Convergent recruitment of adamalysin-like metalloproteases in the venom of the red bark centipede (*Scolopocryptops sexspinosus*)

Schyler A. Ellsworth[a], Gunnar S. Nystrom[a], Micaiah J. Ward[a], Luciana Aparecida Freitas de Sousa[b], Micheal P. Hogan[a], and Darin R. Rokyta[a,*]

[a] Department of Biological Science, Florida State University, Tallahassee, Florida 32306, USA
[b] Laboratório de Imunopatologia, Instituto Butantan, Av. Vital Brazil, 1500, 05503-900 São Paulo, SP, Brazil
[*] Corresponding author: drokyta@bio.fsu.edu

Running title: The venom of *Scolopocryptops sexspinosus*

Keywords: Centipede, venom, transcriptome, proteome

Corresponding author:
Darin R. Rokyta
Florida State University
Department of Biological Science
319 Stadium Dr.
Tallahassee, FL USA 32306-4295
email: drokyta@bio.fsu.edu
phone: (850) 645-8812

Abstract

Many venom proteins have presumably been convergently recruited by taxa from diverse venomous lineages. These toxic proteins have characteristics that allow them to remain stable in solution and have a high propensity for toxic effects on prey and/or potential predators. Despite this well-established convergent toxin recruitment, some toxins seem to be lineage specific. To further investigate the toxic proteins found throughout venomous lineages, venom proteomics and venom-gland transcriptomics were performed on two individual red bark centipedes (*Scolopocryptops sexspinosus*). Combining the protein phenotype with the transcript genotype resulted in the first in-depth venom characterization of *S. sexspinosus*, including 72 venom components that were identified in both the transcriptome and proteome and 1,468 nontoxin transcripts identified in the transcriptome. Ten different toxin families were represented in the venom and venom gland with the majority of the toxins belonging to metalloproteases, CAPS (cysteine-rich secretory protein, antigen 5, and pathogenesis-related 1 proteins), and β-pore-forming toxins. Nine of these toxin families shared a similar proteomic structure to venom proteins previously identified from other centipedes. However, the most highly expressed toxin family, the adamalysin-like metalloproteases, has until now only been observed in the venom of snakes. We confirmed adamalysin-like metalloprotease activity by means of *in vivo* functional assays. The recruitment of an adamalysin-like metalloprotease into centipede venom represents a striking case of convergent evolution.

1 Introduction

Convergent recruitment of homologous toxic proteins has occurred in venoms of taxa throughout the animal kingdom. (Fry et al., 2009a,b; Casewell et al., 2013; Undheim et al., 2014a). Venom proteins are typically recruited from proteins that are secreted, have high solubility and stability, and influence physiological homeostasis (Fry et al., 2009b). These shared characteristics lead to a propensity for toxic effects. Cysteine-rich secretory protein, antigen 5, and pathogenesis-related 1 protein domains (CAPs) exemplify this convergent recruitment into animals venoms. CAPs have been identified in both vertebrate and invertebrate venoms, spanning the venoms of insects, cephalopods, scorpions, centipedes, cone snails, and snakes (Fry et al., 2009a,b). Although CAPs share a highly conserved domain structure, this protein family has a diverse set of functions including proteolytic activity, protease inhibition, and ion-channel regulation (Gibbs et al., 2008). The stable structure of this protein family along with the large breadth of ancestral activity have allowed CAPs to be ubiquitously recruited into venom glands as a toxin.

Centipede venoms characterized to date contain toxins from similar protein families as those identified in many other venomous taxa, including CAPs, hyaluronidase (HYLA), phospholipase A_2 (PLA2), serine proteases (SPs), and various neurotoxic peptides (Fry et al., 2009a,b; Yang et al., 2012; Undheim et al., 2015; Hakim et al., 2015; Ward and Rokyta, 2018). Centipede venoms also contain unique toxin families unknown in other venomous animals. These toxins include scoloptoxins (SLPTXs) and a class of proteins with a common set of domains of unknown function (DUFs) (Yang et al., 2012; Undheim et al., 2014a; Hakim et al., 2015; Undheim et al., 2015; Ward and Rokyta, 2018). Although a rich diversity of novel toxins have been identified in centipedes, most of the research in centipede venoms has focused on a single family, Scolopendridae, in the order Scolopendramorpha (Liu et al., 2012; González-Morales et al., 2014; Undheim et al., 2014a, 2015; Ward and Rokyta, 2018).

To elucidate potential unexplored venom diversity in a non-scolopendrid species, we characterized the venom-gland transcriptome and venom proteome of the red bark centipede (*Scolopocryptops sexspinosus*, Figure 1). *Scolopocryptops sexspinosus* belongs to the family Scoloprocryptopidae (92 currently recognized species) and makes up approximately 12% of the order Scolopendramorpha, which are the largest and most commonly recognized centipedes (Undheim et al., 2015; Bonato et al., 2016). Most of the venom research to date has focused on the family Scolopendridae (Undheim et al., 2015); we provide the first in-depth characterization of a centipede venom from the family Scolopocryptopidae. *Scolopocryptops sexspinosus* is a common centipede that occurs primarily in the eastern United States, ranging from as far north as Ontario, Canada, to Florida, and as far west as Nebraska and Texas (Figure 1; Shelley, 2002). Two independent venom-gland transcriptomes of *S. sexspinosus* were sequenced and quantitative mass spectrometry was conducted on their venoms to provide an in-depth venom characterization of a scolopocryptopid centipede. This venom characterization revealed the

identity and function of a new, highly expressed toxin family for centipedes.

2 Materials and Methods

2.1 Centipedes, venoms, and venom-glands

Two individuals of *S. sexspinosus* were collected in northern Florida in Madison and Leon counties and labeled C0142 and C0184, respectively. Venom and venom-glands were collected from each specimen as described by Ward and Rokyta (2018). Centipedes were briefly anesthetized in CO_2, and venom was extracted through electro-stimulation of the forcipules resulting in a muscle contraction releasing the venom. Venom was then collected in LC/MS grade water, lyophilized, and stored at -80°C until further use. Four days following venom extraction, venom-glands were removed under a stereomicroscope and stored in RNAlater (Qiagen). The extracted venom-glands were stored at 4°C for 24 hours, then stored at -80°C until RNA extraction.

2.2 Transcriptome sequencing

RNA extraction was performed as previously described (Rokyta and Ward, 2017; Ward et al., 2017; Ward and Rokyta, 2018). Venom-gland tissue was first homogenized in 500 μL Trizol using a 20-gauge needle and syringe. An additional 500 μL of Trizol was added to completely dissolve any remaining tissue. The Trizol solution with the homogenized tissue was then transfered to phase lock heavy gel tubes (5Prime) and mixed with 200 μL of chloroform. The tubes were then centrifuged at 12,000×G for 20 minutes to isolate RNA from the DNA and other cellular components. RNA was then precipitated with the addition of isopropyl alcohol, and the pellet was rinsed with 75% ethanol. RNA was resuspended in H_2O, and the quality and concentration was verified using a Total RNA 6000 Pico Bioanalyzer chip (Agilent Technologies) according to the manufacturer's instructions. Approximately 116 ng and 227 ng of RNA were isolated from C0142 and C0184, respectively. RIN scores are not typically used to asses quality of invertebrate RNA because of the comigration of 28s rRNA fragements with the 18s rRNA (Paszkiewicz et al., 2014). Therefore, RNA quality was determined based on the presence of the double peak of the 28s rRNA fragments in relation to the 18s rRNA.

mRNA was isolated from 90–100 ng total RNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs). To achieve a fragment size of approximately 370 base pairs, a fragmentation step of 15.5 minutes was used, consistent with methods previously described (Rokyta and Ward, 2017; Ward et al., 2017; Ward and Rokyta, 2018). Fragmented mRNA was immediately used to construct a cDNA library by using the NEBNext Ultra RNA Library Prep Kit with the High-Fidelity 2X Hot Start PCR Master Mix and Multiplex Oligos for Illumina (New England Biolabs) according to the manufacturer's instructions. Agencourt AMPure XP Purification Beads were used throughout the cDNA purification steps. To determine the quality, concentration, and

average length of the cDNA libraries, each library was analyzed using a High Sensitivity DNA Bioanalyzer Kit (Agilent Technologies). The total cDNA yield for C0142 was 18 ng (20 μL of 4.0 nM) with an average fragment size of 391 bp, and the total cDNA yield for C0184 was 158 ng (20 μL of 34.6 nM) with an average fragment size of 406 bp. To find the amount of amplifiable cDNA, KAPA PCR was performed by the Florida State University Molecular Cloning Facility. The amplifiable concentration for each sample was 14.7 nM and 62.1 nM for C0142 and C0184, respectively. Equal concentrations of amplifiable cDNA of these two libraries and were pooled with other libraries into one sample to be run on the same Illumina sequencing lane. The quality, concentration, and average base pair length of the pooled sample was then verified utilizing a High Sensitivity DNA Bioanalyzer Kit (Agilent Technologies), and KAPA PCR was used again to find the amplifiable concentration of the pooled sample. The pooled sample was sequenced by the Florida State University College of Medicine Translational Laboratory using an Illumina HiSeq 2500 using a 150 paired end read format.

2.3 Proteomics

Proteomic analysis was performed as previously described (Rokyta and Ward, 2017; Ward et al., 2017; Ward and Rokyta, 2018). Protein concentrations of each venom sample were quantified using the Qubit Protein Assay Kit with a Qubit 1.0 Fluorometer (Thermo Fisher Scientific). Five micrograms of whole venom was trypsin digested for mass spectrometry utilizing the Calbiochem ProteoExtract All-in-One Trypsin Digestion Kit (Merck, Darmstadt, Germany). Prepared samples were dried in a speedvac, and held at -20°C until mass-spectrometry analysis.

LC-MS/MS analysis of the digested venom peptides was performed by Florida State University College of Medicine Translational lab. Samples were run in triplicate on a Thermo Q Exactive HF mass spectrometer used in conjunction with a Dionex Ultimate 3000 RSLCnano System. The Q Exactive HF mass spectrometer was used in data-dependent mode controlled by Thermo Excalibur 3.1.66 (Thermo Scientific) software. Data analysis of the raw files was performed using Proteome Discover 1.4 with Sequest HF as the search engine. Proteome Discover searched through custom FASTA data files to discover peptides and used percolator to validate the peptides (Spivak et al., 2009). Identities of proteins and peptides were validated using Scaffold (version 4.3.4, Proteome Software Inc., Portland, OR, USA) software with the protein and peptide thresholds set to 1% false discovery rate and the minimum number of peptides set to one.

To determine the protein abundances, known concentrations of three highly-purified recombinant *Escherichia coli* proteins were included with each sample. Comparing the known concentrations of the *E. coli* proteins to the normalized spectra counts obtained from Scaffold, we calculated conversion factors based on the slope of the best fit line. Spectra counts for each individual protein identified in Scaffold were then converted to a concentration using the conversion factors obtained. The final concentration of each protein was obtained based on the average concentration among the three replicates.

2.4 Transcriptome assembly and analysis

Analysis and assembly of the transciptomic data was performed as previously described (Rokyta and Ward, 2017; Ward et al., 2017; Ward and Rokyta, 2018). Trascriptomic data generated from 150 paired-end sequencing was filtered with Illumina quality filtering. Because Illumina sequencing is heavily biased toward the smallest fragments in the library, a target distribution of approximately 250 base pairs would mean that the final sequenced insert size is, on average, smaller allowing for sufficient 3' read-pair overlap. Paired reads were then merged using PEAR version 0.9.6 (Zhang et al., 2014) and used in the ensuing analyses. The primary transcriptome assembly was generated with DNAStar NGen version 12.3.1 with 10 million merged reads under default settings. Only contigs with at least 200 reads were retained. Multiple search strategies were employed to identify and annotate toxins as we did not expect to find many toxin homologs searching the available public databases. To be conservative in the assignment of transcripts as venom genes, only contigs that were identified in both the transcriptome and proteome were considered toxins for further analysis. The first two strategies used TransDecoder version 2.0.1 (Haas and Papanicolaou, 2016), and the mass-spectrometry results. TransDecoder was used to create a database of predicted protein sequences, with a minimum sequence of 50 amino acids, from each transcriptome assembly. Each database was then used to search against the mass-spectrometry results. These results were then filtered using Scaffold Viewer version 4.6.0 with the protein and peptide false discovery rate set to 1.0% and the minimum number of peptides set to one. To not omit any peptides, a protein database was also created via Transdecoder that included all proteins and peptides of at least 50 amino acids from the six possible reading frames in each contig. Results were then filtered in Scaffold as described above, excluding all of the contigs found in the first strategy. The third strategy attempted to identify toxins based on homology, by using the transcripts generated by NGen in a BLASTX search (version 2.2.30+) against the UniProt animal toxins database (downloaded on November 16, 2015; Jungo et al., 2012). Full length transcripts were annotated if they matched against a known toxin with at least 80% length coverage and had an e-value of less than 0.0001. The fourth strategy utilized a BLASTX search of all of the reads generated by NGen against the National Center for Biotechnology Information (NCBI) non-redundant (nr) protein database (downloaded on November 13, 2015). This analysis created a general database of toxin and non-toxin transcripts. Full length transcripts were included if they matched 90% of the length a known protein and had an e-value of less than 0.0001. The fifth strategy utilized the de novo assembler Extender (Rokyta et al., 2012) to ensure that no high abundance transcripts were missed, by assembling a transcriptome starting from 1,000 random reads as seeds. Random reads were extended using all of the reads if they had an exact match of 120 nucleotides for extension, and if they had phred qualities of $\geq$30. The contigs were then searched against the UniProt animal toxins database using BLASTX.

