Accepted Manuscript
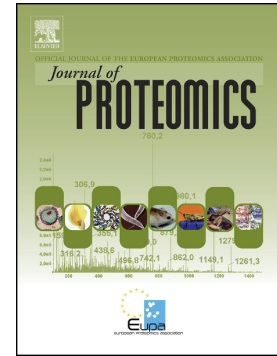
Intact protein mass spectrometry reveals intraspecies variations in venom composition of a local population of Vipera kaznakovi in Northeastern Turkey

Daniel Petras, Benjamin-Florian Hempel, Bayram Göçmen, Mert Karis, Gareth Whiteley, Simon C. Wagstaff, Paul Heiss, Nicholas R. Casewell, Ayse Nalbantsoy, Roderich D. Süssmuth

Intact protein mass spectrometry reveals intraspecies variations in venom composition of a local population of *Vipera kaznakovi* in Northeastern Turkey

Daniel Petras[a,b,*,^], Benjamin-Florian Hempel[a,*], Bayram Göçmen[c], Mert Karis[c,d], Gareth Whiteley[e], Simon C. Wagstaff[f], Paul Heiss[a], Nicholas R. Casewell[e], Ayse Nalbantsoy[g,^] and Roderich D. Süssmuth[a,^]

[a] Technische Universität Berlin, Institut für Chemie, Strasse des 17. Juni 124, 10623 Berlin, Germany.

[b] University of California San Diego, Collaborative Mass Spectrometry Innovation Center, 9500 Gilman Drive, La Jolla, California 92093, United States.

[c] Zoology Section, Department of Biology, Faculty of Science, Ege University, 35100 Bornova, Izmir, Turkey.

[d] Program of Laboratory Technology, Department of Chemistry and Chemical Process Technologies, Acıgöl Vocational School of Technical Sciences, Nevşehir Hacı Bektaş Veli University, Nevşehir, Turkey.

[e] Centre for Snakebite Research & Interventions, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, United Kingdom

[f] Bioinformatics Unit, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, United Kingdom

[g] Department of Bioengineering, Faculty of Engineering, Ege University, Bornova, 35100 Izmir, Turkey.

* These authors contributed equally to this study

^ Correspondence to dpetras@ucsd.edu, analbantsoy@gmail.com or roderich.suessmuth@tu-berlin.de

Abstract

We report on the variable venom composition of a population of the Caucasus viper (*Vipera kaznakovi*) in Northeastern Turkey. We applied a combination of venom gland transcriptomics, and de-complexing bottom-up and top-down venomics. In contrast to sole bottom-up venomics approaches and gel or chromatography based venom comparison, our combined approach enables a faster and more detailed comparison of venom proteomes from multiple individuals. In total, we identified peptides and proteins from 15 toxin families, including snake venom metalloproteinases (svMP; 37.8%), phospholipases A_2 (PLA$_2$; 19.0%), snake venom serine proteinases (svSP; 11.5%), C-type lectins (CTL; 6.9%) and cysteine-rich secretory proteins (CRISP; 5.0%), in addition to several low abundant toxin families. Furthermore, we identified intraspecies variations of the venom composition of *V. kaznakovi*, and find these were mainly driven by the age of the animals, with lower svSP abundance detected in juveniles. On the proteoform level, several small molecular weight toxins between 5 and 8 kDa in size, as well as PLA$_2$s, drove the differences observed between juvenile and adult individuals. This study provides novel insights into the venom variability of *V. kaznakovi* and highlights the utility of intact mass profiling for fast and detailed venom comparisons of snake.

Biological Significance

Population level and ontogenetic venom variation (e.g. diet, habitat, sex or age) can result in a loss of antivenom efficacy against snakebites from wide ranging snake populations. The current state of the art for the analysis of snake venoms are de-complexing bottom-up proteomics approaches. While useful, these have the significant drawback of being time-consuming and following costly protocols, and consequently are often applied to pooled venom samples. To overcome these shortcomings and to enable rapid and detailed profiling of large numbers of individual venom samples, we integrated an intact protein analysis workflow into a transcriptomics-guided bottom-up approach. The application

of this workflow to snake individuals of a local population of *V. kaznakovi* revealed intraspecies variations in venom composition, which are primarily explained by the age of the animals, and highlighted svSP abundance to be one of the molecular drivers for the compositional differences observed.

Highlights

- First venomic analysis of a local population of the Caucasus viper (*Vipera kaznakovi*).

- The venom gland transcriptome of *V. kaznakovi* identified 46 toxin genes relating to 15 venom toxin families.

- Bottom-up venomics revealed the identification of 25 proteins covering 7 toxin families mainly dominated by snake venom metalloproteinases (svMP).

- Population venomics by top-down mass profiling revealed ontogenetic shifts between juvenile and adult snakes.

Keywords: Viperidae, Snake, Middle East, Toxin, Caucasus viper, *Vipera kaznakovi*, Snake Venomics, Transcriptomics, Top-down Venomics, Population Venomics, Venom, Toxin

1. Introduction

Venomics is considered an integrative approach, that can combine proteomics, transcriptomics and/or genomics to study venoms [1]. Although the term was initially used to describe the mass spectrometry-based proteomic characterization of venoms [2,3], genomic [4,5] or more commonly venom gland transcriptomic sequencing [6–14] have also been used to characterize venom compositions. These molecular approaches provide an overview of venom composition by characterizing the nucleotide sequences of venom toxin-encoding genes (among others) and, in the case of transcriptomics, also provide an estimation of their relative expression in the venom gland. Furthermore, (translated) protein sequence databases are crucial for the robust annotation of tandem mass spectra from proteomic analyses in peptide/protein spectrum matching (P*r*SM). A bibliographic search to the keyword "Snake venomics" in PubMed identified 147 hits between 2004 and 2018, including a rapid expansion in the application of venomics approaches in more recent years.

Initial proteomic analyses of snake venoms included the combination of multidimensional separation techniques (chromatographic and gel electrophoresis), N-terminal Edman degradation, and *de novo* sequencing by tandem mass spectrometry of tryptic peptides generated by in-gel digestion of SDS-PAGE bands [2,15]. Since these initial studies, the proteomic characterization of snake venoms has

become more comprehensive due to technical advances in mass spectrometry and next generation nucleotide sequencing. Several complementary strategies were developed to unveil the venom proteomes of more than 100 snake species [16]. Most of these studies applied the so called 'bottom-up' proteomic approaches, whereby intact proteins are typically digested with trypsin before tandem mass spectrometry analysis. Many workflows perform venom decomplexation prior to digestion, either by liquid chromatography (LC) or gel electrophoresis, or a combination of both [17]. The direct, in-solution digestion, or so called 'shotgun proteomics', allows for a fast qualitative overview, but suffers from a less quantitative breakdown of snake venom composition [17,18]. For example, in shotgun experiments, the problem of protein inference often does not permit the differentiation of the numerous toxin isoforms present in venom [19]. Thus, chromatographic or electrophoretic separation of venom samples greatly aids in differentiating between toxin isoforms (encoded by paralogs). In addition, decomplexing prior to trypsin digestion often does not allow for the clear identification of differential post-translational modified variants, so-called proteoforms [20].

To circumvent these limitations, a logical solution would be the of the digestion step and the application of direct analysis of intact proteins by tandem mass spectrometry, so called top-down proteomics. Recently top-down protein analysis has been applied alone, or in combination with other venomics approaches, to study the venoms of the King cobra (*Ophiophagus hannah*) [21,22], the entire genus of mambas (*Dendroaspis* spp.) [23,24], the brown tree snake (*Boiga irregularis*) [6], the Okinawa habu pit viper (*Protobothrops flavoviridis*) [25], and several viper species from Turkey [26–28]. In the case of viperid species, top-down analysis typically only results in partial characterization of the venom, as a number of the main toxin components, such as high molecular weight snake venom metalloproteinases (svMPs) (>30 kDa), are challenging to efficiently ionize by denaturing electrospray ionization (ESI) and might only provide few observable fragments in tandem MS [29]. A possible way to overcome difficulties in terms of ionization of high molecular weight proteins is the application of native ESI, as described by Melani *et al.* [22]. However native top-down mass spectrometry typically requires a special type of mass spectrometer with extended mass range and more extensive sample preparation, which makes this type of analysis more technically challenging.

In the majority of the aforementioned studies, the top-down workflows were performed with a front-end LC-based sample decomplexation. This allows for the generation of MS1 mass profiles (XICs) of intact proteoforms. Typically, the MS1 information is accompanied by tandem MS (MS2) information acquired in data-dependent acquisition (DDA) mode. The MS2 fragment spectra are than matched to a translated transcriptome/genome database in order to identify the proteins. In the case that there are not enough MS2 fragment peaks of a particular proteoform, the intact molecular mass can still enable identification, especially if the intact mass can be associated to masses observed in complementary experiments, such as retention time, mass range of SDS-PAGE and/or bottom-up protein IDs of decomplexed bottom-up venomics [26]. The additional information gained through exact intact protein masses can be particularly informative to differentiate between isoforms or proteoforms. As several studies have shown correlations between different ecological, geographical, genetic and/or developmental factors and the venom proteome, e.g. different diets [30–33], regional separation of populations [10,34–36], sex [37–39] or age [40–43] it would be ideal to characterize venom composition from a representative cohort of individual animals. However, due to the time consuming nature and expensive costs associated with de-complexing bottom-up venom analysis, most venom compositions reported so far [16], were analyzed from a single pool of venom that was sourced from different numbers of individuals. In top-down approaches on the other hand, the simple sample preparation, high sensitivity and fast analysis time allows for a rapid and cheaper comparison of venom composition, and thus seems well suited for large scale quantitative comparisons [44–46] of snake venoms. In addition to providing a better biological understanding of the toxins present in venom [47], and the evolutionary processes underpinning population level venom variations [48], population wide venom analyses would provide important information to better understand regional and intraspecific variations in the venom composition of medically important snake species, which has considerable relevance for the development and clinical utility of antivenom treatment for snakebite [49,50].

In this study, we explored the utility of top-down intact mass profiling to identify intraspecific venom variation by applying it to a local population of the medical relevant Caucasus viper (*Vipera kaznakovi*). The Caucasus viper is a subtropical, medium-sized, viper species with a wide distribution

4

range from the Caucasus Black Sea coastal provinces of Artvin and Rize in northeastern Turkey (**Figure 1**), through Georgia to Russia. A distinctive characteristic of this species is the black coloration with elements of an orange to red zigzag-looking on the dorsal side of the body (**Figure 1**). This species feeds predominately on small vertebrates (mice, lizards etc.) and insects [51].

In a previous shotgun proteomics study of this species, Kovalchuk and coworkers described the venom of *V. kaznakovi* (Krasnodar Territory, near Adler, Russia) to be composed of phospholipase A_2 (PLA$_2$, 19.0%), snake venom metalloproteases (svMP, 16.2%), snake venom serine proteases (svSP, 10.8%), Cysteine-rich secretory proteins (CRISP, 9.7%), C-type lectins (CTL, 12.5%), L-amino acid oxidase (LAAO, 4.0%), vascular endothelial growth factor (VEGF, 4.0%), disintegrins (Dis, 0.5%), phospholipase B (PLB, 0.3%), nerve growth factors (NGF, 0.14%), as well as a number of other venom proteins of lower abundance [52]. Here, we used a combination of venom gland transcriptomics, decomplexing bottom-up proteomics and comparative top-down proteomics to provide a more detailed characterization of the venom composition of *V. kaznakovi*, and to gain first insights into intraspecies variation of its venom composition. Furthermore, our findings highlight the potential of intact protein mass profiling for future population level studies of viperid venoms.

2. Material & Methods

2.1. Sampling

Venom samples of *V. kaznakovi* were collected from 6 adult (2 female, 4 male) and 3 juvenile specimens (unknown sex). All specimens were captured in late June 2015 in their natural habitat and released back into their natural environment after venom extraction. The *V. kaznakovi* individuals were collected in Artvin province in Turkey near the Georgian border, with 6 individuals sampled from Hopa district, 2 individuals from Borçka district and 1 specimen in the Arhavi district. An additional female individual found in Borçka district was collected for venom gland dissection for transcriptomic analysis. Ethical permission (Ege University Animal Experiments Ethics Committee, 2013#049) and special permission (2015#124662) for the sampling of wild-caught *V. kaznakovi* were received from the Republic of Turkey, Ministry of Forestry and Water Affairs. Geographic coordinates for all *V. kaznakovi* individuals are provided in supplemental table 2.

5

2.2. Sample storage and preparation

Crude *V. kaznakovi* venom was extracted by using a parafilm-covered laboratory beaker without exerting pressure on the venom glands. Venom samples were centrifuged at 2000 × g for 10 min at 4 °C to remove cell debris. Supernatants were immediately frozen at −80 °C, lyophilized, and the samples stored at 4 °C until use.

2.3. Determination of lethal dose (LD$_{50}$)

The lethal potency (LD$_{50}$) of pooled *V. kaznakovi* venom to mice (mg/kg) was determined by an up- and down method as recommended by the Organization for Economic Cooperation and Development (OECD) guidelines (Test No. 425) [53,54]. Groups of five Swiss albino mice (n = 15; age, 8 to 10 weeks; female 8 and male 7 individuals) were used per venom dose. Various venom concentrations (5, 2 and 1 mg/kg, milligrams of protein per kg calculated from dry weight venom by Bradford assay) were diluted in ultrapure water to a final volume of 100 μL and injected by the intraperitoneal (IP) route. Control mice (n = 5; female 2 and male 3 individuals) received a single IP injection of sterile saline (0.9%, 100 μL). All assays and procedures involving animals strictly followed the ethical principles in animal research adopted by the Swiss Academy of Medical Sciences [55]. Additionally, they were approved by a local ethics committee (2013#049). Mortality was recorded 24 h after injection, and the median lethal dose was determined by a nonlinear regression fitting procedure in GraphPad Prism 5 (Version 5.01, Inc., San Diego, CA, USA).

