	SVSP-13c	3	8	2015	33.52	96.30
	SVSP-13b	2	5	1975	32.85	95.40
B10	PLA2-1d	2	3	619	10.30	70.50
	SVSP-2b	1	1	501	8.33	80.80
	SVSP-2a	1	1	57	13.73	38.30
	SVSP-2b	1	1	57	13.73	38.30
	SVSP-6	13	19	54	13.01	32.10
B11	SVSP-11b	1	1	46	11.08	29.60
	SVSP-11d	2	2	43	10.36	29.60
	SVSP-5	5	7	28	6.75	22.90
	CTL-8a	1	1	25	6.02	50.40
	SVSP-1	5	8	22	5.30	28.70
B12	LAAO-1e	3	4	36	31.03	15.10
	PDE	12	15	25	21.55	17.30
	SVSP-6	7	10	16	13.79	28.30
	SVSP-2a	1	1	14	12.07	30.40
	CTL-2	4	4	8	6.90	36.50
B13	SVMP11-3a	2	2	6	5.17	7.42
	SVMP11-4a	11	16	58	27.10	21.40
	LAAO-1b/d	3	4	32	14.95	17.50
	LAAO-1e	2	2	28	13.08	20.30
	SVMP11-3a/b	6	10	20	9.35	7.94
B14	CRISP	9	13	19	8.88	41.20
	SVSP-6	7	8	15	7.01	22.90
	SVMP11-4b	1	1	13	6.07	13.40
	SVMP11-3a	25	73	3091	51.92	44.80
	SVSP-13b	1	3	375	6.30	82.20
B15	SVSP-13c	2	4	311	5.22	83.00
	SVMP11-5a	19	32	249	25.49	30.70
	SVMP11-2a	18	28	165	16.89	22.60

**Table 6** (continued)

Fraction	Transcript	Exclusive UPC	Exclusive USC	TSC	% TSC	% Seq. cov.
B17	SVMPIII-4b	1	2	97	9.93	26.30
	SVMPIII-4a	1	1	93	9.52	26.30
	LAAO-1e	4	7	45	4.61	25.30
	SVMPIII-4a	1	4	3619	36.14	53.00
	SVMPIII-4b	3	6	3616	36.11	53.00
B18	SVMPIII-3a	10	35	929	9.28	40.60
	SVMPIII-4b	3	7	3705	40.51	54.00
	SVMPIII-4a	3	7	3684	40.28	54.00

Toxin identification was based on peptide spectral evidence of isolated peaks from RP-HPLC. Proteins were first identified using transcriptome database of mRNA sequences identified from the venom gland. Isolated RP-HPLC peaks are separated by horizontal lines. Abbreviations: CRISP—cysteine-rich secretory protein, CTL—C-type lectin, LAAO—L-amino acid oxidase, MYO—myotoxin-A, NGF—nerve growth factor, PDE—phosphodiesterase, PLA2—phospholipase A<sub>2</sub>; SVMPII—type II snake venom metalloproteinase, SVMPIII—type III snake venom metalloproteinase, SVSP—snake venom serine proteinase, TSC—total Spectrum Count, UPC—unique Peptide Count, USC—unique Spectrum Count, VESP—vespryn, VEGF—vascular endothelial growth factor.

ancestral form may have possessed these subunits with subsequent silencing or loss in the type B lineage, the fact that type B showed its own suite of PLA2s that were not detected in type A suggests a replacement event.

