JOURNAL OF PROTEOMICS 96 (2014) 145–158

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Linking the transcriptome and proteome to characterize the venom of the eastern diamondback rattlesnake (*Crotalus adamanteus*)

Mark J. Margres[a], *James J. McGivern*[a], *Kenneth P. Wray*[a], *Margaret Seavy*[a], *Kate Calvin*[b], *Darin R. Rokyta*[a,*]

[a]*Department of Biological Science, Florida State University, Tallahassee, FL, USA*
[b]*College of Medicine, Florida State University, Tallahassee, FL, USA*

ARTICLE INFO

Article history:
Received 1 October 2013
Accepted 1 November 2013
Available online 12 November 2013

Keywords:
Genotype–phenotype map
Snake venomics
Crotalus adamanteus
Mass spectrometry
Liquid chromatography
Transcriptomics

ABSTRACT

Understanding the molecular basis of the phenotype is key to understanding adaptation, and the relationship between genes and specific traits is represented by the genotype–phenotype map. The specialization of the venom-gland towards toxin production enables the use of transcriptomics to identify a large number of loci that contribute to a complex phenotype (i.e., venom), while proteomic techniques allow verification of the secretion of the proteins produced by these loci, creating a genotype–phenotype map. We used the extensive database of mRNA transcripts generated by the venom-gland transcriptome of *Crotalus adamanteus* along with proteomic techniques to complete the genotype–phenotype map for the *C. adamanteus* venom system. Nanospray LC/MS^E analysis of a whole venom sample identified evidence for 52 of the 78 unique putative toxin transcript clusters, including 44 of the 50 most highly expressed transcripts. Tandem mass spectrometry and SDS-PAGE of reversed-phase high-performance liquid chromatography fractions identified 40 toxins which clustered into 20 groups and represented 10 toxin families, creating a genotype–phenotype map. By using the transcriptome to understand the proteome we were able to achieve locus-specific resolution and provide a detailed characterization of the *C. adamanteus* venom system.

Biological significance
Identifying the mechanisms by which genetic variation presents itself to the sieve of selection at the phenotypic level is key to understanding the molecular basis of adaptation, and the first step in understanding this relationship is to identify the genetic basis of the phenotype through the construction of a genotype–phenotype map. We used the high-throughput venom-gland transcriptomic characterization of the eastern diamondback rattlesnake (*C. adamanteus*) and proteomic techniques to complete and confirm the genotype–phenotype map, providing a detailed characterization of the *C. adamanteus* venom system.

* *Corresponding author at:* Florida State University, Department of Biological Science, 319 Stadium Drive, Tallahassee, FL 32306, USA. Tel.: +1 850 645 8812.
E-mail address: drokyta@bio.fsu.edu (D.R. Rokyta).

1874-3919/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jprot.2013.11.001

1. Introduction

Darwin's theory of natural selection requires that individuals exhibit heritable phenotypic variation, and this variation must result in differential fitness (e.g., survival and/or reproduction) that enables the sorting of phenotypes in the next generation [1]. Identifying the mechanisms of adaptation is contingent on first understanding the molecular basis of the phenotype. This relationship between genes and specific traits is represented by the genotype–phenotype map [2], the construction of which could enable the prediction of phenotypes from genotypes [1]. However, since its first description by Alberch in 1991 [2], it has been stated that the genotype–phenotype map is nonlinear due to canalization [3], phenotypic plasticity [4,5] and the lack of correlation between mRNA levels and protein abundances [2,6–8], among other phenomena. Travisano and Shaw [9] recently argued that the molecular pathway from genotype to phenotype is so complex that it is intractable, will never be reliably predicted, and lacks explanatory value. However, the authors focused on the relationship between genotype and phenotype for morphological traits, ignoring complex biochemical phenotypes such as venoms.

Snake venom glands presumably arose a single time 60–80 million years ago [10,11] (but see [12]), and facilitated a major transition in the feeding ecology of venomous taxa, from purely mechanical means (e.g., constriction) to incapacitation by means of biochemical weaponry. Venoms are integrated systems of proteins with simple biochemical properties and, while the majority of toxins function independently at the biochemical level, they act synergistically to increase the overall efficacy of the venom [13]. Venoms not only immobilize and kill prey, but also aid in digestion [14] and provide a means of defense against predators [15]. Venom is a modular, polygenic trait, with >60 loci potentially contributing to venom composition [16–18]. The specialization of the venom-gland towards toxin production enables the use of transcriptomics to identify a large number of loci that contribute to a specific evolutionarily significant phenotype, while proteomic techniques such as mass spectrometry allow verification of the secretion of the proteins produced by these loci. Correlating specific transcripts with specific toxic proteins while retaining the context of whole venom (e.g., the suite of proteins contributing to venom composition and their biochemical effects) establishes a direct link from genotype to phenotype, creating the genotype–phenotype map. Many proteomic analyses of venoms to date have used mass spectrometry to identify putative toxic proteins, but have relied on public databases rather than a species-specific transcriptome [19–23]. This approach is effective at characterizing venom composition at the gene-family level, but cannot describe venom complexity at either the genomic or proteomic level that could be obtained through the creation of a genotype–phenotype map (e.g., locus-specific resolution). Several studies of the snake venom system have utilized a combined transcriptomic and proteomic approach to correlate specific transcripts with specific proteins to create a genotype–phenotype map but relied on low-throughput sequencing approaches [24–29]. These low-throughput approaches typically generate 1000 ESTs [25,28] and, as a result of low-coverage, identify a small number of full-length transcripts (e.g., 78 total nontoxins and toxins [28]) and fail to detect many of the low-abundance transcripts [24–29]. Although these approaches often correlate specific transcripts with specific proteins [25,28], an unambiguous protein identification requires knowledge of the full range of potential matches (e.g., if 20 serine proteinases are present in the proteome but only a single full-length transcript is sequenced in a low-throughput transcriptome, any proteomic evidence will appear unambiguous). High-throughput, high-coverage transcriptomes should alleviate this bias as these approaches have been shown to identify a great number of full-length transcripts (e.g., 2066–3092) being actively expressed in snake venom glands [16–18], including low-abundance transcripts.

Rokyta et al. [16] recently used a high-throughput transcriptomic approach to generate 95,643,958 pairs of reads and identify 3002 full-length transcripts, including 78 putative toxin clusters (representing 123 unique transcripts grouped into clusters with <1% nucleotide divergence), being actively expressed in the venom gland of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *Crotalus adamanteus* is the largest species of rattlesnake in the world with a record length of 2.44 m and is historically native to seven states in the southeastern United States [30]. *Crotalus adamanteus* primarily consumes small mammals and birds with mice, rats, squirrels, and rabbits forming the majority of the diet [30]. Myotoxin-A (crotamine), snake venom metalloproteinases (SVMPs), C-type lectins (CTLs), and snake venom serine proteinases (SVSPs) dominated the venom-gland transcriptome of *C. adamanteus*. We used the extensive database of mRNA transcripts (2879 nontoxins and 123 toxins) generated by Rokyta et al. [16] along with reversed-phase high-performance liquid chromatography (RP-HPLC), N-terminal sequencing, SDS-PAGE, and two mass spectrometry techniques to construct a genotype–phenotype map for the *C. adamanteus* venom system.

2. Materials and methods

2.1. Venom collection

Rokyta et al. [31,16] sequenced the venom-gland transcriptome of a single *C. adamanteus* from Florida (Wakulla County). The specimen was a female weighing 393 g with a snout-to-vent length (SVL) of 792 mm and a total length (TL) of 844 mm. The animal used as the source of venom for proteomic analyses was a female from Florida (Leon County) collected approximately 35 km from the capture-site of the transcriptome animal and weighed 430 g with a SVL of 724 mm and a TL of 775 mm. Venom was extracted from both animals by electrostimulation under anesthesia [32]. The above procedures and techniques were approved by the Florida State University Institutional Animal Care and Use Committee (IACUC) under protocol #0924.

2.2. Nanospray LC/MS^E

We performed nanospray LC/MS^E using the Synapt G2 HD Mass Spectrometer with an integrated *nano*Acquity UPLC (Waters Corp.) to analyze a whole venom sample from the proteome animal. Digestion of the whole venom sample 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 whole venom digests were adjusted to 3% acetonitrile in LC/MS grade water (J.T. Baker) with 0.1% formic acid. Sample load was approximated at 500 ng and glufibrinopeptide (785.8426 *m/z*, Waters Corp.) was used as the lock mass (external calibrant). Tryptic peptides were separated by reversed-phase chromatography using a Waters *nano*Acquity UPLC BEH130 C18 column with dimensions of 75 μm × 250 mm and 1.7 μm bead size. Gradient conditions were as follows: mobile phase A solvent was 0.1% formic acid, mobile phase B solvent was 0.1% formic acid in acetonitrile, and the column was maintained at 55 °C with a flow rate of 450 nL/min. The column was pre-equilibrated at initial conditions of 7% B and the gradient proceeded 7–35% B over 85 min, 35–50% B over 5 min, 50–80% B over 2 min, and remained at 80% B for 5 min before returning to 7% B over 3 min. Data were acquired in nanoESI Positive mode on a mass range of 50–2000 *m/z* and the time of flight resolution was set at 20,000. The ion source temperature was 80 °C, capillary and cone voltages were 2.8 kV and 30 V, respectively, and nanoflow gas was 0.5 bar. Fragmentation occurred in the trap collision cell with low energy collision set at 4 V and high energy collision set over a ramp of 15–40 V.

