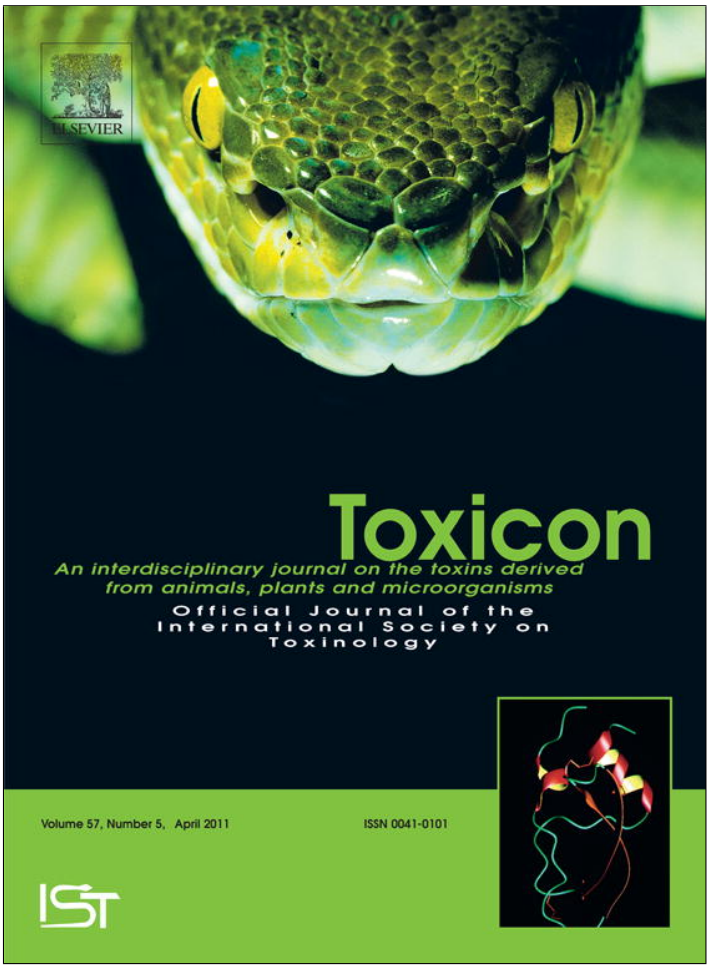

Toxicon 57 (2011) 657–671

Contents lists available at ScienceDirect

Toxicon

journal homepage: www.elsevier.com/locate/toxicon

A high-throughput venom-gland transcriptome for the Eastern Diamondback Rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes

Darin R. Rokyta [a,*], Kenneth P. Wray [a], Alan R. Lemmon [b], Emily Moriarty Lemmon [a], S. Brian Caudle [a]

[a] Department of Biological Science, Florida State University, Tallahassee, FL 32306-4295, USA
[b] Department of Scientific Computing, Florida State University, Tallahassee, FL 32306-4120, USA

ARTICLE INFO

Article history:
Received 24 November 2010
Received in revised form 5 January 2011
Accepted 10 January 2011
Available online 19 January 2011

Keywords:
Crotalus adamanteus
Eastern Diamondback Rattlesnake
Venom
Venom gland
Transcriptome
Positive selection

ABSTRACT

Despite causing considerable human mortality and morbidity, animal toxins represent a valuable source of pharmacologically active macromolecules, a unique system for studying molecular adaptation, and a powerful framework for examining structure-function relationships in proteins. Snake venoms are particularly useful in the latter regard as they consist primarily of a moderate number of proteins and peptides that have been found to belong to just a handful of protein families. As these proteins and peptides are produced in dedicated glands, transcriptome sequencing has proven to be an effective approach to identifying the expressed toxin genes. We generated a venom-gland transcriptome for the Eastern Diamondback Rattlesnake (*Crotalus adamanteus*) using Roche 454 sequencing technology. In the current work, we focus on transcripts encoding toxins. We identified 40 unique toxin transcripts, 30 of which have full-length coding sequences, and 10 have only partial coding sequences. These toxins account for 24% of the total sequencing reads. We found toxins from 11 previously described families of snake-venom toxins and have discovered two putative, previously undescribed toxin classes. The most diverse and highly expressed toxin classes in the *C. adamanteus* venom-gland transcriptome are the serine proteinases, metalloproteinases, and C-type lectins. The serine proteinases are the most abundant class, accounting for 35% of the toxin sequencing reads. Metalloproteinases are the most diverse; 11 different forms have been identified. Using our sequences and those available in public databases, we detected positive selection in seven of the eight toxin families for which sufficient sequences were available for the analysis. We find that the vast majority of the genes that contribute directly to this vertebrate trait show evidence for a role for positive selection in their evolutionary history.

1. Introduction

Venoms can be viewed as inverse biochemical fingerprints of the most fundamental physiological processes sustaining a prey animal's life. In their capacity to interfere with multifarious physiological, neurological, and hemostatic processes, animal toxins contain a wealth of information about their targets and have numerous applications in medical treatments or their design (Harvey et al., 1998; Ménez, 1998; Escoubas and King, 2009). From the perspective of evolutionary biology, snake venoms represent a tremendous opportunity to study a large, integrated system of proteins that contribute to a single, well-defined, ecologically critical phenotypic trait. Venoms function in procuring and digesting prey and in defense, and therefore

* Corresponding author. Tel.: +1 850 645 8812; fax: +1 850 645 8447.
E-mail address: drokyta@bio.fsu.edu (D.R. Rokyta).

0041-0101/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.toxicon.2011.01.008

they directly contribute to fitness. The availability of representative structures for the major toxin classes (see, e.g., Holland et al., 1990; Gomis-Rüth et al., 1993; Zhang et al., 1994; Kumasaka et al., 1996; Gong et al., 1998; Pawelek et al., 2000; Watanabe et al., 2003; Lou et al., 2005) could eventually enable molecular evolutionary study to rival the level of detail found in current research on viral evolution (e.g., Bull et al., 2000; Rokyta and Wichman, 2009), but for a complex trait in vertebrates. Toxin genes show accelerated evolution, exon shuffling, transcriptional splicing, and gene fusion (Pahari et al., 2007). Work on a handful of toxin gene families has demonstrated an important role for positive selection in their evolution (Kordiš and Gubenšek, 2000; Lynch, 2007; Gibbs and Rossiter, 2008), but determining the prevalence of this type of selection in general awaits a more complete study of variation across the large number of genes that contribute to venom, which will first require a complete characterization of all of the venom genes.

A number of research groups have sequenced portions of snake venom-gland transcriptomes to identify the genes contributing to venoms. These previous studies have relied on cloning of cDNA libraries and Sanger sequencing, generating important, but ultimately limited, data. More than half of the expressed sequence tags (ESTs) from these studies have, in most cases, been found to code for toxin genes (see, e.g., Junqueira-de Azevedo and Ho, 2002), and a large proportion of the remaining ESTs to code for genes involved in transcription and translation, cell regulation, and metabolism (Pahari et al., 2007; Leao et al., 2009). Previous work has focused on species native to South America (Junqueira-de Azevedo and Ho, 2002; Ching et al., 2006; Cidade et al., 2006; de Azevedo et al., 2006; Leao et al., 2009), Africa (Francischetti et al., 2004; Wagstaff and Harrison, 2006; Casewell et al., 2009), and China (Qinghua et al., 2006; Zhang et al., 2006). Two North American species (*Sistrurus catenatus* and *Agkistrodon piscivorus*) have been studied in this manner (Pahari et al., 2007; Jia et al., 2008). None of these previous studies used modern high-throughput sequencing technologies, and the majority of mRNAs identified were represented in the data by a single EST each (see, e.g., Qinghua et al., 2006; Leao et al., 2009), indicating low coverage and a high likelihood that many transcripts remained undetected or only partially sequenced. A comparison by Wagstaff and Harrison (2006) between this standard approach to transcriptomics and a proteomic analysis of venom revealed that the standard level of sequence coverage provides a far from complete characterization of snake venoms. The application of next-generation sequencing, though not without its own issues, should alleviate issues of low coverage and provide a more complete characterization of the genes contributing to snake venoms.

The Eastern Diamondback Rattlesnake (*Crotalus adamanteus*) is the largest member of the genus *Crotalus*, a group of New World pit vipers typified by the hollow segments on the end of the tail that form a rattle. The largest individual reliably reported measured 2.44 m, but more commonly, adults average between 1.2 and 1.5 m in length (Klauber, 1997). This species is restricted to the southeastern United States, where it historically occurred in seven states along the southeastern Coastal Plain (Conant and Collins, 1998). Currently, it is listed as endangered in North Carolina and has been essentially extirpated from Louisiana (Palmer and Braswell, 1995; Dundee and Rossman, 1996). Because of its large size, *C. adamanteus* preys primarily on small mammals and birds, seldom taking the various ectothermic prey items that many other rattlesnake species are known to consume. Various mouse and rat species, squirrels, and rabbits form the bulk of the diet, though ground-nesting birds, such as quail, are also routinely consumed (Klauber, 1997).

We present the first, to our knowledge, high-throughput venom-gland transcriptome for a snake species and the first transcriptomic characterization for a species of the genus *Crotalus*. In the work reported here, we focused on characterizing the most abundant toxin-encoding transcripts in the venom-gland transcriptome of *C. adamanteus*. In addition, we provide analyses of the molecular evolutionary forces responsible for the evolution of the identified toxins by testing for evidence of positive selection driving the sequence divergence between *C. adamanteus* and other species with representative sequences in public databases.

2. Materials and methods

2.1. Snake venom-gland preparation

We sequenced the venom-gland transcriptome of a single animal from Florida (Wakulla County). The specimen was an adult female weighing 392.9 g with a snout-to-vent length of 792 mm and a total length of 844 mm. To stimulate transcription in the venom glands, venom was extracted by electrostimulation under anesthesia (McCleary and Heard, 2010). The snake was anesthetized by propofol injection (10 mg/kg). After venom extraction, the animal was allowed to recover for four days so that transcription levels could be maximized (Rotenberg et al., 1971). The snake was euthanized by injection of sodium pentobarbitol (100 mg/kg), and its venom glands were subsequently removed. The above techniques were approved by the Florida State University Institutional Animal Care and Use Committee (IACUC) under protocol #0924. Isolation and purification of venom-gland RNA transcripts were performed with Trizol (Invitrogen), followed by ethanol precipitation and concentration. The RNA extraction procedure involved homogenizing the venom-gland tissue submerged in Trizol, adding 20% chloroform, and centrifugation in phase lock heavy gel tubes (5Prime), which separated the RNA molecules from DNA and other cellular debris. The isolated RNA was pelleted with isopropyl alcohol and washed with 75% ethanol. A secondary ethanol washing/precipitation step performed according to standard techniques removed any additional protein contaminants in the RNA.

2.2. Library construction and sequencing

From the total venom-gland RNA, polyadenylated transcripts were isolated and used for cDNA synthesis. First-strand cDNA synthesis was primed with *N*6 random

primers. Different 454 adapters were ligated to the 5′ and 3′ ends of the cDNA. The cDNA was then amplified with PCR (23 cycles). For sequencing, cDNA in the size range of 600–800 base pairs (including adapters) was eluted from a preparative agarose gel. The nonnormalized, noncloned cDNA library was sequenced with Roche 454 GS FLX technology with Titanium series chemistry (454/Roche, http://www.454.com) in half of a run. Sequencing was performed only from the 5′ ends of fragments.

