Sex-based venom variation in the eastern bark centipede (*Hemiscolopendra marginata*)

Gunnar S. Nystrom[a], Micaiah J. Ward[a], Schyler A. Ellsworth[a], and Darin R. Rokyta[a,*]

[a] Department of Biological Science, Florida State University, Tallahassee, Florida 32306
[*] Corresponding author: drokyta@bio.fsu.edu

Running title: Sex-based variation in *Hemiscolopendra marginata* venom

Keywords: centipede, venom, proteome, transcriptome

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

Abstract

Sexually dimorphic traits are widespread across metazoans and are often the result of sex-specific inheritance or sex-based differences in gene expression. Intersexual differences have even been observed in invertebrate venoms, although the identification of these differences has been limited to the more well-studied groups, such as scorpions and spiders, where sex-based differences in morphology and behavior are apparent. Recent studies on centipede venom have identified evidence of intraspecific variation, but intersexual differences have not been reported. To investigate the potential for sex-based differences in centipede venom composition, we performed reversed-phase high performance liquid chromatography (RP-HPLC) analyses on five male and 15 female eastern bark centipedes (*Hemiscolopendra marginata*) from the Apalachicola National Forest in northern Florida. After detecting a significant sex-based difference in *H. marginata* venom composition, we completed a high-throughput venom-gland transcriptomic and venom proteomic analysis of one male and one female to determine the genetic basis for differences in venom composition. We identified 47 proteomically confirmed toxins and 717 nontoxin transcripts in *H. marginata* venom-glands. Of these proteomically confirmed toxins, the most abundantly expressed in the male venom included ion channel-modulating toxins and toxins so divergent from any characterized homologs that they could not be given a functional classification, whereas the most abundantly expressed in the female venom were γ-glutamyl transferases and CAPs (cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins). These differences were then confirmed by performing replicate LC-MS/MS analyses on the venom from an additional three male and three female *H. marginata*. Our RP-HPLC and high-throughput transcriptomic and proteomic approach resulted in not only an in-depth characterization of *H. marginata* venom, but represents the first example of sex-based variation in centipede venoms.

1 Introduction

Sexually dimorphic traits have been observed in a variety of organisms across the animal kingdom (Glucksmann, 1974; Vollrath and Parker, 1992; Parker, 1992; Owens and Hartley, 1998). These traits include differences in color, size, secondary sexual characteristics, and behavior and are often the result of sex-based differences in gene expression (Ellegren and Parsch, 2007; Mank, 2008). As a biochemical phenotype with direct effects on fitness, venoms represent a unique model for studying the molecular basis of intersexual variation in organisms that experience sex-based differences in life history and behavior. Venoms are a complex mixture of peptides and proteins that exhibit a wide range of toxicity and physiological effects (Casewell et al., 2013). Intraspecific venom variation has been observed in the venom of multiple invertebrate lineages (Herzig et al., 2008; Abdel-Rahman et al., 2011; Undheim et al., 2014; Ward et al., 2018a), and some of this variation has been classified as sex-based. However, the identification of sex-based differences in venom has been limited to the more well-studied groups, such as spiders (Herzig et al., 2008; Binford et al., 2016) and scorpions (D'suze et al., 2015; Ward et al., 2018b). For example, using high performance liquid chromatography and LD_{50} assays in crickets, Herzig et al. (2008) identified sex-based differences in venom yield, complexity, and biological activity in mouse spider (*Missulena)* venoms. Additionally, a proteomic and functional venom comparison across three populations of the Hentz striped scorpion (*Centruroides hentzi*) revealed a significant variation in venom explained by a female-biased population divergence (Ward et al., 2018a).

Recent advancements in centipede (Arthropoda: Chilopoda) venom characterizations have shown that centipede venoms contain a rich diversity of toxins, many of which bear little resemblance to previously identified venom proteins, suggesting their novelty in evolutionary and pharmaceutical analyses (Undheim et al., 2014, 2015; Smith and Undheim, 2018). Toxins frequently identified in these centipede venom characterizations include antimicrobial peptides, β-pore forming toxins, CAPs (cysteine rich secretory [CRISP], antigen 5 [Ag5], and pathogenesis-related 1 [Pr-1]), γ-glutamyl transferases (GGTs), ion-channel modulating toxins, phospholipases, and proteases (Undheim et al., 2014, 2015; Hakim et al., 2015; Ward and Rokyta, 2018). Furthermore, in a previous comparison of *Scolopendra subspinipes* venom, Smith and Undheim (2018) suggest that centipede venoms are more variable than previously expected. High-throughput transcriptomic and proteomic analyses of *Scolopendra viridis* from Florida (Ward and Rokyta, 2018) and Mexico (González-Morales et al., 2014), for example, have recently presented evidence of intraspecific variation in centipede venom.

Although intraspecific variation has been reported in centipedes, sex-based variation in venom has not been examined. To investigate the potential for sex-based differences in centipede venom composition, we sampled multiple individuals of the eastern bark centipede (*Hemiscolopendra marginata*; Figure 1) from the Apalachicola National Forest in northern Florida. *Hemiscolopendra marginata* (Say, 1821), a member of the Scolopendridae family, is a relatively small species of centipede (13–57mm) that con-

sumes invertebrate prey. Although no distinct sexual dimorphism in morphology has been identified in *H. marginata*, male *Scolopendra morsitans* tend to be smaller than females and have a ridge that spans the lateral border of the last pair of legs (Lewis, 1968, 2006). Overall, few sexual dimorphic traits in scolopendromorph centipedes have been identified, but intersexual differences in behavior have been reported. Females of some centipede species, including *H. marginata*, participate in extensive maternal care (Cupul-Magaña et al., 2018). *Hemiscolopendra marginata* venom, although not harmful to humans, has yet to be studied. Therefore, we performed reversed-phase high performance liquid chromatography (RP-HPLC) analyses on multiple individuals of each sex and completed a venom-gland transcriptomic and venom proteomic analysis to identify the genetic basis for potential sex-based differences in *H. marginata* venom. To confirm potential sex-based differences, we also completed replicate proteomic analyses on the venom from three female and three male *H. marginata*.

2 Materials and Methods

2.1 Centipedes, venoms, and venom-glands

Hemiscolopendra marginata were collected from the Apalachicola National Forest from Liberty and Leon counties in northern Florida to assess intersexual differences in centipede venom variation (Figure 1). *Hemiscolopendra marginata* are distributed across much of the southeastern United States (Shelley, 2002) with their range extending into Mexico (Shelley, 2008). Both male and female centipedes were collected by peeling back bark from dead or fallen trees. Centipedes were maintained at Florida State University and sexed under a stereoscopic microscope by exposing internal genitalia. Males were identified based on the presence of the spinning organ behind the second genital sternite (Bonato et al., 2010; McMonigle, 2014). Venom proteome and venom-gland transcriptomic analyses were performed independently on one female and one male *H. marginata* labeled C0150 and C0162, respectively. We also performed replicate LC-MS/MS confirmation analyses on the venom from an additional three males and three females to independently confirm any sex-based differences suggested by the transcriptomics.

Venom and venom-gland extractions were performed as previously described (Ward and Rokyta, 2018). Centipedes were anesthetized with CO_2 for approximately one minute before being Velcro® fitted onto a venom extraction surface, ventral side up. Next, a serrated metal spatula was placed between the forcipules of the centipede and muscle contraction was induced by electrical stimulation at the base of the forcipules to promote ejection of the venom. The venom was lyophilized and stored at –80°C until further use. Four days after venom extraction, the venom-glands were removed under a stereoscopic microscope. Venom-gland tissue was immediately transferred to a 0.7 mL microcentrifuge tube containing 100 μL RNAlater and stored at 4°C overnight before transferring to –80°C for later use. Each centipede specimen was preserved in 95% ethanol and stored at –80°C.

2.2 Reversed-Phase high-performance liquid chromatography

Reversed-phase high-performance liquid chromatography (RP-HPLC) was performed on a single venom sample for the male and female transcriptome individuals and an additional 4 males and 14 females, as previously described by (Ward et al., 2018a). Approximately 7 μg of protein was injected onto a Aeris 3.6 μm C18 column (Phenomenex, Torrance, CA) utilizing 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 with a flow rate of 0.2 mL/min over a 125-minute gradient from 10-75% solution B, followed with a 15-minute wash of 10% B. To visualize and test for potential intraspecific differences in venom composition, we performed a principal component analysis (PCA) using the robCompositions package in R (Templ et al., 2011) and a permutational multivariate analysis of variance (PERMANOVA) on RP-HPLC profiles, using the adonis function from the vegan package in R (Oksanen et al., 2013).

2.3 Transcriptome sequencing

Venom-gland RNA extraction was performed as previously described (Rokyta and Ward, 2017; Ward et al., 2018b; Ward and Rokyta, 2018). First, 100 μL of RNAlater containing the venom-gland tissue was mixed with 500 μL of TriZol (Invitrogen) and homogenized with a sterile 20 gauge needle and syringe. An additional 500 μL of TriZol was added along with 20% choloroform before transferring the tissue mixture to 5Prime phase lock heavy gel tubes. The gel tubes filled with tissue mixture were then centrifuged before adding isopropyl alcohol to isolate the RNA. RNA pellets were washed with 75% ethanol, and purified RNA was washed with 70% ethanol before performing a Qubit RNA quantification (Thermo Fisher Scientific). An RNA quality check was then performed using the RNA 6000 Pico Bioanalyzer Kit (Agilent Technologies) according to manufacturer's instructions. RNA Integrity Numbers (RIN) are difficult to evaluate in invertebrates because the 18S and 28S rRNA fragments often co-migrate (Paszkiewicz et al., 2014). Therefore, RNA quality was evaluated by the existence and abundance of a double peak, corresponding to the vertebrate 18S rRNA peak.

The mRNA was then isolated from approximately 112 ng and 115 ng of total RNA from C0150 and C0162, respectively, using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs), according to manufacturer's instructions. mRNA was fragmented for 15.5 minutes to generate fragment sizes of approximately 370 nucleotides. Isolated mRNA was used to prepare cDNA libraries 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), and Agencourt AMPure XP PCR Purification Beads were utilized to for cDNA purification. A High Sensitivity DNA Bioanalyzer Kit (Agilent Technologies) was used to quality check purified cDNA libraries, according to manufacturer's instructions. Total cDNA yield for C0150 and C0162 were, 920 ng and 169 ng with an average fragment size of 318 and 417 bp, respectively.

Amplifiable concentrations for each sample were determined with KAPA PCR performed by the Florida State University Molecular Cloning Facility, resulting in 730.68 nM and 81.37 nM concentrations for individual C0150 and C0162, respectively. To maximize sequencing space, samples were diluted to ∼5 nM and pooled with additional 5 nM cDNA libraries from other species for sequencing on the same lane. The pooled cDNA library was quality checked with a High Sensitivity DNA Bioanalyzer Kit (Agilent Technologies), and the amplifiable concentration was confirmed with an additional round of KAPA PCR. Sequencing was performed by the Florida State University College of Medicine Translational Laboratory utilizing an Illumina HiSeq 2500.

2.4 Proteomics

Proteomic analyses were completed as previously described (Rokyta and Ward, 2017; Ward et al., 2018b; Ward and Rokyta, 2018). First, a Qubit Protein Assay Kit was used to quantify venom protein samples. Then, approximately 5 μg of whole venom was digested with the Calbiochem ProteoExtract All-in-One Trypsin Digestion Kit (Merch, Darmstadt, Germany), according to manufacturer's instructions. Digested venom proteins were then dried with a SpeedVac.