To generate a consensus transcriptome, annotated contigs from the five different search strategies were combined for each individual. Exact duplicates in the dataset

were then removed utilizing cd-hit-est version 4.6 (Li and Godzik, 2006) on coding sequences. To screen for chimeric sequences or other coverage anomalies, reads were then merged with PEAR version 0.9.6 (Zhang et al., 2014) and aligned using bwa version 0.7.12 (Li, 2013), retaining reads only if there were no mismatches relative to the reference. Alignments were then screened for coverage anomalies that included regions of no coverage and/or multimodal coverage distributions.

A final quality-control step was performed on the assembled transcripts, because the two *S. sexspinosus* RNA-seq libraries were sequenced alongside other RNA-seq libraries from different species. Reads from each library, sequenced in the same lane as the two *S. sexspinosus*, were merged with PEAR version 0.9.6 (Zhang et al., 2014). These reads were then aligned against the coding sequences in the combined transcriptome using bwa version 0.7.12 (Li, 2013). Transcripts were considered contaminants, if there were differences in the full length consensus sequence, and if their abundance was $>100\times$ for another library in comparison to the two *S. sexspinosus* libraries. Contaminants were subsequently removed from the final transcriptome dataset.

Due to the scarcity of proteomically confirmed venom proteins, six additional transcriptome assemblies utilizing a proteomic-driven annotation were employed for each individual (Ward and Rokyta, 2018). Raw reads were first processed and screened for cross-leakage due to demuliplexing and k-mer distributions from each individual were compared against the distributions of other samples that were run in the same lane. 57-mer distributions were generated using jellyfish version 2.2.6 (Marçais and Kingsford, 2011). Reads were removed if $\geq$25% of their length was comprised of 57-mers that had distributions $500\times$ more abundant in another sample than in *S. sexspinosus*. Trim Galore! (Krueger, 2015) was used for adapter and quality trimming with a quality threshold set to a phred of 5 (MacManes, 2014), and we removed any trimmed reads less than 75 nucleotides in length. Reads were merged with PEAR version 0.9.10 (Zhang et al., 2014) utilizing default settings. The first assembly method used was a replicate run of Extender (Rokyta et al., 2012), using 1,000 random seeds with a minimum phred of 30, an overlap of 120 nucleotides, 20 replicates, and using only the merged reads with a minimum phred of 20. This assembly was independent of the first Extender run. Because Extender uses random seeds, the two runs could generate different results. The second and third methods used BinPacker version 1.0 (Liu et al., 2016) and Trinity version 2.4.0 (Grabherr et al., 2011), respectively, using merged and unmerged reads with a k-mer size of 31, and treating all reads as unpaired. The fourth assembly was run using SOAPdenovo-trans version 1.03 (Xie et al., 2014) with merged reads, unmerged reads treated as pairs, and a k-mer size of 127. The fifth assembly was generated with SeqMan NGen version 14.0 using both the merged and unmerged reads, treating all as unpaired. The last assembly method was rnaSPAdes version 3.10.1 (Bankevich et al., 2012) with $k = 127$, using both the merged and unmerged reads, with the unmerged reads treated as paired. Amino-acid sequences from all of the possible reading frames were extracted using the getorf function from EMBOSS version 6.6.0.0 (Rice et al., 2000) with a minimum size of 90 nucleotides and only retained the open reading frames that contained both a start and a stop codon.

The output from each assembly was then clustered to remove exact duplicates using cd-hit version 4.6 (Li and Godzik, 2006) with a sequence identity threshold of 1.0. Each assembly method was then converted into a database and used to search against the mass-spectrometry results as described above.

The final transcriptome for *S. sexspinosus* was achieved by merging the initial consensus transcriptome for each individual with the additional six assembly approaches. Transcripts were then clustered and analyzed using only coding sequences in cd-hit-est using a global sequence identity of 0.98. Using all of the merged reads for each individual, transcriptome abundances were estimated with RSEM (Li et al., 2011) version 1.2.28 utilizing the alignments from bowtie2 (Langmead and Salzberg, 2012) version 2.3.0. All of the transcriptome and proteome abundances were centered logratio transformed (Aitchison, 1986) as described in (Rokyta et al., 2015). This transformation is comparable to a log transformation for linear analyses and does not change rank-based analyses. SignalP version 4.1 was used to determine signal peptides using the default settings (Petersen et al., 2011). Toxin family identification and naming was obtained by searching annotated toxins against previously reported centipede toxins in the NCBI transcriptome shotgun assembly (TSA) database. Because of the number of assembly methods and clustering steps involved in generating the final consensus transcriptome, we performed a BLASTN search to determine which assembly method had assembled each final toxin. We used the final consensus transcriptome to search against all of the contigs from each assembly only accepting matches with full length coverage and a 99% sequence identity. This allowed us to determine which final toxin sequences were generated under each assembly method.

2.5 Reversed-phase high performance liquid chromatography

Reversed-phase high performance liquid chromatography was performed on one venom sample from each individual (Ward et al., 2018; Ward and Rokyta, 2018). Approximately 7 μg of venom was injected onto an Aeris 3.6 μm C18 column (Phenomenex, Torrance, CA) using 0.1% trifluoroacetic acid (TFA) in water for solvent A and 0.06% TFA in acetonitrile for solvent B, on a Waters 2695 Separations Module with a Waters 2487 Dual λ Absorbance Detector. The sample gradient started at 10% B for five minutes with flow rate of 0.2 mL/min, and then changed from 10% B to 55% B over 110 minutes, followed by a wash step of five minutes at 75% B and ten minutes at 10% B.

2.6 Enzymatic activity

Metalloprotease activity of the *S. sexspinosus* venom was tested using the gelatin fluorescein conjugate substrate (Thermo Fisher Scientific, US) from adapted methods previously described (Knittel et al., 2016). *Crotalus adamanteus* venom served as a positive control because of its high metalloprotease activity (Margres et al., 2016). For the assay, venom samples were incubated with 100 μg/mL of gelatin fluorescein conjugate with 50 μg of *S.*

sexspinosus or 20 μg of the *C. adamanteus* venoms in the reaction buffer (50 mM Tris-HCl, 50 mM $CaCl_2$, 1.5 M NaCl, pH 7.6). The assay was performed in the absence and presence of 1,10-Phenanthroline (inhibitor of metalloproteases) or phenylmethylsulfonyl fluoride (inhibitor of serine proteases). All reactions were carried out at 25°C and monitored every three minutes over an hour using a SpectraMax M2 fluorometer (Molecular Devices) at λEM 495 nm and λEX 535 nm. The specific activity of three independent experiments was expressed in relative fluorescence units (RFU/min/μg).

2.7 Hemorrhagic activity

The hemorrhagic activity was based on the method described by Kondo et al. (1960) modified by Gutiérrez et al. (2005). Swiss mice of both sexes (18–20 g body weight) were obtained from Instituto Butantan animal house. The procedures used during the experiments were approved by the Animal Use and Ethic Committee (CEUAIB) of the Instituto Butantan (Protocol 67080408/17). They are in accordance with COBEA guidelines and the National law for Laboratory Animal Experimentation (Law No. 11.794, 8 October 2008). For the assay, 30 μg of *S. sexspinosus* venom and 10 μg of *C. adamanteus* venom was injected intradermally into the dorsal skin of mice both in the absence and presence of 1,10-Phenanthroline. All the samples were incubated for 30 minutes at 37°C either in solution or with 50 μM of 1,10-Phenantroline prior to injection. After 1h, the mice were euthanized in a CO_2 chamber and the dorsal skin was removed. The hemorrhagic area was measured in mm^2.

2.8 Data availability

The raw transcriptome reads were submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject PRJNA340270, BioSamples SAMN10423647 (C0142) and SAMN10423648 (C0184), and SRA accessions SRR8188017 and SRR8188018 for C0142 and SRR8188015 and SRR8188016 for C0184. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al., 2016) partner repository with the dataset identifier PXD011714. The assembled transcripts were submitted to the NCBI Transcriptome Shotgun Assembly database. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/ EMBL/GenBank under the accession GHBZ00000000. The version described in this paper is the first version, GHBZ01000000.

3 Results and Discussion

3.1 Venom-gland transcriptomes

For individual C0142, 12,697,799 read pairs remained after Illumina quality filtering, and 10,300,282 of those were merged. The resulting merged reads had an average length

of 168 base pairs. Merged reads were assembled into the primary C0142 transcriptome and consisted of 3,585 contigs from NGen supported by 6,869,596 reads. Through the MS-directed analysis, 46 unique coding sequences were annotated using TransDecoder and all possible open reading frames (ORFs). Seventy-four unique coding sequences were annotated using BLASTX hits to the UniProt toxins database, and 1,160 unique coding sequences were annotated using BLASTX hits to the NCBI nr database. Using the Extender assembly, another 14 unique coding sequences were annotated by performing a BLASTX search of the UniProt animal toxins database. After screening for duplicates and chimeras, 887 unique coding sequences were identified for C0142.

For individual C0184, 19,186,565 read pairs remained after Illumina quality filtering, and 15,985,989 of those were merged. The resulting merged reads had an average length of 171 base pairs. Merged reads were then assembled into the primary C0184 transcriptome and consisted of 3,830 contigs from NGen supported by 7,234,572 reads. Through the MS-directed analysis, 49 unique coding sequences were annotated using TransDecoder and all possible ORFs. Sixty-two unique coding sequences were annotated using BLASTX hits to the UniProt toxins database, and duplicates from C0142 were excluded. We annotated 1,428 unique coding sequences using BLASTX hits to the NCBI nr database. Using the Extender assembly, another 23 unique coding sequences were annotated by performing a BLASTX search of the UniProt animal toxins database. After screening for duplicates and chimeras, 1,191 unique coding sequences were identified for C0184.

A proteomic-driven annotation using six additional assemblies for each individual was completed because of the dearth of publicly available proteomically confirmed centipede toxins. As described in the methods, raw reads were processed, filtered, and merged. 46 unique coding sequences for C0142, and 38 unique coding sequences for C0184 were annotated only using merged reads from Extender. Then using the other assemblies with unpaired reads, 42 (C0142) and 39 (C0184) unique coding sequences were annotated from the BinPacker assembly, 39 (C0142) and 37 (C0184) unique coding sequences were annotated from the Trinity assembly, and 42 (C0142) and 37 (C0184) unique coding sequences were annotated from the SeqMan NGen assembly. Then treating merged and unmerged reads as paired, 15 (C0142) and 23 (C0184) unique coding sequences were annotated from the SOAPdenovo-trans assembly, and 36 (C0142) and 29 (C0184) unique coding sequences were annotated from the rnaSPAdes assembly. Utilizing six different assemblies resulted in a combined total of 68 and 57 unique coding toxin sequences for C0142 and C0184, respectively.

The final consensus transcriptome consisted of 1,540 unique protein-coding transcripts and was utilized for all subsequent transcript-abundance and LC-MS/MS analyses. Transcripts were separated into two categories: toxins and nontoxins. Seventy-two toxin transcripts were identified in the proteome of one or both *S. sexspinosus*. An overview of which toxins were assembled using each assembly method is shown in Figure 2. Each assembly method assembled an average of 39 toxins for both C0142 and C0184, with SOAPdenovo-trans producing the lowest number of final toxin sequences (20 for

C0142 and 24 for C0184) and NGen producing the highest (50 for C0142 and 52 for C0184). These transcripts accounted for 733,945.84 and 712,844.36 transcripts per million (TPM) for the mapped reads of C0142 and C0184, respectively. The 1,468 nontoxin transcripts identified in *S. sexspinosus* accounted for 266,054.20 and 287,155.47 TPM of the total mapped reads for C0142 and C0184, respectively. These nontoxin transcripts likely encode for proteins that are essential for proper cell function and protein production, but have a low probability of encoding for proteins with toxic functions because they were not detected in proteome.

3.2 A common set of centipede toxins

Even with the limited sampling of centipede venoms, a common set of toxins is emerging that includes β-pore-forming toxins (βPFTxs), CAPs, M12A proteases, SPs, SLPTXs, PLA2s, HYLA, low-density lipoprotein receptor class A repeat domain proteins (LDLA), and DUFs (Undheim et al., 2015). All of these proteins were detected in the venom-gland transcriptome and venom proteome of *S. sexspinosus*.

βPFTxs are non-enzymatic proteins that constitute a major component of centipede venoms (Undheim et al., 2014a, 2015; Ward and Rokyta, 2018). Most of the pore-forming toxins identified in *S. sexspinosus* contain a domain similar to aerolysin toxin, a βPFTx found in bacteria, anemones, and hydras (Sher et al., 2005; Moran et al., 2012; Dal Peraro and Van Der Goot, 2016). Ten different βPFTxs were identified as toxins in *S. sexspinosus* (Table 1). They constituted 13.4% and 10.6% of the toxin transcriptome, and 11.5% and 12.8% of the proteome of C0142 and C0184, respectively. The specific targets and function of βPFTxs in centipede venoms is still unknown, but it has been hypothesized that βPFTxs act to produce edema and myotoxicity (Malta et al., 2008; Undheim et al., 2014a).

