MBE Advance Access published May 20, 2014

For consideration as an Article (Discoveries)

Clawing through evolution: toxin diversification and convergence in the ancient lineage Chilopoda (Centipedes)

Eivind A.B. Undheim[1,2], Alun Jones[1], Karl R. Clauser[3], John W. Holland[1], Sandy S. Pineda[1], Glenn F. King[1]*, Bryan G. Fry[1,2]*

[1]Division of Chemistry and Structural Biology, Institute for Molecular Bioscience, The University of Queensland, St. Lucia, QLD 4072, Australia

[2]Venom Evolution Laboratory, School of Biological Sciences, The University of Queensland, St. Lucia, QLD 4072, Australia

[3]The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

*Corresponding authors: Bryan G. Fry (email: bgfry@uq.edu.au) or Glenn F King (email: glenn.king@imb.uq.edu.au).

Abstract Despite the staggering diversity of venomous animals, there seems to be remarkable convergence in regards to the types of proteins used as toxin scaffolds. However, our understanding of this fascinating area of evolution has been hampered by the narrow taxonomical range studied, with entire groups of venomous animals remaining almost completely unstudied. One such group is centipedes, class Chilopoda, which emerged about 440 mya and may represent the oldest terrestrial venomous lineage next to scorpions. Here we provide the first comprehensive insight into the chilopod "venome" and its evolution, which has revealed novel and convergent toxin recruitments as well as entirely new toxin families among both high and low molecular weight venom components. The ancient evolutionary history of centipedes is also apparent from the differences between the Scolopendromorpha and Scutigeromorpha venoms, which diverged over 430 million years ago, and appear to employ substantially different venom strategies. The presence of a wide range of novel proteins and peptides in centipede venoms highlights these animals as a rich source of novel bioactive molecules. Understanding the evolutionary processes behind these ancient venom systems will not only broaden our understanding of which traits make proteins and peptides amenable to neofunctionalisation but it may also aid in directing bioprospecting efforts.

Introduction

Venoms are key evolutionary innovations in a broad range of animals, with a wide range of animal phyla containing at least one venomous representative (Fry et al. 2009a). This broad phylogenetic spectrum of venomous animals reflects the ecological advantages of venom, which enables the use of biochemical rather than physical means of defense, competitor deterrence, and predation (Casewell et al. 2013; King 2011; Morgenstern and King 2013). Venoms are complex mixtures of inorganic salts, low molecular weight compounds, peptides and proteins, and they can be extremely diverse, with single venoms often containing hundreds of unique toxins (Escoubas et al. 2006; Nascimento, et al. 2006). Despite the incredible diversity of toxins, there is substantial convergence in protein and peptide types recruited into venoms, with most toxins characterized to date belonging to only a handful of protein families (e.g. Fry et al. 2009a; Fry et al. 2009b;

Ruder et al. 2013). The evolutionary trajectories of these convergently recruited toxins provide clues to what makes peptides and proteins amenable for use as toxins. This not only aids in understanding novel toxin families but also provides an understanding of the mechanisms of protein and peptide neofunctionalisation, which can be translated into therapeutic and agrochemical research fields (King 2011; King and Hardy 2013).

However, a major limitation of venom research has been the narrow taxonomical range studied, with entire groups of venomous animals remaining virtually completely unstudied. One such group is centipedes (class Chilopoda), which are one of the four main lineages of myriapods, and at over 440 million years may represent the oldest living terrestrial venomous lineage (Undheim and King 2011). There are about 3300 species of centipede distributed across five extant orders: Scutigeromorpha, Lithobiomorpha, Craterostigmomorpha, Geophilomorpha, and Scolopendromorpha. These are contained within two subclasses, the Notostigmophora (Scutigeromorpha) and Pleurostigmophora (the remaining four), which probably split soon after the emergence of Chilopoda (Murienne et al. 2010). Of the extant orders, the two that are most familiar are also in many ways the most divergent orders of centipede, namely the "house centipedes" (Scutigeromorpha) and the "classic" centipedes (Scolopendromorpha). These orders diverged 430 mya, which probably represents the most ancient divergence within any terrestrial venomous lineage.

With very few exceptions, all centipedes are predators. They prey mainly on arthropods and other invertebrates by subduing them with venom secreted from and injected via a modified first pair of legs called forcipules (Bonato et al. 2010). These are equipped with venom glands, which can essentially be regarded as invaginations of the cuticle containing a high number of weaponized epidermal sub-glands, each with its own extracellular storage space (Dugon and Arthur 2012; Hilken et al. 2005; Lewis 1981; Rosenberg and Hilken 2006). Interestingly, the shape of the forcipules varies substantially between higher-level taxa, particularly Scutigeromorpha and Scolopendromorpha. While Scolopendromorpha are equipped with heavily sclerotized, muscular appendages made for cutting and slicing prey, the forcipules of Scutigeromorpha are slender, delicate, and lack a cutting ridge (Dugon et al. 2012). Despite their difference in ability to overpower prey by physical means, members of both

orders are nevertheless ferocious predators, suggesting there may be compensatory adaptations in the venom of Scutigeromorpha.

In addition to subduing prey, venom is also used in defence and is capable of producing notoriously painful stings in humans (Balit et al. 2004), as well as death in mammals as large as dogs (McKeown 1930). Despite this, however, relatively little is known about centipede venoms. Mohamed et al. (1983) were the first to show enzymatic activity in centipede venom, namely phosphatase and esterase activity from the venom of *Scolopendra morsitans*. Several enzymatic activities were later found in the venoms of three Scolopendromorpha species (Malta et al. 2008), including hyaluronidase, phosphatase, esterase, proteases, and phospholipase activity, suggesting that enzymes play an important role in centipede venoms. These enzymatic activities, along with myotoxicity similar to or exceeding that of the pit viper *Bothrops jararacussu* (Malta et al. 2008) and the characterisation of a GroupX-like phospholipase A_2 (ScolPLA) (González-Morales et al. 2009), painted a picture of an unusual arthropod venom with an apparent superficial convergence to snake venoms.

However, the novelty of centipede venoms was apparent from early studies of their cardiotoxicity and neurotoxicity, where the responsible venom components were identified as being of surprisingly high molecular weight (Gomes et al. 1983; Stankiewicz et al. 1999). Sequencing efforts also confirmed the prevalence of undescribed toxin types; N-terminal sequencing of 24 proteins from two species of *Scolopendra* was only able to identify two CAP [CRiSP (cysteine rich proteins), Allergen (Ag-5), and Pathogenesis-related (PR-1)] proteins, leaving nine out of ten protein families unknown (Rates et al. 2007). Similarly, the two antimicrobial peptides described by Peng *et al.* (2009) bore no significant resemblance to any known protein sequence.

Recently, two studies have provided substantial insight into the composition and activities of venom from *Scolopendra subspinipes*. Yang et al. (2012) described 26 peptides with 2–4 internal disulfide bonds from the venom of *S. subspinipes mutilans*, including modulators of voltage-gated calcium (Ca_V), potassium (K_V) and sodium (Na_V) channels, and thereby provided the first conclusive evidence that centipede venoms were replete with bioactive cysteine-rich peptides. Furthermore, no sequence showed

significant similarity to that of any previously characterized peptide, confirming that centipede venoms are indeed a rich a source of novel toxins. The most comprehensive study to date is by Liu et al. (2012) who published 1122 full-length cDNA sequences encoding 543 different proteins from the venom gland of *S. subspinipes dehaani*, most of which showed little or no sequence similarity to known proteins. Forty peptides and proteins were also purified and characterised from milked venom, with 29 of these related back to their respective transcripts by N-terminal sequencing. In addition to previously described ion channel modulating and PLA_2 activities, this study revealed platelet aggregating, anticoagulant, and trypsin inhibitory activities in centipede venom.

No centipede-venom components have been described from species outside the genus *Scolopendra*, hindering any meaningful evolutionary analyses of these ancient venoms or any real insight into the degree of convergence with other venoms. Most protein families recruited into venoms share a number of characteristics that appear to be important for their ability to successfully be recruited as toxins. These include the presence of an N-terminal signal peptide that is cleaved off during maturation, stabilizing internal disulfide cross-linking, and an ancestral activity involved in a short-term physiological process (Fry et al. 2009a). However, it remains unknown whether centipede venoms show the same degree of convergence as is observed among other venomous animals and hence whether centipedes are subject to the same constraints that appear to govern the recruitment of body proteins into venoms (Fry et al. 2009a).

In addition to two uninvestigated *Scolopendra* species, we here present the first comprehensive overview of the "venomes" of two non-*Scolopendra* Scolopendromorpha and the very first description of non-Scolopendromorpha venom. Emphasizing the Scolopendridae due to their large size, availability, and clinical importance, the species that we selected allow for comparisons at order, sub-family, genus, and species level: *Thereuopoda longicornis* (Scutigeromorpha), *Ethmostigmus rubripes* (Scolopendromorpha; Scolopendridae; Otostigminae), *Cormocephalus westwoodi* (Scolopendridae; Scolopendrinae), *Scolopendra alternans* (Scolopendrinae), and *Scolopendra morsitans* (Scolopendrinae). Representing over 430 million years of evolution, the study of these species provides the first insight into the evolution of venom in centipedes as well as a comparison between two of the most divergent of centipede

species. In addition, we also sequenced the epidermal transcriptome of *E. rubripes* due to the hypothesized epidermal evolutionary origin of centipede venom glands (Dugon and Arthur 2012). We also present the first phylogenetic-based categorization of centipede venom peptides, and propose a new rational nomenclature to facilitate classification of future discoveries in this rapidly growing field. Our results show that centipede toxins exhibit a high degree of convergence with other venomous animals, and thereby contribute to the mounting evidence for a defined set of features that are associated with the weaponization of a protein across all venomous taxa.

Results and Discussion

Using a combined proteomic, transcriptomic and phylogenetic approach we were able to identify a total of 59 unique venom protein and peptide families from the five species of centipede investigated (tables 1 and 2, supplementary files 1–3). These were identified from a total of 25,697 contiguous sequences (contigs) obtained by next-generation sequencing of poly-A enriched RNA extracts from regenerating venom glands and distributed across the species as follows: *E. rubripes* 6980; *C. westwoodi* 1706; *S. alternans* 5044; *S. morsitans* 6029; *T. longicornis* 5938. In addition we obtained 1,216 contigs from the *E. rubripes* epidermis.

2D polyacrylamide gel electrophoresis (PAGE) analysis of crude venoms from *E. rubripes* (fig. 1A), *S. morsitans* (fig. 1B) and *T. longicornis* (fig. 1C) revealed a wide diversity of high molecular weight (HMW) protein families, a feature not commonly associated with arthropod venoms. The venom of *T. longicornis*, in particular, contained a large proportion of HMW proteins, with 11 of 15 families being composed solely of proteins larger than 15 kDa. Furthermore, about one third of all HMW families identified showed no significant sequence similarity to characterized proteins by BLAST and did not contain any functional or structural domains recognized by InterPro. The HMW and LMW families we identified in these centipede venoms are summarized in tables 1 and 2, respectively, while a summary of convergent recruitments is provided in table 3. All complete coding sequences from each family have been individually deposited in

GenBank and named according to the nomenclature proposed below. Tables listing proteomic evidence for submitted sequences can be found in Supplementary file 2.

Metalloproteases

Previous studies have shown that metalloproteases form an important component of Scolopendromorpha venoms (Malta et al. 2008). In support of this, we found that metalloproteases and putative metalloproteases are abundant in the venoms of all species studied, suggesting that this is a plesiomorphic characteristic in centipedes. In *T. longicornis*, astacin-like metalloendoproteases (MEROPS family M12, subfamily A) accounted for ~10% of proteins identified from the venom (fig. 1C). The M12 family is one of the most diverse families of metalloprotease and it has been recruited into the venoms of a wide range of animals including cephalopods, cnidarians, hymenopterans, platypus, reptiles, ticks, and spiders (Barrett et al. 2004; Brust et al. 2013; da Silveira et al. 2007; De Graaf et al. 2010; Francischetti et al. 2003; Ruder et al. 2013; Weston et al. 2013; Wong et al. 2012b). While most of these proteins are members of subfamily M12B, the metalloproteases in the venom from spiders of the genus *Loxosceles* belong to M12A (LALP, *Loxosceles* Astacin Like Protease), as do metalloproteases identified from the nematocysts of the sea anemone *Nematostella vectensis* (Moran et al. 2013). Although their role in envenomation remains to be determined (da Silveira et al. 2007; Trevisan-Silva et al. 2010), members of M12A are widespread throughout eukaryotic and prokaryotic taxa where they perform a range of functions including activation of growth factors, degradation of polypeptides, and processing of extracellular proteins (Bond and Beynon 1995). It has been suggested that proteases in spider venoms may be responsible for degrading the extracellular matrix around synapses, thereby providing the neurotoxic venom components with better access to their molecular targets (King and Hardy 2013). Members of the M12A subfamily can contain a number of additional domains C-terminal to the metalloprotease catalytic domain (Bond and Beynon 1995). Indeed, we detected two variants of M12A proteases from the venom of *T. longicornis*, one that contained an addition C-terminal CUB (Complement subcomponents Clr/Cls, embryonic sea urchin

protein Uegf, BMP-1) domain, and one that contained an epidermal growth factor (EGF) and a CUB domain C-terminal to the M12A active domain.

Although transcripts encoding M12A proteases were found in both *S. alternans* and *S. morsitans*, only putative metalloproteases were found in the venoms of the scolopendrid species. CUB domains were the only recognizable domains in these proteins, but they have some sequence homology to blastula protease 10, an M12A member from sea urchin (UniProt: P42674, E-value 0.001). Given the experimental evidence for the presence of metalloproteases in scolopendrid venoms (Undheim and King 2011) and the lack of other candidate sequences, this suggests that metalloproteases in scolopendrid venoms could be derived members of the M12A subfamily. As in *T. longicornis*, these putative M12A proteases accounted for approximately 10% of identified venom proteins in each of the scolopendrid venoms (fig. 1A,B).

Metalloproteases are involved in skin damage, oedema, blister formation, myonecrosis and inflammation, which is consistent with several of the recurrent symptoms associated with centipede stings (Gutiérrez and Rucavado 2000; Undheim and King 2011). Many members of the M12A subfamily also cleave matrix proteins, and could thereby aid in the spread of other venom components (Bond and Beynon 1995). While the roles of centipede venom metalloproteases remain unknown, the presence and absence of CUB domains suggests they have evolved to exhibit variable substrate specificities, e.g. by interactions with glycoproteins (Bond and Beynon 1995; Bork and Beckmann 1993).

Serine Proteases

We identified two families of serine proteases from scolopendrid venoms, namely trypsin-like S1 and subtilisin-like S8 proteases, supporting previous activity data that suggested the presence of serine protease activity in centipede venom (Malta et al. 2008). While S8 proteases have, to our knowledge, not previously been reported from venoms, S1 proteases are among the most widely recruited proteins into animal venoms and can be found in virtually all studied venomous taxa. As venom components they are involved in a range of functions, including vasodilation, smooth muscle contraction, anticoagulation and immunosuppression (Fry et al. 2009a; Low et al. 2013; Ma et al.

2012; Weston et al. 2013; Wong et al. 2012b). In centipedes, S1 transcripts were found in all species, including the previously published venom-gland transcriptome of *S. subspinipes dehaani* (Liu et al. 2012). Interestingly, these transcripts included both sequences with the characteristic S1 catalytic triad (His-Asp-Ser) and sequences where this triad was interrupted by a Ser to Gly substitution. This mutation has been implicated in the loss of proteolytic activity (Asgari et al. 2003), indicating that centipede S1 proteases may serve at least two types of functions. However, it is uncertain how important these functions are to the toxicity of the centipede venoms studied given that S1 protease could be detected only in low abundance in the venom of *S. morsitans*.