LC-MS/MS was performed by the Florida State University College of Medicine Translational lab as previously described (Rokyta and Ward, 2017; Ward and Rokyta, 2018). Briefly, digested venom protein was resuspended in 0.1% formic acid to a final concentration of 250 ng/μL. Three highly-purified recombinant *Escherichia coli* proteins (Abcam) of known concentration were mixed in specific proportions before digestion to yield final desired concentrations of 2,500 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* protein mix was infused into venom samples prior to LC-MS/MS injection. Approximately 2 μL of each sample was injected into an externally calibrated Thermo Q Exactive HF (high-resolution electrospray tandem mass spectrometer) in conjunction with the Dionex UltiMate3000 RSLCnano System to perform the LC-MS/MS analysis. Starting with LC, all 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.$\times$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 minute linear gradient from 3% to 45% B was implemented 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 nanosprayed into a Q Exactive HF mass spectrometer (Thermo Scientific), which was operated in a data-dependent mode under direct control of the Thermo Excalibur 3.1.66 (Thermo Scientific) throughout chromatographic separation. A data-dependent top-20 method was utilized for the acquisition of MS data, selecting the most abundant, not-yet-sequenced precursor ions from the survey scans (350–1700 m/z). Sequencing was performed with higher energy collisional dissociation fragmentation using a target value of 10^5 ions determined with predictive automatic gain control.

Full scans (350–1700 m/z) were executed at 60,000 resolution in profile mode and MS2 were acquired in centroid mode at 15,000 resolution, as previously described (Rokyta and Ward, 2017; Ward et al., 2018b; Ward and Rokyta, 2018). A dynamic exclusion window of approximately 15 seconds was used and ions with either a single charge, a charge of more than seven, or an unassigned charge were excluded. Measurements were recorded at room temperature with each sample being run and measured in triplicate to promote lable-free quantification and account for machine-related variability between samples. 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. SequestHT search parameters were as follows: 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). Scaffold (version 4.3.4, Proteome Software Inc., Portland, OR, USA) software was used to validate protein and peptide identities. Peptide identities were accepted based on a 1.0% false discovery rate (FDR) using the Scaffold Local FDR algorithm and protein identities were accepted with an FDR of 1.0% and a minimum of one recognized peptide. Proteomic abundance estimates for each individual (C0150 and C0162) were calculated as previously described by Rokyta and Ward (2017) and Ward and Rokyta (2018). Conversion factors were used to convert normalized spectral counts for each venom protein in all replicates to a concentration value. Final concentrations of each sample were averaged across the three replicates for individuals C0150 and C0162.

Replicate LC-MS/MS confirmation analyses for the three additional male (C0537, C0555, C0556) and three female (C0468, C0539, C0553) *H. marginata* venom samples were completed after transcriptome assembly and proteomic annotation of the initial male and female individuals. Approximately 5 μg of whole venom was subject to an in-solution tryspin digestion. In brief, 30 μl of 10 mM DTT was added to each sample. After 10 minutes, the samples were incubated for 60 minutes at 60°C. Next, 30 μl of 50 mM Iodoacetoamide was added to the solution and the samples were incubated at room temperature for 30 minutes before adding 150 μl of 50 mM ammonium bicarbonate. A 0.5 μg trypsin aliquot was then added to each sample to commence digestion. After 18 hours, digestion was terminated using a 1% TFA (in water) solution. Digested venom proteins were then dried with a SpeedVac. LC-MS/MS analysis was done as at the Florida State University College of Medicine Translation Lab as previously described. The resulting raw files were searched with Proteome Discover 2.2 using the consensus transcriptome generated from individuals C0150 and C0162.

2.5 Transcriptome assembly and analysis

Transcriptome assembly and analysis was performed as previously described (Rokyta and Ward, 2017; Ward et al., 2018b; Ward and Rokyta, 2018). Illumina quality filtering was implemented on the transcriptome sequencing data to generate filtered read-pairs

for both individuals. Since we performed 150 paired-end sequencing with a target insert size of 250 nucleotides, we expected most read-pairs to exhibit significant 3' overlap and therefore used PEAR version 0.9.6 (Zhang et al., 2014) to merge reads for further analyses. DNAStar NGen version 12.3.1 with 10 million merged reads and default transcriptome assembly settings was used to generate our primary transcriptome assembly, and only contigs with at least 200 reads were retained. Because we did not expect to find many known homologs of toxins for this species in public databases, multiple search methods were used to identify and annotate proteins in the transcriptome.

Of these methods, two used the whole-venom mass-spectrometry results and the generated protein databases by applying TransDecoder version 2.0.1 (Haas and Papanicolaou, 2016) to the assembled transcriptomes. First a database was generated using the TransDecoder-predicted protein sequences with a minimum length of 50 and the mass-spectrometry results were searched against this database. The results were then filtered using Scaffold Viewer version 4.6.0. To accommodate for 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. For the second strategy, we wanted to ensure that small peptides were not missed by the TransDecoder predictions, so we created an additional database using all possible protein or peptide sequences of at least 50 amino-acids from each of the six potential reading frames. Results were then filtered in Scaffold, and contigs annotated from the previous strategy were removed. Our third method aimed to identify proteins from the transcriptome that displayed homology to known toxins. For this strategy, we peformed 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 annotated full-length putative toxins which showed a match with at least 80% of a known toxin's length. Next, we used a fourth strategy in which we executed a BLASTX search 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 create a general database of all of the toxin and nontoxin transcripts expressed in the venom-glands. We only accepted transcripts with at least 500 reads that matched to at least 95% of the length of a known protein. Finally, for our fifth method, we utilized Extender (Rokyta et al., 2012) to perform a transcriptome assembly from 1,000 random reads to confirm that no high-abundance transcripts were excluded. We only accepted reads with phred qualities of $\geq$30 at all positions that had an exact match of 120 nucleotides for extension. A BLASTX search with the resulting contigs was performed against the UniProt animal toxins database. A final consensus transcriptome was created for this first set of annotated sequences by combining transcripts by individual through clustering based on coding sequences with cd-hit-est version 4.6 (Li and Godzik, 2006) and a sequence identity threshold of 1.0. Then, to screen for chimeras, we aligned the merged reads against the resulting combined set with bwa version 0.7.12 (Li, 2013), and allowed only reads without mismatches relative to the reference. We checked the resulting alignments for regions with no coverage or multimodal coverage distributions. Because we sequenced *H. marginata* RNA-seq libraries in conjuction with libraries

from other species, we checked for cross-contamination between samples by aligning the PEAR-merged reads of each of the other sequenced samples against our transcript set for each individual with bwa version 0.7.12 (Li, 2013), retaining mapped reads with less than four mismatches. We removed transcripts as contaminants if they showed >100× higher coverage for a different library relative to the highest-coverage of comparable *H. marginata* library, had read coverage over the whole coding sequence, and had an absence of homozygous variants relative to the consensus sequence.

As proteomically confirmed centipede venom proteins are scarce in public databases, we included proteomic-driven annotation for six different transcriptome assemblies per individual. First, we processed the raw reads. Then, we screened for and removed any sample cross-leakage attributable to the demuliplexing step by comparing the k-mer distributions for each *H. marginata* sample against each of the other samples that were sequenced in the same lane. We generated 57-mer distributions with jellyfish version 2.2.6 (Marçais and Kingsford, 2011) and identified 57-mers which had a 500× higher abundance in a different sample compared to the sample of interest. Reads for the sample of interest were excluded if ≥25% of their length included these 57-mers. Next, we performed adaptor and quality trimming with Trim Galore! (Krueger, 2015). We used a quality threshold with a phred of 5, and excluded any trimmed reads that were fewer than 75 nucleotides. The reads were then merged with PEAR version 0.9.10 (Zhang et al., 2014) using default settings. For our first transcriptome assembly method, we used Extender (Rokyta et al., 2012) with 1,000 random seeds with a minimum phred of 30, an overlap of 120 nucleotides, 20 replicates, and used only the merged reads that had a minimum phred of 20. We then ran BinPacker version 1.0 (Liu et al., 2016) with a k-mer size of 31, used merged and unmerged reads, and considered all reads as unpaired. We also ran Trinity version 2.4.0 (Grabherr et al., 2011) with a k-mer size of 31 using both merged and unmerged reads, and considered all reads as unpaired. Next, using both merged and umerged reads, we ran SOAPdenovo-trans version 1.03 (Xie et al., 2014) with a k-mer size of 127 but considered unmerged reads as paired. We then ran SeqMan NGen version 14.0 with both merged and unmerged reads, and considered all of the reads unpaired. Finally, we ran rnaSPAdes version 3.10.1 (Bankevich et al., 2012) with $k = 127$, and used both merged and unmerged reads with unmerged reads considered as paired. We used the getorf function from EMBOSS version 6.6.0.0 (Rice et al., 2000) for each assembly using a nucleotide minimum size of 90 and extracted the amino-acid sequences of open reading frames, keeping only those with stop and start codons. To remove exact duplicates within assemblies, we clustered the output of each assembly with cd-hit version 4.6 (Li and Godzik, 2006) setting the sequence identity threshold to 1.0. We then used each generated data set as a database to search our LC/MS/MS results against (see above). With this second set of proteomic-based identifications we then worked to create our final consensus transcriptome. First, we combined each set of putative toxins individually by clustering based on coding sequences with cd-hit-est version 4.6 (Li and Godzik, 2006) and a sequence identity threshold equal to 1.0. To screen for chimeras we aligned the merged reads against the combined set using bwa version 0.7.12 (Li, 2013),

including exact matches only. 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 combined the transcripts for each individual from the first and second (MS-based approach) annotation approaches using cd-hit-est and a sequence identity threshold equal to 0.98 to generate our final consensus transcriptome. We estimated transcript abundances from bowtie2 (Langmead and Salzberg, 2012) version 2.3.0 alignments against coding sequences of the final transcriptome, using RSEM (Li et al., 2011) version 1.2.31. We based alignments on all merged reads for each individual. We then used the centered logratio transformation (Aitchison, 1986) on our transcriptome and proteome abundances as described by Rokyta et al. (2015a). We verified the presence of signal peptides with SignalP version 4.1 under default settings (Petersen et al., 2011).

2.6 Data availability

Raw transcriptome reads have been submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) using the BioProject number PRJNA340270 with BioSamples SAMN10423645 (C0150) and SAMN10423646 (C0162), and SRA accession numbers SRR8188013 and SRR8188014 (C0150) and SRR8188011 and SRR8188012 (C0162). Data from the mass spectrometry proteomics analyses were deposited to the ProteomeXchange Consortium using the PRIDE (Vizcaíno et al., 2016) partner repository and PXD011712 as the dataset identifier. All the assembled transcripts were submitted to the NCBI Transcriptome Shotgun Assembly (TSA) database. Data from this TSA project were deposited at DDBJ/EMBL/GenBank using the accession number GHBY00000000. The version detailed in this paper represents the first version, GHBY01000000.