2.3. Sequence assembly and analysis

We trimmed sequences to remove adapters and low-quality regions using the NGen module of the DNAStar Lasergene software suite. Transcript assembly was performed with NGen at the settings for high-stringency *de novo* transcriptome assembly. The nucleotide consensus sequences from each of the >24,000 contigs with ≥2 reads were searched with blastn against the NCBI nonhuman and nonmouse subset of the EST database (est_others, downloaded August 16, 2010). The sequences were also searched with blastx in all 6 reading frames to the nonredundant protein database (nr, downloaded September 2, 2010). For both searches we used blast version 2.2.23+ (Zhang et al., 2000) and an E-value threshold of 10^{-5}; all other settings remained as default.

We analyzed the blast results to determine which transcripts (i.e., contigs) could be identified as toxin genes. More specifically, we searched the subject description for each match for the following key words: phospholipase, serine, protease, proteinase, metalloproteinase, CRISP, toxin, thrombin, nuclease, nucleotidase, phosphomonoesterase, acetylcholinesterase, oxidase, hyaluronidase, 3ftx, three-finger, cysteine-rich, disintegrin, C-type lectin, and natriuretic. We identified as toxins all transcripts matching a subject whose description contained any one of these keywords (in either the blastn or blastx search). When possible, we assigned transcripts identified as toxins to specific toxin classes on the basis of the most specific keyword found among the matched subject descriptions. For cases in which the toxin classification for the DNA and protein searches conflicted (5 transcripts total), we assigned the transcript to the "Unidentified toxin" category. All other members of this category were transcripts for which "toxin" was the only keyword found in the search results.

Because of the high error rate inherent in 454 sequencing (Emrich et al., 2007), we focused our detailed analysis of transcripts on those whose contigs comprised at least 100 sequencing reads. The consensus sequence from each contig was searched against the NCBI nonredundant protein database with blastx. If no identification could be made on the basis of this search, an additional search of the NCBI nonredundant nucleotide database was performed with blastn. We then searched each of the putative toxin contigs for a substantial open reading frame, especially one with homology to known toxins. Conserved domains were identified with CD-Search of the NCBI Conserved Domain Database. In addition, we screened putative toxins for the presence of a secretion tag with SignalP (Bendtsen et al., 2004; Emanuelsson et al., 2007).

We checked for incorrect assemblies by pooling the constituent sequencing reads of the identified toxin contigs by toxin family and performing a new assembly using the SeqMan Pro module of the DNAStar Lasergene software suite. The resulting contigs were compared to those from the original assembly and used to verify those transcripts. Wary of the high error rate in 454 sequencing (Emrich et al., 2007), we were conservative in identifying a transcript as unique. We only considered coding sequences, and if the predicted protein sequences of two putative transcripts differed by two or fewer amino acid residues, they were considered to be the same. The remote possibility remains that allelic differences exceeded our criterion, causing us to label different alleles as different toxins.

To determine transcript abundances for the identified toxin genes, we used their consensus sequences as an assembly template in NGen. All of the original sequencing reads were assembled against these transcripts with NGen's high-stringency settings. The number of sequencing reads aligning with a given transcript was then used as an estimate of abundance. We calculated average coverage by dividing the combined length of all sequences forming the resulting contigs by the length of the contigs (Table 1).

2.4. Detecting selection

We identified sequences homologous to putative *C. adamanteus* toxins by blasting their coding sequences against the NCBI nonredundant nucleotide sequence database. We attempted to maintain a moderate degree of divergence amongst the sequences analyzed, while having enough sequences for sufficient power. We aimed for a maximum of approximately 10% nucleotide sequence divergence to minimize alignment difficulties and around 10–15 homologous sequences where possible. These values were adjusted as necessary for each analysis. Sequences were aligned with ClustalW (Thompson et al., 1994) on the basis of the amino-acid sequence with the MegAlign module of the DNAStar Lasergene software suite. Gaps and start and stop codons were excluded from all analyses. Model selection was performed with DT-ModSel (Minin et al., 2003). To estimate the maximum-likelihood phylogeny, we used PAUP*, version 4.0b10 (Swofford, 1998), and the iterative search strategy described by Rokyta et al. (2006). We made no effort to quantify nodal support for our estimated phylogenies as our interest was not in the relationships among our sequences or species but only in the parameters of the codon models described below and a likelihood-ratio test for positive selection. These parameter estimates and tests have been found to be insensitive to minor differences in tree topologies (Yang et al., 2000). Note that phylogeny estimation is unnecessary in cases where only three sequences were used in the analysis as only one unrooted topology is possible.

A likelihood-ratio test for positive selection was conducted with codeml from the PAML version 4.4 package (Yang, 1997, 2007) with the maximum-likelihood phylogeny estimated as described above. We tested for the presence of a class of sites experiencing positive selection. Our null model was the nearly-neutral (M1) model in

Table 1
Toxin transcript abundances.

Transcript	GenBank accession #	Sequences	Average coverage	% total sequences	% toxin sequences	% toxin by class
Serine proteinase 1	HQ414117	11597	2449	1.82	7.55	34.88
Serine proteinase 2	HQ414118	3070	677	0.48	2.00	
Serine proteinase 3	HQ414119	5928	677	0.93	3.86	
Serine proteinase 4	HQ414120	431	156	0.07	0.28	
Serine proteinase 5	HQ414121	177	34	0.03	0.11	
Serine proteinase 6	HQ414122	30784	6749	4.84	20.05	
Serine proteinase 7	HQ414123	855	167	0.13	0.56	
Serine proteinase 8	HQ414124	616	175	0.10	0.40	
Serine proteinase 9	HQ414125	111	35	0.02	0.07	
C-type lectin 1	HQ414089	9697	4769	1.53	6.29	22.39
C-type lectin 2	HQ414090	8545	3973	1.34	5.57	
C-type lectin 3	HQ414091	8249	3718	1.30	5.37	
C-type lectin 4	HQ414092	2279	951	0.36	1.48	
C-type lectin 5	HQ414093	2710	1515	0.43	1.77	
C-type lectin 6	HQ414094	863	366	0.14	0.56	
C-type lectin 7	HQ414095	946	512	0.15	0.62	
C-type lectin 8	HQ414096	806	338	0.13	0.53	
C-type lectin 9	HQ414097	306	119	0.05	0.20	
Metalloproteinase 1	HQ414106	7231	1136	1.14	4.71	20.74
Metalloproteinase 2	HQ414107	5029	735	0.79	3.28	
Metalloproteinase 3	HQ414108	6280	783	0.99	4.09	
Metalloproteinase 4	HQ414109	3863	501	0.61	2.52	
Metalloproteinase 5	HQ414110	753	116	0.12	0.49	
Metalloproteinase 6	HQ414111	1866	215	0.29	1.22	
Metalloproteinase 7	HQ414112	823	120	0.13	0.54	
Metalloproteinase 8	HQ414113	609	153	0.10	0.40	
Metalloproteinase 9	HQ414114	933	184	0.15	0.61	
Metalloproteinase 10	HQ414115	317	60	0.05	0.21	
Metalloproteinase 11	HQ414116	4105	683	0.65	2.67	
Phospholipase A_2	HQ414104	13233	6771	2.08	8.62	8.62
L-amino acid oxidase	HQ414099	7795	1112	1.23	5.08	5.08
Cysteine-rich secretory protein	HQ414088	6917	1611	1.09	4.51	4.51
Vespryn	HQ414126	1827	418	0.29	1.19	1.19
Phosphodiesterase 1	HQ414102	901	140	0.14	0.59	1.01
Phosphodiesterase 2	HQ414103	643	139	0.10	0.42	
Nucleotidase	HQ414101	859	143	0.14	0.56	0.56
Myotoxin	HQ414100	452	131	0.07	0.29	0.29
Epidermal growth factor	HQ414087	412	93	0.06	0.27	0.27
Phospholipase B	HQ414105	350	72	0.06	0.23	0.23
Hyaluronidase	HQ414098	345	132	0.05	0.22	0.22
% identified toxin reads = 24.16						

codeml, which allows for a class of sites evolving neutrally ($dN/dS = 1$) and another class with its dN/dS estimated from the data but constrained to be below 1. The alternative model, positive selection (M2), adds a third class with $dN/dS > 1$. To test for positive selection, we compared negative twice the difference in log likelihoods between the models to a χ^2 distribution with 2 degrees of freedom. We further confirmed our results by performing a similar test comparing models M7 (Beta) and M8 (Beta with positive selection), again using a χ^2 distribution with 2 degrees of freedom (Yang and Swanson, 2002). The results of these analyses (not shown) were virtually identical to those we present. To estimate an overall dN/dS, we used the M0 model, which fits a single ratio for all sites and branches, thus averaging dN/dS across sites and time. This ratio has been found to be a poor (i.e., low-power) indicator of positive selection (Crandall et al., 1999), as positive selection usually only affects some subset of sites, but it gives an overall view of the evolutionary forces acting across the entire coding sequence over its history.

2.5. Accession numbers

We have deposited the 40 toxin-encoding transcripts discussed below in the GenBank database under accession numbers HQ414087–HQ414126 (Table 1). The species names and accession numbers for the sequences used to test for selection (other than those we generated) are provided in the Appendix.

3. Results

3.1. Sequencing and assembly results

Our half run on the Roche GS FLX generated 635,484 reads of average length 300 nucleotides after trimming of adapters and low-quality sequence. Of these, 82,621 were singletons, and the remaining 552,863 were assembled into 24,773 contigs of average length 513 nucleotides and comprising on average 22 sequencing reads. We performed an in-depth analysis of those contigs with the highest

coverages and identified as putative toxins. These results are described in the next section. The approach described there was conservative in identifying contigs as toxins and in particular as unique toxins. Here we will describe an analysis that was designed to give a more complete overview of the distribution of toxins in the *C. adamanteus* venom-gland transcriptome. Our assembly protocol was set to a high stringency (see Materials and Methods), so for present purposes we will treat each contig as a unique transcript. Although doing so will obviously lead to an overestimate of the true number of transcripts because of misassembly, error bias leading to clustering, or alternative splicing, this procedure is meant to provide a balancing perspective to the conservative results below.

The 24,773 putative *C. adamanteus* venom-gland transcripts were identified as nontoxins or toxins and to specific toxin classes if applicable, with blastx and blastn searches against the nonredundant NCBI protein database and the nonhuman, nonmouse subset of the EST database. Significant matches were searched for a set of keywords that would identify the transcript as encoding a toxin and would broadly identify the toxin class. Any transcript without a significant match was classified as a nontoxin. We then analyzed these data in two different ways. To get a view of the relative diversity of transcripts in the toxin classes, we examined numbers of transcripts (i.e., contigs). To look at relative expression levels, we examined the numbers of reads (i.e., individual sequences) making up the identified contigs. The results are summarized in Fig. 1. More information on the toxin classes is provided in the next section.

In numbers, the *C. adamanteus* venom-gland transcriptome is dominated by nontoxin transcripts (Fig. 1). Of the 24,773 putative transcripts, 89% were nontoxin. Of those identified as toxins, serine proteinases were the most abundant at 25%, and 16% of the toxin transcripts were unidentified, meaning their best matches were designated toxins, but no other keyword matches were found. The next most abundant classes were the nucleases (14%; these include phosphodiesterases, nucleotidases, etc.), C-type lectins (13%), and the metalloproteinases (11%). The abundances of the remaining classes are illustrated in Fig. 1.