Raw data were generated using MassLynx version 4.1 software (Waters Corp.) and data were processed in ProteinLynx Global SERVER version 3.0. We generated a database specific to *C. adamanteus* by translating the toxin and nontoxin mRNA transcripts identified in the venom-gland transcriptome [16]. Signal peptides, short protein regions that mediate the targeting and transporting of the pre-protein that are cleaved prior to expression [33], were identified by SignalP [34,35] and removed from translated putative toxin transcripts. Proteins were identified using the PLGS Identity^E algorithm to search our database containing 3002 entries specific to *C. adamanteus*. A decoy database was generated and searched in the analysis to test for false-positive identifications. Search parameters allowed for precursor and fragment mass tolerances to be set by the software based on resolution (6 ppm and 15 ppm, respectively), one missed cleavage site, and post-translational modifications of cysteine carbamidomethylation (fixed) and oxidation of methionine (variable). Protein identifications were accepted if they met the following criteria: ≥3 matched peptides, ≥20% sequence coverage, and a higher protein score than the highest scoring decoy identified. Identifications were considered unique if they possessed ≥1 distinguishing peptides.

2.3. Gel filtration chromatography

Gel filtration chromatography was used to separate 2.5 mg of venom into size-selected pools on a Superose 12 column (10/300 GL, GE Healthcare) attached to a Beckman System Gold HPLC. The column was equilibrated in 50 mM ammonium bicarbonate (natural pH ≈ 8.1) at 0.4 mL per minute. The column effluent was monitored at 280 nm and 0.4 mL fractions were collected. Fractions were combined based on the elution profile and these pooled fractions were lyophilized and stored at – 80°C.

2.4. Reversed-phase high-performance liquid chromatography

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 μg of total protein were injected onto a Jupiter C18 column, 250 × 4.6 mm (Phenomenex, Torrence, CA) using the standard solvent system of A = 0.1% trifluoroacetic acid (TFA) in water and B = 0.075% TFA in acetonitrile. After 5 min 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.

RP-HPLC peaks were quantified using the Beckman 32 Karat Software Version 8.0. Relative amounts of individual peaks were determined by measuring the area under each peak relative to the total area of all protein peaks identified. According to the Lambert–Beer law, this relative amount corresponds to the percentage of total peptide bonds in the sample [36], and this measure has been shown to be a useful proxy of the relative amount of a specific protein by weight [37].

2.5. SDS-PAGE analysis

Twelve percent polyacrylamide mini-gels using a 3.9% stacking gel (37.5:1, Amresco Bis-Acryl, Ultra-Pure Grade) at a thickness of 1.0 mm were hand cast (4× Tris–HCl/0.2% SDS in aqueous solution, pH 8.8 for resolving gels, and pH 6.8 for stacking gels) and run using the BioRad Mini-PROTEAN Tetra Cell electrophoresis system. For each RP-HPLC fraction, 350 ng of total protein were prepared 1:1 in SDS-sample buffer [aqueous solution: 4× Tris–HCl/0.2% SDS, pH 6.8; 20% (w/v) glycerol; 4% (w/v) SDS; 2% (v/v) β-mercaptoethanol (final concentration 28 mM)]. Samples were heated to 95 °C for 5 min in sample buffer immediately prior to electrophoresis. SDS-PAGE was conducted using SDS buffer containing 0.125 M Tris base, 0.96 M glycine, and 0.5% (w/v) SDS. Gels were run at 100 V for 2 h and silver stained in accordance with BioRad's Silver Stain Plus protocol. Gel images were produced using the BioRad Molecular Imager ChemiDoc XRS System (BioRad Laboratories, Inc.). Densitometry was performed using Bio Rad's Quantity One software image acquisition and 1-D analysis software. Band densities were measured by taking the average density of the band after accounting for background staining.

2.6. Protein sequencing

Aliquots of RP-HPLC fractions (collected as described above) were spotted onto glass fiber filters for N-terminal sequence analysis. Fractions that appeared to be a mixture of several proteins were dissolved in 0.2 M Tris–HCl (pH 8), 6 M guanidine–HCl, and 1% (v/v) 2-mercaptoethanol prior to incubation at 37 °C for 3 h. Incubation was followed by a 30 minute reaction with 5% (v/v) 4-vinylpyridine at room temperature in the dark. Reduced/alkylated proteins were purified on reversed phase (25–75% B at 0.5% per minute) and either sequenced directly or digested with trypsin as described below.

Fractions that appeared to have blocked N-termini were redissolved in 100 μL of 100 mM ammonium bicarbonate, 0.2 μg trypsin (Promega, Madison, WI) added, and the solution

incubated overnight at 37 °C. Tryptic peptides were separated on a Jupiter Proteo Column (250 × 2 mm, Phenomenex) with a linear gradient of 5–50% B at 0.5% per minute and a flow rate of 0.2 mL per minute. Column effluent was monitored at 220 nm, fractions were collected manually, and selected peptides sequenced. N-terminal sequencing was performed on a Procise 492 cLC (Applied Biosystems/Life Technologies, Carlsbad, CA) using standard sequencing protocols provided by the manufacturer. Peptide sequences were searched against GenBank's non-redundant protein database (nr) using the blastp algorithm (excluding the *C. adamanteus* data of Rokyta et al. [16]) and against our database specific to *C. adamanteus* for comparison.

2.7. Tandem mass spectrometry

Reversed-phase high-performance liquid chromatography fractions and gel filtration chromatography fractions (e.g., size-selected fractions) were collected as described above. 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°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 μm ID × 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 μm OD × 75 μm 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. In order 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.

A Finnigan LTQ linear ion trap mass spectrometer (Thermo Scientific, Waltham, MA) was used to perform tandem mass spectrometry. A nano-electrospray source was used with voltage applied via a liquid junction at the head of the analytical column. Each data dependent acquisition (DDA) scan cycle consisted of one full mass spectrum scan for the *m/z* range of 410–2000 collected in profile mode, one zoom scan in profile mode for the identified dominant peak, and a collision-induced dissociation MS/MS event for that peak acquired in centroid mode. Each peak subjected to a DDA cycle was then added to the dynamic exclusion list for 18 s. The minimum peak intensity for MS/MS to activate was set at 1000 with an activation time of 30 ms and a normalized collision energy of 35%. All samples were run in triplicate.

Tandem mass spectra were extracted by Proteome Discoverer version 1.4.0.288. All MS/MS samples were analyzed using SEQUEST version 1.3.0.339 (Thermo Fisher Scientific, San Jose, CA). SEQUEST searched the 3002 entries in our *C. adamanteus*-specific database (see above) assuming the digestion enzyme trypsin. SEQUEST was searched with a fragment ion mass tolerance of 0.80 Da and a parent ion tolerance of 2.0 Da. Oxidation of methionine (variable), carbamidomethylation of cysteine (fixed), and carboxymethylation of cysteine (fixed) were specified in SEQUEST as potential post-translational modifications. A decoy database was generated and searched in the analysis to test for false-positive identifications. Spectra were also searched against the UniprotKB/SwissProt database containing all reviewed Viperidae protein sequences, excluding *C. adamanteus* (1166 sequences; downloaded August 22, 2013).

Scaffold version 4.0.4 (Proteome Software Inc., Portland, OR) was used to validate MS/MS-based peptide and protein identifications. Individual fraction replicates were combined into a single biosample during analysis. Peptide identifications from both database searches were accepted if they met the following criteria: ≥2 unique peptides, more total peptide spectral matches (PSMs) and total unique peptide counts than the maximum PSMs and unique peptide counts identified for the false-discovery decoys for a given fraction, and ≥20% sequence coverage.

3. Results

3.1. The venom proteome of *C. adamanteus*

Nanospray LC/MS^E analysis of a whole venom sample identified peptide evidence for 52 of the 78 unique toxin transcript clusters, including 44 of the 50 most highly expressed transcripts, that belonged to 12 toxin classes (Fig. 1; Table 1; raw data in Supplementary Table 1). We detected unique peptide evidence for 27 putative toxic proteins, and these unique identifications are indicated by asterisks in Table 1. Although our identification of 52 toxic proteins in the venom may be accurate, it may also be an overestimation as we were unable to distinguish between different toxin family members in all cases due to protein similarity. The 27 identifications with unique evidence are a conservative estimate of the number of toxins contributing to the venom of *C. adamanteus*. We also identified unique evidence for three transcript products in our nanospray LC/MS^E analysis that corresponded to putative nontoxic proteins: phospholipase B (PLB), ectonucleotide pyrophosphatase/phosphodiesterase family member 3 (PDE-3), and reticulocalbin 2 EF-hand calcium-binding domain precursor. These three transcripts were described as putative nontoxins by Rokyta et al. [16], but the detection of these proteins by our nanospray LC/MS^E analysis suggests these proteins are secreted into the venom and may have toxic functions (see Discussion).

Tandem mass spectrometry analysis of the seven gel filtration chromatography fractions identified 28 proteins that clustered into 20 unique groups (Table 2), including CTL-12 and SVMP type III 1, which were not identified in our nanospray LC/MS^E analysis. Proteins that could not be differentiated from one another due to shared peptide evidence were grouped into clusters, and these 20 clusters belonged to six toxin families.

3.2. The quantitative genotype–phenotype map

We identified 21 distinct peaks following RP-HPLC (Fig. 2A). We chose this method for fractionation because it allows quantification of protein amounts by measuring the area under each peak relative to the total area of all protein peaks identified.