Similarly, subtilisin-like S8 protease transcripts were found in the venom gland transcriptomes of all scolopendrid species but they were detected proteomically only in low levels in the venoms of *E. rubripes* and *C. westwoodi* (fig. 1A). All S8 proteases identified belonged to the kexin/pro-protein convertase family, whose members are involved in the activation of growth factors, viral proteins and peptide hormones (Siezen and Leunissen 1997). Although the possibility of a venomous function cannot be discarded, their low abundance suggests that like S1 proteases, S8 proteases are mainly involved in the processing and activation of toxins. In support of this, only weak proteolytic activity was found in the otostigmine centipede *Otostigmus pradoi* after incubating venom with the metal chelator 1,10-phenanthroline (Malta et al. 2008). Interestingly, the presence of serine proteases in the venom suggests that centipede toxins are activated either during storage in the extracellular space, upon venom expulsion, or even both. This again opens the possibility that venom obtained by electrostimulation may contain unprocessed or partially processed toxins due to the involuntary secretion of venom, perhaps explaining the finding by Rates et al. (2007) that the same toxin was present with and without a 10-residue N-terminal tail.

γ-Glutamyl Transpeptidase

γ-Glutamyl transpeptidases (GGTs) are a widespread family of enzymes involved in regulation of oxidative stress and xenobiotic detoxification (Courtay et al. 1992). GGT has previously been reported from the venoms of parasitoid wasps, and is thought to

induce apoptosis of host ovaries by oxidative stress (De Graaf et al. 2010; Falabella et al. 2007). GGT is also found in the venom *S. subspinipes dehaani*, where it induces aggregation of human platelets and hemolysis of red blood cells from mice and rabbits but not humans (Liu et al. 2012). We found GGT-encoding transcripts in all species, albeit only as single partial contigs in *E. rubripes* and *T. longicornis*. In contrast, GGT was highly expressed in *S. alternans*, *S. morsitans*, and *C. westwoodi* venom glands. 2D-PAGE analysis revealed that GGT is abundant in *S. morsitans* venom (fig. 1B), and it was also among the most frequently identified venom proteins from *S. morsitans* and *C. westwoodi* in all mass spectrometry experiments. GGT therefore appears to be an important constituent of scolopendrine venoms and it was probably recruited into the venom subsequent to the split between the two scolopendrid subfamilies approximately 230 mya (Joshi and Karanth 2011).

Glycoside Hydrolases

As the name implies, members of the glycoside hydrolase (GH) superfamily hydrolyze the glycosidic bond between carbohydrates or between a carbohydrate and a non-carbohydrate moiety. Three glycoside hydrolase (GH) families were found in the scolopendrid venoms, namely chitinase (GH family 18), lysozyme (GH family 22), and hyaluronidase (GH family 56), although none of these were present across all species. While chitinases have been found in several venoms and could perhaps aid in digestion of arthropod prey (Balasubramanian et al. 2012; Chen et al. 2008; Fernandes-Pedrosa Mde et al. 2008; Kramer and Muthukrishnan 1998; Vincent et al. 2010), lysozyme hydrolyses β-1,4-links between *N*-acetylmuramic acid and *N*-acetyl-D-glucosamine in the peptidoglycan of bacterial cell walls and may therefore be an antibacterial agent (Chung et al. 1988; Low et al. 2013). Hyaluronidases are found in most animal venoms, where they are often regarded as "spreading factors" that increase the pathological impact of venom components such as haemorrhagins and neurotoxins (Fry et al. 2009a; Girish et al. 2004; Kuhn-Nentwig et al. 2004; Long-Rowe and Burnett 1994; Violette et al. 2012; Wong et al. 2012a). Hyaluronidase activity has previously been described from representatives of both scolopendrid subfamilies (Malta et al. 2008) and our results

confirm the presence of this GH family in scolopendrids as well as its recruitment in an early scolopendrid ancestor.

Phospholipase A_2

PLA_2 is another protein that has been recruited into a wide range of animal venoms, where they display a diverse array of catalytic and derived non-catalytic activities (Fry et al. 2009a; McIntosh et al. 1995). In centipedes, PLA_2 activity has been described from the milked venoms of both the scolopendrid subfamilies Otostigminae (*Otostigmus pradoi*) and Scolopendrinae (*Scolopendra viridicornis* and *S. viridis*) (González-Morales et al. 2009; Malta et al. 2008). However, centipede venom PLA_2 (ScolPLA_2) has only been characterised from *S. viridis* (González-Morales et al. 2009) and *S. subspinipes dehaani* (Liu et al. 2012). We found PLA_2 homologues in the transcriptomes of *S. alternans* and *S. morsitans* as well as in the venom of *S. morsitans* and *E. rubripes*. A heavily mutated transcript in *C. westwoodi* containing several internal stop codons was also found, suggesting a secondary loss of PLA_2 in this species. Phylogenetic analysis revealed that all centipede PLA_2 form a monophyletic sister-clade to Group X-related PLA_2 and were hence recruited into centipede venom at least 230 mya.

Other Enzymes

Glucose dehydrogenase was found to be present in relatively high levels in the venom of all scolopendrids studied. Partial transcripts were also found in *T. longicornis.* and *S. alternans*, suggesting this may be a basal recruitment in centipedes. Glucose dehydrogenase belongs to the glucose-methanol-choline oxidoreductase family, and it is involved in the pentose phosphate pathway (Bak 1967). However, no venomous role has been described for any member of this family and it is unclear how the ancestral pathway may contribute to toxicity or processing of toxins in the venom gland. The abundance in the venom therefore suggests that centipede venom glucose dehydrogenase may have undergone neofunctionalisation.

Non-specific esterases are found in the venoms of a diverse set of taxa, including spiders (Heitz and Norment 1974; Rodrigues et al. 2006), snakes (Sulkowski et al. 1963; Tu and Chua 1966) and octopus (Undheim et al. 2010). Activity has also been described in venom gland extracts from *S. morsitans* (Kass and Batsakis 1979; Mohamed et al. 1983). Indeed, type B carboxyl esterase was found in the venom of *C. westwoodi* and homologous transcripts identified in the remaining species except *E. rubripes*. Although their venomous function is uncertain, venom carboxyl esterases are thought to play a part in the release of endogenous purines during envenomation, which then act as "multitoxins" causing a multitude of pharmacological effects including immobilization through hypotension (Aird 2002; Dhananjaya and D'Souza 2010).

As evident from our 2D-PAGE analysis (fig. 1C), *T. longicornis*. venom contained a number of isoforms of *Porphyromonas*-type peptidyl arginine deiminase (PPAD). Five full-length isoforms were also found in the transcriptome as well as a number of incomplete transcripts. The full-length sequences showed 81% similarity to a putative PPAD from *Pseudomonas entomophila* (UniProt: Q1I4Y0) but no significant homology to any metazoan sequence. PPAD catalyses deamination of the guanidino group on C-terminal Arg residues to yield ammonia and a citrullinated residue (Shirai et al. 2001). It is, however, uncertain what the role of centipede PAD, or centiPAD, is in the venom or venom gland of *T. longicornis* but we speculate that it might be involved in posttranslational modification of toxin arginine residues.

Centipede β-Pore-forming Toxins

Among the abundant proteins in the venoms of all species studied, in particular *T. longicornis* and *E. rubripes*, were large proteins containing pore-forming domains similar to those found in *Aeromonas* aerolysin and *Clostridium* epsilon toxins (fig. 1). Phylogenetic analysis suggests that this family may have been recruited into the venom of centipedes prior to the split of Notostigmophora and Pleurostigmophora over 430 mya (Murienne et al. 2010) and that it has since undergone substantial diversification and gene duplication (fig. 2). Most members of this family contain an aerolysin-like toxin β-complex domain (IPR023307), which appears to be an ancestral trait. The β-complex

domain is a structural domain found in all members of the widely distributed aerolysin-like β-pore-forming toxin (β-PFTx) superfamily. It is directly involved in the pore-forming activity of these toxins via the formation of a transmembrane β-barrel upon assembly of multiple toxin monomers (Knapp et al. 2010). Membrane insertion and subsequent oligomerization of toxin monomers is mediated by binding of toxins to various cell-surface receptors through additional toxin domains. Receptor selectivity may also differ between isoforms, such as in hydralysins (Sher et al. 2005), thereby exhibiting cell-type specificity. The diversity of centipede β-PFTx therefore suggests that these toxins target a range of cell types and tissues, asserting a multitude of toxinological functions. Interestingly, aerolysin requires proteolytic activation in order to oligomerize into the pore-forming heptameric form (Knapp et al. 2010). This can be carried out by a number of proteases, including serine proteases of MEROPS families S1 and S8, both of which are present in Scolopendromorpha venom.

The pore-forming properties of centipede β-PFTx are supported by previous findings that centipede venoms possess highly cytolytic properties (Malta et al. 2008; Quistad et al. 1992). The presence of β-pore formers explains the puzzling results obtained in early studies of the neurotoxic effects of centipede venom, where a venom component of about 80 kDa induced an increase in leak current in the giant axon of the cockroach *Periplaneta americana* (Stankiewicz et al. 1999). The abundance of centipede β-PFTx might also account for the high myotoxic and oedematogenic activities of centipede venoms (Malta et al. 2008), as well as many of the symptoms associated with centipede envenomations (Undheim and King 2011).

CAP Proteins

CAP proteins are among the protein types most widely recruited into animal venoms, where they can act as ion channel modulators, vasodilators, myotoxins, or even proteases (Fry et al. 2009a; Moran et al. 2013; Peichoto et al. 2009; Wong et al. 2012a). In centipedes, CAP proteins have been described from the venoms of *S. angulata*, *S. viridicornis*, and *S. subspinipes dehaani* (Liu et al. 2012; Rates et al. 2007). Supporting the notion that CAP proteins are important constituents of centipede venoms (Liu et al.

2012), we found CAP proteins to be among the most abundant proteins in all scolopendrid venoms and they were also present in scutigerid venom, with multiple isoforms identified in all cases. Phylogenetic analysis revealed CAP proteins were recruited into the venom of centipedes three times; once prior to the split of Notostigmophora and Pleurostigmophora over 430 mya (centipede CAP type 1, centiCAP1), once prior to the division of Scolopendridae about 230 mya (centipede CAP type 2, centiCAP2), and once since the division between *S. morsitans* and the remaining *Scolopendra* spp. within the last 100 mya (centipede CAP type 3, centiCAP3) (fig. 3) (Joshi and Karanth 2011). Interestingly, centiCAP1 proteins diversified both prior to and after the split of Notostigmophora and Pleurostigmophora, although the major radiations probably occurred within a Scolopendridae ancestor. Furthermore, our results suggest that centiCAP1 has been lost in Scolopendrinae. While the activity of centiCAP1 and centiCAP3 remain to be elucidated, centiCAP2 proteins have undergone neofunctionalisation with mutually monophyletic clades exhibiting Ca_V channel inhibiting or trypsin inhibiting activities respectively (Liu et al. 2012). CAP proteins have are also among the principal allergens in vespid and fire ant (*Solenopsis* spp.) venom (Hoffman 2006), and consequently their abundance in scolopendrid venoms might at least partly explain the relatively frequently observed allergic reactions to centipede stings (Undheim and King 2011).

LDLA Domain Containing Proteins

Similar to CAP proteins, we found a large abundance and diversity of proteins containing a low-density lipoprotein receptor Class A repeat (LDLA) domain. The LDLA domain is a structural domain present in a wide range of proteins that signifies a β-hairpin structure followed by a series of β turns (Daly, et al. 1995), although we could not find any representatives from venoms. While the function of these LDLA-containing proteins remains to be determined, their abundance and diversity suggests that they play an important role in centipede venom (fig. 4). Phylogenetic analysis revealed that this protein family was recruited in the venom of an early centipede at least 430 mya, and the

subsequent diversification suggests that several neofunctionalisation events have occurred.

Transferrin

Transferrin was identified in the venoms of *E. rubripes* and *S. morsitans*, and transcripts were found in all species. Invertebrate members of the transferrin superfamily have been implicated in pathways involved in the reaction to secondary infections by binding iron and thereby creating an environment low in free iron that impedes bacterial survival (Ramirez-Gomez et al. 2008). Centipede venom transferrin could therefore potentially function as an antimicrobial agent in the venom gland. Interestingly, draculin, an anticoagulant protein from the venom of the vampire bat *Desmodus rotundus*, has recently been shown to be a member of the transferrin family (Low et al. 2013), and could thereby represent a convergent recruitment between bat and centipede although their functions are likely to differ.

Cystatin

Cystatins are ancestral, potent inhibitors of papain family cysteine proteases, and they have been found in a range of animal venoms including reptiles, *Lonomia* caterpillar bristles, ticks and mosquitoes (Fry et al. 2009a; Fry et al. 2013). Two cystatin isoforms were present in the venom of *E. rubripes*. Both contained the characteristic peptidase interacting sequence Gln-X_{aa}-Val-X_{aa}-Gly as well as the cystatin type-1 like Pro-Gly pair, suggesting that these venom proteins have retained their ancestral function as peptidase inhibitors (Abrahamson et al. 2003).

Unknown High Molecular Weight Venom Proteins

Eleven protein families could not be assigned a putative function or known family. Many of these were both diverse and abundant (Supplementary File 1). In addition, we discovered two protein families that contained domains of unknown function (DUF). The first of these, DUF 3472 (IPR021862), was only found in scolopendrid venoms. No

homologues were found in the transcriptome of *T. longicornis*, nor were any similar arthropod sequences found by BLAST. The second family containing a domain with unknown function, DUF 1397 (IPR009832), was only found in the venom of *T. longicornis*. Similar incomplete transcripts were present in all scolopendrid venom gland transcriptomes which showed significant similarity to hemolymph glycoproteins with unknown function. DUF 1397 has been described mainly from insect haemolymph glycoproteins with unknown function (Samaraweera and Law 1995), suggesting perhaps a recruitment of a hemolymph protein into the venom of *T. longicornis*.

Classification of Peptide Toxin Families

Only one attempt has been made at categorizing centipede-venom peptides made (Liu et al. 2012). Although this categorization provided insight into the pharmacological diversity of centipede venoms, it unfortunately made little sense from a phylogenetic perspective due to the extensive functional radiation exhibited by several toxin families (e.g. SLPTX family 15, see below). Similarly, the nomenclature for naming spider and sea anemone toxins (King et al. 2008; Oliveira et al. 2012) is not able to fully depict the phylogenetic relationships of centipede venom peptides. This is because it relies on activity, organismal taxonomy and general sequence diversity, none of which are able to reflect the phylogenetic relationship of centipede toxins. Due to the practicality of this nomenclature, however, we propose only a minor modification where the phylogenetic family to which the peptide belongs is denoted by a subscript number corresponding to that of the family. This allows venom components to be classified under the same scheme based on activity, organismal taxonomy *and* molecular phylogeny. Using the proposed nomenclature, the recently described voltage gated potassium channel (K_V) inhibitor **κ-SLPTX-Ssm1a** would thus become **κ-SLPTX$_3$-Ssm1a**, indicating that it is a K_V inhibiting member of Scolopendromorpha toxin family number three that was first discovered in *S. subspinipes mutilans* (Yang et al. 2012). As an example, using this system one is able to convey that **μ-SLPTX$_{15}$-Ssd1a** (GenBank: KC144793) and **κ-SLPTX$_{15}$-Ssd1a** (GenBank: KC144556) are both members of the same toxin family,

whereas **κ-SLPTX$_3$-Ssm1a** is not, but at the same time provide information on pharmacological activity and organismal origin of these toxins.

Phylogenetic Distribution of Cysteine-Rich Toxins

According to the nomenclature described above, we classified peptides (<10 kDa) identified from centipede venoms and venom gland transcriptomes into 28 families (table 2). Of these, only two families of cysteine-rich peptides first found in Scolopendromorpha venom were present in both orders (scoloptoxins, or SLPTX). In addition, two cysteine-rich families and one linear peptide family were found only in *T. longicornis* (scutigerotoxins, or SCUTX), compared to 18 cysteine-rich and 5 linear peptide families in the scolopendromorphs.

Although two families of cysteine-rich peptides were found in the transcriptomes of both Scutigeromorpha and Scolopendromorpha species, only one of these could be detected in the venoms of both orders. This family, SLPTX family 1, was first found in the venom of *E. rubripes* and is characterized by the presence of a type 2 chitin-binding (CB_2) domain (IPR002557). The CB_2 domain contains six conserved cysteine residues and is present in a range of proteins including insect peritrophic matrix proteins and animal chitinases. Homologues were found in the venoms of *T. longicornis* and *C. westwoodi*, as well as the *E. rubripes* epidermis transcriptome, and a BLAST search revealed that these are similar to insect epidermal proteins analogous to peritrophins. Sequences identical to the epidermal CB_2 domain-containing transcripts were also found in the venom-gland transcriptome of *E. rubripes*. These were not, however, detected in the venom and contained three CB_2 domains compared to just one in the venom forms (fig. 5), suggesting a secondary loss of domains one and three around the time of recruitment. Interestingly, given the epidermal origin of the centipede venom-gland, this family may represent one of the earliest peptide toxin recruitments in centipedes and perhaps even arthropods.