The CAP family of proteins is extensively distributed in animal venoms (Gibbs et al., 2008; Fry et al., 2009b). This protein family has also been shown to be an abundant component of centipede venoms (Rates et al., 2007; Liu et al., 2012; Undheim et al., 2014a, 2015; Ward and Rokyta, 2018). Based on the phylogeny generated by Joshi and Karanth (2011), Undheim et al. (2014a) divided CAPs into three different classes based on their recruitment into centipede venom. CAP1 was recruited prior to the split of Notostigmophora and Pleurostigmophora, CAP2 prior to the division of Scolopendridae, and CAP3 within the genus *Scolopendra*. All of the CAPs identified in *S. sexspinosus* belong to the class CAP1, which is consistent with the classes described in Undheim et al. (2014a) as *S. sexspinosus* belongs to the family Scolopocryptopidae which is basal to all of the other families in Scolopendramorpha (Joshi and Karanth, 2011). Four transcripts were identified (Table 1) and constituted 10.5% and 2.7% of the toxin transcriptomes of the C0142 and C0184, respectively. CAPs were also abundant in the proteomes of C0142 and C0184 (14.9% and 17.4%, respectively). The activity of CAP1 is still unknown.

Four different classes of proteases were identified as toxins in *S. sexspinosus*: adamalysin-like metalloprotease (discussed below), M12A peptidases, M13 peptidases,

and SPs (Table 1). Besides adamalysins, the M12A proteases were the most abundant with eight unique proteins representing 7.1% and 6.4% of the total toxin transcriptome and 12.8% and 15.1% of the proteome for C0142 and C0184, respectively. M12A proteases are astacin-like metalloendoproteases that are a major component of centipede venoms (Undheim et al., 2014a, 2015; Ward and Rokyta, 2018). The role of these proteases in centipedes is yet to be determined. All of the M12A proteases contained a CUB domain. A single M13 peptidase, or neprilysin-like peptidase, was identified in both individuals and contained a peptidase family M13 domain. This protein does not constitute a large part of the transcriptome or proteome (<1.0% for C0142 and C0184). Six SPs were detected in the toxin transcriptome and the proteome of *S. sexspinosus* including both S1 and S8 proteases (Table 1). Five of the SPs were identified as S1 proteases and contained a trypsin-like serine protease domain and matched to other previously identified centipede toxins. The last SP (Peptidase-1) was an S8 protease. This protease contained an S8 pro-domain, a peptidase S8 family domain, and a proprotein convertase P-domain. The SPs accounted for 1.3% and 2.0% of the toxin transcriptome and 0.4% and 3.3% of the proteome for C0142 and C0184, respectively. The function for SPs in centipede venom is still unknown, but it has been hypothesized that they could function in the activation of other toxins (Undheim et al., 2015).

Scoloptoxins (SLPTXs) are a structurally diverse group of proteins that contains families of centipede toxins characterized as cysteine rich peptides that exhibit a broad range of functions (Yang et al., 2012; Undheim et al., 2014a,b, 2015; Rong et al., 2015). Five different SLPTXs were identified in *S. sexspinosus* (SLPTX1, SLPTX4, SLPTX15, ProtCw1a, SLPTX-1, Table 1). All five SLPTXs were identified in the transcriptome of both individuals, but ProtCw1a (An SLPTX identified from the proteome of *Cormocephalus westwoodi* by Undheim et al., 2014a) and SLPTX4 were only detected in the proteome of C0184. SLPTXs accounted for 11.3% and 53.7% of the toxin transcriptome and 1.1% and 3.7% of the proteome for C0142 and C0184, respectively. Three of the SLPTXs identified were grouped into SLPTX families based on sequence similarity (SLPTX1, SLPTX4 and SLPTX15). Two SLPTXs showed sequence similarities with SLPTXs with no family designation (ProtCw1a and SLPTX-1), but based on their cysteine rich configuration and molecular mass (Undheim et al., 2015), they were grouped into SLPTX families SLPTX8 and SLPTX4 respectively.

PLA2s are commonly recruited into animal venoms, and have a variety of different functions that are described in greater detail in other studies (Fry et al., 2009b). PLA2s have been characterized in other centipedes (Gonzáles-Morales et al., 2009; Liu et al., 2012; Undheim et al., 2014a), including verification of PLA2 enzymatic activity (Gonzáles-Morales et al., 2009; Malta et al., 2008). One transcript was found in both the transcriptome and proteome of each individual *S. sexspinosus* (Table 1).

One hyaluronidase (HYAL) was identified both transcriptomically and proteomically in both individual *S. sexspinosus*. HYAL activity has been observed in centipedes (Malta et al., 2008), and HYALs are generally considered spreading factors in venomous organisms because of their ability to hydrolyze glycosaminoglycans that are distributed in a

variety of tissues and the extracellular matrix of predators and prey (Girish et al., 2004; Undheim et al., 2014a, 2015; de Oliveira-Mendes et al., 2019). The HYAL identified in *S. sexspinosus* may function in a similar role.

LDLAs are an abundant component of centipede venom, but their function is still undetermined (Liu et al., 2012; Undheim et al., 2015; Smith and Undheim, 2018; Ward and Rokyta, 2018). The diversity of LDLAs seen in other centipedes is repeated in *S. sexspinosus*, as six different LDLAs were identified (Table 1). LDLAs accounted for 5.8% and 6.0% of the total toxin transcriptional output for C0142 and C0184, respectively and 4.5% of the total proteomic output for both individuals. LDLA-3 was absent in both the transcriptome and proteome of C0184 and LDLA-6 was absent from the proteome of C0142 (Table 2).

An additional family of toxins found in centipedes are proteins that contain a domain of unknown function (DUF) (Undheim et al., 2014a, 2015). In *S. sexspinosus* four DUFs were identified that all contained DUF3472 and DUF5077 domains (Table 1). DUFs were responsible for 5.3% and 0.9% of the total toxin transcriptional output and 7.1% and 4.8% of the proteomic output for C0142 and C0184, respectively.

Ten other proteins were identified in the transcriptome and proteome of *S. sexspinosus*. These proteins consisted of β-amyloid, protein BAT5 (BAT5), chorion peroxidase (Chorionperoxi), Coatomer subunit β (CoatomerB), fumarylacetoacetase, golgin subfamily A member 2 (GolginA2), leucine-tRNA ligase (LeutRNALiga), mitogen-activated protein kinase kinase kinase 15 (MAPK15), neutral α-glucosidase AB (NeutralphaglucosiAB), and rho guanine nucleotide exchange factor 7 (RhoGuanExchafact7). These proteins were only identified in a single individual with two of them (BAT5 and RhoGuanExchaFact7) only being identified in one of the three proteomic replicates (Table 2). Because all ten proteins displayed low expression in both the proteome and toxin transcriptome of both individuals ($< 0.01\%$), these proteins are most likely not toxins. These proteins likely contribute to the general cell maintenance and protein production in the venom gland and could have leaked into the venom prior to venom gland extraction.

Eleven proteins were generically classified as venom proteins (VPs) due to no detectable sequence homology to any known toxins. VPs found in *S. sexspinosus* accounted for 8.8% and 13.1% of the toxin transcriptional output and 2.5% and 5.2% of the proteomic output for C0142 and C0184, respectively (Figure 3). VP-1 and VP-9 both shared sequence homology with uncharacterized proteins from Rehm et al. (2014) in the TSA database, with VP-1 having a 61% sequence identity to *Scolopendra subspinipes dehaani*, and VP-9 having a 42% sequence identity to *Lithobius forficatus*. VP-3, VP-10 and VP-12 all matched to Unchar-06 protein from *S. subspinipes*, and *Scolopendra alternans* (Undheim et al., 2014a; Smith and Undheim, 2018) and shared above an 88% sequence identity to each other. The other VPs (VP-2, VP-4 – VP-8) did not match to any centipede proteins in the TSA database. VP-9 was the only VP to include a conserved domain, a carbonic anhydrase family domain. Three VPs (VP-8 – VP-10) were only detected in C0184 and VP-12 was only detected in the proteome of C0142 (Table 2). VP-12 was not accounted for in the transcriptome of either individual, probably due

to the similarity between VP-3, VP-10, and VP-12 (> 93% similarity). The similarity between these three toxins could have led to some of the reads being misassigned to either VP-3 or VP-10 making it so VP-12 did not receive any reads. Each VP included a signal peptide of 19–23 amino acids.

3.3 Adamalysin-like metalloproteases

Metalloproteases are an integral venom component in many venomous lineages, including centipedes (Malta et al., 2008; Undheim et al., 2014a, 2015). Centipede zinc-dependent metalloproteases described so far are limited to the M12A proteases or astacin-like metalloendoproteases (Undheim et al., 2014a, 2015; Ward and Rokyta, 2018). M12A proteases were identified in *S. sexspinosus* (see above), however, the major class of toxins detected in *S. sexspinosus* belongs to the M12B metalloproteases, also known as the adamalysins. Based on domain structure, the adamalysins found in *S. sexspinosus* venom are structurally similar to snake venom metalloproteases (SVMPs) and, to our knowledge, are the first to be reported in any centipede lineage. Five different adamalysins were detected in both the transcriptome and proteome of each individual *S. sexspinosus* (C0142 and C0184, Table 1). All adamalysins in *S. sexspinosus* contained a signal peptide of 18 or 26 amino acids, 22 or 23 cysteine residues, and had molecular weights of 65.4 kDa–65.7 kDa (Table 1).

SVMPs have been extensively characterized and are organized into three classes (SVMP I , SVMP II, SVMP III) based on the presence of additional nonprotease domains (Fox and Serrano, 2005, 2008). SVMPs I, II and III are classified based on the presence of a metalloprotease domain (I), with the addition of a disintegrin domain (II), and a cysteine rich domain (III) (Figure 4). Four of the adamalysins (MP-1, MP-2, MP-5, MP-8) detected in *S. sexspinosus* do not resemble any of the SVMPs classifications, based on the presence and absence of additional nonprotease domains following the metalloprotease domain (Fox and Serrano, 2008). MP-4 however, does resemble SVMP I as it only consists of a signal peptide and a metalloprotease domain. The remaining adamalysins all contained an additional cysteine-rich domain following the metalloprotease domain (Figure 4). This domain pattern is not seen in any of the SVMP classifications, but has been seen by Ali et al. (2014) in the salivary glands of ticks. Adamalysins constituted 35.7% and 3.0% of the total toxin transcriptome and 44.3% and 31.6% of the total proteome for C0142 and C0184, respectively. The high expression of adamalysins in the venom of *S. sexspinosus* points toward this convergently recruited toxin having an important ecological function. However, the function of these proteins in tick secretions or centipede venoms has yet to be investigated.

The structure and function of SVMPs has been explored extensively, providing a robust framework for describing evolution of these toxins (Seals and Courtneidge, 2003; Takeda, 2016; Kini, 2018). One mechanism describing the neofunctionalization of SVMPs is the loss of domains following duplication of the ancestral gene (Casewell et al., 2011). The loss of domains seen in the SVMP I and SVMP II have only been identified in the

snake family Viperidae. Venoms from this family have an extensive array of different SVMPs, that have evolved through positive selection (Fox and Serrano, 2005; Rokyta et al., 2011). The adamalysins in *S. sexspinosus* venom have also presumably undergone domain loss from the ancestral adamalysin proteins as they do not have the disintegrin domain and one protein does not contain the disintegrin and the cysteine rich domain (MP-4). The loss of these domains and neofunctionalization of the adamalysins could be associated with ability of the toxin to diffuse faster into different tissues because of its smaller size (Doley and Kini, 2009), while retaining the higher activity that has been seen with the extra domains in SVMP III (Takeda, 2016).

3.4 Functional verification of adamalysins

The enzymatic capacity of a highly expressed adamalysin-like metalloprotease observed in the transcriptome and proteome of *S. sexspinosus* venom was measured by gelatin hydrolysis. Gelatin is a denatured form mainly from type I collagen, and this assay is a common method for examining the metalloprotease activity in different venoms (Feitosa et al., 1998; Camacho et al., 2016; Margres et al., 2016). Metalloproteases are the main enzymes in snake venom responsible for extracellular matrix protein degradation, especially collagen and laminin (Baldo et al., 2010; Freitas-de Sousa et al., 2017). *Scolopocryptops sexspinosus* venom was able to hydrolyze gelatin as shown in Figure 5, however the activity was not abolished when incubated with the metalloprotease inhibitor. Venom activity was only quenched with the addition of the serine protease inhibitor. This suggests that the serine proteases have enzymatic function in the venom of *S. sexspinosus* but that the adamalysins are not active against this substrate. However, it has been shown that *in vitro* catalytic activity of an SVMP on synthetic substrate shows reduced activity even when the same toxin was shown to be highly hemorrhagic *in vivo* (Freitas-de Sousa et al., 2017). Thus, we evaluated the ability to induce hemorrhage *in vivo* with *S. sexspinosus* venom. *Scolopocryptops sexspinosus* venom showed hemorrhagic activity alongside a positive control of *C. adamanteus* venom (Figure 6). The venoms were further treated with a metalloprotease inhibitor that abolished all of the activity for *S. sexspinosus* venom and *C. adamanteus* venom (Figure 6), demonstrating that this activity is induced by adamalysins found in the venom of *S. sexspinosus*. The catalytic effects of SVMPs in snake venom lead to the induction of hemorrhage, apoptosis of endothelial cells, and pro-inflammatory action in envenomated predators or prey (Moura-da Silva et al., 2007). Hemorrhage is the main activity induced by SVMPs and is mainly related to the hydrolysis of capillary basement membrane components (principally collagen IV and laminin). These components are involved in capillary stability and cell anchorage (Shannon et al., 1989), reduction of the basement membrane weakens hemodynamic forces and contributes to capillary distension and consequent extravasation (Gutiérrez et al., 2005). To date, no reports have been published about the symptoms caused from *S. sexspinosus* envenomations. Thus, future work should study the effects of these toxins, especially in the context of animal ecology, due to the high

abundance of adamalysins in *S. sexspinosus* venom.