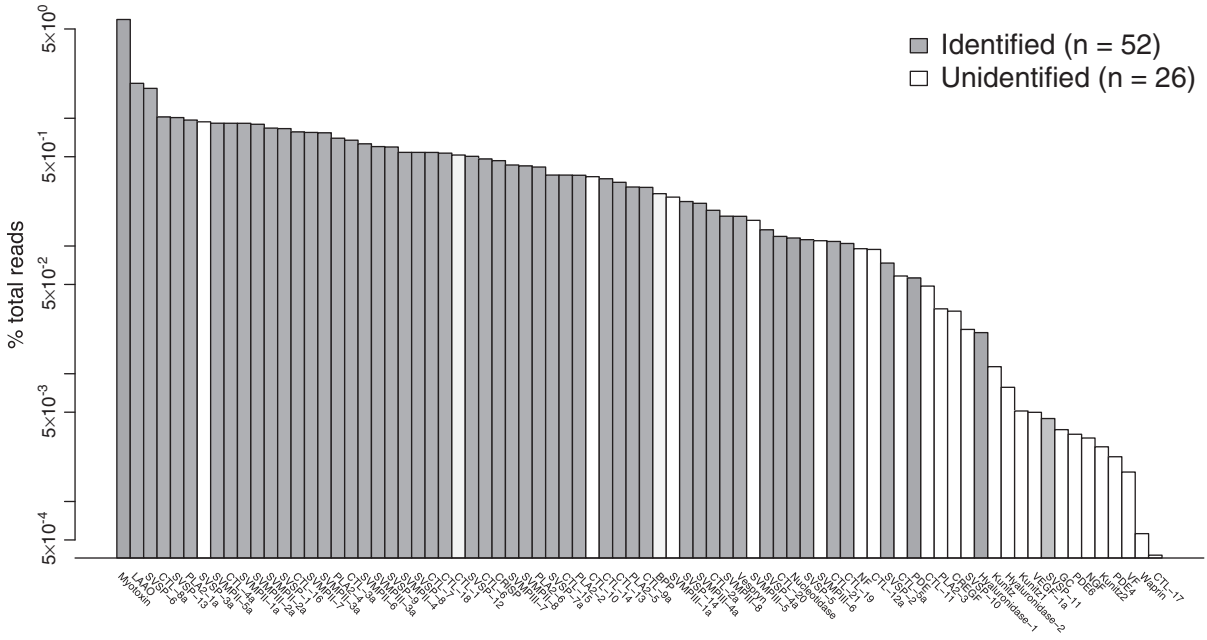

Fig. 1 – The majority of the most highly expressed transcripts identified in the transcriptome of *Crotalus adamanteus* were verified proteomically. Nanospray LC/MS^E analysis of a whole venom sample identified 52 of the 78 toxin clusters from the venom-gland transcriptome of *Crotalus adamanteus*, accounting for the majority of the most highly expressed transcripts. Abbreviations: BPP, Bradykinin-potentiating and C-type natriuretic peptides; CTL, C-type lectin; CREGF, Cysteine-rich with EGF-like domain; CRISP, Cysteine-rich secretory protein; GC, Glutaminyl-peptide cyclotransferase; KUN, Kunitz-type protease inhibitor; LAAO, L-amino-acid oxidase; NGF, Nerve growth factor; NF, Neurotrophic factor; PDE, Phosphodiesterase; PLA$_2$, Phospholipase A$_2$; SVMP, Snake venom metalloproteinase (types II and III); SVSP, Snake venom serine proteinase; VEGF, Vascular endothelial growth factor; VF, Venom factor.

SDS-PAGE analysis of each individual fraction resulted in the identification of nine toxin families on the basis of relative molecular weights according to the transcriptome translation estimates and those described by Mackessy [14] (Fig. 3). Using GenBank's non-redundant protein database (nr) (excluding the toxins identified by Rokyta et al. [16]), N-terminal sequencing identified seven toxin families (Table 3). By using the database generated by translating the mRNA toxin transcripts identified in the venom-gland transcriptome of *C. adamanteus* [16], N-terminal sequencing resulted in the identification of 44 toxins representing the same seven toxin families identified when searching the public database. These 44 proteins clustered into 15 distinct groups on the basis of protein similarity as described above (Table 3).

We next performed tandem mass spectrometry analysis of each RP-HPLC fraction, first searching a UniprotKB/SwissProt database containing all reviewed Viperidae protein sequences, excluding *C. adamanteus*. We identified two toxins belonging to two toxin classes: a phospholipase A$_2$ (PA23_CROAT) in fraction 12 and a CTL (LEC1_BITGA) in fraction 20. By using the database specific to *C. adamanteus*, we identified 36 toxins that grouped into 17 clusters on the basis of shared peptide evidence (Table 4; raw data, including peptide sequences and charges, are included in Supplementary Table 2). SDS-PAGE analysis identified three families that were not detected by tandem mass spectrometry: nucleotidases and hyaluronidases in fraction 17 and PDEs in fraction 16. Based on the nanospray LC/MS^E results, we can infer that these families are represented by PDE/PDE-3 and hyaluronidase-1 (only a single nucleotidase was identified in the transcriptome). By adding the proteins identified in the SDS-PAGE analysis to our tandem mass spectrometry analysis using the species-specific database, the genotype–phenotype map included 40 toxins that clustered into 20 groups and represented 10 toxin families, including CTL-1, which was not detected in our nanospray LC/MS^E analysis.

In total, our three mass spectrometry analyses identified peptide evidence for 58 proteins including 55 of the 78 toxin clusters identified in the venom-gland transcriptome (Table 5). We detected unique evidence for 36 of these 55 putative toxic proteins along with three putative nontoxins (but see Discussion).

3.3. *Transcriptome versus proteome*

We compared the RP-HPLC chromatograms for the female sequenced in the venom-gland transcriptome [16] (Fig. 2A) and the female used in our proteomic investigations (Fig. 2B). Venom compositions for the transcriptome and proteome animals were very similar with the exception of peak six. This peak was absent in the transcriptome animal while present in the proteome animal, and SDS-PAGE, N-terminal sequencing, and mass spectrometry were unable to identify which protein was present in this fraction, potentially due to its absence in the transcriptome animal.

We next compared mRNA levels and protein abundances for the eight most abundant toxin classes in the transcriptome [16]

Table 1 – Nanospray LC/MS^E protein identifications. Asterisks indicate proteins with unique peptide identifications. Protein identifications without asterisks lack a distinguishing peptide but share peptide evidence with multiple proteins. Abbreviations: CTL, C-type lectin; CRISP, Cysteine-rich secretory protein; HYAL, Hyaluronidase; LAAO, L-amino-acid oxidase; MYO, Myotoxin (crotamine); NUC, Nucleotidase; PDE, Phosphodiesterase; PLA$_2$, Phospholipase A$_2$; PLB, Phospholipase B; RET-2, Reticulocalbin 2 EF-hand calcium-binding domain; SVMP, Snake venom metalloproteinase (types II and III); SVSP, Snake venom serine proteinase; VESP, Vespryn (ohanin-like).

