



Toxin Diversity Revealed by a Transcriptomic Study of *Ornithoctonus huwena*

Yiya Zhang¹*, Yong Huang²*, Quanze He³, Jinyan Liu¹, Ji Luo¹, Li Zhu¹, Shanshan Lu¹, Pengfei Huang¹, Xinyi Chen¹, Xiongzi Zeng^{1*}, Songping Liang^{1*}

1 The Key Laboratory of Protein Chemistry and Developmental Biology of Ministry of Education, College of Life Sciences, Hunan Normal University, Changsha, China, **2** State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, China, **3** The State Key Laboratory of Genetic Engineering, Institute of Biomedical Science, Fudan University, Shanghai, China

Abstract

Spider venom comprises a mixture of compounds with diverse biological activities, which are used to capture prey and defend against predators. The peptide components bind a broad range of cellular targets with high affinity and selectivity, and appear to have remarkable structural diversity. Although spider venoms have been intensively investigated over the past few decades, venom strategies to date have generally focused on high-abundance peptides. In addition, the lack of complete spider genomes or representative cDNA libraries has presented significant limitations for researchers interested in molecular diversity and understanding the genetic mechanisms of toxin evolution. In the present study, second-generation sequencing technologies, combined with proteomic analysis, were applied to determine the diverse peptide toxins in venom of the Chinese bird spider *Ornithoctonus huwena*. In total, 626 toxin precursor sequences were retrieved from transcriptomic data. All toxin precursors clustered into 16 gene superfamilies, which included six novel superfamilies and six novel cysteine patterns. A surprisingly high number of hypermutations and fragment insertions/deletions were detected, which accounted for the majority of toxin gene sequences with low-level expression. These mutations contribute to the formation of diverse cysteine patterns and highly variable isoforms. Furthermore, intraspecific venom variability, in combination with variable transcripts and peptide processing, contributes to the hypervariability of toxins in venoms, and associated rapid and adaptive evolution of toxins for prey capture and defense.

Citation: Zhang Y, Huang Y, He Q, Liu J, Luo J, et al. (2014) Toxin Diversity Revealed by a Transcriptomic Study of *Ornithoctonus huwena*. PLoS ONE 9(6): e100682. doi:10.1371/journal.pone.0100682

Editor: Paulo L. Ho, Instituto Butantan, Brazil

Received: March 14, 2014; **Accepted:** May 26, 2014; **Published:** June 20, 2014

Copyright: © 2014 Zhang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. The raw sequencing data can be download from SRA of NCBI using accession numbers: SRP039535.

Funding: Funding came from: 1. The National Basic Research Program of China (973), 2010CB529800 and 2012CB22305 and 2. The Cooperative Innovation Center of Engineering and New Products for Developmental Biology of Hunan Province (20134486). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: xiongzhizeng@yahoo.com.cn (ZX); liangsp@hunnu.edu.cn (SL)

† These authors contributed equally to this work.

Introduction

Spider venoms contain mixtures of compounds with various biological activities that are used to capture prey or to defend against predators [1,2]. Many of these molecules exert their effects by acting selectively and potently on ion channels (e.g., Ca²⁺, Na⁺ or K⁺ voltage-gated ion channels) in cells [3–9]. Owing to their extraordinary chemical and pharmacological complexity, spider venoms have elicited significant interest for use as tools to study neurophysiology and potential lead structures for pharmaceuticals and insecticides [10]. To date, ~40,000 spider species in 109 families, representing 400 million years of evolution, have been described, although venoms from only a few dozen species have been thoroughly investigated [11]. Spider venoms are highly complex mixtures containing, as a conservative estimate, over 300 toxin peptides per species. Hence, the total number of spider toxins could be over 11 million [11]. However, fewer than 1000 representative spider peptide toxins have been characterized and the mechanisms underlying toxin diversity are far from clear.

The majority of toxins found in spider venoms are small, bioactive and heavily post-translationally modified peptides. Disulfide-rich peptides (having two or more disulfide bonds) are known as CKTs (cystine knot toxins) and represent the majority of toxin peptides. Toxin peptides are synthesized in the venom gland as precursor proteins from a single gene comprising a highly conserved signal peptide, propeptide region and a highly variable toxin sequence. These peptides are classified into gene superfamilies according to sequence similarities of the signal peptide in the precursor. Despite the diversity of mature peptides, the molecular mechanisms of transcription preserve the cysteine residues, resulting in a high degree of conservation of the molecular scaffold. So far, over 10 different cysteine patterns have been identified in spider venom, with the number of residues ranging from four to fourteen [12]. Additionally, numerous post-translational modifications (PTMs), including hydroxylation of proline, valine and lysine, carboxylation of glutamate, C-terminal amidation, cyclization of N-terminal glutamine and glycosylation, contribute to the structural variety of the peptides [13,14].

To date, 67 different toxin precursors from *Ornithoctonus huwena* have been identified, based on EST (Expressed sequence tag) sequencing of the cDNA library [15]. Separation of crude venom components using a combination of ion-exchange and reverse-phase high-performance liquid chromatography (HPLC) and 2D gel electrophoresis, followed by silver staining, revealed over 300 protein spots, 133 of which were detected with mass spectrometry [16,17]. The large discrepancy between the gene and mass numbers detected in venom indicates that the low sensitivity of traditional transcriptomic approaches leads to the overlooking of rare sequences, which are transcribed at low levels. The recent availability of second-generation sequencing has facilitated the identification of several toxin-like peptides, significantly accelerating the pace of toxin discovery [18–20]. The 454 Life Sciences pyrosequencing technology is commonly used due to its high-throughput and accuracy comparable to traditional Sanger sequencing [21,22]. We selected this approach, since it generates relatively long readable sequences (on average >300 bp) that encompass the full length of toxin precursors (60–120 amino acids). The technology allows direct identification of toxin precursors and avoids the errors inherent in the assembly of overlapping sequences (contigs) typically required for other second-generation technologies that generate shorter readable sequences (reads).

In the present study, 626 toxin precursors were unambiguously identified and classified into 16 different superfamilies, including six novel superfamilies and six novel cysteine patterns. A surprisingly large number of mutations, incomplete precursor sequences and aberrant sequences (i.e., interrupted or elongated cysteine patterns and highly variable isoforms, including deletions and elongations) were detected. The majority of these result from single amino acid changes and frameshifts. Interestingly, although most unusual toxin variants are expressed at very low levels, they may play an important role in the high rates of evolution of toxin genes within families. Moreover, along with alternative modes of peptide processing, these transcripts may explain the hypervariability of venom peptide and rapid evolution of bioactive peptides.

Materials and Methods

cDNA Library Construction and 454 Sequencing

The tarantula spider, *Ornithoctonus huwena*, is not a protected species and found widely in the Guangxi Province of China. Three tarantula spiders were collected for study. No specific permission was required for these locations/activities. Venom glands of *Ornithoctonus huwena* were obtained two days after being milked via 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. Full-length enriched double-stranded cDNA was synthesized from pooled total RNA using the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen) and NEBNext mRNA Second Strand Synthesis Module (NEB), according to the manufacturer's protocol, and subsequently purified using the QIAquick PCR Purification Kit (Qiagen USA, Valencia, CA). The DNA library was prepared from 300 ng samples using the manufacturer's instructions (Rapid Library Preparation Method, Roche). Sequencing was performed on a Roche GS FLX Titanium sequencer.

Sequence Assembly and Alignment

Sequence reads were trimmed by excluding 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 454 sequence reads were assembled into contigs using CLC Genomics Workbench 3 with its default parameters. Raw reads and contigs were uploaded in a proprietary web-based searchable database. As mentioned previously, such long sequence reads are likely to contain the full nucleic sequences of toxin precursors. Both raw reads and assembled contigs were identified as transcripts. Peptide sequences were identified from the transcript data using tBlastn. The E-value threshold of $e \leq 10^{-5}$ with a bit score >40 was 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 toxin', 'Cellular Proteins', 'Unknown function', 'No Hit'). The identified peptide sequences were aligned using ClustalX 2.0.

Gene Ontology Annotations

Functional characteristics of the transcriptome were predicted using BLAST2GO software [23] with the NCBI non-redundant protein database (cut-off e-value of $\leq 10^{-5}$) using EST 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 [24].

Toxin Identification and Evolutionary Analyses

Known toxins often showed similar sequences, and the new toxin patterns expected a visitor. Moreover, the 'Toxin-like' sequences representing 'no hits' sequences with an abundance of cysteine residues, may encode new toxin peptides. Thus, 'no significant hit' sequences and those displaying high similarity with known toxins presented a huge treasury for toxin identification. Precursors encoded by these cDNA sequences were initially identified using NCBI BLAST, and those without significant BLAST results enriched with cysteine residues were identified as toxin precursors. Signal peptides were 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. Toxin-like proteins were grouped into different toxin families according to sequence similarity. All precursor sequences were aligned using ClustalX 2.0. The resulting alignment was imported into MEGA software to construct a phylogenetic tree with the neighbor-joining method [25], and bootstrap values estimated from 500 replicates.

The nucleotide sequences of superfamilies I, II and XVIII were aligned with ClustalX. The number of synonymous substitutions per synonymous site (Ds) and nonsynonymous substitutions per nonsynonymous site (Dn) were estimated using the original NeiGojobori model [26]. Fisher's exact tests for positive evolution (based on the original Nei-Gojobori model) were performed using MEGA software.

Results

454 Sequencing Statistics and Transcriptome Assembly

The mRNAs of six venom glands from tarantula *Ornithoctonus huwena* were extracted and sequenced using GS FLX technology (454/Roche) following the manufacturer's protocol. Sequencing revealed a total of 123,922 reads (amounting to ~42 Mb) with an average length of ~327 bases per read, ranging from 40 to 836 bp. Raw sequencing data can be downloaded from SRA of NCBI using the accession number SRP039535. Overall, 80,908 reads were assembled into 4,224 contigs, while the rest remained as

singletons. Both raw cDNA reads (>200 bp) and assembled contigs were collected for further analysis.

Representative sequences of each transcript were analyzed with the tBLASTn programs. Significant similarities were evident between 88,182 transcripts and proteins in the UniprotKB, Rепbase and ToxRelDB databases. We observed high identity of 32,267 reads with toxin families, accounting for 29% raw reads. ‘Cellular Proteins’ includes transcripts coding for proteins involved in cellular processes (44%), including peptidases, cell signaling, cell structure and motility, metabolism and protein processing. For precursors with no significant BLAST results, cysteine patterns were extracted and empirically examined. The group ‘Putative toxins’ includes sequences rich in cysteine that display identity to toxins (0.1%). ‘Unknown function’ encompasses ESTs homologous to described sequences with no functional assessment or hypothetical genes (7%). About 20% reads were assigned to the ‘No Hit’ category, indicating no match with currently known sequences. Results are summarized in Figure 1.

Functional Annotation

A search against public databases (nr/NCBI, Swiss-Prot+TREMBL/EMBL) revealed that ~68.8% of all transcripts are 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>). This finding was in agreement with previous data [13,27,28]. According to annotations from GO analyses (Figure 2), transcripts were categorized into 73 biological processes. ‘Metabolic process’ indicating an important metabolic activity, was the most highly represented in *O. huwena* venom gland, similarly observed for snail *Conus consors* [20,29]. For molecular function, binding and catalytic activities rank first, which is related to the high toxin peptide content in the venom gland of *O. huwena*.

Finally, four main categories of cellular components were identified, specifically, cell, extracellular region, organelle and macromolecular complex, all of which were mainly related to structural proteins involved in the secretion and transport of toxic compounds.