3 Results and Discussion

3.1 Sex-based variation in *H. marginata* venom composition

To assess sex-based variation in *H. marginata* venom composition, we performed RP-HPLC analyses on the venom of our two transcriptome individuals and an additional four males and 14 females collected from the Apalachicola National Forest. Although all centipedes were collected using the same methods, we observed a female collecting bias, which may be the result of intersexual differences in behavior or habitat preference. Approximately 14 distinct peak clusters between 20 and 120 minutes were observed on the RP-HPLC profiles (Figure 2). We observed intrasexual variation in RP-HPLC peak abundance and found a significant difference in venom composition between the two sexes (PERMANOVA, $p < 0.001$). Using a variance matrix we determined that the most variable peaks were peaks 2, 10, 11, 12, and 14, which contributed 10.9%, 8.7%, 14.3%, 13.2%, and 20.8% of the variation in venom, respectively. Furthermore,

our PCA analysis showed a distinct separation between male and female *H. marginata* venom composition, with the two most variable principle components accounting for approximately 67% of this variation (Figure 3). As in the variance matrix, the most variable peaks in this analysis included peaks 2, 11, 12, and 14, however, peak 7 instead of peak 10 was the fifth most variable peak. Peaks 2, 10, 12, and 14 were observed at a higher relative abundance in the male venom, while peaks 7 and 11 displayed a higher relative abundance in the female venom (Figure 2). Peaks 2, 7, 10, 11, 12, and 14 were eluted at approximately 32, 58, 76, 81, 87, and 103 minutes, respectively (Figure 2).

3.2 The genetic basis for *H. marginata* venom

To characterize venom expression differences between male and female *H. marginata*, we performed a venom-gland transcriptomic and venom proteomic analysis on one female (C0150) and one male (C0162) *H. marginata*. For the female individual, we generated 14,825,899 raw read pairs after Illumina quality filtering, 12,645,139 of which were merged. We assembled ten million merged reads into a primary female transcriptome of 3,848 contigs using NGen. During our mass spectrometry directed analysis, we annotated 36 unique coding sequences with TransDecoder using all possible open reading frames (ORFs). Using BLASTX hits against the Uniprot toxins database we annotated 50 unique coding sequences. We also annotated 177 unique coding sequences using the BLASTX hits against the NCBI protein database. With the Extender assembly, we annotated an additional seven unique coding sequences by completing a BLASTX search of the UniProt animal toxins database. Removing chimeras and duplicates, we are able to identify a combined total of 498 unique coding sequences for the female individual.

For the male individual, we generated 19,285,984 raw read pairs after Illumina quality filtering. Of these raw read pairs, 15,538,169 were merged. We then used Ngen to assemble ten million merged reads into our primary male transcriptome of 4118 contigs. For our mass spectrometry directed analysis, we annotated 39 unique coding sequences with TransDecoder using all possible ORFs. We then annotated 61 unique coding sequences using BLASTX hits against the UniProt toxins database. After removing duplicates, we annotated 203 unique coding sequences using BLASTX hits against the NCBI nr database. With the Extender assembly, we annotated an additional five unique coding sequences by executing a BLASTX search of the UniProt animal toxins database. With chimeras and duplicates removed, we identified a combined total of 512 unique coding sequences for the male individual.

Our multiple assembly approach resulted in six proteomic driven annotated assemblies per individual. As described above, we processed, filtered, and merged raw reads for each of the six assemblies. For Extender, we annotated only the merged reads, resulting in 20 unique coding sequences for the female and 30 unique coding sequences for the male. Using both merged and unmerged reads (with all reads treated as unpaired) for the BinPacker, SeqMan NGen, and Trinity assemblies, we annotated 36 (C0150) and 33 (C0162), 43 (C0150) and 30 (C0162), and 36 (C0150) and 37 (C0162) unique coding se-

quences. Using merged and unmerged reads, with the unmerged reads treated as paired, we annotated 32 (C0150) and 29 (C0162) unique coding sequences using rnaSPAdes, and 18 (C0150) and 16 (C0162) unique coding sequences using SOAPdenovo-trans. From our six different assemblies, our proteomic-driven annotation resulted in a total of 65 and 57 unique coding sequences for the female and male individuals, respectively.

We combined our annotated transcripts from each individual using both our primary and proteomic-driven annotation strategies. After removing duplicates, we produced a final consensus transcriptome of 764 unique protein-coding transcripts. We used this final consensus transcriptome for the subsequent transcript-abundance estimates and LC-MS/MS analyses for each individual *H. marginata*. We divided transcripts into two classes: toxins and nontoxins. A protein that was proteomically confirmed in the venom of one or both individuals, and thus had a high likelihood of translating into a protein with toxic function, was classified as a toxin. Nontoxin transcripts, however, were classified as proteins that were not detected in the venom proteome of either individual. Although some nontoxins shared homology with toxin-like transcripts from other centipedes in the UniProt animal toxins database, these were not classified as toxins in our analysis because we did not detect them in the venom proteome. Small peptides, for example, might not have been detected due to the extensive proteolytic processing required for the development of their mature peptides (Rokyta and Ward, 2017). We identified 47 toxin transcripts that were found in the venom proteome of at least one of the two *H. marginata* individuals (Table 1), which accounted for approximately 216,915.04 and 389,013.66 transcripts per million (TPM) of the mapped reads in the female and male individual, respectively. The nontoxin transcripts, of which 717 were identified, made up 783,084.96 TPM and 610,986.43 TPM of the total mapped reads in the male and female, respectively. These nontoxins, although likely to encode proteins that regulate essential cellular function and the production of proteins, are unlikely to encode for proteins with high toxicities.

3.3 Transcript and protein abundances across individuals

A comparison of venom-gland transcript abundance for nontoxins showed a strong correlation between individuals (Spearman's rank correlation $\rho = 0.82$, Pearson's rank correlation coefficient $R = 0.77$, and $R^2 = 0.60$; Figure 4). Transcript abundances for toxin-encoding proteins did not show a correlation (Spearman's rank correlation $\rho = 0.05$, Pearson's rank correlation coefficient $R = -0.14$, and $R^2 = 0.02$; Figure 4). Of these toxin-encoding proteins, we observed five outliers including four γ-glutamyl transferases (GGT-1–GGT-4) and one scoloptoxin (SLPTX15-1), all of which will be discussed in the following sections. A toxin-encoding protein was deemed an outlier if it lied outside the 99^{th} percentile of differences between the two toxin measures. Although four of the five outliers (GGT-1–GGT-4) were only identified in the venom and venom-gland transcriptome of the female, SLPTX15-1 was found in the venom of both individuals, but was not detected in the transcriptome of the male (Figure 4; Table 1). To account for

these presence/absence differences, we performed an additional comparison with outliers excluded. In this comparison, we observed a weak positive correlation in toxin transcript abundance across individuals (Spearman's rank correlation $\rho = 0.26$, Pearson's rank correlation coefficient $R = 0.34$, and $R^2 = 0.12$), indicating that differences in toxin transcript abundances between the two individuals could be influenced, but not completely explained, by these presence/absence differences.

A venom proteomic comparison between the two *H. marginata* individuals did not show a strong correlation (Spearman's rank correlation $\rho = 0.43$, Pearson's rank correlation coefficient $R = 0.37$, and $R^2 = 0.14$; Figure 5). Of the 47 proteomically confirmed toxins, 17 were expressed in both individuals (Figure 5). Of the 30 toxins not observed in the venom proteome of both individuals, 11 of these were only detected in the female and 19 were only detected in the male (Table 2). Overall, male and female *H. marginata* displayed substantial proteomic expression differences as predicted from our RP-HPLC analyses.

3.4 Cysteine-rich secretory proteins (CRISPs), antigen 5 (Ag5), and pathogenesis-related 1 (Pr-1) proteins

CAP proteins have been been recruited into the venom of a variety of different animal species, including snakes, the platypus, and sea anemones, (Peichoto et al., 2009; Wong et al., 2012; Moran et al., 2013) and have been described as one of the most abundant protein families in Scolopendrid venoms (Liu et al., 2012; Undheim et al., 2014). Phylogenetic analyses have separated centipede CAP proteins (CAPs) into three distinct classes (CAP1, CAP2, and CAP3) resulting from separate recruitment events (Undheim et al., 2014). Although CAP proteins made up only 18.8% and 2.1% of the transcriptional output for the female and male, respectively, they represented 43.0% and 14.3% of the respective venom proteomes (Figure 6). However, after analyzing our male and female *H. marginata* replicate LC-MS/MS confirmation analyses, we did not observe any significant sex-based difference in CAP proteomic abundances ($p > 0.05$), suggesting that the observed differences may have been the result of between-individual variation.

We identified seven members of the CAP2 class and one CAP3 in the venom of *H. marginata* (Table 1). CAP2s are the most dominant CAP proteins in scolopendrid venoms and have been described as having voltage-gated calcium channel and trypsin inhibitor activities (Undheim et al., 2014, 2015). All of the CAP2 proteins identified in the venom of *H. marginata* contained signal peptides of 15–27 amino acids, 5–7 cysteine residues and had molecular weights ranging between 20.3–22.2 kDA. Each CAP2s in *H. marginata* venom, except CAP2-6, shared sequence identity with a CAP2 from *Cormocephalus westwoodi* (Undheim et al., 2014). CAP2-6 shared 50% sequence identity with a CAP2 from *Scolopendra morsitans* (Undheim et al., 2014).

CAP3 toxins have only been reported from *S. mortisans* venom, and the function of this protein has yet to be characterized (Undheim et al., 2014, 2015). We identified one CAP3 that showed 59% sequence similarity to a CAP3 from *S. morsitans* (Undheim

et al., 2014), but this CAP was only observed in the male venom (Table 1). CAP3-1 contained a 17 amino acid long signal peptide, 36 cysteine residues, and had a molecular weight of 57.4 kDa. As CAP3-1 has a large molecular weight, it is likely that it contributed to the abundance differences seen in peaks 10, 12, or 14. However, CAP3-1 only made up 1.4% of the male venom proteomic output, making it unlikely that it plays a major role in overall venom function.

3.5 γ-glutamyl transferases

Members of the γ-glutamyl transferase (GGT) protein family represent a widespread group of enzymes found in organisms ranging from bacteria to plants and animals (Whitfield, 2001). GGTs play an important role in the detoxification of foreign compounds, the maintenance of intracellular homeostasis, and the transport of amino acids (Courtay et al., 1992; Whitfield, 2001). GGTs are frequently reported from centipede venoms (Liu et al., 2012; Undheim et al., 2014), and GGTs from the venom of *Scolopendra subspinipes dehaani* have been shown to promote human platelet aggregation and the hemolysis of mice and rabbit red blood cells. (Liu et al., 2012). We identified four transcripts encoding for GGTs in the venom of the female *H. marginata*, but did not identify any GGTs in the venom-gland transcriptome or venom proteome of the male (Table 1, Table 2). Our replicate LC-MS/MS confirmation analyses revealed the presence of GGTs in male *H. marginata* venom (Table 6), but we detected a significantly lower GGT proteomic abundance in male venom ($p < 0.01$). All of the GGTs in *H. marginata* venom showed 74–75% sequence identity to a GGT from the venom of *C. westwoodi* (Undheim et al., 2014). Each GGT described here contained a signal peptide of 20 amino acids in length, four cysteine residues, and a had molecular weight ranging from 60.3–60.4 kDa.

