Comparative analyses of putative toxin gene homologs from an Old World viper, Daboia russelii

Neeraja M Krishnan [1] , **Binay Panda** Corresp. [1]

[1] Ganit Labs, Bio-IT Centre, Institute of Bioinformatics and Applied Biotechnology, Bangalore, India

Corresponding Author: Binay Panda
Email address: binay@ganitlabs.in

Availability of snake genome sequences has opened up exciting areas of research on comparative genomics and gene diversity. One of the challenges in studying snake genomes is the acquisition of biological material from live animals, especially from the venomous ones, making the process cumbersome and time-consuming. Here, we report comparative sequence analyses of putative toxin gene homologs from Russell's viper (*Daboia russelii*) using whole-genome sequencing data obtained from shed skin. When compared with the major venom proteins in Russell's viper studied previously, we found 45-100% sequence similarity between the venom proteins and their putative homologs in the skin. Additionally, comparative analyses of 20 putative toxin gene family homologs provided evidence of unique sequence motifs in nerve growth factor (NGF), platelet derived growth factor (PDGF), Kunitz/Bovine pancreatic trypsin inhibitor (Kunitz BPTI), cysteine-rich secretory proteins, antigen 5, andpathogenesis-related1 proteins (CAP) and cysteine-rich secretory protein (CRISP). In those derived proteins, we identified V11 and T35 in the NGF domain; F23 and A29 in the PDGF domain; N69, K2 and A5 in the CAP domain; and Q17 in the CRISP domain to be responsible for differences in the largest pockets across the protein domain structures in crotalines, viperines and elapids from the *in silico* structure-based analysis. Similarly, residues F10, Y11 and E20 appear to play an important role in the protein structures across thekunitzprotein domain of viperids and elapids. Our study highlights the usefulness of shed skin in obtaining good quality high-molecular weight DNA for comparative genomic studies, and provides evidence towards the unique features and evolution of putative venom gene homologs in vipers.

Data deposition: Russell's viper sequence data is deposited in the NCBI SRA database under the accession number SRR5506741 and the individual gene sequences are deposited in the GenBank (accession numbers in Table S1).

Comparative analyses of putative toxin gene homologs from an Old World viper, *Daboia russelii*

Neeraja M Krishnan[1] and Binay Panda[1]*

[1]Ganit Labs, Bio-IT Centre, Institute of Bioinformatics and Applied Biotechnology, Biotech Park, Electronic City Phase I, Bangalore 560100

* Corresponding author: binay@ganitlabs.in

Abstract

Availability of snake genome sequences has opened up exciting areas of research on comparative genomics and gene diversity. One of the challenges in studying snake genomes is the acquisition of biological material from live animals, especially from the venomous ones, making the process cumbersome and time-consuming. Here, we report comparative sequence analyses of putative toxin gene homologs from Russell's viper (*Daboia russelii*) using whole-genome sequencing data obtained from shed skin. When compared with the major venom proteins in Russell's viper studied previously, we found 45-100% sequence similarity between the venom proteins and their putative homologs in the skin. Additionally, comparative analyses of 20 putative toxin gene family homologs provided evidence of unique sequence motifs in nerve growth factor (NGF), platelet derived growth factor (PDGF), Kunitz/Bovine pancreatic trypsin inhibitor (Kunitz BPTI), cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins (CAP) and cysteine-rich secretory protein (CRISP). In those derived proteins, we identified V11 and T35 in the NGF domain; F23 and A29 in the PDGF domain; N69, K2 and A5 in the CAP domain; and Q17 in the CRISP domain to be responsible for differences in the largest pockets across the protein domain structures in crotalines, viperines and elapids from the *in silico* structure-based analysis. Similarly, residues F10, Y11 and E20 appear to play an important role in the protein structures across the kunitz protein domain of viperids and elapids. Our study highlights the usefulness of shed skin in obtaining good quality high-molecular weight DNA for comparative genomic studies, and provides evidence towards the unique features and evolution of putative venom gene homologs in vipers.

Data deposition: Russell's viper sequence data is deposited in the NCBI SRA database under the accession number SRR5506741 and the individual gene sequences are deposited in the GenBank (accession numbers in Table S1).

Introduction

Snake venom genes and their products offer an excellent model system to study gene duplication, evolution of regulatory DNA sequences, and biochemical diversity and novelty of venom proteins. Additionally, snake venoms have tremendous potential in the development of new drugs and bioactive compounds (Vonk et al. 2011). Previous studies have highlighted the importance of gene duplications and/or sub-functionalization (Hargreaves et al. 2014; Malhotra et al. 2015; Rokyta et al. 2011) and transcriptional/post-transcriptional mechanisms (Casewell et al. 2014) contributing towards snake venom diversity. Venom studies, so far, have extensively used data from proteomics experiments alongside individual gene sequences or sequences of particular family members to study variations on gene structure and their sequence composition. Presently, whole genome sequences of several snake species, king cobra *Ophiophagus hannah* (Vonk et al. 2013); Burmese python *Python bivitattus* (Castoe et al. 2013); rattlesnake *Crotalus atrox* (Dowell et al. 2016); Florida pygmy rattlesnake *Sistrurus miliarius barbouri* (Vicoso et al. 2013); garter snake *Thamnophis elegans* (Vicoso et al. 2013); five-pacer viper *Deinagkistrodon acutus* (Yin et al. 2016); pit viper *Protobothops mucrosquamatus* (NCBI Accession PRJDB4386); and corn snake *Pantherophis guttatus* (Ullate-Agote et al. 2014), have either been published or their sequence been made available in the public domain. In addition, genome sequencing efforts are either underway or the sequences of venom-associated genes have been deposited in the databases for a few others (Kerkkamp et al. 2016). Out of the sequenced genomes, only a few have been annotated, or made public, a key requirement for comparative analysis of genes. This, along with the lack of availability of whole genome sequences and/or complete transcript

sequences from venom glands for most snakes has limited studies on toxin gene orthologies and gene variation among venomous snakes.

Four snakes, Russell's viper (*Daboia russelii),* saw-scaled viper (*Echis carinatus*), spectacled cobra (*Naja naja*), and common krait (*Bungarus caeruleus*) are responsible for most snakebite-related mortality in India (Mohapatra et al. 2011; Warrell et al. 2013; Whitaker 2015). Russell's viper is a member of the taxon Viperidae and subfamily Viperinae and is responsible for large numbers of snakebite incidents and deaths in India. Very little is known about the diversity of genes from any viper, including the only viperine where complete genome sequence information is available (European adder, *Vipera berus berus*, https://www.ncbi.nlm.nih.gov/bioproject/170536). Lack of gene annotation from this viper using transcripts obtained from venom glands and other snake species reduces the scope of a detailed comparative study on genes, including the toxin-associated genes. Such a study involving various groups of venomous and non-venomous snakes, in addition to other venomous vertebrates and invertebrates, will facilitate our understanding on the evolution of these genes, their diversity, and function.

One of the challenges in studying the genomes of venomous animals is related to sample acquisition. Additionally, in India, Government permission is required to catch snakes and extract blood samples from them (all snakes are protected in India under the Indian Wildlife Protection Act, 1972). This may be partially circumvented by the use of shed skin that does not require drawing blood or taking any tissue from the animals. However, working with DNA isolated from shed skin has its own challenges. Microbial contamination, lack of full-length DNA in the shed skin cells, rapid degradation of DNA in humid conditions and computational challenges in dealing with short stretches of DNA are some of the bottlenecks for working with DNA from shed skin.

In the current study, we explored the possibility of getting putative toxin gene homolog information from skin-derived low-coverage whole-genome sequencing data from Russell's viper, and performed comparative analysis versus major toxin proteins from a previously studied report (Sharma et al. 2015). We used the coding sequences and annotation from a previously characterized crotaline, a pit viper, *Protobothrops mucrosquamatus* for the analysis. On the venom homologs, we focused our analyses on five key protein domains; nerve growth factor (NGF), platelet derived growth factor (PDGF), Kunitz/Bovine pancreatic trypsin inhibitor (Kunitz BPTI), cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins (CAP) and cysteine-rich secretory protein (CRISP) in Russell's viper. Our study identified the putative venom homologs from skin and the key residues that are changed across the members of Viperinae, Crotalinae and Elapidae that might have contributed towards the evolution of venom in vipers.

Materials and Methods

Russell's viper shed skin and DNA isolation

Freshly shed skin of Russell's viper from Bangalore, India was a gift from Mr. Gerry Martin. The shed skin for the entire snake was obtained, cleaned thoroughly with 70% ethanol and with nuclease-free water 3 times each, dried thoroughly and frozen until the time of extraction of DNA. Genomic DNA was extracted following the protocol of Fetzner (Fetzner 1999) with modifications.

Sequencing, read processing and assembly

Illumina paired-end read libraries (100 base paired-end reads with insert size of 350 bases) were prepared following the manufacturer instructions using amplification free genomic DNA library preparation kit and sequenced using Illumina HiSeq2500 instrument. Archaeal, bacterial and human sequence contamination were removed

from the Russell's viper sequence by DeConSeq (Schmieder & Edwards 2011) using curated and representative genomes (https://www.ncbi.nlm.nih.gov/genome/browse/reference/). Furthermore, the sequenced reads were post-processed to remove unpaired reads and quality analysis was performed using FastQC v0.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The rd_len_cutoff option was exercised during the read assembly step to trim off the low-quality bases, since the per-base quality was found to drop below 28 after the initial 50-70 bases of the read. The Russell's viper read libraries with 26X coverage were assembled using SOAPdenovo2 (r240) (Luo et al. 2012).