2.4. RNA isolation and purification

Venom glands were dissected from a wild caught adult female specimen of *V. kaznakovi* in Kanlıdere, Hopa district (Artvin province) and processed as previously described [9,24]. Briefly, immediately following euthanasia, venom glands were dissected and were immediately flash frozen in liquid nitrogen and stored cryogenically prior to RNA extraction. Venom glands were next homogenized under liquid nitrogen and total RNA extracted using a TRIzol Plus RNA purification kit (Invitrogen), DNAse treated with the PureLink DNase set (Invitrogen) and poly(A) selected using the Dynabeads mRNA DIRECT purification kit (Life Technologies), as previously detailed [9,24].

6

2.5. RNA sequencing, assembly and annotation

RNA-Seq was performed as described in [9,24]. The RNA-Seq library was prepared from 50 ng of enriched RNA material using the ScriptSeq v2 RNA-Seq Library Preparation Kit (epicenter, Madison, WI, USA), following 12 cycles of amplification. The resulting sequencing library was purified using AMPure XP beads (Agencourt, Brea, CA, USA), quantified using the Qubit dsDNA HS Assay Kit (Life Technologies), before the size distribution was assessed using a Bioanalyser (Agilent). The library was then multiplexed and sequenced (alongside other sequencing libraries not reported in this study) on a single lane of an Illumina MiSeq (250 bp paired end sequencing), housed at the Centre for Genomic Research, Liverpool, UK. The *V. kaznakovi* library amounted to 1/6th of the total sequencing lane. The ensuing read data was quality processed by (i) removing the presence of any adapter sequences using Cutadapt (https://code.google.com/p/cutadapt/) and (ii) trimming low quality bases using Sickle (https://github.com/najoshi/sickle). Reads were trimmed if bases at the 3' end matched the adapter sequence for 3 bp or more, and further trimmed with a minimum window quality score of 20. After trimming, reads shorter than 10 bp were removed.

For sequence assembly we used VTBuilder, a *de novo* transcriptome assembly program previously designed for discriminating between multiple related toxin isoforms during the construction of snake venom gland transcriptomes [56], which has previously been utilized for integrating venom toxin gene data with venom proteomic data [9,24]. Paired-end read data was entered into VTBuilder and executed with the following parameters: min. input read length 150 bp; min. output transcript length 300 bp; min. isoform similarity 96%. Assembled contigs were annotated with BLAST2GO Pro v3 [57] using the blastx-fast algorithm with a significance threshold of $1e^{-5}$, to provide BLAST annotations (max 20 hits) against NCBI's non redundant (NR) protein database (41 volumes; Nov 2015) followed by mapping to gene ontology terms, and Interpro domain annotation using default parameters. Following generic annotation, venom toxins were initially identified based on their BLAST similarity to sequences previously identified in the literature or in molecular databases as snake venom toxins, and then manually curated for validation.

2.6. Venom proteomics (bottom-up)

The crude venom (1 mg) was dissolved to a final concentration of 10 mg/ml in aqueous 3% (v/v) acetonitrile (ACN) with 1% (v/v) formic acid (FA) and centrifuged at 16,000 *g* for 5 min to spin down insoluble content. The supernatant was loaded onto reversed-phase HPLC with a Supelco Discovery BIO wide Pore C18-3 column (4.6 x 150 mm, 3 μm particle size) using an Agilent 1260 Low Pressure Gradient System (Agilent, Waldbronn, Germany). The column was operated with a flow rate of 1 mL/min and run with ultrapure water (solution A) and ACN (solution B), both including 0.1% (v/v) FA. A standard separation gradient was used with solution A and solution B, starting isocratically (5% B) for 5 min, followed by linear gradients of 5-40% B for 95 min and 40-70% for 20 min, then 70% B for 10 min, and finally re-equilibration at 5% B for 10 min. Peak detection was performed at $\lambda = 214$ nm using a diode array detector (DAD). After the chromatographic separation of the crude venom, the collected and vacuum-dried peak fractions were submitted to SDS-PAGE gel electrophoresis (12% polyacrylamide). Subsequently, the coomassie-stained bands were excised, and submitted to in-gel trypsin digestion, reduced with fresh dithiothreitol (100 mM DTT in 100 mM ammonium hydrogencarbonate, pH 8.3, for 30 min at 56 °C) and alkylated with iodoacetamide (55 mM IAC in 100 mM ammonium hydrogencarbonate, pH 8.3, for 30 min at 25 °C in the dark). The resulting peptides were then extracted with 100 μL aqueous 30% (v/v) ACN with 5% (v/v) FA for 15 min at 37 °C. The supernatant was vacuum-dried (Thermo speedvac, Bremen, Germany), redissolved in 20 μL aqueous 3% (v/v) ACN with 1% (v/v) FA and submitted to LC-MS/MS analysis.

The bottom-up analyses were performed with an Orbitrap XL mass spectrometer (Thermo, Bremen, Germany) via an Agilent 1260 HPLC system (Agilent Technologies, Waldbronn, Germany) using a reversed-phase Grace Vydac 218MSC18 (2.1 x 150 mm, 5 μm particle size) column. The pre-chromatographic separation was performed with the following settings: after an isocratic equilibration (5% B) for 1 min, the peptides were eluted with a linear gradient of 5-40% B for 10 min, 40-99% B in 3 min, held at 99% B for 3 min and re-equilibrated in 5% B for 3 min.

2.7. Population level venom profiling (top-down)

The top-down MS analysis was performed by dissolving the crude venoms in ultrapure water containing formic acid (FA, 1%) to a final concentration of 10 mg/mL, and centrifuged at 20,000 x g for 5 min. Aliquots of 10 µL dissolved venom samples were submitted to reverse-phase (RP) HPLC-high-resolution (HR)-MS analyses. RP-HPLC-HR-MS experiments were performed on an Agilent 1260 HPLC system (Agilent, Waldbronn, Germany) coupled to an Orbitrap LTQ XL mass spectrometer (Thermo, Bremen, Germany). RP-HPLC separation was performed on a Supelco Discovery Biowide C18 column (300 Å pore size, 2×150 mm column size, 3 µm particle size). The flow rate was set to 0.3 mL/min and the column was eluted with a gradient of 0.1% FA in water (solution A) and 0.1% FA in ACN (solution B): 5% B for 5 min, followed by 5–40% B for 95 min, and 40–70% for 20 min. Finally, the gradient was held isocratic with 70% B for 10 min and re-equilibrated at 5% B for 10 min. ESI settings were: 11 L/min sheath gas; 35 L/min auxiliary gas; spray voltage, 4.8 kV; capillary voltage, 63 V; tube lens voltage, 135 V; and capillary temperature, 330 °C. MS/MS spectra were obtained in data-dependent acquisition (DDA) mode. FTMS measurements were performed with 1 µ scans and 1000 ms maximal fill time. AGC targets were set to 10^6 for full scans and to 3×10^5 for MS/MS scans, and the survey scan as well as both data dependent MS/MS scans were performed with a mass resolution (R) of 100,000 (at m/z 400). For MS/MS the two most abundant ions of the survey scan with known charge were selected. Normalized CID energy was set to 30% for the first, and 35% for the second, MS/MS event of each duty cycle. The default charge state was set to $z = 6$, and the activation time to 30 ms. Additional HCD experiments were performed with 35% normalized collision energy, 30 ms activation time and $z = 5$ default charge state. The mass window for precursor ion selection was set to 2 or 6 m/z. A window of 3 m/z was set for dynamic exclusion of up to 50 precursor ions with a repeat of 1 within 10 s for the next 20 s.

2.8. Bioinformatic analysis

The LC-MS/MS data files (.raw) obtained from the in-gel digestion were converted to mascot generic format (.mgf) files via MSConvert GUI of the ProteoWizard package (http://proteowizard.sourceforge.net; version 3.0.10328) and annotated by DeNovo GUI [58]

9

(version 1.14.5) with a mass accuracy of 10 ppm for precursor mass and 0.2 m/z for fragment peaks. A fixed modification carbamidomethyl cysteine (C +57.02 Da) was selected. Resulting sequence tags were examined manually and searched against the non-redundant Viperidae protein database (taxid: 8689) using the basic local alignment search tool (BLAST) [59].

For peptide spectrum matching, the SearchGUI software tool was used with XTandem! as the search engine [60]. The MS2 spectra were searched against the non-redundant Viperidae protein NCBI (taxid: 8689, 3rd Nov 2017, 1727 sequences), our in-house *Vipera kaznakovi* toxin sequence database (translated from our venom gland transcriptomic analyses; 46 toxin sequences) and a set of proteins found as common contaminants (CRAP, 116 sequences), containing in total 1,889 sequences. Mass accuracy was set to 10 ppm for the precursor mass and 0.2 m/z for the MS2 level. The alkylation of Cys was set as a fixed modification and acetylation of the N-terminus, of Lys, as well as oxidation of Met, were allowed as variable modifications. A false discovery rate (FDR) was estimated through a target-decoy approach and a cut-off of 1% was applied. All PSMs were validated manually and at least two PSMs were required for a protein ID to be considered.

For the top-down data analysis, the .raw data were converted to .mzXML files using MSconvert of the ProteoWizard package (http://proteowizard.sourceforge.net; version 3.0.10328), and multiple charged spectra were deconvoluted using the XTRACT algorithm of the Xcalibur Qual Browser version 2.2 (Thermo, Bremen, Germany). For isotopically unresolved spectra, charge distribution deconvolution was performed using the software tool magic transformer (MagTran).

2.9. Multivariable statistics

Principal component analysis (PCoA), using the relative percentages of the major toxin families as well as different proteoforms as variables, was applied to investigate determinants of compositional variation among venoms. PCoA was performed in R (R Foundation for Statistical Computing, 2016) with the extension Graphic Package rgl, available from https://www.R-Project.org.

2.10. Data sharing

Mass spectrometry proteomics data (.mgf, .raw and results files and search database) have been deposited to ProteomeXchange [61] with the ID PXD010857 via the MassIVE partner repository under project name "Venom proteomics of *Vipera kaznakovi*" and massive ID MSV000082845. Raw sequencing reads and the assembled contigs generated for the venom gland transcriptome (.fastq and .fasta, respectively) have been deposited in the NCBI sequence read archive (SRA) under accession SRR8198764 and linked to the BioProject identifier PRJNA505487.

3. Results and Discussion

3.1. Field work and venom toxicity

During our fieldwork in June 2015 we collected nine *V. kaznakovi* individuals (6 adults and 3 juveniles) in their natural habitat, and extracted their venom before releasing them back into their natural environment. The different *V. kaznakovi* individuals were found in the Hopa (6 spec.), Borçka (2 spec.) and Arhavi (1 spec.) districts of Artvin province (**Figure 1**). The LD_{50} mean values of venom pooled from all collected *V. kaznakovi* individuals was assessed by the intraperitoneal (IP) route using a random sample survey of five Swiss albino mice for three venom dose (5, 2 and 1 mg/kg), which is summarized in supplemental table 1. The LD_{50} value obtained for the pooled *V. kaznakovi* venom was calculated as ~2.6 mg/kg (2.1-3.4 mg/kg) and can be categorized to have slightly weaker toxicity in this model, when compared to other related viper species (0.9-1.99 mg/kg) [62–65].

3.2. Venom gland transcriptomics

The *V. kaznakovi* venom gland transcriptome resulted in 1,742 assembled contigs, of which 46 exhibited gene annotations relating to 15 venom toxin families previously described in the literature (**Figure 2**). The majority of these contigs (33) encode genes expressing toxin isoforms relating to four multi-locus gene families, namely the svMPs, CTLs, svSPs and PLA$_2$s (**Figure 2**). Moreover, these four toxin families also exhibited the highest expression levels of the toxin families identified; in combination accounting for >78% of all toxin expressions (**Figure 2**). These findings are consistent with many prior studies of viperid venom gland transcriptomes [10,12,49,66,67].

The svMPs were the most abundantly expressed of the toxin families detected, accounting for 33.4% of the total toxin expression, and were encoded by 17 contigs (**Figure 2**). However, these contig numbers are likely to be an overestimation of the total number of expressed svMP genes found in the *V. kaznakovi* venom gland, as six of these contigs were incomplete and non-overlapping in terms of their nucleotide sequence, and therefore likely reflect a degree of low transcriptome coverage and/or under-assembly. Of those contigs that we were able to identify to svMP class level (e.g. P-I, P-II or P-III [68,69]), ten exhibited structural domains unique to P-III svMPs, one to P-II svMPs and one to a short coding disintegrin. The svMP contig that exhibited the highest expression level encoded for the sole P-II svMP (5.1% of all venom toxins), whereas the short coding disintegrin, which exhibited 98% identity to the platelet aggregation inhibitor lebein-1-alpha from *Macrovipera lebetina* [70], was more moderately expressed (2.1%). Interestingly, we found no evidence for the representation of the P-I class of svMPs in the *V. kaznakovi* venom gland transcriptome.