The majority of venom SLPTX families identified were recruited prior to the split between the two scolopendrid subfamilies Otostigminae and Scolopendrinae approximately 230 mya (Joshi and Karanth 2011). Only three of these (SLPTX families 4,

5 and 16), show evidence of having undergone radiation prior to this division (figures 6-8). This may, however, be a result of our limited taxonomic sampling within the Otostigminae (*E. rubripes* only), or perhaps a combination of the large number of generations since divergence and positive selection regimes thought to be experienced by toxin peptides during functional radiation (Casewell et al. 2013). Three cysteine-rich families were found to be exclusive to *E. rubripes* (SLPTX families 2, 17, 18), one to *C. westwoodi* (SLPTX family 8), and four to *Scolopendra* (SLPTX families 3, 6, 7, 11). Of these, SLPTX families 6 and 7 were exclusive to *S. morsitans* and *S. subspinipes*, respectively.

In stark contrast, only two cysteine-rich peptide families were found exclusively in *T. longicornis*. Furthermore, both SCUTX families displayed a profound lack of diversity, with one and three complete toxin sequences in SCUTX family 1 and 2, respectively. This was surprising given the absence of other families and the hypothesized dependence on venom for prey capture. While the lack of common structural scaffolds between Scolopendromorpha and Scutigeromorpha can be explained by the long period since organismal divergence, the lack of peptide diversity is puzzling and suggests a difference in importance of venom peptides between the two orders.

Functional Radiation of Cysteine-Rich Toxins

Our phylogenetic classification of toxins clearly shows good examples of functional diversification within several centipede toxin families, despite there only having been a handful of toxins described in terms of activity. Functional radiation is particularly evident in SLPTX family 15, whose family members have evolved to inhibit at least three types of voltage-gated ion channels (Ca_V, K_V, and Na_V; fig. 9). This radiation is likely to have occurred during the approximately 200 million years since the split between *Scolopendra* and *Cormocephalus,* as family 15 appears to have been secondarily lost in the latter. Similar phenomena are apparent among members of SLPTX families 3 and 10, where toxins belonging to the same family and contained in the same venom have evolved different ion channel modulatory activities. Remarkably, in the case of SLPTX family 10, our sequence alignment revealed that a single Asp-Gly mutation next to the

last cysteine residue appears to be responsible for the shift from K_V- to Ca_V-channel inhibition that is apparent from the results of Liu et al. (2012).

The last family (SLPTX family 11) that shows evidence of functional diversification represents a particularly interesting example in terms of centipede toxin evolution. The first toxin described from this family was the K_V-channel inhibitor κ-SLPTX$_{11}$-Ssm3a, a 7.9 kDa toxin with four Cys-residues (Yang et al. 2012). However, our analyses show that this toxin belongs to the same family as three 21–23.5 kDa toxins containing up to 16 cysteine residues described as Group 1–3 K_V-inhibitors by Liu et al. (2012). SLPTX family 11 also contains a similar 22.2 kDa toxin with 16 cysteine residues discovered in the same study that exhibits anticoagulant properties. While the K_V channel inhibitors and anticoagulant protein form reciprocally monophyletic clades, events of extensive N- and C-terminal truncations have taken place in both clades (fig. 10). In addition, we also identified variants from the venom of *S. morsitans* that ranged from 6.9 kDa with six cysteine residues to the ancestral 23 kDa form with 16 cysteine residues. Given the structural importance of cysteine residues due to the formation of stabilizing disulfide bonds, it will be interesting to see what structural and functional alterations and adaptations, if any, have taken place among members of SLPTX family 11.

Linear Venom Peptides

There was a much lower diversity of non-reticulated (linear) peptides compared to cysteine-rich peptides in the venoms studied (table 2). It is noteworthy that two linear peptide hormone families have been recruited into centipede venom, namely insect diuretic hormone (SLPTX family 20) and hypertrehalosaemic hormone (SLPTX family 21). The latter of these was found in the venoms of both *E. rubripes* and *C. westwoodi*, but the loss of phylogenetic signal due to saturating sequence divergence of these small peptides makes it uncertain whether these represent a single or independent recruitment events. Most linear peptides lack the structurally conserved residues equivalent to those found in cysteine-rich peptides. Combined with the high variation in the signal peptide region of centipede-toxin prepropeptides, it is therefore difficult to pinpoint phylogenetic relationships and recruitment events for linear peptides. This could be one reason why we

were unable to detect transcripts coding for peptides related to Scolopin-1 and -2 (Peng et al. 2009) in any of the venoms studied here.

Convergent Recruitment of Peptide Scaffolds

In addition to the 12 convergently recruited HMW protein types (table 3), four LMW venom peptide families showed evidence for convergence with structural scaffolds from other venomous animals: CS-α/β defensin (SLPTX family 2), inhibitory cystine knot (ICK) (SLPTX family 13), Von Willebrand factor type C (VWC)-like (SLPTX family 16), and colipase (SLPTX family 18). Colipase belongs to the same superfamily as the AVIT-toxins from reptile venoms, which are potent agonists of prokineticin receptors and induce smooth muscle contractions and hyperalgesia (Fry 2005; Fry et al. 2009a). These symptoms are consistent with those of centipede stings, suggesting colipase-like toxins identified in the venom of *E. rubripes* may play a similar role. Interestingly, we recovered a homologous sequence in the epidermal transcriptome of the same species suggesting that, as for SLPTX family 1, this may be another example of toxin recruitment from an epidermal gland protein.

A BLAST search of ω-SLPTX$_{13}$-Ssm2a (UniProt: I6S390) revealed low overall sequence similarity (36% identity, 47% positives, E-value 0.83) with U10-lycotoxin-Ls1a (UniProt: B6DD02) from the venom of the wolf spider *Lycosa singoriensis*. However, there are several features suggestive of structural homology, including conserved cysteine pattern, inter-cysteine spacings, and residues surrounding cysteine residues and parts of the signal peptide. The ICK structural family is characterized the possession of an inhibitor cystine-knot motif (Pallaghy et al. 1994; King et al. 2002) that provides these toxins with high levels of chemical, thermal and biological stability (Saez et al. 2010). It is found in several unrelated peptide families in a range of venomous animals (Gracy et al. 2008). However, they are particularly diverse in spider venoms, where the majority of cysteine-rich peptide toxins described to date assume an ICK-type fold (Vassilevski et al. 2009). Unlike spiders, however, centipede ICKs do not appear to constitute the majority of toxin structural folds. Centipede ICK toxins were recruited into the venom of centipedes prior to the split within Scolopendridae about 230 mya. Despite this, members

of centipede ICK show remarkable conservation of the mature peptide region, even more so than the signal peptide region. This is highly unusual for peptide toxins, where the reverse tends to be the case (e.g. Woodward et al. 1990; Sollod et al. 2005; Pineda et al. 2014). A member of the centipede ICK-toxin family found in the venom of *S. subspinipes*, ω-SLPTX$_{13}$-Ssm2a, was recently shown to inhibit Ca_V channels (Yang et al. 2012), and it will be interesting to see whether the conservation of the mature peptide has led to a degree of functional retention across clades.

In addition to convergent recruitment with spider venoms, we found evidence for two toxin recruitments convergent with scorpion venoms. As revealed by BLAST, members of SLPTX family 16 showed significant similarity to putative venom peptides with unknown function described from a venom-gland transcriptome from the scorpion *Opisthacanthus cayaporum* (fig. 8) (Silva et al. 2009). Interestingly, they are also similar to shrimp sequences containing single VWC domains that are likely involved in anti-viral immunity (Chen et al. 2011). SLPTX family 16 may thus represent a recruitment of immune-response components into centipede venom over 230 mya. At least three clades within this family arose prior to the split between the two scolopendrid subfamilies, including one lineage with an additional cysteine residue potentially involved in dimerization and one which exhibits a high variation in cysteine residues (fig. 8).

Another immune-response component recruited into centipede venom is the cysteine stabilized α/β (CS-α/β) defensins of SLPTX family 2. A BLAST search revealed that these are similar to antimicrobial defensins from mollusks and insects. The CS-α/β defensin scaffold has been recruited into several invertebrate venoms, particularly in scorpions where it accounts for the majority of peptide neurotoxins (Fry et al. 2009a). Like knottins, however, this is not the case in centipede venom, as defensins were only observed in the venom and venom-gland transcriptome of *E. rubripes.*

Conclusions

As is evident from figure 11, which summarizes the phylogenetic distribution of toxin types as their earliest respective recruitments, the ancient evolutionary history of centipedes is reflected in their differences in venom composition. Despite having a

venom apparatus unfit to assert physical dominance over prey, Scutigeromorpha appear to have a much lower venom peptide and protein diversity than scolopendromorphs. As a result, the two orders probably employ substantially different venom strategies, with Scutigeromorpha appearing to rely more on non-peptidic compounds enzymatically produced in the gland. In contrast, scolopendrid venoms are replete with ion-channel modulating cysteine-rich peptides, with *Scolopendra* species in particular appearing to harbor a high diversity of cysteine-rich toxins. Although it may be an artifact of our limited taxonomic sampling outside this genus, our data suggest that this is a derived trait in centipedes.

The diversity of peptide structural scaffolds in Scolopendromorpha venoms is remarkable. No one peptide structural family accounted for a significant proportion of the peptides discovered, which is in contrast to other arthropod venoms such as scorpions and spiders, where peptide toxins are dominated by a single structural scaffold (CS-α/β defensins and knottins, respectively). The overall number of toxin scaffolds also sets scolopendrid venoms apart from those of other arthropods, and is more similar to the diversity of scaffolds found in snakes. This divergence in venom composition and complexity highlights centipedes as excellent models to further our understanding of venom evolution.

While we were unable to identify corresponding venom proteins from other taxa for the majority of the toxin families discovered, there are nevertheless a large number of convergently recruited toxin families in centipede venoms. Interestingly, two of the most abundant and diverse HMW venom components across all scolopendrid venoms are among the most commonly recruited toxin types in animal venoms, namely M12 protease and CAP proteins. M12A was also the second most abundant venom component in *T. longicornis*. Furthermore, the vast majority of the novel toxin families showed the typical characteristics of previously described venom proteins (Fry et al. 2009a), namely the presence of a signal peptide and extensive putative internal disulfide cross-linking. Our results therefore add to the mounting evidence that a limited set of constraints govern the chance of recruitment of body proteins into animal venoms.

Lastly, understanding the evolution of these ancient venom systems not only also broadens our understanding of which traits make peptides stable and amenable to neofunctionalisation, but may also aid in directing future biodiscovery efforts. Moreover, the number of venom proteins we identified with no previously known venomous function highlights the power of a combined transcriptomic, proteomic and phylogenetic approach for the investigation of novel venoms. While our results cast light on a hitherto largely neglected group of venomous animals, they raise a large number of questions, such as the underlying reasons for the disparity between Scolopendromorpha and Scutigeromorpha venom composition and diversity. It will also be particularly interesting in future to compare venoms at higher taxonomical levels within Chilopoda. We hope that this contribution will help to spark greater interest in the venoms of these fascinating animals.

Experimental Procedures

Specimen and Venom Collection

Five centipede species were selected to provide both wide taxonomic coverage and a phylogenetic timeline to date evolutionary events. *E. rubripes* and *T. longicornis* were purchased from Mini Beast Wildlife (www.minibeastwildlife.com.au), *S. morsitans* was collected from the Darling Downs region, Queensland, Australia, *C. westwoodi* was collected from the Launceston region, Tasmania, Australia. All species were identified according to Koch (Koch 1983a,b,c) and the Centipedes of Australia online guide (CSIRO Entomology, updated 27 May 2013: http://www.ento.csiro.au/biology/centipedes/centipedeKey.html). *S. alternans* (Haiti) were purchased from La Ferme Tropicale (www.lafermetropicale.com).

Venom was collected from all species except *S. alternans*. Animals were starved for three weeks before milking to allow full regeneration of all venom components. For venom collection from Scolopendromorpha species, specimens were anaesthetised with CO_2, then venom was extracted by electrostimulation (12 V, 1 mA) into pipette tips covering the tip of each forcipule to prevent contamination by other fluids. This approach, however, proved unsuccessful in the case of *T. longicornis* due to their tendency to

autotomize limbs. Instead, specimens were held by a pair of forceps, agitated, and allowed to pierce and envenomate a handheld frame supporting a double layer of parafilm. This was done multiple times for each specimen, allowing up to 15 min between each attempt, until no more venom was secreted. All venom was immediately placed on dry ice, lyophilized, and stored at –80°C until processing. Polyethylene equipment was used to collect and process samples in all cases.

cDNA Library Construction

To effectively sequence and identify proteins from venoms as well as attain data on toxin families and any physiological gene homologs, transcriptomes were constructed by NextGen sequencing of poly-A enriched RNA extracts from regenerating venom glands. In addition to the venom extraction protocol described above, venom glands were further depleted by agitating the centipede and allowing it to envenomate live crickets (*Acheta domesticus*). Venom glands were considered fully depleted once no toxicity could be observed even after allowing the centipede to rest for 10 min. For each species, venom glands from five specimens anaesthetized with CO_2 were dissected out four days after venom depletion, flash frozen, and pooled. Total RNA was extracted using TRIzol (LifeTechnologies) and enriched for mRNA using a DynaBeads Direct mRNA kit (LifeTechnologies). One specimen of *E. rubripes* was washed with 70% ethanol, cuticle dissected from all legs on one side, and mRNA enrichment carried out as stated above.

mRNA was reverse transcribed, fragmented and ligated to a unique 10-base multiplex identifier (MID) tag prepared using standard protocols and applied to one PicoTitrePlate (PTP) for simultaneous amplification and sequencing on a Roche 454 GS FLX+ Titanium platform (Australian Genome Research Facility). Automated grouping and analysis of sample-specific MID reads informatically separated sequences from the other transcriptomes on the plates, which were then post-processed to remove low quality sequences.

Transcriptome Assembly and Functional Annotation

Reads obtained by 454 sequencing were assembled (*de novo*) into contiguous sequences

(contigs) using v 3.4.0.1 of the MIRA software program. Assembly details (number of reads, average read length, number of contigs and average assembled bases per contig) were: *E.rubripes* 72740, 375, 6980, 1035; *E. rubripes* epidermis 28575, 386, 1216, 547; *C. westwoodi* 48041, 376, 1706, 544; *S. alternans* 57175, 355, 5044, 612; *S. morsitans* 93436, 356, 6029, 621; *T. longicornis*. 60008, 369, 5938, 1115. Public access to the data can be found at the National Center for Biotechnology Information (NCBI) under bioprojects, PRJNA200639 (*E. rubripes*), PRJNA213033 (*E. rubripes* epidermis), PRJNA200641 (*C. westwoodi*), PRJNA200753 (*S. alternans*), PRJNA200640 (*S. morsitans*), and PRJNA213032 (*T. longicornis*). Corresponding accession numbers of individually curated sequences deposited with the transcriptome shotgun assembly (TSA) database are GASI01000001–GASI01000195 (*E. rubripes*), GASJ01000001–GASJ01000009 (*E. rubripes* epidermis), GASL01000001–GASL01000050 (*C. westwoodi*), GASK01000001–GASK01000051 (*S. alternans*), GASH01000001–GASH01000185 (*S. morsitans*), and GASR01000001–GASR01000119 (*T. longicornis*).

Assembled contigs were processed using CLC Main Work Bench 6.2 (CLC-Bio) and the Blast2GO bioinformatic suite (Conesa et al. 2005; Götz et al. 2008) to provide BLAST, gene ontology, and domain annotation. BLAST searches were performed by searching nucleotide sequences against the NCBI non-redundant database using tblastx with a significance threshold at a BLAST expect value of 10^{-3}, while domain annotations were obtained using the full suite of applications available through InterPro (Zdobnov and Apweiler 2001). Additionally, proteomically identified sequences were searched against the Universal Protein Resource (UniProt) database using default settings, and locally against a database containing transcriptomes of all species and all publicly available centipede sequences using CLC Main Work Bench.