Identifying toxin gene homologs, coding regions, and predicted gene structures

The DNA sequences for 51 out of 54 venom-associated genes (Fry 2005) from *Protobothrops mucrosquamatus* were downloaded (Table 1). These were used to fish genomic scaffolds bearing highly similar sequences in Russell's viper genome assembly, using BLAST with an E-value threshold of 10^{-3}. The fished scaffolds were then anchored to the respective coding sequences from *Protobothrops mucrosquamatus*, using a discontiguous megaBlast, to determine the correct frame of translation and extract the complete amino acid coding sequence (CDS) corresponding to putative toxin homologs in Russell's viper. We obtained the exon-intron structures for all the putative toxin gene homologs in Russell's viper by aligning the CDS with gene sequences using discontiguous megaBlast and plotted using the tool GSDS2.0 (Hu et al. 2015). The sequences for the Russell's viper putative venom gene homologs were deposited in GenBank and their accession numbers are provided in Table S1.

Comparative analysis between venom proteins and their putative homologs

We obtained the accession IDs for the major toxin families from Russell's viper of the Indian sub-continent (Supplementary Figure S1 in Sharma et al. 2015). Their corresponding protein sequences were matched using blastp with the amino acid sequences from the putative skin homologs. For the genes covered under each family, a percent identity metric, indicative of the extent of sequence similarity between the venom proteins and their skin homologs, was estimated. Similar comparative analyses were performed for king cobra (*Ophiophagus hannah*) using accession IDs provided in Additional File 4 of Tan et al. 2015, and the predicted toxin protein homologs from blood of king cobra (PRJNA201683 from Vonk et al. 2013). Comparative analyses were performed using blastp, with the venom protein sequence as the query, against PRJNA201683.

Comparative analyses of putative venom protein homolog domains

The amino acid sequences of all the Russell's viper's putative toxin homologs were subjected to domain search using Pfam (Finn et al. 2016) (Table S2). All domain sequences were aligned using blastp to non-redundant protein sequences from 18 snake species (Table S3). We wanted to compare the gene structures between the venomous and the non-venomous animals, hence included sequence information from the members of the later group. Five domains (NGF, PDGF, Kunitz BPTI, CAP and CRISP) from four genes (NGF, VEGF, CRISP/Serotriflin, and Kunitoxin) with variability across different snake groups were used for expansive comparative analyses (Table S4). The sequences used were from viperids (taxid: 8689), elapids (taxid: 8602), colubrids (taxid: 8578), boids (taxid: 8572), acrochordids (taxid: 42164), pythonids (taxid: 34894), lizards (squamates (taxid: 8509) minus snakes (taxid: 8570)), crocodiles (taxid: 51964) and testudines (taxid: 8459).

3D structure prediction of the chosen domains

Consensus sequences were determined from NGF, PDGF, Kunitz BPTI, CAP and CRISP domain alignments using Simple Consensus Maker (https://www.hiv.lanl.gov/content/sequence/CONSENSUS/SimpCon.html) for crotalines (CR), viperines (VP) and elapids. The consensus sequences were submitted to the protein fold recognition server (Kelley et al. 2015) using standard mode (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). The best 3D model was further investigated by Phyre2 to analyze the structural model using various open source tools.

Results

Shed skin yielded fairly good quality and high-molecular weight genomic DNA

Genomic DNA isolated from the shed skin of Russell's viper was fairly intact with most of the DNA in the size range of more than 5kbp (Fig. S1). Sequenced short reads were assembled and then used to fish the sequences for the 51 putative toxin genes in Russell's viper (see Materials and Methods). Next, we obtained the exon-intron structures for all putative homologs in Russell's viper by aligning the CDS with gene sequences (Fig. S2). We found the average length of the exons in those sequences to be around 190 nucleotides (nt), matching well with the lengths of other vertebrate exons (Gelfman et al. 2012).

Similarity between venom proteins and their putative skin homologs

For the Russell's viper, we found 45 - 100% sequence similarity between the major venom proteins and their predicted putative skin homologs (Figure 1). The sequences for venom nerve growth factor (VNGF) and its putative skin homolog were identical. Similarly, VEGF and CRISPs from venom gland were highly similar to their putative skin homologs (99% and 92% sequence similarity respectively). Other proteins like, KSPI, SVSPs and PLA2 showed 79%, 74% and 61% sequence identity,

181 respectively (Figure 1 and Fig. S3). In order to find out whether the sequence
182 divergence between some of the venom gland proteins and their predicted putative skin
183 homologs was specific to Russell's viper, we performed similar analysis using venom
184 proteins and their blood homologs from king cobra, *Ophiophagus hannah* (Vonk et al.,
185 2013). In the case of *Ophiophagus hannah*, the differences between toxin proteins and
186 their blood homologs were minor for most families studied (similarity >= 75%), except for
187 PLA2, which had a low similarity of 23% (Fig. S4 and Fig. S5).
188 *Comparative domain analyses*
189 Among the genes, a larger pool of sequences were available only for NGF, PDGF
190 domain of VEGF, Kunitz_BPTI domain of Kunitoxin, CRISP and CAP domains in CRISP
191 and Serotriflin proteins, from various snake groups (Colubridae, Boidae, Pythonidae and
192 Acrochordidae), non-snake reptilian groups (lizards, crocodiles and Testudines),
193 venomous invertebrates (wasps, spiders and scorpions) and venomous vertebrates
194 (fishes and mammals). Therefore, these domains were compared with their putative
195 homologs from Russell's viper. Comparative domain analysis was performed for all
196 putative toxin gene homologs (Fig. S6) across 18 snake species for those where
197 sequence information was available (Table S3). In the case of five domains: CAP and
198 CRISP domains of CRISP and serotriflin genes (L and AL), Kunitz BPTI of kunitoxin (S),
199 NGF (T) and PDGF of VEGFA (AP-AR) and VEGFF (AU), we found that the maximum
200 number of species aligned to their domain sequences. Some protein domains, the
201 CRISP, Kunitz BPTI, guanylate CYC, PDGF of VEGFF and WAP, showed long stretches
202 of mismatches (Fig. S6) compared with Russell's viper sequence. Out of these, only
203 NGF and PDGF domains of VEGF had amino acid changes specific to the members of
204 the group Crotalinae, that were completely absent in any other group used for

comparison, including in lizards, crocodiles, and turtles (Fig. S7). Specific changes in these proteins and their implications are discussed below.

The putative skin-derived NGF gene homolog in Russell's viper is a single exon gene with a 745nt transcript coding for a 244 amino acid protein consisting of a single NGF domain (Figure 2A). The NGF domain bears 28% sequence conservation across all the five vertebrate phyla, namely, fishes, amphibians, reptiles, aves and mammals distributed along the length of the domain (Figure 2B). Thirty-six percent out of these residues are conserved across other venomous vertebrates (fishes, squamates and mammals) and venomous invertebrates (scorpions and wasps) (Figure 2C). Thirteen percent of the skin-derived putative NGF domain residues are variable with respect to the domain sequence in at least one among the NGF sequences in the groups of viperids and elapids (Figure 2D). Although several amino acids in the NGF domain in crotalines seem to have changed from the putative domain in Russell's viper and other vipers of the group Viperinae, their function probably remains unchanged. For example, phenylalanine (F) to isoleucine (I) at position 12 and serine (S) to asparagine (N) at position 19 between the crotalines and viperines does not change the function of the amino acids (from one hydrophobic amino acid to another and from one polar amino acid to another). However, it is also true that F changing to I removes the bulky aromatic ring, whereas S could be a phosphorylated site as opposite to the N in the same position. There are others, for example, threonine (T) and glutamine (Q), at positions 67 and 68, respectively, in the NGF domain of the crotalines, which were only there in that specific group. One of those, a polar amino acid glutamine at position 68, is a very important residue as its corresponding amino acid in any of the other snakes, except in colubrids, is a hydrophobic proline.

In Russell's viper, the putative skin-derived VEGFA gene homolog comprises five exons coding for a 652nt long transcript and a translated protein with two domains: PDGF and VEGF-C (Figure 3A). The PDGF domain sequence exhibits conservation in 65% of its residues across the three vertebrate phyla (reptiles, birds and mammals) (Figure 3B). Since sequence information from fishes and amphibians were not available, we could not include those in the comparison study. Out of the conserved residues in the above said domain, 21% were also conserved in venomous vertebrates (squamates and mammals) and venomous invertebrates (wasps). Fifteen percent of the PDGF domain residues were variable in at least one of the two snake groups: viperids and elapids (Figure 3C and Figure 3D). Like the NGF domain, the evolution of the putative skin-derived PDGF domain in crotalines at certain amino acids is striking. For example, in the crotalines, the position 67 is a polar amino acid tyrosine (Y) while in all other reptiles, venomous invertebrates and mammals; this is primarily a hydrophobic amino acid phenylalanine (F).

The putative skin-derived Kunitoxin gene homolog in Russell's viper is a 3.1kb gene comprising two exons, with a transcript length of 270nt that codes for a 44 amino acids long single Kunitz BPTI domain (Figure 4A). About 29% of the protein domain residues are conserved across the four vertebrate phyla (amphibians, reptiles, aves and mammals) (Figure 4B). Since sequence information from the Kunitz BPTI for fishes was not available, we could not include those in the comparison. Out of these conserved residues, 76% are conserved in venomous vertebrates (squamates and mammals) and venomous invertebrates (scorpions and wasps) (Figure 4C) and 56% of the domain residues are variable in at least one of two snake groups (viperids, and elapids) (Figure 4D). Of the residues that are evolved in the members of Crotalinae, the second residue, a positively charged one, arginine (R) is present only in the members of Viperinae, which

is replaced by a hydrophobic residue, proline (P), in the crotalines and elapids. Residues 14-18 are very polymorphic in the crotalines and elapids, but not so in the viperines.