GGTs made up 14.1% and 42.5% of the total toxin transcriptome and venom proteome outputs for the female venom, respectively (Figure 6). Due to their prevalent expression in only the female venom and large molecular weight, it is likely that the GGTs contribute to the high relative abundance observed in peak 11 (Figure 2). Furthermore, since these proteins make up a large portion of the proteomic output, they likely make a significant contribution to female venom function. Although GGTs from centipede venoms have shown hemolytic activity on mammalian red blood cells (Liu et al., 2012), this likely does not represent the primary function of GGTs in *H. marginata* venom as *H. marginata's* small body size presumably restricts it from consuming larger vertebrate prey. Female centipedes, including *H. marginata*, often provide extensive parental care in which the maternal individual will wrap her body around the brood until the offspring have completed their first molt (Cupul-Magaña et al., 2018). Although this behavior is essential to increasing survival in offspring, it is not known what role, if any, venom may play in maternal care. In the invasive garden ant, *Lasius neglectus*, adult workers secrete venom onto the brood while grooming to provide a chemical defense against fungal infections (Tragust et al., 2013). Female *H. marginata* could be using a similar grooming technique during maternal care, but further experiments aimed at understanding cen-

tipede grooming and the antimicrobial properties of GGTs from centipede venom would be needed to confirm this function. Previously, GGTs have also been identified in the parasitoid wasp, *Aphidius ervi*, where they induce apoptosis in aphid host ovaries (Falabella et al., 2007). Falabella et al. (2007) suggested that GGTs in *A. ervi* may alter glutathione metabolism, resulting in increased levels of oxidative stress in prey. The GGTs observed in *H. marginata* venom may serve a similar function aimed at disrupting prey homeostasis. However, because *H. marginata* lacks distinct sexual dimorphism in size and appearance and information regarding its feeding ecology and behavior is scarce, the exact role that GGTs play in female venom is unclear. Furthermore, Falabella et al. (2007) observed that the GGT found in *A. ervi* was transcribed at higher levels in females than males, which follows the pattern seen in *H. marginata* GGT expression. Thus, the differential expression of GGTs in the female *H. marginata* transcriptome and venom proteome are likely the result of sex-based differences in transcriptional and translational regulatory mechanisms.

3.6 Scoloptoxins

Although ion channel-modulating toxins have been well characterized from a variety of venomous species (Kalia et al., 2015), ion channel modulators have only recently been described in centipede venoms (Yang et al., 2012). Scoloptoxins (SLPTXs), represented by 31 distinct families, are one of the more diverse and abundant toxin classes in centipede venoms (Undheim et al., 2014; Smith and Undheim, 2018). These toxins, consisting of 3–18 cysteine residues and molecular weights ranging from 3–24 kDa, act primarily as modulators of calcium, potassium, and sodium channels (Yang et al., 2012; Undheim et al., 2014; Smith and Undheim, 2018). We identified toxins from five SLPTX families (SLTPX4, SLPTX10, SLPTX11, SLPTX15, SLPTX16) in the venom-gland transcriptome and venom proteome of *H. marginata* (Table 1).

We identified two members of the SLPTX4 family (Table 1) in the venom of *H. marginata*, although these proteins were only proteomically confirmed in the male venom. Toxins in the SLPTX4 family have been shown to act as inhibitors of voltage gated potassium channels (Undheim et al., 2015). Both SLPTX4 toxins identified had four cysteine residues and contained signal peptides of 17 (SLPTX4-1) and 25 (SLPTX4-3) amino acids. SLPTX4-1 shared approximately 56% sequence identity with a SLPTX4 reported from *Ethmostigmus rubripes*, (Undheim et al., 2014) and SLPTX4-3 shared 34% sequence identity with a SLPTX4 found in *S. subspinipes* (Smith and Undheim, 2018).

Members of the SLPTX10 family are characterized as potassium and sodium channel inhibitors (Liu et al., 2012; Undheim et al., 2014, 2015). We identified four members of the SLPTX10 family (Table 1), each of which had six cysteine residues and a signal peptide of 22–24 amino acids in length. Three of these proteins (SLPTX10-1, SLPTX10-2, and SLPTX10-3) shared sequence identity with toxins from *C. westwoodi* (Undheim et al., 2014), and one (SLPTX10-5) shared 51% sequence similarity to U-SLPTX10-Sm2b from *S. morsitans* (Undheim et al., 2014) and had a molecular weight of 8.6 kDa.

The SLPTX11 family, which range in size from 6.7–25.6 kDa and contain 6–19 cysteine residues have been described as having potassium channel inhibiting and anticoagulant activity (Liu et al., 2012; Undheim et al., 2014). We identified five members of the SLPTX11 family (Table 1) in *H. marginata* venom, which all contained 16–17 cysteine residues, a signal peptide of 18–22 amino acids long, and had molecular weights ranging from 21.6–25.1 kDa. Although we identified SLPTX11 expression in the transcriptome of both individuals, SLPTX11s were only present in the venom proteome of the male individual. SLPTX11-3 showed 27% sequence similarity to an SLPTX11 from *S. morsitans* (Undheim et al., 2014). SLPTX11-5, SLPTX11-6, SLPTX11-9, and SLPTX11-10 all shared sequence identity to a protein detected in *S. subspinipes* (Smith and Undheim, 2018).

Toxins in the SLPTX15 family have been characterized as calcium, potassium, and sodium channel antagonists (Liu et al., 2012; Luo et al., 2018). We identified two members of the SLPTX15 family in the venom proteome of *H. marginata* (Table 1). SLPTX15s contained a signal peptide of 23 amino acids, four cysteine residues, and had molecular weights ranging from 5.7–6.1 kDa. SLPTX15-1 shared 61% sequence identity to an SLPTX15 from *C. westwoodi* (Undheim et al., 2014). SLPTX15-3 had 49% sequence similarity to a toxin identified from *S. s. dehaani* (Unpublished data, Rehm et al.)

We identified only one member of the SLPTX16 family (Table 1) in the venom of *H. marginata*, which was identified in the male venom, but absent from the female venom (Table 2). Members of the SLPTX16 family typically contain 3–9 cysteine residues, and range in size from 7.4–13.6 kDa (Undheim et al., 2015). However, the function of toxins within this family has yet to be characterized. With a signal peptide of 25 amino acids, eight cysteine residues, and a molecular weight of 10.2 kDa, SLPTX16-1 has the characteristics expected from a member of the SLPTX16 family. SLPTX16-1 showed approximately 74% sequence similarity to a toxin from *E. rubripes* (Undheim et al., 2014).

Overall, SLPTXs contributed 37.1% and 20.0% of the total toxin transcriptional output and 4.3% and 26.1% of the proteomic output for the female and male, respectively (Figure 6). Our replicate LC-MS/MS confirmation analyses also revealed a significantly higher expression of SLPTXs in the male venom ($p = 0.013$). As SLPTXs displayed a high abundance in male venom, these proteins likely play a major role in venom function. This difference may, in part, be explained by intersexual differences in behavior. For example, in some species of centipedes, males and females show spatial segregation, potentially the result of territoriality or cannabalism of males by females (Dugon and Arthur, 2012). Furthermore, some species of scorpion exhibit significant differences in behavior between males and females (Booncham et al., 2007; Carlson and Rowe, 2009; Carlson et al., 2014), with males having higher rates of dispersal as they search for mates. As the male scorpions are more mobile, they likely experience different selective pressures compared to their female conspecifics (*e.g.* increased predator exposure, male-male competition).

3.7 Other toxins

We identified two neprilysins, or metallo-endopeptidases, in the venom of *H. marginata* (Neprilysin-1 and Neprilysin-12), both of which contained an M13 peptidase domain. These toxins made up less than a percent of the toxin transcriptional output in both individuals and were only identified in the venom proteome of the male (2.5%; Figure 6). Neprilysin-1 and Neprilysin-12 contained signal peptides of 29–31 amino acids long, 10–11 cysteine residues, and molecular weights of 79.3 and 80 kDa, respectively. Neprilysin-1 and Neprilysin-12 showed sequence similarity to toxins from *S. subspinipes* (Smith and Undheim, 2018). Due to their large molecular weight, Neprilysins could be contributing to the abundance differences seen in peaks 12 and 14. However, as these proteins make up a relatively small portion of the male proteome, they likely do not play a large role in overall venom function.

Finally, we identified 18 proteins in *H. marginata* venom for which we were unable to provide a functional classification. Of these proteins, one contained a domain of unknown function (DUF3472-1). DUF3472-1 had an 18 amino acid signal peptide, a molecular weight of 46.3 kDa, and shared 46% with a DUF from *E. rubripes (Undheim et al., 2014)*. DUF3472-1 was only identified in the female venom. We classified the remaining 17 proteins as uncharacterized venom proteins (VPs), which made up 26.3% and 77.0% of the toxin transcriptional output and 8.0% and 57.0% of the proteomic output for the female and male, respectively. Our replicate LC-MS/MS confirmation analyses also revealed a significantly higher expression of VPs in the male venom compared to the female venom ($p<0.01$). Due to the high prevalence in the male venom as opposed to the female venom and intermediate to larger molecular weights, VPs likely contribute to the higher relative abundances seen in the male RP-HPLC venom profiles (*e.g* peaks 10, 12, and 14) and play a major role in overall male venom function. Five of these proteins (VP-7, VP-12, VP-13, VP-16, and VP-18) did not match to any centipede proteins in the TSA database, each of which had a signal peptide of 18–26 amino acids, and a molecular weight of 6.6–40.7 kDa. VP-21 contained 19 amino acid long signal peptide, 4 cysteine residues, had a molecular weight of 18.2 kDa, and shared 34% sequence identity with an uncharacterized protein from *S. s. dehaani* found in the TSA database (Rehm et al., 2014). Of the remaining VPs, nine (VP-4, VP-5, VP-14, VP-15, VP-19, VP-20, VP-23, VP-24, VP-26) showed 29–43% sequence similarity with an undescribed protein from *S. s dehaani* found in the TSA database (Rehm et al., 2014) and two (VP-10 and VP-17) shared 40% sequence identity with a different, uncharacterized protein from *S. s dehaani* found in the TSA database (Rehm et al., 2014). VP-7 and VP-13 shared 91% sequence identity, while VP-4 and VP-23 shared 97% sequence identity. Finally, a group of VPs (VP-10, VP-17, VP-20, and VP-26) all shared at least 90% sequence identity.

3.8 Transcript versus protein abundance estimates

For our comparison of transcript versus protein abundance within individuals, we did not find strong correlations for either the male (Spearman's rank correlation $\rho = 0.23$,

Pearson's rank correlation coefficient $R = 0.34$, and $R^2 = 0.11$; Figure 7) or the female (Spearman's rank correlation $\rho = -0.01$, Pearson's rank correlation coefficient $R = -0.03$, and $R^2 = 0$; Figure 7). The CAP2 and GGT toxin classes demonstrated higher expression in the transcriptome relative to the venom proteome, whereas the SLPTX10 toxin class showed the opposite phenomenon, displaying a higher protein than transcript expression (Figure 7). Expression differences between the transcriptome and venom proteome could be the result of mapping biases (Wang et al., 2009; Peng et al., 2012; Rokyta et al., 2012) or post-translational modifications (Rokyta et al., 2015b; Ward and Rokyta, 2018). Additionally, previous transcriptomic characterizations in scorpions and centipedes have shown similar discrepancies between protein and venom-gland mRNA expression (Nisani et al., 2012; Rokyta and Ward, 2017; Ward et al., 2018b; Ward and Rokyta, 2018). Asynchronous regeneration profiles have also been observed in the venom of centipedes (Cooper et al., 2014) and scorpions (Nisani et al., 2012; Carcamo-Noriega et al., 2019), suggesting temporal expression differences in the regeneration of different venom components. Furthermore, a distinct venom heterogeneity has been identified in some scorpion species (Inceoglu et al., 2003), where venom is often secreted as a continuum of differing composition, providing further evidence for potential differences in invertebrate venom regeneration profiles. Although it has been shown that a post extraction time of four days is optimal to achieve maximum transcription levels in snakes venom-glands (Rotenberg et al., 1971), this may not be the case for invertebrates such as scorpions and centipedes. Ward and Rokyta (2018) argued that these discrepancies could be alleviated by optimizing venom proteomic and venom-gland transcriptomic studies to include time sensitive transcriptomic and proteomic analyses. If there is a temporal expression discrepancy between venom transcript and protein expression, toxins with higher abundance in the transcriptome (*e.g.* SLPTXs) could be upregulated later in the venom regeneration process and toxins with higher abundance in the venom proteome (*e.g.* GGTs and CAPs) could be upregulated earlier.

