ZOOLOGICAL RESEARCH

Comparative analysis of diverse toxins from a new pharmaceutical centipede, *Scolopendra mojiangica*

Zi-Chao Liu[1,3,#], Jin-Yang Liang[2,5,#], Xin-Qiang Lan[1,2,5], Tao Li[1,4], Jia-Rui Zhang[2,7], Fang Zhao[1,4,6], Geng Li[1,4], Pei-Yi Chen[1,4], Yun Zhang[2,*], Wen-Hui Lee[2,*], Feng Zhao[1,4,6,*]

[1] *Key Laboratory of Ethnic Medical Resources Research and Southeast Asian International Cooperation of Yunnan Universities*, *Department of Biology and Chemistry*, *Puer University*, *Puer*, *Yunnan* 665000, *China*

[2] *Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences* & *Yunnan Province*, *Kunming Institute of Zoology*, *Chinese Academy of Sciences*, *Kunming*, *Yunnan* 650223, *China*

[3] *Engineering Research Center for Exploitation and Utilization of Leech Resources in Universities of Yunnan Province*, *School of Agronomy and Life Sciences*, *Kunming University*, *Kunming*, *Yunnan* 650214, *China*

[4] *Key Laboratory of Active Molecules and Drug Development*, *Puer University*, *Puer*, *Yunnan* 665000, *China*

[5] *Kunming College of Life Science*, *University of Chinese Academy of Sciences*, *Kunming*, *Yunnan* 650204, *China*

[6] *Institute of Comparative Study of Traditional Materia Medica*, *Institute of Integrative Medicine of Fudan University*, *Shanghai* 200032, *China*

[7] *Nanshan College*, *Guangzhou Medical University*, *Guangzhou*, *Guangdong* 511436, *China*

ABSTRACT

As the oldest venomous animals, centipedes use their venom as a weapon to attack prey and for protection. Centipede venom, which contains many bioactive and pharmacologically active compounds, has been used for centuries in Chinese medicine, as shown by ancient records. Based on comparative analysis, we revealed the diversity of and differences in centipede toxin-like molecules *between Scolopendra mojiangica*, a substitute pharmaceutical material used in China, and *S. subspinipes mutilans*. More than 6 000 peptides isolated from the venom were identified by electrospray ionization-tandem mass spectrometry (ESI-MS/MS) and inferred from the transcriptome. As a result, in the proteome of *S. mojiangica*, 246 unique proteins were identified: one in five were toxin-like proteins or putative toxins with unknown function, accounting for a lower percentage of total proteins than that in *S. mutilans*. Transcriptome mining identified approximately 10 times more toxin-like proteins, which can characterize the precursor structures of mature toxin-like peptides. However, the constitution and quantity of the toxin transcripts in these two centipedes were similar. In toxicity assays, the crude venom showed strong insecticidal and hemolytic activity. These findings highlight the extensive diversity of toxin-like

Received: 15 October 2019; Accepted: 06 December 2019; Online: 31 December 2019

Foundation items: This work was supported by grants from the Chinese National Natural Science Foundation (81860696, 31560596, 81373945, and 31360516), Yunnan Applied Basic Research Projects (2016FD076), Key Research Program of the Chinese Academy of Sciences (KJZD-EW-L03), "Yunling Scholar" Program, Yunnan Provincial Training Programs of Youth Leader in Academic and Technical Reserve Talent (2019HB058), and Puer University (2017XJKT12 & CXTD011)

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: zhangy@mail.kiz.ac.cn; leewh@mail.kiz.ac.cn; zhaofeng@mail.kiz.ac.cn

DOI: 10.24272/j.issn.2095-8137.2020.019

proteins in *S. mojiangica* and provide a new foundation for the medical-pharmaceutical use of centipede toxin-like proteins.

Keywords: Centipede; Toxins; Pharmaceutical use; Proteotranscriptomic analysis

INTRODUCTION

As one of the oldest and most important predatory arthropods, the centipede has a fossil record that extends back 420 million years (Undheim & King, 2011). Approximately 3 300–3 500 centipede species have been found, with distribution worldwide and in most provinces of China (Rong et al., 2015). Centipede venom, which is secreted from venom glands in the first pair of limbs (Edgecombe & Giribet, 2007), is essential for survival, not only for subduing and killing prey but also for defense against predators.

Animal venom has long been considered a rich source of pharmacological and novel therapeutics (Kalia et al., 2015; Smith et al., 2013; Zhang, 2015). Furthermore, dried centipedes have been used medicinally for centuries, as shown in ancient Chinese medical records. Recently, an increasing number of studies have shown that centipede venom contains various functional components, including a rich reservoir of structural and pharmacological peptides (Hakim et al., 2015; Undheim et al., 2015, 2016). In addition, because of their excellent chemical and pharmacological activities, particularly as neurotoxins and ion channel inhibitors, centipede toxins have received further attention (Liu et al., 2012; Yang et al., 2012, 2013, 2015). Several antimicrobial peptides and specific toxins have also been identified in centipede venom (Chen et al., 2014; Hou et al., 2013; Peng et al., 2010; Yang et al., 2012). Interestingly, centipede toxins are expressed outside the venom gland and are involved in gene recruitment processes (Zhao et al., 2018a). These venom peptides have significant chemical, thermal, and biological stability, which enable researchers to adapt their functions for therapeutic use.

Therefore, centipede venom research is of great interest for investigating putative toxins. These toxins can act on a range of molecular targets, including voltage-gated sodium (Na_V), potassium (K_V), and calcium (Ca_V) channels (Liu et al., 2012; Yang et al., 2012). However, biochemical studies on centipede toxins are not nearly as extensive as studies on other venomous animals, such as snakes, spiders, and scorpions (Undheim et al., 2016), and complete data on centipede venom toxins, peptides, and protein sequences are currently limited to a small number of species (Hakim et al., 2015; Undheim et al., 2016). One potential reason is that most centipede species are considered too small to obtain enough venom for activity testing or high-throughput drug screening. Omics analysis of venom or venom glands is one approach for probing toxin molecular diversity. Specifically, to identify new putative proteins and enable comparison across species, large-scale sequencing of a broad array of centipede venom should be applied to further confirm the complexity of venom (Gonzalez-Morales et al., 2014; Liu et al., 2012; Rong et al., 2015).