Of the total 552,863 reads assembled into contigs, 46% were toxin and 54% nontoxin. This greater proportion of toxins among reads than among transcripts illustrates the overexpression of toxins in the venom gland. Of the 253,968 reads identified as toxins, 28% were serine proteinases, 17% were C-type lectin, 16% were unidentified, and 15% were L-amino acid oxidases. The abundances of the remaining classes are illustrated in Fig. 1.

3.2. Toxin transcripts

Among the 456 contigs with at least 100 sequences, we identified 90 with significant blast matches ($E < 10^{-4}$) to known toxins. Of these, 69 had clear open reading frames,

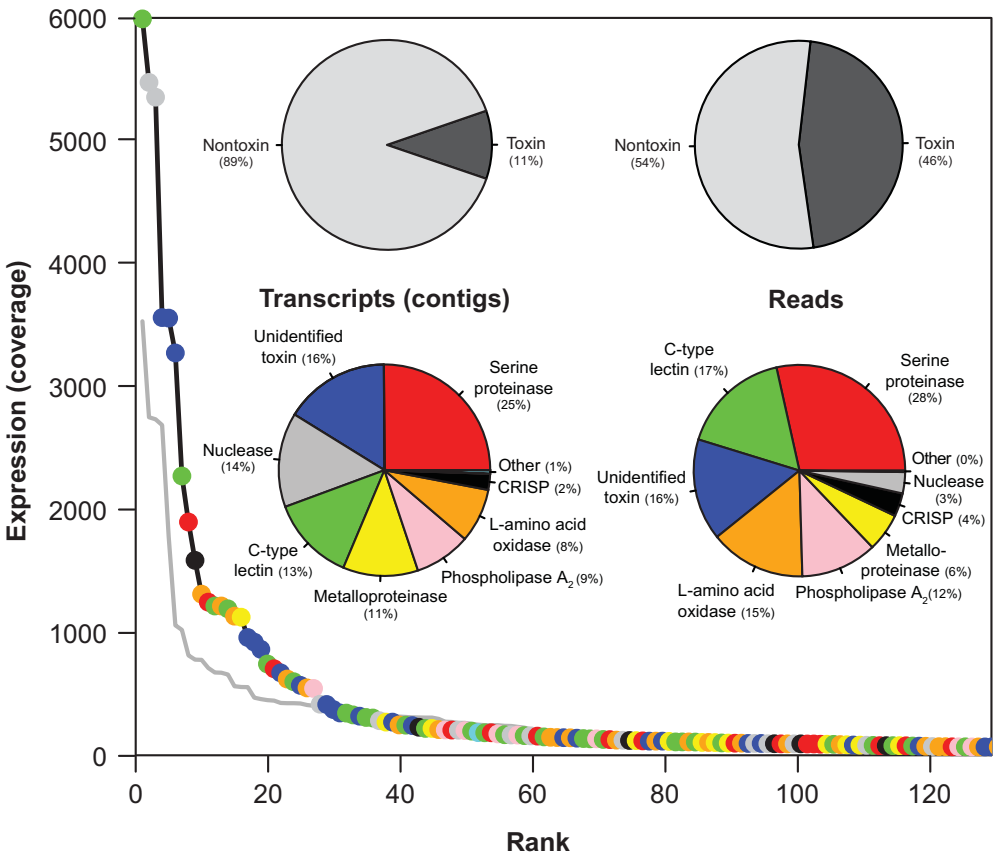

Fig. 1. Abundance of observed transcripts. Pie graphs show the relative abundances of toxin and nontoxin transcripts (left) and reads (right). Curves show the predicted relative expression level for toxin transcripts (black curve) and nontoxin transcripts (gray curve), with transcripts ranked from most (rank = 1) to least abundant. Colors of points on curves and sections of the pie graphs indicate the toxin classes to which the toxins were assigned.

and 40 of these were clearly unique genes. These 40 transcripts are grouped into 11 known classes of toxins and two putative, previously undescribed classes. These 40 transcripts accounted for 24% of the total sequencing reads (Table 1). Thirty of these had complete coding sequences, and 10 had only partial coding sequences. The partial sequences were all 5′ truncated, probably as an artifact of our cDNA library construction and sequencing protocol (see Discussion).

3.2.1. Serine proteinases

Snake-venom serine proteinases (SVSPs) belong to the trypsin family S1 of clan SA and are defined by their conserved catalytic mechanism involving a highly reactive serine residue. SVSPs are known to affect a diversity of substrates that are involved in the coagulation, fibrinolytic, and kallikrein-kinin systems, as well as affecting platelets (Serrano and Maroun, 2005; Phillips et al., 2010). Their effects are the result of the proteolytic degradation of blood components, resulting in an imbalance of the hemostatic system of prey. We identified nine unique SVSPs in the *C. adamanteus* venom-gland transcriptome. Six of these, SVSPs 1–6, had complete coding sequences, and three were partial sequences. The partial sequences were all 5′ truncated. All of the complete coding sequences encode a 258–271 amino-acid precursor with an 18 amino-acid residue signal peptide as predicted by SignalP and a 223–227 amino-acid trypsin-like serine proteinase domain. The nine identified SVSPs accounted for 34.88% of the identified toxin sequencing reads and 8.42% of all reads (Table 1). Individual abundances ranged from 4.84% to 0.02% of the total sequencing reads or 20.05% to 0.07% of the identified toxin reads. In terms of the number of sequencing reads, the SVSPs were by far the most abundant toxin class present in the *C. adamanteus* venom-gland transcriptome.

3.2.2. C-type lectins

We identified nine unique, full-length C-type lectins (CTLs) in the *C. adamanteus* venom-gland transcriptome and labeled them CTL 1–9. In total, they accounted for 22.39% of the identified toxin reads and 5.43% of the total sequencing reads and were therefore the second most abundant class of toxins. Individual abundances ranged from 1.53% to 0.05% of the total reads or 6.29% to 0.20% of the identified toxin reads. The encoded precursor proteins ranged from 142 to 158 amino acids in length, each with a 23–24 amino-acid signal peptide and a 106–119 amino-acid C-type lectin conserved domain. Most snake-venom CTLs exist as heterodimers of paralogous proteins, and these dimers can exist as monomers or oligomers (Du and Clemetson, 2010). More complex arrangements have also been documented. For example, a toxin from *C. atrox* was found be composed of 10 CTLs arranged into two pentamers covalently bound together in a back-to-back, staggered arrangement (Walker et al., 2004). With only transcriptomic data, we could not identify participants in these quaternary structural arrangements. Snake-venom CTLs are nonenzymatic venom components, and their primary pharmacological effect is the disruption of hemostasis. They are known to attack a diverse array of plasma components and blood cells (Du and Clemetson, 2010).

3.2.3. Metalloproteinases

Snake-venom metalloproteinases (SVMPs) are associated with local and systemic hemorrhaging and require Zn^{2+} for catalytic activity (Fox and Serrano, 2010). These enzymes have been found to be myonecrotic, hemorrhagic, fibrinolytic, and inhibitors of platelet aggregation. They are also known to participate in the degradation of the extracellular matrix and contribute to both prey immobilization and digestion (Gutiérrez et al., 2010). We identified 11 unique SVMPs in the *C. adamanteus* venom-gland transcriptome, which we labeled SVMP 1–11. Four of these (SVMP 8–11) are 5′-truncated transcripts with only part of their coding sequences. Overall, the SVMPs account for 20.74% of the identified toxin reads and 5.02% of the total sequencing reads (Table 1). Individual abundances range from 1.14% to 0.05% of the total sequences and 4.71% to 0.21% of the identified toxin sequences. SVMPs are the third most abundant class of toxins identified but also the most diverse in terms of the number of unique types identified. SVMPs 1–3 are all type II as defined by Fox and Serrano (2005) and are therefore composed of a metalloproteinase domain and a disintegrin domain. SVMPs 4–11 are all type III and therefore composed of a metalloproteinase domain, a disintegrin domain, and a cystein-rich domain (Fox and Serrano, 2005, 2010; Gutiérrez et al., 2010). All of the full-length transcripts have a predicted 20 amino-acid signal peptide and a conserved M12B propeptide region.

3.2.4. Phospholipase A_2

We identified a single phospholipase A_2 (PLA_2) transcript in the *C. adamanteus* venom-gland transcriptome, where it is represented by 2.08% of the total sequencing reads and 8.62% of the identified toxin sequences (Table 1). It encodes a precursor protein with 138 amino acids with a 16 amino-acid signal peptide. It belongs to Group II and is a D49 enzyme. Its closest match in the nucleotide database is a PLA_2 transcript from *C. horridus* (GQ168368). These two sequences are 97.1% identical at the amino acid level. PLA_2 enzymes are esterolytic and hydrolyze glycerophospholipids, releasing lysophospholipids and free fatty acids. PLA_2s are known to evince a wide array of pharmacological effects, including presynaptic and postsynaptic neurotoxicity, myotoxicity, cardiotoxicity, and anticoagulant effects. These effects may be independent of or dependent on the enzymatic activity of the protein (Kini, 2003; Doley et al., 2010).

3.2.5. L-amino acid oxidase

Snake-venom L-amino acid oxidases (LAAOs) are flavoproteins generally consisting of two 57–68 kDa identical subunits (Tan and Fung, 2010). They are highly specific for L-amino acids, and hydrophobic amino acids are their best substrates. Snake-venom LAAOs are known to produce a wide array of effects, including platelet aggregation inhibition, platelet aggregation induction, apoptosis, and edema. They have also been noted to be antibacterial and antiviral (Tan and Fung, 2010). These effects appear to be mediated by the H_2O_2 released during the oxidation reaction. The enzymes are thought to attack specific cell types and to generate high local concentrations of H_2O_2 (Tan and Fung, 2010). We found a single LAAO transcript with a full

coding sequence. It was represented by 1.23% of the total reads and 5.08% of the identified toxin reads. The precursor protein is 516 amino acids in length and has an 18 amino-acid signal peptide. The predicted mature protein has a conserved amino oxidase domain. The coding sequence is identical in amino-acid sequence to a *Crotalus atrox* LAAO (AF093248) and 99.9% identical at the nucleotide level. Interestingly, Raibekas and Massey (1998) sequenced an LAAO from *C. adamanteus* that differs at three amino acids from our sequence and is 99.1% identical at the nucleotide level. The existence of multiple versions of LAAOs has been previously noted in *C. adamanteus* (Hayes and Wellner, 1969).

3.2.6. Cysteine-rich secretory protein

We found a single transcript encoding a cysteine-rich secretory protein (CRISP) in the *C. adamanteus* transcriptome. It accounted for 1.09% of the total sequences and 4.51% of the identified toxin sequences. The most similar homolog in the nucleotide database is catrin from *C. atrox* (Yamazaki et al., 2003), which is 94.6% identical to our sequence at the amino-acid level and 97% identical at the nucleotide level. The *C. adamanteus* transcript encodes a 239 amino-acid precursor with the N-terminal 19 amino acids predicted to serve as a signal peptide by SignalP. It has two conserved domains: an SCP-like extracellular protein domain and a CRISP domain. Snake-venom CRISPs are known to inhibit smooth muscle contraction (Yamazaki et al., 2003; Yamazaki and Morita, 2004).

