Toxicon 152 (2018) 121–136

Contents lists available at ScienceDirect

Toxicon

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

Venom-gland transcriptomics and venom proteomics of the giant Florida blue centipede, *Scolopendra viridis*

Micaiah J. Ward, Darin R. Rokyta*

Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA

ARTICLE INFO

Keywords:
Centipede
Venom
Transcriptome
Proteome

ABSTRACT

The limited number of centipede venom characterizations have revealed a rich diversity of toxins, and recent work has suggested centipede toxins may be more rapidly diversifying than previously considered. Additionally, many identified challenges in venomics research, including assembly and annotation methods, toxin quantification, and the ability to provide biological or technical replicates, have yet to be addressed in centipede venom characterizations. We performed high-throughput, quantifiable transcriptomic and proteomic methods on two individual *Scolopendra viridis* centipedes from North Florida. We identified 39 toxins that were proteomically confirmed, and 481 nontoxins that were expressed in the venom gland of *S. viridis*. The most abundant toxins expressed in the venom of *S. viridis* belonged to calcium and potassium ion-channel toxins, venom allergens, metalloproteases, and *β*-pore forming toxins. We compared our results to the previously characterized *S. viridis* from Morelos, Mexico, and found only five proteomically confirmed toxins in common to both localities, suggesting either extreme toxin divergence within *S. viridis*, or that these populations may represent entirely different species. By using multiple assembly and annotation methods, we generated a comprehensive and quantitative reference transcriptome and proteome of a Scolopendromorpha centipede species, while overcoming some of the challenges present in venomics research.

1. Introduction

With a fossil record spanning 420 million years, centipedes are currently represented by approximately 3500 species within five extant orders: Scutigeromorpha, Lithobiomorpha, Craterostigmorpha, Scolopendromorpha, and Geophilomorpha (Edgecombe and Giribet, 2007; Fernández et al., 2014; Undheim et al., 2015b). The most recognizable species belong to the order Scolopendromorpha (Undheim et al., 2015b; Hakim et al., 2015), and the venom of *Scolopendra subspinipes* (formerly *S. subspinipes mutilans*, Smith and Undheim, 2018) is the most widely studied (Yang et al., 2012, 2013; Rong et al., 2015; Luo et al., 2018; Smith and Undheim, 2018; Zhao et al., 2018). Venom characterizations have also been completed for *S. dehaani* (formerly *S. subspinipes dehaani,* Liu et al., 2012) and *S. viridis* (from Morelos, Mexico, González-Morales et al., 2014), as well as a comprehensive five-species comparison by Undheim et al. (2014), which included *Ethmostigmus rubripes*, *Cormocephalus westwoodii*, *S. alternans*, *S. morsitans*, and the scutigeromorph, *Thereuopoda longicornis*. These characterizations of centipede venoms have revealed a rich diversity of toxins, including ion-channel toxins, proteases, *β*-pore forming toxins, CAPs (cysteine-rich secretory proteins [CRISPs], antigen 5 [Ag5], and pathogenesis-related 1 [Pr-1]), antimicrobial peptides, γ-glutamyl transferases, and phospholipases (Undheim et al., 2015b; Hakim et al., 2015).

The genomic, transcriptomic, and proteomic approaches used in venom characterization studies are collectively termed 'venomics' (Ménez et al., 2006) and were recently reviewed by Sunagar et al. (2016), where the challenges of complete and accurate venom characterizations are discussed, with particular attention to transcriptome assembly, annotation, and proteomic abundance methods. Venom-gland transcriptome assemblies are prone to error due to minor sequence variation among toxins that are often collapsed into single transcripts, which has led to incorrect or incomplete venom characterizations (Macrander et al., 2015). Next-generation sequencing (NGS) methods such as Illumina provide an extraordinary amount of data, but have shortened read lengths in comparison to previous sequencing methods (such as Sanger). Based on these assembly challenges, Sunagar et al. (2016) and other transcriptome assembly comparisons (Honaas et al., 2016; Rana et al., 2016; Holding et al., 2018) suggest the use of multiple transcriptome assembly tools, as some may be more optimal than others depending on the organism or tissue. Annotation of toxin transcriptome sequences is often based on

* Corresponding author. Florida State University, Department of Biological Science, 319 Stadium Dr., Tallahassee, 32306-4295, FL USA.
E-mail address: drokyta@bio.fsu.edu (D.R. Rokyta).

https://doi.org/10.1016/j.toxicon.2018.07.030
Received 6 July 2018; Received in revised form 25 July 2018; Accepted 31 July 2018
Available online 04 August 2018
0041-0101/

homology to previously annotated genes, and the putative toxin transcript may or may not be proteomically expressed. Because venom genes are often expressed in other tissues as well as the venom-glands, as is the case in some centipedes (Undheim et al., 2014; Smith and Undheim, 2018; Zhao et al., 2018), the inclusion of putative toxin transcripts, regardless of proteomic expression in the venom, may inflate the true toxin diversity for a given species (Rodríguez de la Vega and Giraud, 2016). Annotation based on homology can also lead to novel toxins being neglected as they often do not have any homologous matches in available databases.

Centipede venom characterizations have been somewhat limited in comparison to other venomous animals (*i.e.* snakes), and many of the venomics challenges discussed by Sunagar et al. (2016) have yet to be addressed. Futhermore, although some of the above-mentioned centipede venom characterizations used single individuals for venom-gland sequencing (Liu et al., 2012; González-Morales et al., 2014), several used pooled venom-gland RNA (Yang et al., 2012; Smith and Undheim, 2018; Zhao et al., 2018), and others did not specify the number of individuals or whether pooling of venom-gland tissue or RNA took place (Undheim et al., 2014; Rong et al., 2015). In all cases, the venom samples used for proteomics were pooled from multiple individuals, although Smith and Undheim (2018) did use RNA and venom from the same five individuals for both transcriptomic and proteomic analysis. The pooling of venom samples and/or RNA is common in venomics, especially in invertebrates where sampling from a single individual does not always provide sufficient yield for adequate analysis. Unfortunately, sample pooling of multiple individuals does not allow for biological replicates for which to examine potential variation.

We completed quantitative transcriptomics and proteomics on two *Scolopendra viridis* (Say, 1821) individuals from North Florida, using an Illumina sequencing platform and independent quantitative mass spectrometry on their venoms. *Scolopendra viridis* are members of Scolopendromorpha (Family: Scolopendridae) and are distributed throughout the southern United States and Mexico (Shelley, 1987, 2002; McAllister et al., 2015; Charruau et al., 2018). Adults average approximately 6 cm in length and are described as either blue or green in color, with the presence or absence of light-brown or golden lateral stripes (Fig. 1, Shelley, 1987; Charruau et al., 2018). Previous venom characterization of *S. viridis* from the southern state of Morelos, Mexico (abbreviated MO) revealed sodium channel ion-toxins, phospholipases, proteases, hyaluronidases and venom allergens (González-Morales et al., 2009, 2014). The MO populations of *S. viridis* are separated from the Northern Florida *S. viridis* populations by approximately 2700 km. Intraspecific venom variation is common across venomous taxa (Coelho et al., 2014; Rokyta et al., 2015b; Margres et al., 2015, 2017; Carcamo-Noriega et al., 2018; Schaffrath et al., 2018) and recently established in *S. subspinipes* (Smith and Undheim, 2018). We therefore compared our newly generated venomics data for *S. viridis* from Florida (abbreviated FL) with the available venomics data for *S. viridis* (MO). Our venom characterization of *S. viridis* overcomes some of the commonly identified challenges in venomics research while providing a complete reference transcriptome and proteome of a scolopendromorph centipede species.

2. Materials and methods

2.1. Centipedes, venoms, and venom-glands

Venom-gland transcriptomic and venom proteomic analyses were performed independently on two individual *S. viridis* labeled C0167 and C0169. These specimens were collected in May of 2015 in the northern Florida counties of Madison (C0167) and Leon (C0169), which are separated by approximately 130 km. Both individuals were identified as male on the basis of two gonopods (sometimes called styles) present on the genitals (Bonato et al., 2010; McMonigle, 2014).

Fig. 1. Dorsal view of a representative adult *S. viridis* from northern Florida.

Venom and venom-glands were extracted using methods similar to those previously described (Rokyta and Ward, 2017; Ward et al., 2018). Both *S. viridis* individuals used in this study were housed under identical conditions, including feeding, venom extractions, and venom-gland removal, for the purpose of biological replication. Prior to venom extraction, centipedes were starved for seven days to ensure ample venom production. Centipedes were briefly anesthetized with CO_2 for approximately one minute prior to being positioned dorsal side down on a Velcro® fitted venom extraction surface. A serrated metal spatula was placed between the forcipules of the centipede, and electrical stimulation was applied at the base of the forcipules to induce a muscle contraction. Venom was released onto the serrated spatula as the centipede grasped the spatula with its forcipules. Venom was then lyophilized and stored at − 80°C until later use. Four days after venom extraction, venom-glands were removed under a stereoscopic microscope with micro-surgical dissection instruments. Prior to gland removal, centipedes were fully anesthetized with CO_2 for approximately 15 min. The body of the centipede was taped to a sterile surface, dorsal side down, such that both forcipules were visible under the microscope. Using a micro-surgical blade, each forcipule was cut starting at the base and curving around the outer edge up to the forcipule tip. Tweezers were used to peal back the forcipule exoskeleton and expose the venom-glands. Venom-glands were removed with micro-surgical tweezers by gently grasping and pulling the gland and surrounding tissues. The tissue was immediately transferred to a 0.7 mL microcentrifuge tube containing 100 μL RNAlater. The gland tissue in RNAlater was then stored at 4 °C overnight and transferred to − 80°C until further use. The body of each centipede specimen was preserved in 95% ethanol and submitted as voucher specimens to the Florida State Collection of Arthropods with the Florida Department of Agriculture and Consumer Services Division of Plant Industry (FDACS-DPI) in Gainesville, FL, under Sample Number E2018-3472-1.

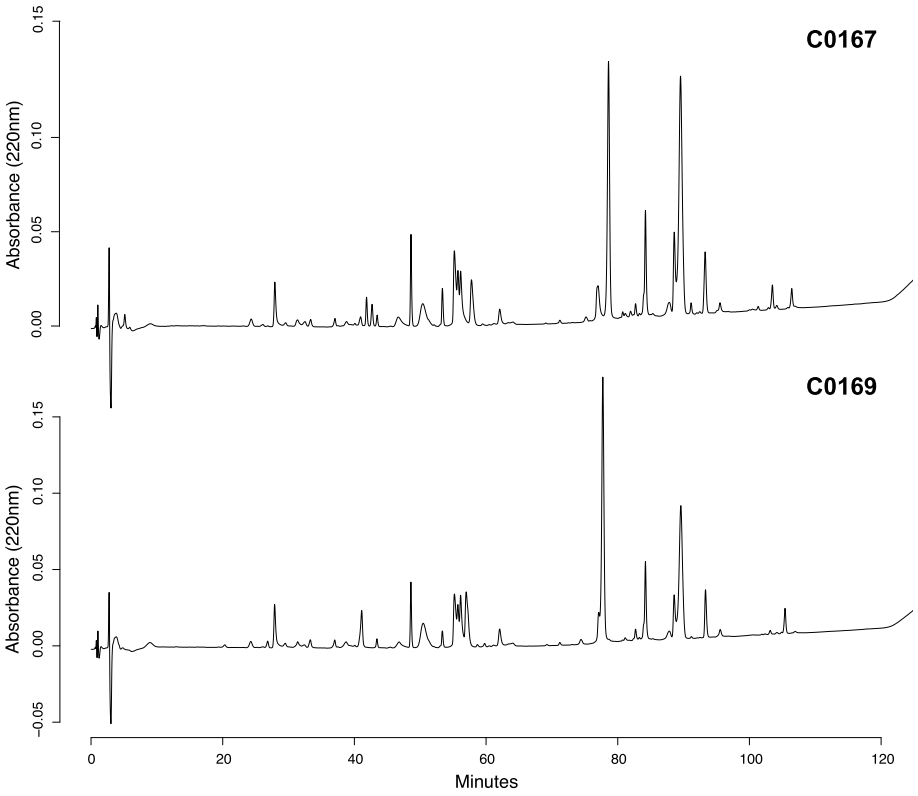

Fig. 2. RP-HPLC profiles for *S. viridis* individuals C0167 and C0169 indicate approximately 12 distinct peak clusters between 20 and 120 min, consistent with the diversity of venom components identified by LC-MS/MS proteomics.

2.2. Transcriptome sequencing

Centipede venom-gland RNA extraction was performed as previously described (Rokyta and Ward, 2017; Ward et al., 2018). In brief, the 100 μL RNAlater containing the venom-gland tissue was mixed with 500 μL of TriZol (Invitrogen) and homogenized with a sterile 20 gauge needle and syringe. Once homogenized, an additional 500 μL of TriZol was added along with 20% chloroform, and the tissue mixture was then transferred to phase lock heavy gel tubes (5Prime). To separate the RNA from DNA and other cellular debris, the gel tubes with the tissue mixture were centrifuged and isopropyl alcohol was used to pellet the isolated RNA. The RNA pellets were then washed with 75% ethanol, and the purified RNA isolate was washed with 70% ethanol prior to Qubit RNA quantification (Thermo Fisher Scientific). RNA quality was checked using the RNA 6000 Pico Bioanalyzer Kit (Agilent Technologies) according to manufacturer's instructions, with a total yield of approximately 855 ng and 96 ng for C0167 and C0169, respectively. Because RIN scores for invertebrate RNA are often difficult to determine due to 28S rRNA fragments that co-migrate with the 18S rRNA (Paszkiewicz et al., 2014), the RNA quality was assessed by the presence and abundance of a double peak that corresponds to the typical 18S rRNA peak.