Previous centipede research has mainly focused on *Scolopendra mutilans* (Zhao et al., 2018a), and occasionally on *S. subspinipes subspinipes*, *S. viridis*, and *S. dehaani* (Liu et al., 2012). To date, however, no comprehensive research has been reported on the new pharmaceutical centipede, *S. mojiangica* (Wang et al., 1997), which is used as a substitute medicinal material in traditional Chinese medicine. Therefore, a fully integrated approach combining transcriptomics and proteomics is essential for understanding the differences among pharmaceutical centipedes, including venom composition and toxin diversity. Here, in-depth proteotranscriptomic analyses (combined proteomic and transcriptomic analyses) were used to study centipede venom, and the protein/peptide composition of the dissected venom gland from *S. mojiangica* was described. Complete comparative analyses of the protein compounds and toxin distribution in the venom or venom gland of *S. mojiangica* and *S. mutilans* were also presented based on RNA-Seq and MS datasets.

MATERIALS AND METHODS

Animals and ethics

Adult *S. mojiangica* (both sexes) were collected from Mojiang (N23°27', E101°41'), Yunnan Province, China. All centipede (*S. mojiangica*) studies were reviewed and approved by the Animal Care and Use Committee of Puer University (ACUP. 531068520180126, approved on 17 September 2018).

Venom collection and sample preparation

The venom of *S. mojiangica* was collected as per our previous method. Briefly, a 3 V alternating current (AC) was used to stimulate the venom glands in the first pair of centipede limbs (Liu et al., 2012). The venom samples were stored at −20 °C until use. A 300 mg *S. mojiangica* venom sample was solubilized in 3 mL of Tris-HCl buffer. The venom solution was then loaded on a Sephacryl S-100HR (HiprepTM26/60, 71-1247-00-EG, GE Healthcare, USA) gel filtration column with a flow rate of 0.5 mL/h. Thirteen peaks (named P1−13) were obtained from this procedure (Supplementary Figure S1).

The proteins/peptides contained in the venom were pre-denatured with 500 μL of 25 nmol/L NH_4HCO_3 and separated with a 3 kDa cut-off ultrafiltration tube. The low molecular weight (<3 kDa) proteins/peptides were collected and desalinated before peptidomic analysis. Proteins/peptides with molecular weights greater than 3 kDa were applied to SDS-PAGE gels for separation. One half of each sample was mixed with extraction buffer (0.25% acetic acid and protease inhibitor cocktail) and disrupted with a sonicator (Hielscher Ultrasound Technology, Germany). To further separate these samples, 12% gel with protein ladder (Thermo, ref. 26614, USA) SDS-PAGE was used, followed by staining with GelCode Blue Stain (Thermo ref. 24592, USA) and destaining with Milli-Q water (Millipore, USA). We excised six bands from each lane for in-

gel trypsin digestion. Samples were extracted with 100% acetonitrile, desalinated, lyophilized, and stored at −80 °C until further electrospray ionization-tandem mass spectrometry (ESI-MS/MS) analysis.

RNA extraction, sequencing, and transcriptome analysis

A total of 260 mg of venom gland was preserved in liquid nitrogen after collection from *S. mojiangica* until use. RNA extraction and cDNA library construction were performed according to our previous work (Zhao et al., 2014a, 2014b). cDNA from the *S. mojiangica* venom gland was sequenced using the Illumina HiSeq™ 2000 (USA), and the short-read assembly program SOAPdenovo-Trans (v1.03) was run with default parameters to complete *de novo* transcriptome assembly. Overlaps with certain lengths and connected paired-end reads were combined in the program to form contigs. The sequence clustering software TGICL was used to splice sequences and remove redundant sequences to produce the complete assembly of contigs of each sample (Pertea et al., 2003), and the longest possible non-redundant unigenes were produced. The TGICL parameters were the same as the parameters used in our previous work (Zhao et al., 2014b).

HPLC fractionation and mass spectrometry

After in-gel digestion, candidate fractionation samples were loaded onto an EASY-nLC HPLC system (Thermo Fisher Scientific, USA) equipped with a binary rapid separation nano-flow pump and ternary loading pump. Mobile phase eluent A (0.1% TFA contained in ddH_2O) and mobile phase eluent B (ACN/ddH_2O/TFA 90/10/0.08% (v/v/v)) were used. Samples were applied to a Thermo Scientific EASY loading column (2 cm×100 μm, 5 μm −C18, USA) by the auto-sampler and analytical column (75 μm×100 mm, 3 μm −C18), respectively, with a flow rate of 250 nL/min. With linear stepwise gradients (0′−5% B, 5′−5% B, 12.5′−20% B, 62.5′−70% B, 63.5′−99% B, 65′−99% B, 66′−5% B and 72′−5% B), we separated the peptides with the column. Starting at 20% eluent B, 1.25 mL/5 min of each fraction was collected and lyophilized.

We selected the data-dependent mode of the Q Exactive instrument (Thermo Finnigan, USA), which then switched between full scan MS and MS/MS acquisition automatically. Based on the predictive automatic gain control (AGC) of the previous full scan, we accumulated 3×10^6 target value ions and acquired 70 000 (m/z 200) resolution of full scan MS spectra (m/z 300−1 800) in the Orbitrap. In addition, 15 s was set as the dynamic exclusion value. We isolated and fragmented the 10 most intense multiply charged ions (z≥2) sequentially by higher-energy collisional dissociation (HCD) with a fixed resolution of 17 500 (m/z 200) and an injection time of 60 ms for the MS2 scanning method. The mass spectrometric conditions were as follows: 2 kV spray voltage, no sheath and auxiliary gas flow, 250 °C heated capillary temperature, 27 eV normalized HCD collision energy, and 0.1% underfill ratio. A total of 1×10^5 counts was set as the ion selection threshold for MS/MS.

Data processing and bioinformatics analysis

Using Proteome Discoverer (version 1.4), RAW data files were produced. Mascot v2.2 was used as the search tool to generate peak lists in our transcriptome database. Trypsin was chosen as an enzyme, and two missed cleavages were allowed. The MS/MS search criteria were as follows: MS polypeptide tolerance 2×10^4 mg/m³ and MS/MS mode 0.1 Da. The aminomethylation of cysteine was statically modified and the oxidation of methionine was dynamically modified. High confidence peptides were used for protein identification, generating a 1% false discovery rate (FDR) threshold. Only unique peptides with high confidence were used for protein identification.

All unigenes in our centipede database were annotated with BLASTX and searched against known databases, as presented in our previous study (Zhao et al., 2014a, 2014b, 2018a). Unigenes were aligned with high-priority databases and annotated with a given description instead of aligning with a low-priority database. Gene Ontology (GO) annotation was carried out using the Blast2GO (Conesa et al., 2005) software suite v2.5.0. In these searches, the BLASTX cut-off was set to $1e^{-6}$. The BLAST tool was used to search the toxin database and annotate the toxin with Tox-Prot in UniProtKB (02 February 2019, 6 822 sequences) and the animal toxin database platform ATDB (He et al., 2008), with the toxins then verified by phylogenetic analyses. The grouped sequences were aligned using MUSCLE v3.8.31 (Edgar, 2010). MrBayes 3.2.7 was used for phylogenetic analyses with maximum likelihood. The values were estimated by ultrafast bootstrap using 10 000 iterations. The resulting trees were analysed with MEGA 7 (Kumar et al., 2016), which was also used to automatically plot expression values and detection in venom.