Spider Toxin Transcript Analysis

Toxin peptides are the most abundant compounds of the spider venom gland. According to precursor sequence identity, 599 toxin precursors were produced from ‘spider toxin peptides’ and 27

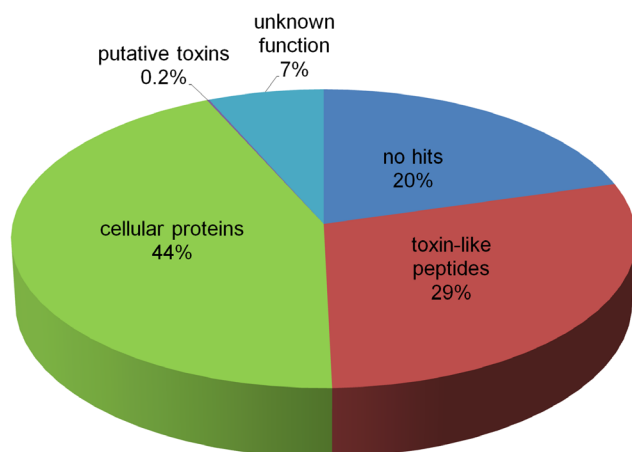


Figure 1. Classification of ESTs from *O. huwena* venom glands.
doi:10.1371/journal.pone.0100682.g001

non-redundant precursor sequences from ‘putative toxin’, yielding 626 non-redundant precursor sequences in total. Interestingly, only a small fraction of the total peptide precursors found in assembled contigs were retrieved from raw data, indicating that genetic diversity is underestimated if only raw reads are analyzed. Several protein and enzyme sequences were additionally identified among the contigs.

Significant differences at the mRNA level were observed among the different toxin precursors. Precursors were most abundantly identified from superfamilies I, II and XVIII, corresponding to 9,286, 6,232 and 4,167 reads, respectively. This finding suggests that toxins more highly expressed at the mRNA level, and tend to be more abundant in the venom isoforms. Linear regression analysis ($r^2 = 0.97$) indicated the highest number of reads in superfamilies with the largest number of precursors (Figure 3A). As shown in Figure 3B, 408 putative toxin precursors only had one cDNA read. These rare transcripts comprised ~65% of the total retrieved putative toxin precursors. We additionally identified 53 high-level (>10 cDNA reads) and 165 low-level precursors (2–10 cDNA reads). The total number of precursors and cDNA reads for each superfamily are plotted in Figure 3C. Seven known superfamilies contained toxin precursors with high-level cDNA reads (>10 reads). Three of these (I, II and XVIII) had more than 9 high-level read precursors, while the remaining (X, XI, XIV and XVI) contained only one high-level read precursor. Precursors in superfamilies XV and XVII displayed low-level or very low-level expression. One precursor was identified in superfamily XIII with three cDNA reads. No high-level precursors were identified in the six putative new superfamilies, and only low or very low-level cDNA reads were observed (Figure 3D).

Toxin precursors with very low or low-level expression often displayed high sequence identity with high-level precursors in the same family or superfamily. Analysis of gene sequences revealed that very low/low level precursors are variants produced by hypermutation, fragment insertion/deletion and mutation-induced premature termination and elongation. Eight toxin cDNA and precursor sequences are aligned in Figure 4. Mutation of the 100th base ‘A’ to ‘C’ caused single substitution of a cysteine residue (HWTX-Ia12) and created an odd number of cysteine residues, leading to cysteine pattern disorder. Moreover, the absence of three bases resulted in frameshift mutation in the signal peptide region without altering the enzyme site and mature peptide region. Premature stop codons (HWTX-Ic3) produced truncated isoforms as well as truncated cysteine patterns. Frameshifts produced highly variable isoforms via deletions/additions in the C-terminal region (HWTX-Id1, Ih10 and Ij1 and Ik1).

Abundant toxin variations were additionally observed in venom. Among the 30 full and 17 partial sequences identified via Edman degradation [17], 40 (30 full and 10 partial) were clustered into eight superfamilies (12 families). As shown in Figure 5, single mutation (in CM5-24.03 vs. HWTX-I, CM7-28.48 vs. CM7-31.11 and CM2-21.5 vs. CM2-21.5) and C-terminal extension (in CM5-19 vs. HWTX-I, CM4-25.6 vs. CM4-12.9) were detected in venom that were possibly produced by diverse precursors. CM3-30.10 and CM3-32.87, generated from the same toxin precursor (HWTX-IIIa1), produced diverse isoforms via alternative cleavage and post-translational modifications (PTMs), previously reported as the major strategies underlying toxin diversity [13]. Different molecular masses were obtained for the same amino acid sequence via post-translational processing (in CM5-23.07 vs. CM5-23.87, CM6-26.65 vs. CM6-27.7 and CM2-37.32 vs. CM2-38.60).

Eight toxin superfamilies identified previously in venom from *Ornithoctonus huwena* were observed in the transcriptome, but their

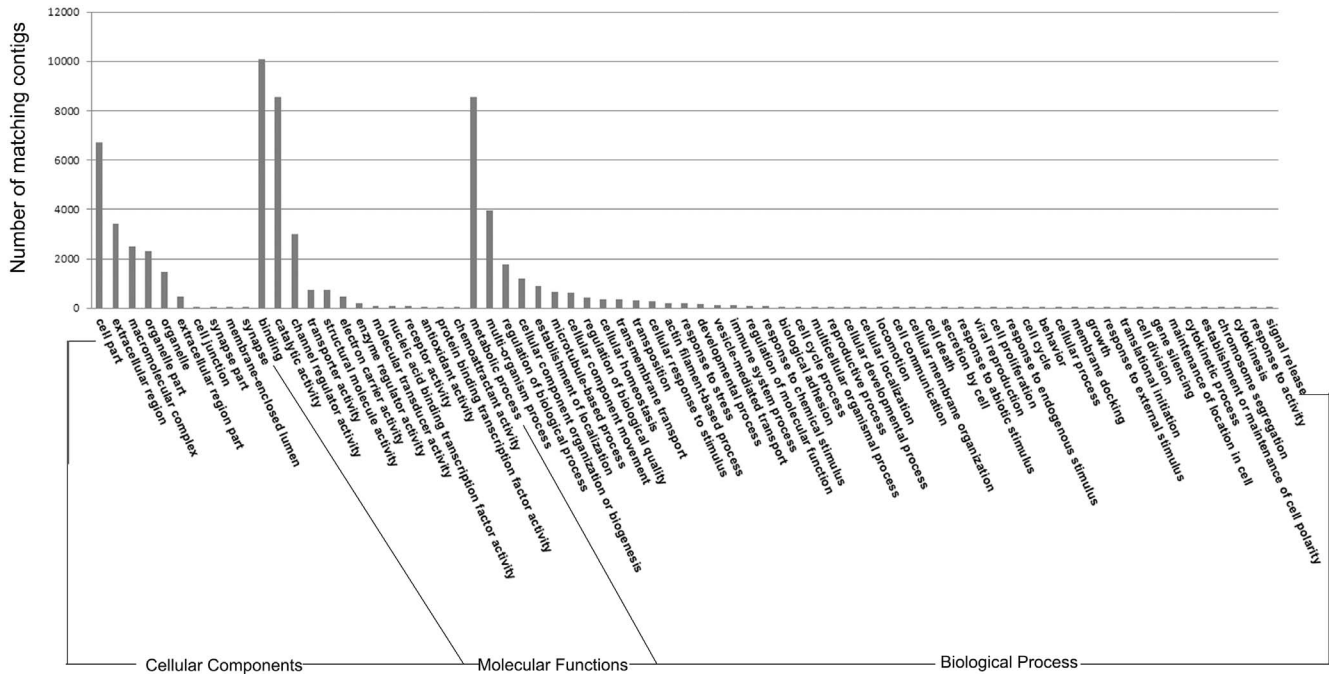


Figure 2. Gene ontology sorted annotations. Results are classified into Biological process, Molecular function and Cellular component at the second level, according to standard gene ontology terms. The different ontology categories are presented on the X-axis. Number of ESTs matching GO annotation terms (prior to clustering) are presented on the Y-axis. As more than 80% of the contigs failed to show association with GO terms, this analysis represents an annotation of the most conserved eukaryotic genes.
doi:10.1371/journal.pone.0100682.g002

mutated isoforms showed considerable variations. Only 15 precursors and 24 putative mature peptides were detected with both approaches, probably caused by the well-known phenomena of intraspecific variations in many species [30–33].

Toxin Precursors and Their Classification

Overall, 626 toxin precursors were categorized into 10 known and 6 putative new gene superfamilies (Table S1). The signal peptides and cysteine patterns are listed in Table 1. The phylogenetic tree shows the toxin precursors originate from three different clades. Members of the HWTX-I superfamily belong to one clade, the HWTX-XI, HWTX-XXII, HWTX-XXIV, HWTX-XIV and HWTX-XV superfamilies and HWT-XXIII, HWTX-XXI superfamilies belong to the second clade, and the others belong to the third clade. Each superfamily was further divided into several distinct families and subfamilies based on the identity of precursor sequences (Figure 6). In addition to the 15 known families, one novel family from the known superfamily XV and eight putative new families from six novel superfamilies (designated XX, XXI, XXII, XXIII, XXIV, XXV) were identified.

The HWTX-I Superfamily

In this superfamily, reads and toxin variants were the most abundant, with a total of 9,286 cDNA reads and 199 precursors, which clustered into five families (25 subfamilies: HWTX-I [a, c~k, m], SHL-I [a~f], HWTX-III [a~d], HWTX-IV [a~c] and HWTX-Va) (Figure 7). The majority of sequences in this superfamily showed high similarity. In particular, their precursors contained a highly conserved cleavage signal “CYASE” for signal peptides and consensus “GEER” cleavage signal for propeptide processing enzyme. Predicted mature peptides comprised 30 to 60 residues, three disulfide bonds and a classic Type I (C-C-CC-C-C)

pattern. In total, 16 sequences displayed missing signal peptides, which was also a common phenomenon in the HWTX-XVI superfamily. Despite high sequence similarities among the five known families [15], the mature peptides of 20 novel subfamilies were highly variable. These novel subfamilies showed truncated or extended C-terminal regions produced by stop codon shifts and fragment insertion/deletion. Peptides of all known subfamilies were expressed at high levels, and those of most novel subfamilies at low levels (<50 reads), except HWTX-If and HWTX-Ih.

Family I contained 11 subfamilies (99 isoforms), including the known HWTX-Ia and HWTX-Ic and nine novel subfamilies. Among the known subfamilies, 31 and 8 isoforms were identified, respectively, although precursors HWTX-Ia4 and HWTX-Ia9 have been detected previously in *O. huwena*. HWTX-Id, HWTX-Ie and HWTX-Ig contained a long cysteine pattern with one or two additional residues in the C-terminal region, compared with HWTX-Ia. The same cysteine arrangement, -C-C-CC-C-C- (cysteine pattern I), was observed in HWTX-If, HWTX-Ih and HWTX-Im. We speculate that these variants potentially inhibit high voltage-activated (HVA) Ca^{2+} channels [34,35]. HWTX-Ih showed a short pattern with cysteine residues missing at the C-terminal, which would break III-VI disulfide bonds. This phenomenon was additionally observed in the known toxins HWTX-Ib and HWTX-Ic [15].

Overall, 16 precursors belonging to the HWTX-III family were classified into one known and three novel subfamilies. HWTX-IIIa was the known subfamily with three toxin precursors, but only one (HWTX-IIIa1) has been detected previously. HWTX-III [b~d] contained an extended C-terminal region and an impaired disulfide bridge with the double cysteine motif missing.