Following quantification and quality analysis, mRNA was isolated from approximately 90–100 ng of total RNA from each C0167 and C0169 using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs). Continuing to follow methods previously described (Rokyta and Ward, 2017; Ward et al., 2018), a fragmentation time of 15.5 min was used to generate fragments of approximately 370 nucleotides (120 bp adapter-ligated), and the isolated mRNA was immediately used in cDNA library preparation 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). Agencourt AMPure XP PCR Purification Beads were used to purify cDNA throughout the protocol following manufacturer's instructions. The quality of the purified cDNA libraries was checked using a High Sensitivity DNA Bioanalyzer Kit (Agilent Technologies) per manufacturer's instructions. The total cDNA yield for C0167 was 503 ng (20 μL of 108.2 nM) with an average fragment size of 393 bp, and the total cDNA yield for C0169 was 350 ng (20 μL of 76.7 nM) with an average fragment size of 386 bp. Amplifiable concentrations of each sample were determined by KAPA qPCR performed by the Florida State University Molecular Cloning Facility, resulting in 72.75 nM and 54.15 nM amplifiable concentrations for C0167 and C0169, respectively. Each sample was diluted to ~5 nM and pooled with other 5 nM cDNA libraries to be run on the same sequencing lane. The quality of the pooled cDNA library sample was checked using a High Sensitivity DNA Bioanalyzer Kit (Agilent Technologies), and the amplifiable concentration was confirmed with an additional round of KAPA qPCR. Sequencing was performed by the Florida State University College of Medicine Translational Laboratory using an Illumina HiSeq 2500.

2.3. Proteomics

Proteomic analyses were completed following methods previously described (Rokyta and Ward, 2017; Ward et al., 2018). The venom used in this analysis was obtained from a single venom extraction for each individual in September 2015. Quantification of venom protein samples was completed using a Qubit Protein Assay Kit and approximately 5 μg of whole venom was digested using the Calbiochem ProteoExtract All-in-One Trypsin Digestion Kit (Merch, Darmstadt, Germany) per manufacturer's instructions, resulting in approximately 4.3 μg of digested venom proteins. Digested protein samples were then frozen and dried using a SpeedVac.

LC-MS/MS was performed by the Florida State University College of Medicine Translational lab following methods previously described (Rokyta and Ward, 2017; Ward et al., 2018). Digested venom protein samples were resuspended in 0.1% formic acid to reach a final concentration of 250 ng/μL. Known concentrations of three highly-purified

recombinant *Escherichia coli* proteins (Abcam) were mixed in specified proportions prior to digestion to yield final desired concentrations of 2500 fmol of P31697 (Chaperone protein FimC), 250 fmol of P31658 (Protein deglycase 1), and 25 fmol of P00811 (Beta-lactamase ampC) per injection. The digested *E. coli* peptide mix was infused into venom samples prior to LC-MS/MS injection. A 2 μL aliquot of each sample was injected into an externally calibrated Thermo Q Exactive HF (high-resolution electrospray tandem mass spectrometer) in conjunction with Dionex UltiMate3000 RSLCnano System to perform LC-MS/MS analysis. Beginning with LC, samples were aspirated into a 50 μL loop and loaded onto the trap column (Thermo μ-Precolumn 5 mm, with nanoViper tubing 30 μm i.d.×10 cm), with a flow rate of 300 nL/min for separation on the analytical column (Acclaim pepmap RSLC 75 μMx 15 cm nanoviper). A 60 min linear gradient from 3% to 45% B was performed using mobile phases A (99.9% H_2O (EMD Omni Solvent) and 0.1% formic acid) and B (99.9% ACN and 0.1% formic acid). The LC eluent was directly nanosprayed into a Q Exactive HF mass spectrometer (Thermo Scientific). The Q Exactive HF was operated in a data-dependent mode and under direct control of the Thermo Excalibur 3.1.66 (Thermo Scientific) throughout the chromatographic separation. A data-dependent top-20 method was used for MS data acquisition, selecting the most abundant, not-yet-sequenced precursor ions from the survey scans (350–1700 m/z). Sequencing was performed using higher energy collisional dissociation fragmentation with a target value of 10^5 ions determined with predictive automatic gain control. Full scans (350–1700 m/z) were performed at 60,000 resolution in profile mode. MS2 were acquired in centroid mode at 15,000 resolution. A 15-second dynamic exclusion window was used and ions with a single charge, a charge more than seven, or an unassigned charge were excluded. All measurements were performed at room temperature, and each sample was run and measured in triplicate to facilitate label-free quantification and account for any machine-related variability between samples. The resulting raw files were searched with Proteome Discoverer 1.4 using SequestHT as the search engine, custom-generated FASTA databases, and percolator to validate peptides. The SequestHT search parameters used were: enzyme name = Trypsin, maximum missed cleavage = 2, minimum peptide length = 6, maximum peptide length = 144, maximum delta Cn = 0.05, precursor mass tolerance = 10 ppm, fragment mass tolerance = 0.2 Da, dynamic modifications, carbamidomethyl +57.021 Da(C) and oxidation +15.995 Da(M). Protein and peptide identities were validated using Scaffold (version 4.3.4, Proteome Software Inc., Portland, OR, USA) software. Peptide identities were accepted based on a 1.0% false discovery rate (FDR) using the Scaffold Local FDR algorithm. Protein identities were also accepted with an FDR of 1.0% and a minimum of one recognized peptide. Additional analyses using a more stringent parameter of three minimum recognized peptides were completed for each individual to investigate discrepancies between transcript and protein abundances.

Estimated proteomic abundances for each individual (C0167 and C0169) were calculated as described by Rokyta and Ward (2017) and Ward et al. (2018). Conversion factors for each of the three replicates were calculated by finding the slope of the best fit line of the known *E. coli* internal standard concentrations and observed normalized spectral counts, with an intercept at the origin. These conversion factors were then used to convert the normalized spectral counts for each venom protein in each replicate to a concentration value. Final concentrations for each sample were averaged across the three replicates per individual (C0167 and C0169).

2.4. Transcriptome assembly and analysis

Transcriptome assembly and analysis was performed following methods previously described (Rokyta and Ward, 2017; Ward et al., 2018). Only read-pairs that passed the Illumina quality filter were used in our analysis. Because we performed 150 paired-end sequencing with a target insert size of 250 nucleotides, we expected most read-pairs to show significant 3′ overlap and therefore used PEAR version 0.9.6 (Zhang et al., 2014) to merge reads for subsequent analyses. We used DNAStar NGen version 12.3.1 with 10 million merged reads and default transcriptome assembly settings to generate our primary transcriptome assembly. We retained only contigs with at least 200 reads. We were not expecting to find many known homologs of toxins for this species in public databases, so multiple search strategies were used to identify and annotate proteins in the transcriptome. Two of the strategies used the whole-venom mass-spectrometry results and the generated protein databases by applying TransDecoder version 2.0.1 (Haas and Papanicolaou, 2016) to our assembled transcriptomes. We first created a database using the TransDecoder-predicted protein sequences with a minimum length of 50 and searched our mass-spectrometry results against this database. We then filtered these results using Scaffold Viewer version 4.6.0. To accommodate possible short proteins in the venom, protein and peptide false-discovery rates were set to 1.0%, and the minimum number of peptides was set to one. In our second strategy, we wanted to ensure that small peptides were not being missed by the TransDecoder predictions, so we generated another database using all possible protein or peptide sequences of at least 50 amino-acids from all six possible reading frames. We filtered the results as before in Scaffold. For this second strategy, all contigs already annotated in the first strategy (*i.e.* those predicted by TransDecoder) were excluded. Our third strategy aimed to identify proteins from the transcriptome with homology to known toxins. To do this, we conducted a BLASTX (version 2.2.30+) search of our transcripts generated by NGen against the UniProt animal toxins database (downloaded on November 16, 2015) and attempted to annotate full-length putative toxins that showed a match against at least 80% of the length of a known toxin. For our fourth strategy, we conducted a BLASTX analysis of the transcripts generated by NGen against the National Center for Biotechnology Information (NCBI) non-redundant (nr) protein database (downloaded on November 13, 2015) to generate a general database of toxins and nontoxins expressed in the venom-glands. Only transcripts comprised of at least 500 reads with a match of at least 95% of the length of a known protein were considered. In our fifth strategy, we used Extender (Rokyta et al., 2012) to assemble the transcriptome from 1000 random reads to better ensure that no high-abundance transcripts were missed. Reads were only used if they had phred qualities of ≥30 at all positions and an exact match of 120 nucleotides for extension. We then searched the resulting contigs against the UniProt animal toxins database with BLASTX. To generate a final consensus transcriptome for this first set of annotated sequences, we first combined transcripts by individual by clustering based on coding sequences with cd-hit-est version 4.6 (Li and Godzik, 2006) with a sequence identity threshold of 1.0. We then screened for chimeras by aligning the merged reads against the resulting combined set with BWA version 0.7.12 (Li, 2013), allowing only reads with no mismatches relative to the reference. Resulting alignments were checked for regions with no coverage or multimodal coverage distributions. Because our RNA-seq libraries were sequenced with libraries from other species, we checked for sample cross-contamination by aligning the PEAR-merged reads of each other sample against our transcript set for each individual with BWA version 0.7.12 (Li, 2013), retaining mapped reads with three or fewer mismatches. Transcripts were removed as contaminants if they showed > 100× higher coverage for another library relative to the highest-coverage of the corresponding *S. viridis* libraries, had coverage over the entire length of the coding sequence, and had no homozygous variants relative to the consensus sequence.

Because of the paucity of proteomically confirmed venom proteins in public databases for centipedes, we added proteomic-driven annotation for six new transcriptome assemblies for each individual. We first processed the raw reads. We screened for and removed sample cross-leakage attributable to the demultiplexing step by comparing *k*-mer distributions for each sample against those of all other samples sequenced in the same lane. We generated 57-mer distributions using

jellyfish version 2.2.6 (Marçais and Kingsford, 2011) and identified 57-mers that were 500× more abundant in another sample relative to the focal sample. Reads for the focal sample were removed if ≥25% of their length was comprised of these 57-mers. We used Trim Galore! version 0.4.4 (Krueger, 2015) for adapter and quality trimming. We set the quality threshold to a phred of 5, and removed any reads less than 75 nucleotides in length after trimming. Reads were merged with PEAR version 0.9.10 (Zhang et al., 2014) with the default settings. We ran Extender (Rokyta et al., 2012) with 1000 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. We ran BinPacker version 1.0 (Liu et al., 2016) with a *k*-mer size of 31 using the merged and unmerged reads and treating all reads as unpaired. We ran Trinity version 2.4.0 (Grabherr et al., 2011) with a *k*-mer size of 31 and using both the merged and unmerged reads, treating all as unpaired. We ran SOAPdenovo-trans version 1.03 (Xie et al., 2014) with a *k*-mer size of 127 and using both the merged and unmerged reads. The unmerged reads were treated as paired in this assembly. We ran SeqMan NGen version 14.0 using both the merged and unmerged reads, treating all as unpaired. We ran rnaSPAdes version 3.10.1 (Bankevich et al., 2012) with $k = 127$, using both the merged and unmerged reads. The unmerged reads were treated as paired in this assembly. For each assembly, we used the getorf function from EMBOSS version 6.6.0.0 (Rice et al., 2000) with a minimum size of 90 nucleotides to extract amino-acid sequences of open reading frames, retaining only those open reading frames with both start and stop codons. We clustered the output of each assembly with cd-hit version 4.6 (Li and Godzik, 2006) with a sequence identity threshold of 1.0 to remove exact duplicates within assemblies. Each resulting data set was used separately as a database against which to search our mass-spectrometry results as described above. To generate a final consensus transcriptome for this second set of proteomic-based identifications, we first combined the six sets of putative toxins by individual by clustering based on coding sequences with cd-hit-est version 4.6 (Li and Godzik, 2006) with a sequence identity threshold of 1.0. We then screened for chimeras by aligning the merged reads against the resulting combined set with BWA version 0.7.12 (Li, 2013), allowing only exact matches. Resulting alignments were checked for regions with no coverage or multimodal coverage distributions. We then combined across individuals using cd-hit-est with a sequence identity threshold of 0.98.

We created the final consensus transcriptome by combining the transcripts for each individual from the first annotation approach with the combined transcripts for the second, purely MS-based approach, using cd-hit-est with a sequence identity threshold of 0.98. Transcript abundances were estimated on the basis of Bowtie2 (Langmead and Salzberg, 2012) version 2.3.0 alignments against the coding sequences of the final consensus transcriptome, using RSEM (Li et al., 2011) version 1.2.31, and alignments were based on all merged reads for each individual. Using the cenLR function in the robCompositions package in R (Templ et al., 2011), we applied the centered logratio (clr) transform (Aitchison, 1986) on all of our transcriptome and proteome abundances as previously described (Rokyta et al., 2015a); this transform is equivalent to a log transform for linear analyses and does not affect rank-based analyses. The presence of signal peptides was verified with SignalP version 4.1 using the default settings (Petersen et al., 2011). In the few cases where a signal peptide was not detected in an identified putative toxin using default settings, the sensitive option in SignalP was also tested. Because the majority of centipede venomics data is available through the NCBI transcriptome shotgun assembly (TSA) database, all putative toxins were searched against this database to follow naming conventions and assign toxin family classification.