Comparative expression analysis was performed as follows: comparison of RNA-Seq data of venom glands of various species was performed using Bowtie v0.12.7 (Langmead et al., 2009) and TopHat v2.0.6 (Trapnell et al., 2009) for mapping. Gene expression values were calculated from the expected number of fragments per kilobase of transcript sequence per million base pairs sequenced (FPKM) (Trapnell et al., 2010). The FPKM values for genes from every tissue were determined by rSeq (Jiang & Wong, 2009). The graphs and statistical analyses were performed using GraphPad Prism v5.0 (La Jolla, USA) and R v3.3.2. Here, $P<0.05$ was considered statistically significant.

Insect bioassays and hemolytic assays

Insect bioassays were performed according to the method in Yang et al. (2012). Freeze-dried crude venom powder was dissolved in insect saline (concentrations in deionized water: 140 mmol/L NaCl, 5 mmol/L KCl, 4 mmol/L $NaHCO_3$, 1 mmol/L $MgCl_2$, 0.75 mmol/L $CaCl_2$, 5 mmol/L HEPES) and injected into grasshoppers (*Locusta migratoria manilensis*; mass 700–900 mg) and mealworms (*Tenebrio molitor* larvae; mass 190–210 mg). Ants (*Tetramorium spp*., adults; mass 35–55 mg) were fed with same venom.

Using human, mouse, and rabbit red blood cells (RBCs),

hemolytic activity was assayed as described previously (Liu et al., 2012; Zhao et al., 2018b). Briefly, serial dilutions of the samples were incubated with washed RBCs (3%) at 37 °C for 30 min and then centrifuged. The resulting supernatant was measured at an absorbance of 540 nm. Maximum hemolysis was determined by adding 1% Triton X-100 to the cell samples.

RESULTS

Phylogeny of scolopendrid centipedes and isolation of venom gland

Original Chinese medicinal centipedes include *S. mutilans*, *S. multidens*, *S. mojiangica*, and *S. negrocapitis* (Wang et al., 1997). Here, we studied the novel substitutional pharmaceutical centipede, *S. mojiangica*, with comparative analysis of active molecules. *Scolopendra mojiangica* showed a relatively close relationship to *S. negrocapitis*, *S. mutilans*, and *S. multidens* (Figure 1A), though a smaller body size than *S. mutilans*, *S. dehaani*, and *S. multidens*. Similar to other species, it also uses venom to attack prey and in defense.

The protocol for isolating venom glands from *S. mojiangica* was described in our previous study (Liu et al., 2012). Healthy adult centipedes (*n*=280) without injury were selected, and the venom glands were dissected from their first pair of limbs. After that, 3 V AC was used to stimulate the venom gland and ensure that more toxins were included, so that proteome coverage could be improved. The isolated venom glands were then further processed (Figure 1B). A portion of each sample was used to obtain the proteome by SDS-PAGE analysis. Protein bands from the venom gland were excised for in-gel digestion and subjected to ESI-MS/MS analysis. The remaining portion of each sample was used to extract RNA, followed by RNA-Seq analysis of the transcriptome.

Proteomic analysis of venom components

A total of 246 proteins were identified in *S. mojiangica* at 95% coverage by ESI-MS/MS analysis (Supplementary Table S1; Figure 2A). In the proteome, 73.6% of proteins (*n*=181) were cellular components and 19.1% of proteins (*n*=47) were unknown functional proteins, which were putative venom toxins. Only 18 proteins were identified as toxin-like proteins, including neurotoxins, K^+ channel inhibitors, and blarina toxins (Figure 2B; Table 1). Although we obtained more proteins in *S. mojiangica* than in *S. mutilans* and *S. viridis* with proteomic analysis, the detected toxin-like proteins in *S. mojiangica* represented a lower percentage of total proteins than those identified in *S. mutilans* in our previous study (Figure 2C). In the venom proteome, most of the identified proteins showed a molecular weight of less than 50 kDa, similar to the proteome of *S. mutilans* (Figure 2D). Thus, the centipedes contained notably small functional molecules for potential pharmaceutical use, as expected. Based on peptide detection, 23.2% of proteins consisted of six or more unique peptides (Supplementary Figure S2). In addition, the more enriched the peptides assembled into proteins, the more comprehensive was the proteome obtained.

Transcriptomic analysis of venom components

We acquired 43 381 437 clean reads assembled into 132 597

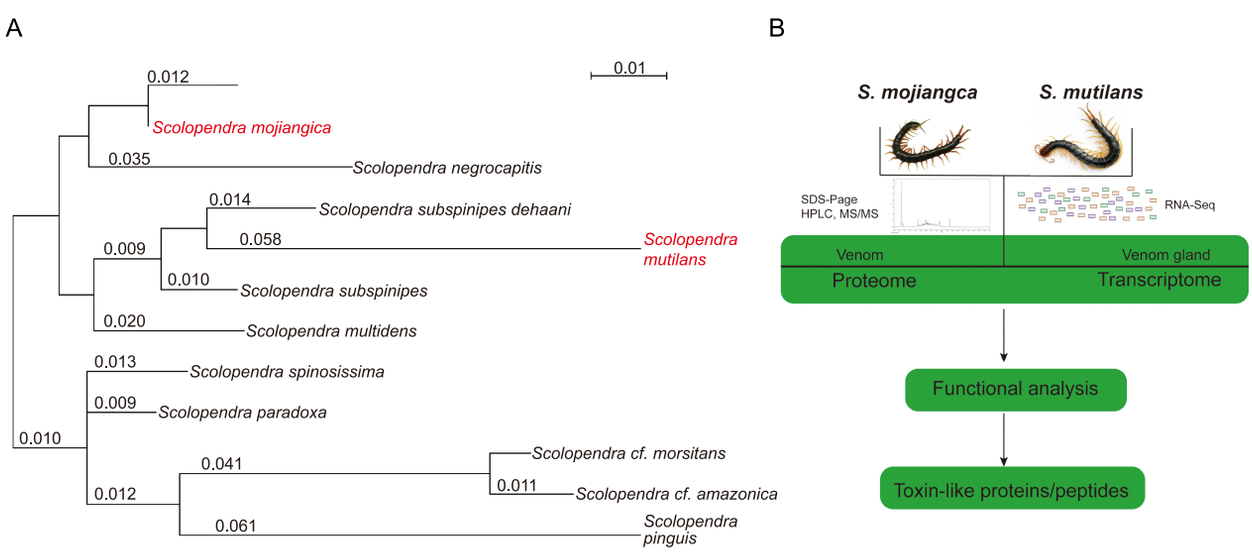

Figure 1 Proteomic and transcriptomic analyses of new pharmaceutical centipede

A: Molecular phylogenetic analysis of centipede, *S. mojiangica*, by maximum likelihood based on *COI* genes. Red labels correspond to two centipedes in our study, and posterior probabilities are assigned to nodes. B: Workflow for proteomic and transcriptomic analyses of centipede, *S. mojiangica*. Venom was processed and subjected to SDS-PAGE followed by in-gel digestion. Samples were then analysed in a separate ESI-MS/MS assay. For transcriptomic analysis, venom glands (not venom) were used for high-throughput sequencing. Functional analysis was combined with proteomic and transcriptomic data.