3.5 Transcript and protein abundances across individuals

Venom-gland transcript abundance comparison between the two individuals showed a strong correlation for nontoxin transcripts (Spearman's rank correlation $\rho = 0.88$, Pearson's rank correlation coefficient $R = 0.86$, and $R^2 = 0.74$; Figure 7). The toxin transcripts from each individual were also positively correlated (Spearman's rank correlation $\rho = 0.46$, Pearson's rank correlation coefficient $R = 0.44$, and $R^2 = 0.19$; Figure 7). Seven of the outliers in Figure 7 (β-PFTx-2, LDLA-3, pM12A-1, pM12A-8, ProtCw1a, S1-4, and VP-8) were only present in the proteome of one individual but not the other (Table 2).

Venom proteomic comparison between the two individuals showed strong agreement (Spearman's rank correlation $\rho = 0.83$, Pearson's rank correlation coefficient $R = 0.81$, and $R^2 = 0.65$; Figure 8), with 48 of the 72 toxins detected in both proteomes. The 24 toxin proteins that were proteomically detected in one individual and completely absent from the other are described in Table 2. Ten of these corresponded to proteins with low transcriptomic and proteomic abundances (less than 0.1%) that are likely proteins used for cellular processes unrelated to venom function (discussed below). The remaining 14 proteins were a β-PFTx, a DUF, two LDLAs, three M12As, a SP, two SLPTX, and four VPs. Eight of these proteins showed low expression levels, and six of the toxins (β-PFTx-2, DUF3472-4, LDLA-3, pM12A-1, pM12A-8, and VP-8) showed moderate abundances. Five of the six toxins that were moderately expressed in one individual, and not at all in the other, were also identified as outliers in the transcriptome comparison. The difference between the two individuals could be attributed to intraspecific variation as the high correlation between the nontoxin transcriptomes, the toxin transcriptomes, and the proteomes, suggests that this is a biological rather than technical, variation.

The reversed-phase high-performance liquid chromatography (RP-HPLC) chromatograms help visualize the complexity of the venom described by the proteomes and transcriptomes of *S. sexspinosus* (Figure 9). Most of the peaks appear to be consistent between the two individuals, with some variation seen in the relative absorbance and different number of peaks. The majority of this variation is seen in peaks at ∼20–40 minutes. Additional variation can be seen around 60 minutes, where C0184 has two different peaks not seen in C0142. The profiles reflect some of the proteomic differences discussed above.

3.6 Transcript versus protein abundance estimates

A positive correlation is observed for both C0142 and C0184 when comparing the transcriptomic abundances with the proteomic abundances within each individual (Figure 10). C0142 shows a strong positive correlation (Spearman's rank correlation $\rho = 0.78$, Pearson's rank correlation coefficient $R = 0.79$, and $R^2 = 0.62$). C0184, however, did

not have as strong of a correlation between the transcriptome and proteome abundances (Spearman's rank correlation $\rho = 0.57$, Pearson's rank correlation coefficient $R = 0.48$, and $R^2 = 0.23$). Three individual proteins in C0184 (βPFTX-11, pM12A-3, and S1-5), were detected at low levels in the transcriptome, but were detected at moderate levels in the proteome (Figure 10). These three proteins shared 80–96% similarity with other proteins in their respective families and were indicated as possible outliers in a Cook's distance test of the regression. Removing these three low abundance proteins changes the correlation (Spearman's rank correlation $\rho = 0.65$, Pearson's rank correlation coefficient $R = 0.7$, and $R^2 = 0.50$).

Although several explanations have been hypothesized for differences between transcriptomes and proteomes, differences should be considered technical or methodological unless otherwise ruled out (Rokyta et al., 2015; Ward and Rokyta, 2018). One potential explanation for the differences observed between transcriptomic and proteomic abundances is the timing of mRNA production in the venom gland following venom extraction (Morgenstern et al., 2011; Ward et al., 2018). One study investigated the timing of mRNA and protein production in the venom gland of snakes and noted that mRNA production was highest between three and seven days after venom extraction (Currier et al., 2012). However, significant individual variation was observed in the timing of mRNA production and in the relative abundance of toxin family transcripts. This type of variation could account for the differences observed between the transcriptomes of C0142 and C0184. The two individual *S. sexspinosus* show a strong agreement on the proteomic level, but not on the transcriptomic level (Figures 7,8). This discrepancy could indicate a difference in the timing of mRNA production and the amount of mRNA toxin expression in C0184, even though both individuals had their venom glands removed four days following venom extraction. Venom from a fully regenerated venom gland was comparable between C0142 and C0184, strengthening the argument for variable transcriptomic expression. Figure 7 shows that the three proteins that were considered outliers (βPFTX-11, pM12A-3, and S1-5) contribute to the low correlation of the transcript and protein abundances in C0184, and they are also considered outliers in the transcriptome comparison between C0142 and C0184. The proteins showing a low abundance in the transcriptome of C0184 may have resulted from a biological difference associated with the timing of mRNA production prior to venom gland removal.

4 Conclusions

Through an in-depth venom characterization of *S. sexspinosus* we identified and described 72 complete toxins through linking the protein phenotype with the transcript genotype, representing the first venom characterization of a scolopocryptopid centipede. Toxin families identified included: β-PFTx, CAPs, M12A peptidases, SPs, M13 peptidases, SLPTXs, PLA2s, LDLAs, HYALs, DUFs, VPs and adamalysins. Adamalysin-like metalloproteases have previously been detected in snake venoms alone, and those detected in *S. sexspinosus* displayed a unique domain structure that is unlike those found

in snakes (SVMPs). The domain structure in the adamalysins in *S. sexspinosus* could represent a neofunctionalization of this toxin family. However, functional assays suggest that the adamalysins in *S. sexspinosus* retain some of the same enzymatic activity as the SVMPs. Adamalysins were the most highly expressed protein family identified in the proteome, constituting nearly half of the proteomic abundance for both individual *S. sexspinosus*. The recruitment of adamalysins into the centipede venom identified here represents a striking case of molecular convergent evolution.

Acknowledgments

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1449440. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Funding for this work was provided by the National Science Foundation (NSF DEB-1145978 and NSF DEB 1638902) and the Florida State University Council on Research and Creativity. We thank Rakesh Singh of the Florida State University College of Medicine Translational Science Laboratory for advice and assistance with proteomic analyses, and Margaret Seavy of the Florida State Molecular Core Facility for her guidance on RP-HPLC parameters. We also thank Pierson Hill for his assistance in collecting specimens.

References

Aitchison, J., 1986. The statistical analysis of compositional data. Chapman and Hall, London.

Ali, A., L. Tirloni, M. Isezaki, A. Seixas, S. Konnai, K. Ohashi, I. d. S. V. Junior, and C. Termignoni, 2014. Reprolysin metalloproteases from *Ixodes persulcatus, Rhipicephalus sanguineus* and *Rhipicephalus microplus* ticks. Experimental and Applied Acarology 63:559–578.

Baldo, C., C. Jamora, N. Yamanouye, T. M. Zorn, and A. M. Moura-da Silva, 2010. Mechanisms of vascular damage by hemorrhagic snake venom metalloproteinases: tissue distribution and in situ hydrolysis. Plos Neglected Tropical Diseases 4:e727.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, et al., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19:455–477.

Bonato, L., A. Chagas Jr, G. Edgecombe, J. Lewis, A. Minelli, L. Pereira, R. Shelley, P. Stoev, and M. Zapparoli, 2016. Chilobase 2.0-a world catalogue of centipedes (chilopoda). Available online at: http://chilobase.biologia.unipd.it [Accessed 03/13/2019] .

Camacho, E., L. Sanz, T. Escalante, A. Pérez, F. Villalta, B. Lomonte, A. G. C. Neves-Ferreira, A. Feoli, J. J. Calvete, J. M. Gutiérrez, et al., 2016. Novel catalytically-inactive PII metalloproteinases from a viperid snake venom with substitutions in the canonical zinc-binding motif. Toxins 8:292.

Casewell, N. R., S. C. Wagstaff, R. A. Harrison, C. Renjifo, and W. Wüster, 2011. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. Molecular Biology and Evolution 28:2637–2649.

Casewell, N. R., W. Wüster, F. J. Vonk, R. A. Harrison, and B. G. Fry, 2013. Complex cocktails: the evolutionary novelty of venoms. Trends in Ecology & Evolution 28:219–229.

Currier, R. B., J. J. Calvete, L. Sanz, R. A. Harrison, P. D. Rowley, and S. C. Wagstaff, 2012. Unusual stability of messenger RNA in snake venom reveals gene expression dynamics of venom replenishment. PLOS One 7:e41888.

Dal Peraro, M. and F. G. Van Der Goot, 2016. Pore-forming toxins: ancient, but never really out of fashion. Nature Reviews Microbiology 14:77.

Doley, R. and R. M. Kini, 2009. Protein complexes in snake venom. Cellular and Molecular Life Sciences 66:2851–2871.

Feitosa, L., W. Gremski, S. S. Veiga, M. C. Q. Elias, E. Graner, O. C. Mangili, and R. R. Brentani, 1998. Detection and characterization of metalloproteinases with gelatinolytic, fibronectinolytic and fibrinogenolytic activities in brown spider (*Loxosceles intermedia*) venom. Toxicon 36:1039–1051.

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

———, 2008. Exploring snake venom proteomes: multifaceted analyses for complex toxin mixtures. Proteomics 8:909–920.

Fry, B., K. Roelants, and J. Norman, 2009a. Tentacles of venom: toxic protein convergence in the kingdom animalia. Journal of Molecular Evolution 68:311–321.

Fry, B. G., K. Roelants, D. E. Champagne, H. Scheib, J. D. Tyndall, G. F. King, T. J. Nevalainen, J. A. Norman, R. J. Lewis, R. S. Norton, et al., 2009b. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. Annual Review of Genomics and Human Genetics 10:483–511.

Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch, 2005. Protein identification and analysis tools on the ExPASy server. Springer.

Gibbs, G. M., K. Roelants, and M. K. O'bryan, 2008. The cap superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins-roles in reproduction, cancer, and immune defense. Endocrine Reviews 29:865–897.

Girish, K., R. Shashidharamurthy, S. Nagaraju, T. V. Gowda, and K. Kemparaju, 2004. Isolation and characterization of hyaluronidase a "spreading factor" from Indian cobra (*Naja naja*) venom. Biochimie 86:193–202.

Gonzáles-Morales, L., E. Diego-García, L. Seqovia, M. del Carmen Gutiérrez, and L. D. Possani, 2009. Venom from the centipede *Scolopendra viridis* Say: purification, gene cloning and phylogenetic analysis of a phospholipase A2. Toxicon 54:8–15.

González-Morales, L., M. Pedraza-Escalona, E. Diego-Garcia, R. Restano-Cassulini, C. V. Batista, M. del Carmen Gutiérrez, and L. D. Possani, 2014. Proteomic characterization of the venom and transcriptomic analysis of the venomous gland from the mexican centipede *Scolopendra viridis*. Journal of Proteomics 111:224–237.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev, 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29:644–652.

Gutiérrez, J. M., A. Rucavado, T. Escalante, and C. Díaz, 2005. Hemorhhage induced by snake venom metalloproteinases: biochemical and biophysical mechanisms involved in microvessel damage. Toxicon 8:997–1011.

Haas, B. and A. Papanicolaou, 2016. Transdecoder (find coding regions within transcripts) Available from http://transdecoder.github.io.

Hakim, M. A., S. Yang, and R. Lai, 2015. Centipede venoms and their components: resources for potential therapeutic applications. Toxins 7:4832–4851.

Joshi, J. and K. P. Karanth, 2011. Cretaceous–tertiary diversification among select scolopendrid centipedes of South India. Molecular Phylogenetics and Evolution 60:287–294.

Jungo, F., L. Bougueleret, I. Xenarios, and S. Poux, 2012. The uniprotkb/swiss-prot tox-prot program: a central hub of integrated venom protein data. Toxicon 60:551–557.

Kini, R. M., 2018. Accelerated evolution of toxin genes: Exonization and intronization in snake venom disintegrin/metalloprotease genes. Toxicon 148:16–25.

Knittel, P. S., P. F. Long, L. Brammall, A. C. Marques, M. T. Almeida, G. Padilla, and A. M. Moura-da Silva, 2016. Characterising the enzymatic profile of crude tentacle extracts from the South Atlantic jellyfish *Olindias sambaquiensis* (Cnidaria: Hydrozoa). Toxicon 119:1–7.

Kondo, H., S. KONDO, H. IKEZAWA, R. MURATA, and A. OHSAKA, 1960. Studies on the quantitative method for determination of hemorrhagic activity of Habu snake venom. Japanese Journal of Medical Science and Biology 13:43–51.

Krueger, F., 2015. Trim Galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Buisulfite-Seq) libraries. URL `https://github.com/FelixKrueger/TrimGalore`.