4 Conclusions

Using RP-HPLC analyses, we observed a significant difference in venom composition between male and female *H. marginata*, representing the first case of sex-based variation in centipede venoms. To further characterize these venom expression differences, we performed a high-throughput venom proteomic and venom-gland transcriptomic analysis of the venom from one male and one female *H. marginata*. We identified a total of 47 proteomically confirmed toxins, including proteins commonly detected in centipede venoms (*e.g.* CAPs, GGTs, DUFS, SLPTXs). We also identified two neprilysins and 17 proteins that we could not provide a functional classification for (VPs). We found discrepancies between transcript and protein abundances similiar to those seen in previous invertebrate venom characterizations, reinforcing the need for time sensitive transcriptomic and proteomic analyses or further investigations into post-translational regulatory mechanisms. Of the toxins observed exclusively in the female venom (Table 2), only DUF3472-1 and

GGT-1–GGT-4 represent distinct protein classes not found in the male. Additionally, of the toxins detected in the venom proteome of just the male venom (Table 2), neprilysins represent proteins from distinct classes not detected in the female venom. However, CAP3-1, SLPTX4s, SLPTX11s, and SLPTX16-1 represent proteins from distinct families within their respective classes that were not identified proteomically in the female. These presence/absence differences most likely contributed to the differences seen in relative peak abundances between the two sexes, but mass spectrometry of fractionated venom components would be needed to confirm proteomic composition of individual peaks. As GGTs accounted for a large portion of the female venom (42.5%), but were either absent from or at a low abundance in the male venom, these proteins could function to induce oxidative stress in female *H. marginata* prey. This presence/absence difference could also partially be explained if venom is involved in the maternal care of *H. marginata*. The abundance of SLPTXs (26.1%) and VPs (57.0%) in the male venom suggest that these proteins likely play a major role in overall venom function. However, as *H. marginata* lacks distinct sexual dimorphism, it is unclear as to what ecological mechanisms regulate this intersexual variation. Further studies aimed at studying the behavioral ecology of centipedes and characterizing sexually dimorphic venom components may help elucidate the mechanisms underlying sex-based variation in centipede venoms.

Acknowledgments

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

References

Abdel-Rahman, M. A., I. M. Abdel-Nabi, M. S. El-Naggar, O. A. Abbas, and P. N. Strong, 2011. Intraspecific variation in the venom of the vermivorous cone snail *Conus vexillum*. Comp. Biochem. Physiol. C 154:318–325.

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

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

Binford, G. J., R. G. Gillespie, and W. P. Maddison, 2016. Sexual dimorphism in venom chemistry in *Tetragnatha* spiders is not easily explained by adult niche differences. Toxicon 114:45–52.

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

Booncham, U., D. Sitthicharoenchai, A. Pradatsundarasar, S. Prasarnpun, and K. Thirakhupt, 2007. Sexual dimorphism in the Asian giant forest scorpion, *Heterometrus laoticus couzijn*, 1981. NU. Int. J. Sci. 4:42–52.

Carcamo-Noriega, E. N., L. D. Possani, and E. Ortiz, 2019. Venom content and toxicity regeneration after venom gland depletion by electrostimulation in the scorpion *Centruroides limpidus*. Toxicon 157:87–92.

Carlson, B. E., S. McGinley, and M. P. Rowe, 2014. Meek males and fighting females: sexually-dimorphic antipredator behavior and locomotor performance is explained by morphology in bark scorpions (*Centruroides vittatus*). PLOS ONE 9:e97648.

Carlson, B. E. and M. P. Rowe, 2009. Temperature and desiccation effects on the antipredator behavior of *Centruroides vittatus* (scorpiones: Buthidae). J. Arachnol. 37:321–331.

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

Cooper, A. M., W. J. Kelln, and W. K. Hayes, 2014. Venom regeneration in the centipede scolopendra polymorpha: Evidence for asynchronous venom component synthesis. Zoology 117:398–414.

Courtay, C., T. Oster, F. Michelet, A. Visvikis, M. Diederich, M. Wellman, and G. Siest, 1992. γ-glutamyltransferase: nucleotide sequence of the human pancreatic cDNA: evidence for a ubiquitous γ-glutamyltransferase polypeptide in human tissues. Biochem. Pharmacol. 43:2527–2533.

Cupul-Magaña, F. G., E. González-Santillán, E. Rodríguez-López, J. Bueno-Villegas, and L. E. Verdín-Huerta, 2018. First record of parental care in the scolopendrid centipede *Hemiscolopendra marginata* (say, 1821) from Mexico (Scolopendromorpha: Scolopendridae). Pan-Pac. Entomol. 94:1–3.

D'suze, G., M. Sandoval, and C. Sevcik, 2015. Characterizing *Tityus discrepans* scorpion venom from a fractal perspective: venom complexity, effects of captivity, sexual dimorphism, differences among species. Toxicon 108:62–72.

Dugon, M. M. and W. Arthur, 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.

Ellegren, H. and J. Parsch, 2007. The evolution of sex-biased genes and sex-biased gene expression. Nat. Rev. Genet. 8:689.

Falabella, P., L. Riviello, P. Caccialupi, T. Rossodivita, M. T. Valente, M. L. De Stradis, A. Tranfaglia, P. Varricchio, S. Gigliotti, F. Graziani, et al., 2007. A γ-glutamyl transpeptidase of *Aphidius ervi* venom induces apoptosis in the ovaries of host aphids. Insect Biochem. Mol. Biol. 37:453–465.

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

Glucksmann, A., 1974. Sexual dimorphism in mammals. Biol. Rev. 49:423–475.

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

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

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

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

Herzig, V., A. A. Khalife, Y. Chong, G. K. Isbister, B. J. Currie, T. B. Churchill, S. Horner, P. Escoubas, G. M. Nicholson, and W. C. Hodgson, 2008. Intersexual variations in northern (*Missulena pruinosa*) and eastern (*M. bradleyi*) mouse spider venom. Toxicon 51:1167–1177.

Inceoglu, B., J. Lango, J. Jing, L. Chen, F. Doymaz, I. N. Pessah, and B. D. Hammock, 2003. One scorpion, two venoms: prevenom of *Parabuthus transvaalicus* acts as an alternative type of venom with distinct mechanism of action. Proc. Natl. Acad. Sci. U.S.A. 100:922–927.

Kalia, J., M. Milescu, J. Salvatierra, J. Wagner, J. K. Klint, G. F. King, B. M. Olivera, and F. Bosmans, 2015. From foe to friend: using animal toxins to investigate ion channel function. J. Mol. Biol. 427:158–175.

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

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

Lewis, J., 1968. Individual variation in a population of the centipede *Scolopendra amazonica* from nigeria and its implications for methods of taxonomic discrimination in the Scolopendridae. Zool. J. Linnean. Soc. 47:315–326.

———, 2006. The biology of centipedes. Cambridge university press.

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

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

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

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

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

Luo, L., B. Li, S. Wang, F. Wu, X. Wang, P. Liang, R. Ombati, J. Chen, X. Lu, J. Cui, et al., 2018. Centipedes subdue giant prey by blocking KCNQ channels. Proc. Natl. Acad. Sci. U.S.A. P. 201714760.

Mank, J. E., 2008. Sex chromosomes and the evolution of sexual dimorphism: lessons from the genome. Amer. Nat. 173:141–150.

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

McMonigle, O., 2014. Centipedes in Captivity: The Reproductive Biology and Husbandry of Chilopoda. Coachwhip Publications Greenville.

Moran, Y., D. Praher, A. Schlesinger, A. Ayalon, Y. Tal, and U. Technau, 2013. Analysis of soluble protein contents from the nematocysts of a model sea anemone sheds light on venom evolution. Mar. Biotechnol. 15:329–339.

Nisani, Z., D. S. Boskovic, S. G. Dunbar, W. Kelln, and W. K. Hayes, 2012. Investigating the chemical profile of regenerated scorpion (*Parabuthus transvaalicus*) venom in relation to metabolic cost and toxicity. Toxicon 60:315–323.

Oksanen, J., F. Blanchet, R. Kindt, P. Legendre, P. Minchin, R. O'Hara, G. Simpson, P. Solymos, M. Stevens, and H. Wagner, 2013. Vegan: community ecology package. R package ver. 2.0–10.

Owens, I. P. and I. R. Hartley, 1998. Sexual dimorphism in birds: why are there so many different forms of dimorphism? Proc. Royal Soc. B 265:397–407.

Parker, G., 1992. The evolution of sexual size dimorphism in fish. J. Fish Biol. 41:1–20.

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

Peichoto, M. E., S. P. Mackessy, P. Teibler, F. L. Tavares, P. L. Burckhardt, M. C. Breno, O. Acosta, and M. L. Santoro, 2009. Purification and characterization of a cysteine-rich secretory protein from *Philodryas patagoniensis* snake venom. Comp. Biochem. Physiol. C 150:79–84.

Peng, Z., Y. Cheng, B. C.-M. Tan, L. Kang, Z. Tian, Y. Zhu, W. Zhang, Y. Liang, X. Hu, X. Tan, et al., 2012. Comprehensive analysis of RNA-seq data reveals extensive RNA editing in a human transcriptome. Nat. Biotechnol. 30:253.

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

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

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

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

Rokyta, D. R., M. J. Margres, and K. Calvin, 2015a. Post-transcriptional mechanisms contribute little to phenotypic variation in snake venoms. G3 5:2375–2382.

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

Rokyta, D. R., K. P. Wray, J. J. McGivern, and M. J. Margres, 2015b. The transcriptomic and proteomic basis for the evolution of a novel venom phenotype within the Timber Rattlesnake (*Crotalus horridus*). Toxicon 98:34–48.

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

Shelley, R., 2008. Revision of the centipede genus *Hemiscolopendra* kraepelin, 1903: Description of *H. marginata* (say, 1821) and possible misidentifications as Scolopendra spp.; proposal of Akymnopellis, n. gen., and redescriptions of its South American components (Scolopendromorpha: Scolopendridae: Scolopendrinae). Int. J. Myriap. 1:171–204.

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

Smith, J. J. and E. A. Undheim, 2018. True Lies: Using Proteomics to Assess the Accuracy of Transcriptome-Based Venomics in Centipedes Uncovers False Positives and Reveals Startling Intraspecific Variation in *Scolopendra subspinipes*. Toxins 10:96.

Templ, M., K. Hron, and P. Filzmoser, 2011. Robcompositions: Robust estimation for compositional data. Manual and Package, version 1.

Tragust, S., B. Mitteregger, V. Barone, M. Konrad, L. V. Ugelvig, and S. Cremer, 2013. Ants disinfect fungus-exposed brood by oral uptake and spread of their poison. Curr. Biol. 23:76–82.