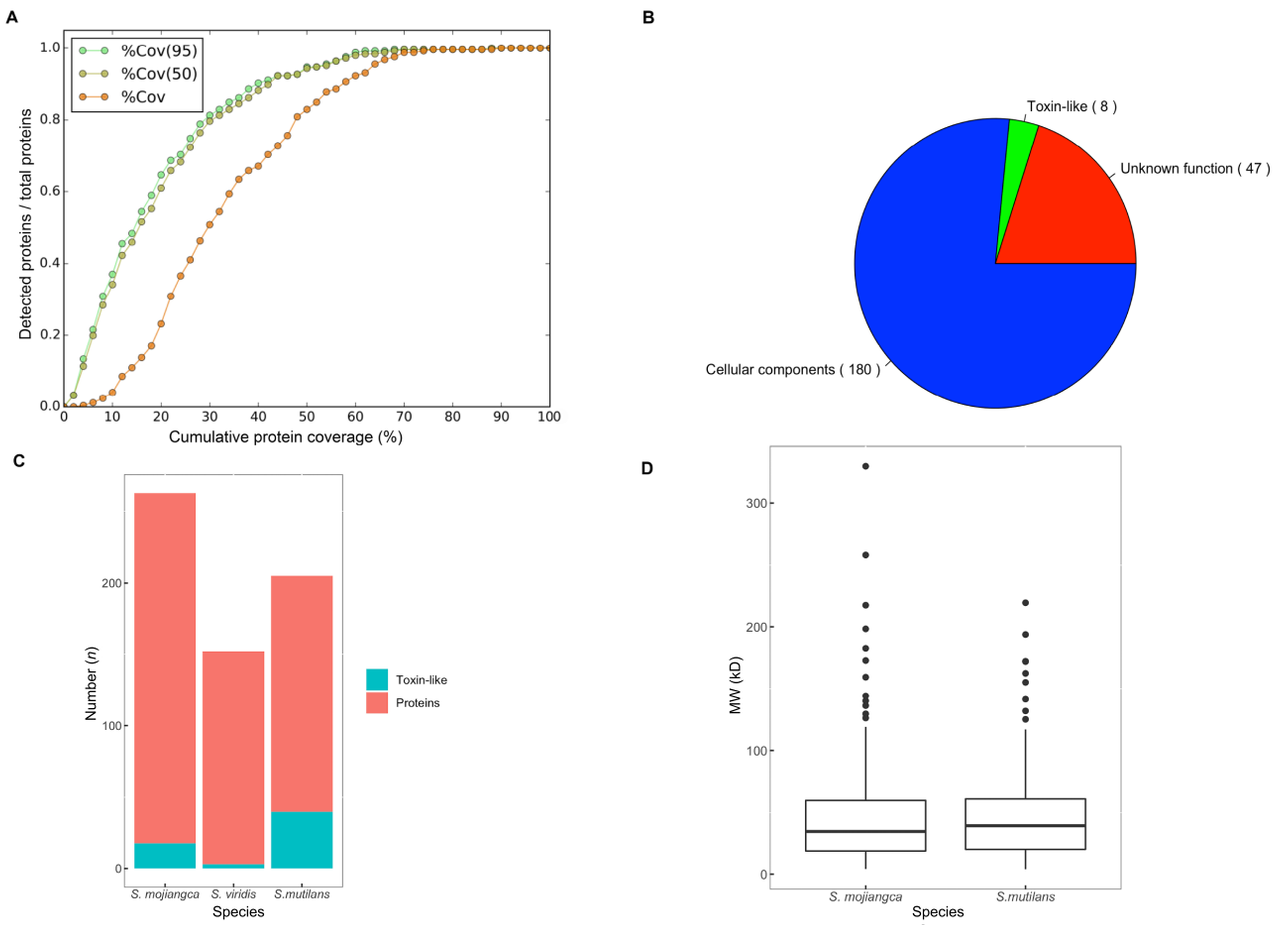

Figure 2 Profiles of *S. mojiangica* proteome

A: Cumulative distribution of protein peptide coverage. Horizontal axis shows protein peptide coverage, and vertical axis shows protein ratio. B: Pie chart of identified proteins from our *S. mojiangica* proteome. C: Comparison of toxin-like proteins determined by proteomic analysis among three centipedes: i.e., *S. mojiangica*, *S. viridis* (Gonzalez-Morales et al., 2014), and *S. mutilans* (Zhao et al., 2018a). D: Distribution of molecular weights of proteome proteins.

contigs from the venom gland using the Trinity program. As a result, the transcriptome data consisted of 107 642 putative gene objects (all unigenes) ranging from 101 bp to 9 184 bp, with an average length of 423 bp. The number of unigenes larger than 500 bp was 24 219. The largest unigenes were 9 184 bp in size, and the N50 of the unigenes was 214 bp (Supplementary Figure S3 and Table S2).

For comparative analysis, the venom gland transcriptome from *S. mojiangica* showed many transcripts (*n*=46 571) with high similarity to those of *S. mutilans*. Notably, however, most transcripts showed low similarity between the two centipede species (Figure 3A). In the transcriptomic expression analysis, the read count of each transcript in *S. mojiangica* and *S. mutilans* showed biases for gene expression, with higher expressed transcripts in *S. mojiangica* (Figure 3B). Functional annotation analyses of these transcripts were combined with Blast searching and phylogenetic analyses to obtain toxin-like unigenes. In total, 410 toxin-like transcripts were identified in the transcriptome of *S. mojiangica*, more than that identified in *S. mutilans* (342 transcripts). Furthermore, these transcripts were divided into 34 categories, mainly consisting of alpha-latrocrustotoxin, delta-latroinsectotoxin, ion channel inhibitors, and alpha-latrotoxin (Figure 3C).

