

Article

# A Combinational Strategy upon RNA Sequencing and Peptidomics Unravels a Set of Novel Toxin Peptides in Scorpion *Mesobuthus martensii*

Ning Luan<sup>1,†</sup>, Wang Shen<sup>1,†</sup>, Jie Liu<sup>2,3,†</sup>, Bo Wen<sup>2,3</sup>, Zhilong Lin<sup>2,3</sup>, Shilong Yang<sup>4,5,6</sup>, Ren Lai<sup>1,4,5,6,\*</sup>, Siqi Liu<sup>2,3,\*</sup> and Mingqiang Rong<sup>4,5,6,\*</sup>

<sup>1</sup> Life Sciences College of Nanjing Agricultural University, Nanjing 210095, Jiangsu, China; 15996268076@163.com (N.L.); shenwangzs@163.com (W.S.)

<sup>2</sup> BGI-Shenzhen, Shenzhen 518083, China; liujie8@genomics.cn (J.L.); wenbo@genomics.cn (B.W.); linzhilong@genomics.cn (Z.L.)

<sup>3</sup> China National GeneBank-Shenzhen, BGI-Shenzhen, Shenzhen, Guangdong 518083, China

<sup>4</sup> Key Laboratory of Animal Models and Human Disease Mechanisms of Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Kunming 650223, Yunnan, China; yslzoology@163.com

<sup>5</sup> United Laboratory of Natural Peptide of University of Science and Technology of China & Kunming Institute of Zoology, Chinese Academy of Science, Kunming 650223, Yunnan, China

<sup>6</sup> Sino-African Joint Research Center, Chinese Academy of Science, Wuhan 430074, Hubei, China

\* Correspondence: rlai72@njau.edu.cn or rlai@mail.kiz.ac.cn (R.L.); siqiliu@genomics.cn (S.L.); rongmingqiang@mail.kiz.ac.cn (M.R.); Tel.: +86-871-6519-6202 (R.L.); +86-10-80485460 (S.L.); +86-871-6519-9086 (M.R.)

† These authors contributed equally to this paper.

Academic Editor: Eivind Undheim

Received: 22 August 2016; Accepted: 21 September 2016; Published: 5 October 2016

**Abstract:** Scorpion venom is deemed to contain many toxic peptides as an important source of natural compounds. Out of the two hundred proteins identified in *Mesobuthus martensii* (*M. martensii*), only a few peptide toxins have been found so far. Herein, a combinational approach based upon RNA sequencing and Liquid chromatography-mass spectrometry/mass spectrometry (LC MS/MS) was employed to explore the venom peptides in *M. martensii*. A total of 153 proteins were identified from the scorpion venom, 26 previously known and 127 newly identified. Of the novel toxins, 97 proteins exhibited sequence similarities to known toxins, and 30 were never reported. Combining peptidomic and transcriptomic analyses, the peptide sequence of BmKKx1 was reannotated and four disulfide bridges were confirmed within it. In light of the comparison of conservation and variety of toxin amino acid sequences, highly conserved and variable regions were perceived in 24 toxins that were parts of two sodium channel and two potassium channel toxins families. Taking all of this evidences together, the peptidomic analysis on *M. martensii* indeed identified numerous novel scorpion peptides, expanded our knowledge towards the venom diversity, and afforded a set of pharmaceutical candidates.

**Keywords:** scorpion; peptidomics; toxins; diversity

## 1. Introduction

Venomous animals are the most fascinating species, using poisons or venoms to capture preys or defend themselves. A large number of species belong to the poisonous animal kingdom, such as chordates (reptiles, fishes, amphibians, mammals), echinoderms (starfishes, sea urchins), mollusks (cone snails, octopi), annelids (leeches), nemertines, arthropods (arachnids, insects, myriapods), and cnidarians (sea anemones, jellyfish, corals) [1]. Animal venoms are a rich source of natural

compounds that have evolved to have high affinity and selectivity for a diverse range of biological targets. Snake venoms are known to contain 100 unique components while venoms from scorpions, spiders, and cone snails have recently been shown to contain from 300 to over 1000 unique molecular entities [2]. The task to survey the complete diversity of venom is not trivial due mainly to the vast complexity of animal venoms and low resolution of separation techniques. It is this amazing complexity and large number of peptides that pose the greatest challenge to venomics [3]. This task can be deeply approached with two different methods: proteomic/peptidomic and transcriptomic (genomic) approaches [4,5]. The proteomic approach is based on the final product (i.e., the proteins and peptides secreted into the venom), while the transcriptomic approach is based on the genomic recipe book of potential toxins. Combining proteomic and transcriptomic approaches together can be the best way to profile the venom toxin composition.

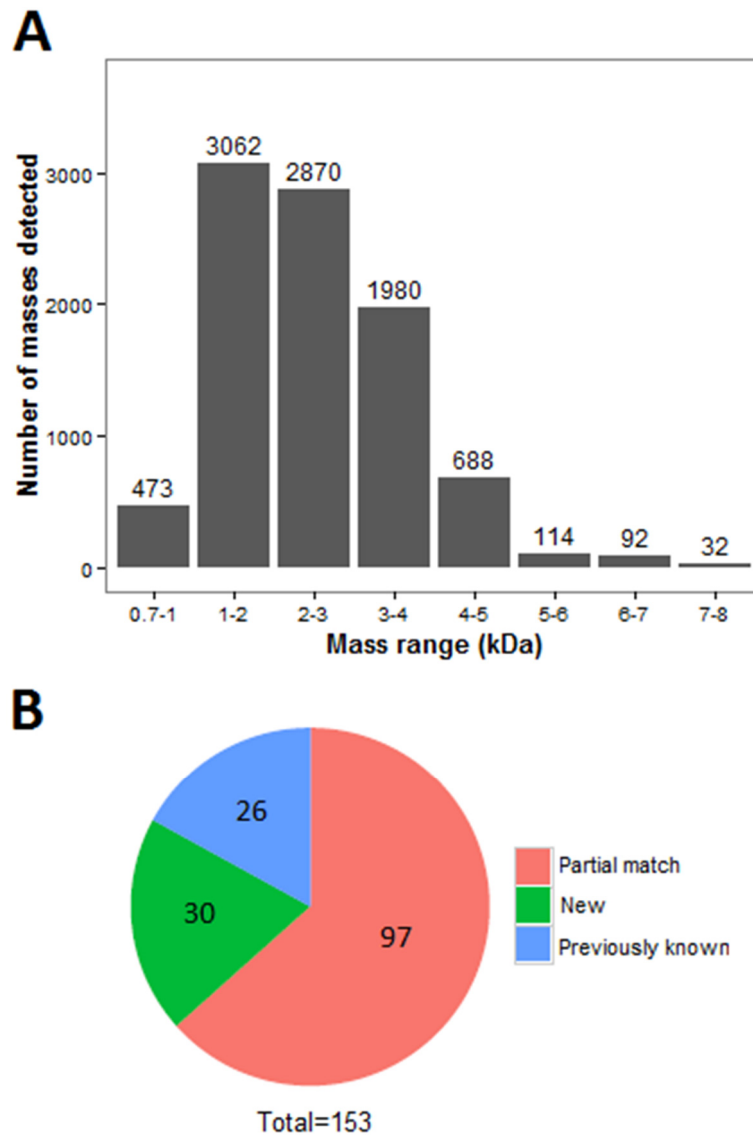
Approximately 2000 described scorpion species are widely distributed in different ecological niches around the world [6]. The combined methods (proteomic and transcriptomic) have been used to unravel the diversity of scorpion venoms. Complex components of Chinese scorpion venom, such as *Mesobuthus martensii* (also known as *Buthus Martensi Karsch*) [7], *Androctonus bicolor*, and *Heterometrus petersii*, were revealed by chromatography (2-DE, SDS-PAGE, and RP-HPLC) with mass spectrometry (ESI-Q-TOF MS and MALDI-TOF MS), and transcriptomics [8–11]. Two-hundred twenty-seven proteins unambiguously have been identified from *M. martensii*. Among those proteins, only 50 peptides are toxins and the others are enzymes or cell-associated proteins. In *Androctonus bicolor*, there are 103 novel putative peptides and only 16 were identified at the protein level.