To determine whether the genes encoding the two subunits of canebrake toxin were present in the genomes of type B animals but not expressed, we adapted the PCR-based assay of [Wooldridge et al. \(2001\)](#) to the two subunits of canebrake toxin and applied it to 17 individuals from the type A (FL) population and nine from the type B (GA) population. Both subunits readily amplified from the ge-

**Table 8**Overall type B *Crotalus horridus* protein identifications.

Transcript	Exclusive UPC	Exclusive USC	TSC	% Prot. seq. cov.
BPP-1a/c	4	5	83	39.30
CRISP	26	42	264	91.00
CTL-2	10	19	161	75.40
CTL-4	10	17	170	88.10
CTL-7	2	2	7	24.40
CTL-8a	2	4	881	100.00
CTL-9a	7	17	1206	95.90
CTL-10a	1	1	178	100.00
CTL-10b	1	1	175	100.00
CTL-11a	1	3	867	100.00
LAAO-1a	2	4	858	73.30
LAAO-1b	5	6	846	73.50
LAAO-1c	2	4	861	71.70
MYO	3	4	32	48.80
NGF-1a	7	12	140	26.90
NUC	4	4	14	9.32
PDE	17	21	40	23.90
*PDE2	2	3	7	5.30
PLA2-1c	1	1	7699	95.90
PLA2-1d	8	17	5654	100.00
PLA2-1e	1	1	6053	92.60
*PLB	13	17	108	26.20
SVMPII-1	1	3	157	20.90
SVMPII-2	3	4	245	23.10
SVMPII-3a	40	91	5048	67.00
SVMPII-4a	28	48	589	67.70
SVMPIII-1	23	35	203	53.80
SVMPIII-2a	27	51	520	46.80
SVMPIII-3a	19	37	276	47.50
SVMPIII-4a	4	8	9579	63.00
SVMPIII-4b	4	8	9653	63.00
SVSP-1	11	16	66	57.50
SVSP-2a	1	1	2142	84.20
SVSP-2b	1	1	2892	84.20
SVSP-4	5	9	51	58.20
SVSP-5	23	45	1305	85.00
SVSP-6	44	83	1882	90.40
SVSP-7	18	35	196	86.80
SVSP-8	3	4	326	36.50
SVSP-9	14	21	667	75.30
SVSP-11b	3	5	547	72.50
SVSP-11d	3	3	617	62.50
SVSP-12a	1	2	900	77.60
SVSP-12d	1	2	856	77.30
SVSP-13b	4	8	5009	95.40
SVSP-13c	6	14	4883	96.30
VEGF-1	10	19	76	65.00
VESP	7	7	16	35.30

Transcripts indicated with "\*" were not identified as toxins in our transcriptomic analysis. Abbreviations: BPP—Bradykinin-potentiating and C-type natriuretic peptides, CTL—C-type lectin, CRISP—cysteine-rich secretory protein, LAAO—L-amino acid oxidase, NGF—nerve growth factor, NUC—nucleotidase, PDE—phosphodiesterase, PLA2—phospholipase A<sub>2</sub>, PLB—phospholipase B, SVMPIII—type III snake-venom metalloproteinase, SVSP—snake venom serine proteinase, TSC—total spectrum count, UPC—unique peptide count, USC—unique spectrum count, VEGF—vascular endothelial growth factor.

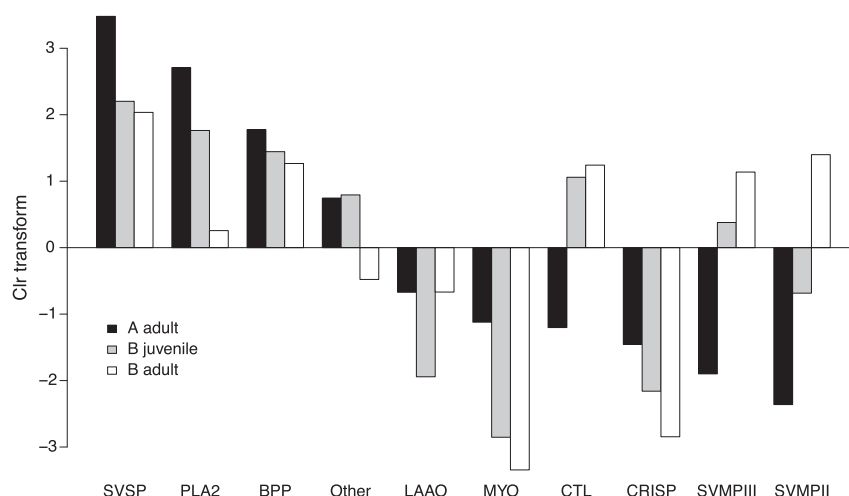
**Table 7**Overall type A *Crotalus horridus* protein identifications.