Protein	Protein score	Matched peptides	Coverage (%)	GenBank accessions
CRISP*	19,930.4	24	90	JU173623
CTL-2a	10,769.1	8	51	JU173638
CTL-2b	10,769.1	8	51	JU173639
CTL-3a	3987.3	11	62	JU173640
CTL-3b	3895.3	10	45	JU173641
CTL-3c	3895.3	10	45	JU173642
CTL-3d	3987.3	11	62	JU173643
CTL-3e	3895.3	10	45	JU173644
CTL-3f	3942.3	8	36	JU173645
CTL-4a	1925.4	8	76	JU173646
CTL-4b	1925.4	6	66	JU173647
CTL-4c	717.7	5	34	JU173648
CTL-4d	1925.4	6	66	JU173649
CTL-4e	1848.5	5	35	JU173650
CTL-6*	22,417.4	10	61	JU173654
CTL-7	3895.3	10	45	JU173655
CTL-8a	4110.4	4	58	JU173656
CTL-8b	4110.4	4	58	JU173657
CTL-9a*	701.2	8	47	JU173659
CTL-9b	610.2	5	28	JU173660
CTL-13*	26,594.5	10	85	JU173628
CTL-14	1848.5	7	49	JU173629
CTL-15	1848.5	5	39	JU173630
CTL-16*	4206.3	5	84	JU173631
CTL-18	1848.5	6	40	JU173633
CTL-19	5488.2	6	40	JU173634
CTL-20	10,845.9	7	71	JU173636
CTL-21	10,769.1	8	53	JU173637
HYAL-1*	735.3	11	27	JU173662
LAAO*	7633.6	33	72	JU173667
MYO*	11,356.4	3	24	JU173668
NUC*	1663.1	28	60	JU173671
PDE*	2274.4	37	61	JU173674
PDE-3*	2389.9	39	60	JU175352
PLA$_2$-1a	21,358.1	10	73	JU173675
PLA$_2$-1b	21,384.8	12	78	JU173676
PLA$_2$-2	20,563.5	6	44	JU173677
PLA$_2$-4	19,178.8	8	60	JU173679
PLA$_2$-5	21,358.1	9	68	JU173680
PLA$_2$-6	19,020.2	9	51	JU173681
PLB*	3324.9	22	54	JU175433
RET-2*	771.0	17	23	JU175619
SVMPII-1a	1534.9	23	36	JU173682
SVMPII-1b	1534.9	23	36	JU173683
SVMPII-1c	1534.9	23	36	JU173684
SVMPII-1d	1534.9	23	36	JU173685
SVMPII-1e	1534.9	23	36	JU173686
SVMPII-2a	8422.6	20	56	JU173687
SVMPII-2b	8422.6	20	56	JU173688
SVMPII-3a	3733.6	23	60	JU173689
SVMPII-3b	3733.6	23	60	JU173690
SVMPII-3c	3733.6	23	60	JU173691
SVMPII-3d	3733.6	23	60	JU173692
SVMPII-4	1542.7	22	33	JU173693
SVMPII-5a	6224.0	20	54	JU173694
SVMPII-5b	6224.0	19	51	JU173695
SVMPII-5c	6224.0	19	51	JU173696
SVMPII-5d	6212.5	18	44	JU173697
SVMPII-5e	6224.0	19	51	JU173698
SVMPII-5f	6224.0	20	54	JU173699
SVMPII-5 g	6216.1	18	50	JU173700
SVMPII-5 h	6215.7	19	46	JU173701
SVMPII-6	6773.8	19	54	JU173702
SVMPII-7	9262.8	19	58	JU173703
SVMPII-8*	3543.8	26	61	JU173704
SVMPIII-2a*	22,964.3	24	63	JU173707
SVMPIII-2b	22,964.3	22	61	JU173708
SVMPIII-2c	22,964.3	23	63	JU173709
SVMPIII-2d*	28,346.4	27	64	JU173710
SVMPIII-2e	21,229.9	18	49	JU173711
SVMPIII-3a*	3447.2	37	61	JU173712
SVMPIII-3b	3462.7	38	63	JU173713
SVMPIII-3c	3441.8	34	53	JU173714
SVMPIII-4a*	6007.1	37	74	JU173715
SVMPIII-4b	5990.7	35	62	JU173716
SVMPIII-7	3448.5	35	53	JU173719
SVMPIII-8*	1317.1	14	25	JU173720
SVSP-1*	77,740.8	15	87	JU173726
SVSP-2*	26,767.0	15	76	JU173727
SVSP-4*	39,005.0	9	75	JU173730
SVSP-5*	17,347.8	10	80	JU173732
SVSP-6	124,414.3	17	90	JU173733
SVSP-7a	15,255.2	8	50	JU173734
SVSP-7b	15,255.2	9	64	JU173735
SVSP-7c	15,255.2	10	66	JU173736
SVSP-8*	17,265.1	11	84	JU173737
SVSP-9*	47,340.3	12	71	JU173738
SVSP-11*	3535.0	11	65	JU173722
SVSP-12*	92,409.2	15	81	JU173723
SVSP-13*	35,050.3	9	58	JU173724
SVSP-14*	15,287.8	12	81	JU173725
VESP*	1763.1	9	43	JU173741

to a χ^2 distribution with seven degrees of freedom (Fig. 4). mRNA abundances were measured as the percentage of reads mapping to a specific transcript, and the most abundant toxin classes in the venom-gland transcriptome were (in order from most to least) SVMPs, CTLs, SVSPs, and myotoxin [16]. Reversed-phase high-performance liquid chromatography peak abundances were calculated as described above. For peaks that contained multiple toxin classes, class percentages were measured by densitometry of SDS-PAGE gels. If a toxin class was identified by tandem mass spectrometry but lacked a distinct gel band, it was arbitrarily assigned 5% of the fraction percentage. The most abundant toxins in the proteome were (in order from most to least) phospholipases A_2 (PLA$_2$s), SVSPs, myotoxin, and SVMPs. Protein and mRNA levels were significantly different (χ^2 =77.754, $p < 0001$, $df = 7$), with the greatest discordance between CTLs, L-amino-acid oxidase, and PLA$_2$s. Protein/mRNA abundances for SVSPs, SVMPs type III, and the individual cysteine-rich secretory protein were well-correlated.

Table 2 – Tandem mass spectrometry analysis of the gel filtration chromatography (e.g., size-selected) fractions. Fractions are in descending order, with fraction A containing the largest proteins. Complex formation, non-specific associations between proteins, and three-dimensional shape may also determine in which fraction specific proteins elute. Abbreviations: CTL, C-type lectin; CRISP, Cysteine-rich secretory protein; LAAO, L-amino-acid oxidase; PLA_2, Phospholipase A_2; SVMP, Snake venom metalloproteinase (types II and III); SVSP, Snake venom serine proteinase.

Fraction	Protein	Total spectra	Unique peptides	Coverage (%)
A	SVSP-6, 13	34	4	38
	SVMPII-2, 7	29	7	24
	SVMPII-2	14	6	24
	CTL-4, 14	17	4	40
	CTL-8, 16	14	3	53
B	SVMPIII-3, 7	936	27	43
	SVMPIII-1	64	12	27
C	LAAO	283	30	63
	SVMPIII-3, 7	237	26	38
	SVMPII-1, 4	205	12	32
	SVMPII-2, 7	151	14	34
	CTL-9	6	2	30
D	CTL-14	253	9	50
	CTL-4	163	6	50
	SVSP-6, 13	175	9	69
	SVSP-7	12	5	26
	SVMPII-1, 4	179	13	35
	SVMPII-2, 7	174	15	34
	SVMPIII-3, 7	117	16	22
	CTL-8, 16	70	4	53
	SVMPIII-4	38	11	24
	LAAO	28	10	26
	CTL-12	5	2	23
E	PLA_2-1	394	9	73
	SVSP-6, 13	228	9	72
	CTL-14	111	9	54
	SVSP-1	106	9	66
	SVMPIII-3, 7	104	14	21
	CRISP	26	4	29
F	CRISP	267	19	65
	SVSP-1	160	10	76
	PLA_2-1, 4, 5	72	5	72
	CTL-6	42	9	67
	CTL-1, 13	34	8	61
	CTL-14	22	3	39
G	SVMPIII-4	50	8	22
	CTL-4, 14	44	4	47
	CTL-6	31	6	51
	CTL-1, 13	26	6	50
	CTL-9	31	5	62
	CRISP	24	6	35
	SVSP-1	24	5	42
	CTL-2, 21	24	4	31
	SVSP-6, 13	19	2	29

4. Discussion

We identified 55 putative toxins in the venom proteome of *C. adamanteus*, the majority of which were PLA_2s, SVMPs, SVSPs, and myotoxin. By using the transcriptome data of Rokyta et al. [16] as a reference database, we were often able to identify the protein products of individual mRNA transcripts rather than simply the gene-family resolution attained through traditional approaches of searching public databases. This species-specific database allowed us to map 40 toxin transcript products to specific RP-HPLC fractions, and correlating specific transcripts with specific toxic proteins while retaining the context of whole venom establishes a direct link from genotype to venom composition, generating a genotype–phenotype map for an ecologically critical trait. We were also able to use stringent parameters when identifying proteins following mass spectrometry, limiting false positive identifications yet retaining the majority of the data (e.g., 40 mapped toxins). For comparison, we used the same parameters when searching against the UniprotKB/Swiss-Prot database containing all reviewed Viperidae protein sequences (excluding *C. adamanteus*) and were only able to identify two toxins.

Although we were able to identify evidence for 55 putative toxic proteins, we were unable to distinguish between different toxin family members in all cases due to protein similarity. This is a limitation inherent in the current methodology as high levels of redundancy are problematic when attempting to distinguish among protein family members. Rokyta et al. [16] grouped toxin transcripts into clusters with <1% nucleotide divergence to account for alleles, recent duplicates, and/or sequencing errors. Our tandem mass spectrometry analyses (Tables 3 and 5) did not have the resolution capable of distinguishing among individual isoforms of toxins (e.g., members of the same transcript cluster). However, our nanospray LC/MS^E analysis was capable of resolving these differences in some, but by no means all, cases based on the presence of a single distinguishing peptide (e.g., CTL-9a in Table 1).

We also identified three putative nontoxins in our LC/MS^E analysis: PLB, PDE-3, and reticulocalbin 2 EF-hand calcium-binding domain precursor. PLB has been previously identified as a putative toxin [31,18,38–40] and possesses hemolytic/cytotoxic activity in *Pseudechis colletti* [38], while PDEs are known toxins whose putative function is to liberate toxic nucleosides [41–43]. PDE activity has been detected in *C. adamanteus* venom [14] while PLB activity awaits verification. PDE-3 was listed as a nontoxin by Rokyta et al. [16] as it lacked a signal peptide (as did PLB), and presumably is not secreted. The discovery of these proteins in the venom, however, warrants further investigation. Reticulocalbin 2 EF-hand calcium-binding domain precursor was a highly expressed nontoxin in the venom-gland transcriptome [16], and its presence in the venom may be due to leakage of a highly expressed housekeeping protein or a potential unknown role in prey incapacitation.