*M. martensii* is one of traditional medicine in China. Fifty toxins have been identified in the venom of this species. The amount of known/identified peptides is far lower than the estimated amount, implying a need for deep analyses in venomics [12]. Here, peptidomic and transcriptomic approaches were employed to fully explore the deep diversity of the venom peptides from this scorpion.

## 2. Results

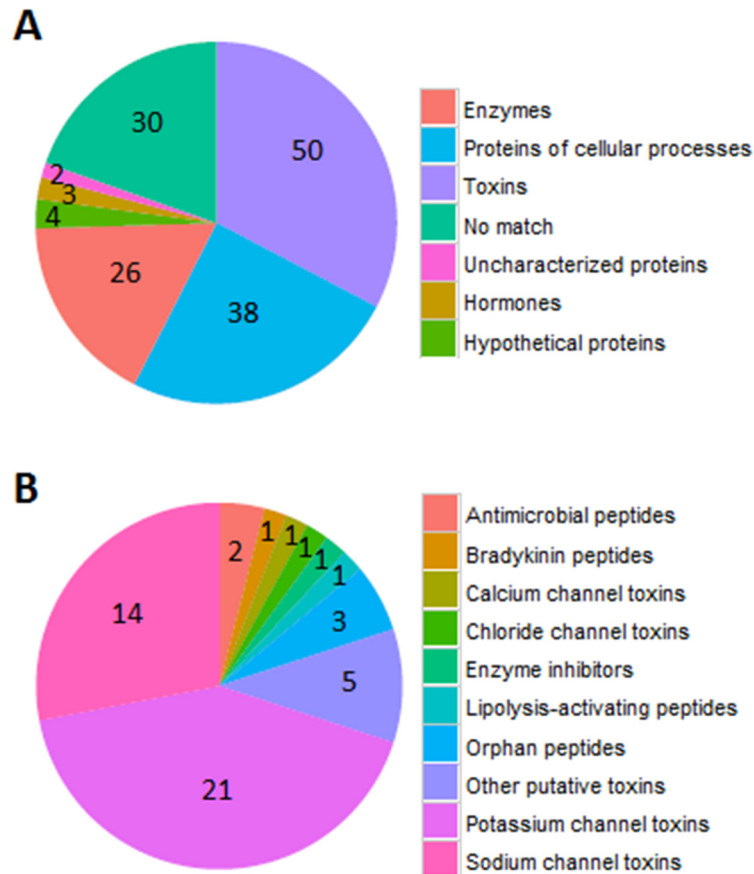
### 2.1. Venomics of the Scorpion Gland

De novo assembly of RNA-Seq data using Trinity software generated 79,804 transcripts. Proteomic analysis of the scorpion *Mesobuthus martensii* has been reported before and proteins from this species were isolated by 2-DE, SDS-PAGE, and RP-HPLC [8]. In order to focus on peptides, fractions with mass weight less than 10 KD (identified by MALDI-TOF MS) were analyzed to uncover the complexity of the venom. The small peptides were detected on a Prominence Nano-HPLC system coupled with Q-Exactive. As shown in Figure 1, most of the fragments (peptide precursor ions detected by LS/MS) have molecular weights from 1000 to 5000 Da, and only a few fragments are larger than 5000 Da. This study focused mainly on peptide toxins and larger proteins (> 10 KD) were filtered manually. The molecular weight of peptides ranges from 0.7 to 8 KD (Figure 1A). The number of peptides with molecular weight from 0.7 to 1, 1 to 2, 2 to 3, 3 to 4, 4 to 5, 5 to 6, 6 to 7, and 7 to 8 KD were 473, 3062, 2870, 1980, 688, 144, 92, and 32, respectively. This indicates that most of peptides are small fragments. Finally, 708 unique peptides were identified and 153 proteins were identified by the two analysis methods (Tables S1 and S2). Among the identified proteins, 26 toxins showed the same sequence to BLAST results, 97 proteins were partially matched, and 30 were novel venom proteins unknown before (Figure 1B).



**Figure 1.** Molecular weights and number of peptides (peptide precursor ions detected by LS/MS). (A) The molecular weights of fragments from scorpion venom. The x-axis represents the molecular weight ( $10^3$ ) and the y-axis are the amounts of fragments; (B) The classification of identified proteins by BLAST results. The known proteins are in blue, partially-matched proteins are red, and novel proteins are green.

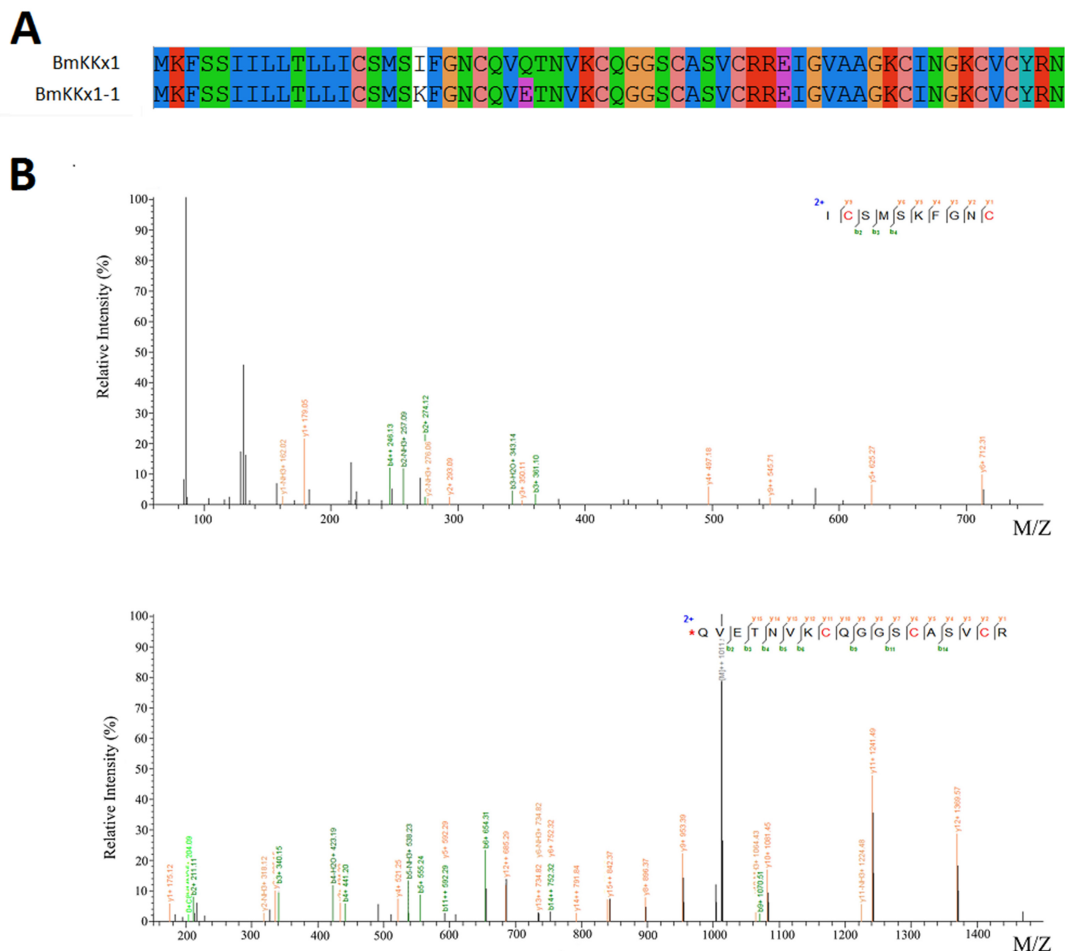
The 26 known and 97 partially-known proteins were categorized to different function groups. Among these proteins, 50 belong to toxins, 38 to cellular process proteins, 26 to peptide fragments of enzymes, three to hormones, four to hypothetical proteins, and two to previously uncharacterized proteins (Figure 2A). Toxins were the major components in venom, which were further divided into 11 function groups, including 21 potassium channel inhibitors, 14 sodium channel toxins, 2 antimicrobial peptides, 1 chloride channel toxin, 1 lipolysis-activating peptide, 1 bradykinin, three orphan peptides, and 5 putative toxins (Figure 2B).