Langmead, B. and S. L. Salzberg, 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357–359.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997 .

Li, M., I. X. Wang, Y. Li, A. Bruzel, A. L. Richards, J. M. Toung, and V. G. Cheung, 2011. Widespread RNA and DNA sequence differences in the human transcriptome. Science 333:53–58.

Li, W. and A. Godzik, 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659.

Liu, J., G. Li, Z. Chang, T. Yu, B. Liu, R. McMullen, P. Chen, and X. Huang, 2016. BinPacker: packing based *de novo* transcriptome assembly from RNA-seq data. PLOS Computational Biology 12:e1004772.

Liu, Z.-C., R. Zhang, F. Zhao, Z.-M. Chen, H.-W. Liu, Y.-J. Wang, P. Jiang, Y. Zhang, Y. Wu, J.-P. Ding, et al., 2012. Venomic and transcriptomic analysis of centipede *Scolopendra subspinipes dehaani*. Journal of Proteome Research 11:6197–6212.

MacManes, M. D., 2014. On the optimal trimming of high-throughput mRNA sequence data. Frontiers in Genetics 5:13.

Malta, M. B., M. S. Lira, S. L. Soares, G. C. Rocha, I. Knysak, R. Martins, S. P. Guizze, M. L. Santoro, and K. C. Barbaro, 2008. Toxic activities of Brazilian centipede venoms. Toxicon 52:255–263.

Marçais, G. and C. Kingsford, 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27:764–770.

Margres, M. J., R. Walls, M. Suntravat, S. Lucena, E. E. Sánchez, and D. R. Rokyta, 2016. Functional characterizations of venom phenotypes in the eastern diamondback rattlesnake (*Crotalus adamanteus*) and evidence for expression-driven divergence in toxic activities among populations. Toxicon 119:28–38.

Moran, Y., D. Fredman, P. Szczesny, M. Grynberg, and U. Technau, 2012. Recurrent horizontal transfer of bacterial toxin genes to eukaryotes. Molecular Biology and Evolution 29:2223–2230.

Morgenstern, D., B. H. Rohde, G. F. King, T. Tal, D. Sher, and E. Zlotkin, 2011. The tale of a resting venom gland: transcriptome of a replete venom gland from the scorpion *Hottentotta judaicus*. Toxicon 57:695–703.

de Oliveira-Mendes, B. B. R., S. E. M. Miranda, D. F. Sales-Medina, B. de Freitas Magalhães, Y. Kalapothakis, R. P. de Souza, V. N. Cardoso, A. L. B. de Barros, C. Guerra-Duarte, E. Kalapothakis, et al., 2019. Inhibition of *Tityus serrulatus* venom hyaluronidase affects venom biodistribution. PLoS neglected tropical diseases 13:e0007048.

Paszkiewicz, K. H., A. Farbos, P. O'Neill, and K. Moore, 2014. Quality control on the frontier. Frontiers in Genetics 5:157.

Petersen, T. N., S. Brunak, G. von Heijne, and H. Nielsen, 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 8:785–786.

Rates, B., M. P. Bemquerer, M. Richardson, M. H. Borges, R. A. Morales, M. E. De Lima, and A. M. Pimenta, 2007. Venomic analyses of *Scolopendra viridicornis nigra* and *Scolopendra angulata* (centipede, Scolopendromorpha): shedding light on venoms from a neglected group. Toxicon 49:810–826.

Rehm, P., K. Meusemann, J. Borner, B. Misof, and T. Burmester, 2014. Phylogenetic position of myriapoda revealed by 454 transcriptome sequencing. Molecular Phylogenetics and Evolution 77:25–33.

Rice, P., I. Longden, and A. Bleasby, 2000. EMBOSS: the European molecular biology open software suite.

Rokyta, D. R., A. R. Lemmon, M. J. Margres, and K. Aronow, 2012. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genomics 13:312.

Rokyta, D. R., M. J. Margres, and K. Calvin, 2015. Post-transcriptional mechanisms contribute little to phenotypic variation in snake venoms. Genes|Genomes|Genetics 5:2375–2382.

Rokyta, D. R. and M. J. Ward, 2017. Venom-gland transcriptomics and venom proteomics of the blackback scorpion (*Hadrurus spadix*) reveal detectability challenges and an unexplored realm of animal toxin diversity. Toxicon 128:23–37.

Rokyta, D. R., K. P. Wray, A. R. Lemmon, E. M. Lemmon, and S. B. Caudle, 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.

Rong, M., S. Yang, B. Wen, G. Mo, D. Kang, J. Liu, Z. Lin, W. Jiang, B. Li, C. Du, et al., 2015. Peptidomics combined with cdna library unravel the diversity of centipede venom. Journal of Proteomics 114:28–37.

Seals, D. F. and S. A. Courtneidge, 2003. The ADAMs family of metalloproteases: multidomain proteins with multiple functions. Genes & Development 17:7–30.

Shannon, J. D., E. N. Baramova, J. Bjarnason, and J. Fox, 1989. Amino acid sequence of a *Crotalus atrox* venom metalloproteinase which cleaves type IV collagen and gelatin. Journal of Biological Chemistry 264:11575–11583.

Shelley, R. M., 2002. A synopsis of the North American centipedes of the order Scolopendromorpha (Chilopoda). Virginia Museum of Natural History.

Sher, D., Y. Fishman, M. Zhang, M. Lebendiker, A. Gaathon, J.-M. Mancheño, and E. Zlotkin, 2005. Hydralysins, a new category of β-pore-forming toxins in cnidaria. Journal of Biological Chemistry 280:22847–22855.

Moura-da Silva, A., D. Butera, and I. Tanjoni, 2007. Importance of snake venom metalloproteinases in cell biology: effects on platelets, inflammatory and endothelial cells. Current Pharmaceutical Design 13:2893–2905.

Smith, J. J. and E. A. Undheim, 2018. True lies: Using proteomics to assess the accuracy of transcriptome-based venomics in centipedes uncovers false positives and reveals startling intraspecific variation in *Scolopendra subspinipes*. Toxins 10:96.

Freitas-de Sousa, L. A., M. Colombini, M. Lopes-Ferreira, S. M. Serrano, and A. M. Moura-da Silva, 2017. Insights into the mechanisms involved in strong hemorrhage and dermonecrosis induced by atroxlysin-Ia, a PI-class snake venom metalloproteinase. Toxins 9:239.

Spivak, M., J. Weston, L. Bottou, L. Käll, and W. S. Noble, 2009. Improvements to the percolator algorithm for peptide identification from shotgun proteomics data sets. Journal of Proteome Research 8:3737–3745.

Takeda, S., 2016. ADAM and ADAMTS family proteins and snake venom metalloproteinases: A structural overview. Toxins 8:155.

Undheim, E. A., A. Jones, K. R. Clauser, J. W. Holland, S. S. Pineda, G. F. King, and B. G. Fry, 2014a. Clawing through evolution: toxin diversification and convergence in the ancient lineage Chilopoda (Centipedes). Molecular Biology and Evolution Pp. 2124–2148.

Undheim, E. A., K. Sunagar, B. R. Hamilton, A. Jones, D. J. Venter, B. G. Fry, and G. F. King, 2014b. Multifunctional warheads: Diversification of the toxin arsenal of centipedes via novel multidomain transcripts. Journal of Proteomics 102:1–10.

Undheim, E. A. B., B. G. Fry, and G. F. King, 2015. Centipede venom: recent discoveries and current state of knowledge. Toxins 7:679–704.

Vizcaíno, J. A., A. Csordas, N. del Toro N, J. A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol, F. Reisinger, T. Ternent, Q. W. Xu, R. Wang, and H. Hermjakob, 2016. 2016 update of the PRIDE database and related tools. Nucleic Acids Research 44:D447–D456.

Ward, M. J., S. A. Ellsworth, M. P. Hogan, G. S. Nystrom, P. Martinez, A. Budheo, R. Zelaya, A. Perez, B. Powell, H. He, et al., 2018. Female-biased population divergence in the venom of the hentz striped scorpion (*Centruroides hentzi*). Toxicon 152:137–149.

Ward, M. J., S. A. Ellsworth, and D. R. Rokyta, 2017. Venom-gland transcriptomics and venom proteomics of the Hentz striped scorpion (*Centruroides hentzi*; Buthidae) reveal high toxin diversity in a harmless member of a lethal family. Toxicon 142:14–29.

Ward, M. J. and D. R. Rokyta, 2018. Venom-gland transcriptomics and venom proteomics of the giant florida blue centipede, *Scolopendra viridis*. Toxicon 152:121–136.

Xie, Y., G. Wu, J. Tang, R. Luo, J. Patterson, S. Liu, W. Huang, G. He, S. Gu, S. Li, X. Zhou, T.-W. Lam, Y. Li, X. Xu, G. K.-S. Wong, and J. Wang, 2014. SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. Bioinformatics 30:1660–1666.

Yang, S., Z. Liu, Y. Xiao, Y. Li, M. Rong, S. Liang, Z. Zhang, H. Yu, G. F. King, and R. Lai, 2012. Chemical punch packed in venoms makes centipedes excellent predators. Molecular & Cellular Proteomics 11:640–650.

Zhang, J., K. Kobert, T. Flouri, and A. Stamatakis, 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614–620.

Figure Legends

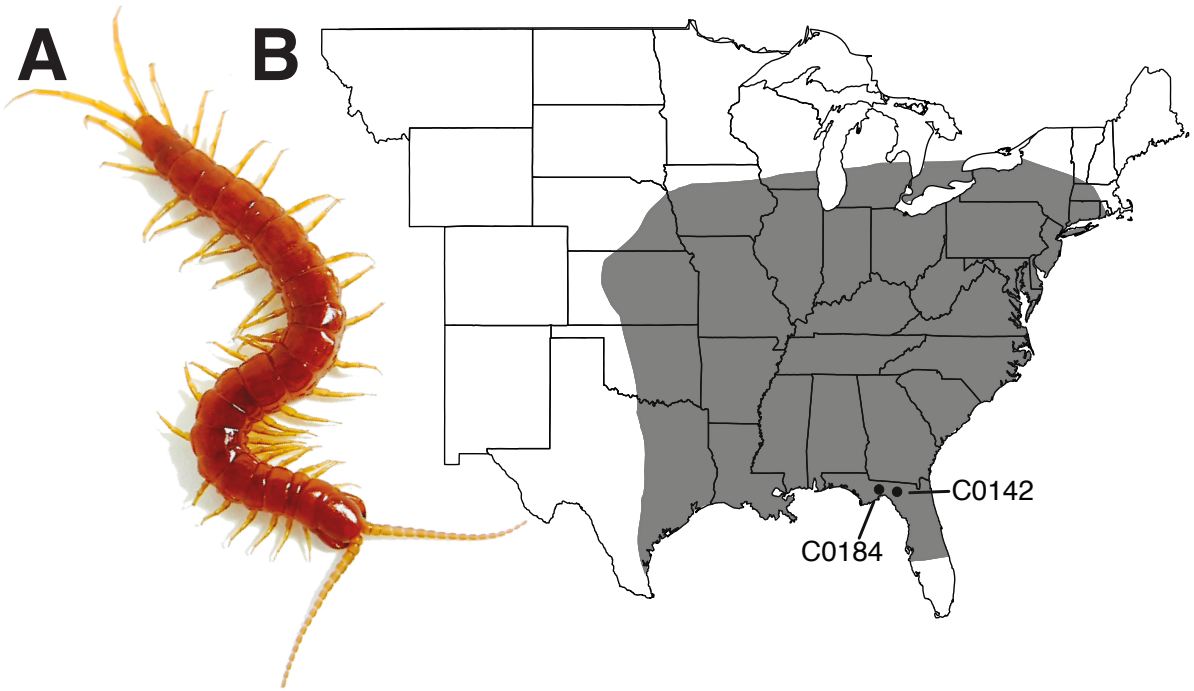

Figure 1. A) Dorsal view of a *S. sexspinosus* from Northern Florida. *Scolopocryptops sexspinosus* can reach a maximum of 69 mm in length and 4 mm in width. B) Range map of *S. sexspinosus*. Points indicate where individuals C0142 and C0184 were collected.

