Molecular diversity and evolutionary trends of cysteine-rich peptides from the venom glands of Chinese spider Heteropoda venatoria

Yiying Ding
Hunan Agricultural University
Kezhi Chen
Hunan Agricultural University
Xuewen Zhang
Hunan Agricultural University
Tiaoyi Xiao
Hunan Agricultural University
Jinjun Chen (chhncjj@126.com)
Hunan Agricultural University https://orcid.org/0000-0001-5495-051X

Research

Keywords: Spider, Venom, Cysteine-Rich Peptide, Molecular Diversity, Evolution

Posted Date: April 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-24220/v1

Abstract

Background: The huntsman spider (Heteropoda venatoria Carl Linnaeus, 1767) in family Sparassidae, is highly valued in tropical and subtropical countries because the species capture and feed on cockroaches and other domestic insect pests. Unlike most other species of Araneomorphae, the huntsman spiders do not use webs to capture prey. Their great speed and strong chelicerae (mouthparts) with toxin glands are used to capture the insects.

Results: We identified 154 novel putative cysteine-rich peptide toxin precursors by analyzing expressed sequence tags of the spider H. venatoria venom gland. The sequences of cysteine-rich peptide precursor revealed 24 families based on the phylogenetics analyses of signal peptide and cysteine framework in mature region, including 8 families of classic Inhibitory cystine knot toxins, 2 families of novel 6-cys motifs, 13 families of long cysteine-rich peptides with 8, 10 and 12-cys, and one family of 2-cys peptides. Intriguingly, four kinds of motifs are first described in spider venom. Furthermore, combining the diverse cysteine-rich peptide sequences of H. venatoria with the sequences from represent spiders explored previously, the dynamic evolutionary trends of venom cysteine-rich peptides were investigated based on the analysis of the structures of precursors and the patterns of cysteine scaffolds in the phylogenetic framework.

Conclusion: H. venatoria is an appropriate intermediate species for the evolutionary analysis of spider peptide toxins from Mygalomorphae to Araneomorphae with a holistic view. This work revealed the dynamic evolutionary trends of venom cysteine-rich peptides of spider: the mature peptides have been developed longer with more cysteines; and the propeptides between the signal and mature peptides have been diminished and even vanished. With respect to potential insecticidal applications, the work provides promising new templates and gene clones for further exploration.

Background

Spiders (*Araneae*), including 48,504 described species grouped into 4,156 genera and 120 families https://wsc.nmbe.ch/, have immense biological and ecological diversity. The vast majority of spiders employ extraordinary chemical and pharmacological complicated venoms to rapidly subdue their prey. Spider venoms contain a plethora of compounds including proteins, peptides, and low-molecular-weight components. The predominant peptides of spider venoms are secretory cysteine-rich peptides (CRPs) with multiple disulfide bonds that serve to provide stability and resistance to protease degradation [1]. There has been tremendous interest in studying the biochemical and structural properties [2, 3], pharmacological applications [4], evolution and diversification of secretory cysteine-rich peptide toxin [5–9]. Up to now, the venoms of snakes and Conus snails are much more deeply understood than those of spiders partly due to the fact that spiders have enormous species and extremely complicated venoms, but with relatively small venom glands filled with venom at microliter level, which imply the high-efficiency of spider toxins in turn [10, 11].

The huntsman spider (*Heteropoda venatoria* Carl Linnaeus, 1767), is highly valued in pantropical countries because the species capture and feed on cockroaches and other domestic insect pests. The spider captures the preys directly instead of by spinning webs, their great speed and strong chelicerae (jaws) are used to capture the insects on which they feed [12]. The venom of *H. venatoria* contains hundreds of peptides with severe toxicities on *Periplaneta americana* (LD_{50}: 28.18 mg/g of body weight), and venom neurotoxins target specific types of insect ion channels and receptors, which have wide applied potential as insecticides in pest control [13]. Many studies have focused on the activity and composition of the venom of *H. venatoria*. However, the diversity of CRPs in the venom as well as the evolutionary relationship in the phylogenetic framework has not been explored.

In the study, Sanger sequencing method was employed for the construction of the venom gland cDNA library and revealed the complex toxin repertoire of *H. venatoria* venom. Furthermore, the dynamic evolution of CRPs is explored by combining the data from well described venom gland ESTs retrieved from ArachnoServer [14, 15]. Our research results contribute to understanding the evolutionary trends and diversity of *H. venatoria* CRPs in spider venom.

Materials And Methods

Spider collection

The spiders *H. venatoria* were collected by sweeping and visually searching from the old buildings in the farm without pesticides for scientific research of Hunan Agricultural University, Changsha, China (28°18'33" N, 113°07'69" E).

cDNA library construction and expression sequence tag sequencing

The cDNA library was prepared and sequenced as previously described [16]. Simply, eight adult female spiders were aggravated to secret their venom gland contents and encourage production of venom transcripts [17]. Four days later the venom glands of the eight spiders were harvested and homogenized in liquid nitrogen and in the presence of TRIzol reagent (Invitrogen). Polyadenylic acid (+) [polyA(+)] - containing RNAs were purified from the total RNA on an oligo(dT)-cellulose affinity column using the mRNA Purification Kit (Promega) according to the manufacturer's protocol. The full length cDNA library was constructed as described in the CreatorTM SMART™ cDNA Library Construction Kit (Clontech). The polymerase chain reaction was performed with the M13 forward and reverse primers from the kit to rapidly screen recombinant clones. The clones containing inserts ≥ 500 base pairs were grown in LB medium containing chloramphenicol (30 mg/mL) in 96-well plates for 16 h. The plasmids were extracted by alkaline lysis and sequenced from the 5'-end on an automated ABI PRISM 3700 sequencer (Perkin Elmer) using the T7 promoter primer and ABI PRISM® Big Dye™ terminator v3.1 ready reaction cycle sequencing kit (Applied Biosystems).

CRPs identification and evolutionary analyses

The sequenced cDNA were trimmed by removal of vector, primer sequences and poly(A) tails with ABI PRISM® DNA Sequencing Analysis Software V.3.3 [18]. The consensus sequences of each cluster were further filtered by screening for homology to ribosomal RNA, mitochondrial DNA and *E. coli* genome sequences[19]. After deleting matches, the remaining sequences were searched against public databases (nr/NCBI, SwissProt/UniProtKB and TrEMBL/UniProtKB) using the BLASTn or BLASTx programs to identify putative functions of the new expression sequence tags (ESTs) [20]. The signal peptides were predicted with the SignalP 4.1 program (http://www.cbs.dtu.dk/services/SignalP/)[21] and SpiderP (http://www.arachnoserver.org) [22]. Furthermore, the putative CRPs were searched in KNOTTIN database (http://knottin.cbs.cnrs.fr) [23, 24]. Multiple sequences of precursors were aligned using the ClustalW program [25]. The resulting alignments were then hand-edited using the BioEdit program (http://www.mbio.ncsu.edu/BioEdit/BioEdit.html). Sequences were aligned using Clustal X 2.0, and gapped positions were omitted from subsequent analyses. The resulting alignments were imported into MEGA software to construct a phylogenetic tree with the neighbor-joining method [26, 27].

Results

The morphology, Habits and phylogenetic status of H. venatoria

The spider *H. venatoria* is a member of family Sparassidae, common name giant crab spider, and takes more than 250 days to complete their life cycle, which leads to the collection of natural venoms is limited greatly [28]. Adult specimens have a flattened body length of about 2.5 cm with 8 long slightly hairy legs spanning 6 to 12 cm. A yellow to cream clypeus just in front of the eyes is one of the main features of *H. venatoria*. The female has a larger abdomen with an overall brown body. Usually, an egg sac up to about 2.5 centimeters wide was carried with her pedipalps under its body. The male has a slender body and longer legs, a distinctive pattern on his carapace. This species is found in many tropical and subtropical regions of the world, and can't survive outside during sub-freezing temperatures. In present study, the spiders were found from basements, barns, and greenhouses of our scientific research farm in summer.

The huntsman spiders do not spin webs. They are known to hunt and feed on living insects with their exceptional agility and speed at night. They can stay and run on a vertical smooth surface, as well as contort and squeeze their large body to fit into surprisingly small cracks and crevices, which give them a strong advantage both in predation and evading predators. Almost as soon as they catch their prey, the spiders paralyze them by injecting with the venoms which are from glands (Fig. 1) extending from the chelicerae into the cephalothorax. The spider is considered a beneficial resident of households because it can hunt pests efficiently and does no harm to people.