Contigs from all transcriptomes were pooled and translated to all reading frames to provide a sequences database for the proteomic identification of venom components. The 6-frame translation retained as many open reading frames (ORFs) from each DNA entry that met the conditions of being longer than 8 amino acid residues and not shorter than the next longest ORF from that DNA entry by a factor of 0.8, yielding a total of 168,927 amino acid sequences. BLAST searches and InterPro scans were conducted on translated sequences identified proteomically to verify that the correct reading frame was annotated.

2D-PAGE Analysis

To visualize HMW venom components, we employed 2D-PAGE. Venom samples (1.5 mg) from a species from each of the major clades (*E. rubripes, S. morsitans* and *T. longicornis*) were directly solubilized in 410 µl of rehydration buffer (8M urea, 100 mM DTT, 4% CHAPS, 110 mM DTT and 0.01% bromophenol blue). The sample was mixed and centrifuged (5 min, 14,000 rpm) to pellet any insoluble material, then 0.5% ampholytes (Biolytes pH 3–10) was added to the supernatant before loading onto isoelectric focusing (IEF) strips (Bio-Rad ReadyStrip, non-linear pH 3–10, 24 cm IPG) for 24 h passive rehydration. Proteins were focused in a Ettan IPGphor3 (GE Lifesciences, USA) with the IEF conditions as follows: 100 V for 1 h, 500 V for 1 h, 1,000 V for 1 h and 8,000 V for 98,400 V/h. After running IEF, IPG strips were equilibrated for 10 min in a reducing equilibration buffer (50 mM Tris-HCl, pH 8.8, 6 M urea, 2% SDS, 30% glycerol, 1.5% DTT) followed by a second incubation for 20 min in an alkylating equilibration buffer (50 mM Tris-HCl, pH 8.8, 6 M urea, 2% SDS, 30% glycerol, 2% iodoacetamide). IPG strips were briefly rinsed in SDS-PAGE running buffer and embedded on top of the 12% polyacrylamide gels (26 × 20 cm Ettan DALTsix Electrophoresis unit, GE Lifesciences, USA) and covered with 0.5% agarose. Second dimension electrophoresis was performed at 10°C for 5 mA per gel for 20 h followed by 20 mA per gel for 14 h. The resulting gels were stained in 0.2% colloidal coomassie brilliant blue G250 (34% methanol, 3% phosphoric acid, 170 g/L ammonium sulfate, 1 g/L Coomassie blue G250) overnight and destained in 1% acetic acid/H_2O. Visible spots were subsequently picked from gels and digested overnight at 37°C using sequencing grade trypsin (Sigma, USA). Briefly, gel spots were washed with ultrapure water, destained (40 mM NH_4CO_3/50% Acetonitrile (ACN)) and dehydrated (100% ACN). Gel spots were rehydrated in 10 µL of 20 µg/ml proteomics-grade trypsin (Sigma-Aldrich) and incubated overnight at 37°C. Digests were eluted by washing the gel spots for 20 min with each of the following solutions: 20 µL 1% formic acid (FA), followed by 20 µL of 5% ACN/0.5% FA.

Mass Spectrometry—AB SCIEX 5600

Reduction and alkylation of cysteine residues in venom proteins and peptides was performed according to the protocol described by Hale et al. (2004), then 4 µg of reduced and alkylated venom was incubated overnight in 10 µL 20 ng/µL trypsin in 40 mM NH_4CO_3 pH 8 at 37°C. Digested and undigested reduced/alkylated samples were resuspended in a final concentration of 0.5% formic acid, 2% ACN, centrifuged for 15 min at 14,000 rcf prior to LC-MS/MS. 0.75 µg was processed by LC-MS/MS, using a Agilent Zorbax stable bond C18 column (2.1 mm × 100 mm, 1.8 µm particle size, 300 Å pore size) at a flow of 400 µl/min and a gradient of 1–40% solvent B (90% acetonitrile [ACN], 0.1% formic acid [FA]) in 0.1% FA over 15 min on a Shimadzu Nexera UHPLC coupled with an AB SCIEX 5600 mass spectrometer equipped with a Turbo V ion source heated to 450°C. MS^2 spectra were acquired at a rate of 20 scans/s, with a cycle time of 2.3 s, and optimised for high resolution. Precursor ions with m/z between 300–1,800 m/z, a charge of +2 to +5, and an intensity of at least 120 counts/s were selected, with a unit mass precursor ion inclusion window of ± 0.7 Da, and excluding isotopes within 2 Da for MS^2.

MS^2 spectra were searched against translated predicted ORFs from all assembled transcriptomes using ProteinPilot v4.5 (AB SCIEX). Searches were run as thorough identification searches, specifying tryptic digestion and alkylation reagent as appropriate. Biological modifications and amino acid substitutions were allowed in order to maximize the identification of protein sequences from the transcriptome despite the inherent variability of toxins, potential isoform mismatch with the transcriptomic data, and to account for experimental artifacts leading to chemical modifications. Spectra were inspected manually to eliminate false positives, excluding spectra with low signal to noise, erroneous modification assignments, and confidence values below 95% unless justifiable by the presence of a-ions. For the tryptic digests, only protein identifications supported by two or more unique high-confidence spectra were included. For identification of 2D gel spots, protein identifications that did not meet the above stringent criteria were still considered valid if that protein had been identified by the other proteomic approaches and was supported by multiple unique spectra primarily penalized

due to low signal-to-noise ratio. The results were further analysed using CLC Main Workbench v6.6.

Mass Spectrometry—Thermo Fisher Orbitrap

Due to the novelty of centipede venoms we decided to complement our proteomic results with an additional mass spectrometry platform in order to maximize the identification of venom components. 1 µg venom was reduced and alkylated according to the protocol described by Hale et al. (2004), then 500 ng reduced/alkylated venom was digested with either trypsin or Glu-C (Sequencing Grade, Promega, Madison, WI), at a ratio of 1:50 enzyme:substrate with vortexing at 37°C for 4 h. Aliquots of each digest were analyzed by automated nano LC-MS/MS using two different reversed phase LC separation methods; a classical HPLC method and an ultra-high pressure UHPLC method. Both LC methods were coupled to the same Orbitrap Elite Fourier transform mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with a nanoflow ionization source (James A. Hill Instrument Services, Arlington, MA).

For the classical HPLC LC-MS/MS method, aliquots of each digest (150 ng) were separated using an Agilent 1200 nano-LC system (Agilent Technologies, Wilmington, DE). Peptides were eluted from a 10-cm column (Picofrit 75 µm ID, New Objectives) packed in-house with reverse-phase resin containing either Zorbax C-18 with 3 µm particles, 300 Å pores (Agilent Technologies) for *C. westwoodi*, *E. rubripes*, and *S. morsitans* venoms, or ReproSil-Pur C18-AQ 3 µm particles, 120 Å pores (Dr. Maisch, Ammerbuch Germany) for *T. longicornis* venom. The 90 min ACN/0.1% FA gradient was run at a flow rate of 200 nl/min to yield ~20 s peak widths. Solvent A was 0.1% FA and solvent B was 90% ACN/0.1% FA. The elution portion of the LC gradient was 3–6% solvent B in 1 min, 6–36% in 40 min, 36–56% in 10 min, 56–90% in 5 min and held at 90% solvent B for 9 min. Data-dependent LC-MS/MS spectra were acquired in ~3 s cycles, with each cycle of the following form: one full Orbitrap MS scan at 60,000 resolution followed by 12 MS/MS scans in the Orbitrap at 15,000 resolution using an isolation width of 3.5 m/z. The top five most abundant precursor ions were each sequentially subjected to CID, HCD, and ETD. Dynamic exclusion was enabled with a

mass width of ±10 ppm, a repeat count of 1, and an exclusion duration of 12 s. Charge-state screening was enabled along with monoisotopic precursor selection and non-peptide monoisotopic recognition to prevent triggering of MS/MS on precursor ions with unassigned charge or a charge state of 1. For CID, the normalized collision energy was set to 30 with an activation Q of 0.25 and activation time of 10 ms. For HCD, the normalized collision energy was set to 35. For ETD, fluoranthene was used as the ETD reagent with an anion AGC target of 500,000 ions, supplemental activation was enabled, and the reaction time was dependent on the precursor charge state (precursor charge state/reaction time in ms: +2/100, +3/66.7, +4/50, +5/40, +6/33.3, etc). All MS/MS spectra were collected with an AGC target ion setting of 100,000 ions. The instrument control software does not currently allow for separate AGC targets for each dissociation mode. Optimal AGC targets would be closer to 30,000 ions for CID and HCD, and 200,000 ions for ETD.

For the UHPLC LC-MS/MS method, aliquots of each tryptic digest (100 ng) were separated using an EASY-nLC 1000 UHPLC system (Proxeon, Thermo Fisher Scientific). The LC system was connected to a custom-fit cross (360 μm, IDEX Health & Science, UH-906x), and columns were heated to 50˚C using column heater sleeves (Phoenix-ST) to prevent overpressuring of columns during UHPLC separation. Peptides were eluted from a 15 cm x 75 μm diameter capillary column (Picofrit 10 μm tip opening, New Objective, PF360-75-10-N-5) packed in-house with reversed phase resin (1.9 μm ReproSil-PurC18-AQ medium, Dr. Maisch GmbH, r119.aq). The mobile phase flow rate was 200 nL/min and comprised of 3% ACN/0.1% FA (Solvent A) and 90% ACN/0.1% FA (Solvent B). For *C. westwoodi*, *E. rubripes*, and *S. morsitans* venoms a 124-min LC-MS/MS method consisted of a 10-min column-equilibration procedure and a 9-min sample-loading procedure for a 2 μL injection. The elution portion of the LC gradient was 0–5% solvent B in 2 min, 5–35% in 90 min, 35–59% in 12 min, 59–90% in 2 min, and held at 90% solvent B for 10 min to yield ~20 s peak widths. For *T. longicornis* venom a 94 min gradient was used with the 5–35% solvent B portion shortened to 60 min. Data-dependent LC-MS/MS spectra were acquired in cycles of ~3.5 s (124-min gradient) or 3 s (94-min gradient), with each cycle was of the following form: one full orbitrap MS scan at 60,000 resolution followed by either 12 (124-min gradient) or 10 (94-min

gradient) HCD MS/MS scans in the orbitrap at 15,000 resolution using an isolation width of 2.5 m/z. Dynamic exclusion was enabled with a mass width of ±20 ppm, a repeat count of 1, and an exclusion duration of 30 s (124-min gradient) or 10 s (94-min gradient). Charge state screening was enabled along with monoisotopic precursor selection and non-peptide monoisotopic recognition to prevent triggering of MS/MS on precursor ions with unassigned charge or a charge state of 1. For HCD MS/MS scans the normalized collision energy was 29, AGC target 50,000 ions, and maximum ion time 200 ms. The mass spectra and the 6-frame translation FASTA file used for searching the spectra associated with this paper may be downloaded from MassIVE (http://massive.ucsd.edu) using the identifier MSV000078573.

All MS data were interpreted using the Spectrum Mill software package v4.1 beta (Agilent Technologies, Santa Clara, CA). Similar MS/MS spectra acquired on the same precursor m/z within ± 30 s were merged. MS/MS spectra with precursor charge >11 and poor quality MS/MS spectra, which failed the quality filter by not having a sequence tag length > 0 (i.e., minimum of two masses separated by the in-chain mass of an amino acid) were excluded from searching. MS/MS spectra were searched against the transcriptomic sequence database described above. Searches were conducted separately for spectra from each dissociation method. Scoring parameters were ESI Orbitrap (CID), ESI Orbitrap-ETD, or ESI Orbitrap-HCD respectively. CID and HCD spectra were allowed ±20 ppm product mass tolerance and 40% minimum matched peak intensity. ETD spectra were allowed ±100 ppm product mass tolerance (due to an apparent instrument acquisition software bug affecting mass calibration of high mass ETD product ions) and 30% minimum matched peak intensity. All spectra were allowed ±20 ppm precursor mass tolerance. Ethanolyl cysteine was a fixed modification. Allowed variable modifications were oxidized methionine, deamidation of asparagine, and pyroglutamic acid modification at N-terminal glutamine with a precursor MH+ shift range of –18 to 64 Da. Tryptic digests were searched with trypsin enzyme specificity with a maximum of three missed cleavages. Glu-C digests were searched with Glu-C with cleavage specificity of post D or E with a maximum of six missed cleavages.

Identities interpreted for individual spectra were automatically designated as confidently assigned using the Spectrum Mill autovalidation module to apply target-decoy based

false-discovery rate (FDR) scoring threshold criteria via a two-step auto threshold strategy at the spectral and protein levels. First, peptide mode was set to allow automatic variable range precursor mass filtering with score thresholds optimized to yield a spectral level FDR of <1.2% for each dissociation method and each precursor charge state in each LC-MS/MS run. Second, protein mode was applied to further filter all the peptide-level validated spectra combined from both LC-MS/MS runs (trypsin and Glu-C digests) derived from a single venom sample using a minimum protein score of 20 and a maximum protein-level FDR of zero. Since the maximum peptide score is 25, the protein level step filters the results so that each identified protein is comprised of multiple peptides unless a single excellent scoring peptide was the sole match. The above criteria yielded false discovery rates of <1.0% for each sample at the peptide-spectrum match level and <1.2 % at the distinct peptide level as estimated by target-decoy-based searches using reversed sequences.

Peptide Classification and Phylogenetic Analysis

Phylogenetic analyses were performed to allow reconstruction of the molecular evolutionary history of each toxin type for which transcripts were bioinformatically discovered. Toxin sequences were identified primarily by identification from proteomic data, but also by manually comparing the translated DNA sequences with previously characterised toxins using a local, low stringency (E value ≤ 10) BLAST search (Altschul et al. 1990) of all previously sequenced centipede toxins in CLC Main Workbench v6.6. For novel toxins, only contigs with more than five reads encoding complete coding sequences with no ambiguous residues were included. Toxins were grouped based on (i) sequence homology, (ii) cysteine pattern similarity, (iii) presence of conserved domains or (iv) common BLAST hit, following an iterative approach of aligning similar sequences, attempted joining of alignments, phylogenetic analysis, and rejection or accepting joined alignments. Sequences were aligned using MAFFT v7 (http://mafft.cbrc.jp/alignment/server/) and corrected manually using CLC Main Workbench v6.6. Alignments are included in Supplementary file 3.

Molecular phylogenetic analysis of toxin transcripts was conducted using their translated

amino acid sequences. Where available and appropriate, comparative sequences from other centipedes, venomous animals, and physiological gene homologs identified from non-venom-gland transcriptomes were included in each dataset as outgroup sequences. When presented as sequence alignments, the signal peptide sequence is shown in lowercase, cysteines are highlighted in black, while > and < indicate incomplete N/5' or C/3' ends, respectively. Datasets were analyzed using Bayesian inference implemented on MrBayes, version 3.2 (Ronquist et al. 2012), using lset rates=gamma with prset aamodelpr=mixed command, which enables the program to optimize between nine different amino acid substitution matrices implemented in MrBayes. The log-likelihood score of each saved tree was plotted against the number of generations to establish the point at which the log likelihood scores reached their asymptote, and the posterior probabilities for clades was established by constructing a majority-rule consensus tree for all trees generated after completed burn-in phase.

Acknowledgements

This study was supported by the Australian Research Council (Discovery Grants DP1095728 and DP130103813 to GFK), The University of Queensland (International Postgraduate Research Scholarship, UQ Centennial Scholarship, and UQ Advantage Top-Up Scholarship to EABU) and the Norwegian State Education Loans Fund (EABU). This work was also supported in part by the Broad Institute of MIT and Harvard, and by grants to Steven A. Carr from the US National Cancer Institute (U24CA160034, part of the Clinical Proteomics Tumor Analysis Consortium initiative) and the National Heart, Lung, and Blood Institute (HHSN268201000033C and R01HL096738).

