Structural and Functional Diversity of Peptide Toxins from Tarantula *Haplopelma hainanum* (Ornithoctonus hainana) Venom Revealed by Transcriptomic, Peptidomic and Patch-clamp Approaches

Yi-Ya Zhang1†, Yong Huang2†, Quan-Ze He1, Ji Luo1, Li Zhu1, Shan-Shan Lu1, Jin-Yan Liu1, Peng-Fei Huang1, Xiong-Zhi Zeng1*, Song-Ping Liang1*

Address:

†These authors contributed equally to this work.
1The Key Laboratory of Protein Chemistry and Developmental Biology of Ministry of Education, College of Life Sciences, Hunan Normal University, Changsha 410081, P. R. China
2State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, 20 Dong-Da Street, Fengtai District, Beijing 100071, China
* Corresponding authors. Tel.: +86 731 88872556; fax: +86 731 88861304.
E-mail addresses: xiongzhizeng@gmail.com (Zeng, X), liangsp@hunnu.edu.cn (S. Liang).

Running title: Structural and Functional Diversity of Peptide Toxins from *Ornithoctonus hainana*

Keywords: 454 sequencing; *Haplopelma hainanum*; spider venom; cysteine pattern; molecular diversity; evolution
Background: Spider venom is a complex mixture and contains over 400 venom peptides per species.
Results: Highly abundant of hypermutation, fragment insertion/deletion and variable post-translational modifications were observed in venom gland, and the highly diverse toxins exhibit diverse functions.
Conclusion: Spider toxins are abundant and variable, which work together generating diverse bioactivities.
Significance: Our study reveals the mechanism of structural and functional diversity of peptide toxins from *Ornithoctonus hainana*.

Abstract
Spider venom is a complex mixture of bioactive peptides to subdue their prey. Early estimates suggested that over 400 venom peptides are produced per species. In order to investigate the mechanisms responsible for this impressive diversity, transcriptomics based on second-generation high-throughput sequencing was combined with peptidomic assays to characterize the venom of the tarantula *Haplopelma hainanum*. The genes expressed in the venom glands were identified and the bioactivity of their protein products was analyzed using the patch-clamp technique. A total of 1,136 potential toxin precursors were identified that clustered into 90 toxin groups, of which 72 were novel. The toxin peptides clustered into 20 cysteine scaffolds that included between 4 to 12 cysteines, and 14 of these groups were newly identified in this spider. Highly abundant toxin peptide transcripts were present and resulted from hypermutation and/or fragment insertion/deletion. In combination with variable post-translational modifications, this genetic variability explained how a limited set of genes can generate hundreds of toxin peptides in venom glands. Furthermore, the intraspecies venom variability illustrated the dynamic nature of spider venom, and revealed how complex components work together generate diverse bioactivities that facilitate adaptation to changing environments, types of prey, and milking regimes in captivity.

Abbreviations: RP-HPLC, reverse phase high-performance liquid chromatography; MALDI-TOF-MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; TTX-S, tetrodotoxin-sensitive; TTX-R, tetrodotoxin-resistant; Nav channels, voltage-activated sodium channels; Kv channels, voltage-activated potassium channels; Cav channels, voltage-activated calcium channels; EGTA, ethylene glycol-bis (β-aminoethyl ether)-N,N,N’,N’-tetraacetic acid; TEA, tetraethylammonium; CCA, a-cyano-4-hydroxycinnamic acid; DRG, dorsal root ganglion; HEPES, N-hydroxyethyl piperazine-N-ethanesulfonic acid.

Introduction
Spiders are one of the most ancient venomous animals that have been roaming the earth for 300 million years and number nearly 40,000 species. Tarantulas are a group of often hairy and large arachnids that comprise more than 860 species. Like other venomous animals, these predators use a complex arsenal of venom to paralyze and kill their prey, and many of these toxins have proved to be invaluable tools for pharmacological studies of voltage-sensitive and ligand-gated ion channels (1-4). The vast majority of spider toxins are cysteine rich polypeptides and their properties and structures have been reviewed in detail (4-7). The molecular diversity of spider toxins has also been investigated and they appear to be based on a limited set of structural scaffolds such as the inhibitor cysteine knot (ICK) and disulfide-directed β-hairpin (DDH) whose cysteines form three to five characteristic disulfide bonds, which is a much smaller number than toxins from other species such as marine cone snails. The lack of a complete genome sequence may be responsible for our currently limited knowledge of the cysteine pattern diversity present in tarantula toxins, and improving this knowledge
will be challenging. Expressed sequence tags (ESTs) are short sequence reads derived from cDNA libraries that are useful tools for the identification of transcripts in species without a fully sequenced genome (8,9). In the past few years, all reports on tarantula toxin transcriptomics have utilized classical cloning and Sanger sequencing strategies (9). In previous work, 420 peptides were detected by mass spectrometry but few could be paired with peptide precursors identified from cDNA and genomic DNA sequencing (10). Furthermore, this limited data focused mainly on highly abundant and smaller toxin precursors, while less prevalent and longer gene sequences were largely ignored, which has proved a barrier to research on the molecular diversity and genetic mechanisms of toxin evolution (11). The relatively new 454 Life Sciences pyrosequencing technology has been successfully implemented in a number of species, including spiders (12-16). This approach provides a more comprehensive landscape of the transcriptomic content of venom glands and has improved technical capabilities that identify longer sequences, a wider range of sensitivity, and greater accuracy than traditional Sanger sequencing (17,18). The longer reads (>300 bp on average) generated with 454 pyrosequencing allows coverage of full-length (60–120 amino acids) precursors, which affords the direct identification of spider toxin precursors and avoids the errors inherent in assembling reads into contigs as typically required for other second-generation technologies that generate shorter read lengths (19). In this study, 1,136 toxin precursors were identified using 454 Life Sciences pyrosequencing and classified into 15 Classes (Class A-O). Sequence analysis revealed that extensive hypermutation and fragment insertion/deletion dramatically increased the molecular diversity of toxin transcripts, and this diversity was further enhanced by highly variable post-translational modifications (PTMs). The high intraspecies variability may explain the transcriptome differences between this work and previous work. Analysis of the bioactivity of 14 of the identified toxins against DRG ion channels using patch-clamping revealed functional diversity. In conclusion, variable peptide processing and selective expression explains how a limited set of gene transcripts can generate hundreds of toxin peptides in spider venom that have diverse activities and cooperate to subdue potential prey species.

Materials and Methods

Materials

Dithiothreitol (DTT), iodoacetamide, and trifluoroacetic acid (TFA) were obtained from Sigma (St. Louis, MO, USA). Acetonitrile (ACN) was a domestic product (chromatogram grade). Milli-Q H₂O was used for the preparation of all buffers. Other chemicals were analytical grade.

cDNA Library Construction and 454 sequencing

The tarantula spiders were collected from Hainan, China. Venom glands of Haplopelma hainanum were obtained 2 days after being milked by electrical stimulation, and ground to fine powder in liquid nitrogen. Total RNA was extracted with TRIZol (Invitrogen, Carlsbad, CA, USA) and used to construct a cDNA library. 5 μg of full-length double-stranded cDNA was then processed by the standard Genome Sequencer library-preparation method using the 454 DNA Library Preparation Kit (Titanium chemistry) to generate single-stranded DNA ready for emulsion PCR (emPCR™). The cDNA library was sequenced according to GS FLX technology (454/Roche).