The CTLs were the next most abundant toxin family, with six contigs representing 27.5% of all toxin gene expression (**Figure 2**). One of these CTLs, which exhibits closest similarity to snaclec-7 from *Vipera ammodtyes* venom (GenBank: APB93444.1), was by far the most abundantly expressed toxin identified in the venom gland transcriptome (15.4% of all toxins) (**Figure 2**). We identified lower expression levels for the multi-locus svSP and PLA$_2$ toxin families, which accounted for 9.2% and 8.1% of the total toxins, expressed in the venom gland transcriptome respectively, and were encoded by seven and three contigs (**Figure 2**). Of the remaining toxin families identified, only two exhibited expression levels >3% of the total toxin expression; CRISPs were encoded by two contigs amounting to 5.4% of total toxin expression, and LAAO by a single contig representing 4.2% (**Figure 2**). The remaining nine, lowly expressed, toxin families identified in the venom gland transcriptome are displayed in Figure 2, and combined amounted to 12.1% of total toxin expression.

3.3 Venom proteomics of pooled venom

To broadly characterize the venom composition of *V. kaznakovi*, we performed bottom-up analysis of pooled venom by reversed phase-HPLC separation (**Figure 3A**) and direct online intact mass analysis by ESI-HR-MS (Figure **3B**). The prominent bands of the subsequent separation by SDS-PAGE

(**Figure 3C**) were excised followed by trypsin in-gel digestion and LC-MS/MS analysis. During the first analysis we did not have a species-specific transcriptome database available, hence the spectra were analyzed by *de novo* sequencing. The resulting sequence tags were searched against the NCBI non-redundant viperid protein database using BLAST [59]. The 57 sequence tags resulted in the identification of 25 proteins covering seven toxin families (**Table 1**), namely svMP, PLA$_2$, svSP, CTL, CRISP, VEGF and LAAO.

De novo sequencing of MS/MS spectra of native small peptides (peaks 1-9) resulted in four additional sequence tags and the identification of a svMP inhibitor (svMP-i) and two bradykinin potentiating peptides (BPP). When we obtained the assembled transcriptome data, we re-analyzed the MS/MS data from the tryptic peptides by peptide spectrum matching (PSM) using the translated protein sequences of the transcriptome as well as the NCBI Viperidae protein database. PSM resulted in 114 peptide matches in total, which doubled the number of annotated spectra in comparison to the *de novo* annotation. The analysis revealed the same seven major toxin families as identified by the tryptic *de novo* tags, but showed 29 identified proteins (compared to 25 by the prior approach) and thus a modest improvement. Not surprisingly, most of the peptide matches were from the transcriptome derived sequences, with only six protein IDs sourced from other viperid sequences in the NCBI database. Relative quantification through integration of the UV-HPLC peaks and densitometric analysis of the SDS-PAGE gels revealed that the most abundant toxin families were svMP (37.7%), followed by PLA$_2$ (19.0%), svSP (9.6%), LAAO (7.1%), CTL (6.9%), CRISP (5.0%), and VEGF (0.3%). In the small molecular mass range (< 2kDa), SVMP-i represented 12.6%, BPP 2.0%, and unknown peptides 4.0% of the overall venom composition (**Figure 3D**).

When comparing the abundance of venom toxins (**Figure 3D**) with transcriptomic predictions of expression levels (**Figure 2A**), we observed an overall positive correlation, but also noted some major differences, particularly relating to the CTLs: transcriptomic expression levels showed CTLs to be the second most abundant toxin family (27.5% of all toxins) while proteomic analysis showed a much lower abundance (6.9%). Interestingly, some of the molecular masses observed for CTLs (~20 kDa) during SDS-PAGE did not correspond to the expected molecular mass derived from the transcriptome sequences. As reported in other studies, we assume that some of the observed CTLs are hetero-dimers

13

[71]. SvMPs showed highly consistent profiles, as both the most abundantly expressed (33.4%) and translated (32.7%) toxin family. Similarly, the svSPs (9.2%) and CRISPs (5.4%) exhibited transcription levels highly comparable to their relative protein abundance in venom (9.6% and 5.02%). A lower transcription level was shown for PLA_2 (8.1%) in contrast to the two times higher protein level (19.0%). As anticipated, with the exception of VEGF (2.0% T; 0.4% P) and svMP-i (1.7%; 12.6%) as part of the peptidic content, other lowly expressed 'toxin' families could not be assigned on the proteomic level.

The observed discrepancies in proteomic abundance and transcriptomic expression (e.g. CTLs and PLA_2s) could be influenced by many factors, e.g. post-genomic factors acting on toxin genes [49], such as the regulation of expression patterns by MicroRNAs (miRNA) [7,72], degradation processes [73], systematic or stochastic variations [74] or technical limitations in our experimental approaches, including the lower sensitivity of the proteomics workflow. Perhaps most importantly, we compared the toxin transcription level of a single individual (adult female) to a pooled venom protein sample (n=9), and thus, while it is possible that these differences are predominately due to the above mentioned regulatory processes, it seems likely that intra-specific venom variations may also influence our findings. Due to sampling/ethical restrictions relating to the sacrifice of multiple individuals, we were unable to sequence venom gland transcriptomes of additional specimens of *V. kaznakovi* to investigate this further.

The previous proteomic characterization of *V. kaznakovi* venom by Kovalchuk and coworkers was performed by in-solution trypsin proteolysis followed by nanoLC-MS/MS [52]. The PSM against a full NCBI Serpentes database identified 116 proteins from 14 typical snake venom protein families. The semi-quantitative venom composition showed PLA_2 (41.0%) as the most abundant component, followed by svMPs (16.2%), CTL (12.5%), svSP (10.8%), CRISP (9.7%), LAAO (4.0%), VEGF (4.0%) and other lowly abundant proteins (< 1%) [52]. The main difference between these findings and those described herein, are the considerably higher levels of PLA_2 and the lower abundance of svMPs (~ 4 fold difference for both protein families). The reasons for the additional detection of lowly abundant proteins could be of technical nature, as the nanoLC-MS/MS and mass spectrometer used in the previous study is typically more sensitive than the LC-MS/MS setup we applied.

Differences in protein abundance could also be the result of the different quantification methods applied (UV abundance vs. summed peptide abundance [52]), but the observed variations could also be biological in nature, i.e. the result of intra-specific venom variation, as the animals were collected in distinct geographic regions (Krasnodar Territory, Russia [51], with a distance of ~ 400 km to our collection site in Turkey). However, as in most other venom proteomics studies, Kovalchuk *et al.* determined toxin composition using a pooled venom sample (15 individuals [52]), which has the potential to offset variation among individuals. Therefore, in order to robustly assess the extent of intra-specific (e.g. population level) venom variation in *V. kaznakovi*, analysis of a representative group of individuals is necessary.

3.4 Population level venom profiling

It is understandable that many venom proteomic studies were undertaken using pooled samples, due to the associated costs and analysis time of decomplexing bottom-up venomics. For example, herein we fractionated pooled venom from *V. kaznakovi* into 25 fractions and further separated the protein containing fractions (MW >5kDa) by SDS-PAGE. This multidimensional separation resulted in 25 digested peptide samples which were analyzed by LC-MS/MS, requiring ~ 10 h MS run time (25 min/sample), and an estimated ~ US$1250 costs (US$50/sample). Multiplying this effort and cost by numerous venom samples from individuals would of course make such a study comparatively expensive. Hence, many previous studies investigating venom variability within a species have used pooled venom for in-depth proteomic analysis, and then illuminated individual variability by the comparison of HPLC chromatograms and/or SDS-PAGE images [50,75–77]. However, such an approach allows at best a comparison at the protein family level (if protein families are clearly separated by HPLC or SDS-PAGE). As an alternative, top-down or shotgun proteomics would facilitate differential comparisons on the protein, or potentially proteoform, level, by performing a single LC-MS/MS run per individual. Taking into account the longer LC-MS run-times (120 min/run) for shotgun or top-down analysis and costs than can be estimated to be ~ US$200 per venom sample – the analysis would be significantly cheaper and faster. However, shotgun approaches are likely to suffer from the aforementioned issues with protein inference, while top-down approaches have the

15

drawback of not resolving high molecular mass proteins. This is particularly the case if the identification and comparison of proteins are based on Protein Spectrum Matching (PrSM), as high molecular weight toxins may not result in isotope resolved peaks and sufficient precursor signal, and thus are unlikely to provide sufficient fragment ions. However, a comparison by MS1 mass profiling only [78] would eliminate the problem of insufficient MS/MS fragments and isotope resolution, as spectra can be easily deconvoluted based on their charge state distribution. Such an approach could be particularly interesting for laboratories that are equipped with low resolution mass spectrometers.

In order to explore the potential of venom comparison by top-down mass profiling, we analyzed the venoms of nine *V. kaznakovi* individuals by LC-MS using the same chromatographic method as for our initial HPLC separation of our decomplexing bottom-up venom analysis. Chromatographic peak extraction of all individuals resulted in 119 consensus extracted ion chromatograms (XIC) or so-called ion features. The alignment of XICs by retention time and mass enabled the comparison of samples between individuals, and a comparison with the mass profile of the pooled venom sample for a protein level annotation. An overview of all resulting features, including annotations, is shown in supplemental table 1. The binary distribution of ion features showed that individual venoms contained between 62 and 107 features, with a slightly higher average feature number in juveniles vs. adults. Comparing the total ion currents (TIC) of the LC-MS runs, the individual with the lowest feature number also had the lowest overall signal. Hence it is likely that the lower number of features in this individual was due to lower overall signal intensity and therefore might not be biologically representative. For further statistical evaluation we normalized feature abundance to TIC. Matching the features to the pooled bottom-up venomics results yielded an annotation rate of between 83.4% and 93.5% of the features (based on XIC peak area). As anticipated, the annotation rate is slightly lower than the relative annotation of the pooled sample (96.0%; based on the UV_{214} peak area).

A comparison of the resulting protein family venom compositions is shown in figure 4 and supplemental table 2. The highest variance was observed for svSP, CTL and LAAO toxin families (**Figure 5A**). Taking the age of the individuals into account, the abundance of svSPs was generally higher in the adult individuals than in the juveniles (average of 21.7% vs. 5.5 %), but no significant difference between male and female individuals, or between different geographic regions was

observed. The svSPs play a significant role in mammalian envenomation by affecting the hemostatic system through perturbing blood coagulation, typically via the inducement of fibrinogenolytic effects [79,80]. Taking this into account, a possible explanation could be that the lower svSP concentration observed in juveniles could be the result of differences in diet, as young animals typically prey on insects, before switching to feed upon small mammals and lizards as they become adults [81–83]. Despite observed variations in abundance, no significant differences were found between the individual groups for the CTL and LAAO toxin families (**Figure 5A**). However, there was evidence that svSP concentration is correlated to levels of LAAO, as the three individuals with the lowest svSP abundance showed the highest content of LAAO (**Figure 5A**). Whether this is a true biological effect or perhaps is the result of differences in ion suppression of the co-eluting compounds requires further investigation with large sample sizes. While we also observed variation between the PLA$_2$ levels identified in the different venoms, which ranged from 6.5-25.1%, in all cases these abundances remained considerably lower than those previously reported by Kovalchuk *et al.* (41%) [52].

In order to investigate the intra-species differences by multivariate statistics we performed a principal coordinate analysis (PCoA) using the Bray-Curtis dissimilarity metric. The PCoA plots of protein-level and proteoform-level data are shown in figure 5. Clustering between adult and juvenile individuals in protein-family level PCoA space was not found to be significant (**Figure 5B**, ADONIS: r^2=0.32, p=0.27). In addition, and as anticipated from the univariate statistics, no separation based on gender or geographical region could be observed. Since a possible explanation for not resolving phenotype differences could be the reduction of variables through the binning of proteoforms, we next used proteoform abundance as the input matrix for PCoA. The outcomes of this second analysis revealed a weak grouping between both juvenile and adults, as well as between male and female snakes (**Figure 5C**, ADONIS: r^2=0.33, p=0.07). To investigating the toxin variants underpinning these separations, we used univariate comparisons of the two groups and plotted the fold change of toxin abundance (log2) vs. the statistical significance (-log10 p-value, t-test) shown in supplemental figure 2. Aside from the aforementioned differences in svSP toxins, the most significant (p-value < 0.05, log2 fold change >2 or <-2) differences observed between juvenile and adult individuals was the higher abundance of small proteins (with masses 7707.26 Da, 5565.02 Da and 5693.10 Da) in the

17

juvenile group, all of which were unidentified in our proteomic analyses. In addition, we observed several smaller peptides (with masses 589.27 Da, 1244.56 Da, and 575.26 Da), as well as a putative PLA_2 (13667.91 Da), that were more abundant in venom of the juveniles. Contrastingly, a putative PLA_2 (13683.86 Da) was of lower abundance in the juvenile group. While we observed fewer significant changes between the venom toxins of the male and female individuals, the observed masses of the differential features indicated that some of those differential toxins belong to different protein families than those involved in differentiating between juvenile and adult snakes. Two toxins with the masses 22829.66 Da and 24641.23 Da were of higher abundant in male individuals, and could be putatively annotated as hetero-dimeric CTLs, while a third toxin (13549.87 Da, putative PLA_2) was also of higher abundance in the males.