TABLES

Table 1 HMW venom proteins in centipede venoms. Summary of HMW proteins identified from centipedes in this study and the type of evidence used for identification: Transcriptome (T), AB-SCIEX 5600 MS (A), Thermo Fisher Orbitrap MS (O), and in-gel digestion (G).

HMW protein family	*S. alternans*	*S. morsitans*	*C. westwoodi*	*E. rubripes*	*T. longicornis*	**Function**
M12A	T	O, T			A, G, O, T	Metalloprotease MEROPS family M12A
Putative M12A	T	A, G, O, T	A, O, T	A, G, O, T		Putative metalloprotease, potentially member of M12A subfamily
Protease S1	T	A, O, T	T	T	T	Serine endoprotease activity, potentially involved in activation of toxins
Protease S8	T	T	O, T	G, T		Serine endoprotease activity, potentially involved in activation of toxins
γ-Glutamyl Transferase	T	A, G, O, T	A, O, T	T	T	Platelet aggregating activity, hemolytic to mouse and rabbit hemocytes
Chitinase			A, O, T	T		Unknown
Lysozyme C		A, O, T		T		Potential antimicrobial component
Hyaluronidase		A, O, T		A, G, O, T		Degrades glycosaminoglycans, potentially facilitating the spread of venom components
Glucose Dehydrogenase	T	A, G, O, T	A, O, T	A, G, O, T	T	Unknown
Carboxyl-Esterase	T		A, O, T	T	T	Unknown
Peptidyl Arginine Deiminase					G, O, T	Venom activity unknown; catalyses deamination of the guanidine group of arginine residues
ScolPLA$_2$		A, O, T	T	A		Venom activity unknown; venom PLA2 can be myotoxic, inflammatory, and neurotoxic
β-Pore-forming Toxin	T	A, G, O, T	A, O, T	A, G, O, T	A, G, O, T	Potential cytotoxin; formation of polymeric pore structures in cell membranes
CAP Type 1				A, G, O, T	O, T	Unknown
CAP Type 2	T	A, G, O, T	A, O, T	A, G, O, T		Ca_V channel inhibitor Group 1 (KC144967); Trypsin inhibitor (KC144061)
CAP Type 3		A, O, T				Unknown
LDLA-repeat	T	A, G, O, T	A, O, T	A, G, O, T	O, T	Unknown
Cystatin				O, T		Potential protease inhibitor
Transferrin	T	O, T	T	G, O, T	T	Potential antimicrobial component
DUF3472	T	A, G, O, T	A, O, T	A, G, O, T		Unknown
DUF1397					A, O, T	Unknown

Table 2 LMW peptides identified from centipede-venoms. Sumamry of LMW venom peptides identified in this study as well as in *S. subspinipes* (Liu et al. 2012; Yang et al. 2012), listed with the type of evidence used for identification: Transcriptome (T), AB-SCIEX 5600 MS (A), Thermo Fisher Orbitrap MS (O), in-gel digestion (G), or Edman degradation (E). Abbreviations used: AVIT, AVIT, Colipase and Prokineticin superfamily; CB_2, Chitin Binding Domain; ICK, inhibitor cystine knot superfamily; VWC, Single Von Willebrand factor type C domain protein.

Toxin family	*S. alternans*	*S. subspinipes*	*S. morsitans*	*C. westwoodi*	*E. rubripes*	*T. longicornis*	**Cysteines/Function**
SCUTX 1						A, O, T	2 Cys
SCUTX 2						A, O, T	8 Cys
SLPTX 1 – CB_2			A, O, T	A, O, T	A, O, T	A, O, T	6 Cys
SLPTX 2 – Defensin					A, T		6 Cys
SLPTX 3	T	E, T	A, T				6 Cys; κ-SLPTX$_3$-Ssm1a
SLPTX 4	T	E, T	A, O, T		A, O, T		4 Cys; K_V inhibitor Group 8 (KC144226)
SLPTX 5	T	E, T	T	T	T		7 Cys; ϖ-SLPTX$_5$-Ssm1a
SLPTX 6			A, O, T				4 Cys
SLPTX 7 – ICK		E, T					6 Cys; κ-SLPTX$_7$-Ssm2a
SLPTX 8	T	T	O, T	A, T	A, T		6 Cys
SLPTX 9	T	E, T	A, O, T		A, O, T		6 Cys; Predicted neurotoxin 5
SLPTX 10	T	E, T	A, O, T	A, O, T	A, G, O, T		6 Cys; K_V inhibitor Group 9 (KC144849); Ca_V inhibitor Goup 4 (KC144448)
SLPTX 11	T	E, T	A, G, O, T				6–18 Cys; K_V inhibitor Groups 1-3 (KC144287, KC144104, KC144040); Anticoagulant (KC144430); κ-SLPTX$_{11}$-Ssm3a
SLPTX 12	T	T	O, T		A, O, T		7 Cys
SLPTX 13	T	E, T	A, O, T	A, T	A, O, T		8 Cys; ω-SLPTX$_{13}$-Ssm2a; Ca_V inhibitor Group 2
SLPTX 14	T	T	O, A, T		T		8 Cys
SLPTX 5	T	E, T	A, O, T	A, O, T	A, O, T		8 Cys; K_V inhibitor Group 10 (KC144556; Na_V inhibitor Group (KC144793); Ca_V inhibitor Group 3 (KC145039); Predicted neurotoxin 4
SLPTX 16 – VWC	T	E, T	A, O, T	O, T	A, O, T		8 Cys
SLPTX 17					O, T		8 Cys
SLPTX 18 – AVIT					A, T		10 Cys; putative colipase. Same superfamily as AVIT-toxins which induce smooth muscle contraction and hyperalgesia
SLPTX 19			A, T			T	12 Cys; putative carboxypeptidase inhibitor
SLPTX 20			A, O, T	A, O, T			6 Cys
SCUTX 3						A, T	Linear; proline rich
SLPTX 21			A, O, T				Linear; diuretic hormone
SLPTX 22				A, T	T		Linear; hypertrehalosaemic hormone
SLPTX 23					A, O, T		Linear; Er 6
SLPTX 24					A, T		Linear; Er 8

Table 3 Proteins and peptides in centipede venoms convergently recruited in other venomous animals.*

	M12	PS1	GGT	PLA_2	Chi	Lys	Hyal	CO-E	β-PF	CAP	Tran	Cys	ICK	Def	VWC	AVIT
Cephalopod	X[1]	X		X	X		X	X		X						
Cnidarian	X[1]	X		X	X		X		X	X			X			
Cone snail				X			X			X						
Fish							X									
Insect (bristle)		X		X								X				
Insect (proboscis)		X		X			X			X		X			X	
Insect (stinger)	X[1]	X	X	X	X		X			X						
Platypus	X[1]	X					X			X				X		
Scorpion		X		X			X			X			X	X	X	
Shrew		X														
Spider	X				X		X	X		X			X			X
Reptile	X[1]	X		X(3)			X	X		X		X(2)		X		X
Tick	X[1]	X		X	X				X	X		X	X			X
Vampire Bat		X				X				X	X					

*Abbreviations used: AVIT, AVIT, Colipase and Prokineticin superfamily; β-PF, β-Pore-forming Toxins; CAP, CRiSP (cysteine rich proteins), Allergen (Ag-5), and Pathogenesis-related (PR-1) proteins; Chi, Chitinase; CO-E, Carboxyl-Esterase; Cys, Cystatin; Def, Defensin; GGT, γ-Glutamyl Transpeptidase; Hyal, Hyaluronidase; ICK, inhibitory cystine knot structural superfamily; Lys, Lysozyme; M12, Metalloprotase family M12; PS1, Serine protease S1; PLA_2, Phospholipase Type A2; Tran, Transferrin; VWC, Single Von Willebrand factor type C domain protein. [1] M12 protease subfamily B

FIGURE LEGENDS

Figure 1: 2D-PAGE gel of venom from three species of centipede. Venoms analysed were from (A) *E. rubripes*, (B) *S. morsitans*, and (C) *T. longicornis*. Proteins identified by in-gel digestion and LC-MS/MS are annotated according to their respective families. The isoelectric point (pI) is indicated above each gel and the molecular mass (kDa) is shown on the right of each gel. Abbreviations: β-pore-forming toxins (β-PFT); cysteine-rich, allergen, and pathogenesis-related (CAP); γ-glutamyl transpeptidases (GGT); glucose dehydrogenase (GDH); hyaluronidase (Hyal); LDLA-domain containing proteins (LDLA); metalloprotease M12A (M12A); peptidyl arginine-deiminase (PAD); putative metalloprotease M12A (pM12A).

Figure 2: Phylogenetic reconstruction of members of the centipede β-pore-forming toxin family. Mid-point rooted phylogenetic tree as estimated by MrBayes showing the reciprocal monophyly of Scutigeromorpha and Scolopendromorpha sequences and the extensive radiation of β-pore-forming toxins that has occurred in both clades. Phylogeny was inferred under the WAG model (posterior probability 1) with rates set to gamma. Posterior probabilities are displayed at each node as a fraction.

Figure 3: Phylogenetic reconstruction of members of the CAP protein superfamily. Mid-point rooted phylogenetic tree as estimated by MrBayes showing the distribution of centipede venom CAP protein types 1, 2, and 3, as well as characterized activity. Phylogeny was inferred under the WAG model (posterior probability 1) with rates set to gamma. Posterior probabilities are displayed at each node.

Figure 4: Phylogenetic reconstruction of the LDLA toxin family. Phylogenetic tree as estimated by MrBayes showing the interleaved distribution of LDLA-proteins from Scolopendridae. Phylogeny was inferred under the Blosum model (posterior probability 1) with rates set to gamma. Posterior probabilities are displayed at each node.

Figure 5: Alignment and phylogenetic reconstruction of the centipede CB_2 Domain family. *Upper panel:* Alignment of epidermal CB_2 domain (CBD) containing proteins from centipede venom and epidermis as well as butterfly (*Danaus plexippus*). CBD are shaded in grey and cysteine residues highlighted in black. *Lower panel:* Phylogenetic reconstruction of the same alignment as estimated by MrBayes is shown on the lower right. Phylogeny was inferred under the WAG model (posterior probability 1) with rates set to gamma. Posterior probabilities are displayed at each node. Numbers correspond to the following sequences: 1. *T. longicornis* GASR01000113 2. *T. longicornis* GASR01000112 3. *T. longicornis* GASR01000111 4. *T. longicornis* GASR01000114 5. *Ethmostigmus rubripes* GASI01000095 6. *Ethmostigmus rubripes* GASI01000096 7. *Ethmostigmus rubripes* GASI01000092 8. *Ethmostigmus rubripes* GASI01000093 9. *Ethmostigmus rubripes* GASI01000094 10.*Cormocephalus westwoodi* GASL01000036 11. *E. rubripes* epidermis GASJ01000006 12. *E. rubripes* epidermis GASJ01000008 13. *E. rubripes* epidermis GASJ01000007 14.*Danaus plexippus* (Butterfly) UniProt: G6D9I4.

Figure 6: Phylogenetic reconstruction of scoloptoxin family 4. Mid-point rooted phylogenetic tree as estimated by MrBayes, showing the non-monophyletic distribution of family members and characterized pharmacological activity. Phylogeny was inferred under the Jones model (posterior probability 0.934) with rates set to gamma. Posterior probabilities are displayed at each node.

Figure 7: Phylogenetic reconstruction of scoloptoxin family 5. Mid-point rooted phylogenetic tree as estimated by MrBayes, showing the non-monophyletic distribution of family members and characterized pharmacological activity. Phylogeny was inferred under the Jones model (posterior probability 0.99) with rates set to gamma. Posterior probabilities are displayed at each node.

Figure 8: Phylogenetic reconstruction of scoloptoxin family 16. Mid-point rooted phylogenetic tree as estimated by MrBayes, showing the non-monophyletic distribution

of family members, schematic cysteine pattern (with additional cysteine in red), and the convergent recruitment of single Von Willebrand factor type C (VWC) domain protein with scorpions. Phylogeny was inferred under the Blosum model (posterior probability 1) with rates set to gamma. Posterior probabilities are displayed at each node.

Figure 9: Phylogenetic reconstruction of scoloptoxin family 15. Mid-point rooted phylogenetic tree as estimated by MrBayes, showing the phylogenetic distribution of characterized pharmacological activities. No activity was found for "Predicted Neurotoxin 4" against Na_V, Ca_V, and K_V channels in rat dorsal root ganglion neurons by Yang et al. (2012). Phylogeny was inferred under the Blosum model (posterior probability 0.95) with rates set to gamma. Posterior probabilities are displayed at each node.

Figure 10: Phylogenetic reconstruction of scoloptoxin family 11. Mid-point rooted phylogenetic tree as estimated by MrBayes, showing the phylogenetic distribution of characterized pharmacological activities and number of cysteines present in the mature toxin. Phylogeny was inferred under the WAG model (posterior probability 1) with rates set to gamma. Posterior probabilities are displayed at each node.

Figure 11: Phylogenetic distribution of centipede toxin families. A representative phylogenetic tree of the species included in this study, showing the phylogenetic distribution of toxin types as their earliest respective recruitments. Abbreviations used: β-PF, centipede β-pore forming toxins; CAP, CRiSP (cysteine rich proteins), allergen (Ag-5), and pathogenesis-related (PR-1) proteins; CO-esterase B, carboxylesterase type B; DUF1397, domain of unknown function 1397; DUF 3472, domain of unknown function 3472; GGT, γ-glutamyl transpeptidase; GDH, glucose dehydrogenase; M12A, metalloprotease M12A; PAD, centipede peptidyl arginine-deiminase; PS1, serine protease S1; PS8, subtilisin-like serine protease S8; SCUTX, scutigerotoxin family; SLPTX, scoloptoxin family.

SUPPLEMENTARY FILES

Supplementary file 1 – Supplementary table S1: Summary of completely novel HMW venom proteins in centipede venoms. Table shows HMW proteins identified from centipedes in this study and the type of evidence used for identification: Transcriptome (T), AB-Sciex 5600 MS (A), Thermo Fisher Orbitrap MS (O), 2D-PAGE and in-gel digestion (G).

Supplementary file 2: Proteomic evidence for protein and peptide identifications of sequences from centipede venoms submitted to GenBank. The full search summary of the Thermo Fisher Orbitrap data is presented, with all curated and submitted sequences ordered first. Names and annotations correspond to those in the sequence database available at MassIVE (http://massive.ucsd.edu).

Supplementary file 3: Alignments used for phylogenetic analysis of centipede venom protein and peptide families.

REFERENCES

Abrahamson M, Alvarez-Fernandez M, Nathanson C-M. 2003. Cystatins. Biochem Soc Symp: 179-199.

Aird SD. 2002. Ophidian envenomation strategies and the role of purines. Toxicon 40: 335-393.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215: 403-410.

Asgari S, Zhang G, Zareie R, Schmidt O. 2003. A serine proteinase homolog venom protein from an endoparasitoid wasp inhibits melanization of the host hemolymph. Insect Biochem Mol Biol 33: 1017-1024.

Bak TG. 1967. Studies on glucose dehydrogenase of *Aspergillus oryzae*. II. Purification and physical and chemical properties. Biochim Biophys Acta 139: 277-293.

Balasubramanian PG, Beckmann A, Warnken U, Schnolzer M, Schuler A, Bornberg-Bauer E, Holstein TW, Ozbek S. 2012. Proteome of *Hydra* nematocyst. J Biol Chem 287: 9672-9681.

Balit CR, Harvey MS, Waldock JM, Isbister GK. 2004. Prospective study of centipede bites in Australia. Clin Toxicol 42: 41-48.

Barrett AJ, Rawlings ND, Woessner JF. 2004. Handbook of Proteolytic Enzymes. London Elsevier.

Bonato L, Edgecombe G, Lewis J, Minelli A, Pereira L, Shelley R, Zapparoli M. 2010. A common terminology for the external anatomy of centipedes (Chilopoda). ZooKeys 69: 17-51.

Bond JS, Beynon RJ. 1995. The astacin family of metalloendopeptidases. Protein Sci 4: 1247-1261.

Bork P, Beckmann G. 1993. The CUB domain: A widespread module in developmentally regulated proteins. J Mol Biol 231: 539-545.