Sequence assembly and alignment

Sequence reads were trimmed by excluding adapters and low-quality regions using the NGen module of the DNAStar Lasergene software suite. Subsequently, assembly was performed with SeqMan pro (DNASTAR, USA) using high stringency de novo transcriptome assembly (100% identity between reads with 50 nucleotide sequence overlap). Similar reads were assembled into contigs
using CLC Genomics Workbench 3 with its default parameters. All assembled contigs and rest reads were searched against the NCBI subset of the EST database with BLASTn. All these sequences were also searched with BLASTx in all 6 reading frames to the non-redundant protein database to found out their correct translation product. For both searches the blast version 2.2.25+ and an E-value threshold of $e \leq 10^{-3}$ with a bit score $>40$ was considered and recorded as a significant match for each query sequence. We analyzed the blast results and used a home-made PERL script to classify representative sequences into five categories (‘Toxin-like’, ‘Putative toxins’, ‘Cellular Proteins’, ‘Unknown function’ and ‘No Hit’).

**Gene Ontology (GO) annotations**

Functional characteristics of the transcriptome were predicted using BLAST2GO software (20) with the NCBI non-redundant protein database (cut-off e-value of $\leq 10^{-5}$) using contigs. Each contig with GI accession (NCBI) of the significant hits retrieved was assigned GO terms according to molecular function, biological process and cellular component ontologies at a level that provides the most abundant category numbers (21).

**Toxin Identification and Evolutionary Analyses**

The identification of toxin peptide sequences was carried out from the raw data using tBlastn and a signal peptide can be predicted with the SignalP 3.0 program (http://www.cbs.dtu.dk/services/SignalP/). The propeptide cleavage site was ascertained from the known start site of previously characterized mature toxins. As mentioned previously, such long sequence reads are likely to contain the full nucleic sequences of toxin precursors. The ‘Toxin-like’ sequences, the sequences representing ‘no hits’ and with an abundance of cysteine residues, may encode new toxin peptides. Toxin precursors were selected out according to the following four parameters: the proteins came from a full open reading frame (ORF) translateing by Geneious software(22); the proteins containing more than 4 Cys residues (23); the proteins containing more than 45 amino acids; The toxin precursors were clustered into toxin groups according to their sequence similarity. All precursor sequences were aligned using the ClustalX. The resulting alignment was imported into MEGA software to construct phylogenetic tree by the neighbor-joining method (24), and bootstrap values were estimated from 500 replicates.

**Electrophysiological assays**

Toxins were purified and identified as described previously (10). Rat DRG neurons were acutely dissociated from 30-day old Sprague–Dawley rats and maintained in short-term primary culture according to the method described by Xiao and Liang (25). Ionic currents were recorded from DRG cells under whole-cell patch-clamp techniques using an Axon 700B patch-clamp amplifier (Axon Instruments, Irvine, CA, USA) at room temperature (20–25°C). Patch pipettes with DC resistance of 2–3 MΩ were fabricated from borosilicate glass tubing (VWR micropipettes; VWR Co., West Chester, PA, USA) using a 2-stage vertical microelectrode puller (PC-10; Narishige, Tokyo, Japan) and fire-polished by a heater (Narishige). Experimental data were acquired and analyzed using the programs Clampfit 10.0 (Axon Instruments) and Sigmaplot 9.0 (Sigma, St. Louis, MO, USA). For Na$^+$ current recording, the pipette solution contained (in mM): CsCl 145, MgCl$_2$ 4, HEPES 10, EGTA 10, Glucose 10, ATP 2 (pH 7.2), and the external solution contained (in mM): NaCl 145, KCl 2.5, CaCl$_2$ 1.5, MgCl$_2$ 2, HEPES 10, D-Glucose 10 (pH 7.4). For K$^+$ currents recording, the pipette solution contained (in mM): KCl 135, KF 25, NaCl 9, CaCl$_2$ 0.1, MgCl$_2$ 1, EGTA 1, HEPES 10 and ATP-Na$_2$ 3, adjusted to pH 7.4 with 1M KOH, and the external bath solution contained (in mM): NaCl 150, KCl 30, CaCl$_2$ 5, MgCl$_2$ 4, TTX 0.3, HEPES 10 and D-glucose 10, adjusted to pH 7.4 with 1M NaOH. For Ca$^{2+}$ current recording, the internal solution contained (in mM): Cs-methane sulfonate 110, phosphocreatine 14, HEPES 10, EGTA 10, ATP-Mg 5, adjusted to pH 7.3 with CsOH, and the external solution contained (in
mM): BaCl₂ 10, tetraethylammonium (TEA)-Cl 125, TTX 0.3 and HEPES 10, adjusted to pH 7.4 with TEA-OH.

Results

454 Sequencing Statistics and Transcriptome Assembly

The mRNAs of six venom glands from tarantula Ornithoctonus hainana were extracted and sequenced using GS FLX technology (454/Roche) following the manufacturer’s protocol. Sequencing revealed a total of 249,549 reads (amounting to ~757 Mb) with an average length of ~328 bases per read (max. 830 bp, min. 40 bp, 6.72% of reads <100 bp, 84.6% of reads 100–500 bp, 8.63% of reads >500 bp). The raw sequencing data can be downloaded from SRA (NCBI) using accession number SRP040123. After removing sequences of low quality, a total of 215,640 reads were assembled into 65,432 contiguous DNA sequences (contigs) with an average length of 625 bp (36.3 reads/contig), with the rest remaining as single reads. Whereas this study focused mainly on toxin peptides, numerous other protein sequences were identified and will be described elsewhere. As outlined in the experimental procedures section, we searched for toxin peptide sequences directly from the sequencing reads, as the average read length of >200 bp allowed the identification of full-length toxin precursors. Toxin peptides were also searched for in the contigs, and no additional toxin peptide sequences were found. In total, 52,570 reads displayed similarity to known peptide toxins or ‘Toxin-Like’ sequences; the ‘Putative Toxins’ category includes sequences rich in cysteine residues and sharing sequence identity with toxins or proteins including the ICK motif (5%) that were not identified by a BLAST search; the ‘Cellular Proteins’ category includes transcripts coding for proteins involved in cellular processes (44%); The ‘Unknown Function’ category includes reads that shared sequence identity with previously described sequences with no functional assessment or hypothetical genes; The ‘No Hit’ category indicates no match with currently known sequences. The results are summarized in Figure 1.

Functional Annotation

A search against publicly available databases (nr/NCBI, Swiss-Prot + TREMBL/EMBL) revealed that 8,773 high-confidence proteins were associated with GO terms and further grouped into Molecular Functions (MF), Biological Process (BP) and Cellular Components (CC) at the second level according to standard gene ontology terms (http://www.geneontology.org). Based on annotations from gene ontology (GO) analysis (Figure 2), transcripts were categorized into 2,610 groups with 80% identity threshold. These included functional annotations for 816 Biological Process (BP), 798 Molecular Function (MF), and 996 Cellular Component (CC) categories (Figure 2). Highly expressed transcripts were enriched in metabolism and translation processes, indicating that venom glands are metabolically active and engaging in the intensive protein synthesis and processing required for venom production. Transcripts associated with binding, catalysis and channel regulation were also highly represented, indicating that venom peptides may play an important role in prey inactivation by binding to ion channels. GO terms related to redox homeostasis and proteolysis were also enriched, which may be related to the extensive post-translational modification of spider toxins that was reported previously (9).

Spider Toxin Transcript Identification and Classification

Although spider toxins contain diverse disulfide bridge patterns and fold into a variety of three-dimensional structures, cysteine-rich domains are a common feature shared by many toxin sequences (26,27). Overall, 1,136 toxin precursors were identified in the venom gland transcriptome, and 65.8% of the mature peptides included two adjacent cysteine residues. From previous proteomics results (10), we were able to predict precursor endoproteolytic and amidation sites with high confidence, and 90 toxins and variants accounted for 95% of all toxin precursors. These
were divided into 14 classes (Class A-N) based on the number of cysteine residues present (Table 1), and 70 sequences did not fall into these categories in Class O (Figure 3). Of the 18 Haplopelma hainanum peptide toxins previously characterized, 11 belonged to Class G along with 252 variants, four belonged to Class E along with 39 variants, and the other three belonged to Class C, L and M along with 19, 25 and 4 variants, respectively.