Transcript	Exclusive UPC	Exclusive USC	TSC	% Prot. seq. cov.
BPP-1a/c	3	3	5	14.30
CRISP	17	26	134	60.20
CTL-10	3	4	13	33.30
CTL-5	8	10	17	68.10
LAAO-1a	16	19	52	18.50
NGF	5	8	58	22.00
NUC	2	2	2	3.39
PDE-1	1	1	11	9.25
PLA2-1a	11	16	662	68.00
PLA2-2a	20	29	1011	90.20
PLA2-3a	3	7	351	70.50
SVMPIII-1	4	4	6	6.44
SVSP-1e	1	1	396	63.70
SVSP-1f	1	1	385	62.10
SVSP-2a	1	1	189	40.40
SVSP-2b/c	2	2	184	40.40
SVSP-3a	3	3	83	47.80
SVSP-4	8	10	66	39.70
SVSP-5	17	26	217	44.60
SVSP-6	5	7	32	27.30
SVSP-7	17	20	52	52.10
SVSP-8	18	34	172	59.20
SVSP-9a	1	2	183	55.60
SVSP-9b/d	1	2	165	55.60
VEGF-1	5	7	21	28.20

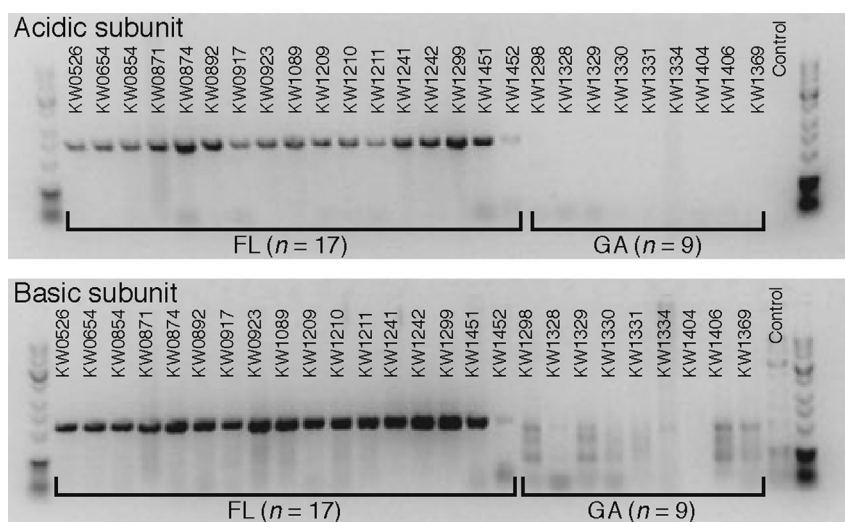
Abbreviations: BPP—Bradykinin-potentiating and C-type natriuretic peptides, CTL—C-type lectin, CRISP—cysteine-rich secretory protein, LAAO—L-amino acid oxidase, NGF—nerve growth factor, NUC—nucleotidase, PDE—phosphodiesterase, PLA2—phospholipase A<sub>2</sub>, SVMPIII—type III snake-venom metalloproteinase, SVSP—snake venom serine proteinase, TSC—total spectrum count, UPC—unique peptide count, USC—unique spectrum count, VEGF—vascular endothelial growth factor.

nomes of all of the type A animals but neither subunit amplified from any of the type B animals ([Fig. 7](#)). We therefore found a distinct, fixed, genetic difference between these populations and determined that the absence of at least two of the more significant type A venom components in type B venoms reflects an absence of the corresponding genes in their genomes.