2.5. Reversed-phase high-performance liquid chromatography

Reversed-phase high-performance liquid chromatography was performed on a single venom sample for each C0167 and C0169. Approximately 7 μg of protein was injected onto a Jupiter 5 μm C18 column (Phenomenex, Torrance, CA) using the standard solvent system of A = 0.1% trifluoroacetic acid (TFA) in water and B = 0.06% TFA in acetonitrile, and a Waters 2695 Separations Module with a Waters 2487 Dual λ Absorbance Detector. Samples were run using a flow rate of 0.2 mL/min over a 125-minute gradient from 10 to 75% solution B, followed by a 15-minute wash of 10% B.

2.6. Data availability

The raw transcriptome reads were submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject PRJNA340270, BioSamples SAMN09042581 (C0167) and SAMN09042582 (C0169), and SRA accessions SRR7102114 (C0167) and SRR7102113 (C0169). 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 PXD009665. 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 GGNE00000000. The version described in this paper is the first version, GGNE01000000.

3. Results and discussion

3.1. Venom-gland transcriptomes

For individual C0167, we generated 17,514,462 raw read pairs after Illumina quality filtering, 14,528,913 of which were merged. Ten million merged reads were then assembled into our primary C0167 transcriptome of 3161 contigs using NGen. In our MS-directed analysis, we annotated 16 unique coding sequences using TransDecoder and eight additional unique coding sequences using all possible open reading frames (ORFs). We annotated 73 unique coding sequences using BLASTX hits to the UniProt toxins database, and 427 unique coding sequences using BLASTX hits to the NCBI nr database. Using our Extender assembly, another 13 unique coding sequences were annotated by performing a BLASTX search of the UniProt animal toxins database. After screening for duplicates and chimeras, we identified a combined total of 445 unique coding sequences for C0167.

For individual C0169, we generated 16,851,368 raw read pairs after Illumina quality filtering, 14,352,884 of which were merged. Ten million merged reads were then assembled into our primary C0169 transcriptome of 1919 contigs using NGen. In our MS-directed analysis, we annotated 23 unique coding sequences using TransDecoder and three additional unique coding sequences using all possible ORFs. We annotated 73 unique coding sequences using BLASTX hits to the UniProt toxins database. After excluding duplicates from the C0167, we annotated 70 unique coding sequences using BLASTX hits to the NCBI nr database. Using our Extender assembly, another 12 unique coding sequences were annotated by performing a BLASTX search of the UniProt animal toxins database. After screening for duplicates and chimeras, we identified a combined total of 103 unique coding sequences for C0169.

Because very few proteomically confirmed centipede toxins are currently available in public databases, we also completed a proteomic-driven annotation using six additional assemblies for each individual. Raw reads were processed, filtered, and merged as described in the methods. Using only merged reads in Extender, we annotated 20 unique coding sequences for C0167 and 18 unique coding sequences for C0169. Using merged and unmerged reads, all treated as unpaired, we annotated 28 (C0167) and 29 (C0169) unique coding sequences using BinPacker, 23 (C0167) and 35 (C0169) unique coding sequences using Trinity, and 24 (C0167) and 36 (C0169) unique coding sequences using SeqMan NGen. Using merged and unmerged reads, treating unmerged reads as paired, we annotated 11 (C0167) and 15 (C0169) unique coding sequences using SOAPdenovo-trans, and 20 (C0167) and 33

(C0169) unique coding sequences using rnaSPAdes. Our proteomic-driven annotation, using six different assemblies, resulted in combined totals of 45 and 46 unique coding sequences for C0167 and C0169, respectively.

The annotated transcripts from each individual *S. viridis*, using both our primary and proteomic-driven annotation methods, were combined and duplicates were removed, producing a final consensus transcriptome of 520 unique protein-coding transcripts. This final consensus transcriptome was used for all subsequent transcript-abundance estimates and LC-MS/MS analyses for each individual. Transcripts were divided into two classes: toxins and nontoxins. Toxin transcripts were classified as such because they were proteomically confirmed in the venom of one or both individuals and therefore had high likelihoods of encoding proteins with toxic functions. Nontoxin transcripts were classified as such because they were not detected in the proteome of either individual. Some nontoxins, however, shared homology with toxin-like transcripts from other centipede species in the UniProt animal toxins database, but these were not classified as toxins in our analysis because they were not proteomically detected. We identified 39 toxin transcripts that were detected in the proteome of one or both *S. viridis* individuals (Table 1), which accounted for 330,236.92 and 213,607.56 transcripts per million (TPM) of the mapped reads in C1067 and C0169, respectively. We identified 481 nontoxin transcripts, including 20 transcripts with homology to toxin-like transcripts from other centipede species. The remaining 461 nontoxin transcripts likely encode proteins with essential cell function and toxin production, but have low likelihoods of encoding proteins or peptides with toxic functions. The nontoxins accounted for 669,763.09 TMP of the total mapped reads in C0167, and 786,392.40 TPM of the mapped reads in C0169.

3.2. Ion-channel scoloptoxins

Scoloptoxins (SLPTXs) were first described by Yang et al. (2012) in *S. subspinipes* and are one of the most diverse and abundant groups of centipede toxins (Undheim et al., 2014). Members of the SLPTX group of toxins have been shown to act on calcium, potassium, and sodium channels (Yang et al., 2012) and have been documented in *Scolopendra*, *Cormocephalus*, and *Ethmostigmus* genera belonging to the order Scolopendromorpha (Undheim et al., 2014). The SLPTX class of toxins currently consists of 31 recognized families with protein sizes of 3–24 kDa that contain 3–18 cysteine (Cys) residues or are classified as linear peptides (Undheim et al., 2014, 2015b; Smith and Undheim, 2018). Of the 31 recognized SLPTX families, we identified six in the venom-gland transcriptome of *S. viridis* (SLPTX1, SLPTX5, SLPTX10, SLPTX13, SLPTX15, and SLPTX16), five of which were also detected in the proteome and confirmed as toxins (SLPTX5 was not detected in the

Table 1
Toxins identified in the venom-gland transcriptome and proteome of *Scolopendra viridis*.

Toxin	Signal peptide	Precursor (aa)	Cysteine Residues	MW (kDa)	C0167 TPM	C0169 TPM	C0167 fmol	C0169 fmol
β-PFTx-1	Yes	345	6	36.9	30,827.18	16,468.06	42.30	383.90
β-PFTx-2	Yes	345	3	36.8	21,269.67	12,484.18	398.68	995.79
β-PFTx-4	Yes	351	5	38.3	11,181.02	6,696.17	–	3.43
β-PFTx-5	Yes	335	3	35.9	11,171.43	6,081.36	6.02	11.57
β-PFTx-6	Yes	363	4	38.9	4,039.95	2,606.46	–	75.38
β-PFTx-7	Yes	327	5	35.8	9,170.00	6,457.00	6.23	–
CAP2-1	Yes	219	6	20.8	11,337.73	6,667.06	1,112.74	2,978.70
CAP2-2	No	211	7	23.4	11,317.97	4,315.59	667.10	1,770.29
Chitinase-1	Yes	377	4	39.1	853.21	418.78	93.51	593.18
DUF3472-1	Yes	433	1	46.8	1,544.83	1,172.14	42.41	331.99
DUF3472-2	Yes	431	–	46.3	5,891.40	2,594.54	187.87	405.83
HYAL-1	Yes	378	5	41.2	110.30	267.98	–	11.53
LDLA-1	Yes	231	7	24.2	5,634.00	4,493.66	–	9.86
LDLA-2	Yes	216	8	22.5	14,162.80	4,743.93	–	54.13
LDLA-3	Yes	185	6	19.0	16,130.11	11,023.70	66.42	70.60
LDLA-4	Yes	216	8	22.4	0.65	4,535.99	–	60.80
LDLA-5	Yes	154	6	15.4	34,976.27	21,298.13	–	31.35
pM12A-1	Yes	427	12	47.5	20,972.31	9,796.91	697.40	2,311.60
pM12A-2	Yes	402	12	43.3	10,439.09	5,597.06	185.43	270.80
pM12A-3	Yes	415	12	43.0	6,889.26	4,557.32	–	42.64
pM12A-4	Yes	393	12	41.7	15,995.65	7,350.18	36.35	25.09
SLPTX1-1	Yes	119	6	10.6	18,090.87	13,433.43	31.44	225.69
SLPTX10-1	Yes	96	6	8.1	4,602.39	4,626.35	16.78	227.67
SLPTX10-2	Yes	114	6	10.8	2,701.08	2,804.20	140.31	1095.58
SLPTX10-3	Yes	80	6	6.4	17,602.61	13,628.32	34.75	208.81
SLPTX10-4	Yes	96	6	8.1	–	516.54	–	257.54
SLPTX10-5	Yes	96	6	8.1	2,090.96	382.41	132.57	291.16
SLPTX10-6	Yes	67	6	4.7	8,238.38	10,192.19	318.38	1,022.54
SLPTX13-2	Yes	73	8	6.2	8,423.77	4,811.58	64.16	152.14
SLPTX15-3	Yes	79	4	6.7	1,577.80	823.54	456.67	620.58
SLPTX15-4	Yes	78	4	6.1	3,641.32	2,566.77	89.19	169.93
SLPTX15-5	Yes	75	4	6.4	9,366.29	7,071.09	65.95	226.39
SLPTX15-6	Yes	87	4	7.8	5,086.35	8,008.28	21.12	233.33
SLPTX16-1	Yes	110	8	8.8	200.07	1,756.24	133.87	76.68
SLPTX16-2	Yes	113	9	9.6	40.11	3.40	56.19	32.97
VP-1	Yes	193	4	19.2	152.79	89.54	10.66	–
VP-2	Yes	217	6	23.7	3,662.61	1,819.88	41.97	171.84
VP-3	Yes	104	2	9.2	597.54	1,328.18	–	140.81
VP-4	Yes	135	8	12.8	247.15	119.42	–	21.25

Cysteine residues and molecular weights were determined using ExPASy ProtParam (Gasteiger et al., 2005) and do not include signal peptides. Abbreviations: *β*-PFTx—*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, pM12A—M12A family of metalloproteases, SLPTX—scoloptoxins, VP—venom protein.

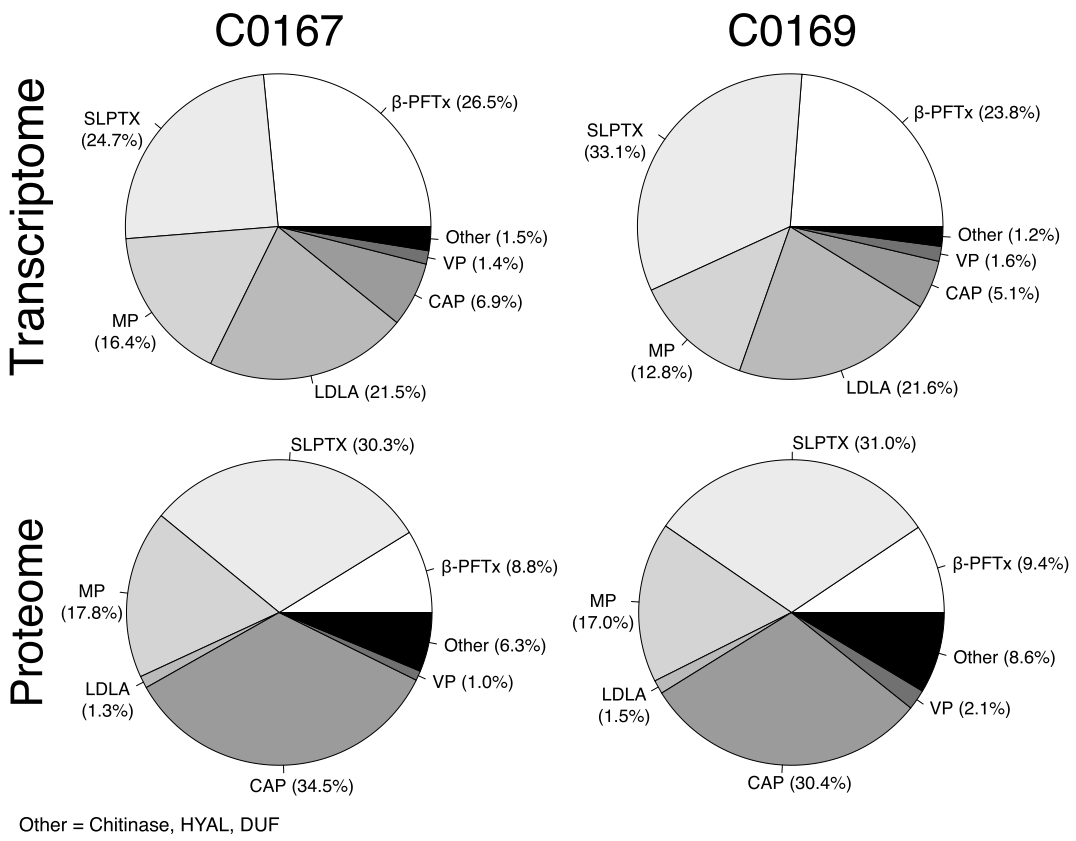

Fig. 3. Toxin class abundances were similar between venom-gland transcriptomes and venom proteomes of both individuals, but the class-level abundance comparisons of transcriptome and proteome did not agree well within individuals. Both LDLA domain containing proteins and *β*-PFTxs are represented considerably less in the proteomes than would be predicted by their abundance in the transcriptomes of each individual. Conversely, CAPs show a much greater representation in the proteomes than would be predicted by their transcriptome presence. Transcriptome abundances were based on transcripts per million (TPM) and percentages refer only to reads mapped to putative toxins (total toxin transcriptional output). Proteome abundances were expressed as molar percentages. Abbreviations: *β*-PFTx—*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—metalloprotease, SLPTX—scoloptoxins, VP—venom protein.

proteome and therefore omitted from the toxin class). The SLPTX toxins accounted for 81,662.00 TPM in C0167 and 70,624.34 TPM in C0169 (*i.e.* 24.73% and 33.06% of the total toxin transcriptional output, respectively; Fig. 3 and Table 1).