We identified 14 precursors in the HWTX-IV family, which were clustered into three subfamilies. Two novel subfamilies (HWTX-IVb and HWTX-IVc) showed C-terminal elongation

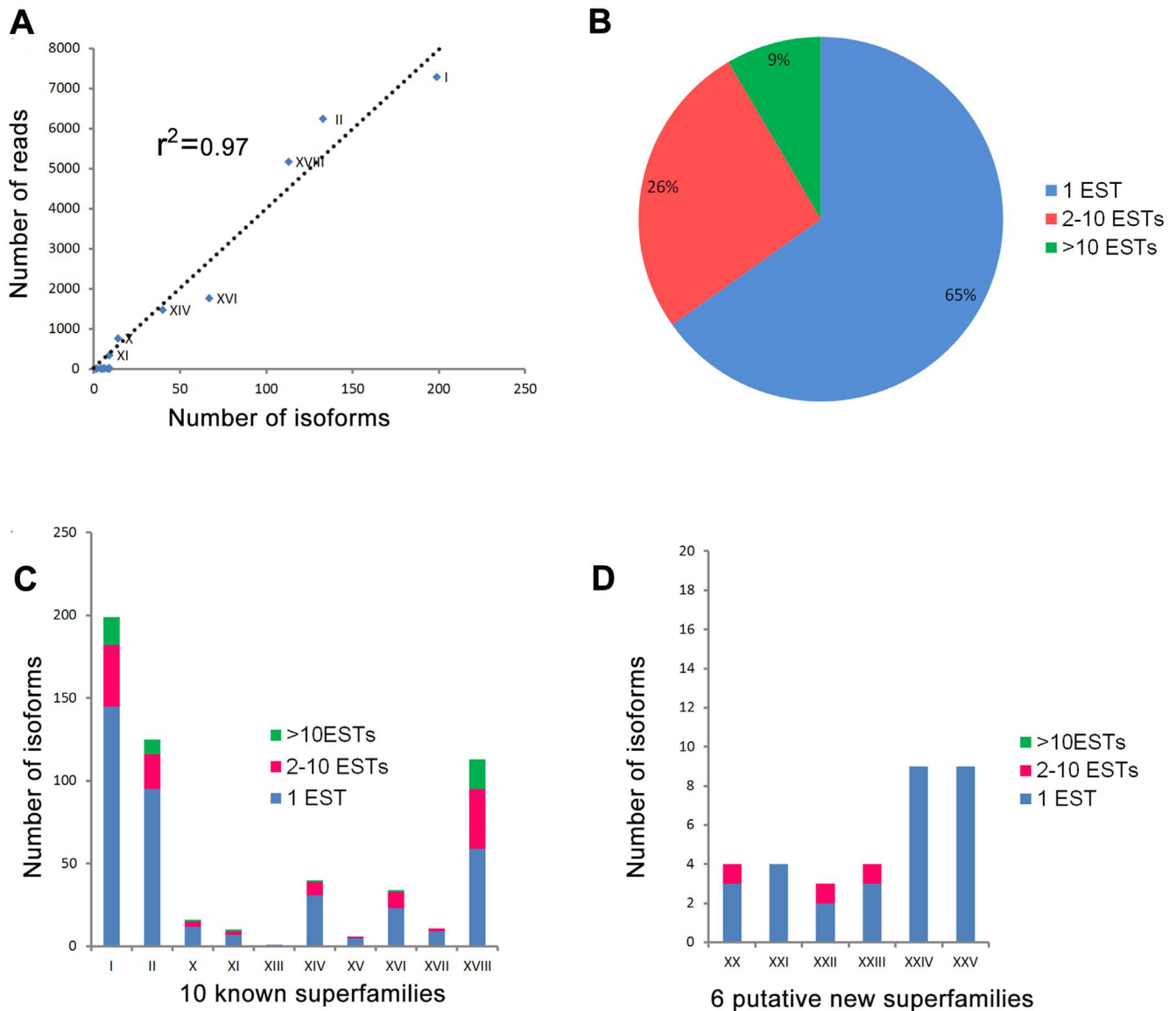


Figure 3. Isoforms and transcription levels of putative toxin peptides. A, The number of precursors and total number of reads per gene superfamily show an apparent correlation. The goodness of fit was $R^2=0.97$, revealing a significant correlation between the two parameters. B, Variations in the transcriptional levels. Only 7% of the putative sequences had more than 10 cDNA reads, 27% had moderate cDNA reads, 27% were present as rare transcripts with only single cDNA read discovered for the full-length precursor. C, Isoforms of 10 known superfamilies. C, Isoforms of six putative new superfamilies. doi:10.1371/journal.pone.0100682.g003

and truncation, respectively. The missing cysteine in HWTX-IVc may lead to loss of inhibitory activity of the peptides on the neuronal tetrodotoxin-sensitive voltage-gated sodium channel. Mature sequences from short and long precursors displayed strikingly similar primary sequences. Accordingly, we speculate that C-terminal diversification results from a simple alteration of the stop codon.

Only one subfamily (HWTX-Va) was detected in HWTX-V containing 10 precursors. These are variants of HWTX-V that show high sequence similarity with HWTX-V, and specifically inhibit high voltage-activated calcium channels in adult cockroach *dorsa*.

SHL-I was additionally identified as an abundant toxic component with 60 toxin precursors. The C-terminal mutation

in SHL-Id and SHL-If destroyed the disulfide bridge via alteration of the sixth cysteine residue to valine.

The HWTX-II Superfamily

HWTX-II was the second most abundant superfamily with 133 isoforms, accounting for 24% toxin reads, and classified into three families (HWTX-II, HWTX-VII and HWTX-XI). In addition to the three known subfamilies, eleven new subfamilies displayed high identity with HWTX-II. HWTX-II [b, g, m], HWTX-VIIb and HWTX-XIc exhibited the same cysteine pattern II (“C-C-C-C-C-C-”) as HWTX-II, with I-III, II-V and IV-VI disulfide connectivity and a DDH three-dimensional structure motif. Despite the C-terminal elongation, these could be the neurotoxins as ion channels antagonists. However, HWTX-II [c~f, h~k, n] had one to three cysteine residues missing, which may significantly

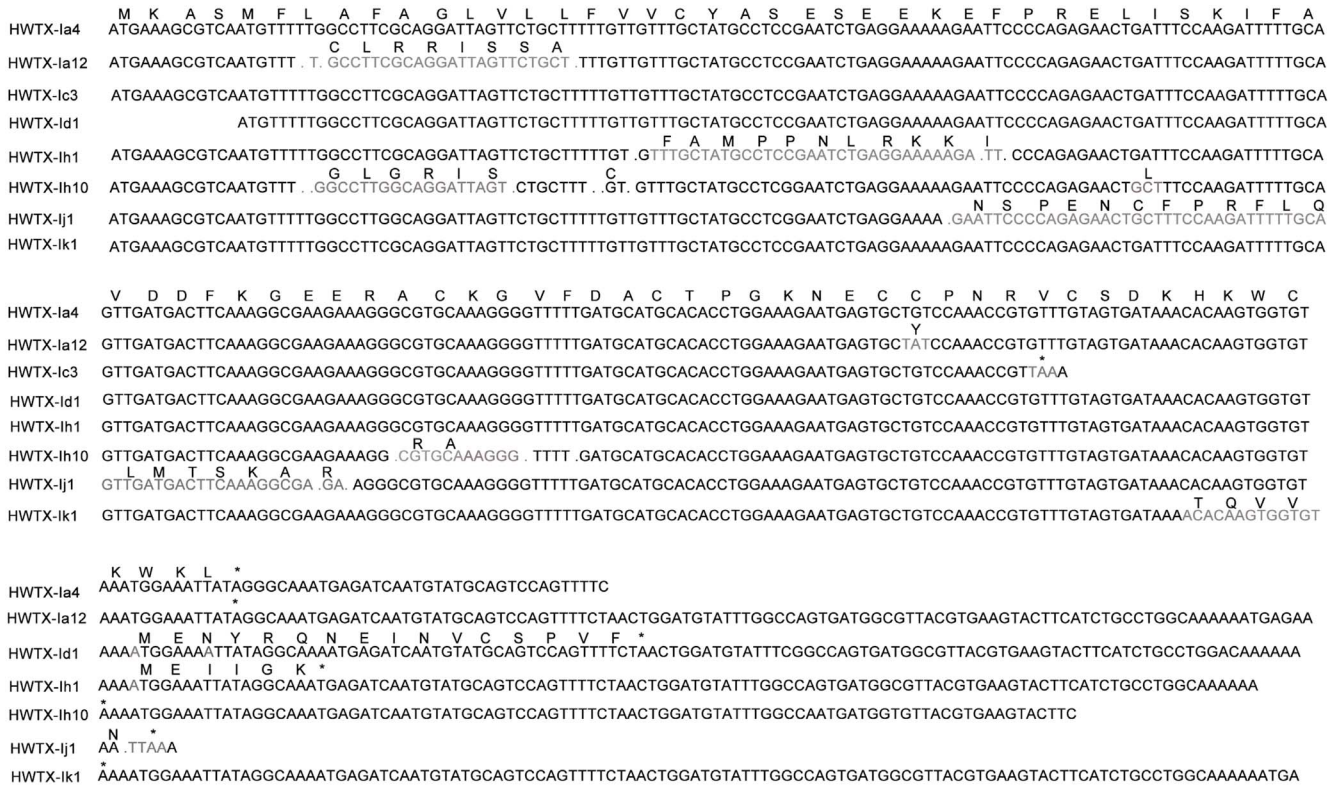


Figure 4. Comparison of the cDNA and precursor sequences of HWTX-Ia4 with the seven similar mutations.
doi:10.1371/journal.pone.0100682.g004

affect disulfide bond formation. HWTX-VII displayed high sequence identity (about 81%) with the HWTX-II family, implying similar molecular folding and bioactivity (Figure 7).

The HWTX-X Superfamily

This superfamily contained 14 members and was classified into four subfamilies. The three new subfamilies showed high sequence identity with the known peptide (HWTX-X) (Figure 7). HWTX-Xb shared the same cysteine pattern I (“-C-C-CC-C-”) as HWTX-X, and an expanded C-terminal region of about 20 residues that did not affect its biological function as an N-type Ca²⁺ channel inhibitor [36]. The main motif “IPCCGVCSHNKCT” in HWTX-Xc was modified to “YHAAECVHIISVPNRRRETILKRC” with a double cysteine residue missing. The ICK motif (inhibitor cystine knot motif) in HWTX-Xc was disrupted and the rest of the cysteine residues could form two disulfide bonds in a way of independent assortment. This phenomenon was additionally observed in the HWTX-Xd subfamily. We speculate that the mutation contributes significantly effect to the functional diversity of HWTX-X.

The HWTX-XI Superfamily

HWTX-XI is a serine protease inhibitor consisting of 55 residues with three disulfide bridges [37]. Previous results indicate that HWTX-XI follows the classical Kunitz architecture formed by three disulfide bridges with a linkage pattern of I-VI, II-IV, and III-V [15]. Overall, nine precursors were detected for the HWTX-XI superfamily, although only Huwentoxin-11g8, a known isoform, was expressed at a level higher than 50 reads [37].

HWTX-XIb contained two peptides lacking the II-IV disulfide bond, with tyrosine replacing the fourth cysteine. Members of this group have been designated ‘sub-Kunitz type’ toxins (cysteine

pattern VIII). The sub-Kunitz type toxin has a Kunitz motif formed by the remaining two disulfide bridges [38]. In the bovine pancreatic trypsin inhibitor (BPTI) reduction experiment, native-like conformation and trypsin inhibitor activity remained for BPTI without the disulfide bond [39]. Similarly, HWTX-XIa lacked two cysteine residues in the C-terminal region. The final member of the HWTX-XI superfamily was HWTX-XIc. Precursor peptides in this subfamily contained a very long mature peptide region (> 110 residues) with the ICK motif (Figure 7).