Class A
Class A contained seven novel transcripts, including HN-Aa and its variants (Figure 4). The signal peptide and propeptide of these precursors are identical, and their cleavage sites are LFA/ED and LESEK, respectively. The mature peptide has twelve Cys residues, which is the highest number in a toxin peptide scaffold in H. hainanum reported to date, although two other spider toxins share this cysteine pattern (28,29). TX-L precursors from Lycosa singoriensis superfamily VI have the cysteines arranged C-CC-C-C-C-C-C-C-C-C-C-C-, whereas HN-Aa has the double cysteine shifted to the left. The double knot toxin (DkTx) from the Earth Tiger tarantula (Selenocosmia huwena) has the pattern C-C-CC-C-C-C-C-C-C-C-C-C-, and this sequence shares 87% similarity with HN-Aa. DkTx selectively and irreversibly activates the capsaicin- and heat-sensitive channel TRPV1 (29). This cysteine pattern therefore indicates that HN-Aa may be a TRPV1 channel activator. The six variants identified showed high similarity with the mature peptide, and the presence of different lengths of mature peptide results in great potential for toxin variability.

Class B
Class B contains only four members, the fewest of all classes (Figure 4). HN-Ba and its three variants include eleven Cys residues, and HN-Ba precursors contain over 100 residues that includes the consensus VIAYA cleavage signal, while the propeptide numbered only four residues. A large number of Gln residues were found in the mature peptide, and the C-terminus is rich in Thr and Ser residues, which is strikingly different from other toxins present in this species. The cysteine pattern is shares homology with the U1-hexatoxin-Hsp201a from the Funnel-web spider (30). Other than this highly conserved cysteine pattern, HN-Ba showed no obvious sequence similarity with other peptide toxins.

Class C
Class C contains 29 transcripts that clustered into the known toxin HN-Ca (HNTX-XIV) and two novel toxins (HN-Cb and HN-Cc; Figure 4), all of which included 10 Cys residues, but Class C toxins were further divided into two clades based on cysteine pattern. HN-Ca and HN-Cb shared the pattern -C-C-CC-C-C-C-C-C-C-C-, and formed the first clade, and these toxins do not have a propeptide but do shared a signal peptide containing the cleavage site ELVSC. The mature HN-Cb peptide is rich in Val residues and has a C-terminus that is approximately 20 longer than HN-Ca (HNTX-XIV), indicating diverse functions. Further research is needed to investigate whether these additional residues are cleaved during post-translational modification or remain and have some additional functional significance. HN-Cc exhibited the cysteine pattern -C-C-C-CC-C-C-C-C-C-, and the signal peptide, propeptide and mature peptide were distinct from HN-C[a-b], although HN-Cc peptides shared sequence similarity and included an extended cysteine variant not present in HNTX-XVIII.

Class D
The four toxin precursors in Class D contain nine cysteine residues and exhibited two different cysteine patterns (Figure 4). HN-Da shares significant similarity with HNTX-IV (31) but has the cysteine pattern -C-C-CC-C-C-C-C-C- with an additional three cysteines at the C-terminal region that also included a particularly large number of Ser residues. HN-Db has the cysteine pattern -C-C-C-CC-C-C-C-C-C- and shared an identical signal peptide and propeptide with HNTX-XVIII, but the mature peptide was different, in particular due to the additional cysteine. The additional C-terminal
residues could be removed to produce a more canonical spider toxin during maturation.

Class E
The mature peptides of Class E displayed two distinct cysteine patterns and the precursors were classified into nine groups (Figure 4). HN-E[a-d, f, h-i] shared the cysteine pattern -C-C-CC-C-C-C-C- and the other two toxins shared the cysteine pattern -C-C-C-CC-C-C-C- (Table 1). HN-E[a-c] share high sequence similarity with members of groups HNTX-I, HNTX-III and HNTX-IV, respectively, and include a 30 residue C-terminal extension. Interestingly, the CIC motif was present in the majority of these sequences, which is typical of µ-AGTXs from the American funnel web spider (Agelenopsis aperta) and PnTx3-3 from the Brazilian armed spider (32-35). We speculated that the cysteines in this motif are not involved in disulfide bond formation because PnTx3-3 only has two disulfide bonds, with the fourth and fifth Cys residues excluded (32). Members of HN-Ed share sequence similarity with HNTX-VII, with the propeptide lost and a longer mature peptide. Toxin precursors lacking propeptides are commonly observed in scorpions but rarely in spiders. Four known toxins, HN-Ef (HNTX-XVII), HN-Eg (HNTX-XVIII), HN-Ee (HNTX-XV) and HN-Eh (HNTX-XX) have been chemically and functionally characterized in our previous work (10). HN-Ef (HNTX-XVII) and HN-Eh (HNTX-XX) exhibited the shortest mature peptides in this Class. The typical motif C_7X1C_7V1X5C_7VIIX1C_7VIII, where X is any residue except cysteine was present in precursors, and was first identified in Mu-agatoxins from Agelenopsis aperta (32). Numerous variants differed by only one or two amino acids, suggesting these may be ideal for studying structure-function relationships. Moreover, the HN-Ei mature peptide is identical with HWTX-XV (Table 2). Whilst interesting, it is not unprecedented for toxins from different venomous animals to share identical sequences.

Class F
Class F includes 15 toxins (149 precursors) and divided into three clades based on the seven cysteine pattern present in these peptides (Figure 4). Ten (HN-F[a-g, k, n-o]) share the -C-C-CC-C-C-C-, pattern, and four (HN-F[f-j, m]) exhibited the -C-C-C-CC-C-C- pattern, while only HN-Fi had the pattern -C-CC-C-C-C-C-. The pattern in HN-Fa is shared with that in U1-theraphotoxin-Pc1a, a toxin that possesses strong in vitro anti-plasmodial activity against the intra-erythrocyte stage of P. falciparum (33,36). HN-F[b-c] share high homology with HNTX-I and HNTX-III, including the conservative sequence NEINACSPVF in the C-terminal region. HN-Fd shares high sequence identity with HNTX-IV, but includes the much longer conserved C-terminal sequence “MRSMYPVQFSNWMYLANGGIMSSTSSACQL MSINK” that is enriched with Ser. While HN-Fe has the same signal peptide as HNTX-VIII, the C-terminus is much longer. The mature peptide of HN-Ff shares minimal sequence identity with HNTX-XIV, and the key Cys residues are changed into other amino acids. HN-Fg-j shares high sequence similarity with HNTX-XV and HNTX-XVIII, but can not form four disulfide bonds as there is a cysteine missing in the C-terminal region. We speculated that despite the high identity with known toxins, these peptides must display a novel cysteine pattern. HN-F[k-m] shares sequence similarity with toxins from Ornithoctonus huwena, and HN-Fk shares sequence similarity with HWTX-VIII, including a high proportion of basic residues in the mature peptides. HN-Fm shares similarity with HWTX-XVIII, and HN-Fl also includes a large number of basic residues, but no homologs were identified from a BLAST search, therefore this appears to be a novel toxin.