The skin-derived putative CRISP gene homolog in Russell's viper is a 25kb long gene, comprises of 8 exons coding for a 787nt transcript and two translated protein domains, CAP and CRISP (Figure 5A). The CAP domain exhibits conservation in 7% of its residues across all the five vertebrate phyla (Figure 5B). Forty-two percent of those residues are conserved across venomous vertebrates (amphibians, squamates and mammals) and venomous invertebrates (scorpions and wasps) (Figure 5C). In addition, there are five residues conserved across all the venomous animals (Figure 5C). Twenty-seven percent of the CAP domain residues are variable in at least one of the three snake groups (Figure 5D). There are several extra residues for the CAP domain in the crotalines and elapids, but not in the viperines. The conserved residues comprised mostly of Cystines and to a lesser extent Asparagines (Figure 5E) across venomous vertebrates (squamates and mammals) (Figure 5F). Sixty percent of the residues in the putative homolog of the CRISP domain are variable in at least one viperine or elapid member with respect to the domain sequence of Russell's viper (Figure 5G).

Next, we explored the role of consensus domain sequences in the putative protein homologs and the possible role of conserved amino acids in those domains across viperids and elapids. We constructed the 3D structure models using Phyre2, followed by Phyre2 investigation, for further analyses on the structural model. As evident from the analyses, amino acid residues 18-19 and 117 of the NGF domain reflected a difference in mutation sensitivity as detected by SusPect algorithm (Yates et al. 2014), especially in the elapids compared to the viperids (Figure 6). There were differences in certain residues across these two groups. Residue 18 is Valine in the viperines and Isoleucine in the elapids; residue 19 is Serine in the viperines and Asparagine in the

crotalines; and residue 117 is Threonine in the elapids and Serine in the crotalines (Figure 6A). This might have implications in the structure of the protein as the largest pockets detected by fpocket algorithm appear to be vastly different among the crotalines, viperines and elapids for the NGF, PDGF, CAP and CRISP domains (Figure 6). The pockets appeared smallest in all cases for the elapids (quantitatively substantial for PDGF, CAP and CRISP: one-fourth that of viperines for PDGF, and two-thirds that of viperines for CAP and CRISP), and largest in the case of viperines (Figure 6). Minor differences in clashes were observed at residues 10,11 and 20 of the Kunitz domain and residue 38 of this domain showed a rotamer conflict in the case of the crotalines (Figure 6C). Similarly, residue 46 of the CAP domain and residues 4 and 31 of the CRISP domain showed rotamer conflict for the viperines (Figure 6D and Figure 6E). The other protein quality and functional parameters were not affected across the 3D structure models for the three snake groups (Fig. S8).

Discussion

Accessibility and affordability of high-throughput sequencing technologies along with the availability of sophisticated computational tools to assemble, annotate and interpret genomes is playing a powerful role in deciphering gene functions and their role in evolution. Snake toxin genes are coded by gene families and produce gene isoforms through the process of duplications (Casewell et al. 2013; Fry 2005). Several studies on the venom-associated proteins from New World vipers have classified the venoms into four groups (type I-IV), based on the relative abundance of toxin families (Calvete 2013; Gibbs et al. 2013; Goncalves-Machado et al. 2016; Jimenez-Charris et al. 2015; Lomonte et al. 2014; Mora-Obando et al. 2014; Pla et al. 2017; Salazar-Valenzuela et al. 2014). The different groups are: snake venom metalloproteinase-predominant (type I), heterodimeric β -neurotoxic PLA2 –rich (type II), serine proteinases and PLA2 (type III)

and type IV, which is similar to type III but with significant higher concentration of snake venom metalloproteinases (Calvete 2017). Russell's viper (*Daboia russelii*) is a Old World pitless viper, characterized by the lack of heat sensing pit organs (Mallow et al. 2003).

There is significant variation in the venom composition of Russell's viper in India (Jayanthi & Gowda 1988; Sharma et al. 2015), making the universal anti-venom less effective against all Russell's viper bites across the country. The variation in the venom composition within the same species is thought to be a result of adaptation in response to the difference in diets (Barlow et al. 2009; Casewell et al. 2013; Daltry et al. 1996). A comparison across four published studies (Kalita et al. 2017; Mukherjee et al. 2016; Sharma et al. 2015; Tan et al. 2015) on Russell's viper venom proteins revealed that the composition of some of the major venom proteins varied significantly (Figure S9). For example, in one study (Mukherjee et al. 2016), VNGF constituted only 0.4% of the venom while in another (Kalita et al. 2017), the same protein constituted 4.8% of the venom. As both studies came from the same lab, there is little chance for any technical or assay-related variability. In the first study, the venom was used from the captive species in a zoo in the USA where the snake was from a Pakistani origin (Mukherjee et al. 2016) while the other study used venom from a commercial source in India (Kalita et al. 2017). This suggests that there is a great deal of variation in the composition of Russell's viper venom collected from different locations, corroborating the earlier results (Jayanthi & Gowda 1988; Sharma et al. 2015). Currently, efforts are underway to collect venoms of Russell's viper from different regions of India in order to understand their composition (Rom Whitaker, and Gerry Martin, *personal communications*).

Studies on venom-associated genes using whole-genome sequencing data in Russell's viper are scarce. One of the reasons is the relative difficulty in accessing

venom glands from snakes. This can be partially addressed by studying their putative homologs from shed skin, which is relatively easy to access. Past studies on the members of Viperidae focused on proteins and used proteomics-based analyses (Gao et al. 2013; Gao et al. 2014; Kalita et al. 2017; Li et al. 2004; Liu et al. 2011; Mukherjee et al. 2016; Sharma et al. 2015; Tan et al. 2017; Tan et al. 2015; Villalta et al. 2012). The only viperine where complete genome sequence information is available is a European adder, *Vipera berus berus* (https://www.ncbi.nlm.nih.gov/bioproject/170536). Although sequence information is available for this species, the annotation is not available and therefore could not be used in our study.

The aim of the current study was two fold. First, as handling and getting biological material from snakes requires specific expertise, we wanted to test whether one can obtain high-molecular weight DNA from shed skin as a source of analyte for genome sequencing. Second, we wanted to study the potential of skin-derived putative toxin gene homologs, as surrogates of their venom gland counterparts through comparative analyses. On the first account, we found the results to be satisfactory. Although shed skin is often contaminated with bacteria and other microorganisms, and the DNA obtained from the shed skin is sheared, we show that one can successfully isolate high-molecular weight genomic DNA from shed skin (Fig. S1). Therefore, shed skin may be an attractive option in the future for generating snake genome data. We showed that the comparisons of amino acid sequence and three dimensional structures of five toxin domains with their putative skin homologs across the major kingdoms of life can generate important information towards understanding the macro- and micro-evolution of these genes. Results from our comparative analyses showed that some of the venom gland proteins are identical or near identical to their putative skin-derived homologs (VNGF, 3FTX and LAAO) but others had low overall similarity (Snaclec and RVV). We

were curious to find out whether the low sequence similarity for some venom proteins with their putative homologs was specific to Russell's viper and how much of the low overall similarity in those proteins was due to the heterogeneity, if any, found among snakes of the same species. Comparative analysis between the toxins and their blood homologs in king cobra provided us with an answer for the first question where 7 out of 8 venom proteins studied (except for PLA2) were very similar to their blood homologs (Fig. S4 and Fig. S5). This suggests that some venom proteins may not be that different from their homologs in other organs. A recent study in python, where the authors argue that the functional evidence of toxic effects on prey and not their expression is the correct criterion to classify proteins as venom toxins (Reyes-Velasco et al. 2015), strengthens this hypothesis further. However, we are aware that this may vary from species to species and in some species the venom proteins may be very different from their homologs in other organs.

In our study, we compared venom proteins described previously (Sharma et al. 2015) using animals captured near Chennai, Tamil Nadu, India with their skin-derived putative homologs from a completely different animal (shed skin was collected in Bangalore, Karnataka, India). The distance between these two places is roughly 350-400km. Therefore; it is possible that in our study, the low similarity in some of the venom proteins with their putative skin homologs could have been due to the variation in the venom composition of the animals in these two locations. Despite this, 50% of the venom proteins studied had >75% and 3 had near perfect sequence similarity with their putative skin homologs. A clear picture will emerge from a direct comparison between the venom proteins and their skin counterparts from the same animal.

From the sequence data, we succeeded in assembling near complete CDS for 20 gene families representing 51 gene homologs (Fry 2005). This highlights the utility of

379 genome sequencing data in inferring putative toxin gene homologs. As the lengths of the
380 putative toxin gene homologs in Russell's viper were much longer than the CDS, the
381 intronic sequences were assembled with gaps. This was primarily due to the low
382 coverage sequencing data used for assembly and the lack of long-insert mate pair
383 sequencing data in our repertoire. The aim of the current study was not to assemble a
384 perfect genome for Russell's viper but to use the low coverage data to fish out putative
385 toxin gene homologs from skin. The mean length of exons for the putative toxin gene
386 homologs in Russell's viper was 190 bases, much smaller compared to the average
387 intron length. In our study, we could assemble exons accurately using short-read
388 sequence data. Interestingly, we found that the AT to GC ratio in the CDS regions
389 (cumulatively for all the 51 toxin gene homologs) was 1:1 whereas it was skewed (the
390 ratio is 1.5:1) for the full gene sequences.