The HWTX-XIV Superfamily

The HWTX-XIV superfamily contained 40 members, which were further classified into six subfamilies. The signal peptides displayed high similarity, although mature peptides contained a highly variable C-terminal region, and all were processed from precursors containing no propeptide sequences (Figure 7). The sequence identities of the mature peptides of HWTX-XIV [b, c] were 76.2% and 79.4%, compared to HWTX-XIVa1, respectively. However, the C-terminal regions lacked two cysteine residues, caused by the removal of two “T” bases and insertion of one “A” base in the cDNA sequence, respectively. Both cDNA sequences of HWTX-XIV [d, f] lacked three cysteines, resulting from the insertion of two “G” and absence of two “A” bases. The mature peptide of HWTX-XIVe showed 43.1% similarity with HWTX-XIVa and cDNA sequence identity of 85.1%, indicating different bioactivities. Although this mutation may not be structurally significant, it contributes significantly to the considerable diversity of toxins peptides in venom gland.

The HWTX-XV Superfamily

The HWTX-XV superfamily included two families with six members. Four precursors were identified for the known HWTX-

Peptide	families	peptide/precursor sequences	annotate
HWTX-I	I	ACKGVFDACIPGKNECCPNRVCSDKHKWCWK	HWTX-I
HWTX-Ia17	I	ACKGVFDACIPGKNECCPNRVCSDKHKWCWK	
CM5-23.07	I	MKASMFALAGLVLLVLLFAMPNLRKRIPQRTAFQDFAVDDFKGEERACKGVFDACIPGKNECCPNRVCSDKHKWCWK	
HWTX-Ia17	I	MKASMFALAGLVLLVLLFAMPNLRKRIPQRTAFQDFAVDDFKGEERACKGVFDACIPGKNECCPNRVCSDKHKWCWK	
CM5-23.87	I	ACKGVFDACIPGKNECCPNRVCSDKHKWCWK	
HWTX-Ia17	I	MKASMFALAGLVLLVLLFAMPNLRKRIPQRTAFQDFAVDDFKGEERACKGVFDACIPGKNECCPNRVCSDKHKWCWK	
CM5-19	I	ACKGVFDACIPGKNECCPNRVCSDKHKWCWK	C-terminal prolong
HWTX-Ia15	I	MKASMFALAGLVLLVLLFAMPNLRKRIPQRTAFQDFAVDDFKGEERACKGVFDACIPGKNECCPNRVCSDKHKWCWK	
CM5-24.03	I	ACKGVFDACIPGKNECCPNRVCSDKHKWCWK	none
CM5-25.14	I	ECKGFGKCVPGKNECCSG.....	none
CM3-30.10	III	DCAGYMRCEKELCCSGYVCSRWKWCVLPAFWR	
HWTX-IIa1	III	MVMNKAAMFLTFAGLVLLVLLFVVCYASESEEKPEPKEMLSSIFAVDNDFKQEERDCAGYMRCEKELCCSGYVCSRWKWCVLPAFWR	
CM3-32.87	III	DCAGYMRCEKELCCSGYVCSRWKWCVLPAFWR	
HWTX-IIa1	III	MVMNKAAMFLTFAGLVLLVLLFVVCYASESEEKPEPKEMLSSIFAVDNDFKQEERDCAGYMRCEKELCCSGYVCSRWKWCVLPAFWR	
CM5-26.05	IV	ECLIEIFKACNPNDQCCSKLWCSR HWTX-IV	
HWTX-IVa	IV	MKASMFALAGLVLLVLLFVVCYASESEEKPEPKEMLSSIFAVDNDFKQEERDCAGYMRCEKELCCSGYVCSRWKWCVLPAFWR	
CM5-27.7	IV	ECLIEIFKACNPNDQCCSKLWCSR SRKTRWCKYQIGK	C-terminal prolong
HWTX-IVa	IV	MKASMFALAGLVLLVLLFVVCYASESEEKPEPKEMLSSIFAVDNDFKQEERDCAGYMRCEKELCCSGYVCSRWKWCVLPAFWR	
CM1-20.09	V	BCRWYLGCCSDGDCCKHLQCHSNYEWCVWDGTF	C-terminal truncate
HWTX-V	V	MKSIVFVFLGALLAVVCSASEDAHKELLEKVVRAVMDKTDVAQEAERECRWFLGGCSQDGDCCCKHLQCHSNYEWCVWDGTF	
CM1-26.26	V	BCRWYLGCCSDGDCCKHLQCHSNYEWCVWDGTF	HWTX-V
HWTX-V	V	MKSIVFVFLGALLAVVCSASEDAHKELLEKVVRAVMDKTDVAQEAERECRWFLGGCSQDGDCCCKHLQCHSNYEWCVWDGTF	
CM2-25.21	SHL	CGRQYKDYTYGCCSGYVCSPTWKWCVRPGR	single mutate
SHL-1b1	SHL	GCLGDKCDYNNGCCSGYVCSRTWKVCLAGP	
CM2-28.05	SHL	GCLGDKCDYNNGCCSGYVCSRTWKVCLAGP	C-terminal truncate
SHL-1a17	SHL	MKTSMFLTLTGLVLLSFLFAMPNLRKEFFKELLSSIFAADSDFKVEERGLGDKCDYNNGCCSGYVCSRTWKVCLAGP	
CM2-30.80	SHL	GCLGDKCDYNNGCCSGYVCSRTWKVCLAGP	
CM2-37.32	SHL	GCLGDKCDYNNGCCSGYVCSRTWKVCLAGP	C-terminal prolong
SHL-1a15	SHL	MKTSMFLTLTGLVLLSFLFAMPNLRKEFFKELLSSIFAADSDFKVEERGLGDKCDYNNGCCSGYVCSRTWKVCLAGP	SHL-I
CM2-38.60	SHL	GCLGDKCDYNNGCCSGYVCSRTWKVCLAGP	
SHL-1a17	SHL	MKTSMFLTLTGLVLLSFLFAMPNLRKEFFKELLSSIFAADSDFKVEERGLGDKCDYNNGCCSGYVCSRTWKVCLAGP	
CM3-28.4	SHL	GCLGDKCDYNNGCCSGYVCSRTWKVCLAGP	
SHL-1a16	SHL	MKTSMFLTLTGLVLLSFLFAMPNLRKEFFKELLSSIFAADSDFKVEERGLGDKCDYNNGCCSGYVCSRTWKVCLAGP	
CM6-33.8	II	FECSSICEIEKGGESCKPKCKGKWCCKFNMCKV	none
CM2-28.48	II	FECSSICEIEKGGESCKPKCKGKWCCKFNMCKV	mutate
HWTX-IIa	II	MKVTLIAILTCAAIVLHTAAEELAEESQLMEVMPDTELAADVDEERLPEFCSPFCEIEKGGESCKPKCKGKWCCKFNMCKV	
CM7-31.11	II	FECSSICEIEKGGESCKPKCKGKWCCKFNMCKV	N-terminal truncate
huwentoxin-II	II	MKVTLIAILTCAAIVLHTAAEELAEESQLMEVMPDTELAADVDEERLPEFCSPFCEIEKGGESCKPKCKGKWCCKFNMCKV	
CM7-34.29	II	MKVTLIAILTCAAIVLHTAAEELAEESQLMEVMPDTELAADVDEERLPEFCSPFCEIEKGGESCKPKCKGKWCCKFNMCKV	HWTX-II
huwentoxin-II	II	MKVTLIAILTCAAIVLHTAAEELAEESQLMEVMPDTELAADVDEERLPEFCSPFCEIEKGGESCKPKCKGKWCCKFNMCKV	
CM6-17.20	X	KCLPPKPCYGAATKIPCCGVCSSHRCI	HWTX-X
HWTX-X	X	MNMKILVLAIVLCLVSTHAERHSKTDMDSPMQLQERKCLPPKPCYGAATKIPCCGVCSSHRCI	
CM7-32.77	X	IDTCLPSDRGRCKASPERVYFNGRTCAKFIYGGCGGNGKFPFQEAAMKRCACA	HWTX-XI
Hw11g8	XIV	MGIAIRLSAVLFLSVLFWVTFPALLSADHHDGRIDTCLPSDRGRCKASPERVYFNGRTCAKFIYGGCGGNGKFPFQEAAMKRCACA	
CM2-19	XIV	BCWSQADCSGDHCCAGSFSKNCRYPGGDGAQCE	C-terminal truncate
HWTX-XIVa4	XIV	MKSLCRKRVLLVCLVVMAMMELVSCBCWSQADCSGDHCCAGSFSKNCRYPGGDGAQCEPRNKYEYVSTGCPDENLICSVINRCQSV	
CM2-21	XIV	BCWSQADCSGDHCCAGSFSKNCRYPGGDGAQCE	C-terminal truncate
HWTX-XIVa4	XIV	MKSLCRKRVLLVCLVVMAMMELVSCBCWSQADCSGDHCCAGSFSKNCRYPGGDGAQCEPRNKYEYVSTGCPDENLICSVINRCQSV	
CM2-21.5	XIV	BCWSQADCSGDHCCAGSFSKNCRYPGGDGAQCEPRNKYEYVSTGCPDENLICSVINRCQSV	mutate
HWTX-XIVa9	XIV	MKVTLVLCVSDHDGDTGSCBCWSQADCSGDHCCAGSFSKNCRYPGGDGAQCEPRNKYEYVSTGCPDENLICSVINRCQSV	
CM2-45.8	XIV	BCWSQADCSGDHCCAGSFSKNCRYPGGDGAQCEPRNKYEYVSTGCPDENLICSVINRCQSV	HWTX-XIV
HWTX-XIVa7	XIV	MTVGGVDGDHDTGSCBCWSQADCSGDHCCAGSFSKNCRYPGGDGAQCEPRNKYEYVSTGCPDENLICSVINRCQSV	
CM3-27.51	XIV	BCASQADCSGDHNLIGSRPEFNRPR.....	mutate
CM4-42.45	XIV	BCASQADCSGDHNLIGSRPEFNRPR.....	mutate
CM3-13	XIX	DCARKEGECVWCKTCCDLDFYCFMFIIPCKGYP	HWTX-XIX
CM2-25.44	XIX	DCARKEGECVWCKTCCDLDFYCFMFIIPCKGYP	mutate
CM5-17.4	XVII	SCKKIKKEI.....	mutate
CM4-12.9	XVII	SCKKIKKEI.....	mutate
HWTX-XVII	XVII	MIMIPSFVITCCNGYLAKIRIVFPVSKDSTATENDTEQRMEIVQVETARACSRQIGDRCKSNDCDGTIVYVGGKEVNVQMDKSSDNDSSERNR	
CM4-14.5	XVII	ACSRQIGDRCKSNDCDGTIVYVGGKEVNVQMDKSSDNDSSERNR	
CM4-25.6	XVII	ACSRQIGDRCKSNDCDGTIVYVGGKEVNVQMDKSSDNDSSERNR	C-terminal prolong
CM5-26.4	XVII	ACSRQIGDRCKSNDCDGTIVYVGGKEVNVQMDKSSDNDSSERNR	C-terminal prolong
CM7-26.55	XVII	ACSRQIGDRCKSNDCDGTIVYVGGKEVNVQMDKSSDNDSSERNR	C-terminal truncate
CM3-24.60	XVII	ACSRQIGDRCKSNDCDGTIVYVGGKEVNVQMDKSSDNDSSERNR	C-terminal truncate
CM3-16.8	XVII	ACSRQIGDRCKSNDCDGTIVYVGGKEVNVQMDKSSDNDSSERNR	C-terminal truncate
HWTX-XVIIa	XVII	MIMIPSFVITCCNGYLAKIRIVFPVSKDSTATENDTEQRMEIVQVETARACSRQIGDRCKSNDCDGTIVYVGGKEVNVQMDKSSDNDSSERNR	
CM1-17.7	XVII	SCAPRETCRNRNLLCCRGECVCPVIGDCPCYGD	HWTX-XVII
HWTX-XVIIa1	XVII	SCAPRETCRNRNLLCCRGECVCPVIGDCPCYGD	
CM5-42.8	XVII	CGVMKACADSDCECFHCFRNVWCGK	none
CM4-22.7	XVII	CGVMKACADSDCECFHCFRNVWCGK	none
CM4-40.45	XVII	IICAPBEGPCVAGIGCCAGLRCGAGLGLAGSQ	none
CM5-16.7	XVII	SVIKEKSTKENDCCCGMST.....	none
CM1-17.81	XVII	NCIGBQVPCDENDPRCCSLVWLKTLHGHWIKSSYCYKCK	none
CM3-13.9	XVII	GDRENSNPDGTTVS.....	none
CM3-8.8	XVII	FVK	none

Figure 5. Full or partial sequences and biological activity of peptides from *O. huwena* found in both transcriptome and venom proteome.
doi:10.1371/journal.pone.0100682.g005

XV family. HWTX-XVa5 displayed a truncated pattern with two cysteine residues missing at the C-terminal region.