Class G
Class G includes the greatest number of peptides, accounting for 36% of all toxins, and 10 of these (HN-Ga (HNTX-I), HN-Gc (HNTX-III), HN-Ge
(HNTX-IV), HN-Gg (HNTX-IX), HN-Gj (HNTX-VII), HN-Gm (HNTX-XIII), HN-Gn (HNTX-X), HN-Gr (HNTX-XIII), HN-Gs (HNTX-XIX), HN-Gu (HNTX-XVI), HN-Go (HNTX-XII)) have been chemically and functionally characterized in our previous work. These ten toxins contain six cysteine residues forming three disulfide bonds. Based on sequence similarity, Class G peptides were further divided into 25 distinct toxin groups (Figure 5). HN-G [b, d, f, i, l, q] are similar to the known toxins HNTX-I, HNTX-III, HNTX-IV, HNTX-IX, HNTX-VII, and HNTX-XII, but with a longer C-terminal region. The three disulfides have a 1-4, 2-5 and 3-6 connectivity that form the distinctive inhibitor cysteine knot (ICK) motif. The majority of members in this class are variants that differ by only one or two amino acids, implying that they are natural mutants. In contrast to the traditional ICK-motif-containing spider toxins, HN-Gt and HN-Gv are similar to HNTX-XVI and HNTX-XVIII, and adopt a -C-C-C-C-C- cysteine pattern. Toxins HN-G[w-y] share high sequence similarity with HWT-Xa, HWTX-XIII, and HWTX-XVI, respectively, while HN-Gw, HN-Gx and HN-Gy differ from HWTX-Xa, HWTX-XIII and HWTX-XVI respectively by only two amino acid residues. The signal peptides and propeptides of these toxins included a signal-peptide cleavage site similar to the CYASE sequence.

Classes H and I

Although most of the known spider toxins contain more than six cysteine residues, the members of classes H and may include four to five cysteines. Class H included five transcripts with five cysteines in a distinctive -CC-CC-C- pattern comprising two contiguous Cys pairs (Figure 5). This cysteine pattern is novel, and homology between HN-Ha with other toxins is very low. HN-Ha groups only accounted for about 0.2 % of the toxin transcripts, and are therefore unlikely to play a key role in the venom function of H. hainanum.

All of the 28 transcripts were classified into six toxin groups according to sequence homology. Three toxins, accounting for a quarter of the Class I sequences, did not generate hits in a BLAST search (Figure 8). HN-Ia shares homology with HWTX-XV from O. huwena, which is a tetrodotoxin (TTX)-sensitive voltage-gated sodium channel (VGSC) inhibitor. The signal peptide and propeptide of HN-Ia are highly variable, but HN-Ia has a highly conserved mature peptide and shares the -C-CC-C- cysteine arrangement with HN-Ic. HN-I[a-b] also share this pattern, but also did not generate hits in a BLAST search. HN-Ie differs from HNTX-III, having a C-terminal mutation and lacking Cys 5 and Cys 6. The first and second Cys are replaced by other residues in the mature peptide of HN-Ia.

Classes J and K

Class J contains a set of eight homologous transcripts (HN-Ja) with the cysteine pattern x3Cx3CxxC4Cx5Cx4Cx5Cx1Cxn, where x is any amino acid and n is an undefined number (Figure 6). In contrast to some other classes, these peptides do not contain the double-cysteine (-CC-) motif, and this cysteine pattern is consistent with that of LSTX-M1 and Omega-agatoxin-1A (28,37), in which ten Cys residues form four intra-chain and one inter-chain disulfide bonds. In contrast, HN-Ja did not show sequence similarity with LSTX-M1 or Omega-agatoxin-1A, but is similar to HNTX-II. Class K contains a set of six transcripts which share the cysteine pattern -C-C-C-C-C-C-. The C-terminal region is elongated compared with HNTX-II family peptides and includes an additional Cys residue (Figure 9). The signal peptide and propeptide was the same for all members of these classes, but with C-terminal regions of the mature peptides were variable.

Class L

Class L comprises 100 transcripts that share the cysteine pattern -C-C-C-C-C-C- (Figure 6). HN-La (HNTX-II) was the most significant component in this class. The signal-peptide and propeptide of
HN-L[b-d] were identical with that of HNTX-II, and the mature peptides were also very similar, and all included the conserved NHHDKIRNRKV sequence motif in the C-terminal region. HN-Le shares high sequence similarity with HWTX-VII, but the third Cys in the mature peptide is replaced by Leu, and the conservative sequence VLKCR is added to the C-terminal region. HN-L[f-i] may be narrowly distributed or even species-specific, since no homologs from other spiders have been reported to date. HN-Lg and HN-Li have identical signal peptides, both lack a propeptide, and both have similar mature peptides that include the Cys pattern x2Cx5Cx12Cx11Cxnx. A Lys-rich motif is also present in the mature peptide of HN-L (Figure 6).

Classes M and N
Mature peptides of Class M shared the cysteine pattern -C-C-C-C-C-, which is characteristic of kunitz-type sequences and is found in HWTX-XI and HWTX-VII from O. huwena (38). HN-M[a-b] share high sequence similarity with HNTX-II, with the third Cys replaced by Asp, Leu or Ile in HN-Ma. Similarly, the fifth cysteine in mature HN-Mb peptides is replaced by other residues, and there are several additional Cys residues in the C-terminal region. HN-M [c, h] share sequence homology with HNTX-VIII, with a double cysteine missing from the cysteine pattern (Figure 10), whereas HN-Md shares high sequence similarity with HNTX-XI, except the HN-Md3 variant has an extended C-terminal region enriched with basic residues. The three toxins HN-M[e-g] were novel and generated no hits in a BLAST search (Figure 6).

Class N toxins exhibited a -C-C-C-C- cysteine pattern, and HN-N [a-h] shared some similarity with HNTX-II, huwentoxin II and HWTX-VII. HN-Ni exhibited no obvious sequence similarity with any known sequences (Figure 6).

Toxin Abundance Among Classes
The abundance of toxin ESTs is reflective of venom composition and provides information on their diversity and evolution. Major differences in transcription levels between different toxin groups and between toxin classes were investigated (Figure 7). Unsurprisingly, most venom peptides were expressed at relatively high levels and were present in a variety of forms at the transcriptional level. This suggests that the more abundant toxin transcripts are mirrored by a greater number of peptide toxins. Moreover, gene classes containing larger numbers of precursors also tended to produce the highest number of total reads. Overall, Class G produced the highest number of reads and the largest number of precursors, while classes E, F, L, M and N were of intermediate abundance (Figure 7A). Analysis of the data showed that the 30 most abundant toxin peptides accounted for 84.7% of the total toxin transcripts (Figure 7B). HN-Gc (HNTX-III) had the highest number of transcripts, which were translated into 82 different HN-Gc variants. A BLAST search revealed that HN-Gc shares considerable sequence identity with 20 toxins from spiders and cone snails (Figure 8), indicating that HN-Gc may be an ancient ancestral toxin from which others evolved. Most toxin precursors included a number of variants that enriches toxin diversity, although there exist only three mutations between HN-Ee, HN-Md and HN-Gn, suggesting that spiders could further improve venom selectively through greater sequence diversification.