The SLPTX1 family is characterized by six Cys residues and a type 2 chitin-binding (CB2) domain (Undheim et al., 2015b). We identified one SLPTX1 in the venom of *S. viridis*, which contained a 21 amino acid signal peptide, six Cys residues, and shared 65% sequence identity to U-SLPTX1-Sm1a from *S. morsitans* (Undheim et al., 2014), where the U prefix indicates unknown function (King et al., 2008). The SLPTX1 identified in *S. viridis* also shared 29% sequence identity with the recently characterized spooky toxin (SsTx) from *S. subspinipes*, which was shown to block KCNQ potassium channels and was extremely lethal in mice (Luo et al., 2018), indicating that the SLPTX1 family may also target potassium channels.

We identified six members of the SLPTX10 family, all of which contained a 21–23 amino acid signal peptide and six Cys residues, as is characteristic of this family (Undheim et al., 2014, 2015b). The SLPTX10 family can inhibit either potassium or calcium channels (Liu et al., 2012; Undheim et al., 2014, 2015b). SLPTX10-1, SLPTX10-3, SLPTX10-4, and SLPTX10-5 all shared over 59% sequence identity to an approximately 8 kDa *Scolopendra* toxin (function unknown) identified in *Scolopendra viridicornis nigra* (Rates et al., 2007). SLPTX10-2 shared 40% sequence identity to an SLPTX10 from *S. morsitans* and SLPTX10-6 shared 60% sequence identity to an SLPTX10 from *C. westwoodi*, neither of which have a known function (Undheim et al., 2014).

One SLPTX13 was identified in the venom of *S. viridis* and contained a 17 amino acid signal peptide and eight Cys residues. The identified SLPTX13 shared 19% sequence identity to SsTx and over 34% sequence identity to two calcium channel inhibitors, *ω*-SLPTX-Ssm2a and *ω*-SLPTX-Ssm2b, from *S. mutilans* and *S. subspinipes*, respectively (Yang et al., 2012). The *S. viridis* SLPTX13 also shared 52% sequence identity to U-SLPTX13-Cw2a from *C. westwoodi*, which contains an inhibitory cysteine knot (ICK) domain (Undheim et al., 2014) suggesting that family 13 of the SLPTXs may act as calcium channel inhibitors.

The SLPTX15 family is characterized by 4–6 Cys residues and have been shown to act on potassium, calcium, and sodium ion channels (Liu et al., 2012; Undheim et al., 2015b). We identified four members of the SLPTX15 family in the venom of *S. viridis*, all of which contained an 18–24 amino acid signal peptide and four Cys residues. The SLPTX15 toxins identified shared 28–48% sequence identity to SsTx, which also contained four Cys residues (Luo et al., 2018), suggesting that the SLPTX15 toxins identified in *S. viridis* may also act as potent inhibitors of potassium channels. SLPTX15-5 and SLPTX15-6 also shared sequence identity to Scolopendra toxins with unknown function identified in *Scolopendra angulata* and *S. v. nigra*, respectively (Rates et al., 2007).

We identified two members of the SLPTX16 family, which can have 3–9 Cys residues (8 is typical) and are characterized with a Von Willebrand factor type C (VWC-like) domain. SLPTX16-1 contained eight Cys residues and a 24 amino acid signal peptide, and SLPTX16-2 contained nine Cys residues and a 28 amino acid signal peptide. Both SLPTX16-1 and SLPTX16-2 shared sequence identity to U-SLPTX16-Sm3a from *S. morsitans* (Undheim et al., 2014), with 26% and 74%, respectively.

3.3. β-Pore-forming toxins

β-pore-forming toxins (β-PFTxs) are one of the most abundant venom components in both Scutigeromorpha and Scolopendromorpha centipedes and are characterized by the presence of an aerolysin-like toxin β-complex domain (Undheim et al., 2014, 2015b). Members of the β-PFTx family have been identified in bacteria, plants, and fungi, where they function as cytolysins via formation of a transmembrane β-barrel (made up of multiple toxin monomers), which inserts into the cell membrane to form a pore (Knapp et al., 2010; Szczesny et al., 2011; Dal Peraro and Van Der Goot, 2016). Although the exact function of β-PFTxs in centipede venom is unknown, formation of the β-barrel is facilitated by proteolytic activation (Knapp et al., 2010; Dal Peraro and Van Der Goot, 2016), suggesting that venom proteases, which are also common in centipede venoms, may be required for β-PFTx function (Undheim et al., 2014, 2015b). Once activated, the β-barrel forms a pore in the cell membrane through recognition of extracellular receptors, and the diversity within this family indicates that β-PFTxs may exhibit cell or receptor-type specificity (Knapp et al., 2010; Szczesny et al., 2011; Undheim et al., 2014), as has been documented in bacterial β-PFTxs (Dal Peraro and Van Der Goot, 2016). Based on the known function of β-PFTxs in other organisms and the symptoms following some centipede stings, β-PFTxs in centipede venoms are hypothesized to cause myotoxic and edematic effects (Malta et al., 2008) and may contribute to other symptoms as well (Undheim and King, 2011; Undheim et al., 2014).

We identified seven β-PFTx transcripts in the venom-gland transcriptome of *S. viridis*, six of which were detected in the proteome and classified as toxins (Table 1). The β-PFTxs accounted for 87,659.25 TPM in C0167 and 50,793.23 TPM in C0169, representing 26.54% and 23.78% of the total toxin transcriptional output, respectively (Fig. 3). Each β-PFTx identified in the venom of *S. viridis* contained an 18–20 amino acid signal peptide and ranged in molecular weight from 35.9 to 38.9 kDa. Although the β-PFTxs identified in *S. viridis* did not match to any Chilopoda or Scolopendra proteins in the NCBI nr database, all but one (β-PFTx-6) shared low coverage and sequence identity (<30%) to hypothetical or uncharacterized proteins in bacteria (β-PFTx-2 and β-PFTx-4), yeast (β-PFTx-1), deer tick (β-PFTx-5), and horseshoe crab (β-PFTx-7). All *S. viridis* β-PFTxs shared sequence identity (27–68%) to β-PFTxs previously identified in *S. morsitans*, *C. westwoodi*, and *E. rubripes* (Undheim et al., 2014).

3.4. Metalloproteases

We identified four metalloproteases (MPs) in the venom-gland transcriptome of *S. viridis*, all of which were detected in the proteome (Table 1). The MPs accounted for 54,296.31 TPM in C0167 and 27,301.47 TPM in C0169, which is 16.44% and 12.78% of the total toxin transcriptional output, respectively (Fig. 3). Metalloproteases are known to cause edema, necrosis, blisters, and inflammation, which are common symptoms of centipede envenomations (Malta et al., 2008; Undheim and King, 2011; Undheim et al., 2015b; Hakim et al., 2015), suggesting metalloproteases are ubiquitous in centipede venoms. Metalloproteases have also been recognized for their role in posttranslational modification leading to activation of other venom proteins (Bond and Beynon, 1995; Chen et al., 2012; Undheim et al., 2014), facilitating the spread and function of synergistic venom components. All four MPs identified in the venom-gland transcriptome of *S. viridis* were members of the M12A family of metalloproteases and shared 52–62% sequence identity to putative M12A (pM12A) toxins identified in *S. morsitans* (Undheim et al., 2014). Each identified MP contained a signal peptide of 20–27 amino acids, 12 Cys residues, and ranged in molecular weight from 43.0 to 47.5 kDa. Because the M12A MPs were the only proteases detected in the venom of *S. viridis*, they are likely responsible for the required activation of the β-PFTxs also present in the venom.

3.5. CAP proteins

The CAP family of proteins include cysteine-rich secretory proteins (CRISPs), antigen 5 (Ag5), and pathogenesis-related 1 (Pr-1) proteins (Gibbs et al., 2008) and have been identified as some of the more highly expressed proteins in centipede venoms (Rates et al., 2007; Liu et al., 2012; Undheim et al., 2014). Centipede CAP proteins have been categorized into three classes, centiCAP1, centiCAP2, and centiCAP3, based on phylogenetic evidence of three separate recruitment events (Joshi and Karanth, 2011; Fernández et al., 2014; Undheim et al., 2014, 2015b). CentiCAP2 is the most prominent CAP identified in centipedes of the Scolopendridae family (Undheim et al., 2014) and some centiCAP2s have been shown to act as calcium-channel toxins or trypsin inhibitors (Liu et al., 2012). We identified two centiCAP2 transcripts, CAP2-1 and CAP2-2, in the venom-gland transcriptome of *S. viridis* that contributed 6.86% and 5.14% of the total toxin transcriptional output in C0167 and C0169, respectively. Despite the relatively small transcriptional output in comparison to the SLPTX and β-PFTx toxin classes, the CAPs represented over 30% of the venom proteome in each C0167 and C0169 (Fig. 3). Both CAP2-1 and CAP2-2 shared 67–83% sequence identity to centiCAP2s identified in *C. westwoodi* (Undheim et al., 2014), but only CAP2-1 contained a signal peptide of 29 amino acids (Table 1). Cap2-1, with a molecular weight of 10.8 kDa, contained six Cys residues, and CAP2-2 had a molecular weight of 23.4 kDa and seven Cys residues. Because CAPs are considered fundamental allergens in bees, wasps, and ants (Hoffman, 2006), CAP2s present in the venom of *S. viridis* could cause similar allergic reactions, which have been observed after some centipede envenomations (Undheim and King, 2011).

3.6. Other toxins

We identified five proteins that contained a low-density lipoprotein receptor Class A repeat (LDLA) domain. In mammals, the LDL receptor functions in cholesterol metabolism and the LDLA domain is present in several other proteins with unrelated or unknown function (Daly et al., 1995). Proteins containing an LDLA domain have been identified in other centipede venoms, although their role in these venoms remains unknown (Undheim et al., 2014, 2015b). The LDLA domain containing protein transcripts were abundant in the venom-gland transcriptome of *S. viridis*, making up over 20% of the total toxin transcriptional output in each C0167 and C0169 (Fig. 3). This level of abundance did not translate to the venom proteome, where the LDLA domain containing proteins made up less than 2% of confirmed toxins (Fig. 3). All five LDLA domain containing proteins identified in *S. viridis* contained a 16–36 amino acid signal peptide, 6–8 Cys residues and shared 50–66% sequence identity to LDLA domain containing proteins identified in other centipede venoms (Undheim et al., 2014). Two *S. viridis* LDLA domain containing proteins, LDLA-2 and LDLA-4, shared over 50% sequence identity to a C1q tumor necrosis factor related protein in humans, which can function in immunity and energy homeostasis (Kishore et al., 2004).

We identified one venom hyaluronidase (HYAL), which contained a 24 amino acid signal peptide and five Cys residues. This HYAL shared 67% sequence identity to an HYAL from *S. morsitans* and over 38% sequence identity to HYALs from several organisms, including the Atlantic horseshoe crab, *Limulus polyphemus*, in a BLASTP search of the nr database. In centipede venoms, HYALs are thought to function as antibacterial agents, in the degradation of glycosaminoglycans, and may enable the spread of other venom proteins (Undheim et al., 2014, 2015b).

One chitinase was identified in the venom of *S. viridis*, which contained a 20 amino acid signal peptide and shared 89% sequence identity to a chitinase identified in *C. westwoodi* (Undheim et al., 2014). The function of chitinase in centipede venom is unknown (Undheim et al., 2014, 2015b), although it may play a role in breaking down chitin exoskeletons of insect prey (Kumar, 2000; Dugon and Arthur, 2012).

We identified six proteins with unknown functional classification. Two of these proteins contained domains with unknown function (DUF), DUF3472-1 and DUF3472-2. DUF3472-1 shared 46% sequence identity to a DUF identified in *E. rubripes*, and DUF3472-2 shared 69% sequence identity to a DUF identified in *S. morsitans* (Undheim et al., 2014). The molecular weights of DUF3472-1 and DUF3472-2 were 46.8 kDa and 46.3 kDa, respectively. The remaining four proteins with unknown function were generically classified as venom proteins (VPs). The VPs identified in *S. viridis* accounted for less than 2% of the toxin transcriptional output in C0167 and C0169 (Fig. 3). VP-1 contained a 19 amino acid signal peptide, four Cys residues, a molecular weight of 19.2 kDa, and shared 63% sequence identity to an uncharacterized protein found in *Scolopendra subspinipes dehaani* in the TSA database (Rehm et al., 2014). VP-2, VP-3, an VP-4 each contained a 23 amino acid signal peptide, and six, two, and eight Cys residues, respectively. VP-2 shared 37% sequence identity with a U-SLPTX from an unknown family in *C. westwoodi* (Undheim et al., 2014), and had a molecular weight of 23.7 kDa. VP-3 and VP-4 had molecular weights of 9.2 kDa and 12.8 kDa, respectively, and shared sequence identity to uncharacterized venom proteins in *S. morsitans* (VP-3; Undheim et al., 2014) and *S. s. dehaani* (VP-4; Rehm et al., 2014).