The only other species (all rattlesnakes) known to have orthologs of the two canebrake toxin subunits are *Crotalus durissus*, *C. scutulatus*, *Crotalus tigris*, and *Sistrurus catenatus*. Only the latter species occurs in sympatry with *C. horridus*, raising the possibility that the origin of type A venom in *C. horridus* involved intergeneric



**Fig. 6.** Some, but not all, of the venom expression differences between type A and type B adults were mirrored in the expression differences between the adult and juvenile type B. Expression levels are displayed as centered logratio transforms of the raw percentages based on alignments of reads to the coding sequences of the superset of toxins from the two adults. Values greater than zero represent highly expressed classes relative to the geometric mean across all classes. The type A adult and type B juvenile showed parallel reductions in snake venom metalloproteinase (SVMP) expression relative to the type B adult and relatively higher levels of phospholipase A<sub>2</sub> (PLA2) expression. The type A adult was unique in showing low levels of C-type lectin (CTL) expression. Abbreviations: BPP—Bradykinin-potentiating and C-type natriuretic peptides, Clr—centered logratio, CRISP—cysteine-rich secretory protein, CTL—C-type lectin, LAAO—L-amino acid oxidase, PLA2—phospholipase A<sub>2</sub>, SVSP—snake venom serine proteinase, SVMPII—snake venom metalloproteinase (type II), SVMPIII—snake venom metalloproteinase (type III).



**Fig. 7.** A PCR-based assay demonstrated that the two canebrake-toxin subunits were present in the genomes of animals from the source population of our type A *Crotalus horridus*, but not in the genomes of the type B population. The presence or absence of these loci appeared to be a fixed difference between these populations. The type A transcriptome animal was KW1089, and the type B transcriptome animal was KW1298.

hybridization. Hybridization has actually been documented in nature between *C. horridus* and *S. catenatus* (Bailey, 1942). Type A transcripts PLA2-2 and PLA2-3 encode the basic and acidic subunits of canebrake toxin and are closely related to the corresponding subunits of Mojave toxin from *C. scutulatus*, crotoxin from *C. durissus*, and sistruxin from *S. catenatus*. The closest match in the National Center for Biotechnology Information (NCBI) nonredundant protein database for type A PLA2-3 (the acidic subunit) was the acidic subunit of sistruxin from *S. catenatus tergeminus* (GenBank accession: AY508693); the coding sequences differed at only six nucleotide positions of 417. The second best match was with the acidic subunit of crotoxin from *C. durissus terrificus* (GenBank accession: X12606), which differed at 13 positions. For the transcript encoding the basic subunit of canebrake toxins (PLA2-2), the best match was from *Crotalus durissus terrificus* (GenBank

accession: X12603) with nine of 417 sites differing. The second best match from *S. catenatus tergeminus* differed at 12 sites. To formally test the hypothesis of a hybrid origin of canebrake toxin, we performed a Bayes factor phylogenetic analysis (Kass and Raftery, 1995; Nylander et al., 2004). We used the acidic and basic subunits of canebrake toxin, Mojave toxin (GenBank accessions U01026 and U01027), crotoxin (GenBank accessions X12606 and X12603), and sistruxin (GenBank accessions AY508693 and AY355170). Under our null model, we enforced *Crotalus* monophyly separately for each subunit. In our alternative model, we constrained the *C. horridus* sequence to be sister to the *S. catenatus tergeminus* sequence for each subunit. We estimated a Bayes factor of 23.8 in favor of the alternative model, which is considered strong evidence against the null model (Kass and Raftery, 1995). A nearly identical result was obtained when the alternative model consisted

of a completely unconstrained search. This result and the absence of these PLA2s in the type B transcriptome and genome (Fig. 7) strongly suggest that the major toxic components of type A venom in *C. horridus* were acquired through intergeneric hybridization. We cannot, however, rule out molecular convergence or introgression from *C. horridus* to *Sistrurus* as explanations for the phylogenetic patterns. The former would have required convergent molecular evolution and loss of both subunit genes in type B *C. horridus*, and the latter would have required intergeneric hybridization and loss of the subunits in type B *C. horridus*. Neither of these possibilities is as parsimonious as our hypothesis of partial intergeneric hybrid origin of type A venom within *C. horridus*.