391 Our study demonstrates the feasibility of *de novo* sequencing and analyses of
392 gene families without prior sequence information and annotation, and without going
393 through the process of designing individual primers for Sanger sequencing. However,
394 there are certain limitations to our study. First, it focuses on the putative toxin gene
395 homologs from skin and not the toxin genes from venom gland. There is a possibility that
396 the toxin genes from the same animal in the venom gland are different from their
397 homologs in the skin, and therefore can only be described as putative. Hence, we can
398 neither be sure of the presence nor the activity of the homologs in the skin. Although the
399 skin-derived transcriptome data will add value to the study, it will still be inadequate.
400 Future transcriptome and proteomics analysis from both the skin and the venom gland,
401 preferably from the same animal, along with their functional studies will only be
402 definitive. Second, like any other annotation-based study, we relied on the quality of
403 existing/prior annotation of toxin-related genes. It is possible that due to sequencing

artefacts, there are errors in the assembled genomic sequence, and therefore in the translated protein sequences inferred in our study. Future studies using high-coverage sequencing data along with data on RNA and protein to derive better gene annotations along with the functional studies on their spatial and temporal expression will point to the true functional significance of skin-derived toxin gene homologs.

Acknowledgements

We thank Kunal Dhas and Arun Hariharan for their help in laboratory experiments.

Literature Cited

Barlow A, Pook CE, Harrison RA, and Wuster W. 2009. Coevolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution. *Proc Biol Sci* 276:2443-2449.

Calvete JJ. 2013. Snake venomics: from the inventory of toxins to biology. *Toxicon* 75:44-62.

Calvete JJ. 2017. Venomics: integrative venom proteomics and beyond. *Biochem J* 474:611-634.

Casewell NR, Wagstaff SC, Wuster W, Cook DA, Bolton FM, King SI, Pla D, Sanz L, Calvete JJ, and Harrison RA. 2014. Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. *Proc Natl Acad Sci U S A* 111:9205-9210.

Casewell NR, Wuster W, Vonk FJ, Harrison RA, and Fry BG. 2013. Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol Evol* 28:219-229.

Castoe TA, de Koning AP, Hall KT, Card DC, Schield DR, Fujita MK, Ruggiero RP, Degner JF, Daza JM, Gu W, Reyes-Velasco J, Shaney KJ, Castoe JM, Fox SE, Poole AW, Polanco D, Dobry J, Vandewege MW, Li Q, Schott RK, Kapusta A, Minx P, Feschotte C, Uetz P, Ray DA, Hoffmann FG, Bogden R, Smith EN, Chang BS, Vonk FJ, Casewell NR, Henkel CV, Richardson MK, Mackessy SP, Bronikowski AM, Yandell M, Warren WC, Secor SM, and Pollock DD. 2013. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc Natl Acad Sci U S A* 110:20645-20650.

Daltry JC, Wuster W, and Thorpe RS. 1996. Diet and snake venom evolution. *Nature* 379:537-540.

Dowell NL, Giorgianni MW, Kassner VA, Selegue JE, Sanchez EE, and Carroll SB. 2016. The Deep Origin and Recent Loss of Venom Toxin Genes in Rattlesnakes. *Curr Biol* 26:2434-2445.

Fetzner JW, Jr. 1999. Extracting high-quality DNA from shed reptile skins: a simplified method. *Biotechniques* 26:1052-1054.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, and Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* 44:6.

Fry BG. 2005. 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* 15:403-420.

Gao JF, Qu YF, Zhang XQ, He Y, and Ji X. 2013. Neonate-to-adult transition of snake venomics in the short-tailed pit viper, Gloydius brevicaudus. *J Proteomics* 84:148-157.

Gao JF, Wang J, He Y, Qu YF, Lin LH, Ma XM, and Ji X. 2014. Proteomic and biochemical analyses of short-tailed pit viper (Gloydius brevicaudus) venom: age-related variation and composition-activity correlation. *J Proteomics* 105:307-322.

Gelfman S, Burstein D, Penn O, Savchenko A, Amit M, Schwartz S, Pupko T, and Ast G. 2012. Changes in exon-intron structure during vertebrate evolution affect the splicing pattern of exons. *Genome Res* 22:35-50.

Gibbs HL, Sanz L, Sovic MG, and Calvete JJ. 2013. Phylogeny-based comparative analysis of venom proteome variation in a clade of rattlesnakes (*Sistrurus* sp.). *PLoS One* 8:e67220.

Goncalves-Machado L, Pla D, Sanz L, Jorge RJ, Leitao-De-Araujo M, Alves ML, Alvares DJ, De Miranda J, Nowatzki J, de Morais-Zani K, Fernandes W, Tanaka-Azevedo AM, Fernandez J, Zingali RB, Gutierrez JM, Correa-Netto C, and Calvete JJ. 2016. Combined venomics, venom gland transcriptomics, bioactivities, and antivenomics of two *Bothrops jararaca* populations from geographic isolated regions within the Brazilian Atlantic rainforest. *J Proteomics* 135:73-89.

Hargreaves AD, Swain MT, Hegarty MJ, Logan DW, and Mulley JF. 2014. Restriction and recruitment-gene duplication and the origin and evolution of snake venom toxins. *Genome Biol Evol* 6:2088-2095.

Hu B, Jin J, Guo AY, Zhang H, Luo J, and Gao G. 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31:1296-1297.

Jayanthi GP, and Gowda TV. 1988. Geographical variation in India in the composition and lethal potency of Russell's viper (*Vipera russelli*) venom. *Toxicon* 26:257-264.

Jimenez-Charris E, Montealegre-Sanchez L, Solano-Redondo L, Mora-Obando D, Camacho E, Castro-Herrera F, Fierro-Perez L, and Lomonte B. 2015. Proteomic and functional analyses of the venom of *Porthidium lansbergii lansbergii* (Lansberg's hognose viper) from the Atlantic Department of Colombia. *J Proteomics* 114:287-299.

Kalita B, Patra A, and Mukherjee AK. 2017. Unraveling the Proteome Composition and Immuno-profiling of Western India Russell's Viper Venom for In-Depth Understanding of Its Pharmacological Properties, Clinical Manifestations, and Effective Antivenom Treatment. *J Proteome Res* 16:583-598.

Kelley LA, Mezulis S, Yates CM, Wass MN, and Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10:845-858.

Kerkkamp HM, Kini RM, Pospelov AS, Vonk FJ, Henkel CV, and Richardson MK. 2016. Snake Genome Sequencing: Results and Future Prospects. *Toxins (Basel)* 8.

Li S, Wang J, Zhang X, Ren Y, Wang N, Zhao K, Chen X, Zhao C, Li X, Shao J, Yin J, West MB, Xu N, and Liu S. 2004. Proteomic characterization of two snake venoms: *Naja naja atra* and *Agkistrodon halys*. *Biochem J* 384:119-127.

Liu S, Yang F, Zhang Q, Sun MZ, Gao Y, and Shao S. 2011. "Anatomical" view of the protein composition and protein characteristics for *Gloydius shedaoensis* snake venom via proteomics approach. *Anat Rec (Hoboken)* 294:273-282.

Lomonte B, Fernandez J, Sanz L, Angulo Y, Sasa M, Gutierrez JM, and Calvete JJ. 2014. Venomous snakes of Costa Rica: biological and medical implications of their venom proteomic profiles analyzed through the strategy of snake venomics. *J Proteomics* 105:323-339.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, and Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18.

Malhotra A, Creer S, Harris JB, and Thorpe RS. 2015. The importance of being genomic: Non-coding and coding sequences suggest different models of toxin multi-gene family evolution. *Toxicon* 107:344-358.

Mallow D, Ludwig D, and Nilson G. 2003. *True vipers : natural history and toxinology of Old World vipers*. Malabar, Fla.: Krieger Pub. Co.

Mohapatra B, Warrell DA, Suraweera W, Bhatia P, Dhingra N, Jotkar RM, Rodriguez PS, Mishra K, Whitaker R, Jha P, and Million Death Study C. 2011. Snakebite mortality in India: a nationally representative mortality survey. *PLoS Negl Trop Dis* 5:e1018.

Mora-Obando D, Diaz C, Angulo Y, Gutierrez JM, and Lomonte B. 2014. Role of enzymatic activity in muscle damage and cytotoxicity induced by Bothrops asper Asp49 phospholipase A2 myotoxins: are there additional effector mechanisms involved? *PeerJ* 2:e569.

Mukherjee AK, Kalita B, and Mackessy SP. 2016. A proteomic analysis of Pakistan *Daboia russelii russelii* venom and assessment of potency of Indian polyvalent and monovalent antivenom. *J Proteomics* 144:73-86.

Pla D, Sanz L, Sasa M, Acevedo ME, Dwyer Q, Durban J, Perez A, Rodriguez Y, Lomonte B, and Calvete JJ. 2017. Proteomic analysis of venom variability and ontogeny across the arboreal palm-pitvipers (genus *Bothriechis*). *J Proteomics* 152:1-12.

Reyes-Velasco J, Card DC, Andrew AL, Shaney KL, Adams RH, Schield DR, Casewell NR, Mackessy SP, Castoe TA. 2015. Expression of Venom Gene Homologs in Diverse Python Tissues Suggests a New Model for the Evolution of Snake Venom. *Mol Biol Evol* 32:173–183.