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

Undheim, E. A., A. Jones, K. R. Clauser, J. W. Holland, S. S. Pineda, G. F. King, and B. G. Fry, 2014. Clawing through evolution: toxin diversification and convergence in the ancient lineage Chilopoda (Centipedes). Mol. Biol. Evol. P. msu162.

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

Vollrath, F. and G. A. Parker, 1992. Sexual dimorphism and distorted sex ratios in spiders. Nature 360:156.

Wang, Z., M. Gerstein, and M. Snyder, 2009. RNA-seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10:57.

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

Ward, M. J., S. A. Ellsworth, and D. R. Rokyta, 2018b. Venom-gland transcriptomics and venom proteomics of the Hentz striped scorpion (*Centruroides hentzi*; Buthidae) reveal high toxin diversity in a harmless member of a lethal family. Toxicon 142:14–29.

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

Whitfield, J., 2001. Gamma glutamyl transferase. Crit. Rev. Clin. Lab. Sci. 38:263–355.

Wong, E. S., D. Morganstern, E. Mofiz, S. Gombert, K. M. Morris, P. Temple-Smith, M. B. Renfree, C. M. Whittington, G. F. King, W. C. Warren, et al., 2012. Proteomics and deep sequencing comparison of seasonally active venom glands in the platypus reveals novel venom peptides and distinct expression profiles. Mol. Cell. Proteom. Pp. mcp–M112.

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

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

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

Figure Legends

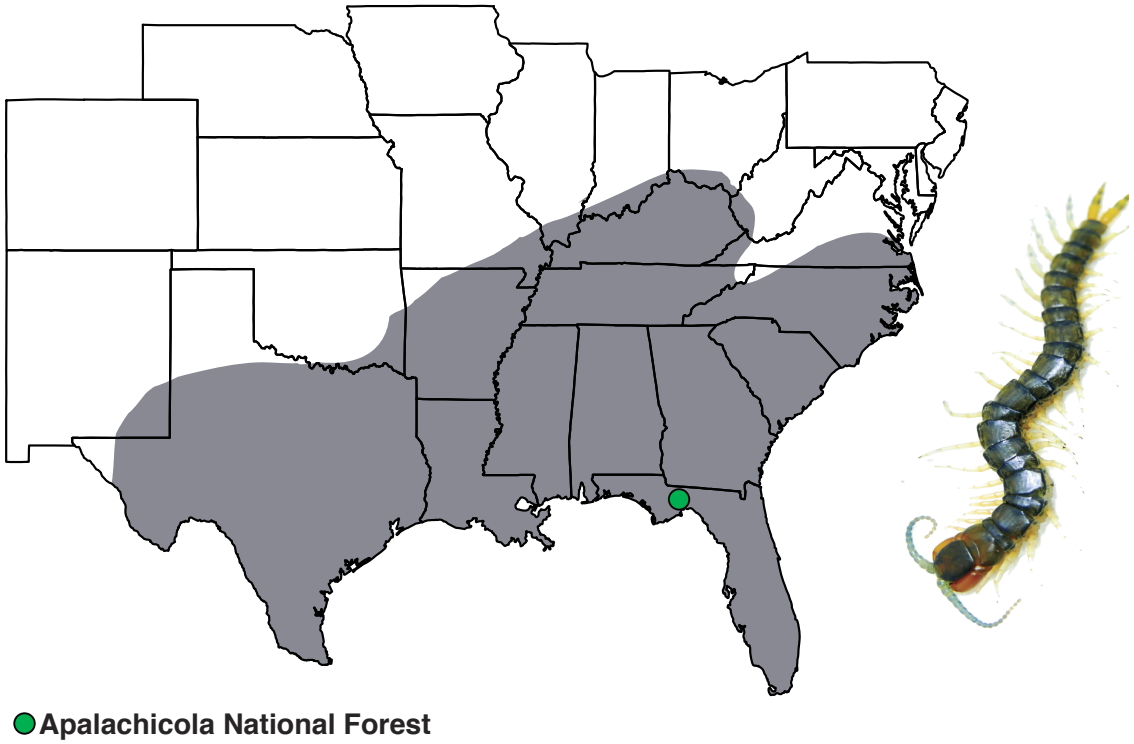

Figure 1. Range of *H. marginata* in the United States (shaded area) and our collection locality in the Apalachicola National Forest in northern Florida. This map only shows *H. marginata's* range in the United States, although the range extends into Mexico. A dorsal view of a representative adult *H. marginata* from northern Florida is shown on the right. Males and females do not show sex-based differences in morphology and range from 13–57mm in length.

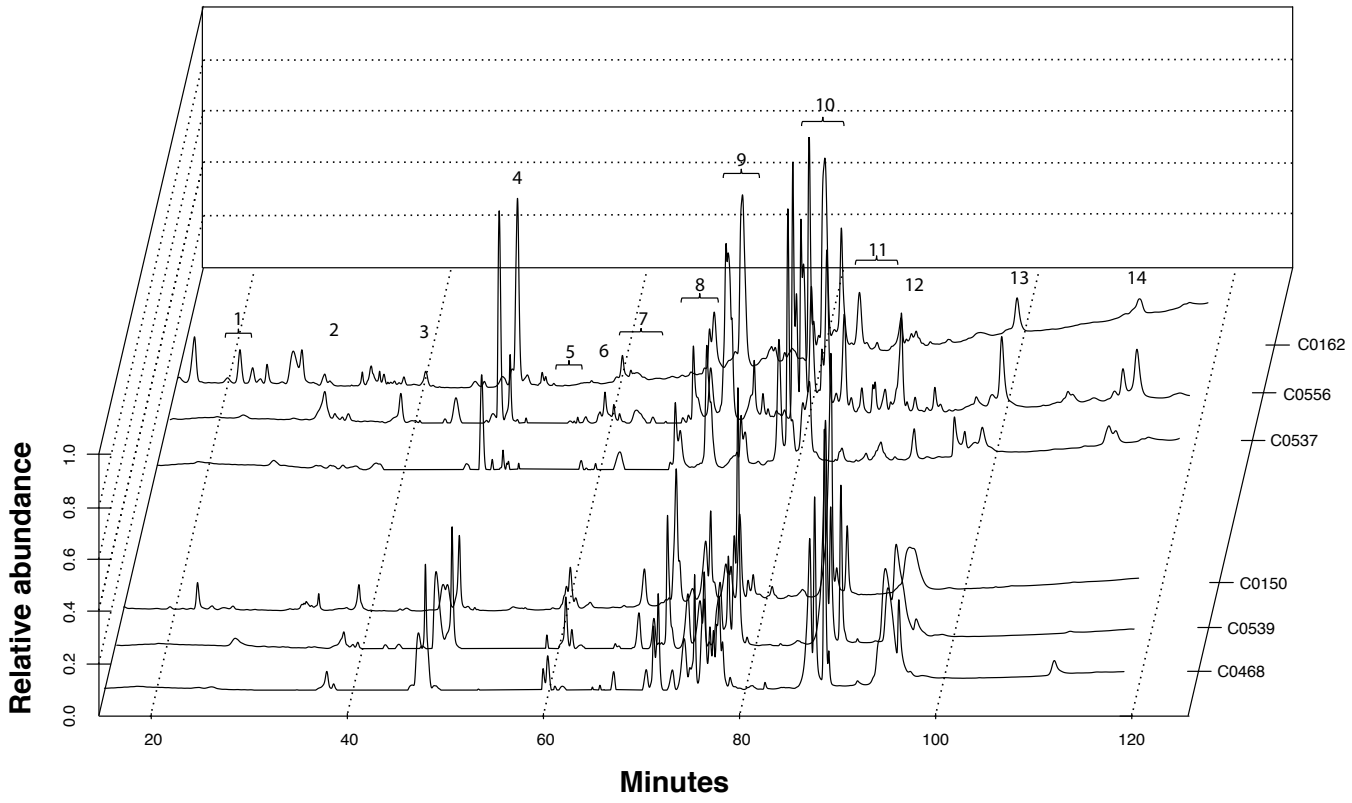

Figure 2. RP-HPLC profiles for male (C0162) and female (C0150) *H. marginata* transcriptome individuals and two additional male (C0537 and C0556) and two additional female (C0468 and C0539) individuals indicated approximately 14 distinct peak clusters between 20–120 minutes. Peaks 2, 10, 11, 12, and 14 were responsible for 10.9%, 8.7%, 14.3%, 13.2%, and 20.8% of the variation in venom and were eluted at approximately 32, 76, 81, 87, and 103 minutes respectively.

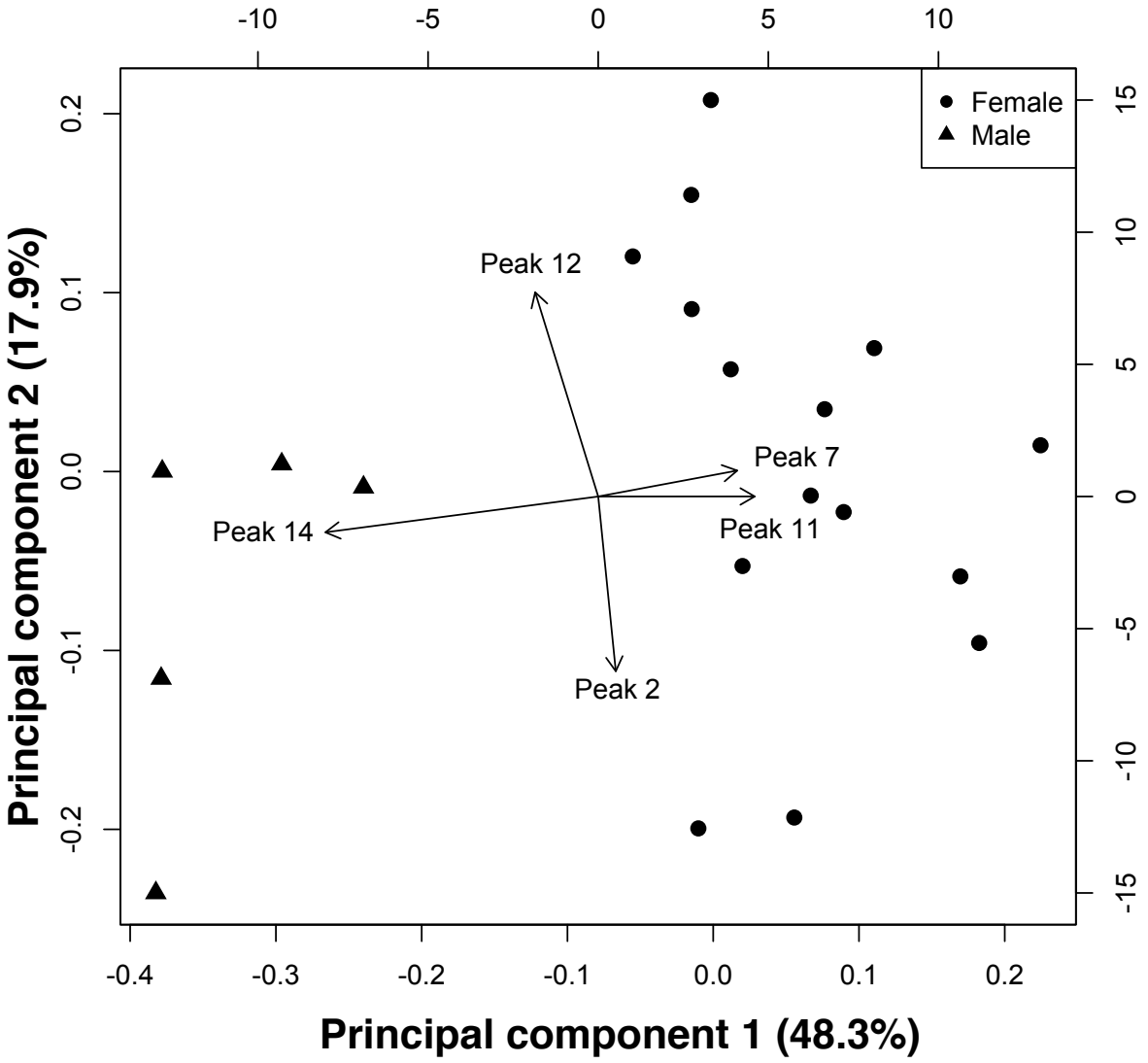

Figure 3. A principal component analysis of the RP-HPLC dataset revealed a separation between male (M) and female (F) individuals in the Apalachicola National Forest population. Sexes are indicated by shape with the five peaks that exhibited the most variation displayed. The scales on the top and right sides of the graph indicate axes for the component loadings.