Brust A, Sunagar K, Undheim EA, et al. 2013. Differential evolution and neofunctionalization of snake venom metalloprotease domains. Mol Cell Proteomics 12: 651-663.

Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG. 2013. Complex cocktails: The evolutionary novelty of venoms. Trends Ecol Evol 28: 219-229.

Chen J, Zhao L, Jiang L, Meng E, Zhang Y, Xiong X, Liang S. 2008. Transcriptome analysis revealed novel possible venom components and cellular processes of the tarantula *Chilobrachys jingzhao* venom gland. Toxicon 52: 794-806.

Chen Y-H, Jia X-T, Zhao L, Li C-Z, Zhang S, Chen Y-G, Weng S-P, He J-G. 2011. Identification and functional characterization of Dicer2 and five single VWC domain proteins of *Litopenaeus vannamei*. Dev Comp Immunol 35: 661-671.

Chung LP, Keshav S, Gordon S. 1988. Cloning the human lysozyme cDNA: Inverted Alu repeat in the mRNA and in situ hybridization for macrophages and Paneth cells. Proc Natl Acad Sci USA 85: 6227-6231.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674-3676.

Courtay C, Oster T, Michelet F, Visvikis A, Diederich M, Wellman M, Siest G. 1992. Gamma-glutamyltransferase: Nucleotide sequence of the human pancreatic cDNA. Evidence for a ubiquitous γ-glutamyltransferase polypeptide in human tissues. Biochem Pharmacol 43: 2527-2533.

da Silveira RB, Wille AC, Chaim OM, et al. 2007. Identification, cloning, expression and functional characterization of an astacin-like metalloprotease toxin from *Loxosceles intermedia* (brown spider) venom. Biochem J 406: 355-363.

Daly NL, Scanlon MJ, Djordjevic JT, Kroon PA, Smith R. 1995. Three-dimensional structure of a cysteine-rich repeat from the low-density lipoprotein receptor. Proc Natl Acad Sci USA 92: 6334-6338.

De Graaf DC, Aerts M, Brunain M, Desjardins CA, Jacobs FJ, Werren JH, Devreese B. 2010. Insights into the venom composition of the ectoparasitoid wasp *Nasonia vitripennis* from bioinformatic and proteomic studies. Insect Mol Biol 19: 11-26.

Dhananjaya BL, D'Souza CJM. 2010. The pharmacological role of nucleotidases in snake venoms. Cell Biochem Funct 28: 171-177.

Dugon MM, Arthur W. 2012. Comparative studies on the structure and development of the venom-delivery system of centipedes, and a hypothesis on the origin of this evolutionary novelty. Evol Dev 14: 128-137.

Dugon MM, Black A, Arthur W. 2012. Variation and specialisation of the forcipular apparatus of centipedes (Arthropoda: Chilopoda): a comparative morphometric and microscopic investigation of an evolutionary novelty. Arthropod Struct Dev 41: 231-243.

Escoubas P, Sollod B, King GF. 2006. Venom landscapes: Mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach. Toxicon 47: 650-663.

Falabella P, Riviello L, Caccialupi P, et al. 2007. A γ-glutamyl transpeptidase of *Aphidius ervi* venom induces apoptosis in the ovaries of host aphids. Insect Biochem Mol Biol 37: 453-465.

Fernandes-Pedrosa Mde F, Junqueira-de-Azevedo Ide L, Goncalves-de-Andrade RM, Kobashi LS, Almeida DD, Ho PL, Tambourgi DV. 2008. Transcriptome analysis of *Loxosceles laeta* (Araneae, Sicariidae) spider venomous gland using expressed sequence tags. BMC Genomics 9: 279.

Francischetti IM, Mather TN, Ribeiro JM. 2003. Cloning of a salivary gland metalloprotease and characterization of gelatinase and fibrin(ogen)lytic activities

in the saliva of the Lyme disease tick vector *Ixodes scapularis*. Biochem Biophys Res Commun 305: 869-875.

Fry BG. 2005. From genome to "venome": Molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Res 15: 403-420.

Fry BG, Roelants K, Champagne DE, et al. 2009a. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. Annu Rev Genomics Hum Genet 10: 483-511.

Fry BG, Roelants K, Norman J. 2009b. Tentacles of venom: toxic protein convergence in the kingdom Animalia. J Mol Evol 68: 311-321.

Fry BG, Undheim EAB, Ali SA, et al. 2013. Squeezers and leaf-cutters: differential diversification and degeneration of the venom system in toxicoferan reptiles. Mol Cell Proteomics

Girish KS, Shashidharamurthy R, Nagaraju S, Gowda TV, Kemparaju K. 2004. Isolation and characterization of hyaluronidase a "spreading factor" from Indian cobra (*Naja naja*) venom. Biochimie 86: 193-202.

Gomes A, Datta A, Sarangi B, Kar PK, Lahiri SC. 1983. Isolation, purification and pharmacodynamics of a toxin from the venom of the centipede *Scolopendra subspinipes dehaani* Brandt. Indian J Exp Biol 21: 203-207.

González-Morales L, Diego-GarcÌa E, Segovia L, Gutiérrez MdC, Possani LD. 2009. Venom from the centipede *Scolopendra viridis* Say: purification, gene cloning and phylogenetic analysis of a phospholipase A2. Toxicon 54: 8-15.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36: 3420-3435.

Gracy J, Le-Nguyen D, Gelly JC, Kaas Q, Heitz A, and Chiche L. 2008. KNOTTIN: the knottin or inhibitor cystine knot scaffold in 2007. Nucl Acids Res 36: D314-D319.

Gutiérrez JM, Rucavado A. 2000. Snake venom metalloproteinases: their role in the pathogenesis of local tissue damage. Biochimie 82: 841-850.

Hale JE, Butler JP, Gelfanova V, You JS, Knierman MD. 2004. A simplified procedure for the reduction and alkylation of cysteine residues in proteins prior to proteolytic digestion and mass spectral analysis. Anal Biochem 333: 174-181.

Heitz JR, Norment BR. 1974. Characteristics of an alkaline phosphatase activity in brown recluse venom. Toxicon 12: 181-187.

Hilken G, Rosenberg J, Brockmann C. 2005. Ultrastructure of the epidermal maxilla II-gland of *Scutigera coleoptrata* (Chilopoda, Notostigmophora) and the ground pattern of epidermal gland organs in Myriapoda. J Morphol 264: 53-61.

Hoffman D. 2006. Hymenoptera venom allergens. Clinic Rev Allerg Immunol 30: 109-128.

Joshi J, Karanth KP. 2011. Cretaceous–Tertiary diversification among select Scolopendrid centipedes of South India. Mol Phylogenet Evol 60: 287-294.

Kass L, Batsakis JG. 1979. Cytochemistry of esterases. Crit Rev Clin Lab Sci 10: 205-223.

King GF. 2011. Venoms as a platform for human drugs: translating toxins into therapeutics. Expert Opin Biol Ther 11: 1469-1484.

King GF, Gentz MC, Escoubas P, Nicholson GM. 2008. A rational nomenclature for naming peptide toxins from spiders and other venomous animals. Toxicon 52: 264-276.

King GF, Hardy MC. 2013. Spider-venom peptides: Structure, pharmacology, and potential for control of insect pests. Annu Rev Entomol 58: 475-496.

King GF, Tedford HW, and Maggio F. 2002. Structure and function of insecticidal neurotoxins from Australian funnel-web spiders. J Toxicol-Toxin Rev 21: 359-389.

Knapp O, Stiles B, Popoff MR. 2010. The Aerolysin-like toxin family of cytolytic, pore-forming toxins. Open Toxinol J 3: 53-68.

Koch L. 1983a. Revision of the Australian centipedes of the genus *Cormocephalus* Newport (Chilopoda: Scolopendridae: Scolopendrinae). Aust J Zool 31: 799-833.

Koch L. 1983b. A taxonomic study of the centipede genus *Ethmostigmus* Pocock (Chilopoda: Scolopendridae: Otostigminae) in Australia. Aust J Zool 31: 835-849.

Koch LE. 1983c. Morphological characters of Australian scolopendrid centipedes, and the taxonomy and distribution of *Scolopendra morsitans* L. (Chilopoda: Scolopendridae: Scolopendrinae). Aust J Zool 31: 79-91.

Kramer KJ, Muthukrishnan S. 1998. Insect Chitinases: Molecular Biology and Potential Use as Biopesticides. Insect Biochem Mol Biol 27: 887-900.

Kuhn-Nentwig L, Schaller J, Nentwig W. 2004. Biochemistry, toxicology and ecology of the venom of the spider *Cupiennius salei* (Ctenidae). Toxicon 43: 543-553.

Lewis JGE. 1981. The biology of centipedes. Cambridge: Cambridge University Press.

Liu Z-C, Zhang R, Zhao F, et al. 2012. Venomic and transcriptomic analysis of centipede *Scolopendra subspinipes dehaani*. J Proteome Res 11: 6197-6212.

Long-Rowe KO, Burnett JW. 1994. Characteristics of hyaluronidase and hemolytic activity in fishing tentacle nematocyst venom of *Chrysaora quinquecirrha*. Toxicon 32: 165-174.

Low DHW, Sunagar K, Undheim EAB, et al. 2013. Dracula's children: The proteomics, transcriptomics and molecular evolution of vampire bat venom proteins. J Proteomics 89: 95-111.

Ma Y, He Y, Zhao R, Wu Y, Li W, Cao Z. 2012. Extreme diversity of scorpion venom peptides and proteins revealed by transcriptomic analysis: Implication for proteome evolution of scorpion venom arsenal. J Proteomics 75: 1563-1576.

Malta MB, Lira MS, Soares SL, Rocha GC, Knysak I, Martins R, Guizze SPG, Santoro ML, Barbaro KC. 2008. Toxic activities of Brazilian centipede venoms. Toxicon 52: 255-263.

McIntosh JM, Ghomashchi F, Gelb MH, Dooley DJ, Stoehr SJ, Giordani AB, Naisbitt SR, Olivera BM. 1995. Conodipine-M, a novel Phospholipase A_2 isolated from the venom of the marine snail *Conus magus*. J Biol Chem 270: 3518-3526.

McKeown KC. 1930. Centipedes and centipede bites. Aust Mus Mag 4: 59-60.

Mohamed AH, Abu-Sinna G, El-Shabaka HA, Abd El-Aal A. 1983. Proteins, lipids, lipoproteins and some enzyme characterizations of the venom extract from the centipede *Scolopendra morsitans*. Toxicon 21: 371-377.

Moran Y, Praher D, Schlesinger A, Ayalon A, Tal Y, Technau U. 2013. Analysis of soluble protein contents from the nematocysts of a model sea anemone sheds light on venom evolution. Mar Biotechnol 15: 329-339.

Morgenstern D, King GF. 2013. The venom optimization hypothesis revisited. Toxicon 63: 120-128.

Murienne J, Edgecombe GD, Giribet G. 2010. Including secondary structure, fossils and molecular dating in the centipede tree of life. Mol Phylogenet Evol 57: 301-313.

Nascimento DG, Rates B, Santos DM, Verano-Braga T, Barbosa-Silva A, Dutra AAA, Biondi I, Martin-Eauclaire MF, Lima MED, Pimenta AMC. 2006. Moving pieces in a taxonomic puzzle: venom 2D-LC/MS and data clustering analyses to infer phylogenetic relationships in some scorpions from the Buthidae family (Scorpiones). Toxicon 47: 628-639.

Oliveira JS, Fuentes-Silva D, King GF. 2012. Development of a rational nomenclature for naming peptide and protein toxins from sea anemones. Toxicon 60: 539-550.

Pallaghy PK, Nielsen KJ, Craik DJ, and Norton RS. 1994. A common structural motif incorporating a cystine knot and a triple-stranded β-sheet in toxic and inhibitory polypeptides. Protein Sci 3: 1833-1839.

Peichoto ME, Mackessy SP, Teibler P, Tavares FL, Burckhardt PL, Breno MC, Acosta O, Santoro ML. 2009. Purification and characterization of a cysteine-rich secretory protein from *Philodryas patagoniensis* snake venom. Comp Biochem Physiol C Toxicol Pharmacol 150: 79-84.

Peng K, Kong Y, Zhai L, Wu X, Jia P, Liu J, Yu H. 2009. Two novel antimicrobial peptides from centipede venoms. Toxicon 55: 274-279.

Pineda SS, Sollod BL, Wilson D, Darling A, Sunagar K, Undheim EAB, Kely L, Antunes A, Fry BG, King GF. 2014. Diversification of a single ancestral gene into a successful toxin superfamily in highly venomous Australian funnel-web spiders. BMC Genomics 15: 177.

Quistad GB, Dennis PA, Skinner WS. 1992. Insecticidal activity of spider (Araneae), centipede (Chilopoda), scorpion (Scorpionidae), and snake (Serpentes) venoms. J Econ Entomol 85: 33-39.

Ramirez-Gomez F, Ortiz-Pineda PA, Rojas-Cartagena C, Suarez-Castillo EC, Garcia-Arraras JE. 2008. Immune-related genes associated with intestinal tissue in the sea cucumber *Holothuria glaberrima*. Immunogenetics 60: 57-71.

Rates B, Bemquerer MP, Richardson M, Borges MH, Morales RAV, De Lima ME, Pimenta A. 2007. Venomic analyses of *Scolopendra viridicornis nigra* and *Scolopendra angulata* (Centipede, Scolopendromorpha): shedding light on venoms from a neglected group. Toxicon 49: 810-826.

Rodrigues MCA, Guimarães LHS, Liberato JL, Polizeli MdLTdM, dos Santos WF. 2006. Acid and alkaline phosphatase activities of a fraction isolated from *Parawixia bistriata* spider venom. Toxicon 47: 854-858.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61: 539-542.

Rosenberg J, Hilken G. 2006. Fine structural organization of the poison gland of *Lithobius forficatus* (Chilopoda, Lithobiomorpha). Nor J Entomol 53: 119-127.

Ruder T, Sunagar K, Undheim EA, Ali SA, Wai T-C, Low DH, Jackson TN, King GF, Antunes A, Fry BG. 2013. Molecular phylogeny and evolution of the proteins encoded by coleoid (Cuttlefish, Octopus, and Squid) posterior venom glands. J Mol Evol: 1-13.

Samaraweera P, Law JH. 1995. Isolation, cloning and deduced amino acid sequence of a novel glycoprotein from the haemolymph of the hawkmoth *Manduca sexta*. Insect Mol Biol 4: 7-13.

Saez NJ, Senff S, Jensen JE, Er SY, Herzig V, Rash LD, and King GF. 2010. Spider-venom peptides as therapeutics. Toxins 2: 2851–2871.

Sher D, Fishman Y, Zhang M, Lebendiker M, Gaathon A, Mancheño J-M, Zlotkin E. 2005. Hydralysins, a new category of β-pore-forming toxins in Cnidaria. J Biol Chem 280: 22847-22855.

Shirai H, Blundell TL, Mizuguchi K. 2001. A novel superfamily of enzymes that catalyze the modification of guanidino groups. Trends Biochem Sci 26: 465-468.

Siezen RJ, Leunissen JAM. 1997. Subtilases: the superfamily of subtilisin-like serine proteases. Protein Sci 6: 501-523.

Silva ÉCN, Camargos TS, Maranhão AQ, Silva-Pereira I, Silva LP, Possani LD, Schwartz EF. 2009. Cloning and characterization of cDNA sequences encoding for new venom peptides of the Brazilian scorpion *Opisthacanthus cayaporum*. Toxicon 54: 252-261.

Sollod BL, Wilson D, Zhaxybayeva O, Gogarten JP, Drinkwater R, King GF. 2005. Were arachnids the first to use combinatorial peptide libraries? Peptides 26: 131–139.

Stankiewicz M, Hamon A, Benkhalifa R, Kadziela W, Hue B, Lucas S, Mebs D, Pelhate M. 1999. Effects of a centipede venom fraction on insect nervous system, a native

Xenopus oocyte receptor and on an expressed *Drosophila muscarinic* receptor. Toxicon 37: 1431-1445.

Sulkowski E, Björk W, Laskowski M. 1963. A specific and nonspecific alkaline monophosphatase in the venom of *Bothrops atrox* and their occurrence in the purified venom phosphodiesterase. J Biol Chem 238: 2477-2486.