Comparative determination of centipede toxins

As expected, we identified 34 kinds of toxin-like unigenes (*n*=342) from the transcriptome of *S. mutilans* using the same annotation method as that of *S. mojiangica* (Figure 4A). In total, 11 of these toxin-like unigenes encoded the most transcripts in the two centipedes. With gene expression analyses, most toxin-like unigenes showed no differential expression between *S. mojiangica* and *S. mutilans*, except for four toxin-like unigenes (i.e., alpha-latrotoxin, hopsarin-D, metalloproteinase, and trocarin) (Figure 4B).

Finally, we determined the toxicity and performed crude isolation of the centipede venom. The crude centipede venom

Table 1 Toxin-like proteins/peptides identified from venom proteome of *S. mojiangica* centipede

Sequence ID	GenBank accession No.	Sequence description	Category	Peptides	E-Value	MW (kD)	Calc. pI	FPKM
ScoMo_singlet48841	AT0003236	Blarina toxin precursor (EC 3.4.21.-)	Blarina toxin	9	1.00E-37	21.61	4.15	92.54
ScoMo_singlet50899	AT0003766	Mucrofibrase-5 precursor (EC 3.4.21.-)	Mucrofibrase-5	11	4.00E-16	14.40	9.93	3 454.74
ScoMo_singlet71394	AT0002263	Pseudechetoxin-like protein precursor	Pseudechetoxin	276	9.00E-42	28.74	9.86	7 195.57
ScoMo_contig2076	gi|429840589	K+ channel inhibitor	Channel inhibitor	617	4.00E-164	62.76	9.15	1.37
ScoMo_singlet78309	AT0000117	Latisemin precursor	Latisemin	412	2.00E-22	20.89	7.96	0.00
ScoMo_contig4762	AT0003236	Blarina toxin precursor (EC 3.4.21.-)	Blarina toxin	108	1.00E-44	28.58	6.5	15 173.32
ScoMo_singlet45908	AT0003741	Thrombin-like enzyme contortrixobin (EC 3.4.21.-)	Serine proteinase	109	1.00E-41	44.94	5.08	1 685.57
ScoMo_singlet67462	AT0000120	Pseudecin precursor	Pseudechetoxin	66	5.00E-32	23.71	8.91	14 111.58
ScoMo_singlet72573	AT0000552	Hopsarin-D (EC 3.4.21.6)	Hopsarin-D	93	1.00E-121	85.15	6.53	132.70
ScoMo_singlet76606	AT0000554	Trocarin precursor (EC 3.4.21.6)	Trocarin	38	3.00E-138	84.92	6.17	60.34
ScoMo_singlet25641	AT0000552	Hopsarin-D (EC 3.4.21.6)	Hopsarin-D	46	5.00E-20	27.21	4.6	184.53
ScoMo_singlet69905	AT0000554	Trocarin precursor (EC 3.4.21.6)	Trocarin	14	4.00E-107	40.69	5.28	1 245.366
ScoMo_singlet57737	AT0003404	Zinc metalloproteinase fibrolase (EC 3.4.24.72)	Metalloproteinase	20	4.00E-16	35.21	8.13	48.71
ScoMo_singlet8256	AT0000762	Alpha-latrocrustotoxin	Alpha-latrocrustotoxin	10	0	50.48	6.79	136.27
ScoMo_singlet68890	AT0000552	Hopsarin-D (EC 3.4.21.6)	Hopsarin-D	13	5.00E-75	42.03	7.88	161.84
ScoMo_singlet7846	gi|392295725	Omega-slptx-ssm2a neurotoxin precursor	Neurotoxin	11	8.00E-36	8.56	4.93	16 647.01
ScoMo_singlet55496	gi|501293796	Cathepsin L	Cathepsin L	180	1.00E-155	37.30	6.35	2.83
ScoMo_singlet39956	AT0000554	Trocarin precursor (EC 3.4.21.6)	Trocarin	12	4E-09	4.64	3.79	5.55

MW: Molecular Weight; Calc. pI: The calculated isoelectric point (pI); FPKM: Fragments Per Kilobase of exon model per Million mapped fragments.

exhibited strong insecticidal action (Figure 5A), and the crude venom had a similar potency as the venom of *S. mutilans*. The crude venom and its fractions eluted from the S-100HR column (Supplementary Figure S1; Figure 5B) showed hemolytic activity. The elution of peak 1 (P1) showed high hemolytic activity on human RBCs when 1 mg/mL protein/peptide was incubated for 4 h. In contrast, peaks 3, 5, and 6 (P3, P5, and P6) had lower hemolytic activity than that of P1 and crude venom.

DISCUSSION

Due to long-term evolutionary fine-tuning, venom toxins exhibit high specificity and potency for molecular targets that are not often found in natural or synthetic small molecules, and thus animal toxins are valuable pharmacological tools (King, 2011, 2013). There are many cases in which venom toxin has been used as a pharmacological molecule, e.g., snake venom, dried toad skin secretions (Chan Su), tarantula venom, and cobra venom used as traditional Ayurvedic, Chinese, Mexican, and Central and South American medicines, respectively (Harvey, 2014; King, 2011). These traditional medicines have been used to treat arthritis, gastrointestinal ailments, asthma, polio, multiple sclerosis, rheumatism, severe pain, and trigeminal neuralgia, or as a diuretic anesthetic and anti-cancer agent. Centipede venom has different biomedical properties and represents a vast reservoir of toxins, similar to venom from other animals. Due to its origins in one of the oldest venomous arthropods, centipede venom displays excellent activities and good prospects for drug development (Undheim et al., 2016; Zhang, 2015). Importantly, the centipede is a traditional Chinese medicine with an application history of more than 2 000 years (Chen & Yu, 1999; Zhao et al., 2018a). In China, pharmaceutically applied centipedes include *S. mutilans*, *S. multidens*, *S. dehaani*, and *S. negrocapitis*, with *S. mojiangica* (Wang et al., 1997) very occasionally used as a substitute. Our results showed that the venom toxicity of this centipede is strong in comparison to that of *S. mutilans*, a commonly used centipede in medicine.