Our results clearly indicated that, in this case of rapid, intra-specific venom evolution, major phenotypic change was accomplished primarily by means of mutations affecting venom-gene expression, which thereby directly affected compositional changes to the venom. Nonsynonymous point mutations were few in number, did not generally affect the more abundant toxins, and affected far fewer loci than expression changes. Whether expression evolution will prove to generally be the dominant mode of venom evolution over short time scales remains to be determined, but compositional differences have been documented proteomically within several species (Gibbs et al., 2009; Sunagar et al., 2014; Margres et al., 2015), suggesting a role for expression evolution. In our case, we saw a global shutdown of SVMP expression and generalized, but paralog specific, reduction in CTL expression. The former may have involved co-option of a pre-programmed ontogenetic expression change, but the latter probably required multiple *cis*-regulatory changes. The replacement of the entire complement of PLA2s in the type A transcriptome could also be regarded as an expression-based change, although this event seems to have been accomplished by means of hybridization. This wealth of potential mechanisms to affect expression changes may bias the evolution of venoms toward changes in venom composition as opposed to changes in the primary sequences of the individual toxic proteins, despite the lack of pleiotropy. Although these patterns may be specific to venoms (or even this single example of intraspecific venom evolution), pleiotropic constraints should generally be stronger for genes contributing to many other complex traits that are not as modular as venom. These constraints should bias genetic pathways of adaptation even more strongly toward expression changes.

## Ethical statement

All techniques involving vertebrate animals were approved by the Florida State University Institutional Animal Care and Use Committee (IACUC) under protocol #0924.

## Acknowledgments

Funding for this work was provided by the National Science Foundation (NSF DEB-1145978). The authors thank Kate Calvin of the Florida State University College of Medicine Translational Science Laboratory and Margaret Seavy of the Florida State University Department of Biological Science Analytical Lab for advice and assistance with proteomic analyses.

## Transparency document

Transparency document related to this article can be found online at <http://dx.doi.org/10.1016/j.toxicon.2015.02.015>.