Trevisan-Silva D, Gremski LH, Chaim OM, da Silveira RB, Meissner GO, Mangili OC, Barbaro KC, Gremski W, Veiga SS, Senff-Ribeiro A. 2010. Astacin-like metalloproteases are a gene family of toxins present in the venom of different species of the brown spider (genus *Loxosceles*). Biochimie 92: 21-32.

Tu AT, Chua A. 1966. Acid and alkaline phosphomonoesterase activities in snake venoms. Comp Biochem Physiol 17: 297-307.

Undheim EAB, Georgieva DN, Thoen HH, Norman JA, Mork J, Betzel C, Fry BG. 2010. Venom on ice: first insights into Antarctic octopus venoms. Toxicon 56: 897-913.

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

Vassilevski A, Kozlov S, Grishin E. 2009. Molecular diversity of spider venom. Biochemistry Mosc 74: 1505-1534.

Vincent B, Kaeslin M, Roth T, et al. 2010. The venom composition of the parasitic wasp *Chelonus inanitus* resolved by combined expressed sequence tags analysis and proteomic approach. BMC Genomics 11: 693.

Violette A, Leonardi A, Piquemal D, et al. 2012. Recruitment of glycosyl hydrolase proteins in a cone snail venomous arsenal: Further insights into biomolecular features of *Conus* venoms. Mar Drugs 10: 258-280.

Weston AJ, Chung R, Dunlap WC, et al. 2013. Proteomic characterisation of toxins isolated from nematocysts of the South Atlantic jellyfish *Olindias sambaquiensis*. Toxicon 71: 11-17.

Wong ES, Morgenstern D, Mofiz E, et al. 2012a. Proteomics and deep sequencing comparison of seasonally active venom glands in the platypus reveals novel venom peptides and distinct expression profiles. Mol Cell Proteomics 11: 1354-1364.

Wong ES, Papenfuss AT, Whittington CM, Warren WC, Belov K. 2012b. A limited role for gene duplications in the evolution of platypus venom. Mol Biol Evol 29: 167-177.

Woodward SR, Cruz LJ, Olivera BM, Hillyard DR. 1990. Constant and hypervariable regions in conotoxin propeptides. EMBO J 9: 1015-1020.

Yang S, Liu Z, Xiao Y, Li Y, Rong M, Liang S, Zhang Z, Yu H, King GF, Lai R. 2012. Chemical punch packed in venoms makes centipedes excellent predators. Mol Cell Proteomics

Zdobnov EM, Apweiler R. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17: 847-848.