4. Concluding remarks

Herein we described comprehensive analyses of the venom composition of *Vipera kaznakovi* by a combination of venom gland transcriptomics and decomplexing bottom-up and top-down venom proteomics. Our findings revealed the presence of 15 toxin families, of which the most abundant toxins were svMPs (37.7%), followed by PLA_2s (19.0%), svSPs (9.5%), CTLs (6.9%) and CRISP (5.0%). Intact mass profiling enabled the rapid comparison of venom sourced from multiple individuals. This population venomics approach enabled higher sensitivity of direct intact protein analysis by LC-MS, in comparison to decomplexing bottom-up venomics, and thus enabled us to work with multiple venom samples of low quantity (< 0.5 mg venom). This approach also permitted us to capture the snakes, perform venom extractions, and then immediately release the animals back into the field. Our findings revealed intraspecies venom variation in *Vipera kaznakovi*, including both ontogenetic differences between juvenile and adult snakes, and to a lesser extent, sexual differences between adult males and females. The highest significant difference in venom proteome composition was observed between the adult and juvenile group, with svSP toxins found to exhibit the greatest variance. However, individuals in all groups showed a generally high relative variance of CTL and LAAO concentrations. svMPs on the other hand seemed to be constantly the most abundant venom

18

component in all *V. kaznakovi* individuals analyzed in our study. However, as the statistical power with a relatively small subject size (n=9) is limited, we acknowledge that extending this study to a larger sample cohort, ideally covering all geographical regions (from Northeastern Turkey to Georgia and Russia) of the *V. kaznakovi* distribution zone, would be revealing. In conclusion, the top-down venom profiling approach that we applied herein appears to be well suited for extensive venom analysis at the population level, and will hopefully enable venom researchers to expand their experimental toolbox towards robust comparisons of intraspecies venom variation, and move beyond the characterization of pooled venoms.

Author contributions

D.P., A.N. and R.D.S. planned the study. D.P., A.N., B.G., M.K. and P.H. collected the animals and prepared venom and venom gland tissue samples. A.N. performed the determination of acute lethal dose. D.P., P.H. and B.-F.H. performed the toxin separation and acquired the mass spectrometry data. G.W., S.C.W. and N.R.C. constructed the transcriptome. D.P., B.-F.H. and N.R.C. performed the data analysis. A.N., N.R.C. and R.D.S. acquired funding and provided materials and instruments for the study. D.P., B.-F.H. and R.D.S. wrote the manuscript. All authors read, discussed and approved the manuscript.

Acknowledgements

References

[1] Calvete JJ, Lomonte B. A bright future for integrative venomics. Toxicon 2015;107(Pt B):159–62.

[2] Juárez P, Sanz L, Calvete JJ. Snake venomics: Characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, N-terminal sequencing, and tandem mass spectrometry analysis. Proteomics 2004;4(2):327–38.

[3] Calvete JJ. Snake venomics - from low-resolution toxin-pattern recognition to toxin-resolved venom proteomes with absolute quantification. Expert Rev Proteomics 2018;15(7):555–68.

[4] Vonk FJ, Casewell NR, Henkel CV, Heimberg AM, Jansen HJ, McCleary RJR et al. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. Proc Natl Acad Sci U S A 2013;110(51):20651–6.

[5] Aird SD, Arora J, Barua A, Qiu L, Terada K, Mikheyev AS. Population Genomic Analysis of a Pitviper Reveals Microevolutionary Forces Underlying Venom Chemistry. Genome Biol Evol 2017;9(10):2640–9.

[6] Pla D, Petras D, Saviola AJ, Modahl CM, Sanz L, Pérez A et al. Transcriptomics-guided bottom-up and top-down venomics of neonate and adult specimens of the arboreal rear-fanged Brown Treesnake, *Boiga irregularis*, from Guam. J Proteomics 2018;174:71–84.

[7] Durban J, Sanz L, Trevisan-Silva D, Neri-Castro E, Alagón A, Calvete JJ. Integrated Venomics and Venom Gland Transcriptome Analysis of Juvenile and Adult Mexican Rattlesnakes *Crotalus simus*, *C. tzabcan*, and *C. culminatus* Revealed miRNA-modulated Ontogenetic Shifts. J Proteome Res 2017;16(9):3370–90.

[8] Aird SD, da Silva NJ, Qiu L, Villar-Briones A, Saddi VA, Pires de Campos Telles M et al. Coralsnake Venomics: Analyses of Venom Gland Transcriptomes and Proteomes of Six Brazilian Taxa. Toxins (Basel) 2017;9(6).

[9] Pla D, Sanz L, Whiteley G, Wagstaff SC, Harrison RA, Casewell NR et al. What killed Karl Patterson Schmidt? Combined venom gland transcriptomic, venomic and antivenomic analysis of the South African green tree snake (the boomslang), *Dispholidus typus*. Biochim Biophys Acta 2017;1861(4):814–23.

[10] Gonçalves-Machado L, Pla D, Sanz L, Jorge RJB, Leitão-De-Araújo M, Alves MLM et al. Combined venomics, venom gland transcriptomics, bioactivities, and antivenomics of two *Bothrops jararaca* populations from geographic isolated regions within the Brazilian Atlantic rainforest. J Proteomics 2016;135:73–89.

[11] Fry BG, Scheib H, L M Junqueira Azevedo I de, Silva DA, Casewell NR. Novel transcripts in the maxillary venom glands of advanced snakes. Toxicon 2012;59(7-8):696–708.

[12] Casewell NR, Harrison RA, Wüster W, Wagstaff SC. Comparative venom gland transcriptome surveys of the saw-scaled vipers (*Viperidae Echis*) reveal substantial intra-family gene diversity and novel venom transcripts. BMC Genomics 2009;10:564.

[13] Rokyta DR, Wray KP, Margres MJ. The genesis of an exceptionally lethal venom in the timber rattlesnake (*Crotalus horridus*) revealed through comparative venom-gland transcriptomics. BMC Genomics 2013;14:394.

[14] Tan CH, Tan KY, Fung SY, Tan NH. Venom-gland transcriptome and venom proteome of the Malaysian king cobra (*Ophiophagus hannah*). BMC Genomics 2015;16:687.

[15] Nawarak J, Sinchaikul S, Wu C-Y, Liau M-Y, Phutrakul S, Chen S-T. Proteomics of snake venoms from Elapidae and Viperidae families by multidimensional chromatographic methods. Electrophoresis 2003;24(16):2838–54.

[16] Tasoulis T, Isbister GK. A Review and Database of Snake Venom Proteomes. Toxins (Basel) 2017;9(9).

[17] Lomonte B, Calvete JJ. Strategies in 'snake venomics' aiming at an integrative view of compositional, functional, and immunological characteristics of venoms. J Venom Anim Toxins Incl Trop Dis 2017;23:26.

[18] Melani RD, Goto-Silva L, Nogueira FCS, Junqueira M, Domont GB. Shotgun Approaches for Venom Analysis. In: Gopalakrishnakone P, Calvete JJ, editors. Venom Genomics and Proteomics: Venom Genomics and Proteomics. Dordrecht: Springer Netherlands; 2014, p. 1–12.

[19] Nesvizhskii AI, Aebersold R. Interpretation of shotgun proteomic data: The protein inference problem. Mol Cell Proteomics 2005;4(10):1419–40.

[20] Smith LM, Kelleher NL. Proteoform: A single term describing protein complexity. Nat Methods 2013;10(3):186–7.

[21] Petras D, Heiss P, Süssmuth RD, Calvete JJ. Venom Proteomics of Indonesian King Cobra, *Ophiophagus hannah*: Integrating Top-Down and Bottom-Up Approaches. J Proteome Res 2015;14(6):2539–56.

[22] Melani RD, Skinner OS, Fornelli L, Domont GB, Compton PD, Kelleher NL. Mapping Proteoforms and Protein Complexes From King Cobra Venom Using Both Denaturing and Native Top-down Proteomics. Mol Cell Proteomics 2016;15(7):2423–34.

[23] Petras D, Heiss P, Harrison RA, Süssmuth RD, Calvete JJ. Top-down venomics of the East African green mamba, *Dendroaspis angusticeps*, and the black mamba, *Dendroaspis polylepis*, highlight the complexity of their toxin arsenals. J Proteomics 2016;146:148–64.

[24] Ainsworth S, Petras D, Engmark M, Süssmuth RD, Whiteley G, Albulescu L-O et al. The medical threat of mamba envenoming in sub-Saharan Africa revealed by genus-wide analysis of venom composition, toxicity and antivenomics profiling of available antivenoms. J Proteomics 2018;172:173–89.

[25] Damm M, Hempel B-F, Nalbantsoy A, Süssmuth RD. Comprehensive Snake Venomics of the Okinawa Habu Pit Viper, *Protobothrops flavoviridis*, by Complementary Mass Spectrometry-Guided Approaches. Molecules 2018;23(8).

[26] Göçmen B, Heiss P, Petras D, Nalbantsoy A, Süssmuth RD. Mass spectrometry guided venom profiling and bioactivity screening of the Anatolian Meadow Viper, *Vipera anatolica*. Toxicon 2015;107(Pt B):163–74.

[27] Hempel B-F, Damm M, Göçmen B, Karis M, Oguz MA, Nalbantsoy A et al. Comparative Venomics of the *Vipera ammodytes transcaucasiana* and *Vipera ammodytes montandoni* from Turkey Provides Insights into Kinship. Toxins (Basel) 2018;10(1).

[28] Nalbantsoy A, Hempel B-F, Petras D, Heiss P, Göçmen B, Iğci N et al. Combined venom profiling and cytotoxicity screening of the Radde's mountain viper (*Montivipera raddei*) and Mount Bulgar Viper (*Montivipera bulgardaghica*) with potent cytotoxicity against human A549 lung carcinoma cells. Toxicon 2017;135:71–83.

[29] Ghezellou P, Garikapati V, Kazemi SM, Strupat K, Ghassempour A, Spengler B. A perspective view of top-down proteomics in snake venom research. Rapid Commun Mass Spectrom 2018.

[30] Daltry JC, Wüster W, Thorpe RS. Diet and snake venom evolution. Nature 1996;379(6565):537–40.

[31] Barlow A, Pook CE, Harrison RA, Wüster W. Coevolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution. Proc Biol Sci 2009;276(1666):2443–9.

[32] Gibbs HL, Mackessy SP. Functional basis of a molecular adaptation: Prey-specific toxic effects of venom from *Sistrurus* rattlesnakes. Toxicon 2009;53(6):672–9.

[33] Gibbs HL, Sanz L, Sovic MG, Calvete JJ. Phylogeny-based comparative analysis of venom proteome variation in a clade of rattlesnakes (*Sistrurus sp.*). PLoS ONE 2013;8(6):e67220.

[34] Alape-Girón A, Sanz L, Escolano J, Flores-Díaz M, Madrigal M, Sasa M et al. Snake venomics of the lancehead pitviper *Bothrops asper*: Geographic, individual, and ontogenetic variations. J Proteome Res 2008;7(8):3556–71.

[35] Huang H-W, Liu B-S, Chien K-Y, Chiang L-C, Huang S-Y, Sung W-C et al. Cobra venom proteome and glycome determined from individual snakes of *Naja atra* reveal medically important dynamic range and systematic geographic variation. J Proteomics 2015;128:92–104.

[36] Jorge RJB, Monteiro HSA, Gonçalves-Machado L, Guarnieri MC, Ximenes RM, Borges-Nojosa DM et al. Venomics and antivenomics of *Bothrops erythromelas* from five geographic populations within the Caatinga ecoregion of northeastern Brazil. J Proteomics 2015;114:93–114.

[37] Pimenta DC, Prezoto BC, Konno K, Melo RL, Furtado MF, Camargo ACM et al. Mass spectrometric analysis of the individual variability of *Bothrops jararaca* venom peptide fraction.

Evidence for sex-based variation among the bradykinin-potentiating peptides. Rapid Commun Mass Spectrom 2007;21(6):1034–42.

[38] Menezes MC, Furtado MF, Travaglia-Cardoso SR, Camargo ACM, Serrano SMT. Sex-based individual variation of snake venom proteome among eighteen *Bothrops jararaca* siblings. Toxicon 2006;47(3):304–12.

[39] Amorim FG, Costa TR, Baiwir D, Pauw E de, Quinton L, Sampaio SV. Proteopeptidomic, Functional and Immunoreactivity Characterization of *Bothrops moojeni* Snake Venom: Influence of Snake Gender on Venom Composition. Toxins (Basel) 2018;10(5).

[40] Mackessy SP. Venom Ontogeny in the Pacific Rattlesnakes *Crotalus viridis helleri* and *C. v. oreganus*. Copeia 1988;1988(1):92.

[41] Gutiérrez JM, Avila C, Camacho Z, Lomonte B. Ontogenetic changes in the venom of the snake *Lachesis muta stenophrys* (bushmaster) from Costa Rica. Toxicon 1990;28(4):419–26.

[42] Guércio RAP, Shevchenko A, Shevchenko A, López-Lozano JL, Paba J, Sousa MV et al. Ontogenetic variations in the venom proteome of the Amazonian snake *Bothrops atrox*. Proteome Sci 2006;4:11.

[43] Mackessy SP, Sixberry NM, Heyborne WH, Fritts T. Venom of the Brown Treesnake, *Boiga irregularis*: Ontogenetic shifts and taxa-specific toxicity. Toxicon 2006;47(5):537–48.

[44] Wu S, Brown JN, Tolić N, Meng D, Liu X, Zhang H et al. Quantitative analysis of human salivary gland-derived intact proteome using top-down mass spectrometry. Proteomics 2014;14(10):1211–22.

[45] Moehring F, Waas M, Keppel TR, Rathore D, Cowie AM, Stucky CL et al. Quantitative Top-Down Mass Spectrometry Identifies Proteoforms Differentially Released during Mechanical Stimulation of Mouse Skin. J Proteome Res 2018;17(8):2635–48.