The lack of correlation between mRNA and protein levels is well-documented and has been attributed to differences in translational efficiency, codon usage/bias, and mRNA versus protein stability [44,7,6,2]. This relationship between mRNA and toxins has been previously demonstrated in venoms [24–29], and we found a significant difference between protein and mRNA levels, with the greatest discordance between CTLs, L-amino-acid oxidase, and PLA_2s. Rodrigues et al. [24] also found poor correlation for CTLs in *Bothropoides pauloensis* but, contrary to our findings, found similar mRNA and protein levels for PLA_2s and L-amino-acid oxidase. CTLs are known to form heterodimers and other complexes [45], and complex formation along with post-translational modifications and/or

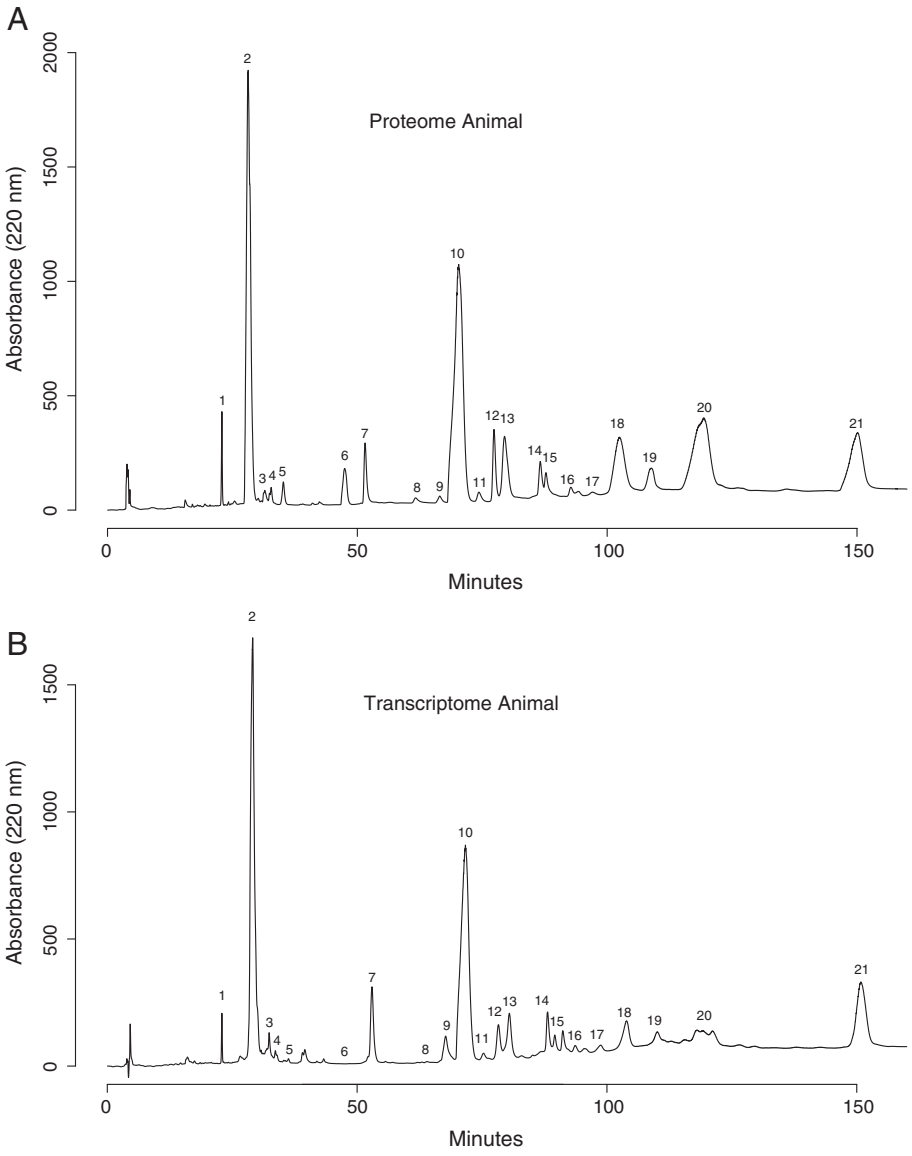

Fig. 2 – The reversed-phase high-performance liquid chromatography profiles of the *Crotalus adamanteus* proteome and transcriptome animals. A) We identified 21 major peaks in our reversed-phase high-performance liquid chromatography analysis of 100 μg of venom from the proteome animal, a juvenile female (72.4 cm snout-vent-length, 77.5 cm total length) from Leon County, FL. These fractions were collected and analyzed by means of N-terminal sequencing, mass spectrometry, and SDS-PAGE. B) Reversed-phase high-performance liquid chromatography reveals the similarity between the venoms of the proteome (A) and transcriptome (B) specimens. The transcriptome animal, a juvenile female (79.2 cm snout-vent-length, 84.4 cm total length) from Wakulla County, FL, lacked peak 6. This peak was present in the proteome animal but was not identified proteomically.

mRNA processing may explain the poor correlation for this particular class in both *C. adamanteus* and *B. pauloensis* [24], suggesting mRNA-protein correlations may be toxin-class specific.

The discrepancy between mRNA and protein abundances [24–29] may also be a function of sequencing depth and proteomic approach (i.e., fractionation prior to mass spectrometry analysis). Rodrigues et al. [24] failed to identify many of the protein products of low-abundance transcripts in the proteome, and argued that these "orphan molecules" may be remnant toxins (i.e., they had a past function in the venom but no longer play a relevant physiological role following envenomation and are thus expressed at extremely low levels and are essentially absent from the proteome). With the greater sequencing depth of Rokyta et al. [16] and the addition of LC/MS^{E} analysis of a whole venom sample, we identified many (but by no means all) low-abundance transcripts in the proteome of *C. adamanteus* including SVSP-11, the 70th most highly expressed transcript (4.46×10^{-3}% total reads, 0.013% toxin reads). However, this transcript was not identified by tandem mass spectrometry analysis of RP-HPLC fractions. Fractionation of whole venom is necessary to look at protein

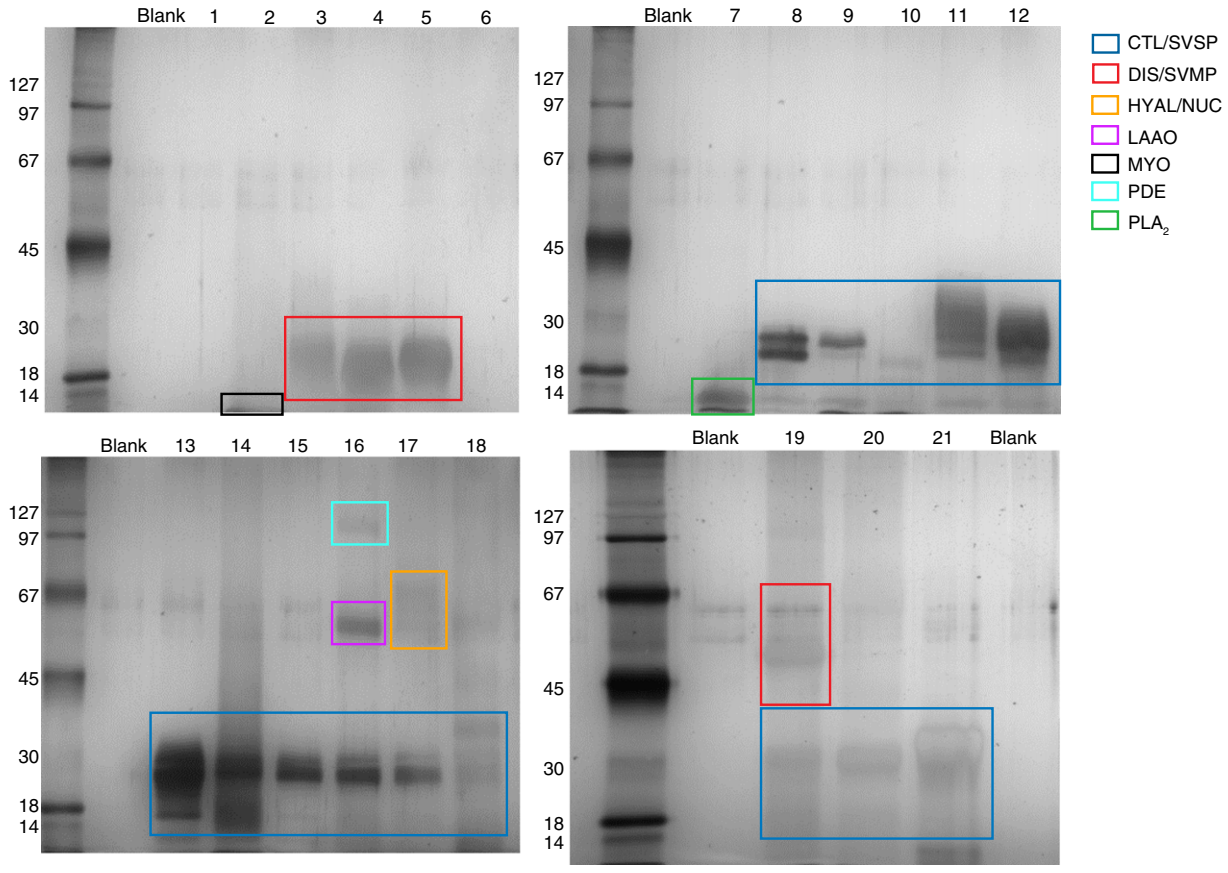

Fig. 3 – SDS-PAGE of the *Crotalus adamanteus* proteome animal. SDS-PAGE of each individual reversed-phase high-performance liquid chromatography fraction resulted in the identification of nine toxin families, including three families that were not detected by tandem mass spectrometry: nucleotidases and hyaluronidases in fraction 17 and phosphodiesterases in fraction 16. The two bands present between the 45–65 kDa markers are artifacts as they occur in the lane loaded with sample buffer. They are commonly found in silver staining procedures, a result of its high sensitivity. Toxin families were identified on the basis of relative molecular weight according to the transcriptome translation estimates. Abbreviations: CTL, C-type lectin; DIS, disintegrins; HYAL, hyaluronidase; LAAO, L-amino-acid oxidase; MYO, myotoxin; NUC, nucleotidase; PDE, Phosphodiesterase; PLA_2, Phospholipase A_2; SVMP, Snake venom metalloproteinase (types II and III); SVSP, Snake venom serine proteinase.

abundance and variation, but for purely characterizing venom composition, LC/MS^E may be more effective as it identified 15 more toxins.