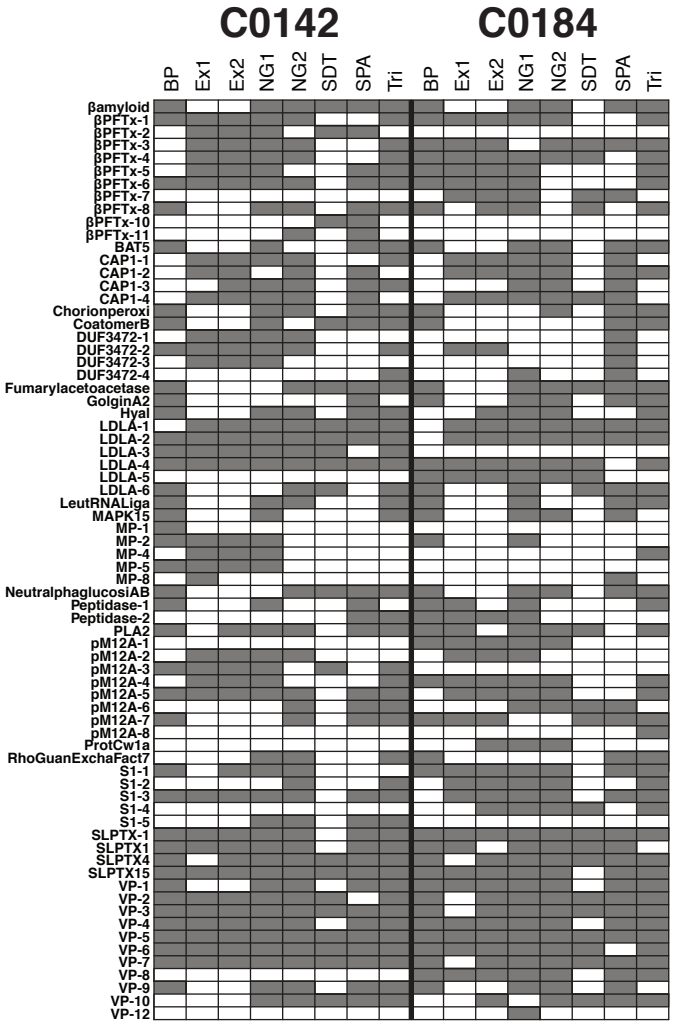

Figure 2. A heat map is shown displaying which assembly method assembled each toxin. Contigs from each assembly that showed a full length sequence and shared a 99% sequence identity to a toxin from the final consensus transcriptome are displayed as a dark rectangle. Abbreviations: BP—BinPacker, Ex—Extender, NG—NGen, SDT—SOAPdenovotrans, SPA—rnaSPAdes, Tri—trinity, BAT5—protein BAT5, BPFTx—β-pore forming toxin, CAP—cysteine-rich secretory protein, antigen 5, and pathogenesis-related 1 protein domains, Chorionperoxi—chorion peroxidase, CoatomerB—Coatomer subunit β, DUF—domain with unknown function, GolginA2—golgin subfamily A member 2, HYAL—hyaluronidase, LDLA—low-density lipoprotein receptor Class A repeat domain, LeutRNALiga—leucine-tRNA ligase, MAP15—mitogen-activated protein kinase kinase kinase 15, MP—adamalysins (M12B metalloproteases), NeutralphaglucosiAB—neutral α-glucosidase AB, Peptidase-1—S8 serine protease, Peptidase-2—M13 peptidase, PLA2—phospholipase A_2, pM12A—peptidase M12a, ProtCw1a—scoloptoxin, RhoGuanExchafact7—rho guanine nucleotide exchange factor 7, S1—S1 serine protease, SLPTX—scoloptoxins, VP—venom protein.

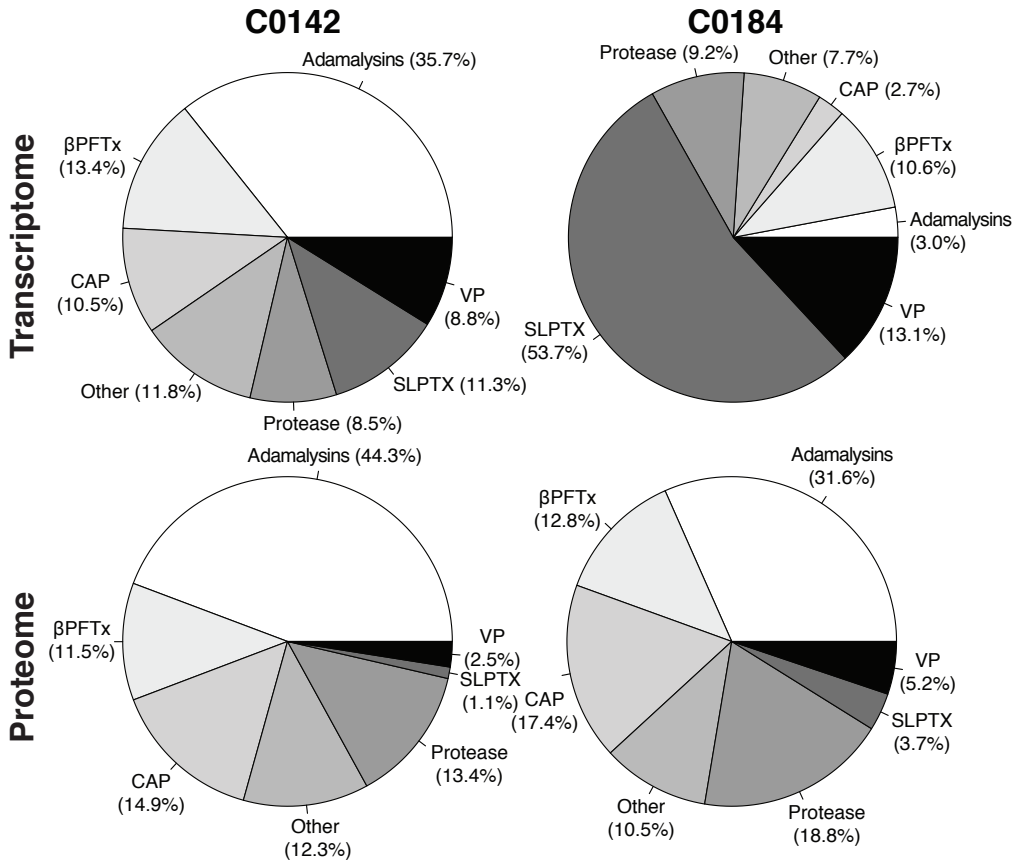

Figure 3. Class level abundance comparisons for the venom-gland transcriptome and venom proteome of C0142 were consistent. The venom proteome of C0142 and C0184 also showed a similar pattern of expression. However, class level abundances for the transcriptome and proteome of C0184 are not consistent. The expression of SLPTXs is 53.7% in the transcriptome, yet only accounts for 3.7% in the proteome. The expression of adamalysins is not very consistent accounting for 3.0% of the transcriptome and changing to 31.6% of the proteome. Abbreviations: Bamyloid—β-amyloid, BAT5—protein BAT5, BPFTx—β-pore forming toxin, CAP—cysteine-rich secretory protein, antigen 5, and pathogenesis-related 1 protein domains, Chorionperoxi—chorion peroxidase, CoatomerB—Coatomer subunit β, DUF—domain with unknown function, GolginA2—golgin subfamily A member 2, HYAL—hyaluronidase, LDLA—low-density lipoprotein receptor Class A repeat domain, LeutRNALiga—leucine-tRNA ligase, MAP15—mitogen-activated protein kinase kinase kinase 15, NeutralphaglucosiAB—neutral α-glucosidase AB, RhoGuanExchafact7—rho guanine nucleotide exchange factor 7, SLPTX—scoloptoxins, SP—serine protease VP—venom protein.

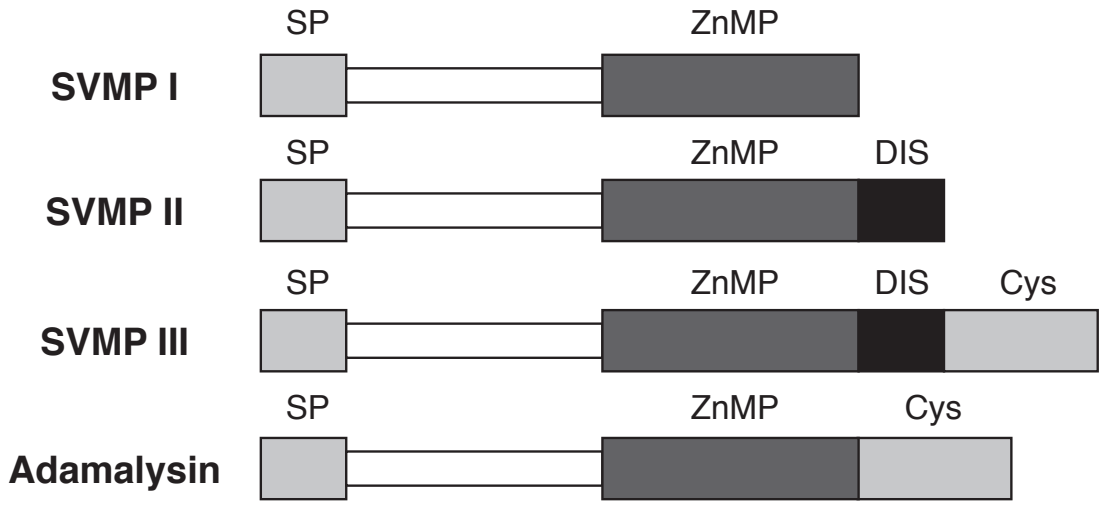

Figure 4. A schematic representation of adamalysin-like metalloproteases. The top three structures represent the snake venom metalloproteases (SVMPs). The bottom structure represents the different domain structure of four of the adamalysin-like proteins in *S. sexspinosus*. One adamalysin in *S. sexspinosus*, MP-4, has a domain structure similar to SVMP I. Abbreviations: SP—signal peptide, ZnMP—Zinc-dependent metalloprotease domain, DIS—disintegrin domain, Cys—ADAM Cysteine-Rich Domain.

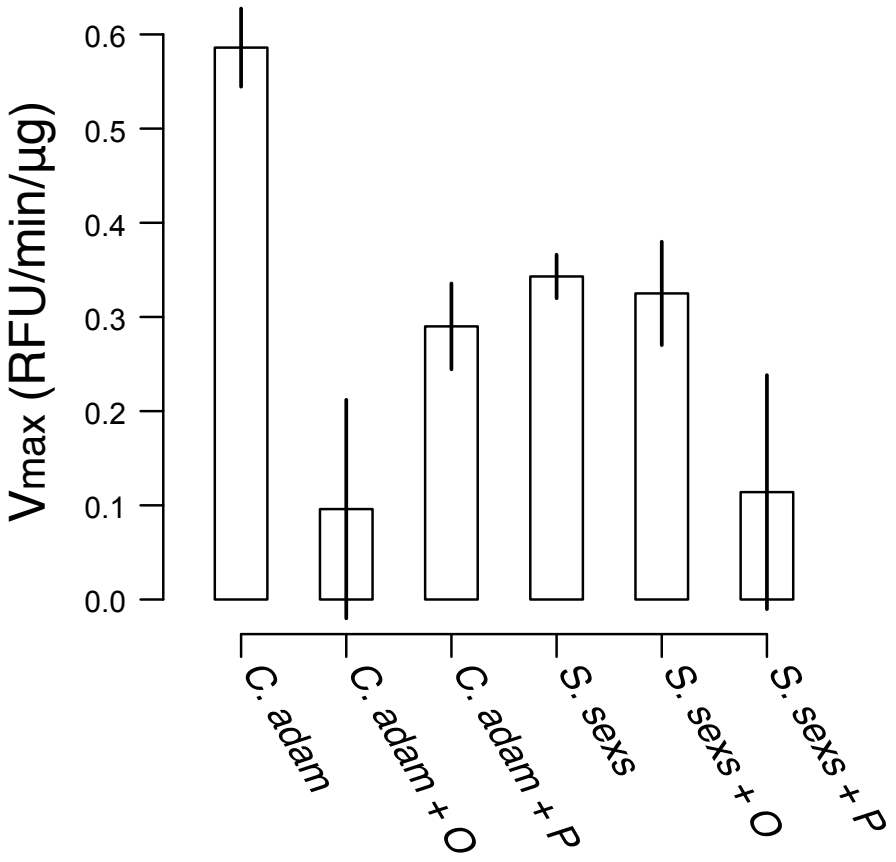

Figure 5. Enzymatic properties of *S. sexspinosus* venom compared against venom from *C. adamanteus* (positive control). Enzymatic activity from each venom was measured through the amount of degraded gelatin fluorescein substrate every minute over a period of three hours. Venom samples were incubated with and without two inhibitors, metalloprotease inhibitors (O) and serine protease inhibitors (P). The maximum rate of the reaction (V_{max}) was calculated for each sample in relative fluorescence units per minute. These values were then corrected for the amount of venom (μg) utilized. The average of three samples per venom are shown along with the 95% confidence intervals. Our positive control (*C. adamanteus*) displayed a high rate of reaction for the gelatin substrate that was quenched completely when inhibited by metalloprotease inhibitors and lowered slightly with serine protease inhibitors. *Scolopocryptops sexspinosus* also showed a high rate of reaction however, metalloprotease inhibitors did not change the activity but serine protease inhibitors quenched the activity of this venom. Abbreviations: *C. adam—C. adamanteus venom*, *S. sexs—S. sexspinosus venom*

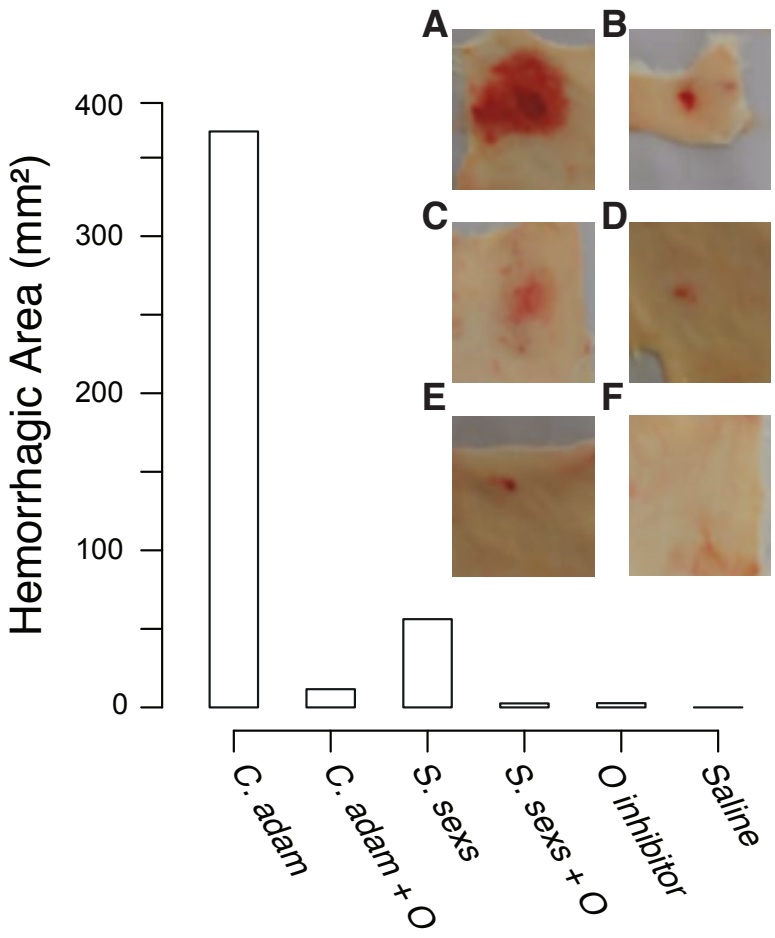

Figure 6. Inhibition of hemorrhagic activity induced by venom from *Scolopocryptops sexspinosus* (*S. sexs*) and *Crotalus adamanteus* (*C. adam*). 30 μg of *S. sexspinosus* venom and 10 μg of *C. adamanteus* venom were incubated for 30 minutes with 50 μM of 1,10-Phenantroline (O) at 37 °C and injected by intradermal route into the dorsal skin of mice. After one hour, the animals were euthanized and the dorsal skin was removed. Results are expressed as a mean of replicated experiments. Insert shows the hemorrhagic area of A) *C. adamanteus* venom, B) *C. adamanteus* venom and inhibitor, C) *S. sexspinosus* venom, D) *S. sexspinosus* venom and inhibitor, E) injection with inhibitor, and F) injection with saline.