The phylogenetic relationships (Fig. 2) amongst the spider families whose venom CRPs have been well described shows *H. venatoria* (in Sparassidae) locate in the middle between Theraphosidae and Lycosidae [16, 29–35]. Compared with Mygalomorphae species (*Haplopelma huwenum* (now named

Haplopelma Schmidti), *Haplopelma hainanum*, *Grammostola rosea*, *Chilobrachys jingzhao*) and Araneomorphae species (*Lycosa singoriensis*, *Dolomedes mizhoanus*, *Araneus ventricosus*), the body size of *H. venatoria*, who belongs to the more primitive Araneomorphae, is medium between the two suborders (Fig. 2). Although sensitive to light like most spiders, *H. venatoria* is a large indoor terrestrial spider and hides under a simple bunker during the day and goes out hunting at night. Unlike web-forming spiders (*A. ventricosus, A. orientalis et al.*) that weave cobwebs before prey and then eat after wrapping its prey with cobwebs, *H. venatoria* subdues prey by long legs, strong chelicerae and complex venom, which seems similar to tarantula's predation.

Based on the comparisons of morphology and predation habits of *H. venatoria* along with other spiders, as well as its phylogenetic classification, we believe that the investigation of the genetic coding products in its venom, the most convergent of spider traits, can contribute to the understanding of Araneae toxins evolution in the context of ecology, as well as the recognition and likely facilitate exploration of the popular spider resources in the cosmotropical regions.

Identification and nomeclature of CRPs

From the cDNA library of *H. venatoria* venom gland, 912 sequenced ESTs and 154 predicted novel CRP precursors were obtained. The cDNA sequences of CRPs have been submitted into the public database (http://www.ncbi.nlm.nih.gov/entrez, GenBank accession numbers: KC145575 - KC145728). The presence and location of signal peptide cleavage sites in the amino acid sequences were predicted with SignalP 4.1 and SpiderP program. A full-length CRPs precursor absolutely contains a signal peptide and a mature peptide, while some CRPs precursors contain a propeptide in addition preceding the mature toxin sequence just as the precursors of the vast majority of spider peptide toxins reported before [36]. Alignment of the resultant amino acid sequences revealed extensive variation in the molecular structure of the transcripts for most CRP types.

Family /cluster identification

In the present study, the sequences of CRP precursors revealed 24 families based on the sequences of signal peptide and cysteine framework. The formation of disulfide bonds stabilizes the three-dimensional (3D) structures of toxins, and is commonly used to classify toxins.

Family A-H. The full primary sequences of the CRPs in the Family A-H are comprised of a signal sequence (19–25 residues) and a propeptide (11–19 residues) preceding the mature toxin sequence. The N- and C-terminus of mature peptides are highly variable regions. The 11 members of Family A are homologs of Kappa-SPRTX-Hv1c, including its five different precursors (κ-SPRTX-Hv1c_1–5). Signal peptide mode of CRPs in the family is 'MKh$_{12}$Sh$_5$', where 'h' indicates hydrophobic residue, the Arabic numerals denote the number of residues, and capital letters indicate the corresponding amino acids. The propeptides of Family A are 19 residues with highly conserved DEQR as an endoproteolytic site preceding the mature peptides, named the Processing Quadruplet Motif (PQM)[37]. Mature peptide mode of family A is 'XCX$_6$CX$_5$CCX$_4$CX$_3$CX$_{4-6}$', where 'X' is any amino acid. On the C-terminal of the mature peptides, there is

'GK' as the amidation site. The characters of the signal peptide, propeptide and mature peptide of family B-H are compared with those of family A, shown in Table 1. The mature peptides of Family A-H show the 'classical' Inhibitory Cystine Knot (ICK) motif containing three disulfide bonds with I–IV, II–V and III–VI connectivity. The first two disulfide bonds (I–IV and II–V) form an embedded ring which is threaded by the third disulfide bond (III–VI). The backbone regions between successive Cys residues are referred to as loops, numbered starting with loop 1 between Cys I and Cys II [38, 39]. There is less amino acid sequence divergence in loop 1 and 3 than in the much more variable loop 2, 4, N- and C-terminus in mature peptides. The precursors of Family A – G have higher similarity with those from the same species than others. Only the sequences of CRPs in the Family H show high homology with U23-ctenitoxin-Pn1a and U4-agatoxin-Ao1a from *Phoneutria nigriventer* and *Agelena orientalis* respectively (Additional file 1: Fig. S1).

Family I. The precursors of Family I have a high content of acidic amino acid in putative mature peptides with a novel Cys scaffold '$CX_5CX_3CX_5CXCXC$'. Since no significant homologous sequence has been found in public protein databases, posttranscriptional processes such as alternative splicing or posttranslational modifications remain uncertain. In this case, the dotted border indicates the putative short propeptides shown in Additional file 2: Fig. S2. The motif of Family I is the first time reported from spider venom.

Family J. Family J includes eleven homologous sequences which are characterized by two consecutive Cys residues in the middle of signal peptide and that is straight followed by the mature region with a cys-scaffold '$CX_{13}CX_2CX_{12}CX_3CX_8C$'. The scaffold in mature peptide has been identified as a conserved domain pfam01147 (representative proteins gi: 221468699, 5921747 shown in Additional file 3: Fig. S3) which includes all known crustacean hyperglycemic hormones (CHHs) found in the sinus gland of isopods and decapods [40] and the molt-inhibiting hormone (MIH) of the lobster *Homarus americanus* [41]. The three disulfide bridges are C_I–C_V, C_{II}–C_{IV}, and C_{III}–C_{VI} [42]. In addition, the amino acid sequences of several translated cDNA (gi: 304306070, 304307035, 304306844, 304306583) from *Loxosceles intermedia* venom gland library [43] are also similar to that of Family J as shown in Additional file 3: Fig. S3. The latrodectins which are identified in widow spiders venom glands also share six conserved cysteines that adopt the same disulfide bond pairing in the mature peptide [44, 45].

Family K, L, M and N. The four families have eight cysteines with a typical motif '$C_IX_6C_{II}X_nC_{III}C_{IV}X_4C_VXC_{VI}X_nC_{VII}XC_{VIII}$' where X is any residue but cystine. However, the amounts and properties of residues in the loops of C_{II}-C_{III}, C_{VI}-C_{VII} and the C-terminus are different from one another. The sequences of the signal and precursor proteins, as well as endoproteolytic sites are also diverse. In Family N, there is a long loop between C_{VI} and C_{VII}, and a very short propeptide preceding the mature region. Especially, there is no propeptide predicted in the precursor of U32-sparatoxin-Hv1a. The amino acids sequences of Family K, L, M and N are aligned with the most similar known homologs, respectively, shown in Additional file 4: Fig. S4.

Family O. Family O includes 20 unique sequences which are highly homologous. It is noteworthy that Family O is a novel venom peptide type with a high expression level in *H. venatoria*. There are twelve residues between the signal peptides and the first Cys, but no usual PQM. The mature region is characterized by a novel eight-Cys scaffold '$C_IX_{21}C_{II}X_4C_{III}X_9C_{IV}X_{10}C_VX_{11}C_{VI}C_{VII}X_4C_{VIII}$'. The transcript of a secretory protein with the identical Cys- scaffold has been identified from black widow spider (gi: 318087504). However, it is hypothesized to be involved in wrapping silk fibers. Moreover, several hypothetical non-secretory proteins from *Amblyomma maculatum* also adopt the same eight-Cys scaffold (Additional file 5: Fig. S5).

Family P. The six precursors are highly homologous with Omega-agatoxin-1A (gi: 2507406) from *Agelenopsis aperta* containing a ten-Cys scaffold '$C_IX_{6/8}C_{II}XC_{III}X_6C_{IV}XC_VX_{7/11}C_{VI}XC_{VII}X_7C_{VIII}X_5C_{IX}X_{19/20}C_X$', so they belong to the omega-agatoxin superfamily which has a particularly interesting feature of the prepropeptide with the occurrence of two glutamate-rich sequences interposed between the signal sequences, the major peptide toxin, and the minor toxin peptide. Heterodimer of the two subunits are linked by a disulfide bond [46] (Additional file 6: Fig. S6).

Family Q and R. Family Q is homologous with U19-ctenitoxin-Pn1a (gi: 50401390), Hainantoxin-XIV-7 (gi: 310946827), HWTX-XIVa2 (gi:166007861) precursor and a toxin-like peptide (gi: 380692240) from *Grammostola rosea*. Family R is homologous with U3-aranetoxin-Ce1a (gi: 27805756). The precursors in both families contain a signal peptide and a mature peptide with a ten-Cys scaffold like '$C_IX_nC_{II}X_4C_{III}C_{IV}X_nC_VX_9C_{VI}X_nC_{VII}XC_{VIII}X_5C_{IX}X_nC_X$'. However, the residues are very different in the loops between cysteines. The loops of C_{IV}-C_V, C_{VI}-C_{VII} and C_{IX}-C_X are longer in Family R than those in Family Q (Additional file 7: Fig. S7).

Family S, T and U. All the three families are composed of a signal peptide and a mature region with a ten-Cys scaffold '$C_IX_7C_{II}X_8C_{III}C_{IV}X_4C_VX_5C_{VI}C_{VII}X_3C_{VIII}X_3C_{IX}X_{17}C_X$', which has a high degree of similarity to the Cystine scaffold of U7-agatoxin-Ao1a (gi: 74845728) and U20-lycotoxin-Ls1a/c/d (gi: 313471673/313471696/313471677) from *Agelena orientalis* and *Lycosa singoriensis* respectively. The amino acids in the loop between C_{VIII} and C_{IX} are conserved in the peptides even from different spider families (Sparassidae, Agelenidae and Lycosoidea). The sequences between C_{IV} and C_V are also conserved in the three families from the same spider *H. venatoria*. However, those in other spaces are less homologous, especially, the amino acid sequences in the N- and C- terminus are much various and diverse in Family S, T and U (Additional file 8: Fig. S8).