Rokyta DR, Wray KP, Lemmon AR, Lemmon EM, and Caudle SB. 2011. A high-throughput venom-gland transcriptome for the Eastern Diamondback Rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes. *Toxicon* 57:657-671.

Salazar-Valenzuela D, Mora-Obando D, Fernandez ML, Loaiza-Lange A, Gibbs HL, and Lomonte B. 2014. Proteomic and toxicological profiling of the venom of *Bothrocophias campbelli*, a pitviper species from Ecuador and Colombia. *Toxicon* 90:15-25.

Schmieder R, and Edwards R. 2011. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One* 6:e17288.

Sharma M, Das D, Iyer JK, Kini RM, and Doley R. 2015. Unveiling the complexities of *Daboia russelii* venom, a medically important snake of India, by tandem mass spectrometry. *Toxicon* 107:266-281.

538 Tan CH, Tan KY, Yap MK, and Tan NH. 2017. Venomics of *Tropidolaemus wagleri*, the
539 sexually dimorphic temple pit viper: Unveiling a deeply conserved atypical toxin
540 arsenal. *Sci Rep* 7:43237.
541 Tan NH, Fung SY, Tan KY, Yap MK, Gnanathasan CA, and Tan CH. 2015. Functional
542 venomics of the Sri Lankan Russell's viper (*Daboia russelii*) and its toxinological
543 correlations. *J Proteomics* 128:403-423.
544 Ullate-Agote A, Milinkovitch MC, and Tzika AC. 2014. The genome sequence of the corn
545 snake (*Pantherophis guttatus*), a valuable resource for EvoDevo studies in
546 squamates. *Int J Dev Biol* 58:881-888.
547 Vicoso B, Emerson JJ, Zektser Y, Mahajan S, and Bachtrog D. 2013. Comparative sex
548 chromosome genomics in snakes: differentiation, evolutionary strata, and lack of
549 global dosage compensation. *PLoS Biol* 11:e1001643.
550 Villalta M, Pla D, Yang SL, Sanz L, Segura A, Vargas M, Chen PY, Herrera M, Estrada
551 R, Cheng YF, Lee CD, Cerdas M, Chiang JR, Angulo Y, Leon G, Calvete JJ, and
552 Gutierrez JM. 2012. Snake venomics and antivenomics of *Protobothrops*
553 *mucrosquamatus* and *Viridovipera stejnegeri* from Taiwan: keys to understand the
554 variable immune response in horses. *J Proteomics* 75:5628-5645.
555 Vonk FJ, Casewell NR, Henkel CV, Heimberg AM, Jansen HJ, McCleary RJ, Kerkkamp
556 HM, Vos RA, Guerreiro I, Calvete JJ, Wuster W, Woods AE, Logan JM, Harrison
557 RA, Castoe TA, de Koning AP, Pollock DD, Yandell M, Calderon D, Renjifo C,
558 Currier RB, Salgado D, Pla D, Sanz L, Hyder AS, Ribeiro JM, Arntzen JW, van
559 den Thillart GE, Boetzer M, Pirovano W, Dirks RP, Spaink HP, Duboule D,
560 McGlinn E, Kini RM, and Richardson MK. 2013. The king cobra genome reveals
561 dynamic gene evolution and adaptation in the snake venom system. *Proc Natl*
562 *Acad Sci U S A* 110:20651-20656.
563 Vonk FJ, Jackson K, Doley R, Madaras F, Mirtschin PJ, and Vidal N. 2011. Snake
564 venom: From fieldwork to the clinic: Recent insights into snake biology, together
565 with new technology allowing high-throughput screening of venom, bring new
566 hope for drug discovery. *Bioessays* 33:269-279.
567 Warrell DA, Gutierrez JM, Calvete JJ, and Williams D. 2013. New approaches &
568 technologies of venomics to meet the challenge of human envenoming by
569 snakebites in India. *Indian J Med Res* 138:38-59.
570 Whitaker R. 2015. Snakebite in India today. *Neurol India* 63:300-303.
571 Yates CM, Filippis I, Kelley LA, and Sternberg MJ. 2014. SuSPect: enhanced prediction
572 of single amino acid variant (SAV) phenotype using network features. *J Mol Biol*
573 426:2692-2701.
574 Yin W, Wang ZJ, Li QY, Lian JM, Zhou Y, Lu BZ, Jin LJ, Qiu PX, Zhang P, Zhu WB, Wen
575 B, Huang YJ, Lin ZL, Qiu BT, Su XW, Yang HM, Zhang GJ, Yan GM, and Zhou Q.
576 2016. Evolutionary trajectories of snake genes and genomes revealed by
577 comparative analyses of five-pacer viper. *Nat Commun* 7:13107.

578 **Figure Legends**

579 **Figure 1:** Sequence identity (%) between the proteins from ten major venom families
580 and their putative skin homologs in Russell's viper. The homolog with the highest identity
581 was considered where more than one homolog was present.

582 **Figure 2:** Comparative analyses of nerve growth factor (NGF). Putative NGF gene
583 homolog, its mRNA, and protein domains in Russell's viper (A) and its comparison with
584 the consensus NGF sequences from all five vertebrate phyla (fishes, amphibians,
585 reptiles, birds and mammals) (B), with venomous (V) vertebrates from multiple phyla of
586 vertebrates and invertebrates (C), and from various reptilian subgroups (D) are shown.
587 The shades of brown and grey in B and C represent conservation to various degrees
588 and variability, respectively. Grey in D represents conserved residues, red represents
589 variable residues in the crotalines (CR), yellow and green represent conserved and
590 variable residues in the viperines (VP), and elapids respectively.

591 **Figure 3:** Comparative analyses of vascular endothelial growth factor - A (VEGF-A).
592 Organization of the putative gene homolog, its mRNA, and protein domains of Russell's
593 viper PDGF domain (A) and its comparison with the consensus sequences from all five
594 vertebrate phyla (fishes, amphibians, reptiles, birds and mammals) (B), from the
595 venomous (V) vertebrates and invertebrates (C), and from various reptilian subgroups
596 (D) are shown. The shades of brown and grey in B and C represent conserved and
597 varying residues, respectively. Grey in D represents conserved residues, red represents
598 variable residues in the crotalines (CR), yellow and green represent conserved and
599 variable residues in viperines (VP), and elapids respectively.

600 **Figure 4:** Comparative analyses of kunitoxin. Organization of the putative gene
601 homolog, its mRNA, and protein domains of Russell's viper (A) and its comparison with
602 the consensus BPTI domain sequences from all five vertebrate phyla (fishes,

603 amphibians, reptiles, birds and mammals) (B), from venomous (V) vertebrates and
604 invertebrates (C), from various reptilian subgroups (D) are shown. The shades of brown
605 and grey in B and C represent conserved and varying residues, respectively. Grey in D
606 represents conserved residues, red represents variable residues in the crotalines (CR),
607 yellow and green represent conserved and variable residues in viperines (VP), and
608 elapids respectively.
609 **Figure 5:** Comparative analyses of CRISP. Organization of putative CRISP gene
610 homolog, its mRNA, and protein domains of Russell's viper (A) and its comparison with
611 the consensus CRISP sequences from all five vertebrate phyla (fishes, amphibians,
612 reptiles, birds and mammals, B and E); from venomous animals (V) vertebrates (fishes,
613 squamates and mammals) and invertebrates (scorpions and wasps, C and F); and from
614 various reptilian subgroups (D and G) are shown. The shades of brown and grey in B, C,
615 E and F represent conserved and varying residues, respectively. Grey in D and G
616 represents conserved residues, red represents variable residues in the crotalines (CR),
617 yellow and green represent conserved and variable residues in viperines (VP), and
618 elapids respectively.
619 **Figure 6:** Three-dimensional protein structural models in NGF (A); PDGF (B); Kunitz
620 BPTI (C); CAP (D); and CRISP (E) across crotalines (CR), viperines (VP) and elapids.
621 The status of the parameters being investigated using Phyre2 are indicated in the color
622 legends on the side.

Table 1(on next page)

Genes and their representative families used in the current study.

The homolog with the highest identity was considered in cases with more than one homolog.