Variation in Toxin Precursor Sequence and Expression
Classes A–N included multiple variants for each toxin sequence, due to random mutations in both prepro and mature toxin regions. This led to broad distributions in the chromatographic fractions due to the close molecular weights of the variants, and is reflected in the diversity of structures and physicochemical properties. Rapid gene duplication accompanied by focal hypermutation of residues encoded in the mature peptide sequence, and the high conservation of cysteine residues due to high codon bias can explain the diversity of the variants (10,11,39). Alignment of the amino acid sequences of HN-Ge and 18 homologous peptides identified
several mutations (Figure 9). All variants except HN-Ge18 and Ge-15 share the same propeptide sequence, and some sequences are highly similar. HN-Gb and HN-Gd have an extended C-terminal region compared with HN-Gc (Figure 5). Alteration of the termination codon, point mutations, insertion and deletions, and shuffling of sections of oligonucleotide sequence are responsible for the diversification of peptides (Figure 10), and is common to the transcriptome of both snakes and spiders (40-42). Mutations appear to occur randomly and display evident sequence diversity without disrupting the ICK scaffold. Unsurprisingly, 321 precursor variants resulted from 18 known toxins, but only 23 of these overlapped. The significance of the molecular evolutionary process is two-fold: most residues mutate randomly, which generates numerous variants, but at the same time, the molecular mechanisms of transcription preserve cysteine residues, resulting in a high conservation of the molecular scaffold (43). Although a high number of toxin variants were present in the venom gland transcriptome of H. hainanum, many of the gene products may not be functionally active, but may play a role in the evolution of future toxin variants. Comparison of proteomic and transcriptomic data revealed that only 15 fully sequenced and 9 partially sequenced venom peptides were identified using the 454 transcriptome approach (Figure 11), and the vast majority of transcripts coding for CKTs did not appear as translated peptides in the venom proteome. The combination of a targeted mutagenic mechanism to generate high variability with the subsequent action of diversifying selection on highly expressed variants might explain the hypervariability (44). Most transcripts are perhaps unlikely to be translated, apart from critically important components. This mechanism might prevent the venom gland from going to the expense of producing toxins unnecessarily, which makes the process more systematic and efficient(45). Studying toxin diversity is of importance for the discovery of novel toxins, and will help to explain toxin evolution.

Modification and Processing of Multiple Peptide Products

The variety of toxin variants in spider venom partly results from the diversity of toxin gene products and partly from post-translational modifications, sequence homologues, and protein degradation (38,41,46,47). For example, F1-29.54-4123 and F1-30.82-4112.8 share the same precursor, but several residues are modified in the C-terminal region of the mature peptides (Figure 11), and amidation provides further diversity. A single glycine is a common amidation signal, as is Gly followed by a single or dibasic endoproteolytic site at the C-terminus (48). Amidation signals were identified at the C-terminus of three toxins (G-K in HN-Ga, Gc, Ge). During post-translational processing, precursors are processed to yield mature truncated toxins. A single peptide (K) or tripeptide (R-M-D) was removed from the C-terminus of the deduced peptides by the precursor processing enzyme in HN-Gg (HNTX-IX) and HN-Go (F3-24.71-4057.6). F3-30.36-3538.0 and F3-25.85-3351.9 are truncated forms of HN-Gm, with a dipeptide (RR) or tripeptide (WRR) deleted from the C-terminus, while F6-25.12-3998.8 and F3-24.71-4057.6 are truncated forms of HN-Go with a tetrapeptide (GRMD) or tripeptide (RMD) deleted from the C-terminus. We speculate that precursors were alternatively spliced and modified during posttranslational processing, as was observed in venoms from O. huwena (41). HNTX-IX shares the same amino acid sequence with HNTX-IX-2, but the molecular weight is higher by 10 Da, and preliminary sequence analysis showed that this may be due to an unknown modification, as observed in O. huwena (49). Despite the very low expression levels of most of these variants, they contributed a great deal to the overall diversity of spider toxins.

Effect of Functional Diversity on Rat DRG Neurons

To explore the functions of these novel spider toxins, we investigated the activities of 14 representative
toxins on Nav, Kv and Cav channels in rat DRG neurons using the whole-cell patch-clamp technique. In total, 11 six cysteine-containing toxins were chosen from Class G. HN-Ef (HNTX-XVII) has four disulfide bonds, HN-La (F8_17.06) includes the DHH motif, and HN-O38 has three cysteine residues that can only form one disulfide. All 14 tested toxins show very weak or lack inhibition on voltage-gated potassium channels, although they exhibits variable inhibitory effects on sodium or calcium channels. Some toxins with high sequence similarity exhibited similar bioactivity against the ion channels (Figure 12). HN-Gc (HNTX-III), HN-Ge (HNTX-IV) and HNTX-V all contain the ICK motif, and all inhibited tetrodotoxin-sensitive (TTX-sensitive (TTX-S) VGSCs by >70%. However, HN-Ga (HNTX-I) also shares this motif but did not affect TTX-S VGSCs. The hydrophilic Asn19 and basic His27 are conserved among the first three toxins, but are replaced by Gly and acidic Asp in HN-Ga, which may be responsible for the different channel inhibition activity. This result indicates that the mutation of key residues could profoundly affect toxin activity. HN-Ga (HNTX-I) also shares this motif but did not affect TTX-S VGSCs. The hydrophilic Asn19 and basic His27 are conserved among the first three toxins, but are replaced by Gly and acidic Asp in HN-Ga, which may be responsible for the different channel inhibition activity. This result indicates that the mutation of key residues could profoundly affect toxin activity. HN-Gu (HNTX-XVI) showed high inhibition against TTX-R VGSCs, and this toxin has an ICK motif, but there are 14 residues between the fourth and fifth cysteines, which is much longer than in other ICK-containing toxins. Moreover, toxins with diverse cysteine patterns could have similar functions. For example, at a concentration of 100 nM, HN-Ge (HNTX-IV) inhibited calcium channels by >50% and TTX-S sodium channels by >30%. Similarly, HN-O38 has only three cysteine residues while F8_17.06 has a complete DDH motif, and both of these toxins have a cysteine pattern distinct from ICK motif toxins, yet all three inhibited calcium currents by >30%.

In summary, the activity of spider toxins on ion channels was highly variable. Toxins with conserved sequence patterns can share similar functions or behave very differently. Toxins exhibiting low activity against DRG ion channels may be important for inactivating prey, but further functional studies are needed.

Discussion

Spider toxins form a huge reservoir of chromosomally-encoded short peptides that exhibit remarkable structural diversity. Many of the more highly abundant toxins have been studied, but the mechanisms of toxin diversity and evolution are far from being well understood. In this work, a high throughput 454 pyrosequencing approach was deployed as it offers the high-throughput, accuracy and relatively long reads (>300 bp on average) required to cover the full length of spider toxin precursors (60–120 amino acids). A total of 1,136 toxin precursors were detected that clustered into 90 toxin peptides, 18 of which have been previously reported, 44 of which were novel but shared sequence similarity with previously reported H. hainanum toxins, 14 of which shared sequence similarity with known toxins from other spiders, and 14 that were novel and did not retrieve any hits following a BLAST search (Figure 14). A total of 20 different cysteine scaffolds were present that contained between four and twelve Cys residues. Six of these cysteine scaffolds have been presented in previous work, but 14 were novel. The cysteine pattern _C-C-CC-C-C_ that folds into the highly stable cysteine knot (ICK) motif was the most abundant. This toxin scaffold is found in a wide range of bioactive peptides in spiders, snakes or other venomous animals (4,50-54), suggesting this structure evolved in an ancient venomous ancestor.

Many toxin variants were identified at the transcriptional level which arose from a limited set of genes through hypermutation and fragment insertion/deletion. Although the mutations occurred randomly, in general sequences diverged but the common cysteine scaffold was conserved. Moreover, alternative cleavage sites, heterogeneous post-translational modifications, and highly variable N- and C-terminal truncations further increase peptide diversity (10).

The abundance of different toxin transcripts was highly variable and reflected in the composition of venom peptides, and his is important for
understanding of toxin diversity and evolution. In general there was a positive correlation between mRNA levels and peptide abundance. Class G contained both the highest number of transcripts and the largest number of peptide precursors. This important toxin class may target channels conserved in various types of prey animal. Toxins of intermediate abundance also may also be related to hunting. A surprisingly large number of transcript sequences were expressed at relatively low levels, many of which differed from more abundant peptides by single amino acid changes, deletions, and frame and stop codon shifts. Some toxin variants were generated from alternative cleavage sites, interrupted or elongated cysteine patterns. The toxins in low abundance might be considered as the track of toxin evolution, namely, natural selection might drive the spider to retain abundant fractions of the venom, which are applied by the spider in prey capture and/or defense.