[46] Ntai I, Toby TK, LeDuc RD, Kelleher NL. A Method for Label-Free, Differential Top-Down Proteomics. Methods Mol Biol 2016;1410:121–33.

[47] Calvete JJ, Casewell NR, Hernández-Guzmán U, Quesada-Bernat S, Sanz L, Rokyta DR et al. Venom Complexity in a Pitviper Produced by Facultative Parthenogenesis. Sci Rep 2018;8(1):11539.

[48] Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG. Complex cocktails: The evolutionary novelty of venoms. Trends Ecol Evol (Amst) 2013;28(4):219–29.

[49] Casewell NR, Wagstaff SC, Wüster W, Cook DAN, Bolton FMS, King SI et al. Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. Proc Natl Acad Sci U S A 2014;111(25):9205–10.

[50] Massey DJ, Calvete JJ, Sánchez EE, Sanz L, Richards K, Curtis R et al. Venom variability and envenoming severity outcomes of the *Crotalus scutulatus scutulatus* (Mojave rattlesnake) from Southern Arizona. J Proteomics 2012;75(9):2576–87.

[51] Starkov VG, Osipov AV, Utkin YN. Toxicity of venoms from vipers of *Pelias* group to crickets *Gryllus assimilis* and its relation to snake entomophagy. Toxicon 2007;49(7):995–1001.

[52] Kovalchuk SI, Ziganshin RH, Starkov VG, Tsetlin VI, Utkin YN. Quantitative Proteomic Analysis of Venoms from Russian Vipers of *Pelias* Group: Phospholipases A_2 are the Main Venom Components. Toxins (Basel) 2016;8(4):105.

[53] Bruce RD. An up-and-down procedure for acute toxicity testing. Fundam Appl Toxicol 1985;5(1):151–7.

[54] Cates CC, McCabe JG, Lawson GW, Couto MA. Core body temperature as adjunct to endpoint determination in murine median lethal dose testing of rattlesnake venom. Comp Med 2014;64(6):440–7.

[55] Ethical principles and guidelines for experiments on animals. Swiss Academy of Medical Sciences. Swiss Academy of Sciences. Experientia 1996;52(1):1–3.

[56] Archer J, Whiteley G, Casewell NR, Harrison RA, Wagstaff SC. VTBuilder: A tool for the assembly of multi isoform transcriptomes. BMC Bioinformatics 2014;15:389.

[57] Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 2008;36(10):3420–35.

[58] Muth T, Weilnböck L, Rapp E, Huber CG, Martens L, Vaudel M et al. DeNovoGUI: An open source graphical user interface for de novo sequencing of tandem mass spectra. J Proteome Res 2014;13(2):1143–6.

[59] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215(3):403–10.

[60] Barsnes H, Vaudel M. SearchGUI: A Highly Adaptable Common Interface for Proteomics Search and de Novo Engines. J Proteome Res 2018;17(7):2552–5.

[61] Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat Biotechnol 2014;32(3):223–6.

[62] Sket D, Gubensek F, Adamic S, Lebez D. Action of a partially purified basic protein fraction from *Vipera ammodytes* venom. Toxicon 1973;11(1):47–53.

[63] Oukkache N, El Jaoudi R, Ghalim N, Chgoury F, Bouhaouala B, Mdaghri NE et al. Evaluation of the lethal potency of scorpion and snake venoms and comparison between intraperitoneal and intravenous injection routes. Toxins (Basel) 2014;6(6):1873–81.

[64] Calderón L, Lomonte B, Gutiérrez JM, Tarkowski A, Hanson LA. Biological and biochemical activities of *Vipera berus* (European viper) venom. Toxicon 1993;31(6):743–53.

[65] Haro L de, Robbe-Vincent A, Saliou B, Valli M, Bon C, Choumet V. Unusual neurotoxic envenomations by *Vipera aspis aspis* snakes in France. Hum Exp Toxicol 2002;21(3):137–45.

[66] Rokyta DR, Lemmon AR, Margres MJ, Aronow K. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genomics 2012;13:312.

[67] Junqueira-de-Azevedo ILM, Bastos CMV, Ho PL, Luna MS, Yamanouye N, Casewell NR. Venom-related transcripts from *Bothrops jararaca* tissues provide novel molecular insights into the production and evolution of snake venom. Mol Biol Evol 2015;32(3):754–66.

[68] Fox JW, Serrano SMT. Insights into and speculations about snake venom metalloproteinase (SVMP) synthesis, folding and disulfide bond formation and their contribution to venom complexity. FEBS J 2008;275(12):3016–30.

[69] Casewell NR, Wagstaff SC, Harrison RA, Renjifo C, Wüster W. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. Mol Biol Evol 2011;28(9):2637–49.

[70] Gasmi A, Srairi N, Guermazi S, Dekhil H, Dkhil H, Karoui H et al. Amino acid structure and characterization of a heterodimeric disintegrin from *Vipera lebetina* venom. Biochim Biophys Acta 2001;1547(1):51–6.

[71] Navdaev A, Clemetson JM, Polgar J, Kehrel BE, Glauner M, Magnenat E et al. Aggretin, a heterodimeric C-type lectin from *Calloselasma rhodostoma* (Malayan pit viper), stimulates platelets by binding to α2β1 integrin and glycoprotein Ib, activating Syk and phospholipase Cγ 2, but does not involve the glycoprotein VI/Fc receptor γ chain collagen receptor. J Biol Chem 2001;276(24):20882–9.

[72] Durban J, Pérez A, Sanz L, Gómez A, Bonilla F, Rodríguez S et al. Integrated "omics" profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, *Crotalus simus simus*. BMC Genomics 2013;14:234.

[73] Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 2012;13(4):227–32.

[74] Ruggles KV, Krug K, Wang X, Clauser KR, Wang J, Payne SH et al. Methods, Tools and Current Perspectives in Proteogenomics. Mol Cell Proteomics 2017;16(6):959–81.

[75] Calvete JJ, Sanz L, Pérez A, Borges A, Vargas AM, Lomonte B et al. Snake population venomics and antivenomics of *Bothrops atrox*: Paedomorphism along its transamazonian dispersal and implications of geographic venom variability on snakebite management. J Proteomics 2011;74(4):510–27.

[76] Galizio NdC, Serino-Silva C, Stuginski DR, Abreu PAE, Sant'Anna SS, Grego KF et al. Compositional and functional investigation of individual and pooled venoms from long-term captive and recently wild-caught *Bothrops jararaca* snakes. J Proteomics 2018;186:56–70.

[77] Margres MJ, Patton A, Wray KP, Hassinger ATB, Ward MJ, Lemmon EM et al. Tipping the scales: The migration-selection balance leans toward selection in snake venoms. Mol Biol Evol 2018.

[78] Fry BG, Wickramaratna JC, Hodgson WC, Alewood PF, Kini RM, Ho H et al. Electrospray liquid chromatography/mass spectrometry fingerprinting of Acanthophis (death adder) venoms: Taxonomic and toxinological implications. Rapid Commun Mass Spectrom 2002;16(6):600–8.

[79] Serrano SMT. The long road of research on snake venom serine proteinases. Toxicon 2013;62:19–26.

[80] Slagboom J, Kool J, Harrison RA, Casewell NR. Haemotoxic snake venoms: Their functional activity, impact on snakebite victims and pharmaceutical promise. Br J Haematol 2017;177(6):947–59.

[81] Dias GS, Kitano ES, Pagotto AH, Sant'anna SS, Rocha MMT, Zelanis A et al. Individual variability in the venom proteome of juvenile *Bothrops jararaca* specimens. J Proteome Res 2013;12(10):4585–98.

[82] Gibbs HL, Sanz L, Chiucchi JE, Farrell TM, Calvete JJ. Proteomic analysis of ontogenetic and diet-related changes in venom composition of juvenile and adult Dusky Pigmy rattlesnakes (*Sistrurus miliarius barbouri*). J Proteomics 2011;74(10):2169–79.

[83] Zelanis A, Tashima AK, Rocha MMT, Furtado MF, Camargo ACM, Ho PL et al. Analysis of the ontogenetic variation in the venom proteome/peptidome of *Bothrops jararaca* reveals different strategies to deal with prey. J Proteome Res 2010;9(5):2278–91.

[84] Geniez P. Snakes of Europe, North Africa & the Middle East: A photographic guide. Princeton: Princeton University Press; 2018.

Table 1. Venom Protein Identifications from *Vipera kaznakovi*. The table shows all protein identification of HPLC fractions (Fig. 3) by LC-MS and LC-MS/MS analysis from pooled venom. Peak numbering corresponds to the UV and MS chromatograms. Sequence tags were obtained by analysis of tryptic peptides by MS/MS de novo sequencing and/or peptide spectrum matching. Molecular weights of intact proteins were determined by SDS-PAGE and intact mass profiling (LC-MS).

Peak	RT	SDS-PAGE Band	Area% SDS-PAGE	Area (214 nm)	Mass [Da] (ESI-MS)	Mass [kDa] (SDS-PAGE)	Sequence [PSM]	Prot-ID	Sequence (de novo)	Identity (blast)	e-value	Accession Number	Protein Family
1	20.1			78423									
					443.2				pEQW	tripeptide SVMPi	-	-	svMP-i
2	27.4			3701									
					432.2					unknown	-	-	Peptide
3	28.2			1928									
					3390.6					unknown			Peptide
4	30.3			11784									
					3942.8				-	unknown	-	-	-
					860.3				EPGEEDW	Bradykinin-potentiating peptide	-	BAN04688.1	BPP
					822.4				pEKWPGPK	Bradykinin-potentiating peptide	-	BAN04688.1	BPP
5	32.6			1075									
					3118.5				-	unknown	-	-	Peptide
					680.3				-	unknown	-	-	Peptide
6	35.7			3117									
					3665.7				-	unknown	-	-	Peptide
7	36.8			2039									
					13848.7				-	unknown	-	-	unknown
8	41.2			18410									
					1101.6				K/QPGPVSV	unkown	-	-	Peptide
9	49.7			2706									
					7228.2				-	unknown	-	-	unkown
					6680.9				-	unknown	-	-	unkown
10	50.4			6454									
		10A	35%			14	PFXEVYQR		PFXEVYQR	Nerve groth factor	2.2E-02	P83942.1	VEGF
							HTVDXQXM*R	T0203_R_0.0314_L_1049	HTVDXQXMR	Nerve groth factor	5.1E-02		
							ETXVPXXQEYPDEXSDXFRPSCVAVXR		-	-	-	-	

Peak	RT	SDS-PAGE Band	Area% SDS-PAGE	Area (214 nm)	Mass [Da] (ESI-MS)	Mass [kDa] (SDS-PAGE)	Sequence [PSM]	Prot-ID	Sequence (de novo)	Identity (blast)	e-value	Accession Number	Protein Family
		10B	65%			15	AAAXCFGENVNTYDKK	F8QN51.1	AAAXCAFGENVNTYDKK	acidic phospholipase A2	3.0E-08	F8QN51.1	PLA$_2$
							pCCFVHDCCYGR		NXFQFGK	acidic phospholipase A2	1.1E+00		
							M*DTYSYSFXNGDXVCGDDDPCXR	T1290_R_0.0575_L_419	MFCAGYXEGGK	cationic trypsin-3-like	4.0E-05	XP_015670852.1	
							SAYGCYCGWGGQGRPQDPTDR		-	-	-	-	
11, 12	61.2 61.6	11,12	100%	456									
				62431		15	AAAXCFGENVN*TYDKK	F8QN51.1	AAAXCAFGENVNTYDKK	acidic phospholipase A2	3.0E-08	F8QN51.1	PLA$_2$
					13557.7; 13540.8; 13523.7		M*DTYSYSFXN*GDXVCDGDDDPCXR		-	-	-	-	
							SAYGCYCGWGGQGRPQDPTDR	T1290_R_0.0575_L_419	-	-	-	-	
							SAXXSYSAYGCYCGWGGQGRPQDPTDR		-	-	-	-	
13	64.4			1798									unknown
					13541.8								
14	69.7			962									unknown
					24671.3								
15	70.5			45974									
		15A	71%		24655.5	25	Q*GCNNNYXK		QGCNNNYXK	cyteine-rich venom protein	3.0E-03	B7FDI1.1	CRISP
							KPEXQN*EXXDXHNSXRR		KPEXQNEXXDXHNSXR	cyteine-rich venom protein	3.0E-05	XP_015678374.1	
							NVDFDSESPR		WTAXXHEWHGEEK	cyteine-rich venom protein	4.0E-07	B7FDI1.1	
							M*EWYPEAAANAER	B7FDI1.1	SVDFDSESPR	cyteine-rich venom protein	2.0E-05	P86537.1	
							SVNPTASNM*XK		-	-	-	-	
							VDFDSESPR		-	-	-	-	
							DFVYGQGASPANAVVGHYTQXVWYK		-	-	-	-	
		15B	29%		13691.8	13	HXSQFGDMXNK		HXSQFGDMXNK	ammodytin I1(A) variant	6.0E-04	CAE47141.1	PLA$_2$
							pCCFVHDCCYGR	Q910A1	-	-	-	-	
							VAAXCFGENM*NTYDQKK		VAAXCAFGENMNTYDQK	ammodytin I1(A) variant	3.0E-09	CAE47176.1	
16, 17	73.3 74.4			56120									
		17A	27%		51761	50	FXTNFKPDCTXXRPSR	T0053_R_0.0734_L_1810	VPXVGVEFWXNR	snake venom metalloproteinase III	5.0E-04	ADW54336.1	svMP
							SECDXPEYCTGK		XVXVVDHSMVEK	snake venom metalloproteinase	9.0E-05	ADI47673.1	
							XGQDXYYCR		-	-	-	-	
							KEN*DVPXPCAPEDVK		-	-	-	-	