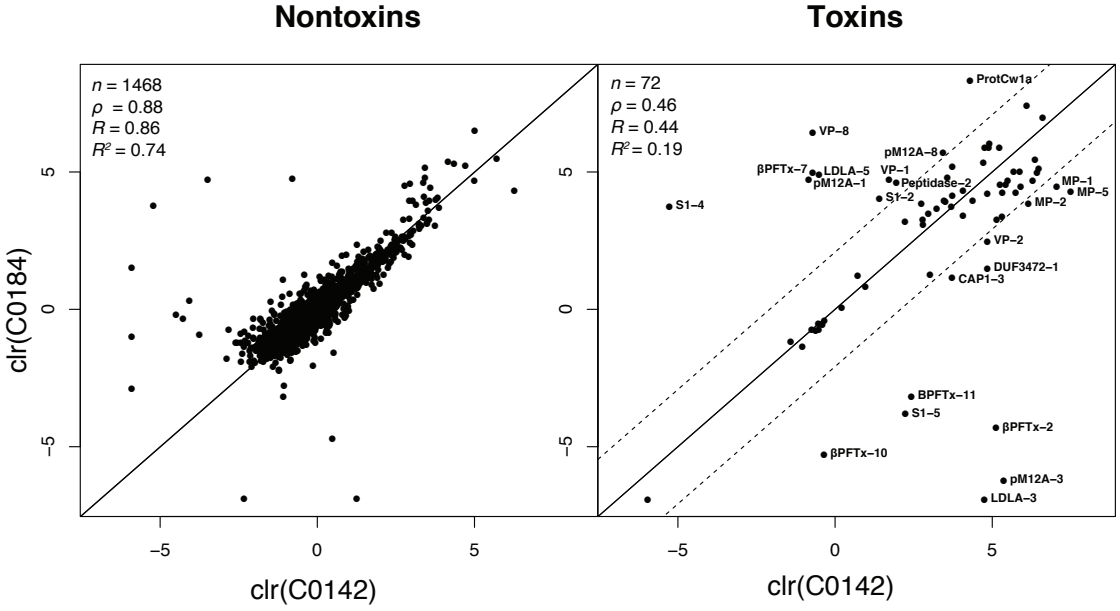

Figure 7. Venom-gland transcript abundance comparison between C0142 and C0184 for the nontoxins was highly correlated, while the transcript comparison for the toxins between the two individuals was less correlated. The dashed lines in the toxins plot represent the 99th percentile of differences between the two nontoxin measures. Labeled points outside the dashed line represent toxins with unusually different expression levels relative to the nontoxins and were considered outliers. Abbreviations: clr—centered logratio transformation, n—number of transcripts, ρ—Spearman's rank correlation coefficient, R—Pearson's correlation coefficient, R^2—coefficient of determination, BPFTx—β-pore forming toxin, CAP—cysteine-rich secretory protein, antigen 5, and pathogenesis-related 1 protein domains, DUF—domain with unknown function, HYAL—hyaluronidase, LDLA—low-density lipoprotein receptor Class A repeat domain, MP—adamalysins (M12B metalloproteases), Peptidase-2—M13 peptidase, pM12A—peptidase M12A, ProtCw1a—scoloptoxin, SLPTX—scoloptoxins, S1—S1 serine protease, VP—venom protein.

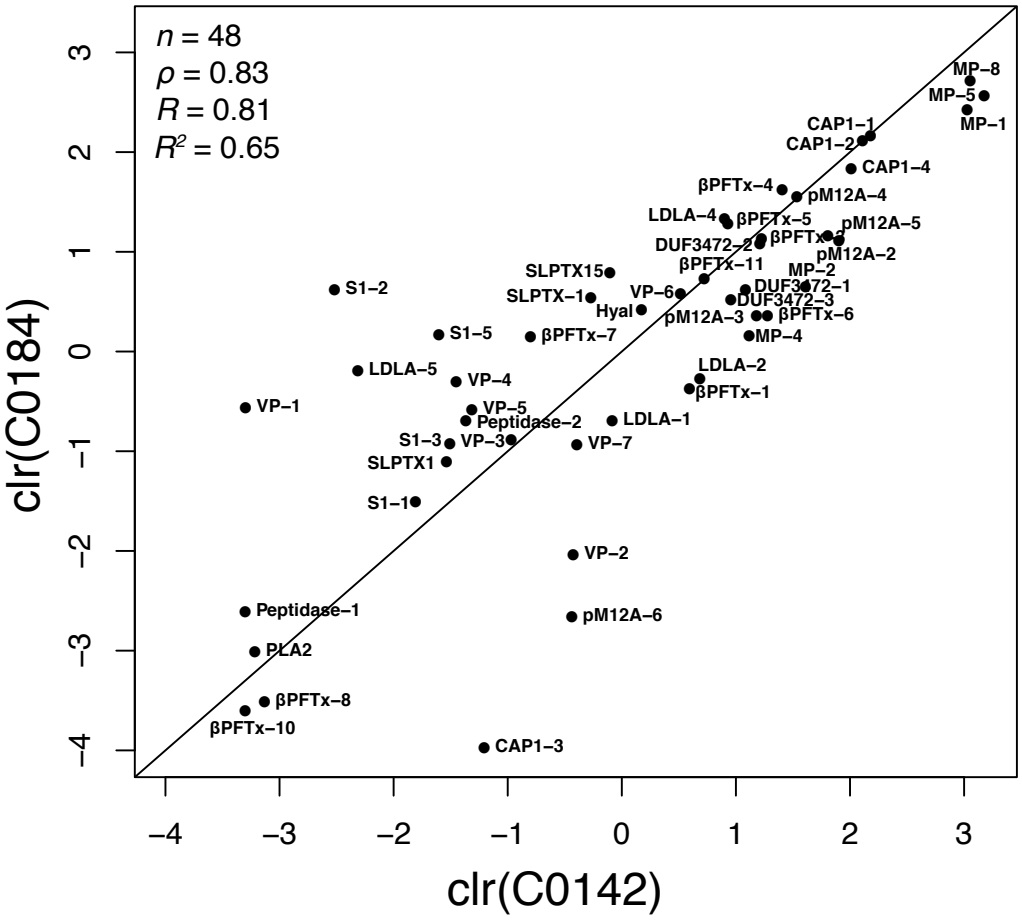

Figure 8. A venom proteomic comparison between the two individuals of *S. sexspinosus* (C0142 and C0184) showed good agreement for proteins detected in both venom proteomes. Table 2 shows the proteomic presence/absence differences between the two individuals, with only eight toxins (β-PFTX-2, DUF3472-4, LDLA-3, pM12A-1, pM12A-8 and VP-8) showing a moderate abundance difference between the two individuals. Abbreviations: clr—centered logratio transformation, n—number of proteins, ρ—Spearman's rank correlation coefficient, R—Pearson's correlation coefficient, R^2—coefficient of determination, BPFTx—β-pore forming toxin, CAP—cysteine-rich secretory protein, antigen 5, and pathogenesis-related 1 protein domains, DUF—domain with unknown function, Hyal—hyaluronidase, LDLA—low-density lipoprotein receptor Class A repeat domain, MP—adamalysins (M12b metalloproteases) pM12A—peptidase M12a, Peptidase-1—S8 serine protease, Peptidase-2—M13 peptidase, PLA2—phospholipase A_2, SLPTX—scoloptoxins, S1—S1 serine protease, VP—venom protein.

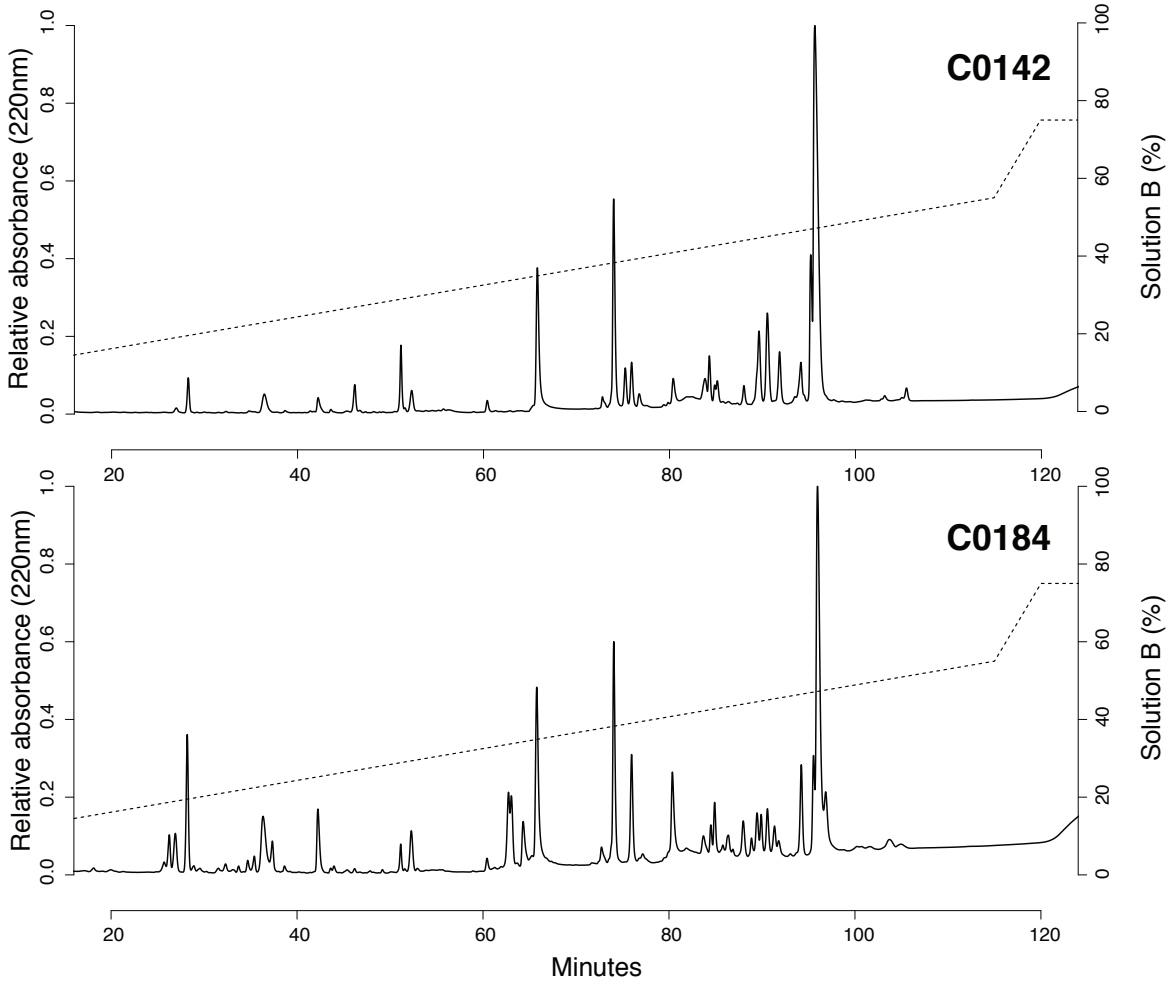

Figure 9. A reversed-phase high-performance liquid chromatography spectrum is shown for a venom sample from each individual *S. sexspinosus*, demonstrating the complexity of the venom. The dashed line indicates the elution gradient used, shown as a percentage of solution B.