HWTX-XXXVIIIa was identified as a new subfamily devoid of the propeptide region, with a similar signal peptide as HWTX-XVa (Figure 7). Members in this family displayed the same cysteine pattern V (“-C-C-C-C-C-C-C-”), distinct from other known toxins in this spider. Moreover, the mature HWTX-XXXVIII peptide contained more than 100 residues, and was considerably longer than HWTX-XV.

The HWTX-XVI Superfamily

In the HWTX-XVI superfamily, 66 precursors were derived from seven subfamilies. We identified 27 precursors for the known HWTX-XVIa subfamily, none of which had been detected previously. From the precursors of six novel subfamilies, mature peptides are diversified through a C-terminal drift. HWTX-XVIIb contained a cropped C-terminal region and HWTX-XVIIc displayed C-terminal elongation, compared with HWTX-XVIa. HWTX-XVI [c, e~g] with a longer mature peptide, showed low cDNA sequence identity with the HWTX-XVIa family. The HWTX-XVI [c, d] and HWTX-XIVe subfamilies only contained

five and four cysteine residues, respectively, which could disrupt the cysteine pattern (Figure 7).

The HWTX-XVII Superfamily

Seven precursors were identified in the HWTX-XVII superfamily, with three subfamilies distinguished based on sequence similarity of the mature peptide. Two known subfamilies, HWTX-XVIIa and HWTX-XVIIb, have been detected previously in the cDNA library of *O. huwena* [15] but only two precursor sequences identified in both our transcriptome study and previous work. Precursors of the novel subfamily, HWTX-XVIIc, contained a much shorter signal peptide “MKDPENSEER” and had no propeptide sequence (Figure 7). Their mature peptides shared low sequence homology with the HWTX-XVIIa family, with a significantly shorter cysteine pattern (“-C-C-CC-C-C-”).

The HWTX-XVIII Superfamily

HWTX-XVIII was the third most abundant superfamily, including 113 sequences. Two known subfamilies, HWTX-XVIIIa and HWTX-XVIIIb, showed higher expression than 100 reads, and ten novel subfamilies displayed lower expression

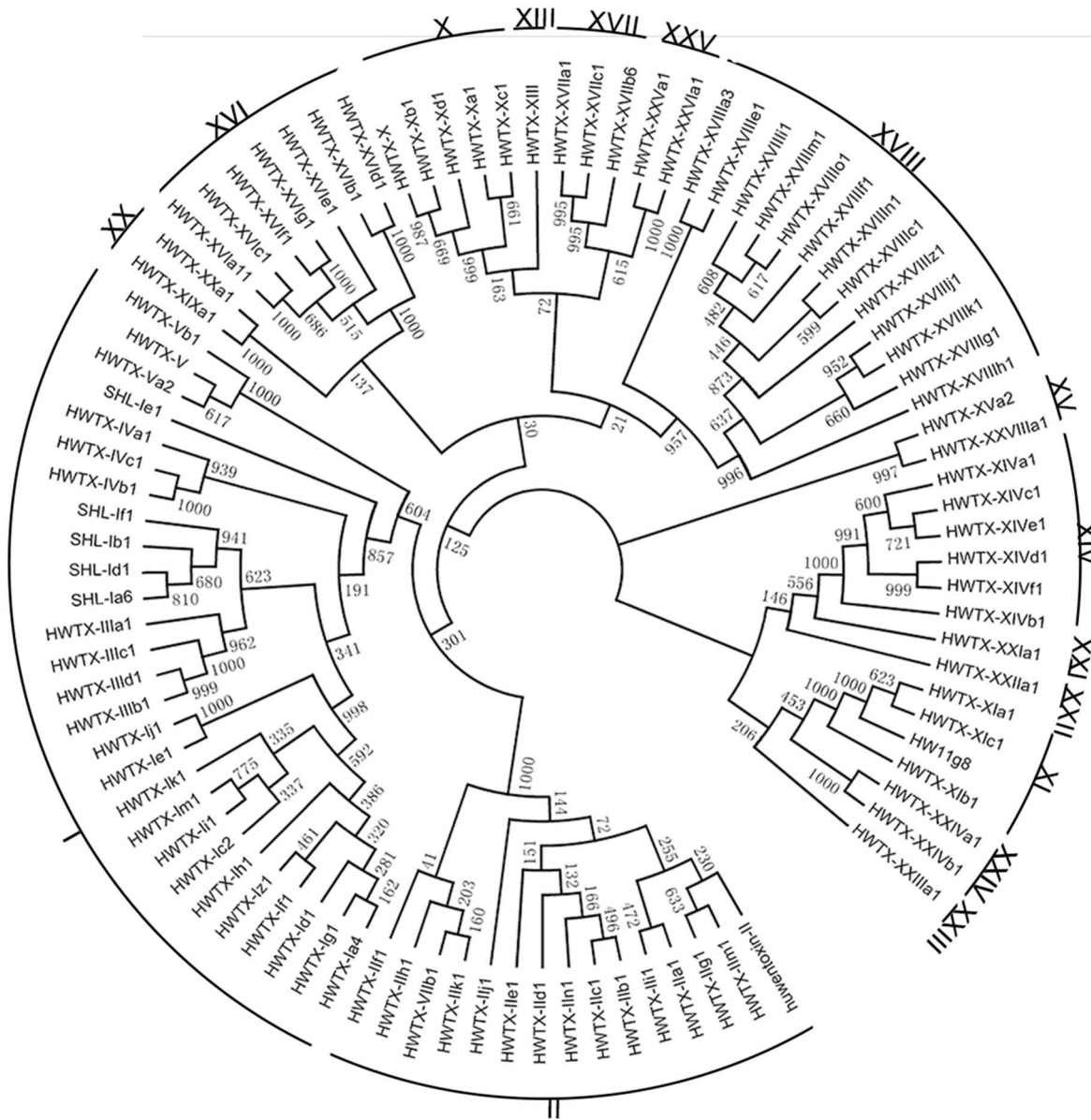


Figure 6. Phylogenetic tree of all venom peptide precursors in the HNTX superfamily from *O. huwena*.
doi:10.1371/journal.pone.0100682.g006

(except HWTX-XVIIIj, HWTX-XVIIIk and HWTX-XVIIIz) (Figure 7). HWTX-XVIII [e~j and o] were the truncated toxins of HWTX-XVIIIa. HWTX-XVIII [k and z] showed different C-termini with cysteine mutations. Loss of cysteine residues led to an odd pattern in these subfamilies, caused by deletion or insertion in the cDNA sequences. This phenomenon was also observed in HWTX-XVIIIc2 and HWTX-XVIIIc3 [15].

The HWTX-XX Superfamily

Six precursors were detected in the novel superfamily HWTX-XX, with two families (HWTX-XIX and HWTX-XX) distinguished based on sequence similarity of the propeptide region (Figure 7). The mature peptide of HWTX-XIX contained 40 amino acids with a conservative cysteine pattern “DCX₆CX₅CCX₆CX₁₄CX” (X is any amino acid), which displayed significant homology with the ICK motif. Other than the highly conserved arrangement of cysteine residues, HWTX-

XIX had no obvious sequence similarity with other toxin peptides. The mature peptide of HWTX-XX was similar to DkTx with the cysteine pattern VII (“-C-C-CC-C-C-C-CC-C-C-”) with the largest number of cysteines in HWTXs. The presence of two double cysteine motifs in this pattern implies that HWTX-XX contains the two disulfides through knots that are separated by a short linker. HWTX-XXa3 was a truncated toxin of HWTX-XXa1 with the codon terminated earlier.

The HWTX-XXIV Superfamily

Nine precursors were identified in this superfamily, which displayed very low expression (one read). The precursor peptide of the HWTX-XXIV superfamily included identical signal peptide and propeptide, but the mature regions in HWTX-XIVa were highly variable and different from those of known toxins. A consistent cysteine pattern “XCX₅₀CX₉CX₇CX₂CX” (X is any

Table 1. Main features of toxins from the venom gland transcriptome of the tarantula *O.huwena*.

superfamily	signal peptides	families	previous data	cysteine framework	Function identity
I	MKASMFALFAGLVLLFVVCYA	HWTX-I	HWTX-I	-c-c-cc-c-c- (I)	mammalian HVA Ca ²⁺ channel initor and TTX-S Na ⁺ channel inhibitor [34,35]
		HWTX-III	HWTX-III	-c-c-cc-c-c- (I)	insect Na ⁺ channel inhibitor [54]
		HWTX-IV	HWTX-IV	-c-c-cc-c-c- (I)	mammalian TTX-S Na ⁺ channel inhibitor [51,56]
		HWTX-V	HWTX-V	-c-c-cc-c-c- (I)	insect Ca ²⁺ channel inhibitor [58]
		SHL	SHL-I	-c-c-cc-c-c- (I)	lectin [53]
II	MKVTLIALITCAAVLVLHTTAA	HWTX-II	HWTX-II	-c-c-c-c-c-c- (II)	
		HWTX-VII	HWTX-VII	-c-c-c-c-c-c- (II)	
X	MNMKILVLVAVLCLVSTHA	HWTX-X	HWTX-X	-c-c-cc-c-c- (I)	mammalian N-type Ca ²⁺ channel inhibitor [36]
XI	MGIARILSAVFLSVLFWTFPALLSAD	HWTX-XI	HWTX-XI	-c-c-c-c-c-c- (II)	K ⁺ channel inhibitor [37]
XIII	MKYAIVLCVIVIVTVVRA	HWTX-XIII	HWTX-XIII	-c-c-cc-c-c- (I)	
XIV	MKVLLVCLVWMMAMMELVSC	HWTX-XIV	HWTX-XIV	-c-c-cc-c-c-c-c-c- (VI)	
XV	MKHFASCIFSVLTVACGVSQT	HWTX-XV	HWTX-XV	-c-c-cc-c-c-c-c- (III)	
		HWTX-XXVIII		-c-c-c-c-c-c-c-c- (V)	
XVI	MNTVRVTFLLVFLAVSLGQA	HWTX-XVI	HWTX-XVI	-c-c-cc-c-c- (I)	
XVII	MKIATFLGSLFLIASYVLICEA	HWTX-XVII	HWTX-XVII	-c-c-cc-c-c-c-c- (III)	
XVIII	MKLSLIIATSLVIAVVA	HWTX-XVIII	HWTX-XVIII	-c-c-c-cc-c-c-c- (IV)	
XX	MKLSNFAVVLVGLFVSVPLFA	HWTX-XX		-c-c-cc-c-c-c-c-c-c- (VII)	
		HWTX-XIX		-c-c-cc-c-c- (I)	
		HWTX-XXI		-c-c-c-c-c-c- (X)	
XXII	MTWLLMVPLMLLSPLLQIAC	HWTX-XXII		-c-c-c-c- (IX)	
XXIII	MIVLGGMAFSVTSMVVACVV	HWTX-XXIII		-c-c-c-c-c- (VIII)	
XXIV	MVCATRRRLIRLSSSG	HWTX-XXIV		-c-c-c-c-c-c- (II)	
XXV		HWTX-XXV		-c-c-c-cc-c-c- (XI)	
		HWTX-XXVI		-c-c-c-c- (IX)	

doi:10.1371/journal.pone.0100682.t001

amino acid, n is any number) (Figure 7) was observed, with the same number of cysteine residues as the sub-Kunitz motif [40].