3.7. Transcript and protein abundances across individuals

The mRNA abundances for nontoxin-encoding proteins were highly correlated between individuals (Spearman's rank correlation $\rho = 0.88$, Pearson's rank correlation coefficient $R = 0.79$, and $R^2 = 0.62$; Fig. 4). The mRNA abundances for toxin-encoding proteins were also highly correlated (Spearman's rank correlation $\rho = 0.89$, Pearson's rank correlation coefficient $R = 0.62$, and $R^2 = 0.38$), with only two outliers outside the 99th percentile of differences in nontoxin measurements, LDLA-4 and SLPTX10-4, both of which were present in the venom of C0169 and completely absent from C0167 (Fig. 4, Table 2). Both LDLA-4 and SLPTX10-4 followed presence/absence patterns previously described (Ward et al., 2018), where an extreme difference in transcript abundance between individuals explains the difference in protein expression. The transcript abundance for LDLA-4 was 0.65 TPM in C0167 and 4535.99 TPM in C0169, and SLPTX10-4 was not detected in the transcriptome of C0167, with an abundance of 516.54 TPM in C0169 (Table 1).

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

Protein	C0167			C0169			Average	
	rep 1	rep 2	rep 3	rep 1	rep 2	rep 3	C0167	C0169
β-PFTx-4	–	–	–	–	–	10.28	–	3.43
β-PFTx-6	–	–	–	77.70	66.19	82.23	–	75.38
β-PFTx-7	14.36	4.34	–	–	–	–	6.23	–
HYAL-1	–	–	–	9.71	9.46	15.42	–	11.53
LDLA-1	–	–	–	14.57	4.73	10.28	–	9.86
LDLA-2	–	–	–	38.85	56.74	66.81	–	54.13
LDLA-4	–	–	–	48.56	56.74	77.09	–	60.80
LDLA-5	–	–	–	19.42	28.37	46.26	–	31.35
pM12A-3	–	–	–	48.56	33.10	46.26	–	42.64
SLPTX10-4	–	–	–	252.52	283.68	236.43	–	257.54
VP-1	14.36	8.68	8.93	–	–	–	10.66	–
VP-3	–	–	–	131.12	137.11	154.19	–	140.81
VP-4	–	–	–	24.28	18.91	20.56	–	21.25

Quantities are given in fmol. Abbreviations: β-PFTx—β-pore forming toxin, HYAL—hyaluronidase, LDLA—low-density lipoprotein receptor Class A repeat domain, pM12A—M12A family of metalloproteases, SLPTX—scoloptoxins, VP—venom protein.

A proteomic (LC-MS/MS) comparison between the two *S. viridis* individuals showed strong agreement (Spearman's rank correlation $\rho = 0.72$, Pearson's correlation coefficient $R = 0.77$, and $R^2 = 0.60$), with 26 of the 39 confirmed toxins expressed in both C0167 and C0169 (Fig. 5). The proteome of C0167 had two toxins that were not detected in C0169 (Table 2), both of which had low abundance in comparison to other toxins expressed in the C0169 proteome (Table 1). The proteome of C0169 had 11 toxins that were not detected in C0167 (Table 2), which were also low in abundance compared to other C0169 toxins, aside from SLPTX10-4 and VP-3, which were both moderately

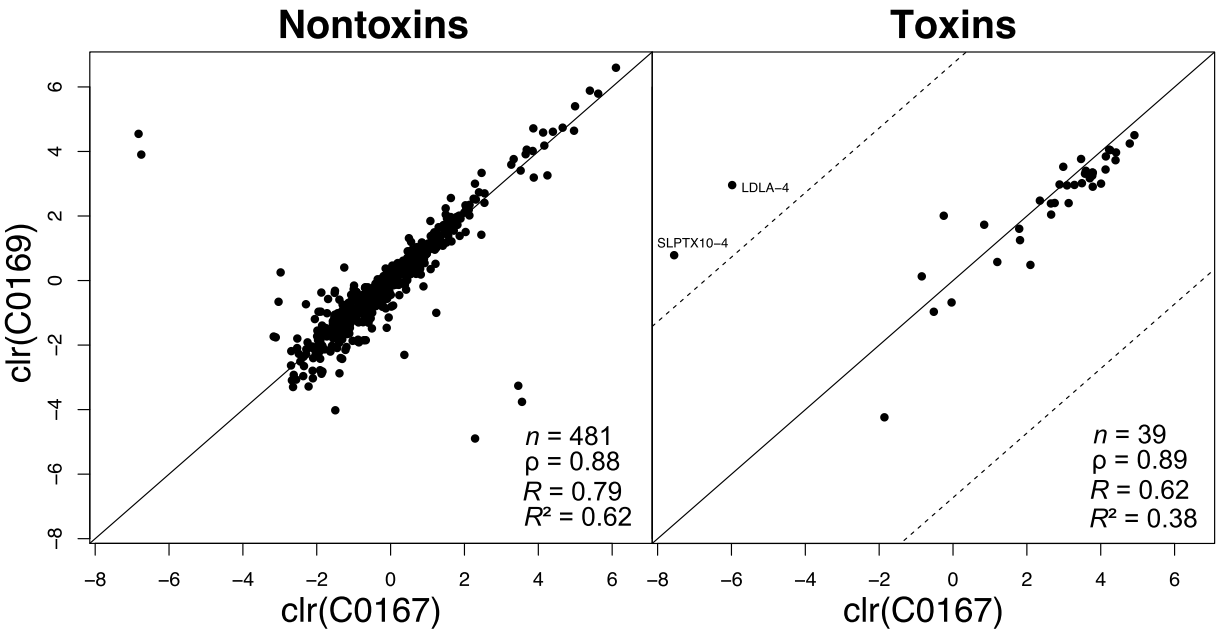

Fig. 4. A venom-gland transcript abundance comparison between *S. viridis* individuals (C0167 and C0169) showed strong agreement. Transcript levels were highly correlated between the venom-gland transcriptomes of the two individuals for both nontoxins and toxins. In the toxin plot, the dashed lines represent the 99th percentile of differences between the two nontoxin measures. Points outside the dashed line therefore represent toxins with unusually different expression levels relative to the nontoxins and are 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, LDLA—low-density lipoprotein receptor Class A repeat domain, SLPTX—scoloptoxins.

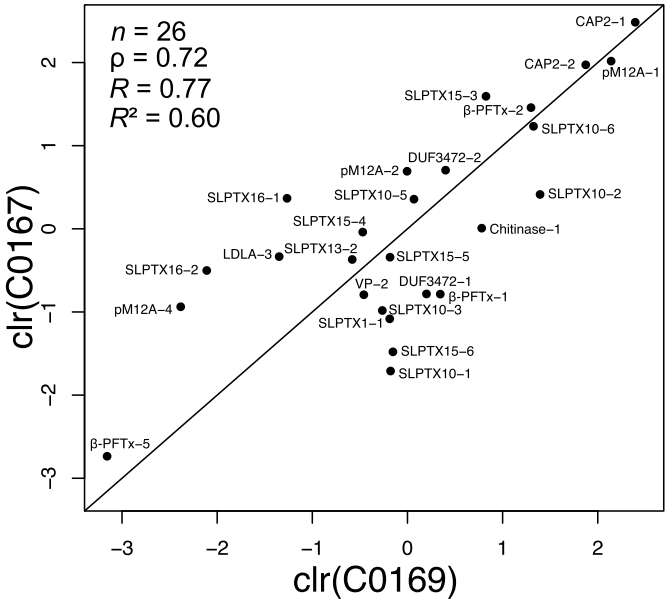

Fig. 5. A venom proteomic comparison between individual *S. viridis* (C0167 and C0169) showed strong agreement for proteins detected in both venom proteomes. Table 2 shows the proteomic presence/absence differences between the two individuals. Abbreviations: clr—centered logratio transformation, *n*—number of proteins, ρ—Spearman's rank correlation coefficient, *R*—Pearson's correlation coefficient, R^2—coefficient of determination, *β*-PFTx—*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, pM12A—M12A family of metalloproteases, SLPTX—scoloptoxins, VP—venom protein.

expressed (Table 1).

The high correlations of nontoxin-encoding transcripts, toxin-encoding transcripts, and protein expression across individuals indicate that any differences observed between the two individuals are biological rather than technical. The presence/absence differences in toxin expression between the two individuals can be explained, in part, by the use of a broad detection threshold (one minimum recognized peptide in Scaffold), although those with comparatively more abundant expression, such as SLPTX10-4 and VP-3, suggest biological venom expression differences that may be attributable to intraspecific expression variation.

RP-HPLC profiles of each *S. viridis* individual are consistent with the complexity of venom composition determined by LC-MS/MS (Fig. 2). Each individual profile indicates approximately 12 distinct peak clusters and we found the same number of toxin families expressed in the venom of each individual (Table 1).

3.8. Transcript versus protein abundance estimates

We found positive correlations in comparing transcriptomic (Fig. 4) and proteomic (Fig. 5) abundances between individuals, but we did not find the same relationship when comparing transcript and protein abundance estimates within either individual (Fig. 6). In our comparison of transcript and protein abundances in C0167, we found correlation coefficients of $\rho = 0.07$, $R = 0.12$, and $R^2 = 0.02$, and in C0169 we found $\rho = 0.09$, $R = 0.18$, and $R^2 = 0.03$ (Fig. 6, A). Discrepancies between the transcriptome and proteome should be assumed to be methodological/technical unless otherwise ruled out (Rokyta et al., 2015a). Technical issues related to protein digestion (Glatter et al., 2012) and ionization (Mirzaei and Regnier, 2006) efficiencies are both potential issues that cannot be ruled out or easily analyzed in terms of their contribution to the transcriptome/proteome discrepancy, although our use of internal standards and biological replicates should minimize these effects. We therefore explored other potential explanations for this discrepancy. One such explanation is that the transcriptome/proteome discrepancy may be a result of our broad protein detection threshold (one minimum recognized peptide in Scaffold). To test this, we completed additional transcript and protein abundance comparisons within each individual using more stringent parameters. We first compared transcript and protein abundance estimates that were proteomically detected in both individuals, thereby removing any proteins with presence/absence differences, based on a threshold of one minimum recognized peptide (Fig. 6, B). Within C0167 we found correlation coefficients of $\rho = 0.00$, $R = 0.04$, and $R^2 = 0.00$, and in C0169 we found $\rho = 0.05$, $R = 0.24$, and $R^2 = 0.06$. We then compared transcript and protein abundance estimates, using all proteomically detected transcripts within each individual, based on a threshold of three minimum recognized peptides (Fig. 6, C). Within C0167 we found correlation coefficients of $\rho = -0.05$, $R = 0.02$, and $R^2 = 0.00$, and in C0169 we found $\rho = -0.11$, $R = -0.12$, and $R^2 = 0.01$. The results of our more stringent analyses suggest that the discrepancies between transcript and protein expression are consistent across individuals and are not due to transcripts and proteins with lower expression, or those that show presence/absence differences between individuals.

Mapping biases, arising from conserved regions of paralogs mapping to an alternate transcript, can result in either an over or underestimation of transcript abundances (Wang et al., 2009; Peng et al., 2012; Rokyta et al., 2012) and may have contributed to the observed transcriptome/proteome discrepancy in *S. viridis*. Some mapping errors may be overcome through the use of biological replicates (Wang et al., 2009) and the two *S. viridis* individuals seemed to agree in relative transcript and protein abundances (Fig. 3) as well as share the apparent discrepancy between them. For instance, *β*-PFTx and LDLA transcripts each represent over 20% of the transcriptome in both C0167 and C0169, and proteomic expression of these toxin classes is reduced by the same magnitude in both individuals. Similarly, CAP transcripts represent only a small fraction of the transcriptome in each individual, but the proteomic expression of CAPs increased by the same magnitude in both individuals. To confirm this agreement, we calculated and compared the differences between mRNA and protein expression clr values for each individual using the proteomically confirmed toxins present in both C0167 and C0169 (represented in Fig. 5). We found a significant correlation when comparing the differences in mRNA and protein expression values across individuals ($p = 6.70 \times 10^{-8}$, $\rho = 0.87$, $R = 0.84$, and $R^2 = 0.71$, Fig. 7), suggesting the transcript to protein abundance discrepancy may be biological versus technical, as is supported by the direction and magnitude of the transcript-to-protein expression levels across individuals.

Post-translational modifications may also result in unexpected protein abundances when compared to the transcriptome, especially in smaller proteins (Rokyta and Ward, 2017; Ward et al., 2018), and may contribute to a potential size bias, where protein abundance estimates are lower or higher than predicted by the apparent transcriptome for small or large proteins, respectively. The toxins with the most drastic expression differences between transcriptome and proteome in *S. viridis*, however, have much larger molecular weights than would be expected if size, and therefore proteomic detection limitations, were a contributing factor in the discrepancy. Nonetheless, we tested whether the transcriptome/proteome discrepancy could be a result of a protein size bias by comparing the transcript and protein abundance differences within each individual to the length of their corresponding coding sequences (CDS length). We did not find a significant correlation between the transcript and protein abundance estimates and the CDS length for either C0167 ($p = 0.36$) or C0169 ($p = 0.28$), suggesting that protein size bias did not contribute to the discrepancy between transcriptome and proteome abundance estimates (Fig. 8).