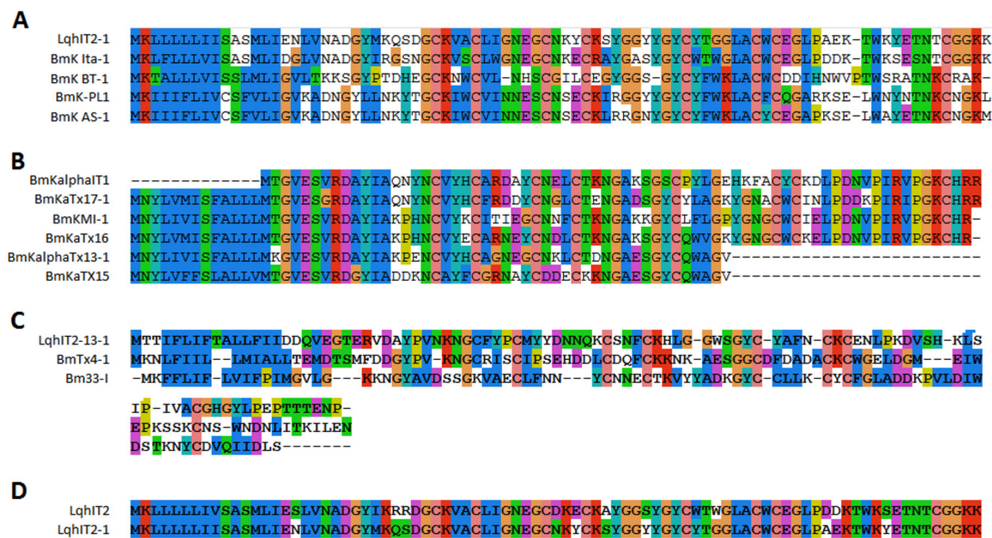


**Figure 2.** The function of identified toxin proteins. (A) The different functions of proteins, including toxins, cellular process proteins, enzyme, hypothetical proteins, hormones, uncharacterized and unknown proteins, are in different colors. (B) The number of different functional proteins from 50 toxins; 21 proteins target potassium channels, 14 proteins have activity on sodium channels, 5 proteins are putative toxins, three are orphan peptides, 2 are antimicrobial peptides, 1 is a calcium channel toxin, 1 is a lipolysis-activating peptide, 1 is bradykinin, 1 is a chloride channel toxin, and 1 is an enzyme inhibitor.

## 2.2. Identification of Peptides by LC-MS/MS Combined with RNA-Seq Data

Some scorpion toxins have been identified by LC/MSMS combined with RNA-Seq data. BmKKx1-1 showed only two different residues (the 18th (I to K) and 25th (Q to E) residues) in the mature fragments compared to BmKKx1. It was very interesting to find that the mature peptide of BmKKx1-1 begin from the 13th residue and formed four disulfide bridges (Figure 3). These results were different from the previous results in that BmKKx1 contained a 22-residue signal peptide and a 38-residue mature peptide with three disulfide bridges [13]. Combining peptidomic and transcriptomic methods was a good way to identify the initiation site of mature peptides.

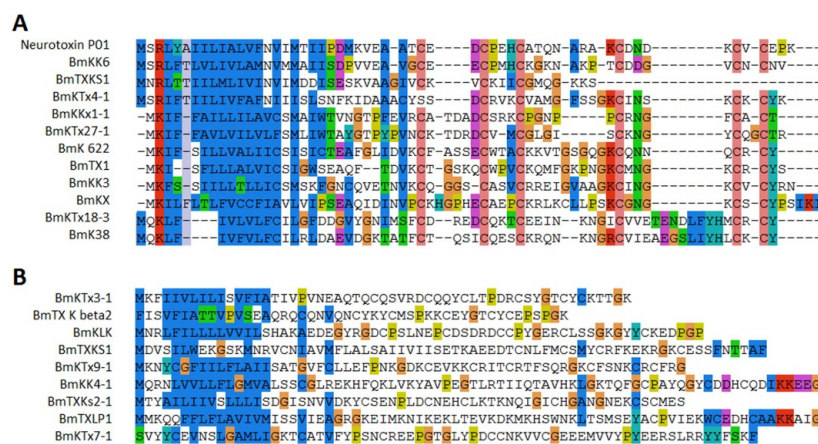




**Figure 4.** Alignment of sodium channel toxins from *M. martensii*. All sodium channel toxins were aligned by similarity. Sequences of sodium channel toxins belong to family 1 (A) and family 2 (B); (C) Three other sodium channel toxins identified from scorpion venom with low similarity were found; (D) Sequence alignment of LqhIT2-1 and LqhIT2.

#### 2.4. Potassium Channel Toxins

Scorpion venom contained many toxins targeting potassium channels. Twenty-one inhibitors were detected in our study, including 11 known and 10 novel peptides. Most of them were medium-long chain peptides with six cysteines forming three disulfide bridges, while two toxins (BmTXKbeta2, BmKlK) were previously-known long chain potassium channel toxins. Four known (BmTx3B, BmTX1, BmKK3 and BmKX) and two novel (BmKKx1-1 and BmKTx27-1) potassium channel toxins exhibited high similarity to each other and were considered as  $\alpha$ -KTxs (Figure 5A). The novel potassium toxin BmKKx1-1 was a mutant to BmKKx1 from the same species, but BmKTx27-1 showed 86.36% similarity to toxin Tx771 from another scorpion species *Buthus occitanus Israelis* [21]. Three toxins (BmKK6, neurotoxin P01 and BmKTx4-1) showed high similarity in the signal peptide and was classified to the same family. Toxin BmKTx4-1 was a novel toxin and exhibited 96.55% similarity to pMeKTx4-1 from *M. eupeus*. It is very interesting that the other five novel toxins (BmKTx7-1, BmKTx30-1, BmKTx18-3, BmKTx9-1 and BmKTx3-1) also showed similarity to the toxins from *M. eupeus* (Figure 5).



**Figure 5.** Alignment of potassium channel toxins from *M. martensii*. (A) Twelve potassium channel toxins aligned based on sequence similarity and all toxins had six cysteines; (B) Other identified potassium channel toxins.

## 2.5. Other Functions

Scorpion venom also contained many other functional toxins. One calcium channel toxin (BmCa1), one chloride channel toxins (Bm12-b), two antimicrobial peptides (marcin-18, BmKn1), and one bradykinin-potentiating peptide (BmKbpp) were detected from *M. martensii* (Table 1) [10,22]. A novel lipolysis-activating peptide (BmLVP2-alpha) has large variation in the signal peptide compared to BmKBT, but only four residues differed in the mature peptide between two peptides [23]. A novel inhibitor of nitric oxide (NO) synthase Makatoxin-2-1 showed amino acid sequence similarities to MKTXII. Three orphan peptides (BmOp-1, BmOp-2 and BmOp-3) exhibited high similarity to AbOp-12 (95%), AbOp-25 (87%), and AbOp-3 (84%) from *Androctonus bicolor* (Table 1). Five peptides (BmPt1-BmPt5) were considered as potential toxins. Those toxins showed similarity to other scorpion species, except BmPt3 and BmPt4. BmPt1 showed 99% identity to venom toxin meu VNP2 from *M. eupeus*. A high identity (91%) was found between BmPt2 and venom protein VP5 from *Odontobuthus doriae*. BmPt5 showed 67% identity to U1-buthitoxin-Hj1b from *Hottentotta judaicus* (Table 1).