Peak	RT	SDS-PAGE Band	Area% SDS-PAGE	Area (214 nm)	Mass [Da] (ESI-MS)	Mass [kDa] (SDS-PAGE)	Sequence [PSM]	Prot-ID	Sequence (de novo)	Identity (blast)	e-value	Accession Number	Protein Family
		16, 17B	73%		13675.9	14	HXSQFGDMXNK		HXSQFGDMXNK	ammodytin I1(A) variant	6.0E-04	CAE47141.1	PLA$_2$
							pCCFVHDCCYGR	Q910A1	-	-	-	-	
							VAAXCFGEN*M*NTYDQKK		VAAXCAFGENMNTYDQK	ammodytin I1(A) variant	3.0E-09	CAE47176.1	
							YMLYSIFDCK		-	-	-	-	
18	80.6			90949									
		18A	18%		N.D.	65	XVXVVDHSM*VTK		XVXVVDHSMVTK	snake venom metalloproteinase group III	1.0E-04	CAJ01689.1	svMP
							YN*SDXTVXR		YNSDXTVXR	snake venom metalloproteinase	2.9E-01	ADI47687.1	
							VPXVGVEXWDHR		VPXVGVEXWDHR	snake venom metalloproteinase	6.5E-02	ADI47590.1	
							pQXVATSEQQR		-	-	-	-	
							VNXXNEM*YXPXNXR	T0033_R_0.0599_L_2024	-	-	-	-	
							KRHDNAQXXTTXDFDGSVXGK		-	-	-	-	
							HSVAXVEDYSPXDR		-	-	-	-	
							FXTNDKPDCTXXRPSR		-	-	-	-	
							KGESYFYCR		KGESYFYCR	snake venom metalloproteinase	9.0E-05	ADI47619.1	
							KENDVPXPCAPEDXK		-	-	-	-	
		18B	8%		51600	50	FXTNFKPDCTXXRPSR		VPXVGVEFWXNR	snake venom metalloproteinase group III	5.0E-04	ADW54336.1	svMP
							SECDXPEYCTGK	T0053_R_0.0734_L_1810	XVXVVDHSMVEK	snake venom metalloproteinase	9.0E-05	ADI47673.1	
							XGQDXYYCR		-	-	-	-	
							KEN*DVPXPCAPEDVK		-	-	-	-	
		18C	53%		30133	35	VXGGDECNXNEHPFXVAXHTAR	T1355_R_0.005_L_400	VXGGDECANXNEHPFXAFVTSDR	snake venom serine proteinase nikobin	2.0E-12	E5AJX2.1	svSP
							FYCAGTLXNQEWVXTAAR		XMGWGTXSSTK	snake venom serine proteinase	3.0E-05	ART88740.1	
							VVCAGXWQGGK		VVCAGXWQGGK	snake venom serine proteinase nikobin	1.0E-04	E5AJX2.1	
							C*AGTXXNQEWVXTAAHCNGK	E5AJX2	XMGWGTXTTTK	snake venom serine proteinase	6.0E-04	ADE45141.1	
							XXPDVPHCANXEXXK		-	-	-	-	
							VHPEXPAK		-	-	-	-	
		18D	20%		17249	14	KTWEDAEKFCTEQAR		WTEDAENFCQK	C-type lectin snaclec-1	1.0E-04	AMB36338.1	CTL
							SPEEVDFM*XK	T0841_R_0.0782_L_536	SPEEVDFMXK	C-type lectin-like protein 2B	1.1E-02	AJO70722.1	
							ADXVWXGXR		HXATXEWXGK	C-type lectin snaclec A16	1.8E-01	B4XSZ1.1	

Peak	RT	SDS-PAGE Band	Area% SDS-PAGE	Area (214 nm)	Mass [Da] (ESI-MS)	Mass [kDa] (SDS-PAGE)	Sequence [PSM]	Prot-ID	Sequence (de novo)	Identity (blast)	e-value	Accession Number	Protein Family
19, 20	85.3 86.6			32261									
		19/20	100%		49299	50	XGQDXYYCR		XGQDXYYCR	zinc metalloproteinase disintegrin-like	1.4E-01	Q0NZX8.1	svMP
							HDNAQXXTAXDFDGPTXGXAHMSSMCQSK		XVXVVDHSMVEK	snake venom metalloproteinase	9.0E-05	ADI47673.1	
							SVAFVEDYSPXDHMVASTMAHEMGHNXGMR		VPXVGVEFWXNR	snake venom metalloproteinase group III	5.0E-04	ADW54336.1	
							FXTNFKPDCTXXRPSR	T0053_R_0.0734_L_1810	pEXVATSEQQSYYDRFR	snake venom metalloproteinase	2.0E-10	ADI47633.1	
							SWVQCESGECCEQCR		-	-	-	-	
							SECDXPEYCTGK		-	-	-	-	
							KEN*DVPXPCAPEDVK		-	-	-	-	
							XYCEXVPN		-	-	-	-	
21	91.3			52573									
		21	100%		57374	60	VGEVNKDPGXXK		VGEVNKDPGXXK	L-amino acid oxidase	7.0E-04	BAN82140.1	LAAO
							VTVXEASER		VTVXEASER	L-amino acid oxidase	5.0E-02	Q6WP39.1	
							NVEEGWYANXGPM*RXPEK	T0018_R_0.0657_L_2450	-	-	-	-	
							HXVVVGAGM*SGXSAAYVXAGAGHK		-	-	-	-	
							SAGQXYEESXKK		-	-	-	-	
							TFCYPSMXQK		-	-	-	-	
					N.D.	50	YNSDXTVXR		YNSDXTVXR	snake venom metalloproteinase	2.9E-01	ADI47687.1	svMP
							XVXVVDHSMVTK		-	-	-	-	
							VPXVGVEXWDHR	T0033_R_0.0599_L_2024	-	-	-	-	
							HSVAXVEDYSPXDR		-	-	-	-	
							KENDVPXPCAPEDXK		-	-	-	-	
22	95.5			15611									
		22A	58%		N.D.	65	XVXVVDHSM*FTK		XVXVVDHSMFTK	zinc metalloproteinase disintegrin-like	7.0E-04	Q9IAX6.1	svMP
							XYEM*VNTXNVVFR		XYEMVNTXNVVFR	snake venom metalloproteinase group III	1.0E-06	AMB36352.1	
							VAXVYXEM*WTNR		KXVYXEMWTNR	snake venom metalloproteinase group III	4.0E-04	AMB36352.1	
							XHSWVECESGECCDQCR	T0039_R_0.0184_L_1944	-	-	-	-	
							AXFGANAAVGQDACFDWNKK		-	-	-	-	
							GTDDFYCR		-	-	-	-	
					8415.9		XFCEIVPNTCK		-	-	-	-	
							XHSWVECESGECCEQCR	T0033_R_0.0599_L_2024	-	-	-	-	

Peak	RT	SDS-PAGE Band	Area% SDS-PAGE	Area (214 nm)	Mass [Da] (ESI-MS)	Mass [kDa] (SDS-PAGE)	Sequence [PSM]	Prot-ID	Sequence (de novo)	Identity (blast)	e-value	Accession Number	Protein Family
		22B	26%		48846	50	YNSDXTVXR	T0033_R_0.0599_L_2024	YNSDXTVXR	snake venom metalloproteinase	2.4E+00	CAJ01688.1	svMP
							HSVAXVEDYSPXDR		STHSPDDPDYGMVDXGTK	zinc metalloproteinase disintegrin-like	2.0E-11	P0DJE2.3	
							XYEM*VNTXNVVFR	T0039_R_0.0184_L_1944	XXCVKPPTGNXXSCK	snake venom metalloproteinase	3.2E-01	AHB62069.1	
		22C	16%		19174	20	TSADYVWXGXWNQR	T0790_R_0.0456_L_550	TSADYVWXGXWNQR	C-type lectin-like protein 2B	4.0E-09	AJO70726.1	CTL
							WTDGSSVXYK		TTDNQWXR	C-type lectin snaclec 7	2.0E-02	Q4PRC6.1	
23	96.4			53265									
		23			48161; 59182	65	-	-	VPXPCANQVXK	snake venom metalloproteinase	7.3E+00	ADI47650.1	svMP
24	104.6			38719									
		24A	80%		57509	65	VPXVGVEXWDHR		VPXVGVEXWDHR	snake venom metalloproteinase	6.5E-02	ADI47590.1	svMP
							pQXVATSEQQR	T0033_R_0.0599_L_2024	-	-	-	-	
							YNSDXTVXR		-	-	-	-	
							VNXXNEMYXPXNXR		-	-	-	-	
							SDPDYAM*VDXGTK		SDPDYAMVDXGTK	snake venom metalloproteinase	7.0E-05	ADI47642.1	
							SVGXXQDYCK	T0039_R_0.0184_L_1944	-	-	-	-	
							CFNYNXQGTENFHCGMENGR		CFNYNXQGTENFHCGMENGR	snake venom metalloproteinase	5.0E-14	AHB62069.1	
		24B	6%		30133	30	XPSSPPSVGSVCR	T0596_R_0.0034_L_647	XPSSPPSVGSVCR	snake venom serine proteinase isoform 7	3.0E-06	ABG26973.1	svSP
							FYCAGYQNNDWDKDXMXXK		FYCAAGYQNNDWDKDXMXXK	snake venom serine proteinase SP-3	8.0E-06	AMB36344.1	
		24C	14%		17061	15	SPEEVDFM*XK		SPEEVDFMXK	C-type lectin-like protein 2B	1.1E-02	AJO70722.1	CTL
							KTWEDAEKFCTEQAR	T0841_R_0.0782_L_536	KTWEDAEKFCTEQAR	C-type lectin snaclec 7	4.0E-11	B4XT06.1	
							GGHXXSXK		HXATXEWXGK	C-type lectin snaclec A16	1.8E-01	B4XSZ1.1	
							ADXVWXGXR		-	-	-	-	
							AWSDEPNCFAAK	T0512_R_0.2394_L_699	-	-	-	-	
							TTDNQWXRR		-	-	-	-	
25	108.7			69780									
		25A	57%		57516	60	HXVVVGAGM*SGXSAAYVXAGAGHK		-	-	-	-	LAAO
							VTVXEASER	T0018_R_0.0657_L_2450	-	-	-	-	
							NVEEGWYANXGPMR		-	-	-	-	
							KFGXQXNEFVQETDNGWYFXK		-	-	-	-	