3.2.7. Vespryn

One transcript was found to be homologous with ohanin from *Ophiophagus hannah* (Pung et al., 2005, 2006) and an ohanin-like transcript from *Lachesis muta* (de Azevedo et al., 2006). The latter is the only previously described putative member of this toxin class from a viperid. The *C. adamanteus* sequence shares 96.3% amino-acid and 98.0% nucleotide sequence identity with the *L. muta* sequence. In addition, it has a four amino-acid insertion relative to the *L. muta* sequence near the N terminus. The *C. adamanteus* sequence shares 78.0% amino acid and 87.3% nucleotide similarity with the sequence of ohanin but has 32 amino acids at the N terminus that are not present in ohanin. *L. muta* has a similar insertion relative to ohanin. The *C. adamanteus* homolog is 222 amino acids in length and has two conserved domains similar to those of ohanin: a PRY domain and an SPRY domain. Pung et al. (2005) isolated the first member of this toxin class, ohanin, in the King Cobra (*O. hannah*). It was found to increase pain sensitivity (induce hyperalgesia) and induce temporary hypolocomotion in mice. Ohanin appears to lack either pre- or postsynaptic neurotoxicity, induces no hemorrhage or necrosis, and is nonlethal. Pung et al. (2006) proposed naming this class of toxins vespryns, and we follow this suggestion. This *C. adamanteus* vespryn transcript accounts for 0.29% of the total sequences and 1.19% of the identified toxin sequencing reads (Table 1).

3.2.8. Phosphodiesterases

We found two distinct phosphodiesterase (PDE) transcripts in the *C. adamanteus* transcriptome. The first, which we denote *C. adamanteus* PDE 1, is full length, and the second, *C. adamanteus* PDE 2, is 5′ truncated. Together, these two transcripts account for 0.24% of all reads and 1.01% of the toxin sequencing reads. *C. adamanteus* PDE 1 is the first full-length toxin PDE in the NCBI database. PDE 1 is 851 amino acids in length and has a 23 amino-acid signal peptide. It represents 0.14% of the total sequencing reads and 0.59% of the toxin reads. Its closest match in the nucleotide database, DQ464266, is a partial phosphodiesterase transcript from the venom gland of *S. catenatus* (Pahari et al., 2007). In the region of overlap, these two sequences are 90% identical at the amino acid level. It has four conserved domains, two of which are Somatomedin-B domains. Somatomedin-B is a protease-inhibiting peptide that is proteolytically excised from vitronectin. The third domain is a 325 amino-acid type I phosphodiesterase/nucleotide pyrophosphatase domain. Enzymes possessing these domains catalyze the cleavage of phosphodiester and phosphosulfate bonds in NAD, deoxynucleotides, and nucleotide sugars. The fourth domain is a 258 amino-acid DNA/RNA nonspecific endonuclease domain. *C. adamanteus* PDE 2 has diverged sufficiently from PDE 1 that alignment of the two is unreliable. Our partial sequence includes the 477 C-terminal amino acids and a conserved DNA/RNA nonspecific endonuclease domain. It represents 0.10% of the total sequencing reads and 0.42% of the identified toxin reads. Nuclease activity is ubiquitous in snake venoms, and the effects of the identified PDEs are probably caused by the liberation of adenosine, a multitoxin, from nucleic acids (Aird, 2002; Dhananjaya et al., 2010;). PDEs catalyze the hydrolysis of phosphodiester bonds resulting in the formation of 5′ mononucleotides, which can subsequently be acted upon by nucleotidases (Aird, 2002). They are known to hydrolyze double-stranded and single-stranded DNA, rRNA, and tRNAs.

3.2.9. Nucleotidase

We found a single partial transcript encoding a nucleotidase, which we will refer to as *C. adamanteus* nucleotidase. It represents 0.14% of the total sequencing reads and 0.56% of the identified toxin reads. Its closest match in the database is an ecto-5′ nucleotidase from *Gloydius blomhoffi*. Our partial sequence contains two conserved domains: a metallophosphatase domain and a nucleotidase domain. Nucleotidases hydrolyze the phosphate at the 5′ carbon of nucleotides. The primary role of nucleotidases in venom has been hypothesized to be in the production of adenosine, which has a variety of toxic effects (Aird, 2002, 2010; Dhananjaya et al., 2010).

3.2.10. Myotoxin

One myotoxin transcript was found in the *C. adamanteus* transcriptome, which we have called *C. adamanteus* myotoxin. The encoded precursor protein is 63 amino-acid residues long, and the first 33 amino acids are predicted to be a signal peptide by SignalP. This transcript represents 0.07% of the total sequences and 0.29% of the toxin sequences. Its predicted sequence is nearly identical to the partial amino-acid sequence of the CAM toxin from *C. adamanteus* (Samejima et al., 1991). Its closest nucleotide match was crotamine from *C. durissus terrificus*. They share

92.2% amino acid and 96.4% nucleotide identity although the *C. adamanteus* sequence is two amino acids shorter at the C terminus. Crotamine is a nonenzymatic myonecrotic polypeptide that acts by reducing resting membrane potential and increasing membrane conductance (Rádis-Baptista et al., 1999). It causes skeletal muscle spasms and spastic paralysis of hind limbs in mice (Oguiura et al., 2005).

3.2.11. Hyaluronidase

We found a single transcript encoding a hyaluronidase in the *C. adamanteus* transcriptome. This transcript, which we have named *C. adamanteus* hyaluronidase, is severely 5′ truncated, encoding only 111 amino acids, despite having 345 matching sequencing reads (0.05% of the total) and 132-fold coverage on average (Table 1). It represents 0.22% of the identified toxin reads. Hyaluronidase activity is known from a variety of viperids, including a close relative of *C. adamanteus*, *C. atrox* (Harrison et al., 2007). Hyaluronidase contributes to the degradation of the extracellular matrix by fragmenting hyaluronic acid. It is generally considered to be a nontoxic "spreading factor" that allows a more rapid diffusion of toxins, but it may contribute to morbidity at the site of the bite (Kemparaju et al., 2010).

3.2.12. Putative toxins

We found two transcripts in the transcriptome that were present in high concentration (as evidenced by their coverage) and that had clear signal peptides but that did not belong to any known toxin family. Given that they are highly expressed genes in the venom gland and are predicted to be secreted from the venom-gland cells, they are probably venom components. Their toxicity, if any, and the nature of their effects remain to be determined. The first new putative toxin is a transcript encoding a gene with a phospholipase B domain. This transcript, which we call *C. adamanteus* phospholipase B (PLB), encodes a precursor 553 amino acids in length. It has a 36 amino-acid signal peptide as determined by SignalP and a 526 amino-acid phospholipase B domain. Proteins with such a domain are known to catalyze the hydrolytic cleavage of both acylester bonds in glycerophospholipids. This transcript was represented by 0.06% of the total sequences (350 sequences) and 0.23% of the identified toxin sequencing reads and 72-fold coverage on average. The second previously undescribed toxin transcript encodes a 364 amino-acid precursor. The N-terminal 29 amino acids are predicted to be a signal peptide by SignalP. Three conserved domains have been detected. The first is a functionally uncharacterized 97 amino-acid DUF3456 domain. The second is a cysteine-rich, Furin-like repeat, and the third is a calcium-binding EGF-like domain. We refer to this gene as *C. adamanteus* epidermal growth factor (EGF). Of the total sequencing reads, 0.06% align to this transcript for 93-fold coverage on average. It accounts for 0.27% of the identified toxin reads.

3.3. Selection

The number of homologous sequences in the database was insufficient for analysis of all classes of toxins found in the *C. adamanteus* venom-gland transcriptome. Our two new putative classes lack sequenced homologs. Both phosphodiesterases lack full-length homologs and are too divergent to use together. Our test statistic, negative twice the difference in log likelihoods, is denoted by Λ, and dN/dS ratios are denoted by ω. Estimated codon-site class frequencies are denoted by p. These represent the estimated proportion of codons falling into each rate class (e.g., purifying selection, neutral, or positive selection). Likelihood-ratio tests of the type we are using can suffer from reduced power with excessive sequence divergence (Anisimova et al., 2001), and sequence alignments can become unreliable as divergence between the sequences increases. Therefore, in some cases in which the divergence between *C. adamanteus* sequences in the same toxin family was high, we ran separate analyses. In all, we conducted 20 tests for positive selection using codeml, so, by Bonferroni correction ($P \leq 0.05/20$), we accepted $P = 0.0025$ as indicating a 5% significance level. See Table 2 for parameter estimates and uncorrected P values. We refer to individual tests by abbreviations for the toxin classes (e.g., SVMP for the snake-venom metalloproteinases) followed by, if necessary, parentheses enclosing the numbers of the *C. adamanteus* sequences included. For example, SVSP(2,5) refers to the test based around *C. adamanteus* snake-venom serine proteinases 2 and 5. The models selected by DT-ModSel for maximum-likelihood phylogeny estimation are listed in Table 2, and the identities and GenBank accession numbers of the homologous sequences used in each analysis are provided in the Appendix.

3.3.1. Serine proteinases

The six *C. adamanteus* SVSPs with complete sequences are all significantly more than 10% divergent from one another except for SVSPs 2 and 5. These two were therefore combined into the same analysis, and each of the remaining four was analyzed separately, for a total of five tests of positive selection for the SVSPs. Note that the sequences used for SVSP(6) and SVSP(2,5) overlapped slightly; one sequence was used in both analyses. This slight lack of independence should not be problematic given the number of other sequences involved and the low P values obtained. In four of these five tests, we could easily reject the null, nearly-neutral model (M1) in favor of the positive-selection model (M2) at a significance level of <0.001 after a Bonferroni correction (Table 2). Considering only the analyses with significant results, we found that, in the SVSPs, 8–30% of the codon sites appeared to have experienced positive selection with $3.47 \leq \omega \leq 5.69$. Notably, the remaining test, SVSP(3), involved the fewest sequences; our inability to reject the null hypothesis may have resulted from insufficient power. Fitting a single dN/dS ratio to all sites and branches with the single-ratio model produced $0.73 \leq \omega \leq 1.28$. In three of the five data sets, we found $\omega > 1$ for the whole protein (i.e., averaged over sites).

3.3.2. C-type lectins

The nine C-type lectins were all sufficiently different from one another to warrant independent analyses. *C. adamanteus* CTL 5 did not have any matches with >80% nucleotide sequence identity, and *C. adamanteus* CTL 6 had only one good match. Therefore, no selection analyses were

Table 2
Summary of codeml tests for positive selection.