Family V. The predicted peptide sequences in Family V have a similar disulfide bonding pattern and structure to U9-agatoxin-Ao1a (gi: 74845712). Their common Cys bonding pattern is '$C_IX_6C_{II}X_3C_{III}XC_{IV}C_VX_5C_{VI}XC_{VII}X_4C_{VIII}XC_{IX}X_8C_XX_6C_{XI}X_{12}C_{XII}$'. However, the sequences are much different in signal peptides, propeptides and loops between the cysteines, and PQM is apparent in U9-agatoxin-Ao1a, rather than in the precursors of Family V from *H. venatoria* (Additional file 9: Fig. S9).

Family W. There are three secretory proteins with a long 12-Cys scaffold in Family W, which have mutual Cys bonding pattern is '$C_IX_7C_{II}X_{23}C_{III}X_9C_{IV}X_7C_VX_{22}C_{VI}X_{15}C_{VII}X_{11}C_{VIII}X_{11}C_{IX}X_8C_XX_8C_{XI}X_{22}C_{XII}$'. The precursors had no homologs when they were aligned against the Database of GenBank, EMBL and DDBJ. However, 2 sequences from spider EST database were matched by using TBLASTN, which have not identified as toxins. The amino acid sequences of gi: 304306221 and gi: 189216028, which are in the cDNA library from *Loxosceles intermedia* venom gland and *Acanthoscurria gomesiana*, respectively, are homologous to U28-sparatoxin-Hv1a with 43% (E-value is 5e-07) and 48% (E-value is 1e-06) positives (Additional file 10: Fig. S10).

Family X. The two precursors in Family X with only two cysteines '$C_IX_6C_{II}X_{16}$' in the mature region which have no significant sequence homolog in the Database of GenBank, EMBL and DDBJ. The propeptides were predicted by using SpiderP shown in Additional file 11: Fig. S11. Noteworthy, there are two different probable cleavage modes.

Phylogenetic study of the CRPs in H. venatoria

The precursor sequences of CRPs from *H. venatoria* venom gland were aligned using Clustal X 2.0. The resulting alignment was imported into MEGA5 software to construct the phylogenetic tree with the neighbor-joining method. The vast majority of the 6-cys ICK motif precursors (Family A, B, C, E, F and G) were defined as the relatively original clade. Only Family D and H, with shorter propeptides (12 aa), were placed outside the "older" clade. Remarkably, Family H, whose signal peptides are different from and longer than other families with 6-cys ICK motif, was located far away from the original clade precursors (Family A, B, C, E, F and G). Intriguingly, four 8-cys ICK-like motif families, L, M, N and K, were put in different clades. It is reasonable to put them in four families although the cys-scaffold of the mature peptide looks similar. Family Q and R, which both adopt a ten-Cys-scaffold '$C_IX_nC_{II}X_4C_{III}C_{IV}X_nC_VX_9C_{VI}X_nC_{VII}XC_{VIII}X_5C_{IX}X_nC_X$' in the mature peptide domains, were also arranged in two far away phylogenetic clades (Fig. 3). The 3 loops (C_{IV}-C_V, C_{VI}-C_{VII} and C_{IX}- C_X) in Family R are longer than those in Family Q.

Common types of spider venom CRPs from multifamily

In present study, the spider CRPs discovered with Sanger sequencing and Edman degradation sequencing methods were gathered from Arachnoserver spider toxins database [14]. There are more than 28 kinds scaffolds of the CRPs from 19 families, 55 species of spiders. The short CRPs with 6-cys are popular in Mygalomorphae, which present account for 82.5% and mainly include three kinds of scaffolds (ICK, disulfide-directed β-hairpin (DDH) and Kunitz). Relatively, the 6-cys peptides account for 22.2% in Araneomorphae which are mainly ICK-like motif peptides, a few of CHH and low molecular weight protein (LMWP) motifs. The long CRPs with more than 6 cysteines account for 17.5% in Mygalomorphae, which present six scaffolds with 8-cys and one with 10-cys. However, the long CRPs are much more popular in Araneomorphae, which present account for 77.8% and include diversified scaffolds (Table 2).

Evolution trend analyses for the propeptides of CRPs and cysteine number in mature peptide

The high-quality cDNA libraries and full-length EST sequences from eight spiders, 4 of Mygalomorphae and 4 of Araneomorphae, were used for the analysis of propeptide and cysteine number in mature domain. The length of propeptide varies among the species. There are longer propeptides in Mygalomorphae than in Araneomorphae. The propeptides longer than 25 aa account for 72.7%, 90%, 53.6% and 53.8% in *H.huwenum*, *H.hainanum, G. rosea* and *Chilobrachys Jingzhao*; respectively. By contrast, there are only 5%, 13.2% and 2.5% propeptides longer than 25 aa in *L. singoriensis, D.mizhoanus* and *A.ventricosus*, respectively, and none is longer than 25 aa in *H. venatoria*. By contrast, the percentages of precursors with a propeptide less than 10 residues are 4.6%, 5.7%, 17.8% and 29.2% in *H. huwenum, H. hainanum, G. rosea* and *C. Jingzhao*, respectively, and 47.7%, 43.4% and 92.5% in three species (*H. venatoria, D. mizhoanus* and *A. ventricosus*) of Araneomorphae. Although the ratio percentage of precursors with shortest propeptide is only 10.4% in *L. singoriensis*, the precursors with a 10–25 aa propeptide account for 84.7% (Fig. 4). As for the cysteine number in mature domain, there are 69.3%, 83.6%, 75%, 76.9% peptides with 6-cys motif in the four species of Mygalomorphae, respectively. However, there is no 6-cys CRPs described in *L. singoriensis* and *D. mizhoanus* so far, and 55% and 10% peptides with 6-cys motif in *H. venatoria* and *A. ventricosus*, respectively. On the contrary, there are more peptides with ≥ 8-cys motif, and the ratios percentages are 45%, 100%, 100% and 90% in *H. venatoria, L. singoriensis, D. mizhoanus* and *A. ventricosus*, respectively (Fig. 5).

Discussion

Molecular diversity of CRPs in the venom glands of H. venatoria

The construction of cDNA library with ESTs Sanger sequencing approach has been proved to be a rapid and reliable method for discovering new genes and obtaining data on the gene expression of CRPs in venom glands, which are characterized as multigenes displaying high similarity in part of their sequences [47]. In our group, second-generation sequencing technologies were applied to explore the diverse peptide toxins in venom of *H. huwenum* [48] and *Latrodectus tredecimguttatus* [49], the sequencing assembly of which strongly relied on the ESTs Sanger sequencing data. Furthermore, the data from Sanger sequencing method usually include the information of 5’ and 3’ untranslated regions which are very important for evolutionary analysis [50]. Also, the ESTs clones facilitate subsequent function research, for example, recombinant expression, transgene, gene modification, etc. Since the transcript of *H. venatoria* venom gland was investigated at the first time, the traditional Sanger sequencing technology was employed. In present study, 154 predicted CRP precursors in 24 families are discovered, among which, there are 8 families of short ICK toxins with 6-cys, 2 novel 6-cys non-ICK motifs, and 13 families of long CRP peptides with 8, 10 and 12-cys. Intriguingly, four novel cysteine scaffolds (Family I, O, W and X) are

first described in spider venom. The spider venom has the disulfide bond-rich inhibitor cystine knok (ICK) motif, which has stable molecular structure and strong tolerance to enzyme digestion [38, 39]. The Cysteine-rich secretory protein (CRISP) family belongs to the CAP protein superfamily, including massive single-stranded secretory proteins of different origins and functions [51]. These proteins play an important role in spider toxins on ion channels and inflammation [52–55]. For example, natrin, a member of the CRISPs family isolated from the venom glands of the Chinese cobra, can inhibit lipid peroxidation and induce apoptosis of tumor cells in rat liver [56–59]. Recently, *Cupiennius salei* CRISP1 and CRISP2 found the C-terminal extension leading to molecular masses twice as high, which identified it in comparable cysteine-rich venom proteins from other spiders, scorpions, and from the salivary glands of ticks, so it might be arachnid-specific [60]. The mature peptides of Family A-H contain three disulfide bonds with 1–4, 2–5, 3–6 connectivity, which provides these toxins with a high degree of chemical, thermal and biological stability [61]. The precursors of Family I have a high content of acidic amino acid and the feature has been noted for numerous spider-toxin propeptide sequences [62], which may be related to specific interactions between the toxin precursor and components of the secretory and/or protein folding pathway in spider venom glands. The diversity of primary structure within the *H. venatoria* venom family suggests that there have been few evolutionary restraints on CRPs diversification outside of the disulfide bridges that direct the 3D fold of these peptides. These findings highlight the extensive diversity of CRPs in *H. venatoria* venoms which can not only provide important data for evolutionary analysis of CRPs in spider venom, but also be exploited as novel therapeutic and biopesticide lead molecules.