In our previous study, the centipede showed diverse protein or peptide components, with the most abundant toxins in the

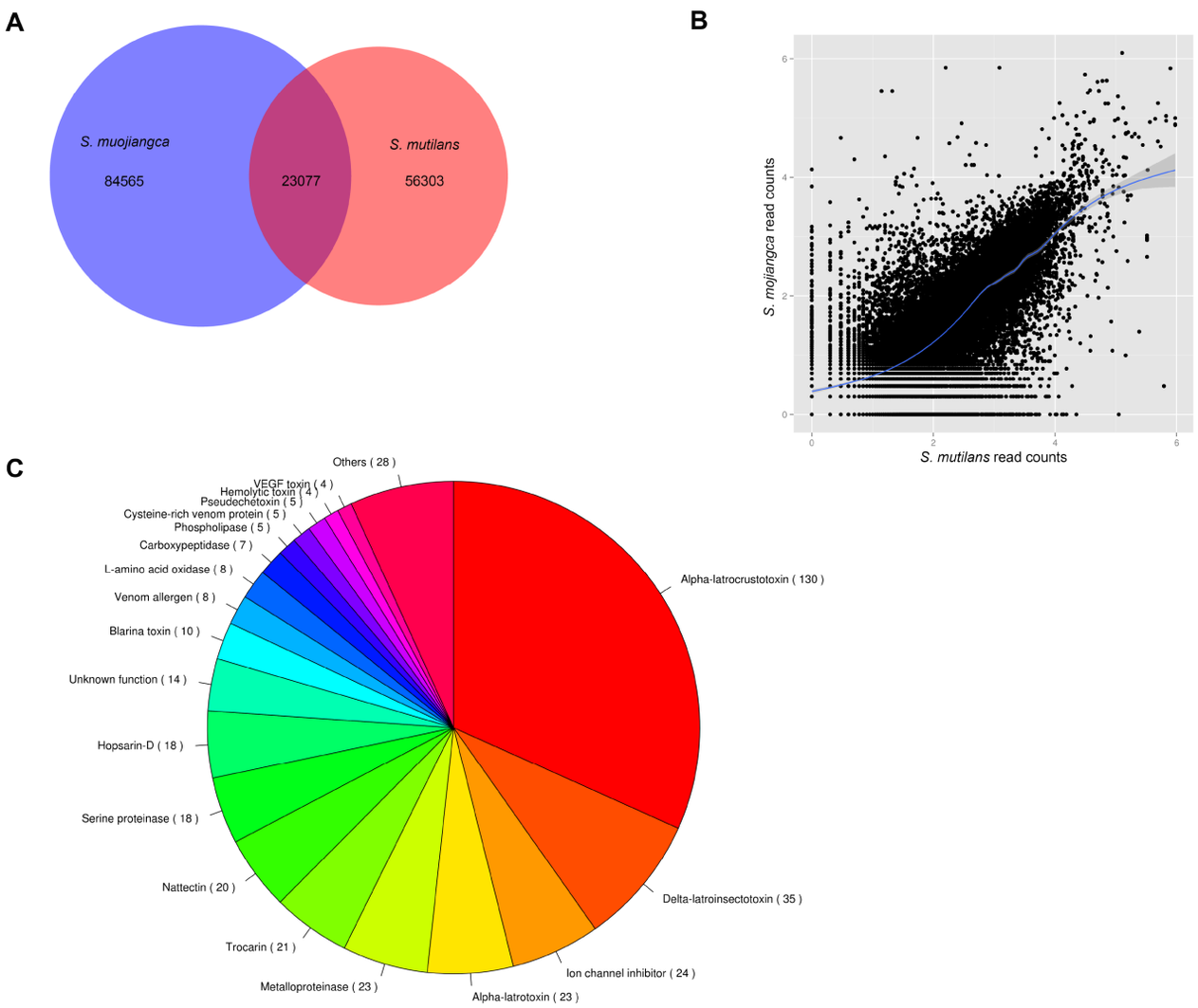

Figure 3 Identification of toxins from transcriptome of venom gland in centipedes

A: Comparison of transcripts identified in venom glands from two centipedes, *S. mojiangica* and *S. mutilans*, with transcriptomic analysis. B: Expression of all transcripts in venom glands of *S. mojiangica* and *S. mutilans*. Read counts reflect quantification accuracy of differential expression by mapping reads to transcripts and read counting. C: Pie chart of venom toxin-like proteins/peptides identified in transcriptomes of *S. mojiangica* and *S. mutilans*. In total, 410 and 342 venom toxin-like proteins/peptides were identified from *S. mojiangica* and *S. mutilans*, respectively, using transcriptomic analysis.

venom and torso tissues found to be more highly expressed than other active molecules using our method (Liu et al., 2012; Zhao et al., 2018a). Here, based on proteomic detection, we showed that the toxin-like proteins in *S. mojiangica* accounted for a lower percentage of total proteins than that in *S. mutilans*. However, there was a similar constitution and quantity of toxin transcripts in these two centipedes. We used high-throughput ESI-MS/MS and RNA-Seq technology to investigate the diversity of novel venom proteins, especially low-abundance peptides/proteins not detected using conventional methods (Savitski et al., 2005). Most of the detected proteins were identified as potentially active molecules with low molecular weights and unknown functions. In addition, each detected protein contained at least six peptides in the proteome dataset. The proteomic results for *S. mojiangica* were very similar to the protein detection results for *S. mutilans*. More than 400 toxin-like proteins/peptides were identified by transcriptome analysis in the centipede, but not detected in the proteome. Thus, most putative toxins in centipede venom may have low levels of expression in *S. mojiangica* and *S. mutilans*. In conclusion, centipede venom contains a surprising variety of toxin-like proteins/peptides.

Regarding toxin distribution, based on transcriptomic analysis, we identified more toxin transcripts in *S. mojiangica*