Similar transcriptome/proteome discrepancies have been observed in snakes, where some toxins with high transcript abundance had low

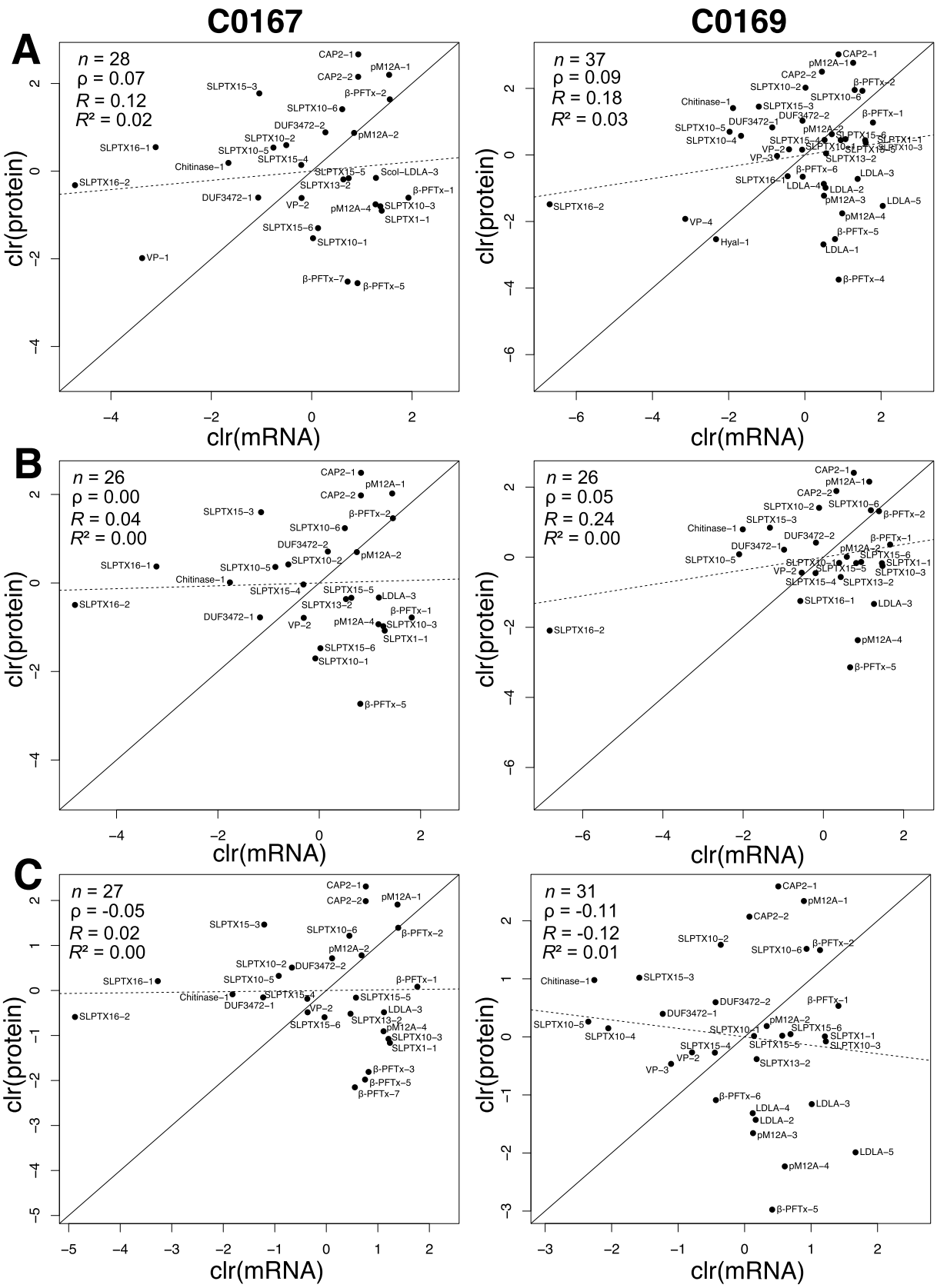

Fig. 6. We found weak agreement between transcript and protein abundances in both individuals using detection thresholds of (A) one minimum recognized peptide, (B) proteins confirmed in both individuals, and (C) three minimum recognized peptides. Abbreviations: clr—centered logratio transformation, *n*—number of transcripts, ρ—Spearman's rank correlation coefficient, *R*—Pearson's correlation coefficient, R^2—coefficient of determination, *β*-PFTx—*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, pM12A—M12A family of metalloproteases, SLPTX—scoloptoxins, VP—venom protein.

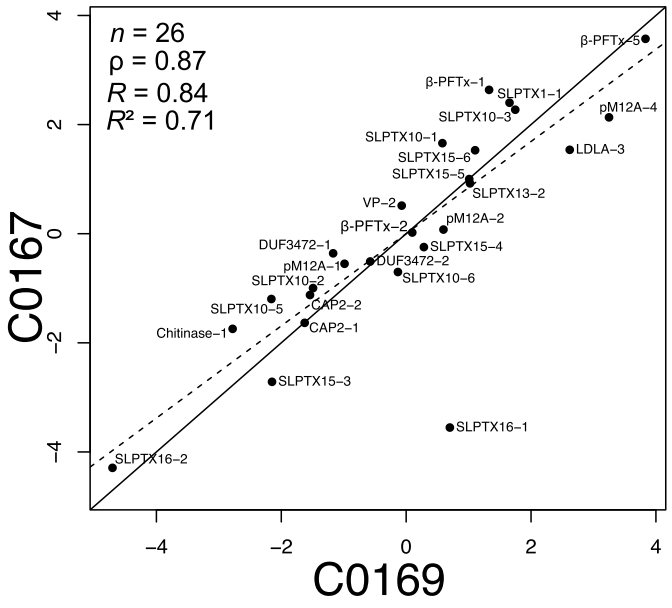

Fig. 7. We found a significant correlation between individuals when comparing mRNA and protein expression differences ($p = 6.70 \times 10^{-8}$), suggesting the discrepancy between transcript and protein abundances is likely biological rather than technical. Abbreviations: CDS–coding sequence, *n*–number of transcripts, ρ–Spearman's rank correlation coefficient, *R*–Pearson's correlation coefficient, R^2–coefficient of determination, *β*-PFTx—*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, pM12A—M12A family of metalloproteases, SLPTX—scoloptoxins, VP—venom protein.

expression in the proteome, and toxins with low transcription abundance had high expression in the proteome (Casewell et al., 2014). These discrepancies were not universal across the snake species analyzed, and even members of the same toxin family exhibited opposing patterns of transcriptome/proteome expression discrepancy in one species compared to another. Casewell et al. (2014) attributed these discrepancies to transcriptional and translational regulatory mechanisms, such as microRNAs that contribute to post-transcriptional repression (Bartel, 2009) and may play a role in the regulation of toxin expression (Durban et al., 2013), as well as post-translational proteolytic processing resulting in multiple protein products encoded by the same gene. In *S. viridis*, transcriptional and translational mechanisms that suppress expression are more likely involved in the regulation of toxins with high abundance in the transcriptome compared to the proteome (*i.e.* *β*-PFTxs and LDLAs), and toxins with low abundance in the transcriptome compared to the proteome (*i.e.* CAPs) may be subject to post-translational proteolytic processing. Translational efficiency, which is determined my multiple processes including mRNA-ribosome binding, availability of amino acids, and translation termination (Wang et al., 2015), may also contribute to the observed transcriptome/proteome discrepancies.

Another possible explanation for the discrepancy between transcript and protein abundances is that the abundance estimates, of either transcripts or proteins, may be influenced by the timing of venom or venom-gland extraction. Imaging mass spectrometry (IMS) of centipede venom glands from multiple species suggests that venom-gland secretory units are toxin specific and that these secretory units are not uniformly distributed throughout the venom glands (Undheim et al., 2015a). Additionally, extreme differences in toxin transcript expression have been shown between venom glands that are actively engaged in venom regeneration and those that are resting or “replete” (Morgenstern et al., 2011). Combined, these invertebrate venom-gland studies suggest that the timing of venom-gland removal affects the abundance of toxin transcripts and that venom-gland secretory units that produce different toxins may also be engaged at different times. Variation in the timing of venom extraction is common in centipede research, where some are starved for three weeks prior to venom extraction (Undheim et al., 2014), some are subject to multiple extractions over a 20-day time period (Liu et al., 2012), and several have not reported starvation or venom extraction frequency. Removal of the venom glands usually occurs 4–5 days after venom extraction (González-Morales et al., 2014; Undheim et al., 2014). In our study, both *S. viridis* were starved for seven days prior to venom extraction, venom glands were removed four days later, and transcriptomes and proteomes were consistent across individuals. CAP toxins were less

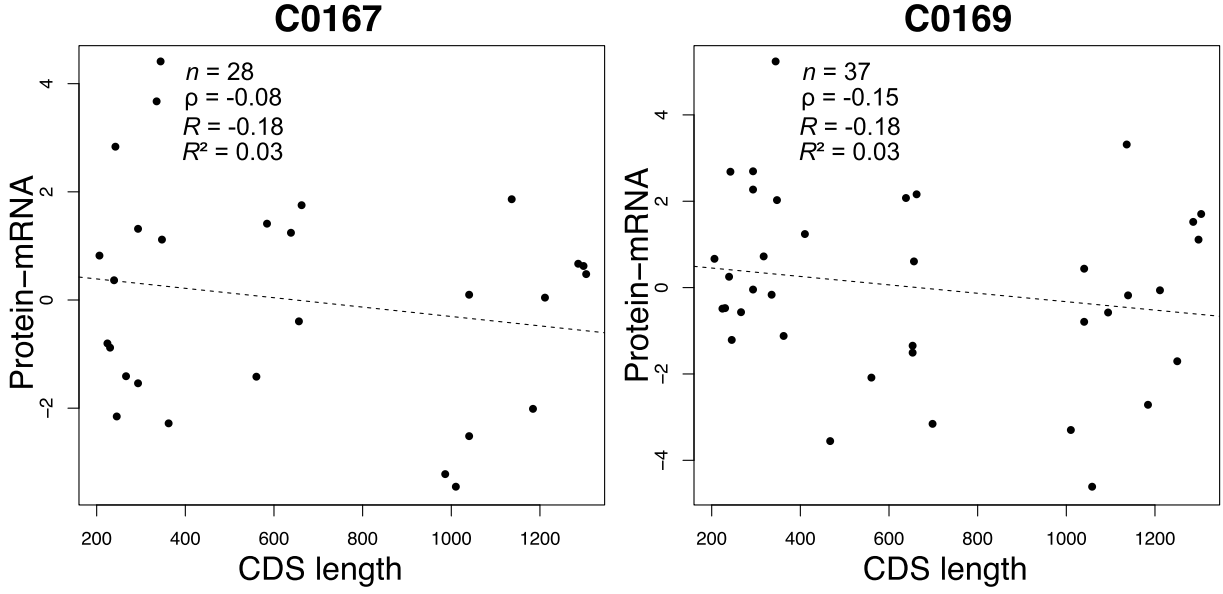

Fig. 8. We found a negative correlation between the discrepancy in protein and transcript abundance estimates and the sizes of the proteins in both *S. viridis* individuals, but these correlations were not significant (C0167: $p = 0.36$; C0169: $p = 0.28$). The lack of significant correlation implies that protein size (CDS length) bias did not contribute to the transcriptome/proteome abundance discrepancy. Abbreviations: CDS–coding sequence, *n*–number of transcripts, ρ–Spearman's rank correlation coefficient, *R*–Pearson's correlation coefficient, R^2–coefficient of determination.

abundant in the transcriptomes compared to the proteomes, and LDLA toxins and β-PFTxs were more abundant in the transcriptomes compared to the proteomes (Fig. 3). If timing-bias does contribute to the transcriptome/proteome discrepancy, our results show that the production of CAP proteins may be upregulated early in the venom regeneration process, and LDLA and β-PFTx protein production begins later. The timing-bias hypothesis could be tested through the comparison of time-series transcriptomes and venom sampling, which would also aid in the optimization of future invertebrate venomics studies.

3.9. Comparison to Scolopendra viridis from Mexico

González-Morales et al. (2009) isolated and characterized a phospholipase (PLA2) from the venom of *S. viridis* collected in the southern state of Morelos, Mexico (MO). In 2014, the same group completed a venom-gland transcriptome and venom proteomic characterization of *S. viridis* from the same locality, using a combination of a cDNA plasmid library and *de novo* MS/MS sequencing of isolated peptides from RP-HPLC fractions and SDS-PAGE gel extracts (González-Morales et al., 2014). The findings of González-Morales et al. (2014) resulted in the deposition of 50 cDNA transcripts to the GenBank database (accession numbers JZ574135–JZ574184), as well as 27 proteomically confirmed and sequenced peptides from gel extracts, and an additional seven N-terminal amino acid sequences from isolated RP-HPLC fractions (totaling 34 peptide sequences). To adequately compare the venom-gland transcriptome and proteome of *S. viridis* from Florida (FL) to *S. viridis* (MO), all 50 cDNA transcripts and 34 peptide sequences were blasted against the NCBI TSA and nr databases, limited to class Chilopoda, in April 2018.

Of the 50 cDNA transcripts available from the venom-gland transcriptome of *S. viridis* (MO), 10 shared sequence identity to putative toxins in the TSA or nr databases (Table 3). The remaining 40 cDNA transcripts available from *S. viridis* (MO) did not share homology with any toxin transcripts in either the TSA or nr databases, although the majority of them shared some level of sequence identity with transcripts from *S. s. dehaani* that were used in a phylogenetic analysis of myriapods (Rehm et al., 2014). Each of the 10 putative toxin transcripts was aligned with toxin transcripts of the same family that were identified in the *S. viridis* (FL) transcriptome. We found nine toxins in the venom of *S. viridis* (FL) that shared over 93% sequence identity to putative toxin transcripts in *S. viridis* (MO) (Table 3), however, only two of these transcripts (JZ574148: pM12A and JZ574169: β-PFTx) were proteomically confirmed in the venom of *S. viridis* (MO) (González-Morales et al., 2014).