Recruitment evens in spider venom

In present study, CHHs-like peptide genes are identified in Family J from the venom gland of *H. venatoria*, and that also widely distributed in very distantly related families: *Agelenidae*, *Sicariidae*, *Theridiidae*, *Sparassidae*, *Pisauridae* and *Nephilidae*. It is strongly suggested that CHHs-like peptides are derived from the superfamily of neuropeptides containing Crustacean Hyperglycemic Hormones (CHH) and independently recruited in spider venom glands [45]. The structure of U1-agatoxin-Ta1a from *Eratigena agrestis* was determined using heteronuclear NMR as a structural homolog of CHH family recently, which confirms the molecular evolutionary analyses indicating that CHHs-like toxins are highly derived members of the ITP/CHH family[63]. The hormone-derived venom peptides, named HAND toxins, are among the most stable peptides ever described, which provide a novel molecular scaffold for engineering drugs and insecticides [63].

Family O from the venom gland of *H. venatoria* with a novel cysteine scaffold in the mature region is also seemly evolved by recruitment of genes encoding normal body proteins followed by extensive duplication and neofunctionalization to play a role in killing and paralyzing prey or defending the organism. A secretory protein with the identical Cys-bone structure has been identified from black widow spider (gi: 318087504). However, it is noted to be involved in wrapping silk fibers. Moreover, several hypothetical non-secretory body proteins from *A. maculatum* also adopt the same eight-Cys scaffold (Fig. S5) without a predicted signal peptide. Therefore, two CRP motifs were discovered as recruitment evens in *H.*

venatoria venom. Another recruitment event, Kunitz-type toxins (KTTs), is characterised by a three stranded antiparallel β-structure as shown in HWTX-XI (κ-theraphotoxin-Hh1a) with the following disulphide bridge arrangement: C_I-C_{VI}, C_{II}-C_{IV}, C_{III}-C_V. This motif was first seen in the bovine pancreatic trypsin inhibitor (BPTI)-like proteinase inhibitor and well known from a diversity of secretory peptides such as toxins of snakes, cattle ticks, cone snails and sea anemones. According to the phylogenetic analyses, it was suggested that an earlier recruitment event for spider toxins than for snake toxins [64]. Recruitment has been considered as a common way in the evolution of toxin diversification[1].

Interspecies evolutionary trend of CRPs in spider venom

Spiders evolved over some 300 million years, and have become the most diverse terrestrial organism group only after insects. Spiders have evolved efficient weapons represented by the venom and/ or the silk for their hunting strategies. The venom system has evolved to restrain prey, defend and deter competitors. However, spiders investigated for venom molecular research are not more than 0.5% of all known species so far, and have focused on many big species and medically important species. With the evolution of the spider from Mygalomorphae to Araneomorphae the spider body size is evolving smaller and the predators have evolved to adopt webs as capture strategies. The species hunting without silk is considered more offensive, and its venoms often show higher complexity and potency [65]. Recent study suggested that peptides (2–15 kDa) in spider venoms are mainly responsible for the paralysis efficacy of the venom [66].

In present study, *H. venatoria* in the family Sparassidae, a relatively primitive species of Araneomorphae, does not rely on web and silk to capture prey that is similar in a way to the predation tactics of cursorial spiders *Lycosidae*, *Hexathelidae* and *Theraphosidae*. It was investigated for a global view of full-length CRPs toxins. The transcript data of *H. venatoria* venom gland along with before work about spider peptide toxins in our group and high-quality cDNA library data in public database provide the opportunity to make a holistic view analysis of the evolution of spider venom CRPs. In Mygalomorphae, the short CRPs with 6-cys are more than 69% of all discovered CRPs in the four species, and mainly include three kinds of scaffolds (ICK, DDH and Kunitz). By contrast, in Araneomorphae, there is no 6-cys CRPs in *L. singoriensis* and *D. mizhoanus*, both are in family Lycosidae which is one most evolutionary at distant end in the spider system. The 6-cys CRPs are 10% and 55% in *A. ventricosus* and *H. venatoria* respectively. *A. ventricosus* is closer to Mygalomorphae than *H. venatoria*, but mainly uses web for predation, which may explain the low percent of 6-cys CRPs in this species. Intriguingly, the 6-cys CRPs in *H. venatoria* have a much shorter propeptide comparing to that in Mygalomorphae. The propeptide that was ever thought as a necessary part of a spider CRPs precursor is short, even disappears in several families of CRPs in Araneomorphae venom. The present study suggests two points of evolutionary trend of CRPs in spider venom from Mygalomorphae to Araneomorphae: the mature peptides have been developed longer with more cysteines; and the propeptides have diminished and even vanished.

Conclusion

The spider *H. venatoria* in the family of Sparassidae is located at the earlier phylogenetic position of Araneomorphae, and as a potential resource for biological control of pest insects. The EST sequencing of cDNA library results revealed the complex toxin repertoire of *H. venatoria* venom with the discovery of 154 predicted CRP precursors in 24 families. Remarkably, four motifs are described for the first time in spider venom. Under considering the spider's characters of morphology, predatory behavior and taxonomic position, the CRPs from 8 families in Araneae were retrieved and analyzed to find out the evolutionary trend rules, which will improve our understanding of the evolution of Araneae venoms and their CRPs toxins. The data of 154 CRPs from *H. venatoria* likely facilitate exploration of the potential sources of agrochemical lead compounds.

Abbreviations

CRPs Cysteine-rich peptides

PQM Processing Quadruplet Motif

ICK Inhibitory Cystine Knot

CHHs Crustacean hyperglycemic hormones

MIH Molt-inhibiting hormone

DDH Disulfide-directed β-hairpin

LMWP Low molecular weight protein

CRISP Cysteine-rich secretory protein

Declarations

Ethics approval and consent to participate

All the experimental animal procedures followed the principles of the Guide for Care and Use of Laboratory Animals and were approved by the Hunan Agricultural University.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from the Hunan Province Department of Education Foundation(16A098); Hunan Provincial Science and Technology Department (2017NK2311); National Natural Science Foundation of China (30900242 and 31972787).

Author Contributions

YD and JC conceived and designed the study; YD constructed the cDNA library with assistance from XZ; KC analyzed the data with assistance from TX; YD wrote the paper with assistance from JC; all authors read and approved the final manuscript.

Acknowledgments

Not applicable.

References

1. Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JD, King GF, et al. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. Annu Rev Genomics Hum Genet. 2009;10:483–511.
2. Gracy J, Chiche L. Structure and modeling of knottins, a promising molecular scaffold for drug discovery. Curr Pharm Des. 2011;17:4337–50.
3. Berkut AA, Peigneur S, Myshkin MY, Paramonov AS, Lyukmanova EN, Arseniev AS, et al. Structure of membrane-active toxin from crab spider Heriaeus melloteei suggests parallel evolution of sodium channel gating modifiers in Araneomorphae and Mygalomorphae. J Biol Chem. 2015;290:492–504.
4. King GF, Hardy MC. Spider-venom peptides: structure, pharmacology, and potential for control of insect pests. Annu Rev Entomol. 2013;58:475–96.
5. Zhu S, Peigneur S, Gao B, Luo L, Jin D, Zhao Y, et al. Molecular diversity and functional evolution of scorpion potassium channel toxins. Mol Cell Proteomics. 2011;10::M110 002832.
6. Garb JE, Hayashi CY. Molecular evolution of alpha-latrotoxin, the exceptionally potent vertebrate neurotoxin in black widow spider venom. Mol Biol Evol. 2013;30:999–1014.
7. He Y, Zhao R, Di Z, Li Z, Xu X, Hong W, et al. Molecular diversity of Chaerilidae venom peptides reveals the dynamic evolution of scorpion venom components from Buthidae to non-Buthidae. J Proteomics. 2013;89:1–14.
8. Junqueira-de-Azevedo IL, Bastos CM, Ho PL, Luna MS, Yamanouye N, Casewell NR. Venom-related transcripts from Bothrops jararaca tissues provide novel molecular insights into the production and evolution of snake venom. Mol Biol Evol. 2015;32:754–66.