Toxin	n/Model	M1: Nearly neutral			−ln*L*	M2: Positive selection				−ln*L*	M0: ω	Λ	*P*[a]
SVSP(1)	11	*p*:	0.31	0.69	2106.05	*p*:	0.26	0.57	0.17	2092.05	1.28	28.00	8.3×10^{-7}*
	K80 + I	ω:	0.00	1.00		ω:	0.00	1.00	4.97				
SVSP(2,5)	12	*p*:	0.54	0.46	3176.63	*p*:	0.51	0.30	0.19	3133.28	1.08	86.70	1.5×10^{-19}*
	K80 + G	ω:	0.08	1.00		ω:	0.17	1.00	4.43				
SVSP(3)	8	*p*:	0.40	0.60	1843.05	*p*:	0.37	0.57	0.06	1839.36	0.73	7.38	0.025
	K80 + G	ω:	0.00	1.00		ω:	0.00	1.00	5.45				
SVSP(4)	12	*p*:	0.39	0.61	2551.51	*p*:	0.66	0.04	0.30	2529.36	1.11	44.30	2.4×10^{-10}*
	K80 + G	ω:	0.01	1.00		ω:	0.32	1.00	3.47				
SVSP(6)	9	*p*:	0.51	0.49	2424.96	*p*:	0.44	0.48	0.08	2402.10	0.79	45.72	1.2×10^{-10}*
	K80 + G	ω:	0.02	1.00		ω:	0.00	1.00	5.69				
CTL(1)	10	*p*:	0.32	0.68	1945.68	*p*:	0.25	0.51	0.24	1918.68	1.57	54.00	1.9×10^{-12}*
	K80 + G	ω:	0.00	1.00		ω:	0.00	1.00	4.54				
CTL(2)	9	*p*:	0.62	0.38	995.74	*p*:	0.57	0.40	0.03	990.08	0.57	11.32	3.5×10^{-3}
	K80 + G	ω:	0.09	1.00		ω:	0.11	1.00	8.92				
CTL(3)	6	*p*:	0.25	0.75	1396.19	*p*:	0.13	0.63	0.24	1385.71	1.46	20.96	2.8×10^{-5}*
	TrNef + I	ω:	0.00	1.00		ω:	0.00	1.00	4.54				
CTL(4)	3	*p*:	0.33	0.67	1089.73	*p*:	0.81	0.00	0.19	1083.21	1.03	13.04	1.5×10^{-3}*
	None	ω:	0.00	1.00		ω:	0.55	1.00	6.05				
CTL(7)	7	*p*:	0.27	0.73	1436.33	*p*:	0.11	0.74	0.15	1419.73	1.48	33.20	6.2×10^{-8}*
	K80 + G	ω:	0.00	1.00		ω:	0.00	1.00	6.59				
CTL(8)	14	*p*:	0.40	0.60	2268.30	*p*:	0.27	0.49	0.24	2239.96	1.24	56.68	4.9×10^{-13}*
	K81 + G	ω:	0.06	1.00		ω:	0.00	1.00	4.17				
CTL(9)	9	*p*:	0.37	0.63	1406.43	*p*:	0.30	0.30	0.40	1381.46	2.24	49.94	1.4×10^{-11}*
	K80 + I + G	ω:	0.00	1.00		ω:	0.00	1.00	6.48				
SVMP(1,2,3)	6	*p*:	0.50	0.50	3434.42	*p*:	0.50	0.37	0.13	3422.39	0.72	24.06	6.0×10^{-6}*
	HKY + G	ω	0.00	1.00		ω:	0.00	1.00	3.90				
SVMP(4,6)	3	*p*:	0.45	0.55	3070.89	*p*:	0.96	0.00	0.04	3062.92	0.72	15.94	3.5×10^{-4}*
	None	ω:	0.00	1.00		ω:	0.51	1.00	16.14				
SVMP(5,7)	7	*p*:	0.43	0.57	4344.98	*p*:	0.44	0.42	0.14	4335.37	0.79	19.22	6.7×10^{-5}*
	HKY + I	ω:	0.00	1.00		ω:	0.00	1.00	3.45				
PLA_2	14	*p*:	0.51	0.49	1324.85	*p*:	0.61	0.12	0.27	1315.26	0.79	19.18	6.8×10^{-5}*
	TrNef + I	ω:	0.01	1.00		ω:	0.11	1.00	3.18				
LAAO	9	*p*:	0.46	0.54	3348.24	*p*:	0.64	0.00	0.36	3331.50	0.98	33.48	5.4×10^{-8}*
	HKY + I	ω:	0.00	1.00		ω:	0.00	1.00	3.06				
CRISP	7	*p*:	0.47	0.53	1799.07	*p*:	0.78	0.00	0.22	1781.98	1.11	34.18	3.8×10^{-8}*
	HKY + I	ω:	0.00	1.00		ω:	0.32	1.00	5.06				
Vespryn	16	*p*:	0.45	0.55	1475.85	*p*:	0.65	0.00	0.35	1469.63	0.84	12.44	2.0×10^{-3}*
	TrNef + G	ω:	0.00	1.00		ω:	0.09	1.00	2.65				
Hyaluronidase	15	*p*:	0.38	0.62	650.14	*p*:	0.96	0.00	0.04	647.36	0.78	5.56	0.062
	K81uf	ω:	0.00	1.00		ω:	0.59	1.00	13.73				

An "*" indicates significance at the 5% level after a Bonferroni correction for 20 tests.
[a] Before correction for multiple comparisons.

undertaken for these particular genes. Eight of the sequences from other snake species used below were used in the analyses of two different *C. adamanteus* CTLs. Because of the paucity of homologous sequences in the database, we had to include more divergent sequences in many of the analyses, which resulted in this slight overlap in sequences used. As our focus is on the strength of evidence for positive selection for the class as a whole, as opposed to the individual paralogs, this slight lack of independence should not undermine our results, especially in light of the *P* values obtained (see Table 2). In six of the seven tests for positive selection for the *C. adamanteus* CTLs, we could reject the nearly-neutral model (M1) in favor of the positive-selection model (M2) at a significance level of <0.05, even after a Bonferroni correction (Table 2). Rejection failed for CTL(2) with $P = 0.07$. Across analyses with significant results, we found 15–40% of the codons to be under positive selection with $4.17 \leq \omega \leq 6.59$. Under the single-ratio model, M0, we found $0.57 \leq \omega \leq 2.24$. Only one of the data sets, CTL(2), had an overall $\omega < 1$.

3.3.3. Metalloproteinases

To analyze the *C. adamanteus* SVMPs, we first excluded the four partial sequences, SVMPs 8–11. We divided the remaining sequences into type II and type III metalloproteinases. All of the type II sequences (SVMPs 1–3) were sufficiently similar to be analyzed together, and the type III sequences were split into two groups: SVMP(4,6) and SVMP(5,7). For all three tests, we could reject the nearly-neutral model in favor of the positive-selection model at a significance level of <0.01 after a Bonferroni correction (Table 2). Overall, we found 4–14% of codon sites to be under positive selection with $3.45 \leq \omega \leq 16.14$. Under the single-ratio model, M0, we found $0.72 \leq \omega \leq 0.79$.

3.3.4. Phospholipase A_2

We had only one PLA_2 from *C. adamanteus*, so combining it with 13 other sequences from the database, we could easily reject the null, nearly-neutral model (M1) in favor of the alternative positive-selection model (M2). We found $P = 0.001$ after a Bonferroni correction (Table 2).

We found strong evidence for positive selection acting on approximately 27% of the sites in this protein with $\omega = 3.18$. Fitting the single-ratio model, M0, yielded $\omega = 0.79$.

3.3.5. L-amino acid oxidase

Using the *C. adamanteus* LAAO together with 8 sequences from the database, we could reject the nearly-neutral model (M1) in favor of the positive-selection model (M2) with $P < 0.001$ after a Bonferroni correction (Table 2). We estimated that 36% of sites are under positive selection with $\omega = 3.06$. Under the single-ratio model, M0, $\omega = 0.98$.

3.3.6. Cysteine-rich secretory protein

A single CRISP was found in the *C. adamanteus* transcriptome, and by analyzing it together with 6 homologous sequences from the database, we found we could reject the nearly-neutral model (M1) in favor of the positive-selection model (M2) with $P < 0.001$ after a Bonferroni correction. We found strong evidence for positive selection acting on approximately 22% of the codon sites in this protein with $\omega = 5.06$. Under the single-ratio model, $\omega = 1.11$ (Table 2).

3.3.7. Vespryn

Analyzing the *C. adamanteus* vespryn sequence with 15 others from the database, we rejected the null, nearly-neutral model (M1) in favor of the alternative positive-selection model (M2) with $P = 0.04$ after a Bonferroni correction for 20 tests. We found evidence that 35% of the codon sites evolved under positive selection with $\omega = 2.65$. Under the single-ratio model, $\omega = 0.84$.

3.3.8. Nucleotidase

A blastn search of the *C. adamanteus* nucleotidase coding sequence resulted in only two significant matches, both from *Gloydius blomhoffi*, which differ by only two nucleotides and only one amino acid residue, so we focus on only one of them (AB332405). Our *C. adamanteus* sequence is 5′ truncated, so this region was excluded from the analysis. The excluded region corresponds to the first 62 amino acid residues of the *G. blomhoffi* sequence. Having only two sequences, we could only use the yn00 program from PAML (Yang, 1997, 2007; Yang and Nielsen, 2000) for maximum-likelihood estimation of the number of synonymous and nonsynonymous substitutions that have accrued since their divergence. We found $dN = 0.02$ and $dS = 0.04$, with standard errors ≤ 0.01. Therefore, $dN/dS = 0.36$, suggesting that these sequences have not diverged as a result of positive selection but have instead primarily experienced purifying selection. Note, however, that some sites might still have experienced positive selection, as the method averages over all sites (similarly to the single-ratio model, M0, in codeml).

3.3.9. Myotoxin

As, effectively, only a single nucleotide sequence closely related to the *C. adamanteus* myotoxin sequence was found in the database (though several very closely related isoforms were present), we could not use codeml for detecting positive selection but instead had to use yn00 to estimate the rates of nonsynonymous (dN) and synonymous (dS) substitution as described above. This approach averages the rates across all sites rather than allowing classes of sites with different rates. Comparing the *C. adamanteus* myotoxin with crotamine from *C. durissus terrificus* (AF053075), we estimated $dN = 0.06$ and $dS = 0.02$, both with a standard error of 0.02. Therefore, we find $dN/dS = 2.8$, suggesting that these sequences have diverged primarily through positive selection.

3.3.10. Hyaluronidase

We analyzed the *C. adamanteus* hyaluronidase gene sequence together with 14 homologs from the database and failed to reject the nearly-neutral model (M1) in favor of the positive-selection model (M2) with $P = 0.06$ before a Bonferroni correction (Table 2). Note, however, that we obtained sequence for only about 25% of the coding sequence for this gene, on the basis of the length of homologous sequences from other species. Codon-based tests for positive selection have been found to have low power for short sequences (Anisimova et al., 2001).

4. Discussion

C. adamanteus is generally considered to be the most dangerous snake in the United States because of its extremely large size and commensurately large venom glands, which can yield in excess of 1 ml of venom (DRR, personal observation). It and the Western Diamondback Rattlesnake (*C. atrox)* are responsible for most snakebite deaths in the United States (Gold et al., 2002). Our results for *C. adamanteus* are the first transcriptomic treatment of a member of the genus *Crotalus* and, to our knowledge, the first application of next-generation sequencing technology to the study of snake venom-gland transcriptomes.