The lack of correlation between mRNA and protein levels may also be due to our methods of quantification. Rokyta et al. [16] used the percentage of reads mapping to a particular transcript as a measure of its abundance, arguing that this proxy provides a measure of the investment made in expressing a specific transcript, which reflects the energetic cost of generating the subsequent protein. This is consistent with previous transcriptomic work with snake venom glands [46] and comparable to the methodology for measuring protein abundance used in this and previous studies [37,47]. However, it may be argued that average coverage may be more appropriate for measuring transcript abundance as it accounts for differences in transcript lengths, but even transcripts from the same gene family can vary in total length due to differences in the lengths of untranslated regions and open-reading frames as well as biases inherent with the assembly process. Although the Lambert–Beer Law has been demonstrated to be a useful measure of protein abundance, densitometry of SDS-PAGE gels following silver staining can be problematic. We chose silver staining for its high sensitivity as all proteomic analyses performed in this study used a single venom sample from one individual. Silver staining minimized the risk of low-abundance proteins going undetected however, unlike coomassie, silver stain does not stain all proteins equally and may have biased our abundance estimates (e.g., proteins that bind ions, such as SVMPs, are often negatively stained). Transcript and protein abundances were measured as percentages, and the amounts of each transcript or protein covaried with other transcripts or proteins within each molecular class (e.g., mRNA or protein), respectively. Therefore, neither mRNA nor protein quantities were independent values (e.g., as one toxin transcript increases, other transcripts must decrease). This generation of closed data causes several statistical issues [48–52], and makes the comparison of RNA and protein abundances problematic. Sequencing depth, proteomic resolution, method of quantification, and the type of data produced limit our ability

Table 3 – N-terminal sequencing of the reversed-phase high-performance liquid chromatography fractions. (KR) indicates the sequenced amino acid is either lysine (K) or arginine (R). X indicates an unidentified amino acid. Abbreviations: CTL, C-type lectin; CRISP, Cysteine-rich secretory protein; LAAO, L-amino-acid oxidase; MYO, Myotoxin; PLA_2, Phospholipase A_2; SVMP, Snake venom metalloproteinase (types II and III); SVSP, Snake venom serine proteinase.

Fraction #	% Absorbance (220 nm)	N-terminal sequence	Fragment sequence	blastp result	Transcriptome blast result
1	0.95	EI(WR)SDGDL		Unknown	SVMPIII-3, 7
2	22.31	YKRXHKKGGHHF		MYO	MYO
3	0.35	HSWVECESGVCC		SVMP	SVMPIII-3, 7
		EAGEECDCGSPR		SVMP	SVMPIII-3, 7
4	0.31	DIISPPVX		SVMP	SVMPIII-3, 7
		EAGEEXDX		SVMP	SVMPIII-3, 7
		SPPVXGNELL		SVMP	SVMPIII-1, 3, 7
5	0.67	DIISPPVCGNELLEA		SVMP	SVMPIII-3, 7
6	1.83	Blocked	(KR)KGXEPK	Unknown	Unknown
			(KR)EGGFXR	Unknown	Unknown
			(KR)FLLXPSR	Unknown	Unknown
			(KR)LGXEPLWK	Unknown	Unknown
7	1.95	SVDFDSESPRKKEIQ		CRISP	CRISP
8	0.33	VVGGDEXNINEHRSL		SVSP	SVSP-5, 11
9	0.33	SLVQFETLIMKVAKR		PLA_2	PLA_2-1, 2, 4–6
10	26.60	SLVQFETLIMKV		PLA_2	PLA_2-1, 2, 4–6
11	0.51	VIGGDEXNINEHRSL		SVSP	SVSP-7, 8
		SLVQFETLIMKVAKR		PLA_2	PLA_2-1, 2, 4–6
12	2.60	VIGGDEXNINEHRFL		SVSP	SVSP-1, 3, 6, 12–14
13	4.51	VIGGDEXNINEHRFL		SVSP	SVSP-1, 3, 6, 12–14
14	1.63	DCPSDWSSYEGHCYR		CTL	CTL-1, 6, 13
		DLKCPPTWSSTRQYC		CTL	CTL-3, 7
		VIGGDECNINEHRFL		SVSP	SVSP-1, 3, 6, 12–14
		KGISYIWIGLRV		CTL	CTL-4, 14, 15, 18
15	1.27	DCPSDWSSYEGHCYR		CTL	CTL-1, 6, 13
		DLKCPPTWSSTRQYC		CTL	CTL-3, 7
		VIGGDECNINEHRFL		SVSP	SVSP-1, 3, 6, 12–14
16	0.60	VIGGDEXNINEHRFL		SVSP	SVSP-1, 3, 6, 12–14
		DCPSDWSSYEXH		CTL	CTL-1, 6, 13
		KGISYIWIGLRV		CTL	CTL-4, 14, 15, 18
		AHDRNXLEE		LAAO	LAAO
17	0.21	VIGGDEXNIN		SVSP	SVSP-1, 3, 6–9, 12–14
		DXPSDWSSYE		CTL	CTL-1, 6, 13
		KGISYIWIGL		CTL	CTL-4, 14, 15, 18
18	6.63	Blocked	(KR)IFPCAPQDK	SVMP	SVMPIII-2, 8
			(KR)LFCVLGPTGNTISCQATSSQ	SVMP	SVMPIII-2d
19	1.78	Blocked	(KR)ATDLLR	SVMP	SVMPIII-4
			(KR)ACSNGQCVDVTT	SVMP	SVMPIII-4
20	14.28	HLNLNPEEQRYIELV	RYIELVIVA	SVMP	SVMPII-2, 5–7
			(KR)TWVYEIVNTLNE	SVMP	SVMPIII-3, 7
21	8.99	Blocked	(KR)GDWNNDICTGQSAECPN	SVMP	SVMPII-2, 3, 6, 7
			(KR)ETVLMNR	SVMP	SVMPII-1, 3, 4, 6, 8
Total % =	98.64				

to understand the relationship between RNA and protein abundances so severely that the claims of discordance [7,8,2] may be just as biased as the assessment of the correlation itself. Until these issues are remedied, comparisons between RNA and protein abundances should be viewed with caution.

Although our joint transcriptomic and proteomic approach constructed a genotype–phenotype map containing 40 loci for an evolutionarily significant phenotype, it is only a snap-shot of the venom system as a single animal was used for proteomic and transcriptomic analyses, respectively. Venom has been shown to exhibit geographic and ontogenetic variation [53–55,14,56,57], and biological replicates and time-series data are needed to map the relationship between venom genes and toxins throughout the life histories and across the ranges of venomous taxa.

Previous proteomic approaches have relied on public databases rather than a species-specific transcriptome to identify putative toxic proteins [19–23], effectively characterizing venom composition at a broad-scale but often failing to differentiate between members of large gene families. Several studies of the snake venom system have attempted to provide a more detailed description of venom complexity by using transcriptomics to generate a species-specific reference database, but relied on low-throughput sequencing approaches. This limited their ability to identify many of the low-abundance transcript products in the proteome [24–29], providing a biased assessment

Table 4 – Tandem mass spectrometry analysis of the reversed-phase high-performance liquid chromatography fractions. The asterisks seen in fractions 3–5 indicate that the percent coverage is greater than detected while the molecular weight is much lower as these fractions represent disintegrins, functional domains of snake venom metalloproteinases that are proteolytically cleaved posttranslationally to produce a free disintegrin. For clusters of proteins (e.g., fraction 3 containing SVMPIII-3 and SVMPIII-7), the number of spectra and unique peptides identified are shared among all members of the cluster. Percent coverage includes all peptide identifications, not just unique peptides, for the given fraction. Abbreviations: CTL, C-type lectin; CRISP, Cysteine-rich secretory protein; LAAO, L-amino-acid oxidase; MYO, Myotoxin; PLA_2, Phospholipase A_2; SVMP, Snake venom metalloproteinase (types II and III); SVSP, Snake venom serine proteinase.

Fraction #	Protein	Molecular weight (kDa)	Total spectra	Unique peptides	Coverage (%)
1	None	–	–	–	–
2	MYO	8	289	3	36
3	SVMPIII-3, 7	69*	737	6	17*
4	SVMPIII-3, 7	69*	883	9	22*
5	SVMPIII-3, 7	69*	1073	14	23*
6	None	–	–	–	–
7	CRISP	27	957	10	55
	PLA_2-5	15	16	2	20
8	CRISP	27	151	2	20
	SVSP-5, 8	28	477	4	26
9	None	–	–	–	–
10	PLA_2-1, 2, 4–6	15	3,083	8	73
11	SVSP-7, 14	27	31	4	38
12	SVSP-1, 9	29	2,440	2	52
13	SVSP-6, 12	29	682	5	51
14	SVSP-1, 9	29	384	3	36
	CTL-6	16	1,247	6	47
	SVSP-6, 12	29	70	3	24
	CTL-7	17	178	3	22
	CTL-4, 14, 15	17	37	2	30
15	SVSP-1, 9	29	322	3	36
	CTL-6	16	1,128	5	38
	SVSP-6, 12	29	76	2	24
	CTL-1, 13	17	1,048	5	41
	CTL-7	17	118	4	38
	CTL-4, 14, 15	17	28	2	30
16	SVSP-1, 9	29	164	4	42
	SVSP-6, 12	29	186	4	40
	CTL-4, 14, 15	17	38	2	28
	LAAO	59	272	9	28
17	SVSP-1, 9	29	385	4	42
	SVSP-6, 12	27	192	3	40
18	SVMPIII-2	68	1,544	10	20
	SVSP-6, 12	29	21	4	35
	CTL-4, 14, 15	17	50	3	34
19	SVMPIII-4	68	1,159	17	42
	CTL-4, 14, 15	17	262	3	40
20	SVMPIII-3, 7	69	1,447	14	36
	SVMPII-1, 4, 5	54	381	8	29
	SVSP-6, 12	29	45	2	22
	SVMPII-5	54	346	8	29
21	SVMPII-3, 6	55	2,696	13	32
	SVMPII-8	55	1,075	7	21
	CTL-4, 14, 15	17	99	3	40
	CTL-2, 20, 21	18	46	3	37
	SVSP-6, 12	29	11	2	22

of venom composition. We used the high-throughput transcriptome data of Rokyta et al. [16] as a reference database to correlate specific transcripts with specific proteins while retaining the context of whole venom, establishing a direct link from genotype to venom composition. We were able to verify the presence of the majority of the toxin transcript products in the venom, generating a genotype–phenotype map for an ecologically critical trait.