Gene	Species with the available sequence information	Protein Family
ACHE	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	Acetylcholinesterase
ADAM11		ADAM (disintegrin/ metalloprotease)
ADAM17		
ADAM19		
ADAM23		
PROK1	*Protobothrops mucrosquamatus, Ophiophagus hannah* and *Python bivitattus*	AVIT (prokinectin)
PROK2	*Protobothrops mucrosquamatus and Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	
CPAMD8	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	Complement C3
crotasin	*Protobothrops mucrosquamatus*	Crotasin/ beta defensin
CST1	No sequence information is available in any of the four species	Cystatin
CST3	*Ophiophagus hannah*	
CST4	No sequence information is available in any of the four species	
CSTA	*Protobothrops mucrosquamatus* and *Thamnophis sirtalis*	
EDN1	*Protobothrops mucrosquamatus* and *Python bivittatus*	Endothelin
EDN3	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus*, *Thamnophis sirtalis* and	
F5		Factor V
F10	*Ophiophagus hannah*	Factor X
KLKB1	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus, Thamnophis sirtalis* and	Kallikrein
KLK14	*Ophiophagus hannah*	
kunitoxin	*Protobothrops mucrosquamatus, Python bivittatus* and *Ophiophagus hannah*	Kunitz-type protease inhibitor
LYNX1	*Ophiophagus hannah*	LYNX/SLUR
CLEC3A	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	Lectin
CLEC3B	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	
CLEC11A	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	
CLEC16A	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	

CLEC19A	*Protobothrops mucrosquamatus and Python bivittatus*	
NPR1	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	Natriuretic peptide
NPR2	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	
NPR3		
NGF	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus, Thamnophis sirtalis, Protobothrops flavoviridis, Crotalus horridus, Sistrurus miliarius barbouri* and *Boa constrictor*	Beta-nerve growth factor
PLAA	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	Phospholipase A (2)
PLA2R1		
PLA2G1B	*Python bivittatus, Thamnophis sirtalis* and *Protobothrops mucrosquamatus*	
PLA2G10	*Protobothrops mucrosquamatus, Protobothrops flavoviridis, Thamnophis sirtalis, Ophiophagus hanna* and *Python bivittatus*	
PLA2G12A	*Python bivittatus, Thamnophis sirtalis* and *Protobothrops mucrosquamatus*	
PLA2G12B	*Protobothrops mucrosquamatus, Ophiophagus hannah* and *Python bivittatus*	
PLA2G15	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus and Thamnophis sirtalis*	
PLA2G3	*Protobothrops mucrosquamatus, Python bivittatus* and *Ophiophagus hannah*	
PLA2G4A	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	
PLA2G4C	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	
PLA2G6	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	
PLA2G7		
SPSB4	*Protobothrops mucrosquamatus* and *Thamnophis sirtalis*	SPIa/Ryanodine
SPSB3	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	
SPSB1	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	

1

VEGFA1	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus, Thamnophis sirtalis, Crotalus horridus* and *Protobothrops flavoviridis*	Vascular endothelial growth factor (VEGF)
VEGFA2		
VEGFA3		
VEGFB	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus, Thamnophis sirtalis, Crotalus horridus, Protobothrops flavoviridis* and *Sistrurus miliarius barbouri*	
VEGFC	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	
VEGFF	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus, Thamnophis sirtalis* and *Protobothrops flavoviridis*	
WAP	*Protobothrops mucrosquamatus, Python bivittatus, Thamnophis sirtalis* and *Ophiophagus hannah*	Whey acidic protein/secretory leukoproteinase inhibitor
WFIKKN1		
WFIKKN2		
CRISP	*Protobothrops flavoviridis, Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus, Thamnophis sirtalis, Crotalus horridus, Calloselasma rhodostoma, Sistrurus miliarius barbouri* and *Deinagkistrodon acutus*	CRISP

Figure 1(on next page)

Sequence identity (%) between the proteins from ten major venom families and their putative skin homologs in Russell's viper.

The homolog with the highest identity was considered where more than one homolog was present.

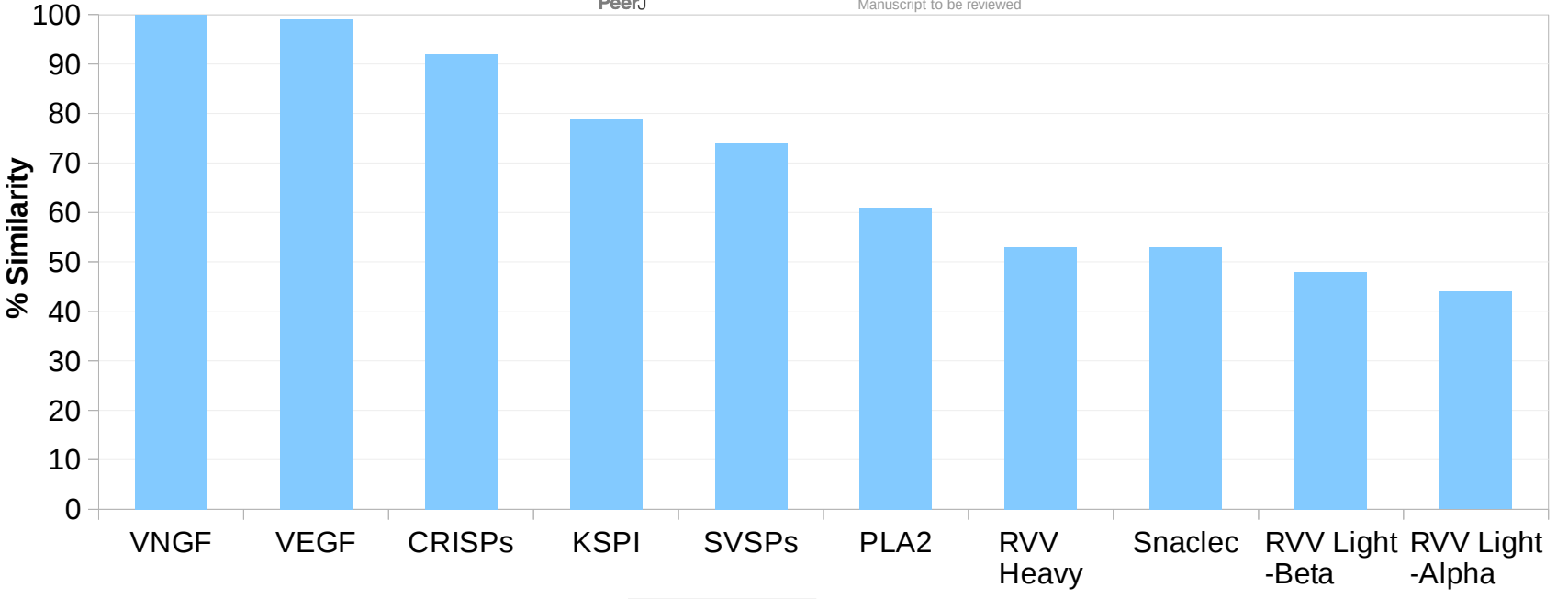

% Similarity
100
90
80
70
60
50
40
30
20
10
0
VNGF
VEGF
CRISPs
KSPI
SVSPs
PLA2
RVV Heavy
Snaclec
RVV Light -Beta
RVV Light -Alpha

Figure 2

Comparative analyses of nerve growth factor (NGF).

Putative NGF gene homolog, its mRNA, and protein domains in Russell's viper (A) and its comparison with the consensus NGF sequences from all five vertebrate phyla (fishes, amphibians, reptiles, birds and mammals) (B), with venomous (V) vertebrates from multiple phyla of vertebrates and invertebrates (C), and from various reptilian subgroups (D) are shown. The shades of brown and grey in B and C represent conservation to various degrees and variability, respectively. Grey in D represents conserved residues, red represents variable residues in the crotalines (CR), yellow and green represent conserved and variable residues in the viperines (VP), and elapids respectively.