Peak	RT	SDS-PAGE Band	Area% SDS-PAGE	Area (214 nm)	Mass [Da] (ESI-MS)	Mass [kDa] (SDS-PAGE)	Sequence [PSM]	Prot-ID	Sequence (de novo)	Identity (blast)	e-value	Accession Number	Protein Family
		25A	57%		57516	60	VGEVNKDPGXXK		-	-	-	-	LAAO
							SAGQXYEESXK	T0018_R_0.0657_L_2450	-	-	-	-	
							XFFAGEYTANAHGWXDSTXK		-	-	-	-	
							VPXVGVEXWDHR		VPXVGVEXWDHR	snake venom metalloproteinase	6.5E-02	ADI47590.1	svMP
							pQXVATSEQQR		-	-	-	-	
							XVXVVDHSMVTK		NPCQXYYTPR	snake venom metalloproteinase	6.0E-05	AGL45259.1	
							YNSDXTVXR	T0033_R_0.0599_L_2024	YNSDXTVXR	snake venom metalloproteinase	2.9E-01	ADI47687.1	
							VNXXNEM*YXPXNXR		-	-	-	-	
							HDNAQXXTTXDFDGSVXGK		-	-	-	-	
							HSVAXVEDYSPXDR		-	-	-	-	
							KENDVPXPCAPEDXK		-	-	-	-	
		25B	16%		30131	30	-	-	XPSSPPSVGSVCR	snake venom serine proteinase isoform 7	3.0E-06	ABG26973.1	svSP
		25C	27%		N.D.	15	-	-	-	-	-	-	CTL
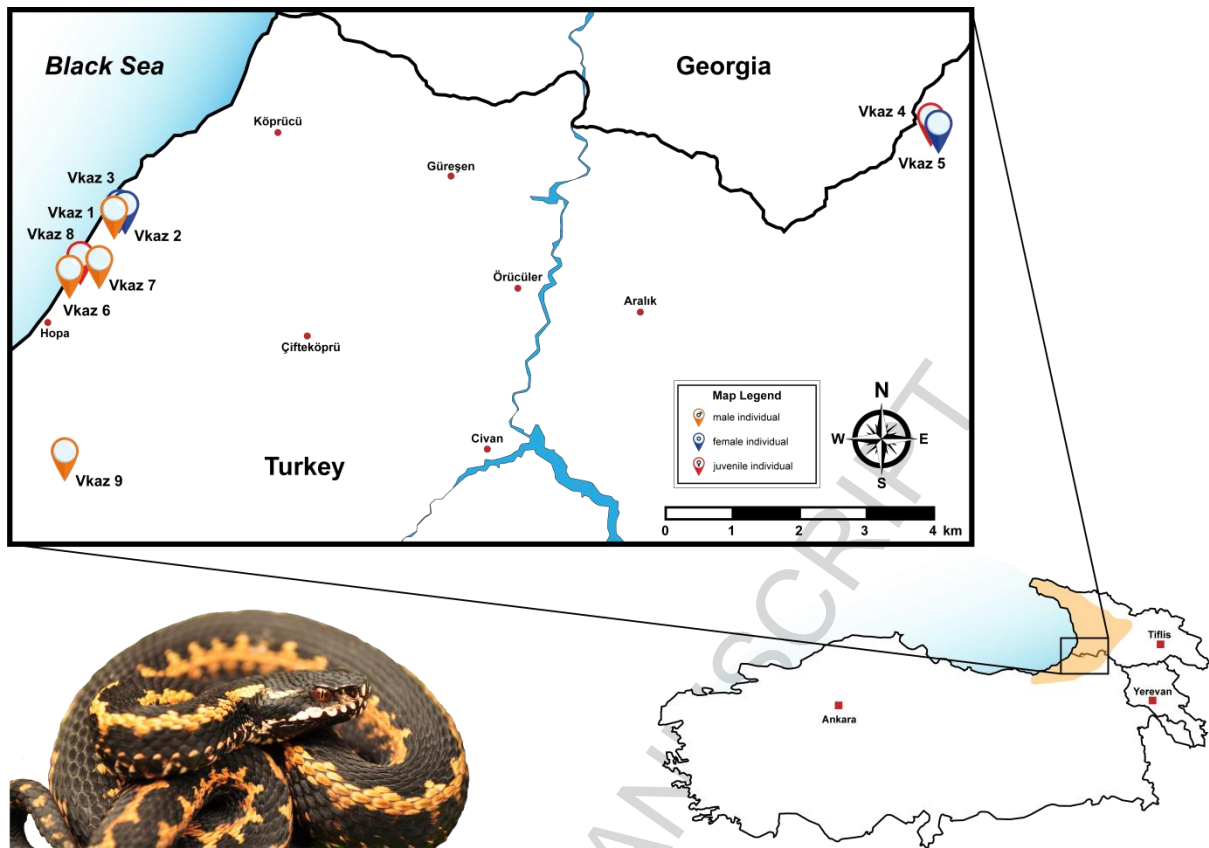
Figure 1. The geographical distribution and sampling localities of *Vipera kaznakovi*. The distribution area of the Caucasus viper (*Vipera kaznakovi*, genus *Viperidae*) is highlighted on the map in the lower right corner and adapted from Geniez *et al*. [84]. The locations and sex/age of the collected individuals are marked on the map (orange – adult male, red – adult female, blue - juvenile).
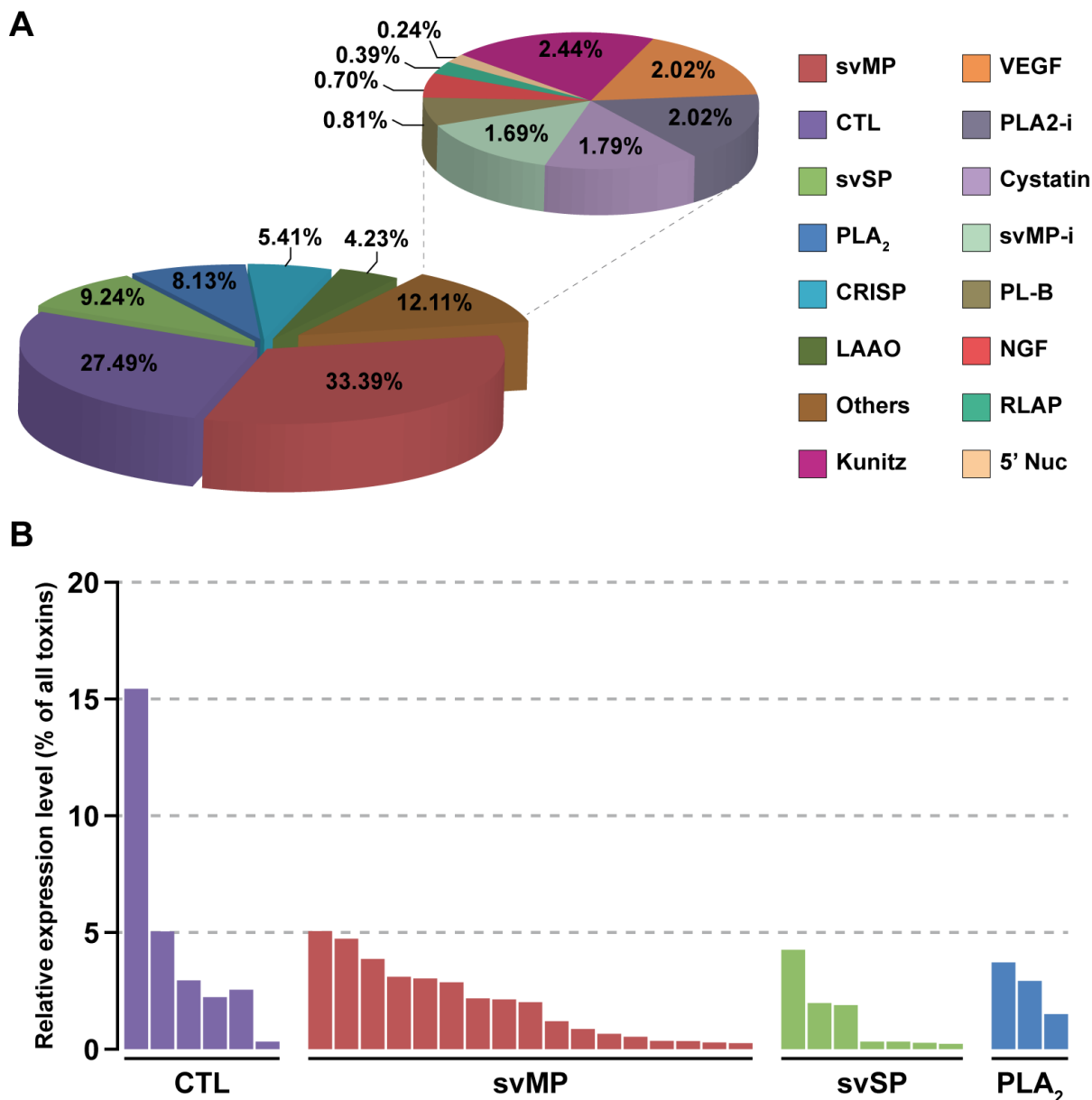
Figure 2. The relative expression levels of toxin families identified in the *Vipera kaznakovi* venom gland transcriptome. A The left pie chart shows the relative expression levels of the major toxin families, each of which accounts for greater than 4% of all toxins encoded in the venom gland. The right pie chart shows the relative expression levels of the remaining toxin families, which in combination account for 12.11% of all toxins encoded in the venom gland ("others"). Percentage values on both charts reflect the expression level of each toxin family as a percentage of the total expression of all identified toxin families. **B** The relative expression levels of individual contigs encoded by the most abundantly expressed toxin families (CTL, svMP, svSP and PLA$_2$). Key: svMP – snake venom metalloproteinase; CTL – C-type lectin; svSP – snake venom serine protease; PLA$_2$ – phospholipase A$_2$; CRISP – cysteine-rich secretory protein; LAAO – L-amino acid oxidase; kunitz – kunitz-type inhibitors; VEGF – vascular endothelial growth factor; PLA2-i – PLA$_2$ inhibitors, SVMP-i – SVMP inhibitors PLB – phospholipase B; NGF – nerve growth factor; RLAP – renin-like aspartic proteases; 5' Nuc – 5' nucleotidase.
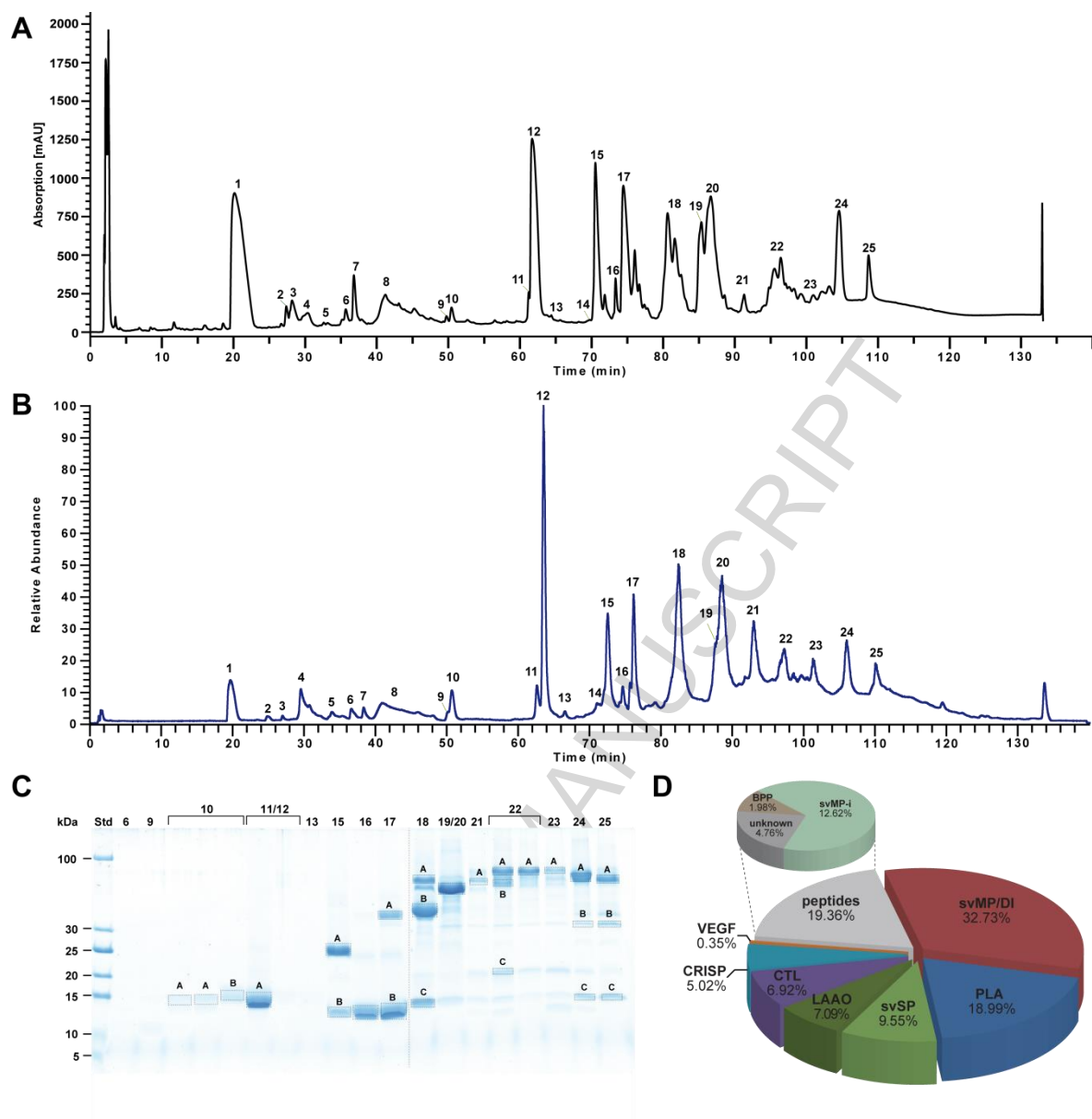
Figure 3. Bottom-up snake venomics of *Vipera kaznakovi*. A Venom separation of *V. kaznakovi* was performed by a Supelco Discovery BIO wide Pore C18-3 RP-HPLC column and UV absorbance measured at λ = 214 nm. **B** Total ion current (TIC) profile of crude *V. kaznakovi* venom. The peak nomenclature is based on the chromatogram fractions. **C** The RP-HPLC fractions (indicated above the lane) of the *V. kaznakovi* venom was analysed by SDS-PAGE under reducing conditions (Coomassie staining). Alphabetically marked bands per line were excised for subsequent tryptic in-gel digestion. **D** The relative occurrence of different toxin families of *V. kaznakovi* are represented by the pie chart. Identification of snake venom metalloproteinase (svMP, red), phospholipases A$_2$ (PLA$_2$, blue), snake venom serine proteinase (svSP, green), C-type lectin like proteins (CTL, purple), cysteine rich secretory proteins (CRISP, light blue), bradykinin-potentiating peptides (BPP, light brown), vascular endothelial growth factors (VEGF-F, red), unknown proteins (n/a, black) and peptides (grey). The *de novo* identified peptides are listed in supplemental table 1.
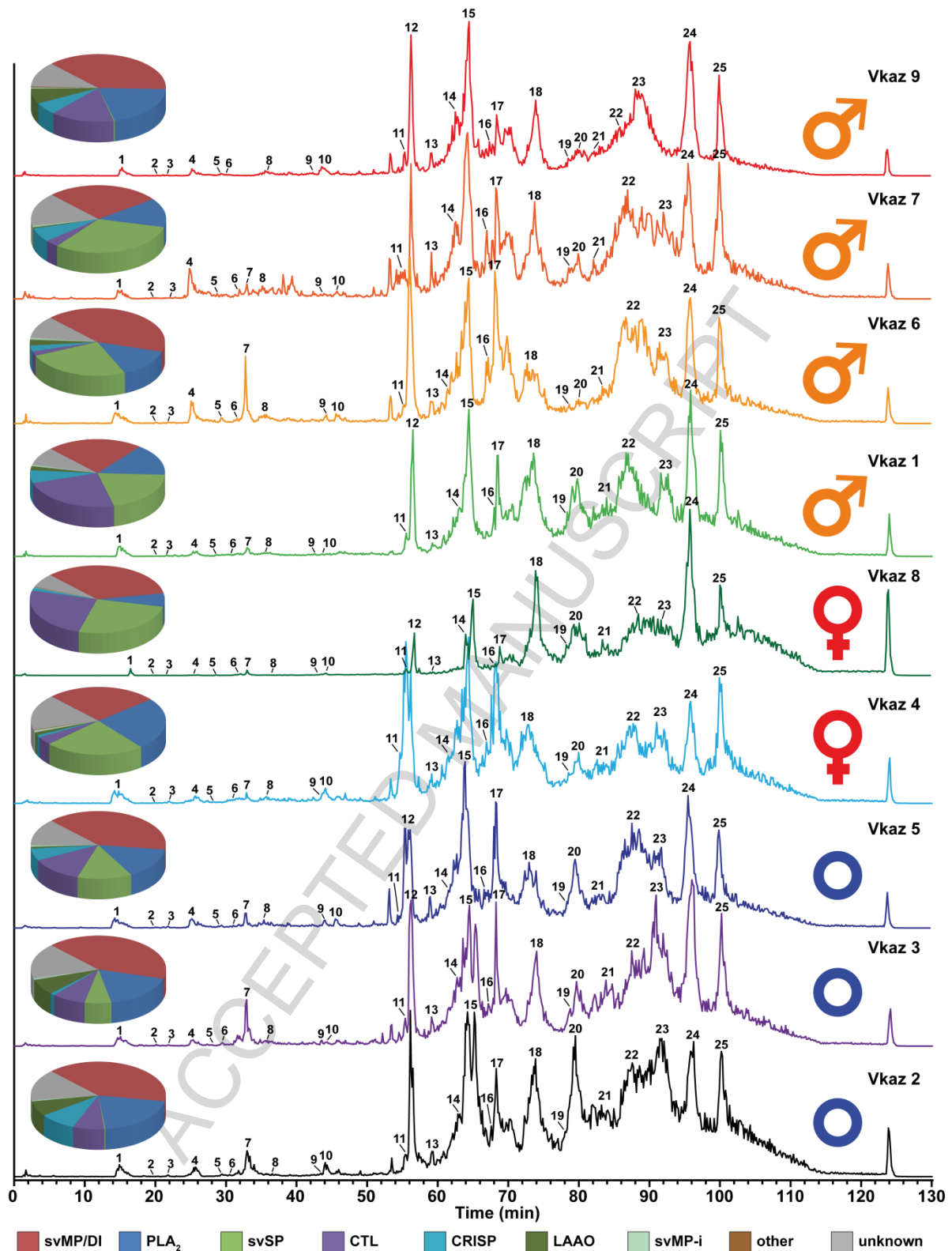
Figure 4. Intact molecular mass profiles of venom from several individuals of *V. kaznakovi*. The total ion counts (TIC) of native, crude venoms from several *V. kaznakovi* individuals were measured by HPLC-ESI-MS. The relative abundance was set to 100% for the highest peak. The peak nomenclature is based on the chromatogram fractions and is shown in figure 3A. The identified molecular masses of intact proteins and peptides are listed in supplemental table 2. The intact molecular mass profiling includes three juveniles of unknown sex (blue circle), and two female (red Venus symbol) and four male (orange Mars symbol) adult individuals.
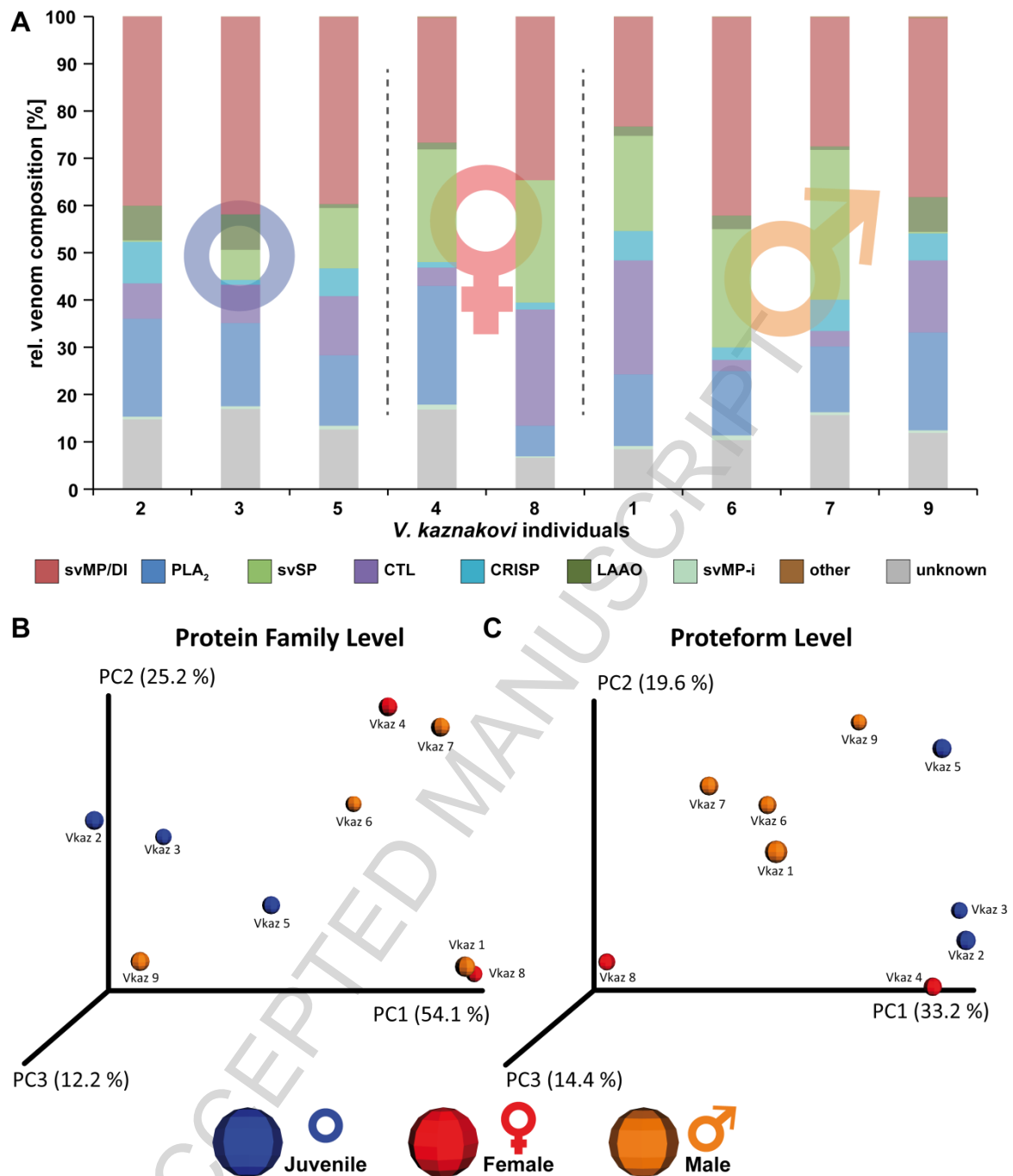
Figure 5. Principal Component Analysis (PCoA) and relative compositions of the individual *V. kaznakovi* venoms. **A** The proteome overview includes three juveniles of unknown sex (blue circle), and two female (red Venus symbol) and four male (orange Mars symbol) adult individuals. **B** and **C** The compositional similarity of venom is displayed through Bray-Curtis-Faith distance in PCoA space. Toxin similarity is visualized at the protein family level (**B**) and proteoform level (**C**).
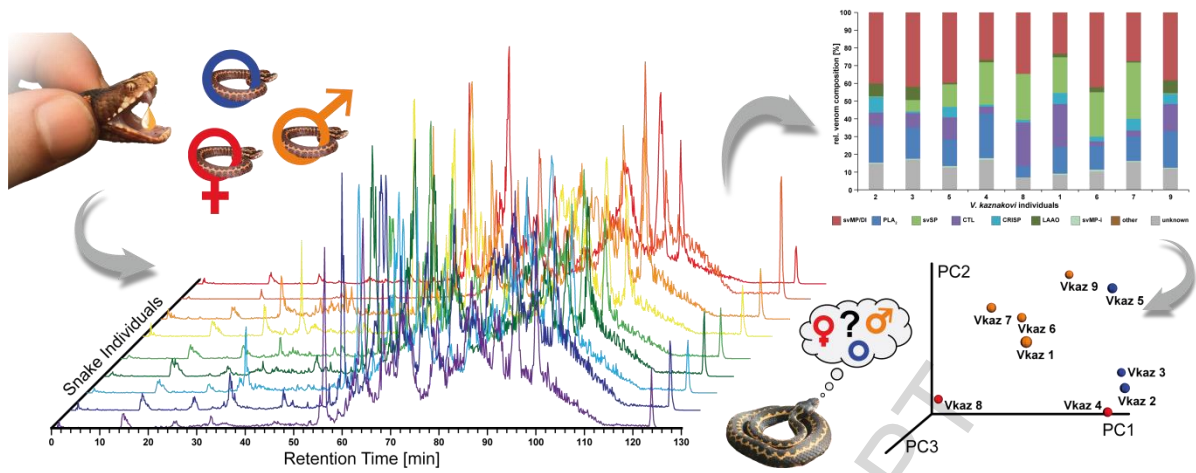
Graphical Abstract