## References

- Aitchison, J., 1986. *The Statistical Analysis of Compositional Data*. Chapman and Hall, London.
- Bailey, R.M., 1942. An intergeneric hybrid rattlesnake. *Am. Nat.* 76, 376–385.
- Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340, 783–795.
- Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., Boultell, J.M., Bryant, J., Carter, R.J., Keira Cheetham, R., Cox, A.J., Ellis, D.J., Flatbush, M.R., Gormley, N.A., Humphray, S.J., Irving, L.J., Karvelashvili, M.S., Kirk, S.M., Li, H., Liu, X., Maisinger, K.S., Murray, L.J., Obradovic, B., Ost, T., Parkinson, M.L., Pratt, M.R., Rasolonjatovo, I.M.J., Reed, M.T., Rigatti, R., Rodighiero, C., Ross, M.T., Sabot, A., Sankar, S.V., Scally, A., Schroth, G.P., Smith, M.E., Smith, V.P., Spiridou, A., Torrance, P.E., Tzonev, S.S., Vermaas, E.H., Walter, K., Wu, X., Zhang, L., Alam, M.D., Anastasi, C., Aniebo, I.C., Bailey, D.M.D., Bancarz, I.R., Banerjee, S., Barbour, S.G., Baybayan, P.A., Benoit, V.A., Benson, K.F., Bevis, C., Black, P.J., Boodhun, A., Brennan, J.S., Bridgman, J.A., Brown, R.C., Brown, A.A., Buermann, D.H., Bundu, A.A., Burrows, J.C., Carter, N.P., Castillo, N., Chiara, M., Catenazzi, E., Chang, S., Neil Cooley, R., Crake, N.R., Dada, O.O., Diakoumakos, K.D., Dominguez-Fernandez, B., Earnshaw, D.J., Egbujor, U.C., Elmore, D.W., Etchin, S.S., Ewan, M.R., Fedurco, M., Fraser, L.J., Fuentes Fajardo, K.V., Scott Furey, W., George, D., Gietzen, K.J., Goddard, C.P., Golda, G.S., Granieri, P.A., Green, D.E., Gustafson, D.L., Hansen, N.F., Harnish, K., Haudenschild, C.D., Heyer, N.I., Hims, M.M., Ho, J.T., Horgan, A.M., Hoshler, K., Hurwitz, S., Ivanov, D.V., Johnson, M.Q., James, T., Huw Jones, T.A., Kang, G.-D., Kerelska, T.H., Kersey, A.D., Khrebukova, I., Kindwall, A.P., Kingsbury, Z., Kokko-Gonzales, P.L., Kumar, A., Laurent, M.A., Lawley, C.T., Lee, S.E., Lee, X., Liao, A.K., Loch, J.A., Lok, M., Luo, S., Mammen, R.M., Martin, J.W., McCauley, P.G., McNitt, P., Mehta, P., Moon, K.W., Mullens, J.W., Newington, T., Ning, Z., Ling Ng, B., Novo, S.M., O'Neill, M.J., Osborne, M.A., Osnowski, A., Ostadan, O., Paraschos, L.L., Pickering, L., Pike, A.C., Pike, A.C., Chris Pinkard, D., Pliskin, D.P., Podhasky, J., Quijano, V.J., Raczky, C., Rae, V.H., Rawlings, S.R., Chiva Rodriguez, A., Roe, P.M., Rogers, J., Rogert Bacigalupo, M.C., Romanov, N., Romieu, A., Roth, R.K., Rourke, N.J., Ruediger, S.T., Rusman, E., Sanches-Kuiper, R.M., Schenker, M.R., Seoane, J.M., Shaw, R.J., Shiver, M.K., Short, S.W., Sizto, N.L., Sluis, J.P., Smith, M.A., Ernest Sohna Sohna, J., Spence, E.J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C.L., Turcatti, G., vandeVondele, S., Verhovskiy, Y., Virk, S.M., Wakelin, S., Walcott, G.C., Wang, J., Worsley, G.J., Yan, J., Yau, L., Zuerlein, M., Rogers, J., Mullikin, J.C., Hurler, M.E., McCooke, N.J., West, J.S., Oaks, F.L., Lundberg, P.L., Klennerman, D., Durbin, R., Smith, A.J., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53–59.