The HWTX-XXV Superfamily

This novel superfamily contained 10 precursors that clustered into two subfamilies (XXV and XXVI). Members of HWTX-XXV did not contain a signal peptide and the propeptide sequence was identified as “MTREETQSLGEHEKDEEVTGSEER”. The mature peptide of the HWTX-XXV family contained 75 residues with a consensus cysteine pattern XI (“-C-C-C-CC-C-C-”) containing an odd number of cysteine residues instead of the usual even number expected for the formation of internal disulfide bonds. The HWTX-XXVI sequence exhibited the same signal peptide and propeptide as HWTX-XXV, but the cysteine pattern IX (“-C-C-C-C-”) was considerably shorter.

HWTX-XIII, HWTX-XXI, HWTX-XXII and HWTX-XXIII Superfamilies

Four novel superfamilies (XIII, XXI, XXII and XXIII) displayed low-level expression and contained one, five, two and four precursors respectively. HWTX-XIII was composed of 77 residues, and the hypothetical mature peptide contained 37 residues with three disulfide-bridged motifs [15]. Signal peptides of HWTX-XXI superfamily members were similar, with identical cysteine patterns (“X₄CX₂CX₉CX₃CX₂CX₂CX₇CX_n”, whereby

X is any amino acid, n is any number) but highly variable mature peptides. The mature peptide of HWTX-XXIIa4 contained 131 residues. A single point mutation may destroy the disulfide bridge by changing the fourth cysteine residue to methionine. Members of the HWTX-XXIII superfamily showed high sequence identity, except for the presence of different residues (leucine or glutamine) at the end of the mature peptide (Figure 7). Precursors of the HWTX-XXII superfamily had identical signal peptides and propeptides with a cysteine pattern of “-C-C-C-C-” (IX), which was not common in other spiders. This positional pattern of cysteines is also found in Kappa-hefutoxin-1, a potassium channel inhibitor that belongs to the short scorpion toxin superfamily (kappa-KTx family).

Different Evolution Strategies of Superfamilies I, II and XVIII

Propeptides displayed high sequence similarity in each superfamily. The contrasting high diversity of mature peptides may be a strategy of toxin evolution. To further explore evolutionary patterns, the Dn/Ds ratios from precursor peptides were examined. The majority of Dn/Ds ratios of superfamily II and XVIII were located in the region lacking constraints, indicating neutral evolution (Figure 8. B, C). Few of the Dn/Ds ratios within small genetic distances (<0.2) were significant (>1) implying positive selection evolution. The decrease in ratio at larger genetic

Superfamily I	Signal peptide	Propeptide	Mature peptide	
HWTX-Ia4	MSAMFALAGLVLLVFCVCA	ESSEKFPRELISK	IKAVDD.FGGEERAKGKVFIACTGKNECCPN.RVSD..FKHKMKNK	81
HWTX-Ic2	MSFFPCSFNALS..LLFLVFCVCA	ESSEKFPRELISK	IKAVDD.FGGEERAKGKVFIACTGKNECCPN.RVSD..FKHKMKNK	90
HWTX-Ie1	MYCTEVSMFALAGLVLLVFCVCA	ESSEKFPRELISK	IKAVDD.FGGEERAKGKVFIACTGKNECCPN.RVSD..FKHKMKNK	77
HWTX-Ie1	MYCTEVSMFALAGLVLLVFCVCA	ESSEKFPRELISK	IKAVDD.FGGEERAKGKVFIACTGKNECCPN.RVSD..FKHKMKNK	100
HWTX-Ie1	MYCTEVSMFALAGLVLLVFCVCA	ESSEKFPRELISK	IKAVDD.FGGEERAKGKVFIACTGKNECCPN.RVSD..FKHKMKNK	105
HWTX-Ie1	MYCTEVSMFALAGLVLLVFCVCA	ESSEKFPRELISK	IKAVDD.FGGEERAKGKVFIACTGKNECCPN.RVSD..FKHKMKNK	108
HWTX-Ie1	MYCTEVSMFALAGLVLLVFCVCA	ESSEKFPRELISK	IKAVDD.FGGEERAKGKVFIACTGKNECCPN.RVSD..FKHKMKNK	110
HWTX-Ie1	MYCTEVSMFALAGLVLLVFCVCA	ESSEKFPRELISK	IKAVDD.FGGEERAKGKVFIACTGKNECCPN.RVSD..FKHKMKNK	84
HWTX-Ie1	MYCTEVSMFALAGLVLLVFCVCA	ESSEKFPRELISK	IKAVDD.FGGEERAKGKVFIACTGKNECCPN.RVSD..FKHKMKNK	77
HWTX-Ie1	MYCTEVSMFALAGLVLLVFCVCA	ESSEKFPRELISK	IKAVDD.FGGEERAKGKVFIACTGKNECCPN.RVSD..FKHKMKNK	70
HWTX-Ie1	MYCTEVSMFALAGLVLLVFCVCA	ESSEKFPRELISK	IKAVDD.FGGEERAKGKVFIACTGKNECCPN.RVSD..FKHKMKNK	82
HWTX-Ie1	MYCTEVSMFALAGLVLLVFCVCA	ESSEKFPRELISK	IKAVDD.FGGEERAKGKVFIACTGKNECCPN.RVSD..FKHKMKNK	88
HWTX-Ie1	MYCTEVSMFALAGLVLLVFCVCA	ESSEKFPRELISK	IKAVDD.FGGEERAKGKVFIACTGKNECCPN.RVSD..FKHKMKNK	99
HWTX-Ie1	MYCTEVSMFALAGLVLLVFCVCA	ESSEKFPRELISK	IKAVDD.FGGEERAKGKVFIACTGKNECCPN.RVSD..FKHKMKNK	104
HWTX-Ie1	MYCTEVSMFALAGLVLLVFCVCA	ESSEKFPRELISK	IKAVDD.FGGEERAKGKVFIACTGKNECCPN.RVSD..FKHKMKNK	98

... (Tables for Superfamily II, Superfamily X, Superfamily XI, Superfamily XIV, Superfamily XV, Superfamily XVI, Superfamily XVII, Superfamily XVIII, Superfamily XX, Superfamily XXIV, Superfamily XXV, and Superfamilies continue) ...

Figure 7. Sequence alignment of representative venom peptide precursors in superfamilies I-XXV from *O.huwena*. Signal peptides and propeptides are shown in framed boxes, and cysteines of mature peptides are highlighted. doi:10.1371/journal.pone.0100682.g007

distances was attributed to saturation. Half the Dn/Ds ratios of superfamily I were significant (>1), showing more positive selection evolution than superfamilies II and XVIII (Figure 8A). Additionally, we examined the null hypothesis of neutral evolution

with Fisher's exact test (cut-off of 0.05) with the Nei and Gojori model [26]. No significant results were observed with superfamilies I, II and XVIII to reject a null hypothesis. In the HWTX-II superfamily, positive selection was detected in a clade containing

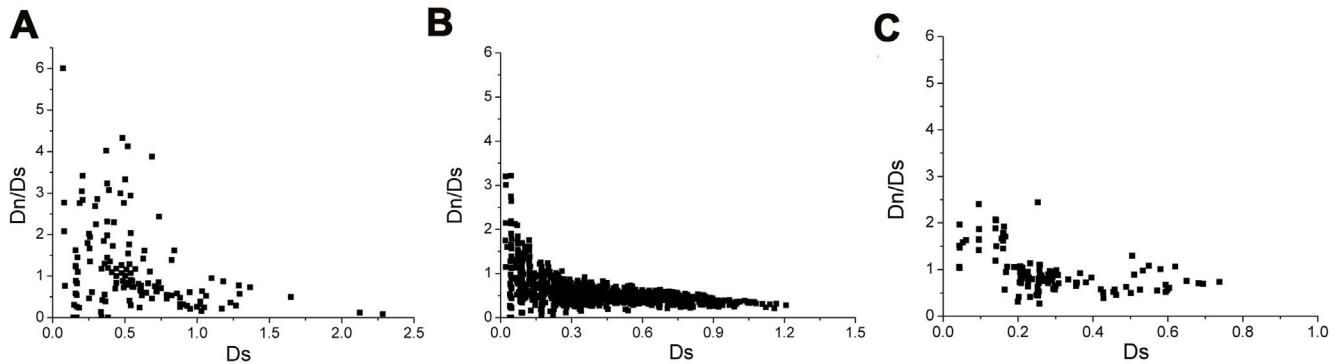


Figure 8. Nonsynonymous versus synonymous substitutions in superfamilies I, II and XVIII. The number of nonsynonymous substitutions per nonsynonymous site (Dn) over the number of synonymous substitutions per synonymous site (Ds) for interfamilial comparisons of (A) superfamily I, (B) superfamily II, (C) superfamily XVIII precursor peptide-encoding domains. $Dn/Ds > 1$ (positive or diversifying selection), $1 > Dn/Ds > 0.27$ (lack of constraints) and $Dn/Ds < 0.27$ (purifying selection). doi:10.1371/journal.pone.0100682.g008

four toxins [a, i, g and m], but not other clades. In the HWTX-XVIII superfamily, positive selection was detected in a clade containing three toxins, HWTX-XVIII [h, j and k], while HWTX-XVIIIc was identified in another clade.

Discussion

As reported previously, most spider peptide toxins are identified at three different levels: transcriptomic, peptidomic, and genomic [13,27,41–44]. However, the information obtained from venom gland cDNA libraries and protein sequencing is limited and may be biased towards the components expressed in high abundance [11,45,46]. Recently, next-generation sequencing technology resulted in an explosion of sequence data for toxin transcripts, both in terms of number and breadth [20,47–49]. The new 454 life Sciences pyrosequencing technology generates relatively long readable stretches (on average > 300 bp) that cover the full length of toxin precursors (60–120 residues). This approach allows the direct identification of peptide precursors and avoids errors inherent to the assembly of overlapping sequences (contigs) typically required for other second-generation technologies that generate shorter reads.

In this study, 123,922 Expressed Sequence Tags (ESTs) were obtained using 454 high-throughput sequencing technology. A total of 626 putative toxin precursors (containing 398 mature peptides) were unambiguously retrieved from transcriptomic data, among which 85 toxin subfamilies within ten known superfamilies and six new superfamilies were analyzed in detail. In the representative precursors of all superfamilies from *O. huwena*, primary structure analysis disclosed the presence of the PQM motif (Processing Quadruplet Motif) [45], except for families XI, XIV, XXIII, XXIV lacking the propeptide.