4.1. Sequencing and analysis techniques

Our library preparation and sequencing protocols introduced several biases into our results, although these should have little effect on our conclusions. Our cDNA library preparation resulted in a bias toward reads in the middles of transcripts and low coverage of the 5′ and 3′ ends. In particular, we observed a tendency for 5′ truncation of some transcripts despite overall moderate coverage. As our focus is on identifying the genes being transcribed and sequencing their coding sequences, this bias is not overly problematic, aside from resulting in a small number of 5′-truncated coding sequences. For sequencing, we used cDNA fragments in the size range of 600–800 base pairs (including adapters). Although this procedure makes optimal use of the current 454 read lengths (approximately 400 nucleotides), it effectively precludes sequencing of transcripts less than about 400 bases long. Very short transcripts may therefore have been missed. We were, however, able to sequence a transcript with a coding sequence of just 189 bases (myotoxin), suggesting that this may not have been a serious limitation.

In our analysis of transcripts, we excluded from consideration all unassembled sequencing reads, which accounted for 13% of the total. These singletons were of significantly lower quality than the sequences included in contigs. Their average length was 106 bases shorter, and

their average quality scores were much lower (28 for assembled and 23 for singletons). Because singletons outnumbered contigs three to one and their quality was significantly lower, their inclusion in the analysis would bias the results toward genes expressed at low levels on the basis of low-quality data.

We were unable to detect significant open reading frames in 21 of the 90 high-coverage toxin contigs. Nearly all of these matched untranslated regions of known toxin transcripts and were therefore probably either assembly detritus or alternative splicings of the 5′- or 3′-untranslated regions. Alternatively, given the bias in our sequencing protocol toward the middles of transcripts, they may simply represent unconnected fragments of other transcripts that happen to lack coding sequences. In a couple of instances, we found multiple fragments of coding sequence interspersed with noncoding sequence, probably indicating unprocessed transcripts.

Single amino-acid differences can potentially have dramatic effects on the properties of a toxin, but we chose to be conservative in designating a transcript as representing a new gene in our analysis of high-coverage toxin contigs; two transcripts differing at two or fewer amino-acid sites were considered to be the same. We know that 454 sequencing is error prone and also that the individual we sequenced could have been heterozygous at some loci. This combination makes difficult determination of whether two similar transcripts represent different alleles or paralogous genes. Because we were trying to determine the number of members of each toxin class, our null assumption was that two versions with very minute sequence differences were different alleles of the same gene.

4.2. Snake venom-gland transcriptomes

We sequenced mRNA transcripts from the venom glands of a single adult female specimen. We elected not to pool across multiple specimens so as to not lose information on relative abundances and to reduce the problems with multiple alleles. Nearly half (46%) of the sequencing reads from the *C. adamanteus* venom-gland transcriptome were identified as representing toxins, but at the level of contigs (our approximation for transcripts), the nontoxins outnumbered the toxins by approximately 9:1 (Fig. 1). Clearly, therefore, the snake-venom toxin genes are massively overexpressed in the venom glands. Among these toxins, the snake-venom serine proteinases were the most highly expressed, followed by the C-type lectins (Fig. 1 and Table 1). The snake-venom metalloproteinases were found to be the most diverse class, with at least 11 unique members. In total, we identified representatives of 11 different known toxin families in the *C. adamanteus* venom-gland transcriptome: serine proteinases, C-type lectins, metalloproteinases, phospholipases A_2, L-amino acid oxidases, cysteine-rich secretory proteins, vespryns, phosphodiesterases, nucleotidases, myotoxins, and hyaluronidases. We also identified two new putative classes of toxins: a phospholipase B and an epidermal growth factor. We based our hypothesis of toxin function on their overexpression in the venom glands and the clear evidence of signal peptides for both proteins, but confirmation will require a proteomic analysis (see, e.g., Calvete et al., 2007) of *C. adamanteus* venom, which is currently under way. In our detailed analysis of the most highly expressed transcripts, we identified 40 toxin genes, which accounted for 24% of the total sequencing reads. Thirty had complete coding sequences, and 10 had partial coding sequences, probably because of the bias inherent in our cDNA library construction protocol (see Materials and Methods). These 40 transcripts probably represent the majority of the expressed toxin genes (see below).

Symptoms of *C. adamanteus* bites in humans include instant and intense pain at the site of the bite, bleeding from the bite wound, intense internal pain, hemorrhage from the mouth, reduced blood pressure and pulse, swelling and discoloration at the site of the bite, difficulty breathing, muscle twitching and spasms, and paralysis of the legs (Klauber, 1997). These symptoms are consistent with the effects of the identified toxins on the basis of homology with sequences from other species that have been characterized. The preponderance of these genes encode proteins that interfere with the hemostatic systems (SVSPs and CTLs) and induce hemorrhage (SVMPs).

A wide diversity of snake species has been subjected to venom-gland transcriptome sequencing, including a number of elapids (see, e.g., Leao et al., 2009), viperids (see, e.g., Wagstaff and Harrison, 2006; Zhang et al., 2006), and colubrids (see, e.g., Ching et al., 2006). The closest relative to *C. adamanteus* treated in this manner is the Desert Massasauga (*Sistrurus catenatus edwardsii*; Pahari et al., 2007), which diverged from *C. adamanteus* approximately 12.7 million years ago (Douglas et al., 2006). These two transcriptomes show remarkable concordances in patterns of toxin families. In both, the serine proteinases appear to be the dominant toxin family, indicating that the disruption of the prey's hemostatic system is a primary mechanism of prey incapacitation for both species. Metalloproteinases are both diverse and abundant in the two species. For each of the two species, only a single CRISP, PLA_2, and LAAO were identified. In contrast, the C-type lectins are diverse and abundant in the *C. adamanteus* venom-gland transcriptome but not in the *S. catenatus* transcriptome, indicating a major shift in venom composition in the lineage leading to one of these two species. The next most closely related species with some transcriptome sequencing is the Western Cottonmouth (*Agkistrodon piscivorus leucostoma*; Jia et al., 2008), which shares a common ancestor with *C. adamanteus* approximately 22 million years ago (Douglas et al., 2006). For this species, the venom composition appears to be drastically different; PLA_2 is the most abundant and diverse toxin class.

Our results represent the most complete transcriptomic characterization of toxin sequences in snake venom glands to date. For the work presented here, we focused only on those sequences that contributed to proteins found in the venom. Future work will address the nontoxin transcripts expressed in the *C. adamanteus* venom gland. Despite generating over 600,000 sequences with an average length of 300 nucleotides, we probably have not completely characterized all of the venom genes, partly because of the high inherent error rate in 454 sequencing; high levels of coverage are required for confidence in the resulting

sequences. We therefore made no attempt to identify transcripts that were extremely similar to one another, to identify different alleles of the same genes, or to look for alternatively spliced transcripts. We also only focused our in-depth analysis on those transcripts with high coverage. Therefore, at this stage we cannot claim to have completely characterized the expressed toxin genes, despite having identified 40 different venom transcripts. Our use of Roche 454 sequencing technology was motivated by the relatively long current average read lengths (approximately 400 bases before quality trimming) compared to other technologies. Analysis of this data set makes clear that coverage, rather than read length, is the primary limitation for venom-gland transcriptomes. We are therefore currently exploring the use of Illumina technology (Illumina, Inc., http://www.illumina.com), which produces shorter but significantly more reads, in characterizing venom-gland transcriptomes. In addition, we must confirm the presence of the proteins encoded by our identified transcripts in the venom of this species. This work is currently under way.

4.3. Toxin molecular evolution

Snake venom is an exceptional trait in evolutionary genetics for a number of reasons. Most measurable, evolutionarily significant traits in complex organisms are the results of long and poorly characterized developmental pathways, so understanding the mechanisms by which genetic variation manifests itself to the sieve of selection at the phenotypic level is next to impossible. Venom and its delivery apparatus (see Mackessy and Baxter, 2006) virtually define the viperid snakes and dictate their ecology, natural history, and evolution. Venom is a complex, polygenic phenotype with a well-defined and ecologically critical function that can nonetheless be decomposed into simpler, independent components whose effects are manifested as easily measurable, biochemical properties.

The direct connection between venom genes and fitness probably explains the ubiquity of positive selection in our identified toxin genes. The vast majority of toxin genes we were able to test for evidence of positive selection did, in fact, show such evidence (Table 2). We performed 20 tests for positive selection, covering 8 of the 11 previously described toxin families found in *C. adamanteus*. Seventeen of these tests were significant, indicating clear evidence that positive selection contributes to the evolutionary history of 7 of the 8 families. In addition, we were able to estimate *dN*/*dS* for two additional families and found evidence for positive selection in one of them. Overall, we found evidence for positive selection acting in 8 of the 11 toxin families, excluding only the nucleotidases, phosphodiesterases, and hyaluronidases. Remarkably, nearly all of the genes that contribute to the venom phenotype appear to have diverged in part as a result of positive selection.

The tests we conducted do not necessarily indicate that positive selection has been acting in the *C. adamanteus* lineage. The method averages the *dN*/*dS* across the entire estimated tree and therefore indicates only a sustained, elevated rate of nonsynonymous substitutions at some proportion of codon sites. All we can say for certain, on the basis of these test results, is that some specified proportion of sites have experienced positive selection in their history. The source for the continued elevated rate of nonsynonymous substitutions could be an arms race between snakes and their prey, as prey are known, in some cases, to be able to evolve some degree of resistance to snake venoms (Biardi et al., 2005).

Previous work has demostrated the presence of positive selection, or at least an elevated substitution rate, in a handful of venom toxin classes. For example, rapid evolutionary change has been demonstrated in disintegrins (Soto et al., 2006) and serine proteinases (Deshimaru et al. (1996), and positive selection has been detected in PLA_2s (Kordiš and Gubenšek 2000; Lynch, 2007; Gibbs and Rossiter, 2008) and disintegrins (Juárez et al., 2008). Our goal is to study molecular evolutionary patterns in all of the genes that contribute to a single, defined, evolutionarily critical phenotype, and the work reported here is the first attempt to examine evolutionary forces acting across the suite of genes contributing to venom. Even though this goal remains to be completely achieved, awaiting at minimum a complete *C. adamanteus* transcriptome and a larger set of homologous sequences from closely related species, we have managed to provide evidence for widespread positive selection in at least seven classes of toxins and spanning multiple paralogous versions within classes.

Acknowledgments

The authors would like to thank Darryl Heard for providing training to DRR and KPW in the electrostimulation technique for venom extraction and Kathleen Harper for advice on reptile anesthesia and euthanasia.

Conflict of interest statement

The authors declare that they have no conflicts of interest.

Appendix. Sequences used for detecting selection

The species names and accession numbers for the sequences used to test for selection (other than those we generated) are as follows. They are grouped by analyses (see Table 2 and the Results section). The percentage is the minimum percentage identity used for including a sequence in the analysis and was adjusted to give adequate sequences for the analyses.

SVSP(1): 89%; *Crotalus durissus* (EU360954), *Viridovipera stejnegeri* (AF395774, AF395779), *Protobothrops mucrosquamatus* (X83224, AF098262, X83222, TMU31417, X83225, X83223, X83221).

SVSP(2,5): 88%; *C. durissus* (EU360952, EU360951), *Sistrurus catenatus* (DQ464247), *V. stejnegeri* (AF545576), *Cryptelytrops albolabris* (EF690365, EF690366), *Gloydius halys* (AF056033, AF015727), *Trimeresurus gramineus* (D67083), *G. ussuriensis* (AUU32937).

SVSP(3): 88%; *C. durissus* (DQ164401), *S. catenatus* (DQ464242), *Lachesis stenophrys* (DQ247723), *Deinagkistrodon acutus* (AF333768, AJ001209, AJ006471, EF101918).