- Biardi, J.E., Chien, D.C., Coss, R.G., 2005. California ground squirrel (*Spermophilus beecheyi*) defenses against rattlesnake venom digestive and hemostatic toxins. *J. Chem. Ecol.* 31, 2501–2518.
- Biardi, J.E., Coss, R.G., 2011. Rock squirrel (*Spermophilus variegatus*) blood sera affects proteolytic and hemolytic activities of rattlesnake venoms. *Toxicon* 57, 323–331.
- Brown, W.S., 1993. Biology, status, and management of the timber rattlesnake (*Crotalus horridus*): a guide for conservation. *Soc. Study Amphib. Reptiles Circ.* 22, 78.
- Calvete, J.J., Sanz, L., Cid, P., de la Torre, P., Flores-Díaz, M., Santos, M.C.D., Borges, A., Brema, A., Angulo, Y., Lomonte, B., Alape-Girón, A., Gutiérrez, J.M., 2010. Snake venomomics of the central american rattlesnake *Crotalus simus* and the south american *Crotalus durissus* complex points to neurotoxicity as an adaptive paedomorphic trend along *Crotalus* dispersal in South America. *J. Proteome Res.* 9, 528–544.
- Carroll, S.B., 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134, 25–36.
- Casewell, N.R., 2012. On the ancestral recruitment of metalloproteinases into the venom of snakes. *Toxicon* 60, 449–454.
- Casewell, N.R., Wagstaff, S.C., Harrison, R.A., Renjifo, C., Wüster, W., 2011. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. *Mol. Biol. Evol.* 28, 2637–2649.
- Durban, J., Pérez, A., Sanz, L., Gómez, A., Bonilla, F., Rodríguez, S., Chacón, D., Sasa, M., Angulo, Y., Gutiérrez, J.M., Calvete, J.J., 2013. Integrated “omics” profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, *Crotalus simus simus*. *BMC Genomics* 14, 234.
- Flynt, A.S., Lai, E.C., 2008. Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nat. Rev. Genet.* 9, 831–842.
- Fox, J.W., Serrano, S.M.T., 2005. Structural considerations of the snake venom metalloproteinases, key members of the M12 reprotolysin family of metalloproteinases. *Toxicon* 45, 969–985.
- Gibbs, H.L., Sanz, L., Calvete, J.J., 2009. Snake population venomomics: proteomics-based analyses of individual variation reveals significant gene regulation effects on venom protein expression in *Sistrurus* rattlesnakes. *J. Mol. Evol.* 68, 113–125.
- Glenn, J.L., Straight, R.C., Wolf, T.B., 1994. Regional variation in the presence of canebrake toxin in *Crotalus horridus* venom. *Comp. Biochem. Physiol. C* 107, 337–346.
- Hendon, R.A., Fraenkel-Conrat, H., 1971. Biological roles of the two components of crotoxin. *Proc. Natl. Acad. Sci. U. S. A.* 68, 1560–1563.