**Table 1.** Scorpion toxins of other function.

Name of Toxin	Identity	Evalue	Function	Sequence
BmCa-1	100	2.76E-41	Calcium channel toxin	MNTFVVVFLLLTAILCHAEHALDETARGCNRL NKKCNSDGDCCRYGERCISTGVNYCRPDFGP
Bm12-b	100	3.14E-36	Chloride channel toxin	MKFLYGIVFIALFLTVMFATQTDGCGPCFTTD ANMARKCRECCGGNGKCFGPQCLCNRE
Marcin-18	100	9.09E-48	Antimicrobial peptide	MQFKKQLMVIFLAYFLVNESEAFFGH LFKLATKIIPSLFRRKNQRSRSMKRDLE NLFDPYQRNLELDRLKQLP
BmKn1	100	8.51E-43	Antimicrobial peptide	MKSQTFLLFLVLLLAISQSEAFIGAVAGLLSKIFGK RSMRDMDTMKYLYDPSLSAADLKTQLMENY
BmOp-1	95.38	3.48E-34	Orphan peptide	MKPRTFLVFLVLCILVMDAVVARRCGGNRRRVI KMVRKLRPLVKVIRIIRKIRGRAPCGGSRLT
BmOp-2	86.59	9.63E-46	Orphan peptide	MNKSLIILVAVVLSLWSAAEARQRRQVYC SNSPKGCQVRCWSWFGKREIADTVEFPGLK DEDDEAILKLAIEELKNE
BmOp-3	83.76	5.15E-59	Orphan peptide	MKTTGILFVCAIISLYLEAESASIHSEL QSNFHKLRRKQIHGRFGDYIPRGDLAQG KVSTGVEMFFPVDDDDGEETRDTRDDDDGEQ TSREDDDDGLTSRRAAPRERHVNCNKRHY
BmLVP2-alpha	85.71	8.46E-59	Lipolysis-activating peptide	MMTLALFGIIFLFLSLIGSIHGDDPEPGDYPTNVYGNK YYCTILGENEYCRKICKLHGVSYGYCNSRCWCEY LEDKDVTIWNAVKNHCTNLLYPNGK
Bpp BmK3	100	1.06E-44	Bradykinin-potentiating peptide	MNKKTLVIFVTVMLIVDEVNSFRFGSFLKVKWKS AKKLRSGKQQLLDYANKVLNGPEEEAAPAERRR
Makatoxin-2-1	90.91	6.3329E-39	Inhibitor of nitric oxide (NO) synthase	MNYLIVISFALLMTGVESGRDAYIADSENCTYFCGA NPYCNDLCTKNGAKSGYQWVWGKYGNCGCW
BmPt2	91.25	1.62E-47	Venom protein	MKSAVIVAITLCLFLNLYANAQKDCSL PVDTRGRGKGFWRYYYNKNKSTCESFI YGGVGGNKNNFLNIENCKKICKAKNC
BmPt3	93.68	3.94E-59	Putative toxin-like peptide	MVKMQVIFIAFIACSMVYGDSLSPWNEG DTYYGCQRQTDFECNKICKLHLATGGGSC YQPFVFKLCRCIGIDYDNSFFFGALEKQCPKRE
BmPt4	78.75	1.77E-31	Venom peptide	AGRFDVSAAGAGRFDVLEAGAGRFDV ASAAGAGRFDVSAAGAGRFDPTLEAG AGRFNPALAEAGAGRFDPAEAEAGAGRF
BmPt5	66.67	4.99E-33	Venom toxin	MFFLLILLALIPMIKLEKECAMESLQGE FPRKDKGYLYECGEDIKCKICKKNHGRS ENDGKCCYGNCFCTGLHGKNIKHQ

## 2.6. Novel Venom Peptide

Peptide/protein sequences were searched on the NCBI nr database using the online BLAST with an *e*-value cutoff set to  $<10^{-5}$ . Thirty peptides showed no similarity to any known proteins, implying those peptides were novel. Both long- and short-chain peptides were found in this new group and may

represent the sequence diversity of unknown peptides of scorpions. BmVt18 contained 60 residues and the mature peptide sequence was PSHRILYNIKIGLYCFKFLYRRKTAYSIIICNCYLNTRLRLR. Four cysteines were found in the mature sequence, indicating this peptide may be a potential toxin. Most of the novel venom peptides did not resolve in full sequences, which may be due to the low abundance in venom (Table 2).

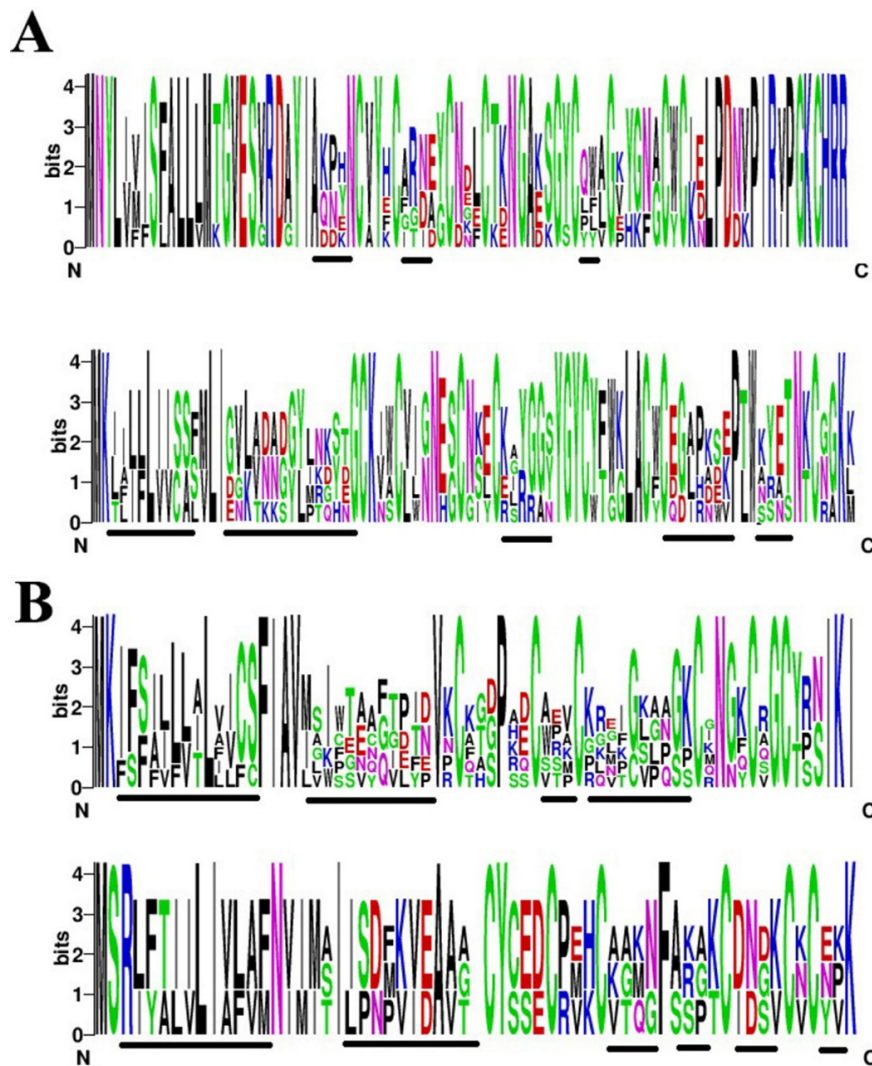
**Table 2.** New peptides/proteins with unknown function.