SVSP(4): 89%; *C. atrox* (AF227154), *S. catenatus* (DQ464240, DQ464241), *T. gramineus* (D67081), *V. stejnegeri* (AF545578, AF395780), *Bothrops jararaca* (AB178322), *D. acutus* (AF159058, EF101917), *T. flavoviridis* (D67078), *G. blomhoffi* (AF017737).

SVSP(6): 89%; *G. ussuriensis* (AUU32937), *S. catenatus* (DQ464238), *G. halys* (AJ001210, AJ006472), *V. stejnegeri* (AF395776), *B. jararacussu* (AY251282), *C. atrox* (AF227153), *G. shedaoensis* (AY434726).

Note that the *G. ussuriensis* sequence was also used for SVSP(2,5).

CTL(1): 83%; *S. catenatus* (DQ464258), *G. halys* (AF190827), *G. blomhoffi* (AF125309), *T. flavoviridis* (TFLFIXA, AB046491), *V. stejnegeri* (AF354911), *D. acutus* (AF176420, AY091758), *B. jararaca* (AY962524).

CTL(2): 90%; *G. halys* (AF197915), *V. stejnegeri* (AF354914, AF354913), *T. flavoviridis* (TFLFIXB), *D. acutus* (AY091761, AB036881, AF350324), *G. blomhoffi* (AF125310).

CTL(3): 84%; *S. catenatus* (DQ464257, DQ464256), *G. blomhoffi* (AB019616), *C. durissus* (AF541884), *T. jerdonii* (GU146050).

CTL(4): 83%; *V. stejnegeri* (AF354924), *C. durissus* (AF541883).

CTL(7): 87%; *C. durissus* (AF541884), *P. mucrosquamatus* (AY390534), *T. jerdonii* (GU136388), *T. flavoviridis* (AY149340), *S. catenatus* (DQ464256), *G. blomhoffi* (AB019616).

CTL(8): 84%; *C. durissus* (AF541883), *P. mucrosquamatus* (AY871785), *Cryptelytrops albolabris* (EF690367), *T. jerdonii* (GU146049), *D. acutus* (AF540645, AY091762, AF102901), *V. stejnegeri* (AF354923, AF354922, AF354919, AF354918), *T. flavoviridis* (TFLFIXA), *G. halys* (AF190827).

CTL(9): 83%; *C. durissus* (AF541881, Y16349), *P. mucrosquamatus* (AY390534, AY871786), *D. acutus* (AF176421, AY091756, AF540647), *T. jerdonii* (GU136388).

CRISP: 90%; *C. atrox* (AY181983), *Agkistrodon piscivorus* (AY181982), *S. catenatus* (DQ464263), *G. blomhoffi* (AF384218), *T. flavoviridis* (AF384219), *T. jerdonii* (AY261467).

Hyaluronidase: 90%; *Echis carinatus* (DQ840261, DQ840262), *Bitis arietans* (DQ840260, DQ840259, DQ840258, DQ840257, DQ840256), *E. pyramidum* (DQ840254, DQ840255, DQ840253), *Cerastes cerastes* (DQ840252, DQ840251, DQ840250), *E. ocellatus* (DQ840249).

LAAO: 90%; *C. atrox* (AF093248), *C. adamanteus* (AF071564), *S. catenatus* (DQ464267), *V. stejnegeri* (AY338966, AY277739), *G. blomhoffi* (AB072392), *G. halys* (AY450403), *Calloselasma rhodostoma* (AJ271725).

PLA_2: 90%; *C. horridus* (GQ168368, GQ168369), *C. atrox* (AF269131), *C. viridis* (AF403134, AY120876, AY120875, AY120877, AF403135, AF403137, AF403136), *C. durissus* (GQ466583), *S. catenatus* (DQ464264, AY508692).

SVMP(1,2,3): 90%; *C. atrox* (GQ451438), *G. halys* (AY071905), *T. jerdonii* (AY364231).

SVMP(4,6): 90%; *A. piscivorus* (GQ451441).

SVMP(5,7): 90%; *C. atrox* (AB042840), *G. halys* (AY149647), *D. acutus* (DQ263750), *T. flavoviridis* (AB051849), *V. stejnegeri* (DQ335449).

Vespryn: 85%; *L. muta* (DQ396476), *Ophiophagus hannah* (AY351433), *Pseudechis porphyriacus* (EU024762), *Demansia vestigiata* (DQ917532, DQ917531), *Hoplocephalus stephensii* (EU024764), *Tropidechis carinatus* (EU024761), *Rhinoplocephalus nigrescens* (EU024760, EU024759), *Pseudechis australis* (EU024763), *Oxyuranus scutellatus* (EU024755), *Notechis scutatus* (EU024753, EU024754), and *Pseudonaja textilis* (EU024752, EU024751).

References

Aird, S.D., 2002. Ophidian envenomation strategies and the role of purines. Toxicon 40, 335–393.

Aird, S.D., 2010. The role of purine and pyrimidine nucleosides in snake venoms. In: Mackessy, S.P. (Ed.), Handbook of Venoms and Toxins of Reptiles. CRC Press, Boca Raton, Florida, pp. 393–419.

Anisimova, M., Bielawski, J.P., Yang, Z., 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Molecular Biology and Evolution 18, 1585–1592.

Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. Journal of Molelcular Biology 340, 783–795.

Biardi, J.E., Chien, D.C., Coss, R.G., 2005. California ground squirrel *(Spermophilus beecheyi)* defenses against rattlesnake venom digestive and hemostatic toxins. Journal of Chemical Ecology 31, 2501–2518.

Bull, J.J., Badgett, M.R., Wichman, H.A., 2000. Big-benefit mutations in a bacteriophage inhibited with heat. Molecular Biology and Evolution 17, 942–950.

Calvete, J.J., Juárez, P., Sanz, L., 2007. Snake venomics. Strategy and applications. Journal of Mass Spectrometry 42, 1405–1414.

Casewell, N.R., Harrison, R.A., Wüster, W., Wagstaff, S.C., 2009. Comparative venom gland transcriptome surveys of the saw-scaled vipers (Viperidae: *Echis)* reveal substantial intra-family gene diversity and novel venom transcripts. BMC Genomics 10, 564.

Ching, A.T.C., Rocha, M.M.T., Paes Leme, A.F., Pimenta, D.C., Furtado, M.F.D., Serrano, S.M.T., Ho, P.L., Junqueira-de Azevedo, I.L.M., 2006. Some aspects of the venom proteome of the Colubridae snake *Philodryas olfersii* revealed from a Duvernoy's (venom) gland transcriptome. FEBS Letters 580, 4417–4422.

Cidade, D.A.P., Simao, T.A., Dávila, A.M.R., Wagner, G., Junqueira-de Azevedo, I.L.M., Ho, P.L., Bon, C., Zingali, R.B., Albano, R.M., 2006. *Bothrops jararaca* venom gland transcriptome: analysis of the gene expression pattern. Toxicon 48, 437–461.

Conant, R., Collins, J.T., 1998. A Field Guide to Reptiles and Amphibians of Eastern and Central North America, third ed. Houghton Mifflin Harcourt, New York, New York.

Crandall, K.A., Kelsey, C.R., Imamichi, H., Lane, H.C., Salzman, N.P., 1999. Parallel evolution of drug resistance in HIV: failure of non-synonymous/synonymous substitution rate ratio to detect selection. Molecular Biology and Evolution 16, 372–382.

de Azevedo, I.L.M.J., Ching, A.T.C., Carvalho, E., Faria, F., Nishiyama Jr., M.Y., Ho, P.L., Diniz, M.R.V., 2006. *Lachesis muta* (Viperidae) cDNAs reveal diverging pit viper molecules and scaffolds typical of cobra (Elapidae) venoms: implications for snake toxin repertoire evolution. Genetics 173, 877–889.

Deshimaru, M., Ogawa, T., ichi Nakashima, K., Nobuhisa, I., Chijiwa, T., Shimohigashi, Y., Fukumaki, Y., Niwa, M., Yamashina, I., Hattori, S., Ohno, M., 1996. Accelerated evolution of crotalinae snake venom gland serine proteases. FEBS Letters 397, 83–88.

Dhananjaya, B.L., Vishwanath, B.S., D'Souza, C.J.M., 2010. Snake venom nucleases, nucleotidases, and phosphomonoesterases. In: Mackessy, S.P. (Ed.), Handbook of Venoms and Toxins of Reptiles. CRC Press, Boca Raton, Florida, pp. 155–171.

Doley, R., Zhou, X., Kini, R.M., 2010. Snake venom phospholipase A_2 enzymes. In: Mackessy, S.P. (Ed.), Handbook of Venoms and Toxins of Reptiles. CRC Press, Boca Raton, Florida, pp. 173–205.

Douglas, M.E., Douglas, M.R., Schuett, G.W., Porras, L.W., 2006. Evolution of rattlesnakes (Viperidae; Crotalus) in the warm deserts of western North America shaped by Neogene vicariance and Quaternary climate change. Molecular Ecology 15, 3353–3374.

Du, X.-Y., Clemetson, K.J., 2010. Reptile C-type lectins. In: Mackessy, S.P. (Ed.), Handbook of Venoms and Toxins of Reptiles. CRC Press, Boca Raton, Florida, pp. 359–375.

Dundee, H.A., Rossman, D.A., 1996. The Amphibians and Reptiles of Louisiana. Louisiana University Press, Baton Rouge, Louisiana.

Emanuelsson, O., Brunak, S., von Heijne, G., Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocols 2, 953–971.

Emrich, S.J., Barbazuk, W.B., Li, L., Schnabel, P.S., 2007. Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Research 17, 69–73.

Escoubas, P., King, G.F., 2009. Venomics as a drug discovery platform. Expert Review of Proteomics 6, 221–224.

Fox, J.W., Serrano, S.M.T., 2005. Structural considerations of the snake venom metalloproteinases, key members of the M12 reprolysin family of metalloproteinases. Toxicon 45, 969–985.

Fox, J.W., Serrano, S.M.T., 2010. Snake venom metalloproteinases. In: Mackessy, S.P. (Ed.), Handbook of Venoms and Toxins of Reptiles. CRC Press, Boca Raton, Florida, pp. 95–113.

Francischetti, I.M.B., My-Pham, V., Harrison, J., Garfield, M.K., Ribeiro, J.M. C., 2004. *Bitis gabonica* (Gaboon viper) snake venom gland: toward a catalog for the full-length transcripts (cDNA) and proteins. Gene 337, 55–69.

Gibbs, H.L., Rossiter, W., 2008. Rapid evolution by positive selection and gene gain and loss: PLA_2 venom genes in closely related *Sistrurus* rattlesnakes with divergent diets. Journal of Molecular Evolution 66, 151–166.

Gold, B.S., Dart, R.C., Barish, R.A., 2002. Bites of venomous snakes. New England Journal of Medicine 347, 347–356.

Gomis-Rüth, F.-X., Kress, L.F., Bode, W., 1993. First structure of a snake venom metalloproteinase: a prototype for matrix metalloproteinases/collagenases. EMBO Journal 12, 4151–4157.

Gong, W., Zhu, X., Liu, S., Teng, M., Niu, L., 1998. Crystal structures of Acutolysin A, a three-disulfide hemorrhagic zinc metalloproteinase from the snake venom of *Agkistrodon acutus*. Journal of Molelcular Biology 283, 657–668.