All mature peptides in superfamilies contained 4–12 cysteine residues, which could be classified as one of 11 types (pattern I–XI). These include five cysteine patterns observed earlier in *O. huwena* and six new patterns. Cysteine pattern I was the most common. This pattern folds into the highly stable ICK motif found in a wide range of bioactive peptides in both animal and plant kingdoms [36,50–54], indicating early evolution in the speciation of spider. As shown in Table 1, nine families shared cysteine pattern I and displayed voltage-gated sodium, potassium and calcium channel inhibition and lectin activity, indicating that ICK motif toxins are central to the success of spider evolution [36,55–58], including *O. huwena*, which uses the multiple target strategy to disrupt neuronal functions of prey and/or predators. Six novel

cysteine patterns have been identified for the first time in *O. huwena*, including three odd patterns. Cysteine pattern VII (“-C-C-CC-C-C-C-C-CC-C-C-”) with 12 residues, found in mature peptides of family XX, was the longest cysteine arrangement in *O. huwena*. Two contiguous cysteine residues were present twice in these sequences (-C3C4- and -C9C10-). These sequences showed high similarity with DKTx, which forms two independently folded domains connected by a kinked tether [59]. Cysteine pattern V (“-C-C-C-C-C-C-C-C-”) was conserved in scorpion toxin BmKAEP2 [60], kurtotoxin [61], L-4CC-Alpha toxin [62], and MkTx I [61,63,64]. Scorpion toxin displays a cysteine arrangement with I–VIII, II–V, III–VI, IV–VII disulfide bonding patterns. Moreover, other cysteine patterns of eight residues were evident. Delta-MSTX-Mb1a from Eastern mouse spider contains the cysteine pattern “-C(X₄₋₈)C(X₄₋₈)CCC(X₂₋₃)C(X₉₋₁₅)C(X₉₋₁₅)C-” and a I–IV, II–VI, III–VII, V–VIII disulfide bonding pattern [65]. The cysteine arrangement “-C(X₄₋₈)C(X₄₋₈)CC(X₁)CC(X₄₋₈)C(X₄₋₈)C-” with a I–IV, II–VI, III–VII, V–VIII disulfide bonding pattern has been reported in Iota-conotoxin RXIA (r11a) and I-superfamily conotoxins from *Comus radiatus* [66,67]. Moreover, short disintegrin CV from Sahara sand viper and Disintegrin pyramidin-A from *Echis pyramidum* leakeyi share the cysteine pattern “-C(X₄₋₈)CC(X₂₋₃)C(X₄₋₈)C(X₉₋₁₅)C(X₄₋₈)C(X₁)C-” with a I–IV, II–VI, III–VII, V–VIII disulfide bonding pattern [68]. IX, X and VIII contained an odd number of cysteine residues instead of the usual even number expected for the formation of internal disulfide bonds. The potassium channel toxin-like peptide, MeuKTX-1 [69], displays a cysteine pattern similar to that of X. The cysteine pattern VIII was a classical sub-Kunitz type motif, similar to that observed previously in *H. hainanum*. The cysteine pattern IX (“-C-C-C-C-”) was also present in the mature peptides of families XXII and XXVI. This structure is rarely found in spider venom, but frequently observed in conopeptides [19,20]. The various cysteine patterns may underlie the diverse functions of toxin peptides. Some spiders use neurotoxins as the main weapons to specifically target the nervous system for killing or paralyzing prey. Other assistant toxins, such as channel TRPV inhibitors or lectin, may enhance the toxicity of venom by binding to their targets.

In this study, we estimated the level of precursor transcription from the number of reads for each transcript. Remarkably, the transcriptome of the *O. huwena* venom gland was predominated by three toxin superfamilies (I, II and XVIII), both in terms of the mRNA level and number of peptide isoforms present, suggesting

an important role in prey capture and/or defense. Indeed, these three superfamilies account for at least 81% of all readable toxin cDNA sequences. This level of transcription for superfamilies was also accompanied by a high number of isoforms (71.2%). Homology analysis revealed important clues regarding toxin evolution (Figure 8). Superfamilies I, II and XVIII have been found in several genera, indicating that these toxins may have been derived from different ancestors. The Dn/Ds ratio exploring evolutionary patterns further showed that superfamily I has more positive selection than II and XVIII. In contrast, most toxins in the novel superfamily were found in very low abundance, indicative of assistant roles in venom. Moreover, the transcript levels may reflect those of the corresponding toxin peptides found in crude venom to a certain extent. For instance, HWTX-Ia transcript was the most highly expressed in the *O.huwena* venom gland, and its corresponding peptide, HWTX-I, the most abundant component detected in venom [50]. While the majority of toxin peptides are produced by precursors with high-level reads, surprisingly, some peptides are produced by precursors with low- and extremely low-level reads, and even some that could not be confirmed at the transcriptome level. For example, superfamily-I isoforms, CM5-24.03(HWTX-Ia) and CM3-13, previously discovered in the venom of *O.huwena* were not detected in the venom gland transcriptome. The results suggest that these toxin peptides are expressed randomly in venom, depending on environmental changes. In contrast, sequences belonging to superfamilies X and XI, which target N-type Ca²⁺ and K⁺ channels, were observed at low transcription levels, but reasonably abundant at the peptide level. One theory is that evolutionary pressures influence the level of expressed toxin peptides [70], but we believe that this is only a partial explanation, since evolution is a lengthy process that cannot create such differences between individuals within a short time. As described previously, highly expressed toxins detected with the cDNA library, as well as all full-length sequences and majority (10/17) of partial sequences identified via Edman degradation were detected via 454 sequencing, implying that highly expressed transcripts and abundant peptides are conservative constituents of venom. Unexpectedly, abundant transcriptomic profile differences were detected with the 454 transcriptome and cDNA library approaches. The dynamics of transcriptional changes demonstrate that venom samples of individual *O.huwena* are not constant in composition and vary

dramatically with time. Moreover, this considerable variation may be associated with changes in diet, environment or replete/milk [30,32,33,71]. Evolutionary innovations are proposed to be a result of infidelity of transcription and heterogeneity inherent to most biological processes, leading to genetic and phenotypic variations [72].

The large number of toxin variants are identified at transcriptional level is an expected finding. Notably, the majority of these modified mRNAs are also transcribed at low levels, which explains why these rare sequences have eluded detection in previous studies using traditional transcriptomic approaches. Toxin gene sequences include single base mutations, deletions, insertions and frame shifts, generating amino acid mutations, insertions and deletions, alternative cleavage sites and cysteine patterns, and highly variable isoforms within families that could be identified at the peptide level. Together with the variable peptide processing described previously, this mechanism contributes to the hypervariability of venom peptides and their ability to evolve rapidly. Post-translational modifications are relatively common in spider venom peptides [13,73,74], and play an important role in the modulation of biological activity. The commonly observed C-terminal amidation and trimming may generate additional isoforms. Based on the present results, we propose that both genetic and post-translational modifications contribute to overall toxin diversity. Therefore, background genetic diversity, in addition to the generation of highly variable transcripts and peptide processing, appears to underlie overall venom peptide diversity. In conclusion, the numerous toxins in spider venom and modification mechanisms have enabled the spider to adapt to its specific environment during evolution.

Supporting Information

Table S1 626 toxin precursors were categorized into 16 gene superfamilies.
(XLS)

Author Contributions

Conceived and designed the experiments: XZ SL. Performed the experiments: Y.Zhang J.Liu J.Luo LZ SL PH XC. Analyzed the data: Y.Zhang. Contributed reagents/materials/analysis tools: YH QH. Wrote the paper: Y.Zhang.

References

- Vassilevski AA, Kozlov SA, Grishin EV (2009) Molecular diversity of spider venom. *Biochemistry (Mosc)* 74: 1505–1534.
- Isbister GK, Fan HW (2011) Spider bite. *Lancet* 378: 2039–2047.
- Dutertre S, Lewis RJ (2010) Use of venom peptides to probe ion channel structure and function. *J Biol Chem* 285: 13315–13320.
- Milescu M, Bosmans F, Lee S, Alabi AA, Kim JI, et al. (2009) Interactions between lipids and voltage sensor paddles detected with tarantula toxins. *Nat Struct Mol Biol* 16: 1080–1085.
- Swartz KJ (2007) Tarantula toxins interacting with voltage sensors in potassium channels. *Toxicon* 49: 213–230.
- Milescu M, Vobecky J, Roh SH, Kim SH, Jung HJ, et al. (2007) Tarantula toxins interact with voltage sensors within lipid membranes. *J Gen Physiol* 130: 497–511.
- Phillips LR, Milescu M, Li-Smerin Y, Mindell JA, Kim JI, et al. (2005) Voltage-sensor activation with a tarantula toxin as cargo. *nature* 436: 857–860.
- Lee SY, MacKinnon R (2004) A membrane-access mechanism of ion channel inhibition by voltage sensor toxins from spider venom. *nature* 430: 232–235.
- Rash LD, Hodgson WC (2002) Pharmacology and biochemistry of spider venoms. *Toxicon* 40: 225–254.
- Vetter I, Davis JL, Rash LD, Anangi R, Mobli M, et al. (2011) Venomics: a new paradigm for natural products-based drug discovery. *Amino Acids* 40: 15–28.
- Liang S (2008) Proteome and peptidome profiling of spider venoms. *Expert Rev Proteomics* 5: 731–746.
- Klint JK, Senff S, Rupasinghe DB, Er SY, Herzig V, et al. (2012) Spider-venom peptides that target voltage-gated sodium channels: pharmacological tools and potential therapeutic leads. *Toxicon* 60: 478–491.
- Jiang L, Zhang D, Zhang Y, Peng L, Chen J, et al. (2010) Venomics of the spider *Ornithoctonus huwena* based on transcriptomic versus proteomic analysis. *Comp Biochem Physiol Part D Genomics Proteomics* 5: 81–88.
- Jin AH, Dutertre S, Kaas Q, Lavergne V, Kubala P, et al. (2013) Transcriptomic messiness in the venom duct of *Conus miles* contributes to conotoxin diversity. *Mol Cell Proteomics*.
- Jiang L, Peng L, Chen J, Zhang Y, Xiong X, et al. (2008) Molecular diversification based on analysis of expressed sequence tags from the venom glands of the Chinese bird spider *Ornithoctonus huwena*. *Toxicon* 51: 1479–1489.
- Liang S, Li X, Cao M, Xie J, Chen P, et al. (2000) Identification of venom proteins of spider *S. huwena* on two-dimensional electrophoresis gel by N-terminal microsequencing and mass spectrometric peptide mapping. *J Protein Chem* 19: 225–229.
- Yuan C, Jin Q, Tang X, Hu W, Cao R, et al. (2007) Proteomic and peptidomic characterization of the venom from the Chinese bird spider, *Ornithoctonus huwena* Wang. *J Proteome Res* 6: 2792–2801.
- Durban J, Juarez P, Angulo Y, Lomonte B, Flores-Diaz M, et al. (2011) Profiling the venom gland transcriptomes of Costa Rican snakes by 454 pyrosequencing. *BMC Genomics* 12: 259.
- Lluisma AO, Milash BA, Moore B, Olivera BM, Bandyopadhyay PK (2012) Novel venom peptides from the cone snail *Conus pulicarius* discovered through