Name of Toxin	Sequence
BmVt1	KIVVKYVKQRIADHIPRRNFKIHKRKFANNIYDSCII
BmVt2	AGRFDPALAAGARGYDPYYVGGPLGETKERIYIDE
BmVt3	MRALTIFGVIFACLLLFRTAVADHDGDDHNEELEEDVKQEKKILGPYSNMGRRTSCKIWIWQMEEAQRPS
BmVt4	IIHQIFQILAALEKLNLMTKFLYLMIKAMIFLVEKLVVVEQNNVTWIHVRIMAIVMICGINTFVNVDHFLFLE
BmVt5	MEKILLNQFIVFIVGIFGMRDVNSEYVSENFVKFPNYISENKDIQNKNNARNLIPDIKSSLKETETAESKRRNV DLINFSEKVKLSDESKKSFFERTAVIHVLNKNENYDKNSKNTLNENLQPSSSNINKLKVTEISNNMNL
BmVt6	KVLEEEAGEGEEEEENWQLMLQISVLFYHHRDKIVYIYDL
BmVt7	ILSNKPLKSSRNDELHKLFLNKNNDNQPIKEFLGNVEKNNILSFPQNILPVYHPDFANMFDEISFQTNTVP
BmVt8	KQEVDVLIQHRQQEQDVLTYQQQERDVLQIH
BmVt9	MKVLVLLLVVALVAAALADKRDNIKRDYGGVGGYGGGGRGGFGGGGRGGFGG GYGGGRGGGIGGGRGGFGGGRGGFGGGRGGFGGGYGGGKYGGKYGGK
BmVt10	TMTFNRYSLHHFRPCYVFIFLMSTIFWTLTNAEDNIKSTIDRDKRAPQLYSFGLGKKSYPVSDSDNIES
BmVt11	TRSSTPERFSTRDLISSFSSASSVSSRCSAMSASEVCCCSRAMSWANWVRRRSSSPTRSLARCSSRSSI
BmVt12	MKCILINFLSLMLLANYSFGNEEQKKNMRALYPRKFYVERLRNDIVEEPFKKKNAYVTGG
BmVt13	MRVDNSAVLLVLIYLAKLSLSAPVDHGKEENPDPSFNIDVEYK NYLEEVARLLENDEDFINQIKDVTHTSKDFQPEDISF
BmVt14	MRVVMLIFFLLCQALNLALVLQSEINFYDSVETDLRSAELNLDIWNED RKISRFLIEEEEISREISKYENAREISKRFLILVGTWCGAGNKA
BmVt15	RRGTAEIREGRFGISRRAEEADLRREPRYQPVYQHEVQQDSYQGNQDSLQNNAAQQRSHYQTYDQQLGR
BmVt16	MKILSFIPVFLCLGILNISCAENDLKRILNTSDEDELEKRWPK
BmVt18	DPPECTCPHGKFKFDHGLMKCIAGPEINCDSEIEGKGTWTAFCHEENMVPCCKHNIG DPPECTCPHGKFKFDHGLMKCIAGPEINCDSEIEGKGTWTAFCHEENMVPC
BmVt21	KDVNNLTYLIKKKLLEIIPKYKSPSIFSKMPIFPHENKESLFPKYKLYQ
BmVt22	IILVGYCVKVMEDLVIAIFGNWHVGMIFIGFQLGLERPTNVVQNNTSIYVTF
BmVt23	MIKENMKWLIYLLIPLSHSVSSREGYRLNEQETYAADITPPHPQKHPYWRRLWENQPPEQKDL YQNSYKNPFYLLDRKQLKNDDEFFELPNQAYLPLKEVGTNNRGADKIHSKSKIPEGHITVK
BmVt25	MKVVLVTFLCILLLSNQNGLNDAGKKEIQAVYRRKAYADPAKNDDVEIVDHNFFRFRSEELN
BmVt26	MKQFMFYFGIVIFATLLANRDVMAQNNKLVILPGFIPEITQK
BmVt27	NSGLSLFGMKIIGYIFGFALICSLPNQNVCHENEGQKRKIDAENDDLKQRVFPRVRFSLSDEEKRLQQR LSPAVLASLTGLEKQRLSRETLEDFSWNTEKRFDAIFGSLPDDTKERIDSAAALASLAENEERFYSTT
BmVt29	LLIWSLLPASPPSTKCVVFFMPPLGEDSLKGQRKLFATLKFPLTV
BmVt30	VSPPTPDAILNELKLRDGRIVEFVHQLVSVLPLRAMVFNELTPRSSELNLDISNMRRRSLQDYELK

2.7. Amino Acid Variability in Peptide Toxins

Peptide toxins from scorpion were rich in sodium and potassium channel toxins. Two sodium channel toxin families and two potassium channel toxin families were found in this work. Although the toxins belong to the same families, the sequences seemed variable in many regions. In order to analyze the diversity of toxins, we used WebLogo, a web-based application designed to generate sequence logos to compare conservation and variety of toxin families. As showed in the Figure 6A, the first sodium channel toxin family had high conservation except for charged residues in the middle of their peptide sequences. Another sodium channel toxin family exhibited a diversity of signals and propeptides. Three highly variable positions (marked with underline) containing hydrophobic and charged amino acid (Figure 6A). Scorpion venom was very rich in potassium channel toxins. As shown in Figure 6B, two families of potassium channel toxins exhibited a high variety in signal peptides and mature peptides, except six cysteine residues (Figure 6B).





**Figure 6.** Sequence logos of toxins from scorpion. (A) Sodium channel toxins from family 1 (upper) and family 2 (lower); (B) Potassium channel toxins from Figure 5A. Each logo consists of stacks of letters and the overall height of each stack indicates the sequence conservation at that position (measured in bits).

### 3. Discussion

Scorpion venom contains a larger number of peptide toxins, which show diversity in structure and biological activity. Hundreds of proteins have been reported, including the inhibitors of the potassium channel, sodium channel, and chloride channel, antimicrobial peptides, lipolysis-activating peptide, and bradykinin [8]. Such a wide range of peptides and proteins with different biological functions make scorpion venoms a valuable source of new compounds both for use in basic research and for the development of new medicines [24]. It also stimulates us to thoroughly investigate the diversity of the scorpion toxins of *M. martensii*.

In the past years, transcriptomic analysis was the most useful method to identify the sequences of peptide toxins from venom glands. The signal peptides, propeptides, and mature peptides of those toxins were determined by bioinformatics. BmKKx1 was first found from the cDNA library and showed 84% identity to BmTx3 [25]. It was thought that the mature sequence of BmKKx1 was QVQTNVKCQGGSCASVCRREIGVAAGKICINGKVCYRN. High accuracy and throughput mass spectrometry provides a powerful platform to investigate the trace peptide. In this study, a fragment (ICSMSKFGNC) of a novel peptide BmKKx1-1 was determined at the protein level by mass spectrometry. This result indicates that BmKKx1-1 should start at the 13th residue and contain eight

cysteines. Thus, peptidomics combined with transcriptomics, may be an accurate method to obtain the correct mature sequence of undetermined toxins.

Thousands of peptides fragments were found by mass spectrometry, and the molecular weights of most fragments ranged from 1000 to 5000 Da. The total number of peptides/proteins identified in this work was 153, indicating that some of the toxins may contain more than one fragment, such as BmKKx1-1 (Figure 2). Peptidomic methods could unsuspectedly determine partial sequences of peptides, but it seems difficult to get the full sequence of long toxins. Peptidomics combined with transcriptomics may be a good way to obtain the correct and full sequence.