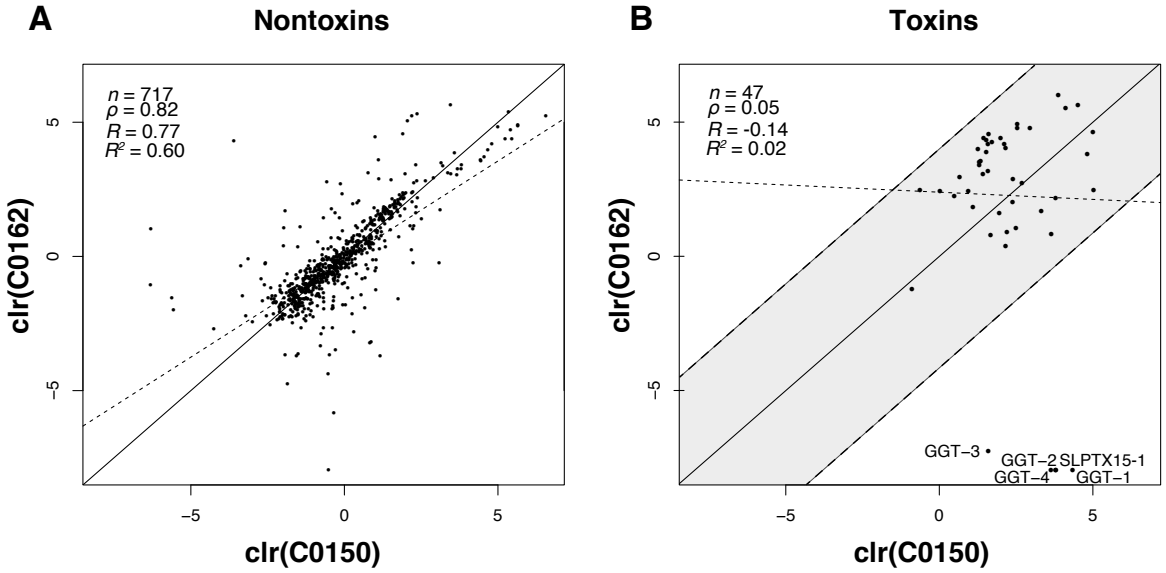

Figure 4. A venom-gland nontoxin transcript abundance comparison between *H. marginata* individuals (C0150; female, and C0162; male) showed strong agreement (A), but the toxin transcript abundance comparison between individuals were not correlated (B), indicating a difference in venom composition between males and females. Solid lines represent a correlation coefficient of one, while the short dashed lines represent the line of best fit. In the toxin transcript plot, labeled transcripts (GGT-1–GGT-4, and SLPTX15-1) are outside the 99$^{\text{th}}$ percentile of differences (shaded region between long dashed lines) between the two nontoxin measures. As such, they 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, GGT—γ-glutamyl transferase, SLPTX—scoloptoxin.

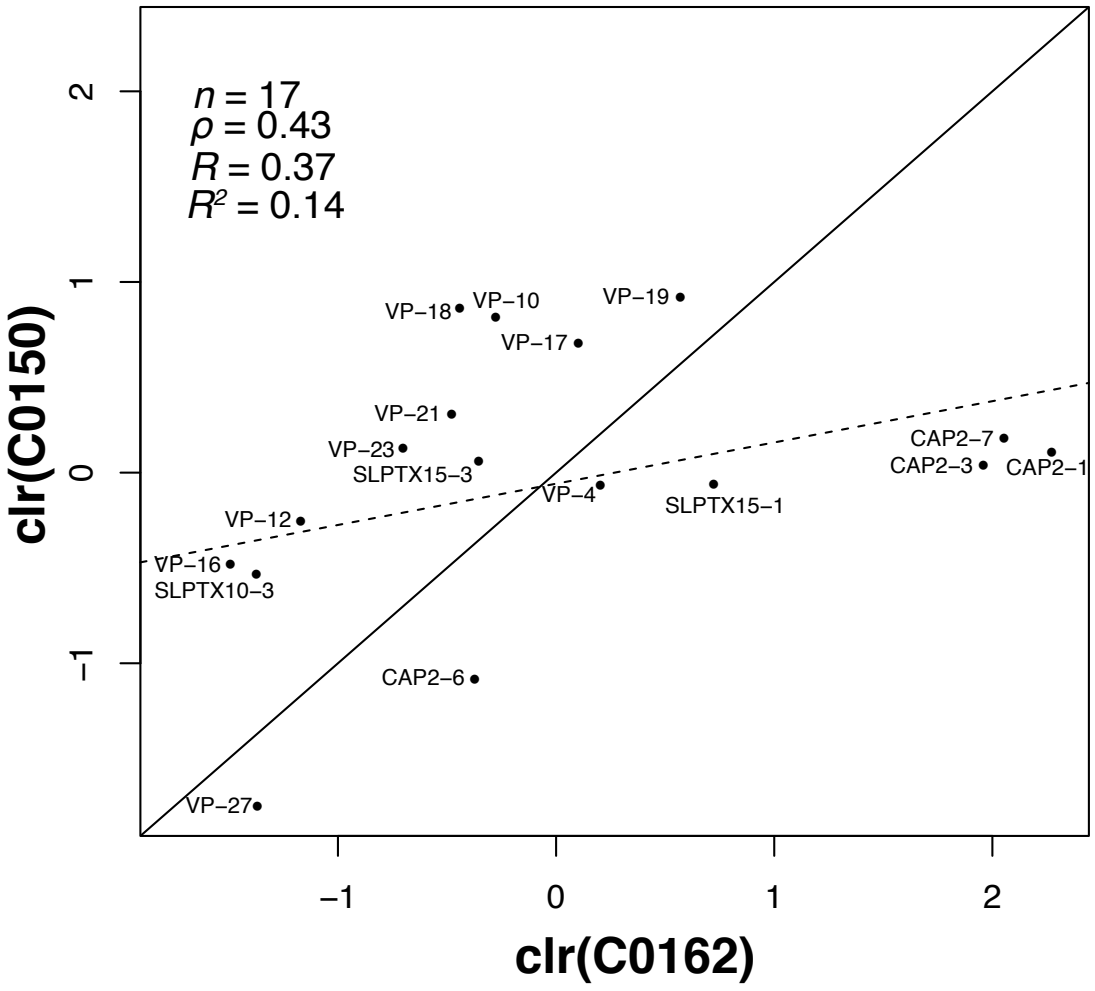

Figure 5. A venom proteomic comparison between an individual female (C0150) and male (C0162) *H. marginata* showed weak agreement for proteins detected in both venom proteomes. Table 2 shows the proteomic presence/absence differences between the two individuals. The solid line represents a correlation coefficient of one, while the dashed line reperesents the line of best fit. 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, CAP—cysteine-rich secretory protein, antigen 5, and pathogenesis-related 1 protein domains, SLPTX—scoloptoxin, VP—uncharacterized venom protein.

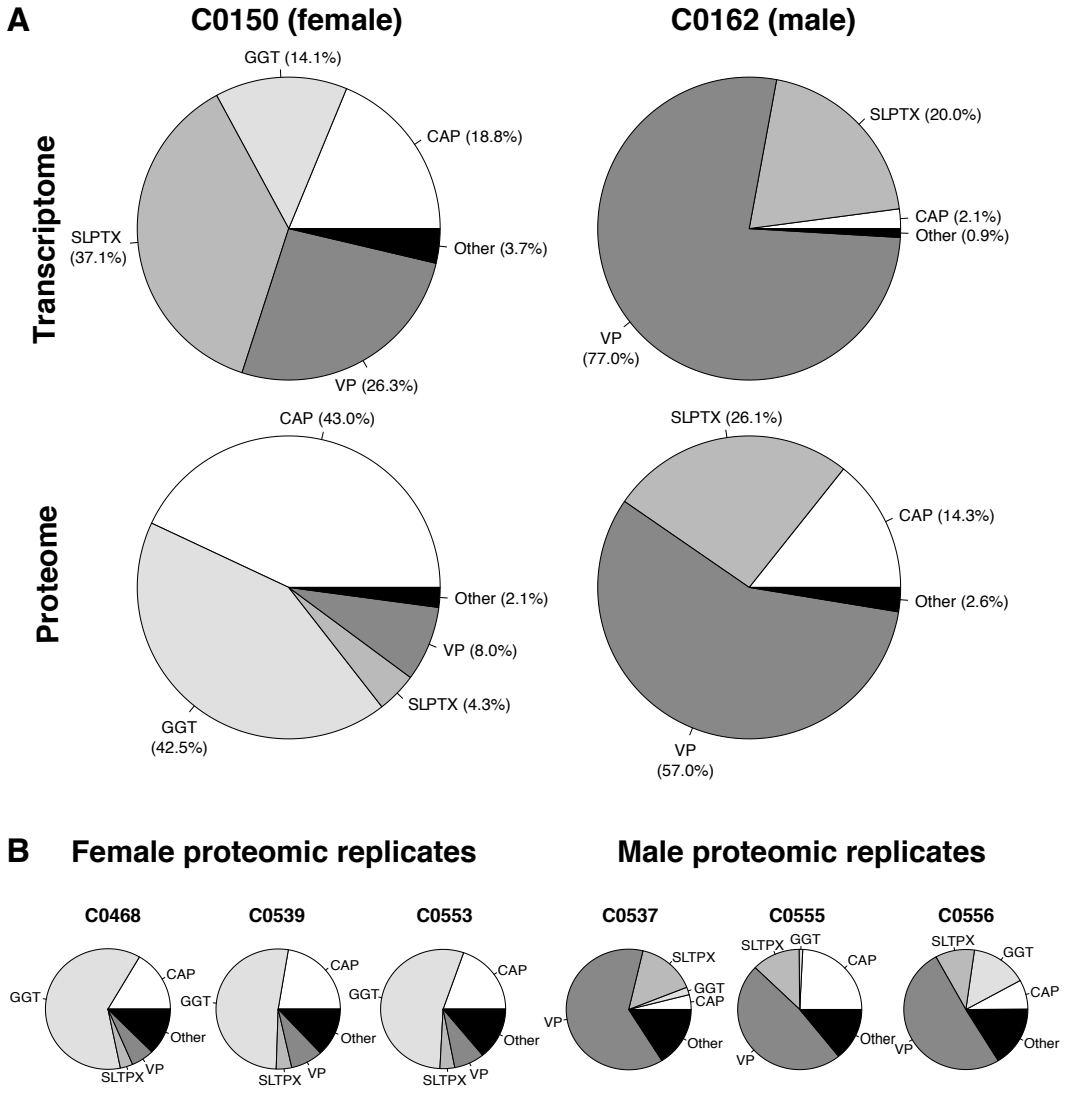

Figure 6. Pie graphs showing class level protein abundances between the two venom-gland transcriptomes and venom proteomes (A), and for three additional male and female proteomic replicates (B). We observed differences in class-level abundances between the two venom-gland transcriptomes and two venom proteomes, but these class-level comparisons for each individual showed weak agreement between the transcriptome and respective venom proteome. In both individuals, CAPs show a larger representation in the venom proteomes than predicted by their respective abundances. Additionally, SLPTXs displayed a greater abundance in the transcriptome compared to the respective venom proteomes. GGTs in the venom proteome of the female showed a considerably larger representation in the venom proteome than expected from transcript abundance. In the female and male proteomic replicates, GGTs were more highly expressed in female venoms and SLPTXs and VPs were more highly expressed in male venoms. Transcriptome abundances were based on transcripts per million (TPM) and percentages refer only to reads mapped to putative toxins (total toxin transcriptional output). Proteomic abundances were expressed as molar percentages. Abbreviations: GGT—γ-glutamyl transferase, CAP—cysteine-rich secretory protein, antigen 5, and pathogenesis-related 1 protein domains, SLPTX—scoloptoxin, VP—uncharacterized venom protein.