9. Puillandre N, Bouchet P, Duda TF Jr, Kauferstein S, Kohn AJ, Olivera BM, et al. Molecular phylogeny and evolution of the cone snails (Gastropoda, Conoidea). Mol Phylogenet Evol. 2014;78:290–303.
10. Vassilevski AA, Kozlov SA, Grishin EV. Molecular diversity of spider venom. Biochemistry. 2009;74:1505–34.
11. Lucia Kuhn-Nentwig RScaWN. Venom Composition and Strategies in Spiders: Is Everything Possible?; 2011.
12. Edwards GB. Huntsman Spider, Heteropoda venatoria (Linnaeus) (Arachnida: Araneae: Sparassidae). University of Florida IFAS; 2009.
13. Huang Y, Wu X, Zhang P, Duan Z, Zhou X, Chen M, et al. Peptide-rich venom from the spider Heteropoda venatoria potently inhibits insect voltage-gated sodium channels. Toxicon. 2017;125:44–9.
14. Herzig V, Wood DL, Newell F, Chaumeil PA, Kaas Q, Binford GJ, et al. ArachnoServer 2.0, an updated online resource for spider toxin sequences and structures. Nucleic Acids Res. 2011;39::D653–7.
15. Pineda SS, Chaumeil PA, Kunert A, Kaas Q, Thang MWC, Le L, et al. ArachnoServer 3.0: an online resource for automated discovery, analysis and annotation of spider toxins. Bioinformatics. 2018;34:1074–6.
16. Chen J, Zhao L, Jiang L, Meng E, Zhang Y, Xiong X, et al. Transcriptome analysis revealed novel possible venom components and cellular processes of the tarantula Chilobrachys jingzhao venom gland. Toxicon. 2008;52:794–806.
17. Binford GJ, Bodner MR, Cordes MH, Baldwin KL, Rynerson MR, Burns SN, et al. Molecular evolution, functional variation, and proposed nomenclature of the gene family that includes sphingomyelinase D in sicariid spider venoms. Mol Biol Evol. 2009;26:547–66.
18. Chou HH, Holmes MH. DNA sequence quality trimming and vector removal. Bioinformatics. 2001;17:1093–104.
19. Sorek R, Safer HM. A novel algorithm for computational identification of contaminated EST libraries. Nucleic Acids Res. 2003;31:1067–74.
20. Nishikawa T, Ota T, Kawai Y, Ishii S, Saito K, Yamamoto J, et al. Database and analysis system for cDNA clones obtained from full-length enriched cDNA libraries. In Silico Biol. 2002;2:5–18.
21. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785–6.
22. Wong ES, Hardy MC, Wood D, Bailey T, King GF. SVM-based prediction of propeptide cleavage sites in spider toxins identifies toxin innovation in an Australian tarantula. PLoS One. 2013;8:e66279.
23. Gracy J, Le-Nguyen D, Gelly JC, Kaas Q, Heitz A, Chiche L. KNOTTIN: the knottin or inhibitor cystine knot scaffold in 2007. Nucleic Acids Res. 2008;36::D314–9.
24. Postic G, Gracy J, Perin C, Chiche L, Gelly JC. KNOTTIN: the database of inhibitor cystine knot scaffold after 10 years, toward a systematic structure modeling. Nucleic Acids Res. 2018;46::D454–8.

25. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23:2947–8.
26. Kumar S, Nei M, Dudley J, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform. 2008;9:299–306.
27. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28:2731–9.
28. Sugumaran MP, Kumar MG, Ramasamy K, Vincent S. Observations on life cycle of certain spiders from Western Ghats of Tamil Nadu. J Environ Biol. 2004;25:325–8.
29. Jiang L, Chen J, Peng L, Zhang Y, Xiong X, Liang S. Genomic organization and cloning of novel genes encoding toxin-like peptides of three superfamilies from the spider Orinithoctonus huwena. Peptides. 2008;29:1679–84.
30. Tang X, Zhang Y, Hu W, Xu D, Tao H, Yang X, et al. Molecular diversification of peptide toxins from the tarantula Haplopelma hainanum (Ornithoctonus hainana) venom based on transcriptomic, peptidomic, and genomic analyses. J Proteome Res. 2010;9:2550–64.
31. Zhang Y, Chen J, Tang X, Wang F, Jiang L, Xiong X, et al. Transcriptome analysis of the venom glands of the Chinese wolf spider Lycosa singoriensis. Zoology (Jena). 2010;113:10–8.
32. Kimura T, Ono S, Kubo T. Molecular Cloning and Sequence Analysis of the cDNAs Encoding Toxin-Like Peptides from the Venom Glands of Tarantula Grammostola rosea. Int J Pept. 2012;2012:731293.
33. Xu X, Wang H, Zhang F, Hu Z, Liang S, Liu Z. A Comparative Analysis of the Venom Gland Transcriptomes of the Fishing Spiders Dolomedes mizhoanus and Dolomedes sulfurous. PLoS One. 2015;10:e0139908.
34. Jiang L, Liu C, Duan Z, Deng M, Tang X, Liang S. Transcriptome analysis of venom glands from a single fishing spider Dolomedes mizhoanus. Toxicon. 2013;73:23–32.
35. Duan Z, Cao R, Jiang L, Liang S. A combined de novo protein sequencing and cDNA library approach to the venomic analysis of Chinese spider Araneus ventricosus. J Proteomics. 2013;78:416–27.
36. Escoubas P. Molecular diversification in spider venoms: a web of combinatorial peptide libraries. Mol Divers. 2006;10:545–54.
37. Kozlov S, Malyavka A, McCutchen B, Lu A, Schepers E, Herrmann R, et al. A novel strategy for the identification of toxinlike structures in spider venom. Proteins. 2005;59:131–40.
38. Norton RS, Pallaghy PK. The cystine knot structure of ion channel toxins and related polypeptides. Toxicon. 1998;36:1573–83.
39. Craik DJ, Daly NL, Waine C. The cystine knot motif in toxins and implications for drug design. Toxicon. 2001;39:43–60.
40. Davey ML, Hall MR, Willis RH, Oliver RW, Thurn MJ, Wilson KJ. Five Crustacean Hyperglycemic Family Hormones of Penaeus monodon: Complementary DNA Sequence and Identification in Single

Sinus Glands by Electrospray Ionization-Fourier Transform Mass Spectrometry. Mar Biotechnol (NY). 2000;2:80–91.

41. Aguilar MB, Falchetto R, Shabanowitz J, Hunt DF, Huberman A. Complete primary structure of the molt-inhibiting hormone (MIH) of the Mexican crayfish Procambarus bouvieri (Ortmann). Peptides. 1996;17:367–74.
42. Katayama H, Nagata K, Ohira T, Yumoto F, Tanokura M, Nagasawa H. The solution structure of molt-inhibiting hormone from the Kuruma prawn Marsupenaeus japonicus. J Biol Chem. 2003;278:9620–3.
43. Gremski LH, da Silveira RB, Chaim OM, Probst CM, Ferrer VP, Nowatzki J, et al. A novel expression profile of the Loxosceles intermedia spider venomous gland revealed by transcriptome analysis. Mol Biosyst. 2010;6:2403–16.
44. Gasparini S, Kiyatkin N, Drevet P, Boulain JC, Tacnet F, Ripoche P, et al. The low molecular weight protein which co-purifies with alpha-latrotoxin is structurally related to crustacean hyperglycemic hormones. J Biol Chem. 1994;269:19803–9.
45. McCowan C, Garb JE. Recruitment and diversification of an ecdysozoan family of neuropeptide hormones for black widow spider venom expression. Gene. 2014;536:366–75.
46. Santos AD, Imperial JS, Chaudhary T, Beavis RC, Chait BT, Hunsperger JP, et al. Heterodimeric structure of the spider toxin omega-agatoxin IA revealed by precursor analysis and mass spectrometry. J Biol Chem. 1992;267:20701–5.
47. Kaas Q, Craik DJ. Bioinformatics-Aided Venomics. Toxins (Basel). 2015;7:2159–87.
48. Zhang Y, Huang Y, He Q, Liu J, Luo J, Zhu L, et al. Toxin diversity revealed by a transcriptomic study of Ornithoctonus huwena. PLoS One. 2014;9:e100682.
49. He Q, Duan Z, Yu Y, Liu Z, Liang S. The venom gland transcriptome of Latrodectus tredecimguttatus revealed by deep sequencing and cDNA library analysis. PLoS One. 2013;8:e81357.
50. Ikeda N, Chijiwa T, Matsubara K, Oda-Ueda N, Hattori S, Matsuda Y, et al. Unique structural characteristics and evolution of a cluster of venom phospholipase A2 isozyme genes of Protobothrops flavoviridis snake. Gene. 2010;461:15–25.
51. Yamazaki Y, Morita T. Structure and function of snake venom cysteine-rich secretory proteins. Toxicon. 2004;44:227–31.
52. Udby L, Bjartell A, Malm J, Egesten A, Lundwall A, Cowland JB, et al. Characterization and localization of cysteine-rich secretory protein 3 (CRISP-3) in the human male reproductive tract. J Androl. 2005;26:333–42.
53. Guo M, Teng M, Niu L, Liu Q, Huang Q, Hao Q. Crystal structure of the cysteine-rich secretory protein stecrisp reveals that the cysteine-rich domain has a K + channel inhibitor-like fold. J Biol Chem. 2005;280:12405–12.
54. Jamsai D, Bianco DM, Smith SJ, Merriner DJ, Ly-Huynh JD, Herlihy A, et al. Characterization of gametogenetin 1 (GGN1) and its potential role in male fertility through the interaction with the ion

channel regulator, cysteine-rich secretory protein 2 (CRISP2) in the sperm tail. Reproduction. 2008;135:751–9.