Biological Significance

Population level and ontogenetic venom variation (e.g. diet, habitat, sex or age) can result in a loss of antivenom efficacy against snakebites from wide ranging snake populations. The current state of the art for the analysis of snake venoms are de-complexing bottom-up proteomics approaches. While useful, these have the significant drawback of being time-consuming and following costly protocols, and consequently are often applied to pooled venom samples. To overcome these shortcomings and to enable rapid and detailed profiling of large numbers of individual venom samples, we integrated an intact protein analysis workflow into a transcriptomics-guided bottom-up approach. The application of this workflow to snake individuals of a local population of *V. kaznakovi* revealed intraspecies variations in venom composition, which are primarily explained by the age of the animals, and highlighted svSP abundance to be one of the molecular drivers for the compositional differences observed.

Figure 1
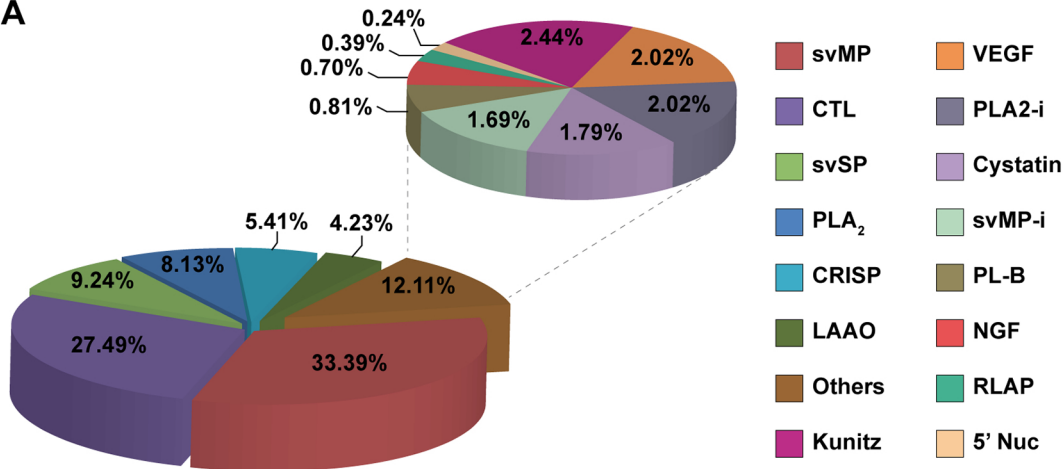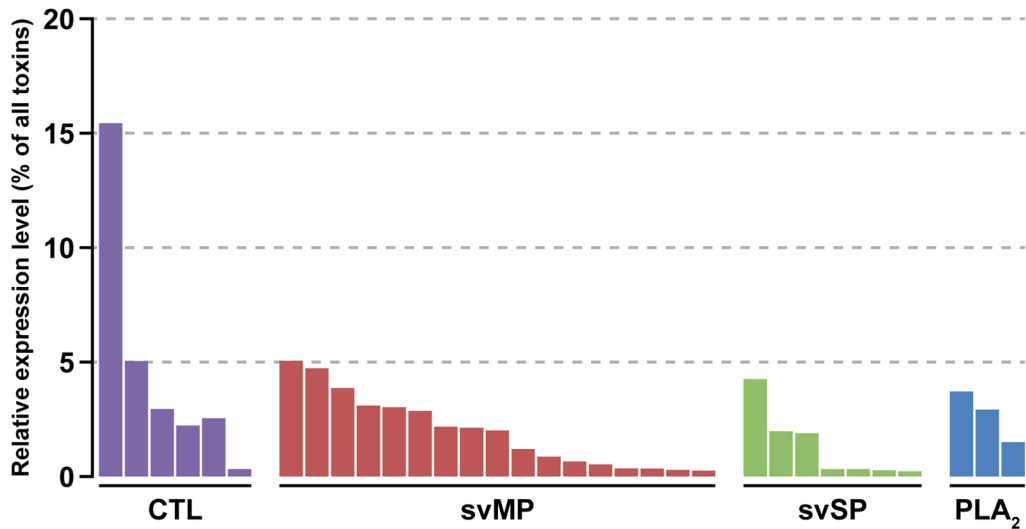
Figure 2

Figure 3

Figure 4

Figure 5