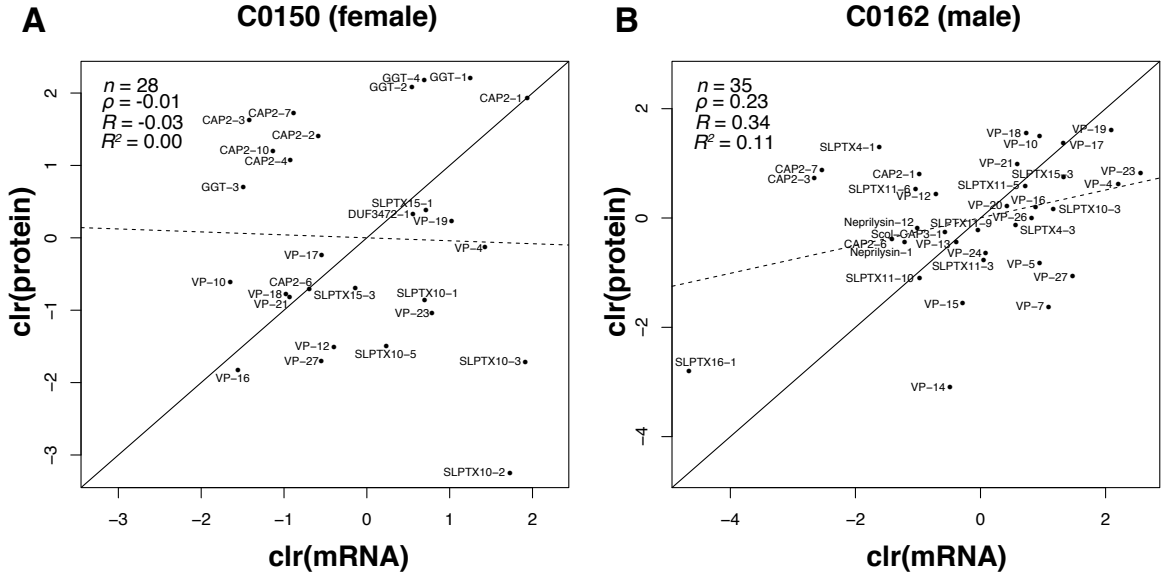

Figure 7. Protein and transcript abundance levels did not show agreement in the female (A). However, a weak agreement in protein and transcript abundance levels was observed in the male individual (B). Table 2 shows the proteomic presence/absence differences between the two individuals. Solid lines represent a correlation coefficient of one, while the dashed lines represent the line of best fit. 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, CAP—cysteine-rich secretory protein, antigen 5, and pathogenesis-related 1 protein domains, SLPTX—scoloptoxin, VP—uncharacterized venom protein.

Tables

Table 1. Toxins identified in the venom-gland transcriptome and venom proteome of *H. marginata*. (C0150; female, C0162; male)

Toxin	Signal peptide (aa)	Precursor (aa)	Cysteine Residues	MW (kDa)	C0150 TPM	C0162 TPM	C0150 fmol	C0162 fmol
CAP2-1	15	212	7	22.2	28,631.26	1,870.11	1,561.89	210.75
CAP2-2	23	212	6	21.0	2,288.89	450.76	923.15	–
CAP2-3	25	212	5	20.6	995.16	349.36	1,140.70	196.76
CAP2-4	24	211	6	20.7	1,633.87	228.80	656.43	–
CAP2-6	23	210	6	20.8	2,054.76	1,211.64	110.78	64.01
CAP2-7	23	212	6	20.8	1,701.66	395.53	1,255.79	226.65
CAP2-10	27	212	5	20.3	1,328.06	799.36	746.78	–
CAP3-1	17	520	36	57.4	2,064.69	2,831.44	–	72.65
DUF3472-1	18	434	1	46.3	7,192.02	362.09	314.76	–
GGT-1	20	572	4	60.3	14,371.58	–	2,055.47	–
GGT-2	20	573	4	60.4	7,107.46	–	1,810.27	–
GGT-3	20	572	4	60.4	926.41	0.11	455.60	–
GGT-4	20	573	4	60.3	8,257.11	–	1,999.01	–
Neprilysin-1	29	717	11	80.0	307.70	1,478.48	–	61.16
Neprilysin-12	31	710	10	79.3	484.41	1,813.91	–	78.26
SLTPX4-1	17	65	4	5.4	564.78	990.62	–	344.79
SLTPX4-3	25	66	4	4.5	664.24	8,733.33	–	83.99
SLPTX10-1	23	115	6	10.9	8,292.55	1,363.30	95.71	–
SLPTX10-2	24	67	6	4.7	23,174.42	7,168.40	8.81	–
SLPTX10-3	22	80	6	6.3	27,939.75	15,978.93	40.75	111.06
SLPTX10-5	23	99	6	8.6	5,199.88	855.57	50.90	–
SLPTX11-3	18	225	16	22.7	691.96	5,245.99	–	44.19
SLPTX11-5	20	233	17	24.7	918.03	10,213.32	–	170.80
SLPTX11-6	20	235	17	24.8	192.67	1,770.97	–	160.87
SLPTX11-9	22	224	16	21.6	688.51	4,777.88	–	75.50
SLPTX11-10	22	238	16	25.1	100.36	1,883.91	–	31.48
SLPTX15-1	23	76	4	6.1	8,425.26	–	331.61	177.92
SLTPX15-3	23	74	4	5.7	3,581.59	18,786.69	112.90	200.69
SLPTX16-1	25	117	8	10.2	76.99	47.14	–	5.72
VP-4	29	188	4	18.0	17,169.54	45,229.07	197.21	176.73
VP-5	22	181	4	18.2	1,383.54	12,811.90	–	41.28
VP-7	19	80	4	6.7	940.50	14,796.49	–	18.66
VP-10	19	179	4	18.2	792.95	12,834.92	122.04	426.96
VP-12	18	372	14	40.7	2,766.53	2,451.47	49.96	146.66
VP-13	21	80	4	6.6	776.22	3,384.38	–	61.21
VP-14	19	186	4	20.0	363.31	3,062.69	–	4.29
VP-15	19	183	4	18.5	911.72	3,753.50	–	19.88
VP-16	26	98	0	8.1	868.12	12,047.33	36.20	116.74
VP-17	19	182	4	18.4	2,383.55	18,728.12	178.36	373.19
VP-18	21	133	8	12.8	1,553.34	10,391.09	103.44	446.93
VP-19	19	171	4	16.9	11,473.86	40,373.07	284.52	474.16
VP-20	19	182	4	18.2	867.82	7,616.59	–	118.23
VP-21	19	178	4	18.2	1,619.79	8,966.08	99.72	256.27
VP-23	19	178	4	18.1	9,049.93	64,526.73	79.75	214.84
VP-24	22	527	12	60.5	717.57	5,414.70	–	49.83
VP-26	19	178	4	18.0	1,046.44	11,311.60	–	95.46
VP-27	22	196	4	19.8	2,374.28	21,706.29	40.94	32.77

Cysteine residues and molecular weights were determined using ExPASy ProtParam (Gasteiger et al., 2005) with signal peptides excluded. Abbreviations: CAP—cysteine-rich secretory protein, antigen 5, and pathogenesis-related 1 protein domains, DUF—domain with unknown function, GGT—γ-glutamyl transferase, SLPTX—scoloptoxin, VP—uncharacterized venom protein.

Table 2. Presence/absence differences between venom proteomes.

	C0150 (Female)			C0162 (Male)			Average	
Protein	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	C0150	C0162
CAP2-2	942.99	907.16	919.29	—	—	—	923.15	—
CAP2-4	650.22	642.95	676.12	—	—	—	656.43	—
CAP2-10	1,141.06	—	1,099.26	—	—	—	746.78	—
CAP3-1	—	—	—	77.85	71.47	68.63	—	72.65
DUF3472-1	305.72	312.66	325.89	—	—	—	314.76	—
GGT-1	2,066.83	2,008.04	2,091.56	—	—	—	2,055.47	—
GGT-2	1,851.57	1,774.68	1,804.57	—	—	—	1,810.27	—
GGT-3	—	—	1,366.81	—	—	—	455.60	—
GGT-4	1,985.01	2,008.04	2,003.99	—	—	—	1,999.01	—
Neprilysin-1	—	—	—	60.55	71.47	51.48	—	61.16
Neprilysin-12	—	—	—	77.85	88.28	68.63	—	78.26
SLPTX4-1	—	—	—	363.29	327.91	343.17	—	344.79
SLPTX4-3	—	—	—	82.17	88.28	81.50	—	83.99
SLPTX10-1	111.95	92.48	82.69	—	—	—	95.71	—
SLPTX10-2	—	26.42	—	—	—	—	8.81	—
SLPTX10-5	—	74.86	77.83	—	—	—	50.90	—
SLPTX11-3	—	—	—	43.25	37.84	51.48	—	44.19
SLPTX11-5	—	—	—	181.65	189.18	141.56	—	170.80
SLPTX11-6	—	—	—	151.37	163.96	167.29	—	160.87
SLPTX11-9	—	—	—	73.52	71.47	81.50	—	75.50
SLPTX11-10	—	—	—	38.92	12.61	42.90	—	31.48
SLPTX16-1	—	—	—	—	—	17.16	—	5.72
VP-5	—	—	—	47.57	46.24	30.03	—	41.28
VP-7	—	—	—	25.95	—	30.03	—	18.66
VP-13	—	—	—	56.22	63.06	64.34	—	61.21
VP-14	—	—	—	—	—	12.87	—	4.29
VP-15	—	—	—	12.97	25.22	21.45	—	19.88
VP-20	—	—	—	116.77	113.51	124.40	—	118.23
VP-24	—	—	—	47.57	50.45	51.48	—	49.83
VP-26	—	—	—	95.15	88.28	102.95	—	95.46

Quantities provided in fmol. Abbreviations: CAP—cysteine-rich secretory protein, antigen 5, and pathogenesis-related 1 protein domains, DUF—domain with unknown function, GGT—γ-glutamyl transferase, SLPTX—scoloptoxin, VP—venom protein.