Gutiérrez, J.M., Rucavado, A., Escalante, T., 2010. Snake venom metalloproteinases: biological roles and participation in the pathophysiology of envenomation. In: Mackessy, S.P. (Ed.), Handbook of Venoms and Toxins of Reptiles. CRC Press, Boca Raton, Florida, pp. 115–138.

Harrison, R.A., Ibison, F., Wilbraham, D., Wagstaff, S.C., 2007. Identification of cDNAs encoding viper venom hyaluronidases: cross-generic sequence conservation of full-length and unusually short variant transcripts. Gene 392, 22–33.

Harvey, A.L., Bradley, K.N., Cochran, S.A., Rowan, E.G., Pratt, J.A., Quillfeldt, J.A., Jerusalinsky, D.A., 1998. What can toxins tell us for drug discovery? Toxicon 36, 1635–1640.

Hayes, M.B., Wellner, D., 1969. Microheterogeneity of L-amino acid oxidase. Journal of Biological Chemistry 244, 6636–6644.

Holland, D.R., Clancy, L.L., Muchmore, S.W., Ryde, T.J., Einspahr, H.M., Finzel, B.C., Heinrikson, R.L., Watenpaugh, K.D., 1990. The crystal structure of a lysine 49 phospholipase A_2 from the venom of the cottonmouth snake at 2.0-Å resolution. Journal of Biological Chemistry 265, 17649–17656.

Jia, Y., Cantu, B.A., Sánchez, E.E., Pérez, J.C., 2008. Complementary DNA sequencing and identification of mRNAs from the venomous gland of *Agkistrodon piscivorus leucostoma*. Toxicon 51, 1457–1466.

Juárez, P., Comas, I., González-Candelas, F., Calvete, J.J., 2008. Evolution of snake venom disintegrins by positive Darwinian selection. Molecular Biology and Evolution 25, 2391–2407.

Junqueira-de Azevedo, I.L.M., Ho, P.L., 2002. A survey of gene expression and diversity in the venom glands of the pitviper snake *Bothrops insularis* through the generation of expressed sequence tags (ESTs). Gene 299, 279–291.

Kemparaju, K., Girish, K.S., Nagaraju, S., 2010. Hyaluronidases, a neglected class of glycosidases from snake venom: beyond a spreading factor. In: Mackessy, S.P. (Ed.), Handbook of Venoms and Toxins of Reptiles. CRC Press, Boca Raton, Florida, pp. 237–258.

Kini, R.M., 2003. Excitement ahead: structure, function and mechanism of snake venom phospholipase A_2 enzymes. Toxicon 42, 827–840.

Klauber, L.M., 1997. Rattlesnakes: Their Habits, Life Histories, and Influence on Mankind, second ed. University of California Press, Berkeley, California.

Kordiš, D., Gubenšek, F., 2000. Adaptive evolution of animal toxin multigene families. Gene 261, 43–52.

Kumasaka, T., Yamamoto, M., Moriyama, H., Tanaka, N., Sato, M., Katsube, Y., Yamakawa, Y., Omori-Satoh, T., Iwanaga, S., Ueki, T., 1996. Crystal structure of H_2-proteinase from the venom of *Trimeresurus flavoviridis*. Journal of Biochemistry 119, 49–57.

Leão, L.I., Ho, P.L., Junqueira-de Azevedo, I.L.M., 2009. Transcriptomic basis for an antiserum against *Micrurus corallinus* (coral snake) venom. BMC Genomics 10, 112.

Lou, Z., Hou, J., Liang, X., Chen, J., Qiu, P., Liu, Y., Li, M., Rao, Z., Yan, G., 2005. Crystal structure of a non-hemorrhagic fibrin(ogen)olytic metalloproteinase complexed with a novel natural tri-peptide inhibitor from venom of *Agkistrodon acutus*. Journal of Structural Biology 152, 195–203.

Lynch, V.J., 2007. Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A_2 genes. BMC Evolutionary Biology 7, 2.

Mackessy, S.P., Baxter, L.M., 2006. Bioweapons synthesis and storage: the venom gland of front-fanged snakes. Zoologischer Anzeiger 245, 147–159.

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.

Ménez, A., 1998. Functional architectures of animal toxins: a clue to drug design? Toxicon 36, 1557–1572.

Minin, V., Abdo, Z., Joyce, P., Sullivan, J., 2003. Performance-based selection of likelihood models for phylogeny. Systematic Biology 52, 647–683.

Oguiura, N., Boni-Mitake, M., Rádis-Baptista, G., 2005. New view on crotamine, a small basic polypeptide myotoxin from South American rattlesnake venom. Toxicon 46, 363–370.

Pahari, S., Mackessy, S.P., Kini, R.M., 2007. The venom gland transcriptome of the desert massasauga rattlesnake *(Sistrurus catenatus edwardsii)*: towards an understanding of venom composition among advanced snakes (superfamily Colubroidea). BMC Molecular Biology 8, 115.

Palmer, W.M., Braswell, A.L., 1995. Reptiles of North Carolina. University of North Carolina Press, Chapel Hill, North Carolina.

Pawelek, P.D., Cheah, J., Coulombe, R., Macheroux, P., Ghisla, S., Vrielink, A., 2000. The structure of L-amino acid oxidase reveals the substrate trajectory into an enantiomerically conserved active site. EMBO Journal 19, 4204–4215.

Phillips, D.J., Swenson, S.D., Francis, S., Markland, J., 2010. Thrombin-like snake venom serine proteinases. In: Mackessy, S.P. (Ed.), Handbook of Venoms and Toxins of Reptiles. CRC Press, Boca Raton, Florida, pp. 139–154.

Pung, Y.F., Wong, P.T.H., Kumar, P.P., Hodgson, W.C., Kini, R.M., 2005. Ohanin, a novel protein from king cobra venom, induces hypolocomotion and hyperalgesia in mice. Journal of Biological Chemistry 280, 13137–13147.

Pung, Y.F., Kumar, S.V., Rajagopalan, N., Fry, B.G., Kumar, P.P., Kini, R.M., 2006. Ohanin, a novel protein from king cobra venom: its cDNA and genomic organization. Gene 371, 246–256.

Qinghua, L., Xiaowei, Z., Wei, Y., Chenji, L., Yijun, H., Pengxin, Q., Xingwen, S., Songnian, H., Guangmei, Y., 2006. A catalog for transcripts in the venom gland of the *Agkistrodon acutus*: identification of the toxins potentially involved in coagulopathy. Biochemical and Biophysical Research Communications 341, 522–531.

Rádis-Baptista, G., Oguiura, N., Hayashi, M.A.F., Camargo, M.E., Grego, K.F., Oliveira, E.B., Yamane, T., 1999. Nucleotide sequence of crotamine isoform precursors from a single South American rattlesnake *(Crotalus durissus terrificus)*. Toxicon 37, 973–984.

Raibekas, A.A., Massey, V., 1998. Primary structure of the snake venom L-amino acid oxidase shows high homology with the mouse B cell interleukin 4-induced Fig1 protein. Biochemical and Biophysical Research Communications 248, 476–478.

Rokyta, D.R., Wichman, H.A., 2009. Genic incompatibilities in two hybrid bacteriophages. Molecular Biology and Evolution 26, 2831–2839.

Rokyta, D.R., Burch, C.L., Caudle, S.B., Wichman, H.A., 2006. Horizontal gene transfer and the evolution of microvirid coliphage genomes. Journal of Bacteriology 188, 1134–1142.

Rotenberg, D., Bamberger, E.S., Kochva, E., 1971. Studies on ribonucleic acid synthesis in the venom glands of *Vipera palaestinae* (Ophidia, Reptilia). Biochemical Journal 121, 609–612.

Samejima, Y., Aoki, Y., Mebs, D., 1991. Amino acid sequence of a myotoxin from venom of the Eastern Diamondback Rattlesnake (*Crotalus adamanteus*). Toxicon 29, 461–468.

Serrano, S.M.T., Maroun, R.C., 2005. Snake venom serine proteinases: sequence homology vs. substrate specificity, a paradox to be solved. Toxicon 45, 1115–1132.

Soto, J.G., Powell, R.L., Reyes, S.R., Wolana, L., Swanson, L.J., Sanchez, E.E., Perez, J.C., 2006. Genetic variation of a disintegrin gene found in the American copperhead snake *(Agkistrodon contortrix)*. Gene 373, 1–7.

Swofford, D.L., 1998. Phylogenetic Analysis Using Parsimony* (PAUP*), Version 4.0. Sinauer Associates, Sunderland, MA.

Tan, N.-H., Fung, S.-Y., 2010. Snake venom L-amino acid oxidases. In: Mackessy, S.P. (Ed.), Handbook of Venoms and Toxins of Reptiles. CRC Press, Boca Raton, Florida, pp. 221–235.

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 Research 22, 4673–4680.

Wagstaff, S.C., Harrison, R.A., 2006. Venom gland EST analysis of the saw-scaled viper, *Echis ocellatus*, reveals novel $\alpha_9\beta_1$ integrin-binding motifs in venom metalloproteinases and a new group of putative toxins, renin-like aspartic proteases. Gene 377, 21–32.

Walker, J.R., Nagar, B., Young, N.M., Hirama, T., Rini, J.M., 2004. X-ray crystal structure of a galactose-specific C-type lectin possessing a novel decameric quaternary structure. Biochemistry 43, 3783–3792.

Watanabe, L., Shannon, J.D., Valente, R.H., Rucavado, A., Alape-Girón, A., Kamiguti, A.S., Theakston, R.D.G., Fox, J.W., Gutiérrez, J.M., Arni, R.K., 2003. Amino acid sequence and crystal structure of BaP1,

a metalloproteinase from *Bothrops asper* snake venom that exerts multiple tissue-damaging activities. Protein Science 12, 2273–2281.
Yamazaki, Y., Morita, T., 2004. Structure and function of snake venom cysteine-rich secretory proteins. Toxicon 44, 227–231.
Yamazaki, Y., Hyodo, F., Morita, T., 2003. Wide distribution of cysteine-rich secretory proteins in snake venoms: isolation and cloning of novel snake venom cysteine-rich secretory proteins. Archives of Biochemistry and Biophysics 412, 133–141.
Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences 13, 555–556.
Yang, Z., 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24, 1586–1591.
Yang, Z., Nielsen, R., 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Molecular Biology and Evolution 17, 32–43.
Yang, Z., Swanson, W.J., 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Molecular Biology and Evolution 19, 49–57.
Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.-M.K., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 431–449.
Zhang, D., Botos, I., Gomis-Rüth, F.-X., Doll, R., Blood, C., Njoroge, F.G., Fox, J.W., Bode, W., Meyer, E.F., 1994. Structural interaction of natural and synthetic inhibitors with the venom metalloproteinase, atrolysin C (form d). Proceedings of the National Academy of Sciences of the USA 91, 8447–8451.
Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000. A greedy algorithm for aligning DNA sequences. Journal of Computational Biology 7, 203–214.
Zhang, B., Liu, Q., Yin, W., Zhang, X., Huang, Y., Luo, Y., Qiu, P., Su, X., Yu, J., Hu, S., Yan, G., 2006. Transcriptome analysis of *Deinagkistrodon acutus* venomous gland focusing on cellular structure and functional aspects using expressed sequence tags. BMC Genomics 7, 152.