55. Pathak BR, Breed AA, Apte S, Acharya K, Mahale SD. Cysteine-rich secretory protein 3 plays a role in prostate cancer cell invasion and affects expression of PSA and ANXA1. Mol Cell Biochem. 2016;411:11–21.
56. Wang F, Li H, Liu MN, Song H, Han HM, Wang QL, et al. Structural and functional analysis of natrin, a venom protein that targets various ion channels. Biochem Biophys Res Commun. 2006;351:443–8.
57. Wang J, Shen B, Guo M, Lou X, Duan Y, Cheng XP, et al. Blocking effect and crystal structure of natrin toxin, a cysteine-rich secretory protein from Naja atra venom that targets the BKCa channel. Biochemistry. 2005;44:10145–52.
58. Wang YL, Kuo JH, Lee SC, Liu JS, Hsieh YC, Shih YT, et al. Cobra CRISP functions as an inflammatory modulator via a novel Zn2+- and heparan sulfate-dependent transcriptional regulation of endothelial cell adhesion molecules. J Biol Chem. 2010;285:37872–83.
59. Lu S, Lu R, Song H, Wu J, Liu X, Zhou X, et al. Metabolomic study of natrin-induced apoptosis in SMMC-7721 hepatocellular carcinoma cells by ultra-performance liquid chromatography-quadrupole/time-of-flight mass spectrometry. Int J Biol Macromol. 2019;124:1264–73.
60. Kuhn-Nentwig L, Langenegger N, Heller M, Koua D, Nentwig W. The Dual Prey-Inactivation Strategy of Spiders-In-Depth Venomic Analysis of Cupiennius salei. Toxins (Basel). 2019;11.
61. Saez NJ, Senff S, Jensen JE, Er SY, Herzig V, Rash LD, et al. Spider-venom peptides as therapeutics. Toxins. 2010;2:2851–71.
62. Liang S. An overview of peptide toxins from the venom of the Chinese bird spider Selenocosmia huwena Wang [= Ornithoctonus huwena (Wang)]. Toxicon. 2004;43:575–85.
63. Undheim EA, Grimm LL, Low CF, Morgenstern D, Herzig V, Zobel-Thropp P, et al. Weaponization of a Hormone: Convergent Recruitment of Hyperglycemic Hormone into the Venom of Arthropod Predators. Structure. 2015;23:1283–92.
64. Yuan CH, He QY, Peng K, Diao JB, Jiang LP, Tang X, et al. Discovery of a distinct superfamily of Kunitz-type toxin (KTT) from tarantulas. PLoS One. 2008;3:e3414.
65. Michálek O, Řezáč M, Líznarová E, Symondson WOC, Pekár S. Silk versus venom: alternative capture strategies employed by closely related myrmecophagous spiders. Biol J Lin Soc. 2018;126:545–54.
66. Pekar S, Bocanek O, Michalek O, Petrakova L, Haddad CR, Sedo O, et al. Venom gland size and venom complexity-essential trophic adaptations of venomous predators: A case study using spiders. Mol Ecol. 2018;27:4257–69.
67. Kuhn-Nentwig L, Largiader CR, Streitberger K, Chandru S, Baumann T, Kampfer U, et al. Purification, cDNA structure and biological significance of a single insulin-like growth factor-binding domain protein (SIBD-1) identified in the hemocytes of the spider Cupiennius salei. Insect Biochem Mol Biol. 2011;41:891–901.

Tables

Table 1 Sequence diversity of the 6-cys ICK motif toxins in *H. venatoria.*

Family	Signal peptide mode	Length of propeptides/ PQM	Mature peptide mode	C-terminal propeptides
A	$MKh_{12}S\ h_5$	19/DEQR	$XCX_6CX_5CCX_4CX_3CX_{4\text{-}6}$	G(K)
B	$MKT\ h_{11}S\ h_5$	15/AVAR	$X_2CX_6CX_5CCX_4CX_3CX_3$	G
C	MKh_{18}	15/VAAR	$X_5CX_6CX_6CCX_3CX_3CX_5$	GK(K)
D	$MKIT\ h_{15}$	11/VQAR	$XCX_6CX_6CCX_4CX_4CX_6$	GK
E	$MKTTh_3Th_6Sh_5$	15-17/ATGR	$X_3CX_6CX_5CCX_4CX_5CX$	W/O
F	$MKTTh_3Th_6Sh_5$	15/VTGR	$X_{2\text{-}4}CX_6CX_5CCX_4CX_5CX_3$	$RKX_{4\text{-}5}$
G	$MKh_5Th_6Sh_5$	15/ITVR	$X_{2\text{-}3}CX_6CX_5CCX_4CX_9CX_4$	G
H	$MKSSh_7Th_4Sh_2EFTRS$	12/VQER	$X_2CX_6CX_4CCX_4CX_8CX_3$	W/O

In the signal peptide mode, ‘h’ indicates hydrophobic residue, and the Arabic numerals denote the number of residues. Capital letters indicate the corresponding amino acids. ‘C’ is cystine, ‘X’ is any residue but cystine, and ‘W/O’ denotes no putative C-terminal propeptides.

Table 2 Characteristics and distribution of common motifs of CRPs in Mygalomorphae and Araneomorphae.

	Mygalomorphae	Araneomorphae
Total number of 6-Cys CRPs /Percentage	597/82.5%	140/22.2%
Total number of > 6-Cys CRPs /Percentage	127/17.5%	520/77.8%
Main motifs = 6-Cys/ number	ICK-like/476 DDH/99 Kunitz/22	ICK-like/118 CHH/20 LMWP/2
Main motifs > 6-Cys/ number	C…C…CXCC…C…C…C/36 C…C…CCC…C…C…C/6 C…C…C…CC…C…C…C/2 C…C…CC…CC…C…C/9 C…C…CC…C…C…C…C/15 C…C…CC…CXC…CXC/23 C…C…CC…C…C…CXC… C…C/35 CRISP >14C/1	C…CXC…CXC…CXC…C/5 C…CXC…CXC…CXC…C…C…C/12 C…CXC…CXC…C…C…C/1 C…C…CC…C…CC…C…C…C/74 C…C…CC…C…C…CXC/1 C…C…CC…CXC…CXC/254 C…C…CC…C…C…CXC…C…C/38 C…CXC…CC…CXC…CXC…C/1 C…C…CXCC…CXC…CXC…C/34 C…C…CXCC…CXC…CXC…C…C…C/18 C…C…CXCC…C…C…CXC…C…C…C/7 C…C…CXCC…C…C…C…C…C/12 C…C…C…C…C…C…CXC…C…C/2 C…C…C…CXC…CXC…C…C…CXC…C/24 C…C…CXCCXC…CXC…CXC…C…C/8 C…C…CC…CXC…C…C…CXC…CXC/7 C…C…CXCC…CXC…CXC…C…C…C…C…C/7 C…C…C…C…C…C…C…C… C…C…C…C/3 CRISP >14C 14

Capital letters indicate the corresponding amino acids.‘C’ is cystine, and ‘X’ is any residue but cystine.

Additional File

Additional file 1: Fig. S1 Homology alignment of family A-H ICKs with 6-cys motif. The putative signal peptides and propeptides are indicated by dark box. C-terminally, prosequences are indicated by gray box. Conserved cysteine residues are indicated by filled with black.

Additional file 2: Fig. S2 Multiple sequence alignment of Family I precursors and amino acid sequences from *H. venatoria*. The putative signal peptides and propeptides are boxed and cysteines of mature peptide are in gray shadow. Gaps (dashes) were introduced to maximize the deduced polypeptide sequence similarities. Rectangles denote C-terminal amidation.

Additional file 3: Fig. S3 Multiple sequence alignment of Family J precursors and amino acid sequences from *H. venatoria*. The putative signal peptides and propeptides are boxed and cysteines of mature peptide are in gray shadow. Gaps (dashes) were introduced to maximize the deduced polypeptide sequence similarities. Rectangles denote C-terminal amidation.

Additional file 4: Fig. S4 Multiple sequence alignment of Family K-N precursors and amino acid sequences from *H. venatoria*. The putative signal peptides and propeptides are boxed and cysteines of mature peptide are in gray shadow. Gaps (dashes) were introduced to maximize the deduced polypeptide sequence similarities. Rectangles denote C-terminal amidation.

Additional file 5: Fig. S5 Multiple sequence alignment of Family O precursors and amino acid sequences from *H. venatoria*. The putative signal peptides and propeptides are boxed and cysteines of mature peptide are in gray shadow. Gaps (dashes) were introduced to maximize the deduced polypeptide sequence similarities. Rectangles denote C-terminal amidation.

Additional file 6: Fig. S6 Multiple sequence alignment of Family P precursors and amino acid sequences from *H. venatoria*. The putative signal peptides and propeptides are boxed and cysteines of mature peptide are in gray shadow. Gaps (dashes) were introduced to maximize the deduced polypeptide sequence similarities. Rectangles denote C-terminal amidation.