- next-generation sequencing of its venom duct transcriptome. *Mar Genomics* 5: 43–51.
20. Terrat Y, Biass D, Dutertre S, Favreau P, Remm M, et al. (2012) High-resolution picture of a venom gland transcriptome: case study with the marine snail *Conus* consors. *Toxicon* 59: 34–46.
 21. Zhang J, Chiodini R, Badr A, Zhang G (2011) The impact of next-generation sequencing on genomics. *J Genet Genomics* 38: 95–109.
 22. Liu L, Li Y, Li S, Hu N, He Y, et al. (2012) Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012: 251364.
 23. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.
 24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25: 25–29.
 25. Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9: 299–306.
 26. Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 5: 150–163.
 27. Fernandes-Pedrosa Mde F, Junqueira-de-Azevedo Ide L, Goncalves-de-Andrade RM, Kobashi LS, Almeida DD, et al. (2008) Transcriptome analysis of *Loxosceles laeta* (Araneae, Sicariidae) spider venomous gland using expressed sequence tags. *BMC Genomics* 9: 279.
 28. Chen J, Zhao L, Jiang L, Meng E, Zhang Y, et al. (2008) Transcriptome analysis revealed novel possible venom components and cellular processes of the tarantula *Chilobrachys jingzhao* venom gland. *Toxicon* 52: 794–806.
 29. Nisani Z, Dunbar SG, Hayes WK (2007) Cost of venom regeneration in *Parabuthus transvaalicus* (Arachnida: Buthidae). *Comp Biochem Physiol A Mol Integr Physiol* 147: 509–513.
 30. Jakubowski JA, Kelley WP, Sweedler JV, Gilly WF, Schulz JR (2005) Intraspecific variation of venom injected by fish-hunting *Conus* snails. *J Exp Biol* 208: 2873–2883.
 31. Davis J, Jones A, Lewis RJ (2009) Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. *Peptides* 30: 1222–1227.
 32. Dutertre S, Biass D, Stocklin R, Favreau P (2010) Dramatic intraspecific variations within the injected venom of *Conus* consors: an unsuspected contribution to venom diversity. *Toxicon* 55: 1453–1462.
 33. Zelanis A, Andrade-Silva D, Rocha MM, Furtado MF, Serrano SM, et al. (2012) A transcriptomic view of the proteome variability of newborn and adult *Bothrops jararaca* snake venoms. *PLoS Negl Trop Dis* 6: e1554.
 34. Zhou PA, Xie XJ, Li M, Yang DM, Xie ZP, et al. (1997) Blockade of neuromuscular transmission by huwentoxin-I, purified from the venom of the Chinese bird spider *Selenocosmia huwena*. *Toxicon* 35: 39–45.
 35. Peng K, Chen XD, Liang SP (2001) The effect of Huwentoxin-I on Ca(2+) channels in differentiated NG108–15 cells, a patch-clamp study. *Toxicon* 39: 491–498.
 36. Liu Z, Dai J, Dai L, Deng M, Hu Z, et al. (2006) Function and solution structure of Huwentoxin-X, a specific blocker of N-type calcium channels, from the Chinese bird spider *Ornithoctonus huwena*. *J Biol Chem* 281: 8628–8635.
 37. Peng K, Lin Y, Liang SP (2006) Nuclear magnetic resonance studies on huwentoxin-XI from the Chinese bird spider *Ornithoctonus huwena*: 15N labeling and sequence-specific 1H, 15N nuclear magnetic resonance assignments. *Acta Biochim Biophys Sin (Shanghai)* 38: 457–466.
 38. Yuan CH, He QY, Peng K, Diao JB, Jiang LP, et al. (2008) Discovery of a distinct superfamily of Kunitz-type toxin (KTT) from tarantulas. *PLoS ONE* 3: e3414.
 39. Zakharaeva E, Horvath MP, Goldenberg DP (2008) Functional and structural roles of the Cys14-Cys38 disulfide of bovine pancreatic trypsin inhibitor. *J Mol Biol* 382: 998–1013.
 40. Tang X, Zhang Y, Hu W, Xu D, Tao H, et al. (2010) Molecular diversification of peptide toxins from the tarantula *Haplopelma hainanum* (*Ornithoctonus hainana*) venom based on transcriptomic, peptidomic, and genomic analyses. *J Proteome Res* 9: 2550–2564.
 41. Diego-Garcia E, Peigneur S, Waekens E, Debaveye S, Tytgat J (2010) Venom components from *Citharus chrysus* spider (Family Theraphosidae): exploring transcriptome, venomics, and function. *Cell Mol Life Sci* 67: 2799–2813.
 42. Liao Z, Cao J, Li S, Yan X, Hu W, et al. (2007) Proteomic and peptidomic analysis of the venom from Chinese tarantula *Chilobrachys jingzhao*. *Proteomics* 7: 1892–1907.
 43. Duan Z, Cao R, Jiang L, Liang S (2012) A combined de novo protein sequencing and cDNA library approach to the venom analysis of Chinese spider *Aranus ventricosus*. *J Proteomics*.
 44. Zhang Y, Chen J, Tang X, Wang F, Jiang L, et al. (2010) Transcriptome analysis of the venom glands of the Chinese wolf spider *Lycosa singoriensis*. *Zoology (Jena)* 113: 10–18.
 45. Kozlov S, Malyavka A, McCutchen B, Lu A, Schepers E, et al. (2005) A novel strategy for the identification of toxinlike structures in spider venom. *Proteins* 59: 131–140.
 46. Pan Z, Barry R, Lipkin A, Soloviev M (2007) Selection strategy and the design of hybrid oligonucleotide primers for RACE-PCR: cloning a family of toxin-like sequences from *Agelena orientalis*. *BMC Mol Biol* 8: 32.
 47. Violette A, Biass D, Dutertre S, Koua D, Piquemal D, et al. (2012) Large-scale discovery of conopeptides and conoproteins in the injectable venom of a fish-hunting cone snail using a combined proteomic and transcriptomic approach. *J Proteomics*.
 48. Laverge V, Dutertre S, Jin AH, Lewis RJ, Taft RJ, et al. (2013) Systematic interrogation of the *Conus marmoreus* venom duct transcriptome with ConoSorter reveals 158 novel conotoxins and 13 new gene superfamilies. *BMC Genomics* 14: 708.
 49. Margres MJ, Aronow K, Loyacano J, Rokyta DR (2013) The venom-gland transcriptome of the eastern coral snake (*Micrurus fulvius*) reveals high venom complexity in the intragenomic evolution of venoms. *BMC Genomics* 14: 531.
 50. Liang S (2004) An overview of peptide toxins from the venom of the Chinese bird spider *Selenocosmia huwena* Wang [= *Ornithoctonus huwena* (Wang)]. *Toxicon* 43: 575–585.
 51. Peng K, Shu Q, Liu Z, Liang S (2002) Function and solution structure of huwentoxin-IV, a potent neuronal tetrodotoxin (TTX)-sensitive sodium channel antagonist from Chinese bird spider *Selenocosmia huwena*. *J Biol Chem* 277: 47564–47571.
 52. Zhang D, Liang S (1993) Assignment of the three disulfide bridges of huwentoxin-I, a neurotoxin from the spider *selenocosmia huwena*. *J Protein Chem* 12: 735–740.
 53. Lu S, Liang S, Gu X (1999) Three-dimensional structure of *Selenocosmia huwena* lectin-I (SHL-I) from the venom of the spider *Selenocosmia huwena* by 2D-NMR. *J Protein Chem* 18: 609–617.
 54. Huang RH, Liu ZH, Liang SP (2003) Purification and characterization of a neurotoxic peptide huwentoxin-III and a natural inactive mutant from the venom of the spider *Selenocosmia huwena* Wang (*Ornithoctonus huwena* Wang). *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)* 35: 976–980.
 55. Wang M, Rong M, Xiao Y, Liang S (2012) The effects of huwentoxin-I on the voltage-gated sodium channels of rat hippocampal and cockroach dorsal unpaired median neurons. *Peptides* 34: 19–25.
 56. Xiao Y, Jackson JO, 2nd, Liang S, Cummins TR (2011) Common molecular determinants of tarantula huwentoxin-IV inhibition of Na⁺ channel voltage sensors in domains II and IV. *J Biol Chem* 286: 27301–27310.
 57. Wang RL, Yi S, Liang SP (2010) Mechanism of action of two insect toxins huwentoxin-III and hainantoxin-VI on voltage-gated sodium channels. *J Zhejiang Univ Sci B* 11: 451–457.
 58. Deng M, Luo X, Meng E, Xiao Y, Liang S (2008) Inhibition of insect calcium channels by huwentoxin-V, a neurotoxin from Chinese tarantula *Ornithoctonus huwena* venom. *Eur J Pharmacol* 582: 12–16.
 59. Bohlen CJ, Priel A, Zhou S, King D, Siemens J, et al. (2010) A bivalent tarantula toxin activates the capsaicin receptor, TRPV1, by targeting the outer pore domain. *Cell* 141: 834–845.
 60. Zhou XH, Yang D, Zhang JH, Liu CM, Lei KJ (1989) Purification and N-terminal partial sequence of anti-epilepsy peptide from venom of the scorpion *Buthus martensii* Karsch. *Biochem J* 257: 509–517.
 61. Sidach SS, Mintz IM (2002) Kurtoxin, a gating modifier of neuronal high- and low-threshold Ca channels. *J Neurosci* 22: 2023–2034.
 62. Sharma M, Ethayathulla AS, Jabeen T, Singh N, Sarvanan K, et al. (2006) Crystal structure of a highly acidic neurotoxin from scorpion *Buthus tamulus* at 2.2 Å resolution reveals novel structural features. *J Struct Biol* 155: 52–62.
 63. Gong J, Kini RM, Gwee MC, Gopalakrishnakone P, Chung MC (1997) Makatoxin I, a novel toxin isolated from the venom of the scorpion *Buthus martensii* Karsch, exhibits nitroergic actions. *J Biol Chem* 272: 8320–8324.
 64. Chuang RS, Jaffe H, Cribbs L, Perez-Reyes E, Swartz KJ (1998) Inhibition of T-type voltage-gated calcium channels by a new scorpion toxin. *Nat Neurosci* 1: 668–674.
 65. Gunning SJ, Chong Y, Khalife AA, Hains PG, Broady KW, et al. (2003) Isolation of delta-missulenatoxin-Mb1a, the major vertebrate-active spider delta-toxin from the venom of *Missulena bradleyi* (Actinopodidae). *FEBS Lett* 554: 211–218.
 66. Jimenez EC, Shetty RP, Lirazan M, Rivier J, Walker C, et al. (2003) Novel excitatory *Conus* peptides define a new conotoxin superfamily. *J Neurochem* 85: 610–621.
 67. Buczek O, Wei D, Babon JJ, Yang X, Fiedler B, et al. (2007) Structure and sodium channel activity of an excitatory II-superfamily conotoxin. *Biochemistry* 46: 9929–9940.
 68. Sanz L, Bazaa A, Marrakchi N, Perez A, Chenik M, et al. (2006) Molecular cloning of disintegrins from *Cerastes vipera* and *Macrovipera lebetina* transmediterranea venom gland cDNA libraries: insight into the evolution of the snake venom integrin-inhibition system. *Biochem J* 395: 385–392.
 69. Gao B, Zhu S (2012) Accelerated evolution and functional divergence of scorpion short-chain K⁺ channel toxins after speciation. *Comp Biochem Physiol B Biochem Mol Biol* 163: 238–245.
 70. Duda TF, Jr., Palumbi SR (2000) Evolutionary diversification of multigene families: allelic selection of toxins in predatory cone snails. *Mol Biol Evol* 17: 1286–1293.
 71. Morgenstern D, Rohde BH, King GF, Tal T, Sher D, et al. (2011) The tale of a resting gland: transcriptome of a replete venom gland from the scorpion *Hottentotta judaicus*. *Toxicon* 57: 695–703.
 72. Tawfik DS (2010) Messy biology and the origins of evolutionary innovations. *Nat Chem Biol* 6: 692–696.

73. Rong M, Duan Z, Chen J, Li J, Xiao Y, et al. (2013) Native pyroglutamation of huwentoxin-IV: a post-translational modification that increases the trapping ability to the sodium channel. *PLoS ONE* 8: e65984.
74. Escoubas P, Quinton L, Nicholson GM (2008) Venomics: unravelling the complexity of animal venoms with mass spectrometry. *J Mass Spectrom* 43: 279–295.