Only 18 of the identified toxin genes have been described in previous work. We hypothesize that toxin variants arose by random mutations and focal hypermutations from rapid gene duplication, and post-translational modifications provide further structural diversity. These processes allow the venom transcriptome to rapidly adapt to changes in the environment or type of prey (55,56). It has been suggested that it is the cumulative effect of individual spider toxins that is responsible for the lethality, and different toxin groups exist in different spiders.

Analysis of toxin activity on DRG ion channels showed that some of the toxins with similar sequence patterns generally exhibited similar functions, however, with some exceptions, some of such toxins alternatively displayed distinct functions. Moreover, some of the peptides with more divergent sequences can also share similar functions if key residues are conserved. The diversity of venom toxins results in many peptides working together to improve the efficiency by which prey are incapacitated.

In conclusion, a combination of transcriptomics, peptidomics and electrophysiology revealed the impressive diversity of spider toxin peptides at the transcriptional and peptide structural and functional levels.

**Acknowledgements**
We are thankful to Dr. Zhonghua Liu and Ying Wang for knowledge support. We thank The Cooperative Innovation Center of Engineering and New Products for Developmental Biology of Hunan Province (20134486). This work was supported by a grant of the National Basic Research Program of China (973), 2010CB529801.
Reference


Roche-454/Sanger transcriptome sequences. BMC Bioinformatics 12, 453


**Figure legends**

**Figure 1.** Functional classification of *Ornithoctonus hainana* venom gland transcripts. The graph shows the relative proportion of different types of transcripts belonging to the categories Toxin-like, Putative toxin, Cellular Proteins, Unknown and non-matched sequences, and those that generated no hits following a BLAST search.

**Figure 2.** Gene ontology-based annotations. Results are split into Biological process, Molecular function and Cellular component based on standard gene ontology terms. The different ontology categories are represented on the X-axis, and the number of ESTs matching GO annotation terms (prior to clustering) is represented on the Y-axis.
Figure 3. EST processing, toxin identification and classification.

Figure 4. Alignment of *Ornithoctonus hainana* venom gland toxin precursors from Classes A-F. Signal peptides and propeptides are shown in framed boxes. Identical residues are shaded. Homology with other spider toxins is shown with black dots.

Figure 5. Alignment of *Ornithoctonus hainana* venom gland toxin precursors from Classes G-I. Signal peptides and propeptides are shown in framed boxes. Identical residues are shaded. Homology with other spider toxins is shown with black dots.

Figure 6. Alignment of *Ornithoctonus hainana* venom gland toxin precursors from Classes J-N. Signal peptides and propeptides are shown in framed boxes. Identical residues are shaded. Homology with other spider toxins is shown with black dots.

Figure 7. Transcriptional profiling of *Ornithoctonus hainana* toxins. ESTs and toxin precursors are shown in black and grey, respectively. (A) Different transcriptional levels between Classes. (B) The 30 most abundant toxins (X-axis) account for 84.7% of total toxins.

Figure 8. Comparison of 21 selected spider toxin amino acid sequences. HN-Gc, HN-Ga and HN-Ge are from *Haplopelma hainanum*. HWTX-I, HWTX-IVb and SHT-I are from *Haplopelma schmidti*. Haplotoxin-2 and Haplotoxin-1 are from *Haplopelma lividum*. Gtx1-15 is from *Grammostola rosea*. JZTX-24, JZTX-18, JZTX-III and JZTX-25 are from *Chilobrachys guangxiensis*. TITx1 is from *Theraphosa blond*. CcoTx1 and CcoTx2 are from *Ceratogyrus marshalli*. PaurTx3 is from *Paraphysa scrofa*. Pterinotoxin-1 is from *Pterinochilus murinus*. VSTX3 is from *Grammostola rosea*. ArMKLT2-0311 is from *Conus arenatus*.

Figure 9. Alignment of HN-Ge group toxin precursors.

Figure 10. Comparison of cDNA and protein sequences of HN-Gc and HN-Gd. A C-terminal extension in HN-Gd caused by a cytosine insertion in the cDNA sequence leads to the delayed termination of the coding region.

Figure 11. Full and partial sequences in the *Ornithoctonus hainana* venom transcriptome and proteome.

Figure 12. Functional diversity of *Ornithoctonus hainana* the CKTs. (A) Effect of 14 selected toxins on tetrodotoxin-sensitive (TTX-S) voltage-gated sodium channels (VGSCs), TTX-resistant (TTX-R) VGSCs, voltage-gated calcium channels and voltage-gated potassium channels in rat dorsal root ganglion (DRG) neurons. Sodium current traces were evoked by a 50 ms depolarization to ~10 mV from a holding potential of ~80 mV. Calcium currents were evoked by a 100 ms depolarization to 0 mV from a holding potential of ~40 mV. Potassium currents were elicited at +30 mV from a holding potential of ~90 mV. (B) Sequence alignment of 14 representative cysteine knot toxin (CKT) precursors.

Figure 13. Phylogenetic analysis of toxins sharing high homology with known *Haplopelma hainanum* toxins using the neighbor-joining method.
Figure 14. Comparison of toxins identified in this and previous studies. Previous studies include 16 toxin families from 11 superfamilies.