The TSA and nr database blast results of sequenced peptides from *S. viridis* (MO) revealed putative toxin matches for seven of the 34 proteomically confirmed peptides (Table 3). The remaining 27 peptides did not have matches in either the TSA or nr databases. Included in the seven putative toxins were members of the β-PFTx, CAP2, DUF3472, γ-glutamyl transferase (GGT), PLA2, and SLPTX13 families, and those with familial matches to toxins in the venom of *S. viridis* (FL) were aligned to assess similarity. Three of the peptides from the venom of *S. viridis* (MO) matched with four corresponding *S. viridis* (FL) toxins: β-PFTx-4, CAP2-1, CAP2-2, and DUF3472-2 (Table 3), each with over 75% sequence identity and over 95% query coverage. The *S. viridis* (MO) peptides corresponding to SLPTX13, PLA2, and GGT toxins did not match to proteins present in the venom of *S. viridis* (FL), however, both the SLPTX13 and PLA2 matched with sequences identified in the venom-gland transcriptome of *S. viridis* (FL) that were not considered toxins in our analysis because they were not proteomically detected. We did not detect any GGT transcripts in the venom-gland transcriptome of *S. viridis* (FL).

Functional analysis of *S. viridis* (MO) venom revealed the presence of toxins capable of inhibiting human sodium channels, but not

Table 3
Putative and confirmed toxins identified in *S. viridis* (MO).

Putative toxin transcripts				
GenBank	TSA & nr Blast Results		*S. viridis* (FL)	Alignment
Accession	(best match)	Source	Toxin Homology	Score
JZ574142	U-SLPTX10-Cw1b	(Undheim et al., 2014)	93% ID to SLPTX10-4	427
			94% ID to SLPTX10-5	438
JZ574146	U-SLPTX13-Sm1a-2	(Undheim et al., 2014)	–	–
JZ574148	Scol-pM12A-14	(Undheim et al., 2014)	94% ID to pM12A-4	870
JZ574149	U-SLPTX1-Cw1	(Undheim et al., 2014)	97% ID to SLPTX-1	599
JZ574151	SLPTX17	(Smith and Undheim, 2018)	–	–
JZ574164	U-SLPTX10-Cw2a	(Undheim et al., 2014)	96% ID to SLPTX10-3	399
JZ574166	U-SLPTX10-Sm2b	(Undheim et al., 2014)	94% ID to SLPTX10-1	438
JZ574169	β-PFTx	(Smith and Undheim, 2018)	96% ID to β-PFTx-2	699
JZ574180	LDLA-42	(Undheim et al., 2014)	93% ID to LDLA-5	680
JZ574184	SLPTX15	(Smith and Undheim, 2018)	98% ID to SLPTX15-5	394

Proteomically confirmed toxins		
Amino Acid Sequence	TSA & nr Blast Results (best match)	*S. viridis* (FL) Toxin Match
NLGLLNWSGGFDLTWDR	Scol-β-PFTx-45	β-PFTx-4
YVFCQYGPGGNYLNQPIYK	Scol-CAP2-22	CAP2-1, CAP2-1
KLSSEGGLSIAVPGELR	Scol-GGT-4	n/a
LHHQLLPDEIEYESKFPNEILEK	Scol-GGT-8	n/a
FDTCCPGYQCLGCNPDGRGCKCQ	U-SLPTX13-Sa1a	SLPTX13-1 (transcript)
SLWNFFFMTFIGGKRAPWKYDGYGN	Scol-PLA2-3	PLA2-1 (transcript)
SAEDELAAVEKAGNGVLV	Scol-DUF3472-2	DUF3472-2

Putative toxin transcripts: Ten of the 50 cDNA transcripts available for *S. viridis* (MO) were identified as putative toxins based on homology to Chilopoda sequences in the TSA and nr databases (best match shown). Putative toxin transcripts were aligned with *S. viridis* (FL) transcripts of the same toxin family. **Proteomically confirmed toxins:** Seven of the 34 sequenced peptides from the venom of *S. viridis* (MO) were matched with Chilopoda toxins in the TSA and nr databases (best match shown). Sequences were then aligned with members of the same toxin family that were confirmed in the venom of *S. viridis* (FL), with toxin matches above 75% sequence identify and over 95% query coverage shown. Toxins that were proteomically confirmed in the venoms of both *S. viridis* (FL) and *S. viridis* (MO) are shown in bold. Abbreviations: β-PFTx—β-pore forming toxin, GGT—γ-glutamyl transferase, LDLA—low-density lipoprotein receptor Class A repeat domain, pM12A—M12A family of metalloproteases, SLPTX—scoloptoxins.

potassium channels, and at least two proteins with hyaluronidase activity (with no corresponding peptide sequences reported) (González-Morales et al., 2014). The venom of *S. viridis* (FL) did not contain toxins with known sodium channel specificity, but did contain several members of the SLPTX family of toxins known for acting on calcium and potassium channels. Only one hyaluronidase was detected in the proteome of C0169 (but not C0167) *S. viridis* (FL), indicating that *S. viridis* (MO) has a greater expression of HYALs compared to *S. viridis* (FL).

Overall, we found five proteomically confirmed toxins similar to both *S. viridis* (FL) and *S. viridis* (MO) venoms, including two β-PFTxs, one CAP2, one DUF3472, and one pM12A, as well as a handful of similar cDNA transcripts encoding putative toxins. The toxins and transcripts in common to both localities of *S. viridis* share high levels of similarity, however, these still represent only a small fraction of toxins and transcripts in either population (*i.e.* 5 of 39 toxins in the present study). Some of the discrepancy between the venoms of *S. viridis* (FL) and *S. viridis* (MO) may be explained by the differences in transcriptomic and proteomic methods and greater sampling would be required for confirmation, however, our comparative results suggest these venoms are distinct to their respective regions and represent evidence of either extreme intraspecific venom variation in *S. viridis*, or the existence of an entirely different species. Although current taxonomical classifications based on morphology have assigned *S. viridis* distributed from Southern Mexico through the Southwestern United States and as far east as Florida, Georgia and South Carolina, as the same species (Shelley, 2002; Charruau et al., 2018), *S. viridis* was formerly recognized as multiple species and subspecies that have since been synonymized (Shelley, 2002; Bonato et al., 2016), including the subspecies *S. viridis tolteca* located in Morelos, Mexico (Bücherl, 1941; Bonato et al., 2016). Some molecular phylogenetic centipede studies have included *S. viridis* (Regier et al., 2005; Vahtera et al., 2012), but the determined phylogenetic placements seem to be conflicting. Although Regier et al. (2005) did not report the collection locality of the *S. viridis* used in their analysis, they did recover the close relationship between *S. viridis* and *S. polymorpha* as has been determined morphologically (Shelley, 2002). Using a combination of morphology and molecular data, Vahtera et al. (2012) placed *S. viridis* (from New Mexico) in a sister clade to other members of the *Scolopendra* genus (although *S. polymorpha* was not included), and suggested *S. viridis* and other New World *Scolopendra* may at some point be removed from this genus. Additional taxonomical work, using both morphological and sequence data, should be completed within the Scolopendromorpha order and in Chilopoda more generally. For this reason, we encourage the submission of voucher specimens to the appropriate regional collections in all future venomics studies.

4. Conclusions

In our complete venom characterization of two *S. viridis* individuals from Florida, we overcame some of the commonly identified challenges of venomics research by utilizing multiple assembly tools and by performing both transcriptomic and proteomic-driven annotation methods to generate a comprehensive, quantitative reference transcriptome and proteome for a Scolopendromorpha centipede species. Consistent with more recent centipede venom characterizations (Smith and Undheim, 2018; Zhao et al., 2018), transcripts were only classified as toxins if they were proteomically confirmed in the venom. We identified 39 confirmed toxins, the most abundant of which were members of the ion-channel targeting SLPTX families, pore forming β-PFTxs, MPs, and CAPs. Other identified toxins included members of the LDLA family, as well as a chitinase, a HYAL, a handful of toxins with domains of unknown function (DUFs), and those that were unable to be classified by homology (VPs). Of the 39 confirmed toxins, 26 were present in the venom of both individuals and the remaining 13 were present in one individual and completely absent from the other. Because the two *S. viridis* individuals in this study were sampled from two separate North Florida locations, the presence/absence of some toxins indicates possible venom expression variation within *S. viridis*, although this cannot be determined with a sample size of two. Additionally, the presence/absence differences between the two individuals did not have an effect on the agreement of their overall venom composition and representation of toxin families. We found weak agreement between the transcriptome and proteome expression in both individuals, which may be attributed to technical biases, biological processes, or a combination of both. Potential technical biases in our approach could result from protein digestion or ionization efficiencies, and potential biological processes that may contribute to the observed transcriptome/proteome discrepancy include transcriptional or translational regulatory mechanisms, translational efficiency, post-translational proteolytic processing, or a timing-bias in protein and transcript expression at the time of venom extraction or venom-gland removal, respectively.

Using the venom characterization data produced in our study and data previously published by González-Morales et al. (2014), we were able to compare the venom-gland transcriptomes and venom proteomes of *S. viridis* (FL) and *S. viridis* (MO). We found five proteomically confirmed toxins in the venoms of *S. viridis* from both regions, along with a handful of homologous transcripts encoding putative toxins. The venom of *S. viridis* (MO) contained toxins capable of inhibiting human sodium channels, but did not have an effect on human potassium channels, and had a high expression of PLA2s and HYALs. The ion-channel SLPTX toxins identified in the venom of *S. viridis* (FL) have not been shown to act on sodium channels, but rather calcium and potassium channels. Additionally, PLA2 expression was detected in *S. viridis* (FL) venom, and the expression of HYALs was much lower in comparison to *S. viridis* (MO). Our comparative results indicate that the venoms of *S. viridis* (FL) and *S. viridis* (MO) exhibit extreme levels of variation between the two regions. Although greater sampling and consistent transcriptomic and proteomic methods would be required for confirmation, the existence of a different species or subspecies of *S. viridis* is likely the most reasonable explanation for this variation.

Although the pooling of individuals is common in venomics research, we have shown that using single venom samples and venom glands from independent individuals provides comparable results to those with pooled sampling, while enabling biological and technical replication. Future directions of centipede venom studies should aim to include individual characterizations from a range of populations, investigating potential timing-bias in mRNA and toxin expression, expanding into other Chilopoda orders, and ideally including molecular phylogenetic analyses, as the evolution of toxins and diversification of species within this ancient lineage is more complex than previously thought.

Ethical statement

Reporting standards: The authors declare that our manuscript describes original research and every effort was made to ensure the accuracy of the results and the account.

Data Access and Retention: The authors declare that all data is reported and provided within the manuscript and there is no additional data to make available.

Originality and Plagiarism: The authors declare that our manuscript is an original work with proper citations as needed.

Multiple, Redundant or Concurrent Publication: The authors declare that the data and work described in our manuscript has not and will not be submitted for consideration to another journal.

Acknowledgment of Sources: The authors have provided proper acknowledgment of sources to the best of their abilities.

Authorship of Paper: Both authors of the manuscript made significant contributions, and no one making significant contributions was excluded from authorship. Both authors have seen and read the submitted version of the manuscript and have approved submission.

Hazards and Human or Animal Subjects: This work did not

involve the use of vertebrate animals.

Disclosure and Conflicts of Interest: The authors declare no conflicts of interest.

Fundamental Errors in Published Works: If a fundamental error or inaccuracy is discovered in the results described by our literature review, the authors will immediately notify the editor or publisher.

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, 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, Michael Hogan for photography, and Schyler Ellsworth and Gunnar Nystrom for sex determination.

Transparency document

Transparency document related to this article can be found online at https://doi.org/10.1016/j.toxicon.2018.07.030.

References

Aitchison, J., 1986. The Statistical Analysis of Compositional Data. Chapman and Hall, London.

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

Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. Cell 136, 215–233.

Bonato, L., Chagas Junior, A., Edgecombe, J., Lewis, G.D., Minelli, A., Pereira, L., Shelley, R., Stoev, P., Zapparoli, M., 2016. ChiloBase 2.0 - a World Catalogue of Centipedes (Chilopoda). http://chilobase.biologia.unipd.it.

Bonato, L., Edgecombe, G.D., Lewis, J.G., Minelli, A., Pereira, L.A., Shelley, R.M., Zapparoli, M., 2010. A common terminology for the external anatomy of centipedes (Chilopoda). ZooKeys 17–51.

Bond, J.S., Beynon, R.J., 1995. The astacin family of metalloendopeptidases. Protein Sci. 4, 1247–1261.

Bücherl, W., 1941. Catá two kilos of the neotropical zone. Mem. Butantan Inst. 15, 251–372.

Carcamo-Noriega, E.N., Olamendi-Portugal, T., Restano-Cassulini, R., Rowe, A., Uribe-Romero, S.J., Becerril, B., Possani, L.D., 2018. Intraspecific variation of *Centruroides sculpturatus* scorpion venom from two regions of Arizona. Arch. Biochem. Biophys. 638, 52–57.

Casewell, N.R., Wagstaff, S.C., Wüster, W., Cook, D.A.N., Bolton, F.M.S., King, S.I., Pla, D., Calvete, J.J., Harrison, R.A., 2014. Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. Proc. Natl. Acad. Sci. U. S. A. 111, 9205–9210.

Charruau, P., Gagnon, R., Dupuis, F., 2018. Record of *Scolopendra viridis* say, 1821 (Chilopoda: Scolopendromorpha: Scolopendridae) from banco chinchorro biosphere reserve, quintana roo, Mexico. Dugesiana 25, 33–35.

Chen, S.-L., Li, Z.-S., Fang, W.-H., 2012. Theoretical investigation of astacin proteolysis. J. Inorg. Biochem. 111, 70–79.