- Hoekstra, H.E., Coyne, J.A., 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61, 995–1016.
- Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogeny. *Bioinformatics* 17, 754–755.
- Jansa, S.A., Voss, R.S., 2011. Adaptive evolution of the venom-targeted vWF protein in opossums that eat pitvipers. *PLoS One* 6, e20997.
- Kass, R.E., Raftery, A.E., 1995. Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795.
- Keller, A., Nesvizhskii, A.I., Kolker, E., Aebersold, R., 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74, 5383–5392.
- Klauber, L.M., 1997. *Rattlesnakes: Their Habits, Life Histories, and Influence on Mankind*, second ed. University of California Press, Berkeley, California.
- Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D.W., 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* 23, 1891–1901.
- Mackessy, S.P., 2008. Venom composition in rattlesnakes: trends and biological significance. In: Hayes, W.K., Beaman, K.R., Cardwell, M.D., Bush, S.P. (Eds.), *The Biology of Rattlesnakes*. Loma Linda University Press, Loma Linda, CA, pp. 495–510.
- Man, O., Pilpel, Y., 2007. Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species. *Nat. Genet.* 39, 415–421.
- Margres, M.J., Aronow, K., Loyacano, J., Rokyta, D.R., 2013. The venom-gland transcriptome of the eastern coral snake (*Micrurus fulvius*) reveals high venom complexity in the intragenomic evolution of venoms. *BMC Genomics* 14, 531.
- Margres, M.J., McGivern, J.J., Seavy, M., Wray, K.P., Facente, J., Rokyta, D.R., 2015. Contrasting modes and tempos of venom expression evolution in two snake species. *Genetics* 199, 165–176.
- Margres, M.J., McGivern, J.J., Wray, K.P., Seavy, M., Calvin, K., Rokyta, D.R., 2014. Linking the transcriptome and proteome to characterize the venom of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *J. Proteomics* 96, 145–158.
- McCleary, R.J.R., Heard, D.J., 2010. Venom extraction from anesthetized Florida cottonmouths, *Agkistrodon piscivorus conanti*, using a portable nerve stimulator. *Toxicon* 55, 250–255.
- McDiarmid, R.W., Campbell, J.A., Touré, T., 1999. Snake Species of the World: a Taxonomic and Geographic Reference. In: *Herpetologists' League*, vol. 1. Washington, D. C.
- McGee, L.W., Aitchison, E.W., Caudle, S.B., Morrison, A.J., Zheng, L., Yang, W., Rokyta, D.R., 2014. Payoffs, not tradeoffs, in the adaptation of a virus to ostensibly conflicting selective pressures. *PLoS Genet.* 10, e100461.
- McGivern, J.J., Wray, K.P., Margres, M.J., Couch, M.E., Mackessy, S.P., Rokyta, D.R., 2014. RNA-seq and high-definition mass spectrometry reveal the complex and divergent venoms of two rear-fanged colubrid snakes. *BMC Genomics* 15, 1061.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628.
- Nesvizhskii, A.I., Keller, A., Kolker, E., Aebersold, R., 2003. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 75, 4646–4658.
- Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. *Syst. Biol.* 53, 47–67.
- Rokyta, D.R., Joyce, P., Caudle, S.B., Wichman, H.A., 2005. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. *Nat. Genet.* 37, 441–444.
- Rokyta, D.R., Lemmon, A.R., Margres, M.J., Aronow, K., 2012. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics* 13, 312.
- Rokyta, D.R., Wray, K.P., Lemmon, A.R., Lemmon, E.M., Caudle, S.B., 2011. A high-throughput venom-gland transcriptome for the eastern diamondback rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes. *Toxicon* 57, 657–671.
- Rokyta, D.R., Wray, K.P., Margres, M.J., 2013. The genesis of an exceptionally lethal venom in the timber rattlesnake (*Crotalus horridus*) revealed through comparative venom-gland transcriptomics. *BMC Genomics* 14, 394.
- Rotenberg, D., Bamberger, E.S., Kochva, E., 1971. Studies on ribonucleic acid synthesis in the venom glands of *Vipera palaestinae* (Ophidia, Reptilia). *Biochem. J.* 121, 609–612.
- Stechert, R., 1982. Historical depletion of timber rattlesnake colonies in New York state. *Bull. N. Y. Herpetol. Soc.* 17, 23–24.
- Straight, R.C., Glenn, J.L., 1989. Isolation and characterization of basic phospholipase (PLA<sub>2</sub>) and acidic subunits of canebrake toxin from *Crotalus horridus atricaudatus* venom using HPLC. *Toxicon* 27, 80.
- Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Tyler-Smith, C.L.C., Carter, N., Scherer, S.W., Tavaré, S., Deloukas, P., Hurler, M.E., Dermizakis, E.T., 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315, 848–853.
- Sunagar, K., Undheim, E.A., Scheib, H., Gren, E.C., Cochran, C., Person, C.E., Koludarov, I., Kelln, W., Hayes, W.K., King, G.F., Antunes, A., Fry, B.G., 2014. Intraspecific venom variation in the medically significant southern pacific rattlesnake (*Crotalus oreganus helleri*): biodiversity, clinical and evolutionary implications. *J. Proteomics* 99, 68–83.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Tyning, T.F., 1990. Conservation of the Timber Rattlesnake in the Northeast. Massachusetts Audubon Society, Lincoln, MA.
- Vonk, F.J., Casewell, N.R., Henkel, C.V., Heimberg, A.M., Jansen, H.J., McCleary, R.J.R., Kerkkamp, H.M.E., Vos, R.A., Guerreiro, I., Calvete, J.J., Wuster, W., Woods, A.E., Logan, J.M., Harrison, R.A., Castoe, T.A., de Koning, A.P.J., Pollock, D.D., Yandell, M., Calderon, D., Renjifo, C., Currier, R.B., Salgado, D., Pla, D., Sanz, L., Hyder, A.S., Ribeiro, J.M.C., Arntzen, J.W., van den Thillart, G.E.E.J.M., Boetzer, M., Pirovano, W., Dirks, R.P., Spink, H.P., Duboule, D., McGlinn, E., Kini, R.M., Richardson, M.K., 2013. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc. Natl. Acad. Sci. U. S. A.* 110, 20651–20656.
- Weinreich, D.M., Delaney, N.F., DePristo, M.A., Hartl, D.L., 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312, 111–114.
- Wooldridge, B.J., Pineda, G., Banuelas-Ornelas, J.J., Dagda, R.K., Gasanov, S.E., Rael, E.D., Lieb, C.S., 2001. Mojave rattlesnakes (*Crotalus scutulatus scutulatus*) lacking the acidic subunit DNA sequence lack Mojave toxin in their venom. *Comp. Biochem. Physiol. B* 130, 169–179.