Table 1. The main features of *Haplopelma hainanum* venom gland toxins

Table 2. Mature peptides from different tarantula species
<table>
<thead>
<tr>
<th>Class</th>
<th>cysteine</th>
<th>precursors</th>
<th>Name</th>
<th>similar sequences</th>
<th>scaffold(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>12 (CC)</td>
<td>7</td>
<td>HN-Aa</td>
<td>double-knot toxin</td>
<td>C-C-CC-C-C-C-C-CC-C-C-C-C [1]</td>
</tr>
</tbody>
</table>
| B     | 11 (CC)  | 4          | HN-Ba         | U34-TRTX                              | C-C-CC-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-
<table>
<thead>
<tr>
<th>H</th>
<th>5 (CC)</th>
<th>2</th>
<th>HN-Ha</th>
<th>non</th>
<th>CC-CC-C [13]</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>4 (CC)</td>
<td>28</td>
<td>HN-I[a~f]</td>
<td>HWTX-XV, HNTX-III</td>
<td>C-C-CC [14], C-CC-C [15]</td>
</tr>
<tr>
<td>J</td>
<td>8</td>
<td>8</td>
<td>HN-Ja</td>
<td>HNTX-II</td>
<td>C-C-C-C-C-C-C-C-C [16]</td>
</tr>
<tr>
<td>K</td>
<td>7</td>
<td>6</td>
<td>HN-Ka</td>
<td>HNTX-II</td>
<td>C-C-C-C-C-C-C-C-C [17]</td>
</tr>
<tr>
<td>L</td>
<td>6</td>
<td>94</td>
<td>HN-L[a~i]</td>
<td>HNTX-II, HWTX-VII</td>
<td>C-C-C-C-C-C-C-C-C [18]</td>
</tr>
<tr>
<td>M</td>
<td>5</td>
<td>132</td>
<td>HN-M[a~h]</td>
<td>HNTX-II, HNTX-VIII, HNTX-XI</td>
<td>C-C-C-C-C [19]</td>
</tr>
<tr>
<td>Amino acid sequences</td>
<td>aa</td>
<td>Peptide toxin</td>
<td>species</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>----</td>
<td>---------------</td>
<td>------------------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>CIGEGVPCDENDPRCCSGLVCLKPTLHGIWYKSYCYKK</td>
<td>39</td>
<td>HWTX-XVIa1</td>
<td>Haplopelma schmidti</td>
<td>(38)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HNTX-XVI</td>
<td>Haplopelma hainanum</td>
<td>(10)</td>
<td></td>
</tr>
<tr>
<td>CIGEGVPRDENDPRCCSGLVCLKPTLHGIWYKSYCYKK</td>
<td>39</td>
<td>HWTX-XVIa5</td>
<td>Haplopelma schmidti</td>
<td>(38)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HNTX-XVI-5</td>
<td>Haplopelma hainanum</td>
<td>(10)</td>
<td></td>
</tr>
<tr>
<td>CIGEGVPCDENDPRCCSGLVCLKPALHGIWYKSYCYKK</td>
<td>39</td>
<td>HWTX-XVIa10</td>
<td>Haplopelma schmidti</td>
<td>(38)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HNTX-XVI-16</td>
<td>Haplopelma hainanum</td>
<td>(10)</td>
<td></td>
</tr>
<tr>
<td>ECRWYLGGCSQGDGCKHLCQCHSNYEWCVWDGTFS</td>
<td>36</td>
<td>HWTX-V</td>
<td>Haplopelma schmidti</td>
<td>(57–59)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HNTX-IX-2</td>
<td>Haplopelma hainanum</td>
<td>(10)</td>
<td></td>
</tr>
<tr>
<td>DCAGYMRECKEKLCCSGYCVSSRWKWCVLPAVPW</td>
<td>33</td>
<td>HWTX-III</td>
<td>Haplopelma schmidti</td>
<td>(53)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HNTX-VIII</td>
<td>Haplopelma hainanum</td>
<td>(10)</td>
<td></td>
</tr>
<tr>
<td>SCAKPGEMCMRIKCCDGQCGCNRTGRCFCK</td>
<td>31</td>
<td>HWTX-XVIIb1</td>
<td>Haplopelma schmidti</td>
<td>(38)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HNTX-XX</td>
<td>Haplopelma hainanum</td>
<td>(10)</td>
<td></td>
</tr>
<tr>
<td>IFECVFSCDIEKEGPCKPKGEKKCSGGKWCKIKLCKLI</td>
<td>39</td>
<td>TXP1</td>
<td>Brachypelma smithi</td>
<td>(60)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ESTx2</td>
<td>Aphonopelma californicum</td>
<td>(61)</td>
<td></td>
</tr>
<tr>
<td>YCQKWMWTCDEERKCCEGLVCLRWCRLKRIINM</td>
<td>31</td>
<td>MTx2</td>
<td>Grammostola rosea</td>
<td>(62)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PaTX2</td>
<td>Paraphysa scrofa</td>
<td>(63)</td>
<td></td>
</tr>
<tr>
<td>Peptide</td>
<td>Species</td>
<td>Accession</td>
<td>Source</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGGWMACKADSDCCETFHCTRFNVCAGK</td>
<td>HWTX</td>
<td>28</td>
<td>Haplopelma schmidti (43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CM3-42.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADCVGDGQRCADWAGPYCCSGYSCRSMPYCRCRSDS</td>
<td>CT-II</td>
<td>38</td>
<td>Hololena curta (65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mu-agatoxin-3</td>
<td></td>
<td>Agelenopsis aperta (66)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EYCYPKRRRVCVTTEQCKPYDTVNNFACGMGAWPEDKKRKVKKNVICYCID</td>
<td>HN-Ei4</td>
<td>57</td>
<td>Haplopelma hainanum (This work)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NELTLCTR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCFKEKWKCPKSKAPEALKCKKPSIKQQKCVRE</td>
<td>HN-Gw2</td>
<td>35</td>
<td>Haplopelma hainanum (This work)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KCKPNGQKCESPTECCSMACSSKSEEEVETHLHFG</td>
<td>HN-Gx4</td>
<td>38</td>
<td>Haplopelma hainanum (This work)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KCKPNGQKCESPAAECSCMACSSKSEEEVETHLHFG</td>
<td>HWTX-XIII</td>
<td>38</td>
<td>Haplopelma schmidti (38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNGRDVPCDPDPKNNRCCSGLCLKPYLHGIWQDYCYVEKSGR</td>
<td>HN-Gy3</td>
<td></td>
<td>Haplopelma hainanum (This work)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNGKDVPCDPDPKNNRCCSGLCLKPYLHGIWQDYCYVEKSGR</td>
<td>HWTX-XVI</td>
<td></td>
<td>Haplopelma schmidti (38)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1
Figure 8

<table>
<thead>
<tr>
<th>Name</th>
<th>Peptide sequences</th>
<th>Length (aa)</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>HN-Gc_HNTX-III</td>
<td>GCALGFDFDDCTPGKNECCP...NYACSSKHKWCKVYLGK</td>
<td>35</td>
<td>100%</td>
</tr>
<tr>
<td>HN-Ga_HNTX-I</td>
<td>GCALGFDDDCVPGKNECCS...GYACNSRDKWCKVLLGK</td>
<td>35</td>
<td>74.3%</td>
</tr>
<tr>
<td>HWTX-I</td>
<td>ACAKGVFDACCTPGKNECCP...NRCVSDKHKWCKWKL</td>
<td>33</td>
<td>65.7%</td>
</tr>
<tr>
<td>Haplotoxin-2</td>
<td>AGLVLVTCTPGKDECCP...NIIVCGSIKIKWCKYKI</td>
<td>33</td>
<td>0.2%</td>
</tr>
<tr>
<td>SHT-I</td>
<td>ECGLIFKACCTPGKNECCS...NRCVSRGKTRWSKWKL</td>
<td>33</td>
<td>51.4%</td>
</tr>
<tr>
<td>Gtx1-15</td>
<td>DGALFGMKCCPNDNCRCR...PNLVCSRTHKWCKYFGK</td>
<td>36</td>
<td>50.0%</td>
</tr>
<tr>
<td>HN-Ga_HNTX-IV</td>
<td>ECALFDGKGCANPSNDDCCKSSNLVCSRKHRCWCKYEIGK</td>
<td>37</td>
<td>48.6%</td>
</tr>
<tr>
<td>JZTX-24</td>
<td>VORGYGLPCTPEKNDCCQ...RLYCSQHRCLCSVKAGK</td>
<td>34</td>
<td>48.5%</td>
</tr>
<tr>
<td>TITx1</td>
<td>AACGLMFESCOPDNKCCP...NRECNRKHWWCKYKLW</td>
<td>35</td>
<td>47.2%</td>
</tr>
<tr>
<td>CcoTx2</td>
<td>DCLGFSDKCDPKDNKCCK...NYCRTSSRDRWKYKL</td>
<td>33</td>
<td>45.7%</td>
</tr>
<tr>
<td>PaurTx3</td>
<td>DCLGFSDKCDPKDNKCCK...NYCRTSSRDRWKYKL</td>
<td>33</td>
<td>44.4%</td>
</tr>
<tr>
<td>Haplotoxin-1</td>
<td>AGCLFGGKDCCPNDNCRCR...PNLVCSRTHKWCKYGW</td>
<td>35</td>
<td>43.2%</td>
</tr>
<tr>
<td>CcoTx1</td>
<td>DCLGFSDKCDPKDNKCCK...NYCRTSSRDRWKYKL</td>
<td>33</td>
<td>42.8%</td>
</tr>
<tr>
<td>JZTX-18</td>
<td>DCLGFSDKCDPKDNKCCK...NYCRTSSRDRWKYKL</td>
<td>33</td>
<td>42.1%</td>
</tr>
<tr>
<td>TITx2</td>
<td>DCLGFSDKCDPKDNKCCK...NYCRTSSRDRWKYKL</td>
<td>33</td>
<td>42.1%</td>
</tr>
<tr>
<td>JZTX-III</td>
<td>DGEQCGFVWKKGRKKPPCCGYYACSKTWGCAVEAP</td>
<td>36</td>
<td>40.5%</td>
</tr>
<tr>
<td>Ptorinotoxin</td>
<td>DCLGFSDKCDPKDNKCCK...GRKGNRKDKWCKYKL</td>
<td>34</td>
<td>38.8%</td>
</tr>
<tr>
<td>VSTX3</td>
<td>DCLGFSDKCDPKDNKCCK...GRKGNRKDKWCKYKL</td>
<td>34</td>
<td>37.1%</td>
</tr>
<tr>
<td>HWTX-IVb</td>
<td>EGKELFKANPNNDCCCKSLV...SRHKTRWCKYQIGK</td>
<td>37</td>
<td>35.1%</td>
</tr>
<tr>
<td>JZTX-25</td>
<td>DCLGFSDKCDPKDNKCCK...GRKGNRKDKWCKYKL</td>
<td>34</td>
<td>30.5%</td>
</tr>
<tr>
<td>ArMKLT2-0311</td>
<td>MAERCGSFGSDGI...PATHTDCGSE...ECFGFEDMGLCT</td>
<td>34</td>
<td>23.8%</td>
</tr>
</tbody>
</table>