Additional file 7: Fig. S7 Multiple sequence alignment of Family Q and R precursors and amino acid sequences from *H. venatoria*. The putative signal peptides and propeptides are boxed and cysteines of mature peptide are in gray shadow. Gaps (dashes) were introduced to maximize the deduced polypeptide sequence similarities. Rectangles denote C-terminal amidation.

Additional file 8: Fig. S8 Multiple sequence alignment of Family S-U precursors and amino acid sequences from *H. venatoria*. The putative signal peptides and propeptides are boxed and cysteines of mature peptide are in gray shadow. Gaps (dashes) were introduced to maximize the deduced polypeptide sequence similarities. Rectangles denote C-terminal amidation.

Additional file 9: Fig. S9 Multiple sequence alignment of Family V precursors and amino acid sequences from *H. venatoria*. The putative signal peptides and propeptides are boxed and cysteines of mature peptide are in gray shadow. Gaps (dashes) were introduced to maximize the deduced polypeptide sequence similarities. Rectangles denote C-terminal amidation.

Additional file 10: Fig. S10 Multiple sequence alignment of Family W precursors and amino acid sequences from *H. venatoria*. The putative signal peptides are boxed and cysteines of mature peptide are in gray shadow. Gaps (dashes) were introduced to maximize the deduced polypeptide sequence similarities. Rectangles denote C-terminal amidation.

Additional file 11: Fig. S11 Multiple sequence alignment of Family X precursors and amino acid sequences from *H. venatoria*. The putative signal peptides and propeptides are boxed and cysteines of mature peptide are in gray shadow. Rectangles denote C-terminal amidation.

Figures

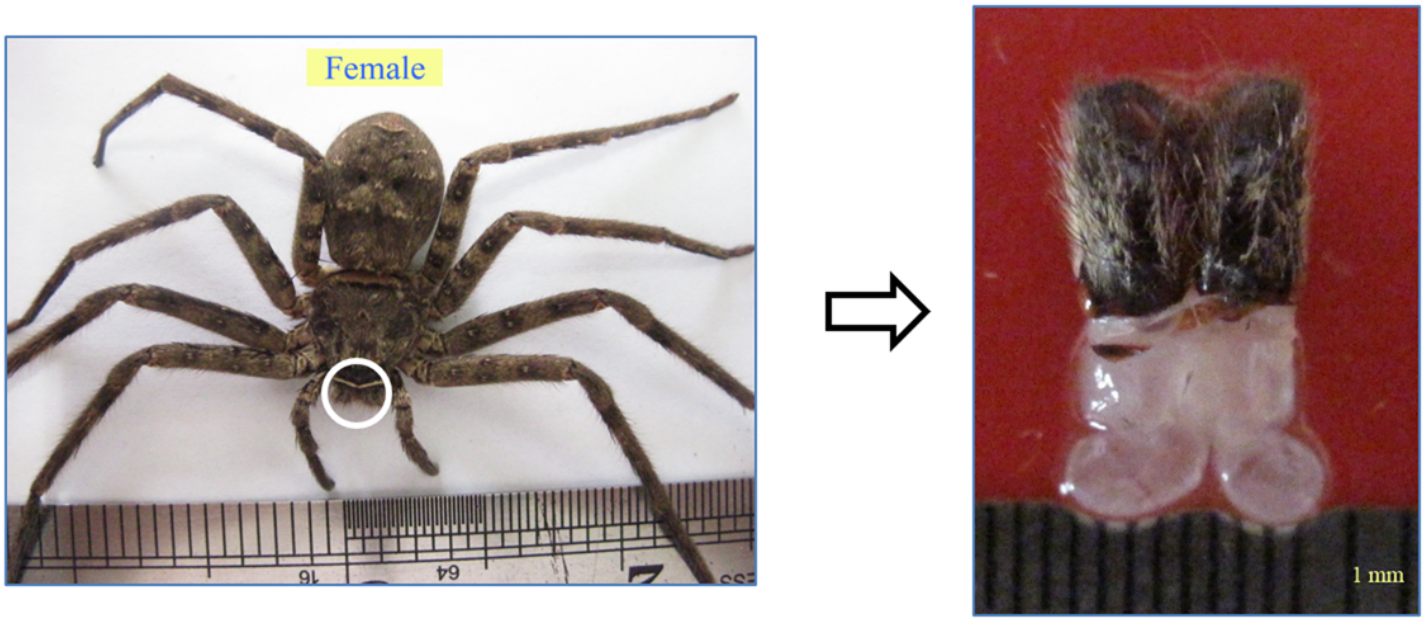

Figure 1

The morphological structure of H. venatoria venom glands.

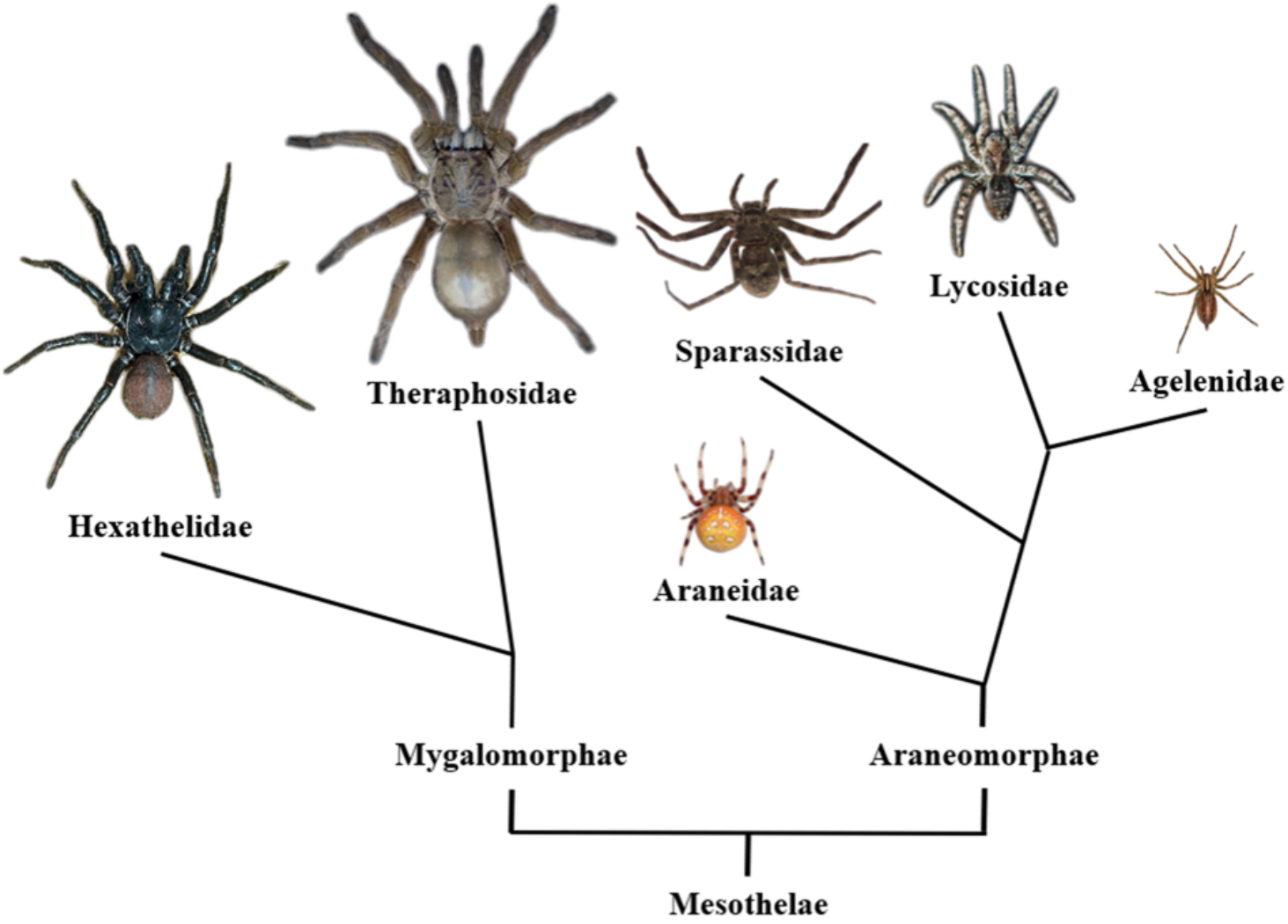

Figure 3

Simplified phylogenetic tree of spider familieswith the physique in relative size of the representative species of each family. Modified from Kuhn-Nentwig et al. (2011) [67]. Hadronyche versuta in Hexathelidae, Chilobrachys Jingzhao in Theraphosidae, Araneus trifolium in Araneidae, Heteropoda venatoria in Sparassidae, Lycosa singoriensis in Lycosidae and Agelenopsis potteri in Agelenidae are mentioned in the present work.

Figure 5

Evolutionary relationships of CRPs from H. venatoria venom glands. Evolutionary analyses were conducted using the Neighbor-Joining method in MEGA5.