5. Conclusions

By using the transcriptome to understand the proteome, we were able to achieve locus-specific rather than gene-family resolution, providing a detailed characterization of the *C. adamanteus* venom system. Nanospray LC/MS^E coupled with the species-specific transcriptome proved effective at describing venom composition

Table 5 – Total protein identifications. An asterisk indicates protein identifications with unique peptide evidence and an X indicates protein identifications that lack a distinguishing peptide but share peptide evidence with multiple proteins. For unambiguous identifications the fraction in which the proteins were identified are listed in parentheses. Abbreviations: CTL, C-type lectin; CRISP, Cysteine-rich secretory protein; HYAL, Hyaluronidase; LAAO, L-amino-acid oxidase; MYO, Myotoxin (crotamine); NUC, Nucleotidase; PDE, Phosphodiesterase; PLA_2, Phospholipase A_2; PLB, Phospholipase B; RET-2, Reticulocalbin 2 EF-hand calcium-binding domain; SVMP, Snake venom metalloproteinase (types II and III); SVSP, Snake venom serine proteinase; VESP, Vespryn (ohanin-like).

Protein	LC/MS^E (whole venom)	MS/MS (gel filtration chromatography fractions)	MS/MS (RP-HPLC fractions)
CRISP	*	* (E, F, G)	* (7, 8)
CTL-1		X	X
CTL-2	X	X	X
CTL-3	X		
CTL-4	X	* (D)	X
CTL-6	*	* (F, G)	* (14, 15)
CTL-7	X		* (14, 15)
CTL-8	X	X	
CTL-9	*	* (C, G)	
CTL-12		* (D)	
CTL-13	*	X	X
CTL-14	X	* (D, E, F)	X
CTL-15	X		X
CTL-16	*	X	
CTL-18	X		
CTL-19	X		
CTL-20	X		X
CTL-21	X	X	X
HYAL-1	*		
LAAO	*	* (C, D)	* (16)
MYO	*		* (2)
NUC	*		
PDE	*		
PDE-3	*		
PLA_2-1	X	* (E)	X
PLA_2-2	X		X
PLA_2-4	X	X	X
PLA_2-5	X	X	* (7)
PLA_2-6	X		X
PLB	*		
RET-2	*		
SVMPII-1	X	X	X
SVMPII-2	X	* (A)	
SVMPII-3	X		X
SVMPII-4	X	X	X
SVMPII-5	X		* (20)
SVMPII-6	X		X
SVMPII-7	X	X	
SVMPII-8	*		* (21)
SVMPIII-1		* (B)	
SVMPIII-2	*	X	X
SVMPIII-4	*	* (D, G)	* (19)
SVMPIII-7	X	X	X
SVMPIII-8	*		
SVSP-1	*	* (G, F)	X
SVSP-2	*		
SVSP-4	*		
SVSP-5	*		X
SVSP-6	X	X	X
SVSP-7	X	* (D)	X
SVSP-8	*		X
SVSP-9	*		X
SVSP-11	*		
SVSP-12	*		X
SVSP-13	*	X	
SVSP-14	*		X
VESP	*		

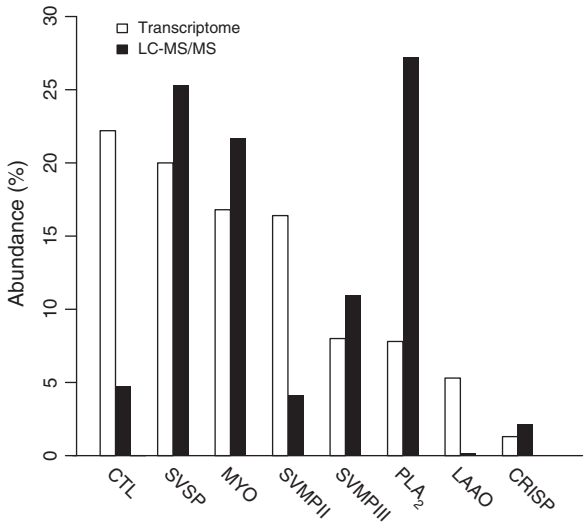

Fig. 4 – A comparison of transcriptome and proteome toxin abundances revealed differences between mRNA and protein levels in the venom of *Crotalus adamanteus*. Protein and mRNA levels were significantly different ($\chi^2 = 71.623$, $p < 0.0001$, $df = 7$), with the greatest discordance between C-type lectins, L-amino-acid oxidase, and phospholipases A_2. The lack of correlation between mRNA and protein levels may be due to differences in mRNA and protein stability, differences in the translational efficiency of specific transcripts due to codon bias or other factors, or may be an artifact of the method of quantification. Transcript abundances were determined by the percentage of reads mapping to a specific transcript, and the relative amounts of individual reversed-phase high-performance liquid chromatography peaks were determined by measuring the area under each peak relative to the total area of all protein peaks identified. If a peak contained multiple proteins, toxin class percentages were quantified by densitometry. Transcript and protein abundances were measured as percentages and, therefore, mRNA/protein quantities were not independent values. This generation of closed data causes several statistical issues, and makes the comparison of RNA and protein abundances problematic. Abbreviations: CTL, C-type lectin; CRISP, Cysteine-rich secretory protein; LAAO, L-amino-acid oxidase; MYO, myotoxin (crotamine); PLA_2, Phospholipase A_2; SVMP, Snake venom metalloproteinase (types II and III); SVSP, Snake venom serine proteinase.

and may be the most efficient way to identify the majority of proteins present in venoms. To understand venom variation and toxin abundances, fractionation followed by tandem mass spectrometry is necessary. The addition of a species-specific transcriptome enables the identification of gene expression variation among individuals and populations through the creation of a genotype–phenotype map, work that is currently underway. While N-terminal sequencing and SDS-PAGE provided independent confirmation of the mass spectrometry results, these analyses did so at much lower resolution and may not be cost-effective or necessary.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2013.11.001.

Acknowledgments

The authors thank Jordan Sirosky for help in acquiring specimens. The authors state they have no conflict of interests. Funding for this work was provided to DRR by the National Science Foundation (DEB 1145987).