Figure 9
Figure 13
Figure 14

14 toxins have homology with known toxins from other spider

14 toxin without any blast hit

-4 toxin similar to known Hainantoxin

90 toxin groups from 454 transcriptome
Structural and functional diversity of peptide toxins from tarantula *Haplopelma hainanum* (*Ornithoctonus hainana*) venom revealed by transcriptomic, peptidomic, and patch clamp approaches.

Yi-Ya Zhang, Yong Huang, Quan-Ze He, Ji Luo, Li Zhu, Shan-Shan Lu, Jin-Yan Liu, Peng-Fei Huang, Xiong-Zhi Zeng, and Song-Ping Liang

PAGE 14194:

The description of spider toxin transcript identification and classification contained errors. "70" should be changed to "57" in the following sentence. "These were divided into 14 classes (class A-N) based on the number of cysteine residues present (Table 1), and 70 sequences did not fall into these categories in class O (Fig. 3)."

PAGES 14196, 14197, AND 14201:

Table 1 has been corrected. "70 sequences" has been changed to "57 sequences" on Fig. 3, and legends have been added to Fig. 7. These corrections do not change the interpretation of the results or the conclusions of this work.

### Table 1

The main features of *H. hainanum* venom gland toxins

<table>
<thead>
<tr>
<th>Class</th>
<th>cysteine</th>
<th>precursors</th>
<th>Name</th>
<th>similar sequences</th>
<th>scaffold(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>12 (CC)</td>
<td>7</td>
<td>HN-Aa</td>
<td>double-knot toxin</td>
<td>C-C-CC-C-C-CC-C-C-C-C [1]</td>
</tr>
<tr>
<td>B</td>
<td>11 (CC)</td>
<td>4</td>
<td>HN-Ba</td>
<td>U34-TRTX</td>
<td>C-C-CC-C-C-C-CC-C-C-C-C [2]</td>
</tr>
<tr>
<td>C</td>
<td>10 (CC)</td>
<td>29</td>
<td>HN-C[a-c]</td>
<td>HNTX-XIV, HNTX-XVII</td>
<td>C-C-CC-C-CC-C-C-C-C-C-C-C-C [3], C-C-C-CC-C-CC-C-C-C-C-C-C [4]</td>
</tr>
<tr>
<td>D</td>
<td>9 (CC)</td>
<td>4</td>
<td>HN-D[a-b]</td>
<td>HNTX-XVIII, HNTX-IV</td>
<td>C-C-CC-C-C-CC-C-C-C-C-C-C [5], C-C-C-CC-C-CC-C-C-C-C [6]</td>
</tr>
<tr>
<td>E</td>
<td>8 (CC)</td>
<td>84</td>
<td>HN-E[a-i]</td>
<td>HNTX-I, HNTX-III, HNTX-IV, HNTX-XII, HNTX-XV, HNTX-XVII, HNTX-XVIII, HNTX-XX, HWTX-XV,</td>
<td>C-C-CC-C-CC-C-C-C-C [7], C-C-C-CC-C-CC-C-C [8]</td>
</tr>
<tr>
<td>F</td>
<td>7 (CC)</td>
<td>150</td>
<td>HN-F[a-l]</td>
<td>U1-TRTX-Pc1a, HNTX-I, HNTX-III, HNTX-IV, HNTX-VIII, HNTX-XIV, HNTX-XV, HNTX-XVIII, HWTX-XIII, HWTX-XVII</td>
<td>C-C-CC-C-C-C-C [9], C-C-C-CC-C-C-C-C-C-C [10]</td>
</tr>
<tr>
<td>H</td>
<td>5 (CC)</td>
<td>2</td>
<td>HN-Ha</td>
<td>non</td>
<td>CC-CC-C [13]</td>
</tr>
<tr>
<td>I</td>
<td>4 (CC)</td>
<td>28</td>
<td>HN-[a-f]</td>
<td>HWTX-XV, HNTX-III</td>
<td>C-C-CC [14], C-CC-C [15]</td>
</tr>
<tr>
<td>J</td>
<td>8</td>
<td>8</td>
<td>HN-Ja</td>
<td>HNTX-II</td>
<td>C-C-C-C-C-C-C-C-C-C-C-C-C-C-C [16]</td>
</tr>
<tr>
<td>K</td>
<td>7</td>
<td>6</td>
<td>HN-Ka</td>
<td>HNTX-II</td>
<td>C-C-C-C-C-C-C-C-C-C-C-C-C-C-C [17]</td>
</tr>
<tr>
<td>L</td>
<td>6</td>
<td>94</td>
<td>HN-L[a-i]</td>
<td>HNTX-II, HWTX-VII</td>
<td>C-C-C-C-C-C-C-C-C-C-C-C-C-C-C-C [18]</td>
</tr>
<tr>
<td>M</td>
<td>5</td>
<td>132</td>
<td>HN-M[a-h]</td>
<td>HNTX-II, HNTX-VIII, HNTX-XI</td>
<td>C-C-C-C-C-C [19]</td>
</tr>
</tbody>
</table>

Authors are urged to introduce these corrections into any reprints they distribute. Secondary (abstract) services are urged to carry notice of these corrections as prominently as they carried the original abstracts.
Additions and Corrections

249,549 reads
removal of low-quality sequences
245,732 reads
contigs assembly
length of > 200 bp
65,432 contigs 202,193 reads

blast
blast hit
toxin cysteine
cysteine rich
not hit
cellular protein
cysteine less
unknown

1136 CKT precursors
cluster
90 toxin groups 57 singletons
Class A-N Class O

Fig. 3

A

B

Fig. 7
Structural and Functional Diversity of Peptide Toxins from Tarantula Haplopelma hainanum (Ornithoctonus hainana) Venom Revealed by Transcriptomic, Peptidomic and Patch-clamp Approaches

Yi-Ya Zhang, Yong Huang, Quan-Ze He, Ji Luo, Li Zhu, Shan-Shan Lu, Jin-Yan Liu, Peng-Fei Huang, Xiong-Zhi Zeng and Song-Ping Liang

J. Biol. Chem. published online March 13, 2015

Access the most updated version of this article at doi: 10.1074/jbc.M114.635458

Alerts:

• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/early/2015/03/13/jbc.M114.635458.full.html#ref-list-1