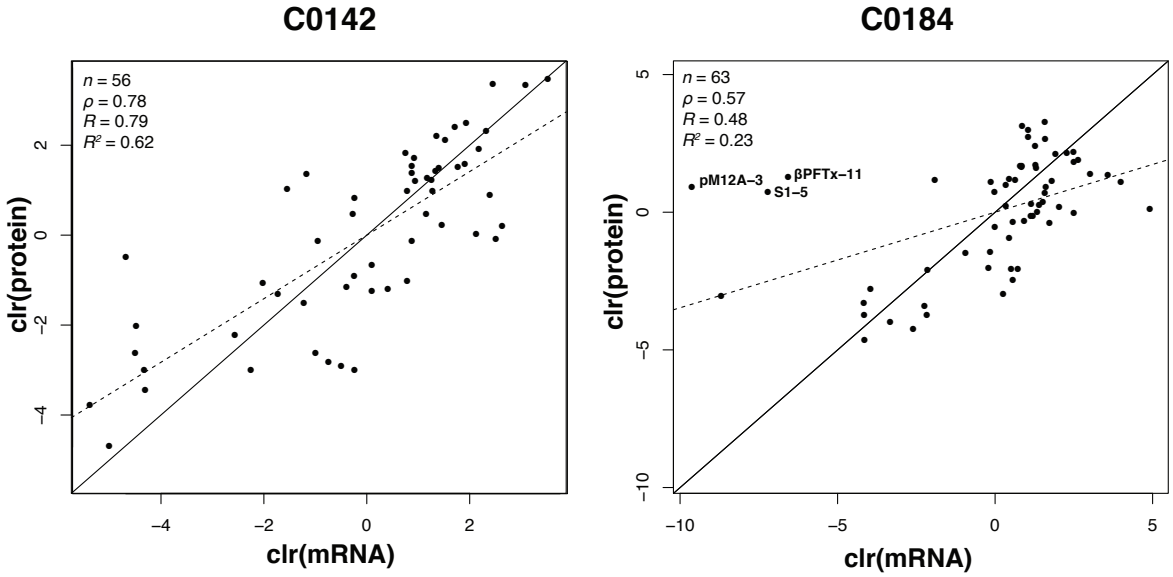

Figure 10. Transcript and protein abundances were positively correlated, but the agreement in individual C0184 was weaker compared to C0142. The weak correlation in C0184 could be attributed to the timing of mRNA expression in the venom gland following venom extraction. This timing bias would influence the transcriptome of one individual yet leave the proteomic evidence consistent among both individuals leading to a low correlation in C0184 and a higher correlation for C0142. Proteins identified as potential outliers using a Cook's distance test of regression are labeled. The line of best fit for each individual is indicated by a dashed line. Abbreviations: clr—centered logratio transformation, n—number of transcripts, ρ—Spearman's rank correlation coefficient, R—Pearson's correlation coefficient, R^2—coefficient of determination, BPFTx—β-pore forming toxin, pM12A—peptidase M12a, S1—S1 serine protease.

Tables

Table 1. Toxins identified and proteomically confirmed in the venom of *Scolopocryptops sexspinosus*.

Toxin	Signal peptide (aa)	Precursor (aa)	Cysteine Residues	MW (kDa)	C0142 TPM	C0184 TPM	C0142 fmol	C0184 fmol
β-PFTx-1	20	326	5	33.9	35,818.07	12,699.72	1,053.17	374.02
β-PFTx-2	22	344	5	36.5	10,396.53	0.74	687.49	–
β-PFTx-3	17	330	5	35.5	7,871.28	3,695.34	1,978.73	1,687.44
β-PFTx-4	17	333	5	35.7	8,233.93	20,163.89	2,369.94	2,767.86
β-PFTx-5	22	344	5	36.6	11,550.57	20,226.04	1,472.66	1,963.53
β-PFTx-6	27	345	5	35.8	22,119.27	8,371.62	2,083.44	775.53
β-PFTx-7	22	344	5	36.4	30.27	8,110.92	260.93	628.85
β-PFTx-8	21	332	3	34.7	1,564.21	2,144.22	25.37	16.27
β-PFTx-10	21	332	3	34.6	43.40	0.28	21.41	14.77
β-PFTx-11	17	333	5	35.7	699.84	2.33	1,197.24	1,134.78
β-amyloid	22	655	15	73.5	76.89	59.39	–	5.83
BAT5	–	546	7	61.2	29.46	26.49	–	3.04
CAP1-1	22	231	9	22.9	22,645.16	4,773.01	5,146.91	4,732.68
CAP1-2	18	231	10	23.4	18,215.00	8,198.09	4,796.01	4,520.54
CAP1-3	18	231	10	23.8	2,565.31	176.76	173.83	10.21
CAP1-4	18	231	10	23.4	33,334.74	5,940.57	4,339.56	3,399.41
Chorionperoxi	19	801	20	87.1	36.34	32.36	31.23	–
CoatomerB	–	952	18	106.9	36.64	26.15	–	7.38
DUF3472-1	26	428	1	44.4	7,891.46	245.79	1,719.75	1,006.78
DUF3472-2	25	427	2	44.3	19,324.82	3,850.90	1,948.00	1,600.87
DUF3472-3	25	426	1	44.3	10,589.39	1,460.71	1,512.70	917.74
DUF3472-4	25	427	1	44.4	1,020.42	1,197.04	1,661.44	–
Fumarylacetoacetase	–	420	7	46.9	161.79	124.43	–	4.51
GolginA2	–	826	4	94.9	33.43	25.81	–	11.56
Hyal	21	371	5	40.5	2,497.87	2,342.45	690.04	825.13
LDLA-1	17	204	8	21.7	14,111.43	5,188.64	534.10	272.82
LDLA-2	17	203	8	21.8	11,775.20	5,252.18	1,151.71	412.47
LDLA-3	18	195	8	20.5	7,188.71	–	1,142.44	–
LDLA-4	18	192	8	19.9	8,418.64	23,244.45	1,430.42	2,055.78
LDLA-5	18	194	8	20.2	37.08	7,586.01	57.42	449.59
LDLA-6	18	192	8	20.4	1,001.80	1,427.36	–	72.76
LeutRNALiga	–	1,172	19	134.7	41.56	31.85	–	19.03
MAPK15	–	1,309	24	148.9	15.06	16.95	9.72	–
MP-1	26	611	23	65.6	71,639.70	4,782.90	11,996.48	6,124.06
MP-2	18	609	22	65.7	28,919.34	2,605.83	2,906.19	1,038.48
MP-4	18	607	22	65.7	12,518.85	1,639.04	1,776.18	638.30
MP-5	26	611	23	65.4	111,038.32	3,956.78	13,931.60	7,043.95
MP-8	26	611	23	65.4	38,046.66	8,079.08	12,303.69	8,223.20
NeutralphaglucosiAB	18	915	5	103.9	44.26	36.66	13.67	–
Peptidase-1	27	698	9	74.7	2,589.61	3,448.28	21.40	40.05
Peptidase-2	19	730	10	81.9	434.65	5,511.94	148.20	271.61
PLA2	21	161	10	16.2	1,991.97	2,921.10	23.32	26.79
pM12A-1	19	418	14	45.1	26.65	6,133.77	–	1,554.26
pM12A-2	22	426	14	46.3	12,720.19	3,913.00	3,895.17	1,656.50
pM12A-3	19	413	14	44.3	13,288.45	0.11	1,893.20	777.09
pM12A-4	19	416	14	44.5	6,967.39	11,363.99	2,691.29	2570.72
pM12A-5	22	426	14	46.0	15,056.29	6,020.37	3,540.04	1,744.35
pM12A-6	20	421	14	45.7	1,270.27	195.62	375.36	38.26
pM12A-7	22	428	12	46.3	1,209.91	1,816.79	31.14	–
pM12A-8	19	419	14	44.9	1,914.33	16,321.30	–	2,642.45
ProtCw1a	19	219	6	24.1	4,527.07	227,991.32	–	348.76
RhoGuanExchaFact7	–	600	11	68.4	21.90	14.12	3.97	–
S1-1	18	267	10	28.4	966.74	2,604.80	95.41	120.34
S1-2	20	274	11	28.5	252.69	3,151.41	46.89	1,010.16
S1-3	18	274	10	29.0	4,934.69	2,921.94	128.86	215.49

S1-4	18	274	10	28.4	0.32	2,350.09	–	388.49
S1-5	20	270	10	28.2	581.28	1.23	116.89	647.03
SLPTX-1	23	78	4	6.3	27,486.89	90,778.45	442.48	935.21
SLPTX1	23	131	6	12.0	3,616.27	1,659.83	124.98	179.55
SLPTX4	32	70	4	4.0	2,079.38	2,778.93	–	39.98
SLPTX15	22	75	4	5.8	45,514.80	59,915.29	522.93	1,193.78
VP-1	23	82	8	6.5	345.41	6,318.63	21.49	309.74
VP-2	23	91	4	7.8	7,868.70	651.87	379.30	71.16
VP-3	20	180	4	17.6	3,621.79	4,217.22	220.48	223.98
VP-4	20	101	6	9.4	2,208.32	6,727.25	136.27	401.62
VP-5	24	99	10	8.5	7,212.75	20,285.28	155.97	304.09
VP-6	22	106	10	9.5	2,582.21	10,067.75	971.73	968.32
VP-7	20	88	4	7.5	40,342.36	9,358.68	391.88	213.73
VP-8	23	108	6	9.7	30.45	34,121.38	–	1,234.27
VP-9	19	299	6	30.9	127.77	191.79	–	7.38
VP-10	21	181	4	17.5	571.64	1,348.05	–	41.52
VP-12	21	181	4	17.5	–	–	136.60	–

Cysteine residues and molecular weights were determined using ExPASy ProtParam (Gasteiger et al., 2005) and do not include signal peptides. Abbreviations: BAT5—protein BAT5, BPFTx—β-pore forming toxin, CAP—cysteine-rich secretory protein, antigen 5, and pathogenesis-related 1 protein domains, Chorionperoxi—chorion peroxidase, CoatomerB—Coatomer subunit β, DUF—domain with unknown function, GolginA2—golgin subfamily A member 2, HYAL—hyaluronidase, LDLA—low-density lipoprotein receptor Class A repeat domain, LeutRNALiga—leucine-tRNA ligase, MAP15—mitogen-activated protein kinase kinase kinase 15, MP—adamalysins (M12b metalloproteases), NeutralphaglucosiAB—neutral α-glucosidase AB, Peptidase-1—S8 serine protease, Peptidase-2—M13 peptidase, PLA2—phospholipase A_2, pM12A—peptidase M12a, ProtCw1a—scoloptoxin, RhoGuanExchafact7—rho guanine nucleotide exchange factor 7, S1—S1 serine protease, SLPTX—scoloptoxins, VP—venom protein.

Table 2. Presence/absence differences in the two venom proteomes.

	C0142			C0184			Average	
Protein	rep 1	rep 2	rep 3	rep 1	rep 2	rep 3	C0142	C0184
β-amyloid	–	–	–	12.94	4.56	–	–	5.83
BAT5	–	–	–	–	9.12	–	–	3.04
β-PFTx-2	681.30	659.39	721.79	–	–	–	687.49	–
Chorionperoxi	29.61	35.64	28.44	–	–	–	31.23	–
CoatomerB	–	–	–	8.61	4.56	8.98	–	7.38
DUF3472-4	1,724.03	1,663.26	1,597.03	–	–	–	1,661.44	–
Fumarylacetoacetase	–	–	–	–	4.56	8.98	–	4.51
GolginA2	–	–	–	30.16	–	4.51	–	11.56
LDLA-3	1,131.54	1,176.14	1,119.64	–	–	–	1,142.44	–
LDLA-6	–	–	–	68.93	72.97	76.37	–	72.76
LeutRNALiga	–	–	–	34.49	9.12	13.49	–	19.03
MAPK15	–	17.79	11.35	–	–	–	9.72	–
NeutralphaglucosiAB	23.70	5.95	11.35	–	–	–	13.67	–
pM12A-1	–	–	–	1,490.92	1,537.03	1,634.83	–	1,554.26
pM12A-7	29.61	29.69	34.12	–	–	–	31.14	–
pM12A-8	–	–	–	2,546.63	2,658.98	2,721.75	–	2,642.45
ProtCw1a	–	–	–	344.72	355.75	345.81	–	348.76
RhoGuanExchaFact7	–	11.90	–	–	–	–	3.97	–
S1-4	–	–	–	392.10	351.19	422.18	–	388.49
SLPTX4	–	–	–	43.10	31.93	44.93	–	39.98
VP-8	–	–	–	1,249.61	1,240.57	1,212.64	–	1,234.27
VP-9	–	–	–	8.61	4.56	8.98	–	7.38
VP-10	–	–	–	43.10	41.05	40.42	–	41.52
VP-12	136.27	148.53	125.01	–	–	–	136.60	–

Quantities are given in fmol. Abbreviations: BAT5—protein BAT5, BPFTx—β-pore forming toxin, Chorionperoxi—chorion peroxidase, CoatomerB—Coatomer subunit β, DUF—domain with unknown function, GolginA2—golgin subfamily A member 2, LDLA—low-density lipoprotein receptor Class A repeat domain, LeutRNALiga—leucine-tRNA ligase, MAP15—mitogen-activated protein kinase kinase kinase 15, NeutralphaglucosiAB—neutral α-glucosidase AB, pM12A—peptidase M12a, ProtCw1a—scoloptoxin, RhoGuanExchafact7—rho guanine nucleotide exchange factor 7, S1—S1 serine protease, SLPTX—scoloptoxins, VP—venom protein.