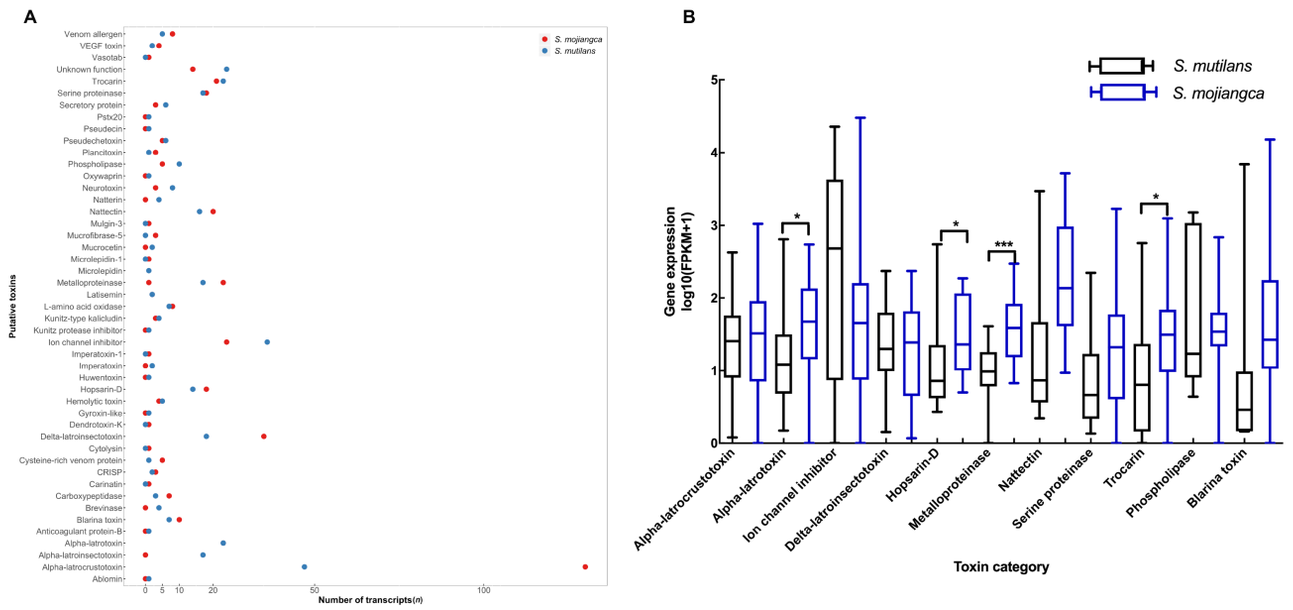

Figure 4 Comparison of toxin-like molecules distributed in centipedes *S. mojiangica* and *S. mutilans*

A: Distribution of identified toxin-like molecules in *S. mojiangica* and *S. mutilans*. Toxin-like transcripts (*n*=410) in *S. mojiangica* were divided into 34 categories. Blue dots represent transcripts of *S. mutilans* and red dots represent transcripts of *S. mojiangica*. B: Main components of toxin-like molecules expressed in *S. mojiangica* and *S. mutilans*. Transcriptomic analysis showed only four types of toxin-like molecules with differential gene expression between *S. mojiangica* and *S. mutilans*.

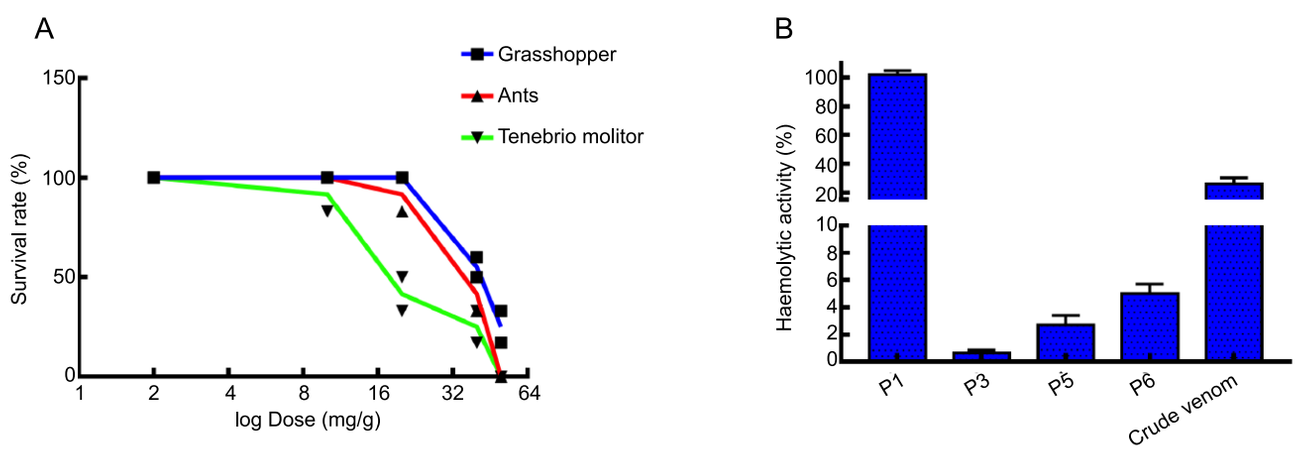

Figure 5 Insecticidal activity of crude centipede venom

A: Insecticidal activity of crude centipede venom. B: Hemolytic activity of elution of crude centipede venom. Peaks 1, 3, 5, and 6 at concentrations of 1 mg/mL were incubated with human red blood cells for 30 min at 37 °C, and absorbance of supernatant was measured at 540 nm.

than in *S. mutilans*. Most toxins did not show significantly differential expression between *S. mojiangica* and *S. mutilans*, including that of ion channel inhibitors and serine proteinases. The centipede *S. mojiangica* demonstrated higher gene expression of metalloproteinase, trocarin, hopsarin-D, and alpha-latrotoxin compared to *S. mutilans*. Therefore, *S. mojiangica* could be substituted for *S. mutilans* in medical use. These results indicate that *S. mojiangica* venom could be a rich source of pharmacologically and medically useful compounds.

Usually, we can obtain approximately 0.2–0.5 mg of crude venom from a single adult *S. mutilans* centipede over a period of two weeks. However, one adult *S. mojiangica* yielded less than 0.1 mg of crude venom in the same period. Therefore, it was difficult to study the venom components, including their pharmaceutical activity or medicinal application. In addition to the current annotation methods of centipede toxins, our results revealed that a wide variety of toxin-like active molecules were

expressed in the venom gland by combining Blast alignment with the existing toxin databases and phylogenetic reconstruction of toxin relationships. Theoretically, this method may produce false positives, especially for proteins with low abundance and expression when using high-throughput proteomic and transcriptomic analyses with ESI-MS/MS and RNA-Seq technology. However, we used previously established approaches to maximize the search for functional proteins. Our results provide good evidence that the use of this substitute medicinal centipede is an appropriate medical option. Importantly, our data provide important clues to improve the use of the centipede as a traditional Chinese medicine.

CONCLUSIONS

Here, we used omics techniques to determine the profiles of venom components and toxin-like molecules in a new pharmaceutical centipede, *S. mojiangica*. We performed in-depth proteomic analysis of venom and deduced full-length protein sequences by combining proteome and transcriptome databases. We obtained more than 400 toxin-like molecules with potent activity. With gene expression and inter-species comparative analysis, we identified a broad and diverse composition of toxin-like molecules, which may play key roles in the functions of centipede venom. Our results indicate that this centipede is valuable for medicinal use and drug development, like other centipede species. Furthermore, our methods could improve the application of the centipede as a traditional Chinese medicine.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

F.Z., W.H.L., and Y.Z. conceived the study. F.Z. wrote the paper and performed data and bioinformatics analyses. Z.C.L., J.Y.L., X.Q.L., T.L., J.R.Z., F.Z., G.L., and P.Y.C. performed all experiments. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank the members of our research groups for providing technical assistance and participating in discussions.

REFERENCES

Chen K, Yu B. 1999. Certain progress of clinical research on Chinese integrative medicine. *Chinese Medical Journal (Engl)*, **112**(10): 934−937.