Although larger proteins were eliminated in the process, some proteins, such as those involved in cellular processes, enzymes, and hormones, were still identified in the scorpion venom. This may be caused by alternative cleavage of the larger proteins and the fragments existed in the venom. Among the identified toxins, 26 toxins, including sodium channel, potassium channel, and calcium channel toxins, have been detected before. The other 127 proteins were identified for the first time. The novel toxin BmKK4-1 showed 98% similarity to the known scorpion toxins BmKK4, and only one residue is different between the two toxins [26]. The similarity of other novel 96 toxins to known toxins ranged from 42% to 98%. Most of the similar toxins were identified from the same scorpion species, *M. martensii*. However, BmTX4-1 exhibits 92% similarity to BmTX4 from *M. eupeus*, and four other potassium channel toxins (BmKTx30-1, BmKTx18-3, BmKTx9-1 and BmKTx3-1) are also proved to show amino acid sequence similarities to toxins from *M. eupeus* [27], suggesting that the two scorpions had a closer relative relationship. The novel toxins found increase the diversity of the scorpion toxins from *M. martensii*. The highly-conserved toxins (LqhIT2-1 and LqhIT2, BmTX4-1 and BmTX4, BmKTx30-1 and BmKTx30, BmKTx18-3 and BmKTx18, BmKTx9-1 and BmKTx9, and BmKTx3-1 and BmKTx3) from *M. martensii*, *M. eupeus*, and *Leiurus quinquestriatus hebraeus* may provide an example that those toxins under the heavy constraints of negative selection [28].

Highly-conserved and variable regions were obviously found among 24 toxins, which belong to two sodium channel and two potassium channel toxin families. Cysteines are highly-conserved in all toxin families, and other residues are varied in mature peptides. Sodium channel toxins are more conserved than potassium channel toxins in mature regions. The first sodium channel toxin family only shows small changes in three short regions (Figure 6A), while most of the residues in the mature region of the potassium channel toxin family are variable. Those results imply high evolutionary changes and diversity in scorpion venom.

#### 4. Conclusions

In summary, we used peptidomics combined with transcriptomics to uncover the diversity of toxins from scorpion *M. martensii*. Our data indicated that at least 127 novel proteins were found in this scorpion and broadened our knowledge about the diversity of venom. The greatly diverse toxins from *M. martensii* will provide candidates for medical pharmaceuticals.

#### 5. Materials and Methods

##### 5.1. Venom Collection

Adult *M. martensii* scorpions (both sexes,  $n = 100$ ) were purchased from Shandong Province, China. Crude venom of the scorpion was collected manually by electrical stimulation of the telson using a 25–50 V and 15–40 Hz pulse current. Protease inhibitor cocktail (Roche, Basel, Switzerland) was added to crude venom according the manufacturer's instructions [29]. Crude venoms were stored at  $-80\text{ }^{\circ}\text{C}$  until further use.

##### 5.2. Transcriptome Sequencing

Total RNA was extracted from the venom glands of 10 scorpions using Trizol reagent (Invitrogen, Carlsbad, CA, USA) and purified using RNeasy Animal Mini Kit (Qiagen, Valencia, CA, USA)

according to the manufacturer's instructions. RNA integrity number (RIN) values were measured for all samples. The RIN of *M. martensii* was 7.7.

mRNA was isolated from the total RNA using poly-T oligo-attached magnetic beads (Invitrogen, Carlsbad, CA, USA) and was fragmented into short fragments (200–700 bp) using divalent cations under elevated temperature. Short fragments were purified and resolved with EB buffer (Qiagen, Valencia, CA, USA) for end reparation and single nucleotide A (adenine) addition. After that, the short fragments were connected with adapters. Suitable fragments were selected as templates for PCR amplification and abundance normalization. During the QC steps, an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) and an ABI StepOnePlus Real-Time PCR System (ABI, Waltham, MA, USA) were used in quantification and qualification of the sample library. At last, the library was sequenced using Illumina HiSeq™ 2000 (Illumina, San Diego, CA, USA). After sorting, cleaning and trimming of the reads, sequence assembly was carried out using Trinity (trinityrnaseq\_r2012-06-08) [30]. The FASTAQ data has been deposited to NCBI (accession: PRJNA341817).

### 5.3. Venom Sample Preparation

Crude venom was suspended in 8 M urea in 0.1 M Tris-HCl (Sigma, St. Louis, MO, USA), pH 8.5, and centrifuged at 16,000 g for 10 min at 4 °C. Protein concentrations were detected using Bradford reagent (Sigma, St. Louis, MO, USA). Denatured proteins and peptides were reduced with 10 mM DTT (Amresco, Solon, OH, USA) at 56 °C for 1 h. After cooling to room temperature, the proteins were alkylated with 55 mM IAA (Sigma, St. Louis, MO, USA) in the dark at room temperature for 45 min [31].

### 5.4. Solid-Phase Extraction of Venom Peptides

The alkylated venom solution was diluted with 8 M urea to 1 mL, and then fractionated in a Strata-X C18 (Phenomenex, Torrance, CA, USA) column previously conditioned with methanol. After loading the venom solution the column was washed with 0.1% FA (J&K, Beijing, China) in 5% ACN (Thermo Fisher scientific, Waltham, MA, USA) and eluted with 80% ACN. The eluates were dried in a SCANVAC (LaboGene, Lyngø, Denmark) concentrator. The dried venom eluates were stored at –20 °C for further analysis.

Venom peptide concentrations were determined using a NanoDrop A280 system (Thermo Fisher Scientific, Waltham, MA, USA), and peptide amounts in the eluates were estimated based on absorbance at 280 nm using the nanodrop system (Thermo Scientific) [32]. HILIC separation was performed on a 20 AD HPLC system (Shimadzu, Tokyo, Japan) using ZIC®-HILIC 250 × 4.6 mm, 5 µm, 200@PEEK HPLC Column SeQuant(TM). Briefly, 200 µg mixed peptides were diluted with buffer A and loaded onto the HILIC column at 0.5 mL/min. Buffer A was composed of 80% can and 0.1% TFA (J&K, Beijing, China), and buffer B was composed of 0.1% TFA. After a 10 minute post injection hold at 100% A, peptides were eluted with a gradient from 0 to 40% B in 10 min followed by 40% to 90% B in 2 min and a 10 min hold to end. The absorbance at 214 nm was monitored, and a total of 10 fractions were collected along the gradient. Collected fractions were dried in a SCANVAC concentrator and stored at –20°C until use for nano LC-MS/MS analysis.

### 5.5. Nano LC-MS/MS Analysis

Liquid chromatography and mass spectrometry were performed on a prominence nano-HPLC system (Shimadzu, Tokyo, Japan) coupled with Q-Exactive (Thermo Fisher Scientific, Waltham, MA, USA). The peptides were separated by nano-LC on an in-house packed Ultimate XB-C18 column (3 µm, 12 cm × 75 µm, Welch Materials) at a flow rate of 300 nl/min. Each fraction was dissolved in 0.1% FA, then injected and eluted using a gradient of 5%–30% solvent B (95% acetonitrile, 0.1% FA) over 40 min. The mass spectrometry was operated in data-dependent mode, automatically switching between MS and MS2 acquisition. Survey full-scan MS spectra ( $m/z$  350–1800) were acquired in the Orbitrap with a resolution of 70,000. The 20 most intense ions were sequentially isolated and fragmented by high energy dissociation (HCD). Peptides with unassigned charge states, as well as less than +2, or more

than +6, were excluded from fragmentation. Fragment spectra were recorded in the Orbitrap mass analyzer with resolution of 17,500. The dynamic exclusion was enabled with repeat count two, and an exclusion duration 8 s.