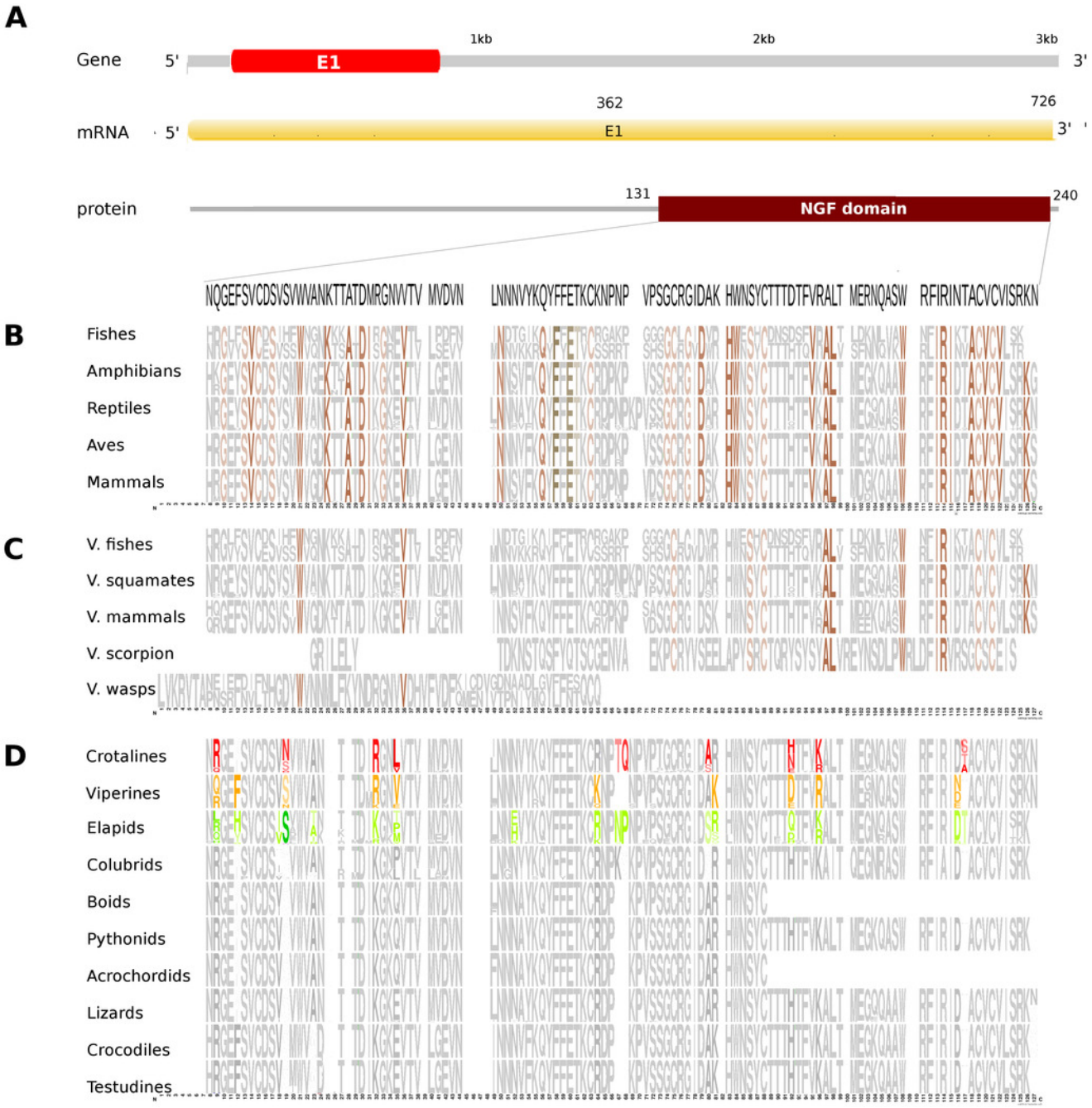

A
Gene
5'
E1
1kb
2kb
3kb
3'
mRNA
5'
362
726
E1
3'
protein
131
240
NGF domain
NQGEFSVCDSVSVWVANKTTATDMRGNVVTV MVDVN
LNNNVYKQYFFETKCKNPNP
VPSGCRGIDAK HWNSYCTTTDTFVRALT MERNQASW
RFIRINTACVCVISRKN
B
Fishes
Amphibians
Reptiles
Aves
Mammals
C
V. fishes
V. squamates
V. mammals
V. scorpion
V. wasps
D
Crotalines
Viperines
Elapids
Colubrids
Boids
Pythonids
Acrochordids
Lizards
Crocodiles
Testudines

Figure 3

Comparative analyses of vascular endothelial growth factor - A (VEGF-A).

Organization of the putative gene homolog, its mRNA, and protein domains of Russell's viper PDGF domain (A) and its comparison with the consensus sequences from all five vertebrate phyla (fishes, amphibians, reptiles, birds and mammals) (B), from the venomous (V) vertebrates and invertebrates (C), and from various reptilian subgroups (D) are shown. The shades of brown and grey in B and C represent conserved and varying residues, respectively. Grey in D represents conserved residues, red represents variable residues in the crotalines (CR), yellow and green represent conserved and variable residues in viperines (VP), and elapids respectively.

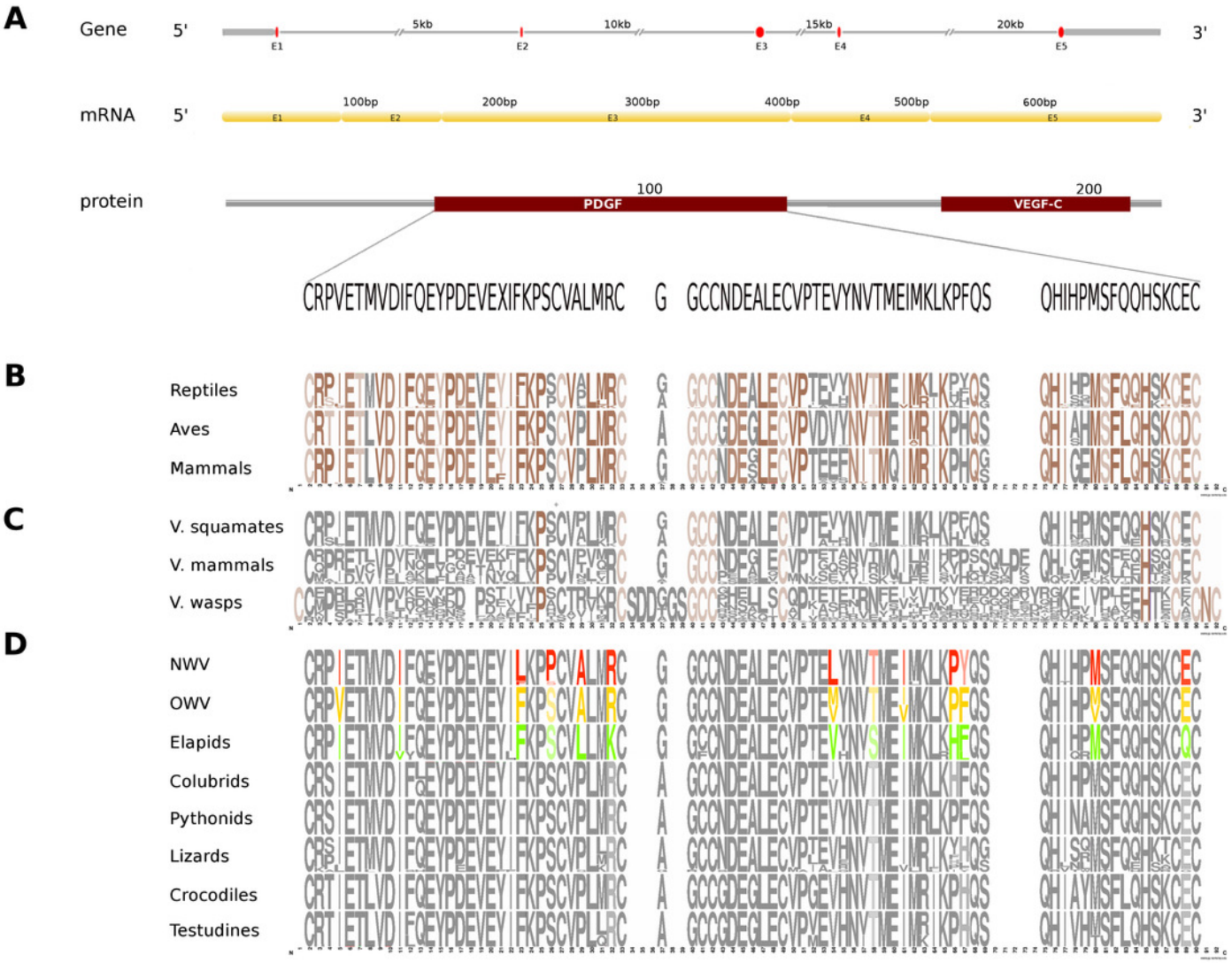

A
Gene
5'
5kb
10kb
15kb
20kb
3'
E1
E2
E3
E4
E5
mRNA
5'
100bp
200bp
300bp
400bp
500bp
600bp
3'
protein
100
200
PDGF
VEGF-C
CRPVETMVDIFQEYPDEVEXIFKPSCVALMRC G GCCNDEALECVPTEVYNVTMEIMKLKPFQS QHIHPMSFQQHSKCEC
B
Reptiles
Aves
Mammals
C
V. squamates
V. mammals
V. wasps
D
NWV
OWV
Elapids
Colubrids
Pythonids
Lizards
Crocodiles
Testudines

Figure 4

Comparative analyses of kunitoxin.

Organization of the putative gene homolog, its mRNA, and protein domains of Russell's viper (A) and its comparison with the consensus BPTI domain sequences from all five vertebrate phyla (fishes, amphibians, reptiles, birds and mammals) (B), from venomous (V) vertebrates and invertebrates (C), from various reptilian subgroups (D) are shown. The shades of brown and grey in B and C represent conserved and varying residues, respectively. Grey in D represents conserved residues, red represents variable residues in the crotalines (CR), yellow and green represent conserved and variable residues in viperines (VP), and elapids respectively.