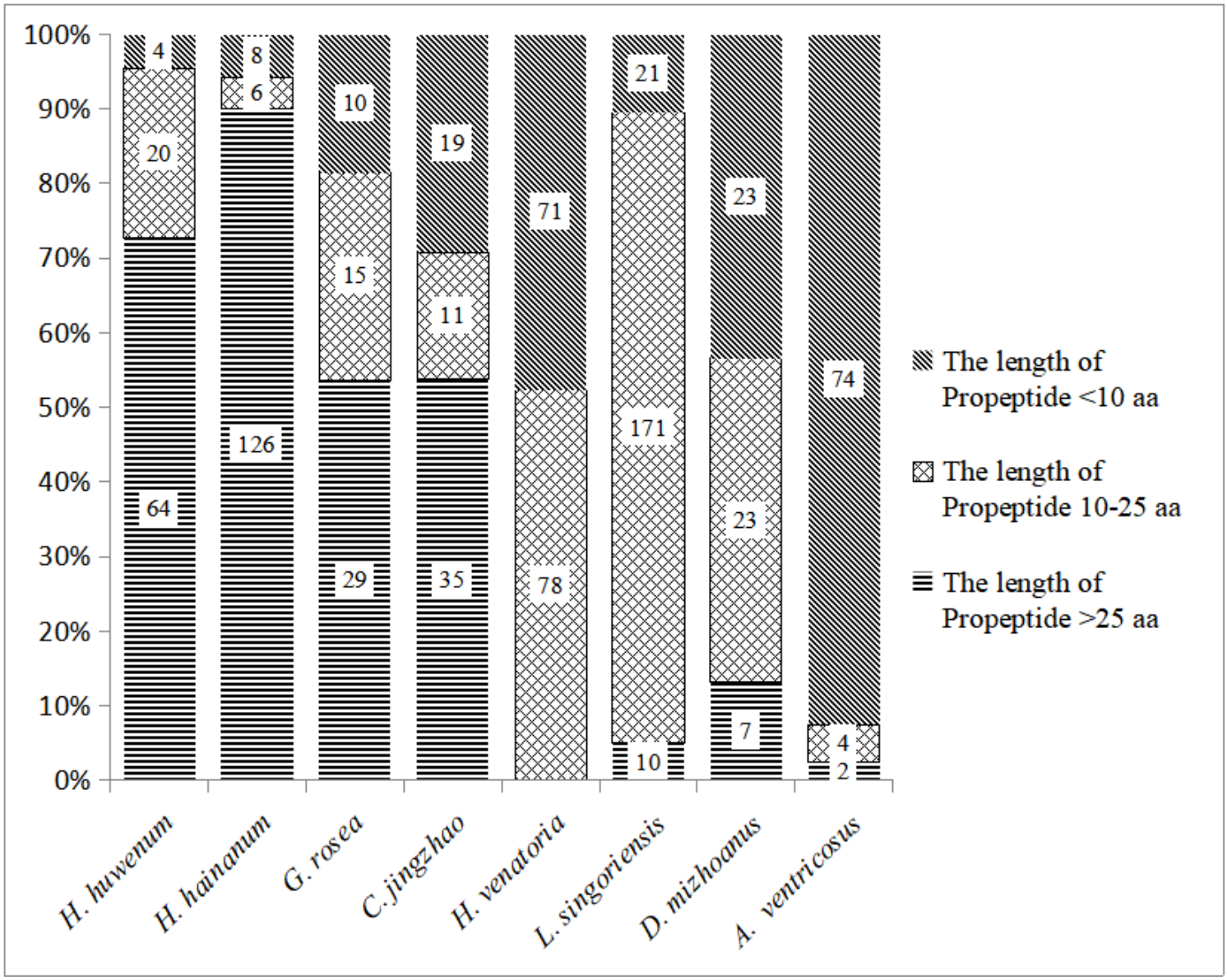

Figure 7

The distribution of the length of propeptide in eight spider species. The species of Mygalomorphae are H. huwenum, H. hainanum, G. rosea and C. Jingzhao, and the species of Araneomorphae are L. singoriensis, D. mizhoanus, A. ventricosus and H. venatoria.

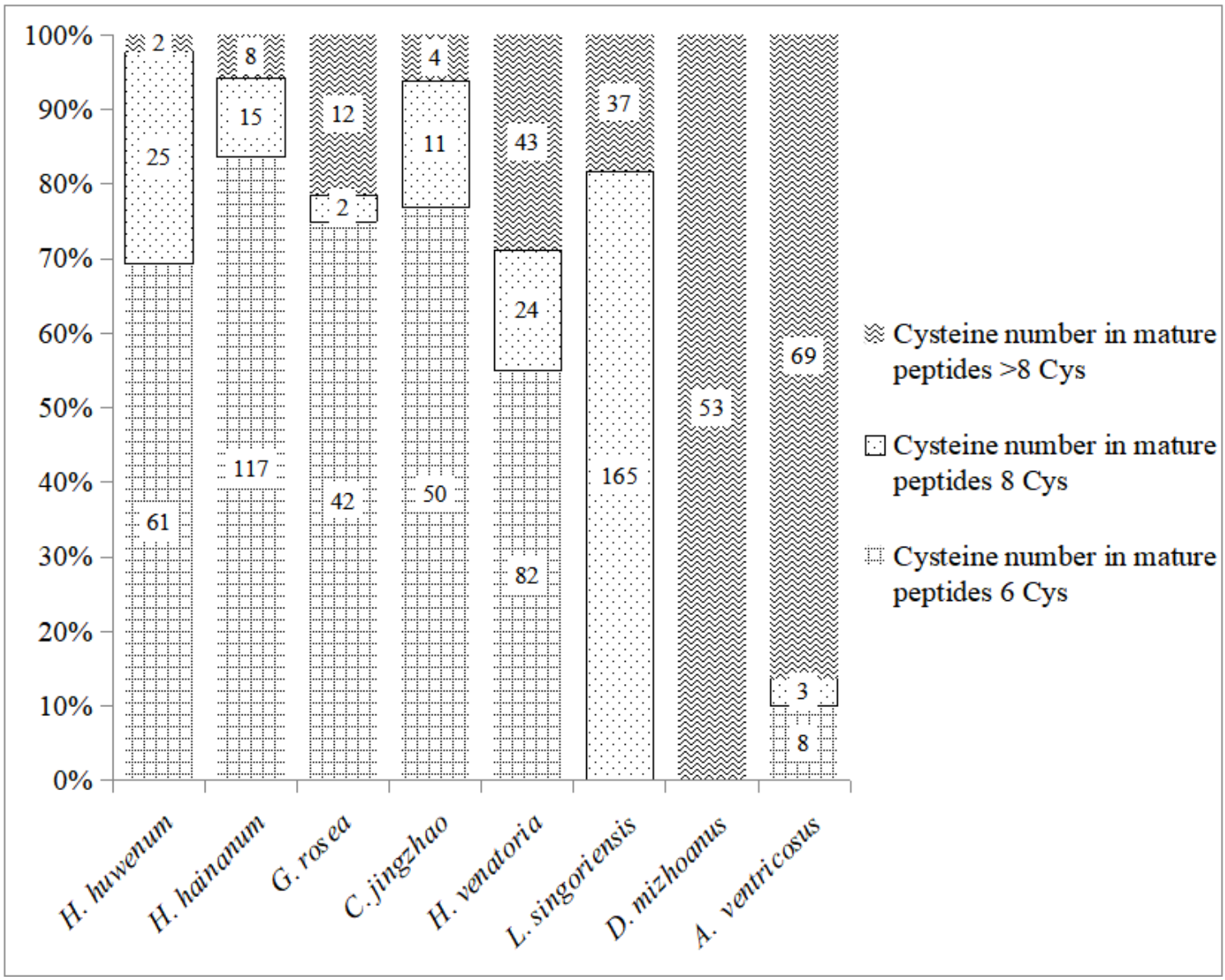

Figure 9

The distribution of cysteine number in mature peptide in eight spider species. The species of Mygalomorphae are H. huwenum, H. hainanum, G. rosea and C. Jingzhao, and the species of Araneomorphae are L. singoriensis, D. mizhoanus, A. ventricosus and H. venatoria.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Fig.S2.tif
- Fig.S9.tif
- Fig.S9.tif
- Fig.S10.tif

- Fig.S11.tif
- Fig.S5.tif
- Fig.S6.tif
- Fig.S7.tif
- Fig.S8.tif
- Fig.S2.tif
- Fig.S3.tif
- Fig.S4.tif
- Fig.S1.tif
- Fig.S6.tif
- Fig.S7.tif
- Fig.S8.tif
- Fig.S10.tif
- Fig.S3.tif
- Fig.S11.tif
- Fig.S3.tif
- Fig.S4.tif
- Fig.S5.tif
- Fig.S8.tif
- Fig.S9.tif
- Fig.S10.tif
- Fig.S11.tif
- Fig.S1.tif
- Fig.S2.tif
- Fig.S4.tif
- Fig.S5.tif
- Fig.S6.tif
- Fig.S7.tif
- Fig.S1.tif