### 5.6. Peptide Identification and Bioinformatic Analysis

The raw MS/MS data were converted into MGF format by Proteome Discoverer 1.2 (Thermo Fisher Scientific, Waltham, MA, USA). The assembled transcript sequences were translated to protein sequences using PGA with the six frame method as the searching database [33]. The MS/MS spectra were searched by Mascot (v2.3.02, MatrixScience; London, UK) against the database. Mascot parameters were set as follows: no enzyme selected; fixed modifications of Carbamidomethyl (C); variable modifications consisting of Gln->pyro-Glu (N-term Q) and Oxidation (M); peptide charge, 2+, 3+, and 4+; 10 ppm of peptide mass tolerance; 0.05 Da of fragment mass tolerance. The automatic Mascot decoy database search was performed. The Mascot results were processed by IQuant [34] utilizing MascotPercolator to re-score the peptide spectrum matches (PSMs) [35], and false discovery rates (FDR) were obtained using the target-decoy strategy with a peptide-level FDR equal to or less than 0.01. Finally, the identified peptide sequences were assembled into a set of confident proteins using the Occam's razor approach implemented in IQuant.

Annotation of identified protein sequences was searched against the NCBI NR database with an *e*-value cutoff set to  $<10^{-5}$  for comparison using the online Blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), and optimized manually. The resulting gene superfamily sequences were further aligned using the ClustalW alignment tool in the MEGA package (version 6.06) [36]. Aligned toxin sequences were used to create sequence logos by WebLogo (version 2.8.2, University of California, Berkeley, CA, USA, 2016) [29,30,37,38].

### 5.7. Ethics Statement

The Animal Care and Use Committee of The Kunming Institute of Zoology, Chinese Academy of Sciences, reviewed and approved all of the procedures used in this study (2015-249). All experiments were performed according to good practices of laboratory animal management.

**Supplementary Materials:** The following are available online at [www.mdpi.com/2072-6651/8/10/286/s1](http://www.mdpi.com/2072-6651/8/10/286/s1) (Tables S1 and S2).

**Acknowledgments:** This work was supported by NSFC (31260208, U1132601, 31200590, 81573320), CAS (SAJC201308, SAJC201606) and Yunnan Province (2011CI139, 2012BC009, P0120150010), Youth Innovation Promotion Association of CAS, Jiangsu Province (Q0201600440), the International Science & Technology Cooperation Program of China (2014DFB30020), Chinese National Basic Research Programs (2014CBA02002-A, 2014CBA02005) and the National High-Tech Research and Development Program of China (2012AA020202).

**Author Contributions:** Participated in research design: Mingqiang Rong, Bo Wen, Ren Lai, Siqu Liu. Conducted experiments: Ning Luan, Wang Shen, Bo Wen, Jie Liu, Zhilong Lin, Shilong Yang. Performed data analysis: Mingqiang Rong, Bo Wen, Ren Lai, Siqu Liu. Wrote or contributed to the writing of the manuscript: Mingqiang Rong, Bo Wen, Ren Lai, Siqu Liu.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Abbreviations

The following abbreviations are used in this manuscript:

LC MS/MS	Liquid chromatography–mass spectrometry/mass spectrometry
2-DE	Two-dimensional gel electrophoresis
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide gel electrophoresis
RP-HPLC	Reversed-Phase High-Performance Liquid Chromatography
ESI-Q-TOF MS	Electrospray ionization tandem quadrupole/orthogonal-acceleration time-of-flight mass spectrometer
MALDI-TOF MS	Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry

## References

1. Utkin, Y.N. Animal venom studies: Current benefits and future developments. *World J. Biol. Chem.* **2015**, *6*, 28–33. [[CrossRef](#)] [[PubMed](#)]
2. Quintero-Hernandez, V.; Jimenez-Vargas, J.M.; Gurrola, G.B.; Valdivia, H.H.; Possani, L.D. Scorpion venom components that affect ion-channels function. *Toxicon* **2013**, *76*, 328–342. [[CrossRef](#)] [[PubMed](#)]
3. Vetter, I.; Davis, J.L.; Rash, L.D.; Anangi, R.; Mobli, M.; Alewood, P.F.; Lewis, R.J.; King, G.F. Venomics: A new paradigm for natural products-based drug discovery. *Amino. Acids* **2011**, *40*, 15–28. [[CrossRef](#)] [[PubMed](#)]
4. Prashanth, J.R.; Lewis, R.J.; Dutertre, S. Towards an integrated venomics approach for accelerated conopeptide discovery. *Toxicon* **2012**, *60*, 470–477. [[CrossRef](#)] [[PubMed](#)]
5. Dutertre, S.; Jin, A.H.; Kaas, Q.; Jones, A.; Alewood, P.F.; Lewis, R.J. Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol. Cell. Proteom.* **2013**, *12*, 312–329. [[CrossRef](#)] [[PubMed](#)]
6. Cao, Z.; Yu, Y.; Wu, Y.; Hao, P.; Di, Z.; He, Y.; Chen, Z.; Yang, W.; Shen, Z.; He, X.; et al. The genome of mesobuthus martensii reveals a unique adaptation model of arthropods. *Nat. Commun.* **2013**, *4*, 2602. [[CrossRef](#)] [[PubMed](#)]
7. Qi, J.-X.; Zhu, M.-S.; Lourenço, W.R. Redescription of mesobuthus martensii martensii (karsch, 1879)(scorpion: Buthidae) from China. *Revista Ibérica Aracnología* **2004**, *10*, 137–144.
8. Xu, X.; Duan, Z.; Di, Z.; He, Y.; Li, J.; Li, Z.; Xie, C.; Zeng, X.; Cao, Z.; Wu, Y.; et al. Proteomic analysis of the venom from the scorpion mesobuthus martensii. *J. Proteom.* **2014**, *106*, 162–180. [[CrossRef](#)] [[PubMed](#)]
9. Zhang, L.; Shi, W.; Zeng, X.C.; Ge, F.; Yang, M.; Nie, Y.; Bao, A.; Wu, S.; E, G. Unique diversity of the venom peptides from the scorpion androctonus bicolor revealed by transcriptomic and proteomic analysis. *J. Proteom.* **2015**, *128*, 231–250. [[CrossRef](#)] [[PubMed](#)]
10. Xu, J.; Zhang, X.; Guo, Z.; Yan, J.; Yu, L.; Li, X.; Xue, X.; Liang, X. Short-chain peptides identification of scorpion buthus martensi karsch venom by employing high orthogonal 2D-HPLC system and tandem mass spectrometry. *Proteomics* **2012**, *12*, 3076–3084. [[CrossRef](#)] [[PubMed](#)]
11. Ma, Y.; Zhao, Y.; Zhao, R.; Zhang, W.; He, Y.; Wu, Y.; Cao, Z.; Guo, L.; Li, W. Molecular diversity of toxic components from the scorpion heterometrus petersii venom revealed by proteomic and transcriptome analysis. *Proteomics* **2010**, *10*, 2471–2485. [[CrossRef](#)] [[PubMed](#)]
12. Ma, Y.; He, Y.; Zhao, R.; Wu, Y.; Li, W.; Cao, Z. Extreme diversity of scorpion venom peptides and proteins revealed by transcriptomic analysis: Implication for proteome evolution of scorpion venom arsenal. *J. Proteom.* **2012**, *75*, 1563–1576. [[CrossRef](#)] [[PubMed](#)]
13. Zeng, X.C.; Luo, F.; Li, W.X. Molecular dissection of venom from chinese scorpion mesobuthus martensii: Identification and characterization of four novel disulfide-bridged venom peptides. *Peptides* **2006**, *27*, 1745–1754. [[CrossRef](#)] [[PubMed](#)]
14. Gordon, D.; Karbat, I.; Ilan, N.; Cohen, L.; Kahn, R.; Gilles, N.; Dong, K.; Stuhmer, W.; Tytgat, J.; Gurevitz, M. The differential preference of scorpion alpha-toxins for insect or mammalian sodium channels: Implications for improved insect control. *Toxicon* **2007**, *49*, 452–472. [[CrossRef](#)] [[PubMed](#)]
15. Tan, Z.Y.; Mao, X.; Xiao, H.; Zhao, Z.Q.; Ji, Y.H. Buthus martensi karsch agonist of skeletal-muscle RYR-1, a scorpion active polypeptide: Antinociceptive effect on rat peripheral nervous system and spinal cord, and inhibition of voltage-gated Na(+) currents in dorsal root ganglion neurons. *Neurosci. Lett.* **2001**, *297*, 65–68. [[CrossRef](#)]
16. Zlotkin, E.; Gurevitz, M.; Fowler, E.; Adams, M.E. Depressant insect selective neurotoxins from scorpion venom: Chemistry, action, and gene cloning. *Arch. Insect Biochem. Physiol.* **1993**, *22*, 55–73. [[CrossRef](#)] [[PubMed](#)]
17. Gurevitz, M.; Karbat, I.; Cohen, L.; Ilan, N.; Kahn, R.; Turkov, M.; Stankiewicz, M.; Stuhmer, W.; Dong, K.; Gordon, D. The insecticidal potential of scorpion  $\beta$ -toxins. *Toxicon* **2007**, *49*, 473–489. [[CrossRef](#)] [[PubMed](#)]
18. Zhu, S.Y.; Li, W.X.; Zeng, X.C.; Liu, H.; Jiang, D.H.; Mao, X. Nine novel precursors of buthus martensii scorpion  $\alpha$ -toxin homologues. *Toxicon* **2000**, *38*, 1653–1661. [[CrossRef](#)]
19. Ji, Y.H.; Mansuelle, P.; Terakawa, S.; Kopeyan, C.; Yanaihara, N.; Hsu, K.; Rochat, H. Two neurotoxins (bmk i and bmk ii) from the venom of the scorpion buthus martensi karsch: Purification, amino acid sequences and assessment of specific activity. *Toxicon* **1996**, *34*, 987–1001. [[PubMed](#)]