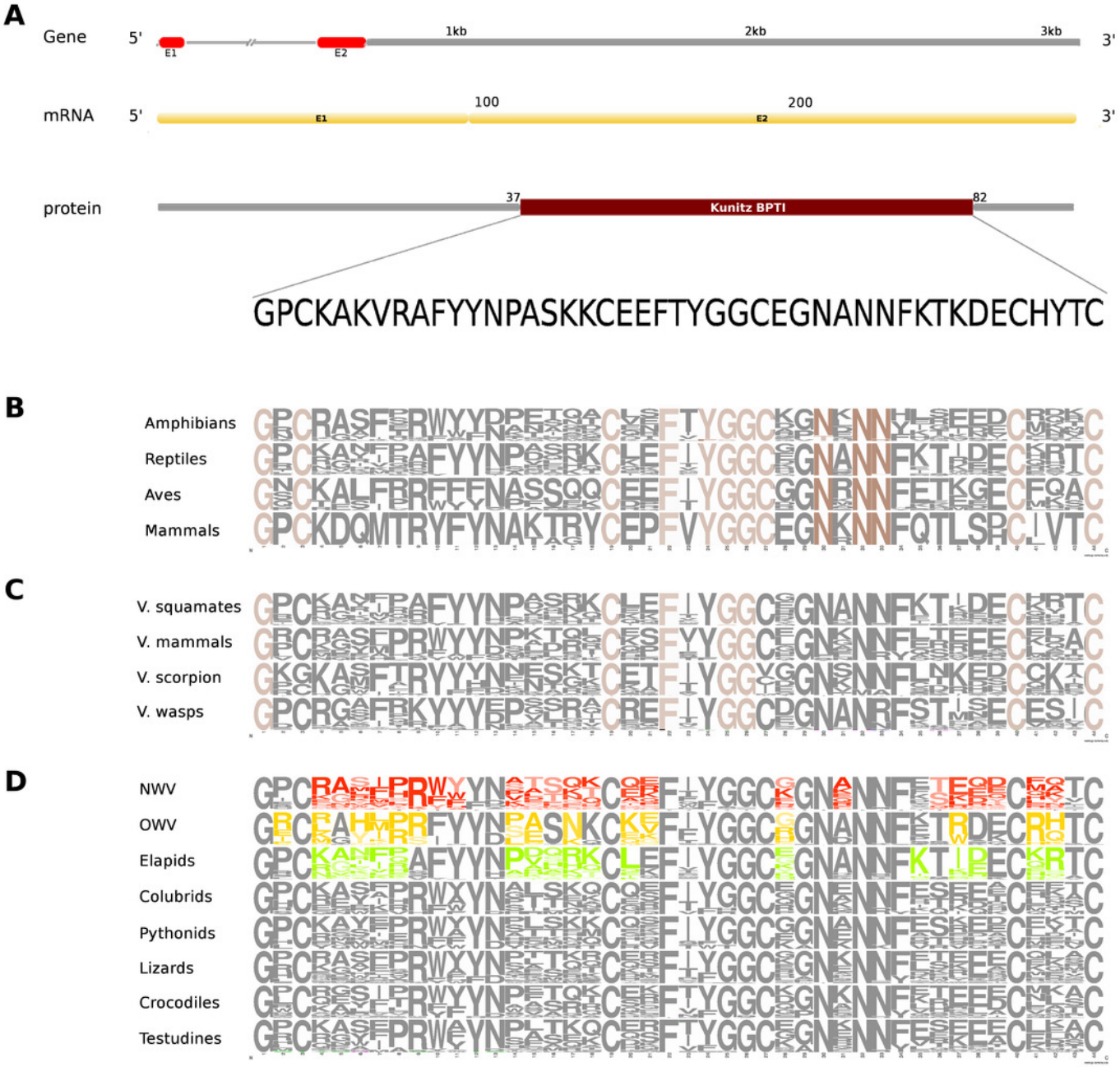

A
Gene
5'
E1
E2
1kb
2kb
3kb
3'
mRNA
5'
E1
E2
100
200
3'
protein
37
Kunitz BPTI
82
GPCKAKVRAFYYNPASKKCEEFTYGGCEGNANNFKTKDECHYTC
B
Amphibians
Reptiles
Aves
Mammals
C
V. squamates
V. mammals
V. scorpion
V. wasps
D
NWV
OWV
Elapids
Colubrids
Pythonids
Lizards
Crocodiles
Testudines

Figure 5

Comparative analyses of CRISP.

Organization of putative CRISP gene homolog, its mRNA, and protein domains of Russell's viper (A) and its comparison with the consensus CRISP sequences from all five vertebrate phyla (fishes, amphibians, reptiles, birds and mammals, B and E); from venomous animals (V) vertebrates (fishes, squamates and mammals) and invertebrates (scorpions and wasps, C and F); and from various reptilian subgroups (D and G) are shown. The shades of brown and grey in B, C, E and F represent conserved and varying residues, respectively. Grey in D and G represents conserved residues, red represents variable residues in the crotalines (CR), yellow and green represent conserved and variable residues in viperines (VP), and elapids respectively.

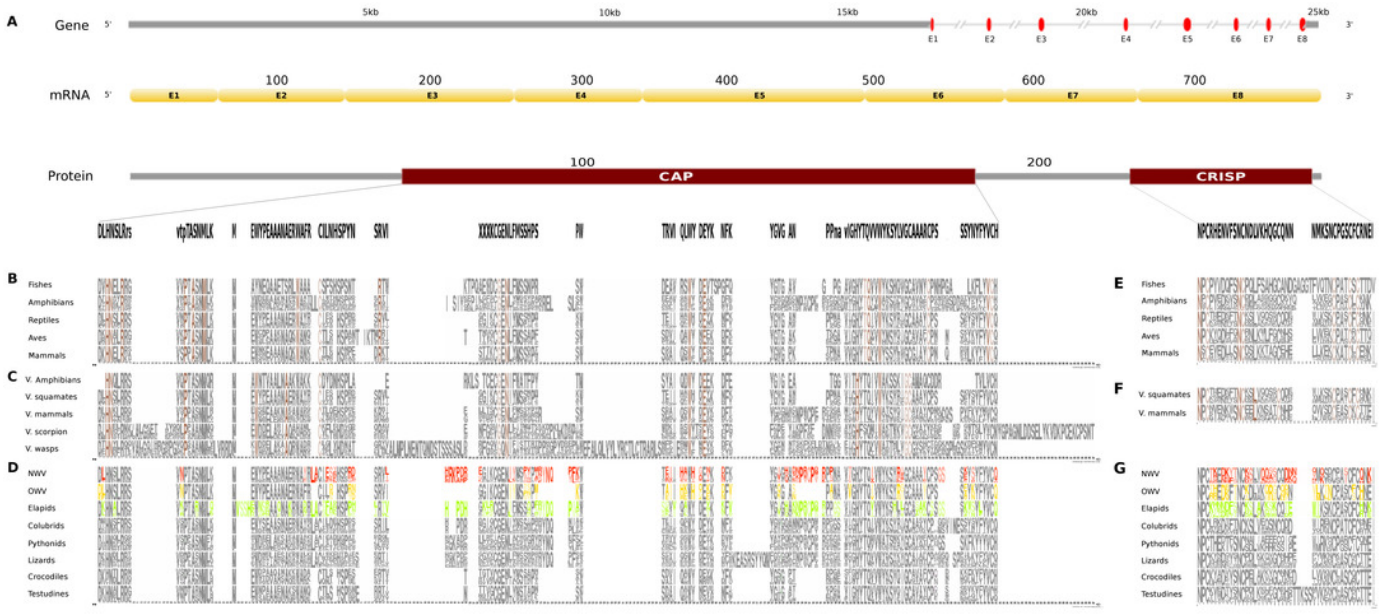
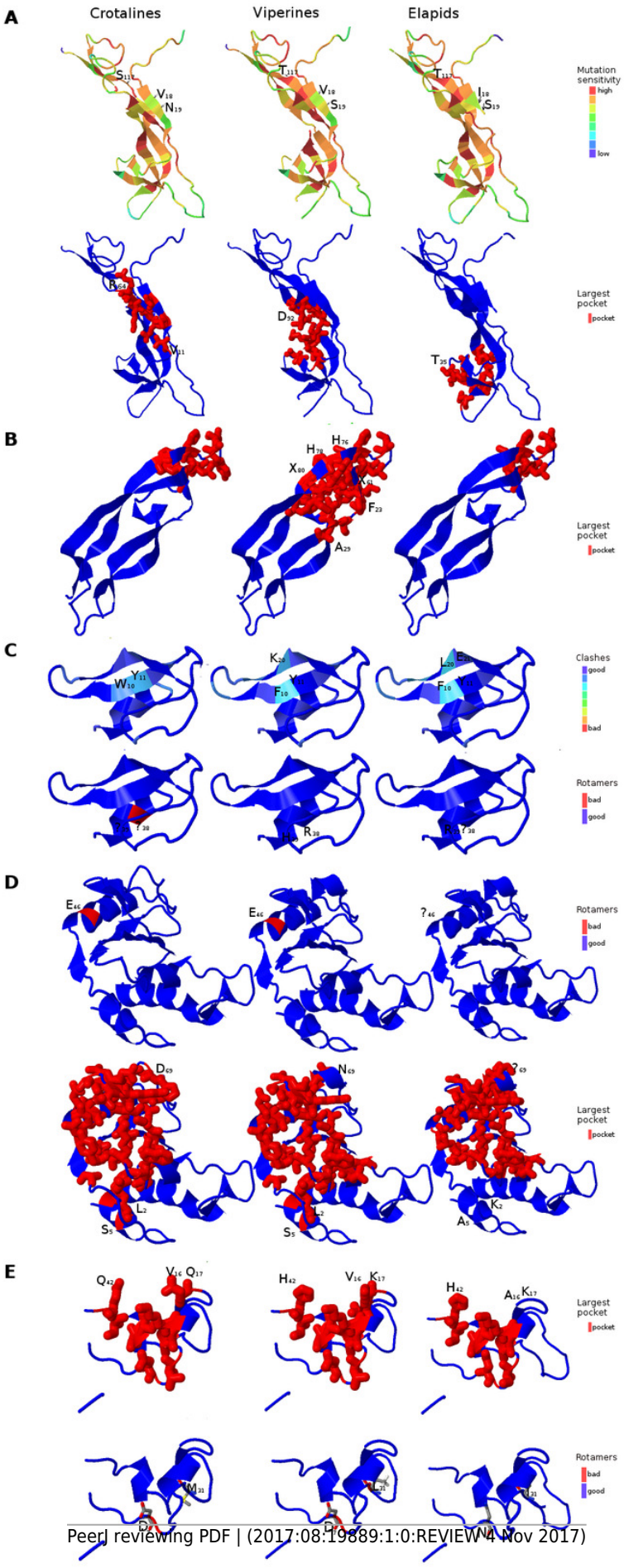


Figure 6

Three dimensional protein structural models in NGF (A); PDGF (B); Kunitz BPTI (C); CAP (D); and CRISP (E) across crotalines (CR), viperines (VP) and elapids.

The status of the parameters being investigated using Phyre2 are indicated in the color legends on the side.


A
Crotalines
Viperines
Elapids
Mutation sensitivity
high
low
Largest pocket
pocket
B
Largest pocket
pocket
C
Clashes
good
bad
Rotamers
bad
good
D
Rotamers
bad
good
Largest pocket
pocket
E
Largest pocket
pocket
Rotamers
bad
good