Coelho, P., Sousa, P., Harris, D., van der Meijden, A., 2014. Deep intraspecific divergences in the medically relevant fat-tailed scorpions (*Androctonus*, Scorpiones). Acta Trop. 134, 43–51.

Dal Peraro, M., Van Der Goot, F.G., 2016. Pore-forming toxins: ancient, but never really out of fashion. Nat. Rev. Microbiol. 14, 77.

Daly, N.L., Scanlon, M.J., Djordjevic, J.T., Kroon, P.A., Smith, R., 1995. Three-dimensional structure of a cysteine-rich repeat from the low-density lipoprotein receptor. Proc. Natl. Acad. Sci. U. S A. 92, 6334–6338.

Dugon, M.M., Arthur, W., 2012. Prey orientation and the role of venom availability in the predatory behaviour of the centipede *Scolopendra subspinipes mutilans* (Arthropoda: Chilopoda). J. Insect Physiol. 58, 874–880.

Durban, J., Pérez, A., Sanz, L., Gómez, A., Bonilla, F., Rodríguez, S., Chacón, D., Sasa, M., Angulo, Y., Gutiérrez, J.M., Calvete, J.J., 2013. Integrated omics profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, *Crotalus simus simus*. BMC Genom. 14, 234.

Edgecombe, G.D., Giribet, G., 2007. Evolutionary biology of centipedes (myriapoda: Chilopoda). Annu. Rev. Entomol. 52, 151–170.

Fernández, R., Laumer, C.E., Vahtera, V., Libro, S., Kaluziak, S., Sharma, P.P., Pérez-Porro, A.R., Edgecombe, G.D., Giribet, G., 2014. Evaluating topological conflict in centipede phylogeny using transcriptomic data sets. Mol. Biol. Evol. 31, 1500–1513.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server. Springer.

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

Glatter, T., Ludwig, C., Ahrné, E., Aebersold, R., Heck, A.J., Schmidt, A., 2012. Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. J. Proteom. Res. 11, 5145–5156.

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

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

Haas, B., Papanicolaou, A., 2016. Transdecoder (Find Coding Regions within Transcripts). http://transdecoder.github.io.

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

Hoffman, D.R., 2006. Hymenoptera venom allergens. Clin. Rev. Allergy Immunol. 30, 109–128.

Holding, M.L., Margres, M.J., Mason, A.J., Parkinson, C.L., Rokyta, D.R., 2018. Evaluating the performance of *de novo* assembly methods for venom-gland transcriptomics. Toxins 10. https://doi.org/10.3390/toxins10060249.

Honaas, L.A., Wafula, E.K., Wickett, N.J., Der, J.P., Zhang, Y., Edger, P.P., Altman, N.S., Pires, J.C., Leebens-Mack, J.H., dePamphilis, C.W., 2016. Selecting superior *de novo* transcriptome assemblies: lessons learned by leveraging the best plant genome. PLoS One 11, e0146062.

Joshi, J., Karanth, K.P., 2011. Cretaceous–Tertiary diversification among select scolopendrid centipedes of South India. Mol. Phylogenet. Evol. 60, 287–294.

King, G.F., Gentz, M.C., Escoubas, P., Nicholson, G.M., 2008. A rational nomenclature for naming peptide toxins from spiders and other venomous animals. Toxicon 52, 264–276.

Kishore, U., Gaboriaud, C., Waters, P., Shrive, A.K., Greenhough, T.J., Reid, K.B., Sim, R.B., Arlaud, G.J., 2004. C1q and tumor necrosis factor superfamily: modularity and versatility. Trends Immunol. 25, 551–561.

Knapp, O., Stiles, B., Popoff, M.R., 2010. The aerolysin-like toxin family of cytolytic, pore-forming toxins. Open Toxinol. J. 3, 53–68.

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. https://github.com/FelixKrueger/TrimGalore.

Kumar, M.N.R., 2000. A review of chitin and chitosan applications. React. Funct. Polym. 46, 1–27.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. 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., Wang, I.X., Li, Y., Bruzel, A., Richards, A.L., Toung, J.M., Cheung, V.G., 2011. Widespread RNA and DNA sequence differences in the human transcriptome. Science 333, 53–58.

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

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

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

Luo, L., Li, B., Wang, S., Wu, F., Wang, X., Liang, P., Ombati, R., Chen, J., Lu, X., Cui, J., Lu, Q., Zhang, L., Zhou, M., Tian, C., Yang, S., Lai, R., 2018. Centipedes subdue giant prey by blocking KCNQ channels. Proc. Natl. Acad. Sci. U. S A. https://doi.org/10.1073/pnas.1714760115. 201714760.

Macrander, J., Broe, M., Daly, M., 2015. Multi-copy venom genes hidden in *de novo* transcriptome assemblies, a cautionary tale with the snakelocks sea anemone *Anemonia sulcata* (Pennant, 1977). Toxicon 108, 184–188.

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

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

Margres, M.J., McGivern, J.J., Seavy, M., Wray, K.P., Facente, J., Rokyta, D.R., 2015. Contrasting modes and tempos of venom expression evolution in two snake species. Genetics 199, 165–176.

Margres, M.J., Wray, K.P., Hassinger, A.T., Ward, M.J., McGivern, J.J., Moriarty Lemmon, E., Lemmon, A.R., Rokyta, D.R., 2017. Quantity, not quality: rapid adaptation in a polygenic trait proceeded exclusively through expression differentiation. Mol. Biol. Evol. 34, 3099–3110.
McAllister, C.T., Robison, H.W., Connior, M.B., 2015. New geographic distribution records for centipedes (Chilopoda: Scolopendromorpha) from Oklahoma. In: Proceedings of the Oklahoma Academy of Science, vol. 94.
McMonigle, O., 2014. Centipedes in Captivity: the Reproductive Biology and Husbandry of Chilopoda. Coachwhip Publications.
Ménez, A., Stöcklin, R., Mebs, D., 2006. "Venomics" or: the venomous systems genome project. Toxicon 47, 255–259.
Mirzaei, H., Regnier, F., 2006. Enhancing electrospray ionization efficiency of peptides by derivatization. Anal. Chem. 78, 4175–4183.
Morgenstern, D., Rohde, B.H., King, G.F., Tal, T., Sher, D., Zlotkin, E., 2011. The tale of a resting venom gland: transcriptome of a replete venom gland from the scorpion *Hottentotta judaicus*. Toxicon 57, 695–703.
Paszkiewicz, K.H., Farbos, A., O'Neill, P., Moore, K., 2014. Quality control on the frontier. Front. Genet. 5, 157.
Peng, Z., Cheng, Y., Tan, B.C.-M., Kang, L., Tian, Z., Zhu, Y., Zhang, W., Liang, Y., Hu, X., Tan, X., Guo, J., Dong, Z., Liang, Y., Bao, L., Wang, J., 2012. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat. Biotechnol. 30, 253.
Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.
Rana, S.B., Zadlock IV, F.J., Zhang, Z., Murphy, W.R., Bentivegna, C.S., 2016. Comparison of *de novo* transcriptome assemblers and k-mer strategies using the killifish, *Fundulus heteroclitus*. PLoS One 11, e0153104.
Rates, B., Bemquerer, M.P., Richardson, M., Borges, M.H., Morales, R.A.V., Lima, M.E.D., Pimenta, A.M.C., 2007. Venomic analyses of *Scolopendra viridicornis* and *Scolopendra angulata* (centipede, Scolopendromorpha): shedding light on venoms from a neglected group. Toxicon 49, 810–826.
Regier, J.C., Wilson, H.M., Shultz, J.W., 2005. Phylogenetic analysis of Myriapoda using three nuclear protein-coding genes. Mol. Phylogenet. Evol. 34, 147–158.
Rehm, P., Meusemann, K., Borner, J., Misof, B., Burmester, T., 2014. Phylogenetic position of myriapoda revealed by 454 transcriptome sequencing. Mol. Phylogenet. Evol. 77, 25–33.
Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European Molecular Biology Open Software Suite.
Rodríguez de la Vega, R.C., Giraud, T., 2016. Intragenome diversity of gene families encoding toxin-like proteins in venomous animals. Integr. Comp. Biol. 56, 938–949.
Rokyta, D.R., Lemmon, A.R., Margres, M.J., Aronow, K., 2012. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genom. 13, 312.
Rokyta, D.R., Margres, M.J., Calvin, K., 2015a. Post-transcriptional mechanisms contribute little to phenotypic variation in snake venoms. Genes Genomes Genet. 5, 2375–2382.
Rokyta, D.R., Ward, M.J., 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., Wray, K.P., McGivern, J.J., Margres, M.J., 2015b. The transcriptomic and proteomic basis for the evolution of a novel venom phenotype within the Timber Rattlesnake (*Crotalus horridus*). Toxicon 98, 34–48.
Rong, M., Yang, S., Wen, B., Mo, G., Kang, D., Liu, J., Lin, Z., Jiang, W., Li, B., Du, C., Yange, S., Jiang, H., Feng, Q., Xu, X., Wang, J., Lai, R., 2015. Peptidomics combined with cDNA library unravel the diversity of centipede venom. J. Proteom. 114, 28–37.
Say, T., 1821. Descriptions of the myriapodae of the United States. J. Acad. Nat. Sci. Phila. 2, 102–114.
Schaffrath, S., Prendini, L., Predel, R., 2018. Intraspecific venom variation in southern African scorpion species of the genera *Parabuthus, Uroplectes* and *Opistophthalmus* (scorpiones: buthidae, scorpionidae). Toxicon 144, 83–90.
Shelley, R.M., 1987. The Scolopendromorph Centipedes of North Carolina, with a Taxonomic Assessment of *Scolopocryptops gracilis Peregrinator* (Crabill)(Chilopoda: Scolopendromorpha). Florida Entomologist, pp. 498–512.
Shelley, R.M., 2002. A Synopsis of the North American Centipedes of the Order Scolopendromorpha (Chilopoda). Virginia Museum of Natural History.
Smith, J.J., Undheim, E.A., 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. https://doi.org/10.3390/toxins10030096.
Sunagar, K., Morgenstern, D., Reitzel, A.M., Moran, Y., 2016. Ecological venomics: how genomics, transcriptomics and proteomics can shed new light on the ecology and evolution of venom. J. Proteom. 135, 62–72.
Szczesny, P., Iacovache, I., Muszewska, A., Ginalski, K., Van Der Goot, F.G., Grynberg, M., 2011. Extending the aerolysin family: from bacteria to vertebrates. PLoS One 6, e20349.
Templ, M., Hron, K., Filzmoser, P., 2011. robCompositions: an R-package for Robust Statistical Analysis of Compositional Data. John Wiley and Sons.
Undheim, E.A., Hamilton, B.R., Kurniawan, N.D., Bowlay, G., Cribb, B.W., Merritt, D.J., Fry, B.G., King, G.F., Venter, D.J., 2015a. Production and packaging of a biological arsenal: evolution of centipede venoms under morphological constraint. Proc. Natl. Acad. Sci. U. S A. 112, 4026–4031.
Undheim, E.A., Jones, A., Clauser, K.R., Holland, J.W., Pineda, S.S., King, G.F., Fry, B.G., 2014. Clawing through evolution: toxin diversification and convergence in the ancient lineage Chilopoda (Centipedes). Mol. Biol. Evol. 31, 2124–2148.
Undheim, E.A.B., Fry, B.G., King, G.F., 2015b. Centipede venom: recent discoveries and current state of knowledge. Toxins 7, 679–704.
Undheim, E.A.B., King, G.F., 2011. On the venom system of centipede (Chilopoda), a neglected group of venomous animals. Toxicon 57, 512–524.
Vahtera, V., Edgecombe, G.D., Giribet, G., 2012. Evolution of blindness in scolopendromorph centipedes (Chilopoda: Scolopendromorpha): insight from an expanded sampling of molecular data. Cladistics 28, 4–20.
Vizcaíno, J.A., Csordas, A., del Toro N, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q.W., Wang, R., Hermjakob, H., 2016. 2016 update of the PRIDE database and related tools. Nucleic Acids Res. 44, D447–D456.
Wang, J., Wu, G., Chen, L., Zhang, W., 2015. Integrated analysis of transcriptomic and proteomic datasets reveals information on protein expressivity and factors affecting translational efficiency. In: Microarray Data Analysis. Springer, pp. 123–136.
Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63.
Ward, M.J., Ellsworth, S.A., Rokyta, D.R., 2018. 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.
Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T.-W., Li, Y., Xu, X., Wong, G.K.-S., Wang, J., 2014. SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. Bioinformatics 30, 1660–1666.
Yang, S., Liu, Z., Xiao, Y., Li, Y., Rong, M., Liang, S., Zhang, Z., Yu, H., King, G.F., Lai, R., 2012. Chemical punch packed in venoms makes centipedes excellent predators. Mol. Cell. Proteom. 11, 640–650.
Yang, S., Xiao, Y., Kang, D., Liu, J., Li, Y., Undheim, E.A.B., Klint, J.K., Rong, M., Lai, R., King, G.F., 2013. Discovery of a selective NaV1.7 inhibitor from centipede venom with analgesic efficacy exceeding morphine in rodent pain models. Proc. Natl. Acad. Sci. U. S. A. 110, 17534–17539.
Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30, 614–620.
Zhao, F., Lan, X., Li, T., Xiang, Y., Zhao, F., Zhang, Y., Lee, W., 2018. Proteotranscriptomic analysis and discovery of the profile and diversity of toxin-like proteins in centipede. Mol. Cell. Proteom. P. https://doi.org/10.1074/mcp.RA117.000431.