20. Xiong, Y.M.; Lan, Z.D.; Wang, M.; Liu, B.; Liu, X.Q.; Fei, H.; Xu, L.G.; Xia, Q.C.; Wang, C.G.; Wang, D.C.; et al. Molecular characterization of a new excitatory insect neurotoxin with an analgesic effect on mice from the scorpion *buthus martensi karsch*. *Toxicon* **1999**, *37*, 1165–1180. [[CrossRef](#)]
21. Kozminsky-Atias, A.; Bar-Shalom, A.; Mishmar, D.; Zilberberg, N. Assembling an arsenal, the scorpion way. *BMC Evol. Biol.* **2008**, *8*, 333. [[CrossRef](#)] [[PubMed](#)]
22. Zhijian, C.; Yun, X.; Chao, D.; Shunyi, Z.; Shijin, Y.; Yingliang, W.; Wenxin, L. Cloning and characterization of a novel calcium channel toxin-like gene *bmca1* from chinese scorpion *mesobuthus martensii karsch*. *Peptides* **2006**, *27*, 1235–1240. [[CrossRef](#)] [[PubMed](#)]
23. Nie, Y.; Zeng, X.C.; Luo, X.; Wu, S.; Zhang, L.; Cao, H.; Zhou, J.; Zhou, L. Tremendous intron length differences of the *bmkb1* and a novel *bmkb1*-like peptide genes provide a mechanical basis for the rapid or constitutive expression of the peptides. *Peptides* **2012**, *37*, 150–156. [[CrossRef](#)] [[PubMed](#)]
24. Billen, B.; Bosmans, F.; Tytgat, J. Animal peptides targeting voltage-activated sodium channels. *Curr. Pharm. Des.* **2008**, *14*, 2492–2502. [[CrossRef](#)] [[PubMed](#)]
25. Huys, I.; Xu, C.Q.; Wang, C.Z.; Vacher, H.; Martin-Eauclaire, M.F.; Chi, C.W.; Tytgat, J. BmTx3, a scorpion toxin with two putative functional faces separately active on A-type K<sup>+</sup> and HERG currents. *Biochem. J.* **2004**, *378*, 745–752. [[CrossRef](#)] [[PubMed](#)]
26. Zhang, N.; Chen, X.; Li, M.; Cao, C.; Wang, Y.; Wu, G.; Hu, G.; Wu, H. Solution structure of BmKK4, the first member of subfamily  $\alpha$ -KTx 17 of scorpion toxins. *Biochemistry* **2004**, *43*, 12469–12476. [[CrossRef](#)] [[PubMed](#)]
27. Kuzmenkov, A.I.; Vassilevski, A.A.; Kudryashova, K.S.; Nekrasova, O.V.; Peigneur, S.; Tytgat, J.; Feofanov, A.V.; Kirpichnikov, M.P.; Grishin, E.V. Variability of potassium channel blockers in *mesobuthus eupeus* scorpion venom with focus on Kv1.1: An integrated transcriptomic and proteomic study. *J. Biol. Chem.* **2015**, *290*, 12195–12209. [[CrossRef](#)] [[PubMed](#)]
28. Sunagar, K.; Moran, Y. The rise and fall of an evolutionary innovation: Contrasting strategies of venom evolution in ancient and young animals. *PLoS Genet.* **2015**, *11*, e1005596. [[CrossRef](#)] [[PubMed](#)]
29. Yang, S.; Liu, Z.; Xiao, Y.; Li, Y.; Rong, M.; Liang, S.; Zhang, Z.; Yu, H.; King, G.F.; Lai, R. Chemical punch packed in venoms makes centipedes excellent predators. *Mol. Cell. Proteom.* **2012**, *11*, 640–650. [[CrossRef](#)] [[PubMed](#)]
30. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [[CrossRef](#)] [[PubMed](#)]
31. Duan, Z.; Cao, R.; Jiang, L.; Liang, S. A combined de novo protein sequencing and cDNA library approach to the venom analysis of chinese spider *araneus ventricosus*. *J. Proteom.* **2013**, *78*, 416–427. [[CrossRef](#)] [[PubMed](#)]
32. Tashima, A.K.; Zelanis, A.; Kitano, E.S.; Ianzer, D.; Melo, R.L.; Rioli, V.; Sant’anna, S.S.; Schenberg, A.C.; Camargo, A.C.; Serrano, S.M. Peptidomics of three bothrops snake venoms: Insights into the molecular diversification of proteomes and peptidomes. *Mol. Cell. Proteom.* **2012**, *11*, 1245–1262. [[CrossRef](#)] [[PubMed](#)]
33. Wen, B.; Xu, S.; Zhou, R.; Zhang, B.; Wang, X.; Liu, X.; Xu, X.; Liu, S. PGA: An R/Bioconductor package for identification of novel peptides using a customized database derived from RNA-Seq. *BMC Bioinform.* **2016**, *17*, 244. [[CrossRef](#)] [[PubMed](#)]
34. Wen, B.; Zhou, R.; Feng, Q.; Wang, Q.; Wang, J.; Liu, S. Iquant: An automated pipeline for quantitative proteomics based upon isobaric tags. *Proteomics* **2014**, *14*, 2280–2285. [[CrossRef](#)] [[PubMed](#)]
35. Brosch, M.; Yu, L.; Hubbard, T.; Choudhary, J. Accurate and sensitive peptide identification with Mascot Percolator. *J. Proteom. Res.* **2009**, *8*, 3176–3181. [[CrossRef](#)] [[PubMed](#)]
36. Kumar, S.; Nei, M.; Dudley, J.; Tamura, K. Mega: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.* **2008**, *9*, 299–306. [[CrossRef](#)] [[PubMed](#)]
37. Zhang, S.; Gao, B.; Zhu, S. Target-driven evolution of scorpion toxins. *Sci. Rep.* **2015**, *5*, 14973. [[CrossRef](#)] [[PubMed](#)]
38. Armon, A.; Graur, D.; Ben-Tal, N. ConSurf: An algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J. Mol. Biol.* **2001**, *307*, 447–463. [[CrossRef](#)] [[PubMed](#)]