Chen M, Li J, Zhang F, Liu Z. 2014. Isolation and characterization of SsmTx-I, a Specific Kv2.1 blocker from the venom of the centipede *Scolopendra Subspinipes Mutilans* L. Koch. *Journal of Peptide Science*, **20**(3): 159−164.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**(18): 3674−3676.

Edgar RC. 2010. Quality measures for protein alignment benchmarks. *Nucleic Acids Research*, **38**(7): 2145−2153.

Edgecombe GD, Giribet G. 2007. Evolutionary biology of centipedes (Myriapoda: Chilopoda). *Annual Review of Entomology*, **52**: 151−170.

Gonzalez-Morales L, Pedraza-Escalona M, Diego-Garcia E, Restano-Cassulini R, Batista CV, Gutierrez Mdel C, Possani LD. 2014. Proteomic characterization of the venom and transcriptomic analysis of the venomous gland from the Mexican centipede *Scolopendra viridis*. *Journal of Proteomics*, **111**: 224−237.

Hakim MA, Yang S, Lai R. 2015. Centipede venoms and their components: resources for potential therapeutic applications. *Toxins (Basel)*, **7**(11): 4832−4851.

Harvey AL. 2014. Toxins and drug discovery. *Toxicon*, **92**: 193−200.

He QY, He QZ, Deng XC, Yao L, Meng E, Liu ZH, Liang SP. 2008. ATDB: a uni-database platform for animal toxins. *Nucleic Acids Research*, **36**.

Hou H, Yan W, Du K, Ye Y, Cao Q, Ren W. 2013. Construction and expression of an antimicrobial peptide scolopin 1 from the centipede venoms of *Scolopendra subspinipes mutilans* in *Escherichia coli* using SUMO fusion partner. *Protein Expression and Purification*, **92**(2): 230−234.

Jiang H, Wong WH. 2009. Statistical inferences for isoform expression in RNA-Seq. *Bioinformatics*, **25**(8): 1026−1032.

Kalia J, Milescu M, Salvatierra J, Wagner J, Klint JK, King GF, Olivera BM, Bosmans F. 2015. From foe to friend: using animal toxins to investigate ion channel function. *Journal of Molecular Biology*, **427**(1): 158−175.

King G. 2013. Venoms to drugs: translating venom peptides into therapeutics. *Australian Biochemist*, **44**(3): 13−16.

King GF. 2011. Venoms as a platform for human drugs: translating toxins into therapeutics. *Expert Opinion on Biological Therapy*, **11**(11): 1469−1484.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, **33**(7): 1870−1874.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, **10**(3): R25.

Liu ZC, Zhang R, Zhao F, Chen ZM, Liu HW, Wang YJ, Jiang P, Zhang Y, Wu Y, Ding JP, Lee WH, Zhang Y. 2012. Venomic and transcriptomic analysis of centipede *Scolopendra subspinipes dehaani*. *Journal of Proteome Research*, **11**(12): 6197−6212.

Peng K, Kong Y, Zhai L, Wu X, Jia P, Liu J, Yu H. 2010. Two novel antimicrobial peptides from centipede venoms. *Toxicon*, **55**(2-3): 274−279.

Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J. 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*, **19**(5): 651−652.

Rong M, Yang S, Wen B, Mo G, Kang D, Liu J, Lin Z, Jiang W, Li B, Du C, Yang S, Jiang H, Feng Q, Xu X, Wang J, Lai R. 2015. Peptidomics combined with cDNA library unravel the diversity of centipede venom. *Journal of Proteomics*, **114**: 28−37.

Savitski MM, Nielsen ML, Kjeldsen F, Zubarev RA. 2005. Proteomics-grade de novo sequencing approach. *Journal of Proteome Research*, **4**(6):

2348−2354.

Smith JJ, Herzig V, King GF, Alewood PF. 2013. The insecticidal potential of venom peptides. *Cellular and Molecular Life Sciences: CMLS*, **70**(19): 3665−3693.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**(9): 1105−1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol*, **28**(5): 511−515.

Undheim EA, Fry BG, King GF. 2015. Centipede venom: recent discoveries and current state of knowledge. *Toxins (Basel)*, **7**(3): 679−704.

Undheim EA, Jenner RA, King GF. 2016. Centipede venoms as a source of drug leads. *Expert Opinion on Drug Discovery*, **11**(12): 1139−1149.

Undheim EA, King GF. 2011. On the venom system of centipedes (Chilopoda), a neglected group of venomous animals. *Toxicon*, **57**(4): 512−524.

Wang K, Fang H, Ye M, Chen H, Zhu Y, Fang H. 1997. Investigation on the resources of medicinal centipedes and identification on their commodities. *Journal of Chinese Medicinal Materials*, **20**(9): 450−452.

Yang S, Liu Z, Xiao Y, Li Y, Rong M, Liang S, Zhang Z, Yu H, King GF, Lai R. 2012. Chemical punch packed in venoms makes centipedes excellent predators. *Molecular and Cellular Proteomics*, **11**(9): 640−650.

Yang S, Xiao Y, Kang D, Liu J, Li Y, Undheim EA, Klint JK, Rong M, Lai R, King GF. 2013. Discovery of a selective NaV1.7 inhibitor from centipede venom with analgesic efficacy exceeding morphine in rodent pain models. *Proceedings of the Natlional Academy of Sciences of the Untied States of America*, **110**(43): 17534−17539.

Yang S, Yang F, Wei N, Hong J, Li B, Luo L, Rong M, Yarov-Yarovoy V, Zheng J, Wang K, Lai R. 2015. A pain-inducing centipede toxin targets the heat activation machinery of nociceptor TRPV1. *Nature Communications*, **6**: 8297.

Zhang Y. 2015. Why do we study animal toxins?. *Zoological Research*, **36**(4): 183−222.

Zhao F, Guo X, Wang Y, Liu J, Lee WH, Zhang Y. 2014a. Drug target mining and analysis of the Chinese tree shrew for pharmacological testing. *PLoS One*, **9**(8): e104191.

Zhao F, Lan X, Li T, Xiang Y, Zhao F, Zhang Y, Lee WH. 2018a. Proteotranscriptomic analysis and discovery of the profile and diversity of toxin-like proteins in centipede. *Molecular and Cellular Proteomics*, **17**(4): 709−720.

Zhao F, Lan XQ, Du Y, Chen PY, Zhao J, Zhao F, Lee WH, Zhang Y. 2018b. King cobra peptide OH-CATH30 as a potential candidate drug through clinic drug-resistant isolates. *Zoological Research*, **39**(2): 87−96.

Zhao F, Yan C, Wang X, Yang Y, Wang G, Lee W, Xiang Y, Zhang Y. 2014b. Comprehensive transcriptome profiling and functional analysis of the frog (*Bombina maxima*) immune system. *DNA Research*, **21**(1): 1−13.