REFERENCES

[1] Landry C, Rifkin S. The genotype–phenotype maps of systems biology and quantitative genetics: distinct but complementary. In: Soyer OS, editor. Evolutionary Systems Biology, Advances in Experimental Medicine and BiologyNew York, New York: Springer; 2012. p. 371–98.
[2] Alberch P. From genes to phenotype: dynamical systems and evolvability. Genetics 1991;84:5–11.
[3] Waddington C. Canalization of development and the inheritance of acquired characters. Nature 1942;150:563–5.
[4] Gause G. Problems of evolution. Trans Conn Acad Sci 1947;37:17–68.
[5] Bradshaw A. Evolutionary significance of phenotypic plasticity in plants. Adv Genet 1965;13:115–55.
[6] Ghazalpour A, Bennett B, Petyuk V, Orozco L, Hagopian R, Mungrue I, et al. Comparative analysis of proteome and transcriptome variation in mouse. PLoS Genet 2011;7(6): e1001393.
[7] Gygi S, Rochon Y, Franza R, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 1999;19(3):1720–30.
[8] Haider S, Pal R. Integrated analysis of transcriptomic and proteomic data. Curr Genomics 2013;14:91–110.
[9] Travisano M, Shaw R. Lost in the map. Evolution 2012;67(2):305–14.
[10] Fry BG, Wüster W. Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. Mol Biol Evol 2004;21(5):870–83.
[11] Fry BG. From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Res 2005;15:403–20.
[12] Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SFR, et al. Early evolution of the venom system in lizards and snakes. Nature 2006;439:584–8.
[13] Doley R, Kini RM. Protein complexes in snake venom. Cell Mol Life Sci 2009;66:2851–71.
[14] Mackessy SP. Venom composition in rattlesnakes: trends and biological significance. In: Hayes WK, Beaman KR, Cardwell MD, Bush SP, editors. The Biology of Rattlesnakes. Loma Linda, California: Loma Linda University Press; 2008. p. 495–510.
[15] Jansa SA, Voss RS. Adaptive evolution of the venom-targeted vWF protein in Opossums that eat pitvipers. PLoS One 2011;6(6):e20997.
[16] Rokyta DR, Lemmon AR, Margres MJ, Aronow K. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genomics 2012;13:312.
[17] Rokyta DR, Wray KP, Margres MJ. The genesis of an exceptionally deadly venom in the timber rattlesnake (*Crotalus horridus*) revealed through comparative venom-gland transcriptomics. BMC Genomics 2013;14:394.
[18] Margres MJ, Aronow K, Loyacano J, Rokyta DR. The venom-gland transcriptome of the eastern coral snake (*Micrurus fulvius*) reveals high venom complexity in the intragenomic evolution of venoms. BMC Genomics 2013;14:531.
[19] Madrigal M, Sanz L, Flores-Diaz M, Sasa M, Nunez V, Alape-Girón A. Snake venomics across genus *Lachesis*. Ontogenetic changes in the venom composition of *Lachesis stenophrys* and comparative proteomics of the venoms of adult *Lachesis melanocephala* and *Lachesis acrochorda*. J Proteomics 2012;77:280–97.
[20] Calvete JJ, Sanz L, Pérez A, Borges A, Vargas AM, Lomonte B, et al. Snake population venomics and antivenomics of *Bothrops atrox*: paedomorphism along its transamazonian dispersal and implications of geographic venom variability on snakebite management. J Proteomics 2011;74:510–27.
[21] Fernandez J, Alape-Girón A, Angulo Y, Sanz L, Gutierrez J, Calvete JJ, et al. Venomic and antivenomic analyses of the Central American coral snake, *Micrurus nigrocinctus* (Elapidae). J Proteome Res 2011;10:1816–27.
[22] Boldrini-França J, Corrêa-Netto C, Silva MMS, Rodrigues RS, Torre PDL, Pérez A, et al. Snake venomics and antivenomics of *Crotalus durissus* subspecies from Brazil: assessment of geographic variation and its implication on snakebite management. J Proteomics 2010;73:1758–76.
[23] Calvete JJ, Borges A, Segura Á, Flores-Díaz M, Alape-Girón A, Gutiérrez JM, et al. Snake venomics and antivenomics of *Bothrops colombiensis*, a medically important pitviper of the *Bothrops atrox-asper* complex endemic to Venezuela: Contributing to its taxonomy and snakebite management. J Proteomics 2009;72:227–40.
[24] Rodrigues RS, Boldrini-França J, Fonseca FPP, de la Torre P, Henrique-Silva F, Sanz L, et al. Combined snake venomics and venom gland transcriptome analysis of *Bothropoides pauloensis*. J Proteomics 2012;75:2707–20.
[25] Corrêa-Netto C, Junqueira-de-Azevedo IdLM, Silva D, Ho P, Leitão-de Araújo M, Alves M, et al. Snake venomics and venom gland transcriptomic analysis of Brazilian coral snakes, *Micrurus altirostris* and *M. corallinus*. J Proteomics 2011;74:1795–809.
[26] Sanz L, Escolano J, Ferritti M, Biscoglio MJ, Rivera E, Crescenti EJ, et al. Snake venomics of the South and Central American bushmasters. Comparison of the toxin composition of *Lachesis muta* gathered from proteomic versus transcriptomic analysis. J Proteomics 2008;71:46–60.
[27] Calvete JJ, Marcinkiewicz C, Sanz L. Snake Venomics of *Bitis gabonica gabonica*. Protein family composition, subunit organization of venom toxins, and characterization of dimeric disintegrins Bitisgabonin-1 and Bitisgabonin-2. J Proteome Res 2007;6:326–36.
[28] Wagstaff SC, Sanz L, Juárez P, Harrison RA, Calvete JJ. Combined snake venomics and venom gland transcriptomic analysis of the ocellated carpet viper, *Echis ocellatus*. J Proteomics 2009;71:609–23.
[29] Vaiyapuri S, Wagstaff S, Harrison R, Gibbins J, Hutchinson E. Evolutionary analysis of novel serine proteases in the venom gland transcriptome of *Bitis gabonica rhinoceros*. PLoS One 2011;6(6):e21532.

[30] Klauber LM. Rattlesnakes: their habits, life histories, and influence on mankind. Berkeley, California: University of California Press; 1997.
[31] Rokyta DR, Wray KP, Lemmon AR, Lemmon EM, Caudle SB. A high-throughput venom-gland transcriptome for the eastern diamondback rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes. Toxicon 2011;57:657–71.
[32] McCleary RJR, Heard DJ. Venom extraction from anesthetized Florida cottonmouths, *Agkistrodon piscivorus conanti*, using a portable nerve stimulator. Toxicon 2010;55:250–5.
[33] Choo K, Tan T, Ranganathan S. A comprehensive assessment of N-terminal signal peptides prediction methods. Asia Pacific Bioinformatics Network (APBioNet) Eighth International Conference on Bioinformatics (InCoB2009). Nucleic Acids ResSingapore: BMC Bioinformatics; 2009.
[34] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–95.
[35] Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2007;2(4):953–71.
[36] McNaught A, Wilkinson A. Compendium of chemical terminology: IUPAC recommendations, vol. 1669. Oxford, United Kingdom: Oxford Blackwell Science; 1997.
[37] Gibbs HL, Sanz L, Calvete JJ. Snake population venomics: proteomics-based analyses of individual variation reveals significant gene regulation effects on venom protein expression in *Sistrurus* rattlesnakes. J Mol Evol 2009;68:113–25.
[38] Bernheimer A, Linder R, Weinstein S, Kim KS. Isolation and characterization of a phospholipase B from venom of Collett's Snake, *Pseudechis colletti*. Toxicon 1987;25(5):547–54.
[39] Fry BG, Casewell NR, Wüster W, Vidal N, Young B, Jackson TNW. The structural and functional diversification of the Toxicofera reptile venom system. Toxicon 2012;60:434–48.
[40] Chatrath S, Chapeaurouge A, Lin Q, Lim T, Dunstan N, Mirtschin P, et al. Identification of a novel protein from the venom of a cryptic snake *Drysdalia coronoides* by a combined transcriptomics and proteomics approach. J Proteome Res 2011;10:739–50.
[41] Aird SD. Ophidian envenomation strategies and the role of purines. Toxicon 2002;40:335–93.
[42] Aird SD. The role of purine and pyrimidine nucleosides in snake venoms. In: Mackessy SP, editor. Handbook of Venoms and Toxins of Reptiles. Boca Raton, Florida: CRC Press; 2010. p. 393–419.
[43] Dhananjaya BL, Vishwanath BS, D'Souza CJM. Snake venom nucleases, nucleotidases, and phosphomonoesterases. In: Mackessy SP, editor. Handbook of Venoms and Toxins of Reptiles. Boca Raton, Florida: CRC Press; 2010. p. 155–71.
[44] Pigliucci M. Genotype–phenotype mapping and the end of the 'genes as a blueprint' metaphor. Phil Trans R Soc B 2010;365:557–66.
[45] Walker JR, Nagar B, Young NM, Hirama T, Rini JM. X-ray crystal structure of a galactose-specific C-type lectin possessing a novel decameric quaternary structure. Biochemistry 2004;43:3783–92.
[46] Pahari S, Mackessy SP, Kini RM. The venom gland transcriptome of the Desert Massasauga Rattlesnake (*Sistrurus catenatus edwardsii*): towards an understanding of venom composition among advanced snakes (Superfamily Colubroidea). BMC Mol Biol 2007;8:115.
[47] Calvete JJ, Fasoli E, Sanz L, Boschetti E, Righetti PG. Exploring the venom proteome of the western diamondback rattlesnake, *Crotalus atrox*, via snake venomics and combinatorial peptide ligand library approaches. J Proteome Res 2009;8:3055–67.
[48] Aitchison J. The statistical analysis of compositional data. London, United Kingdom: Chapman and Hall; 1986.
[49] Aitchison J, Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V. Logratio analysis and compositional distance. Math Geol 2000;32(3):271–5.
[50] Aitchison J, Egozcue JJ. Compositional data analysis: where are we and where should we be heading? Math Geol 2005;37(7):829–50.
[51] Pawlowsky-Glahn V, Buccianti A, editors. Compositional Data Analysis. Theory and Applications Chichester, United Kingdom: John Wiley & Sons; 2011.
[52] Templ M, Hron K, Filzmoser P. robCompositions: an R-package for robust statistical analysis of compositional data. In: Pawlowsky-Glahn V, Buccianti A, editors. Compositional Data Analysis, Theory and ApplicationsChichester, United Kingdom: John Wiley & Sons; 2011. p. 341–55.
[53] Mackessy SP. Venom ontogeny in the pacific rattlesnakes *Crotalus viridis helleri* and *C. v. oreganus*. Copeia 1988;1988:92–101.
[54] Guércio RAP, Shevchenko A, Schevchenko A, López-Lozano JL, Paba J, Sousa MV, et al. Ontogenetic variations in the venom proteome of the Amazonian snake *Bothrops atrox*. Proteome Sci 2006;4:11.
[55] Alape-Girón A, Sanz L, Escolano J, Flores-Díaz M, Madrigal M, Sasa M, et al. Snake venomics of the lancehead pitviper *Bothrops asper*: geographic, individual, and ontogenetic variations. J Proteome Res 2008;7:3556–71.
[56] Durban J, Perez A, Sanz L, Gomez A, Bonilla F, Rodriguez S, et al. Integrated "omics" profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, *Crotalus simus simus*. BMC Genomics 2012;14:234.
[57] Calvete JJ, Sanz L, Cid P, de la Torre P, Flores-Díaz M, Santos MCD, et al. Snake venomics 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 2010;9:528–44.