FIGURE 1

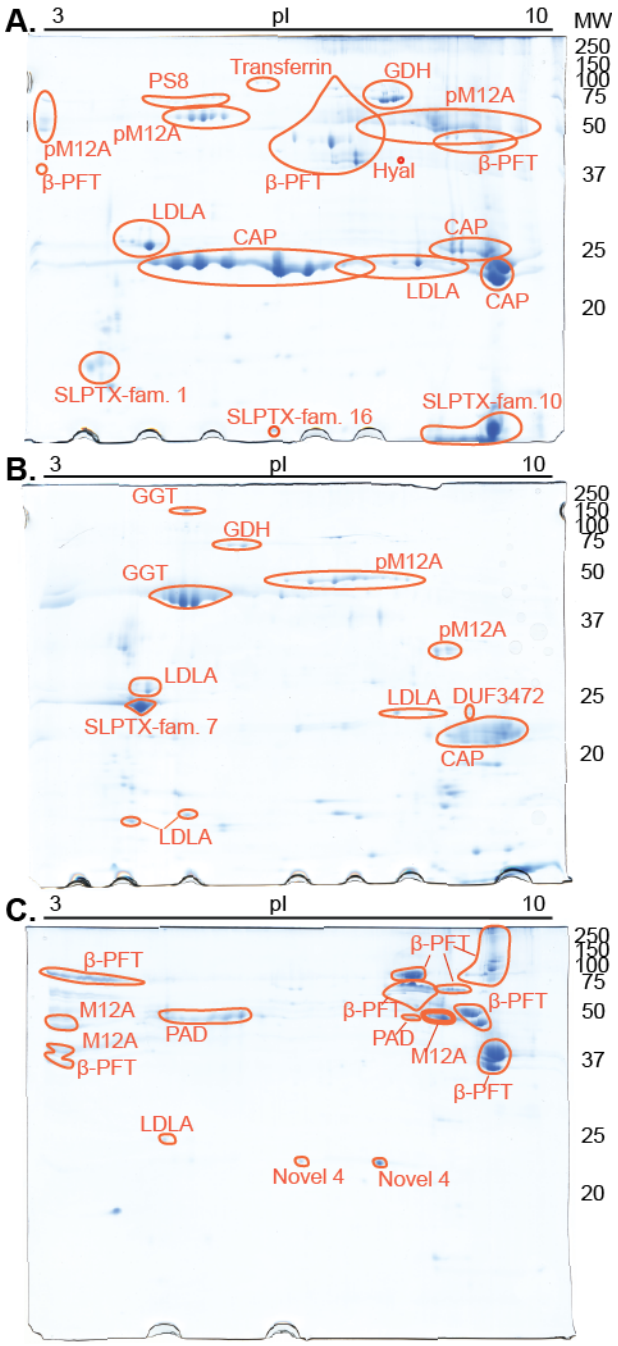

FIGURE 2

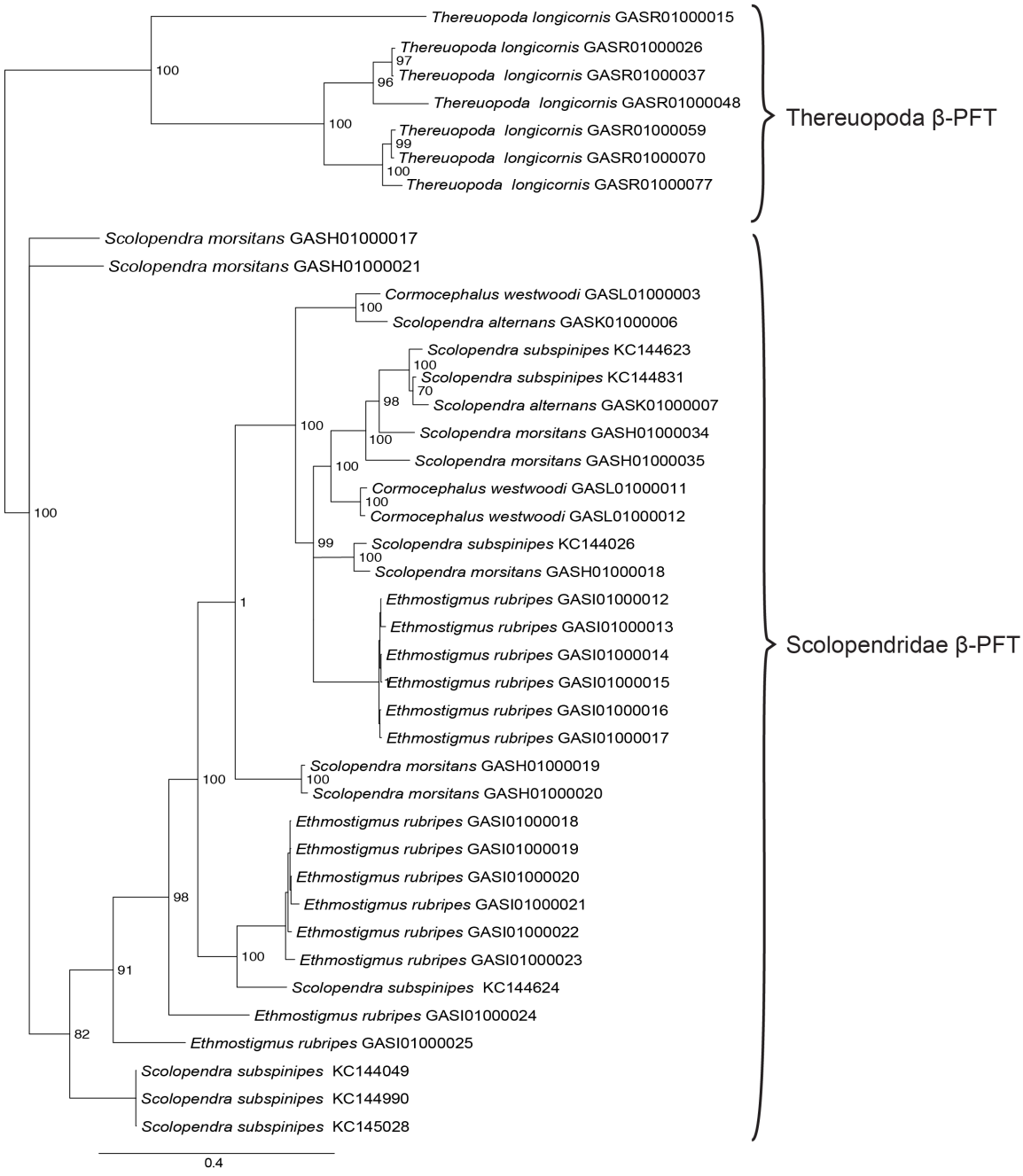

FIGURE 3

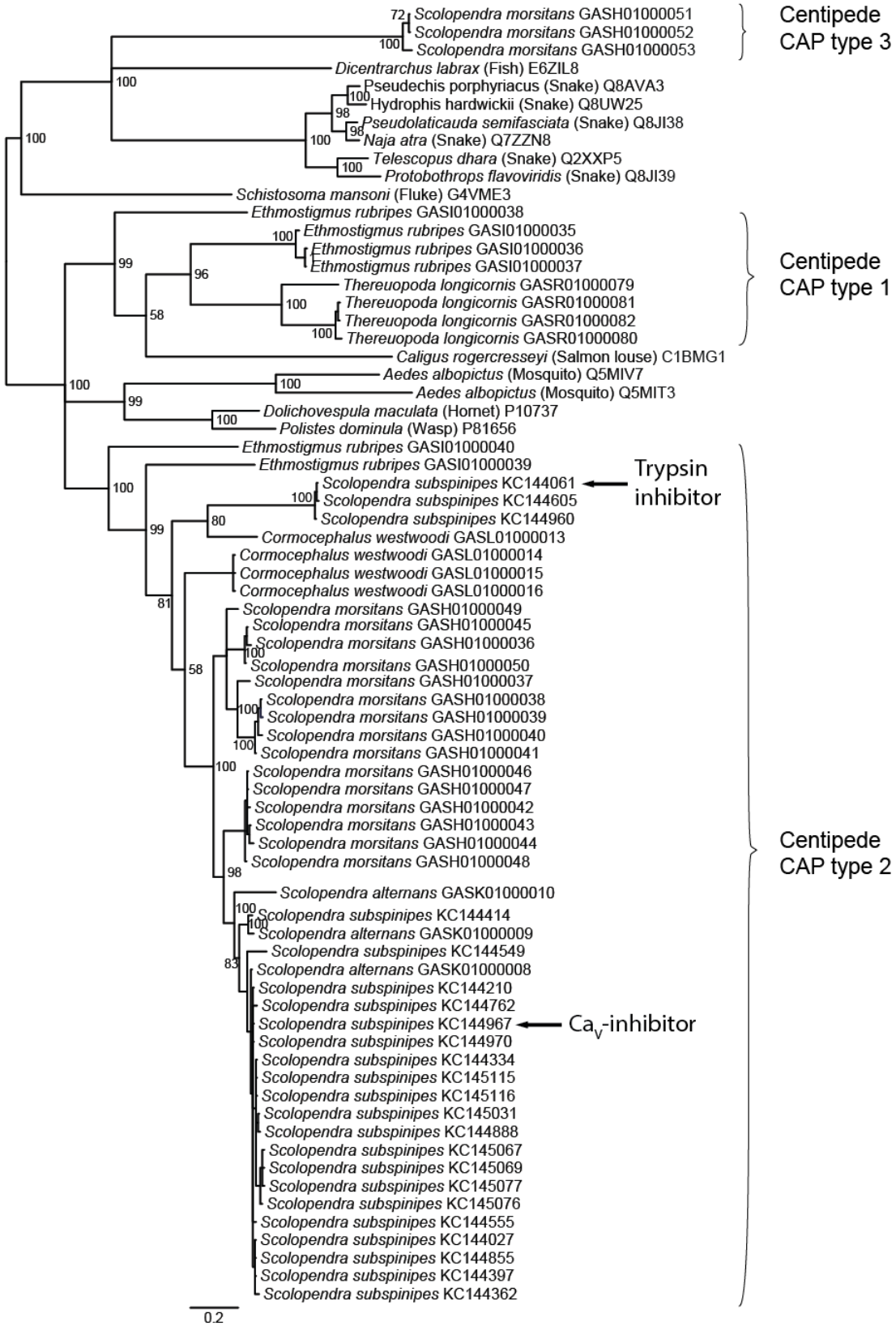

FIGURE 4

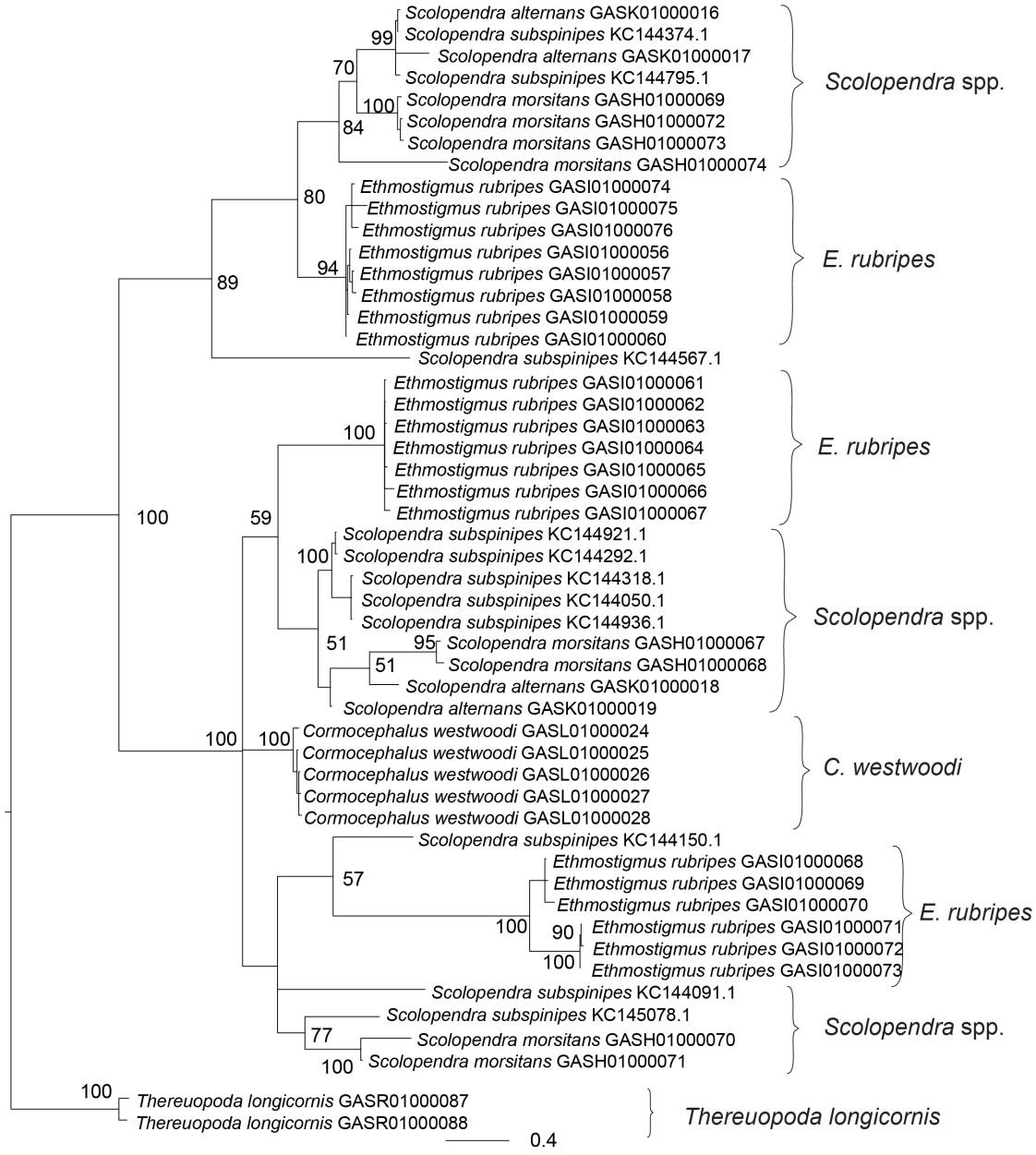

FIGURE 5

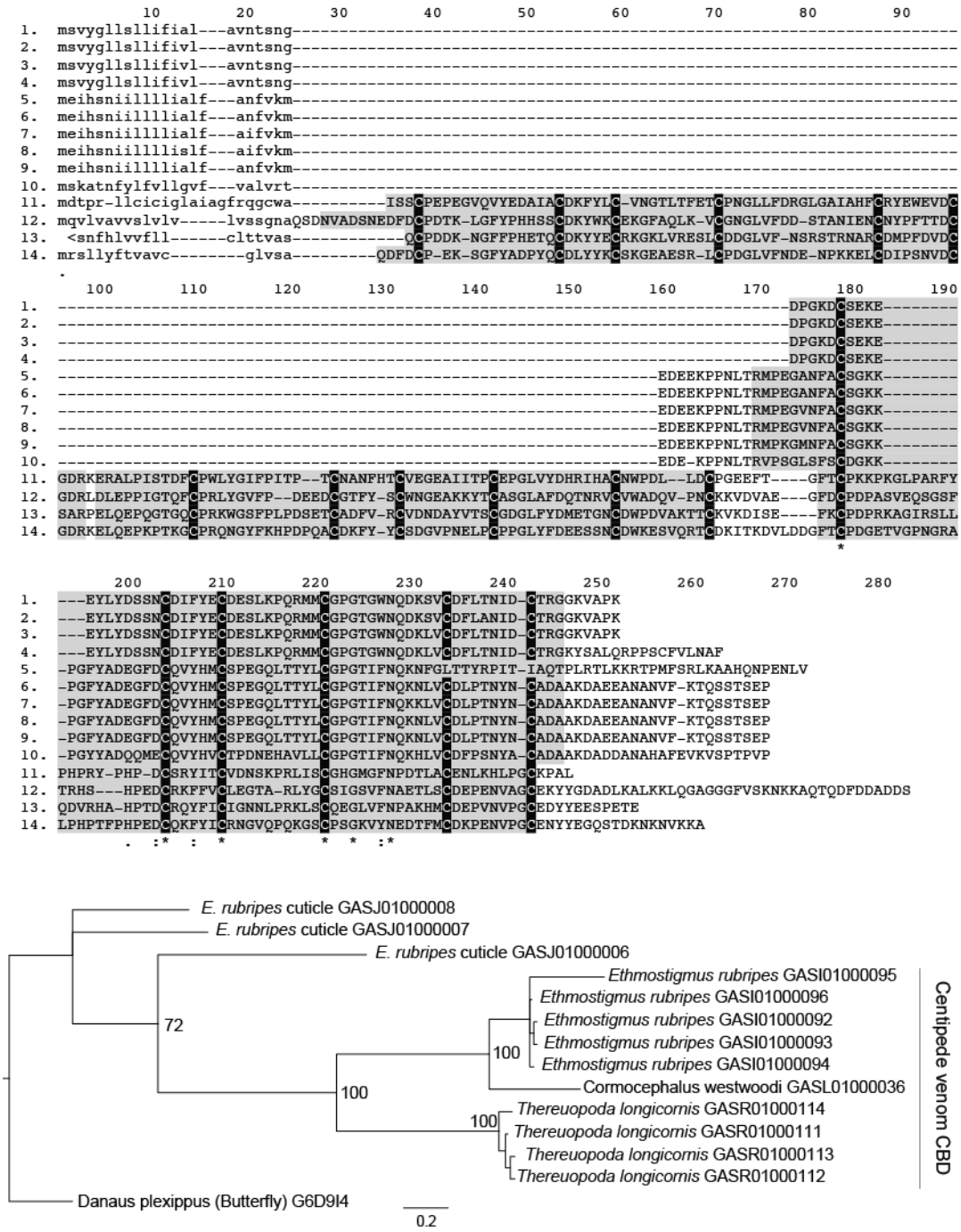

FIGURE 6

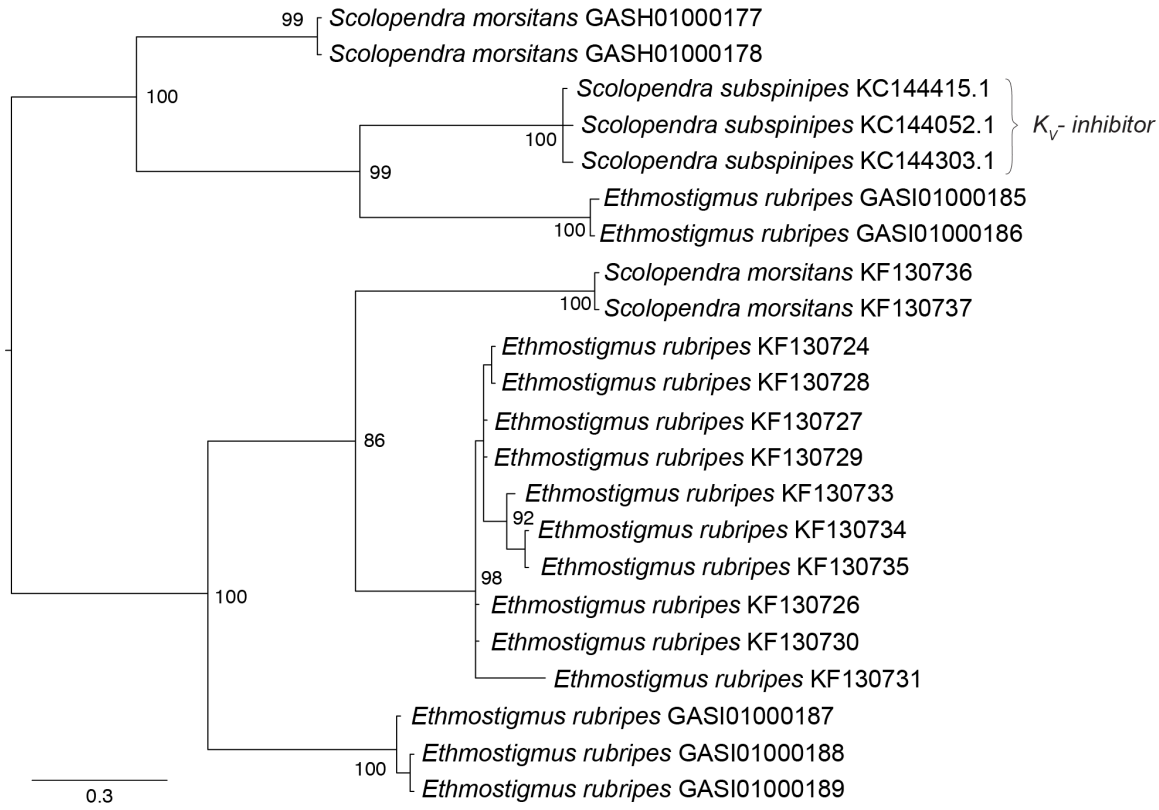

FIGURE 7

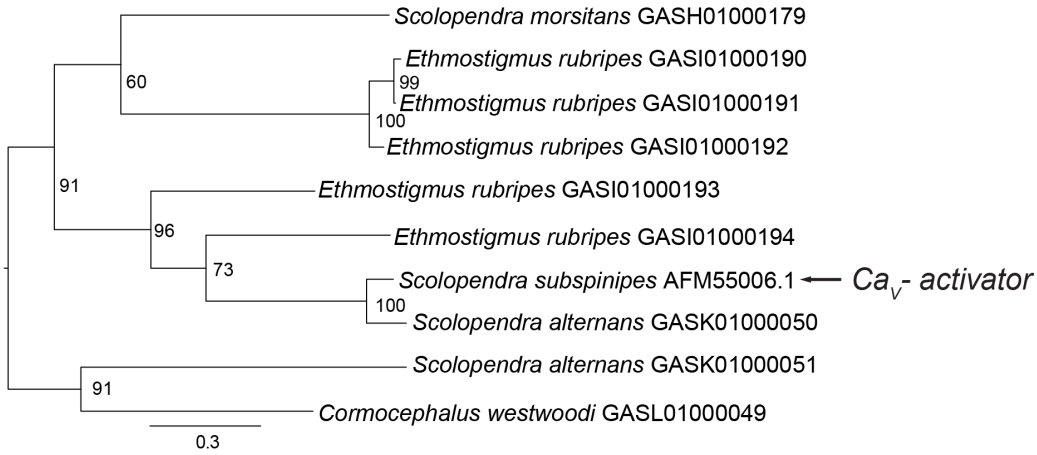

FIGURE 8

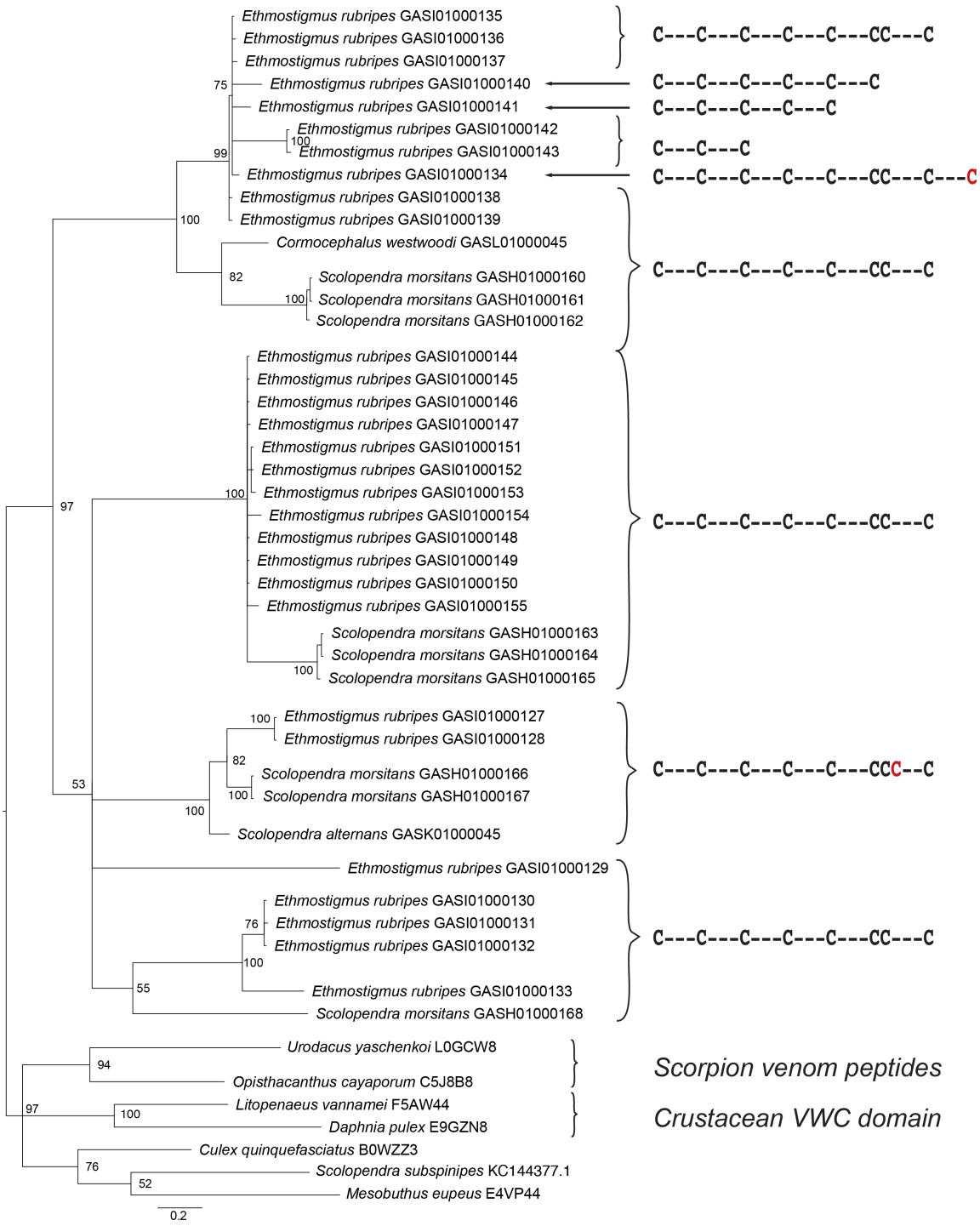

FIGURE 9

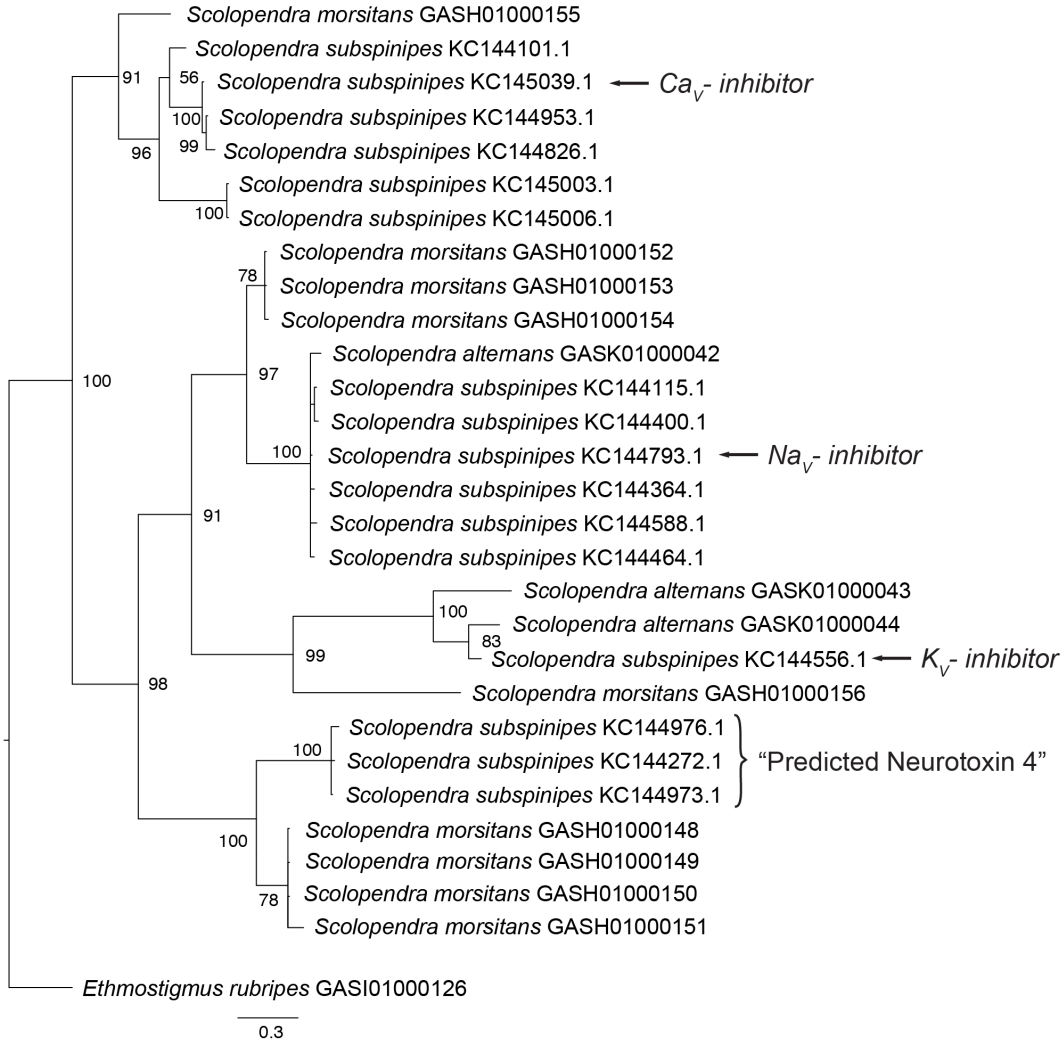

FIGURE 10

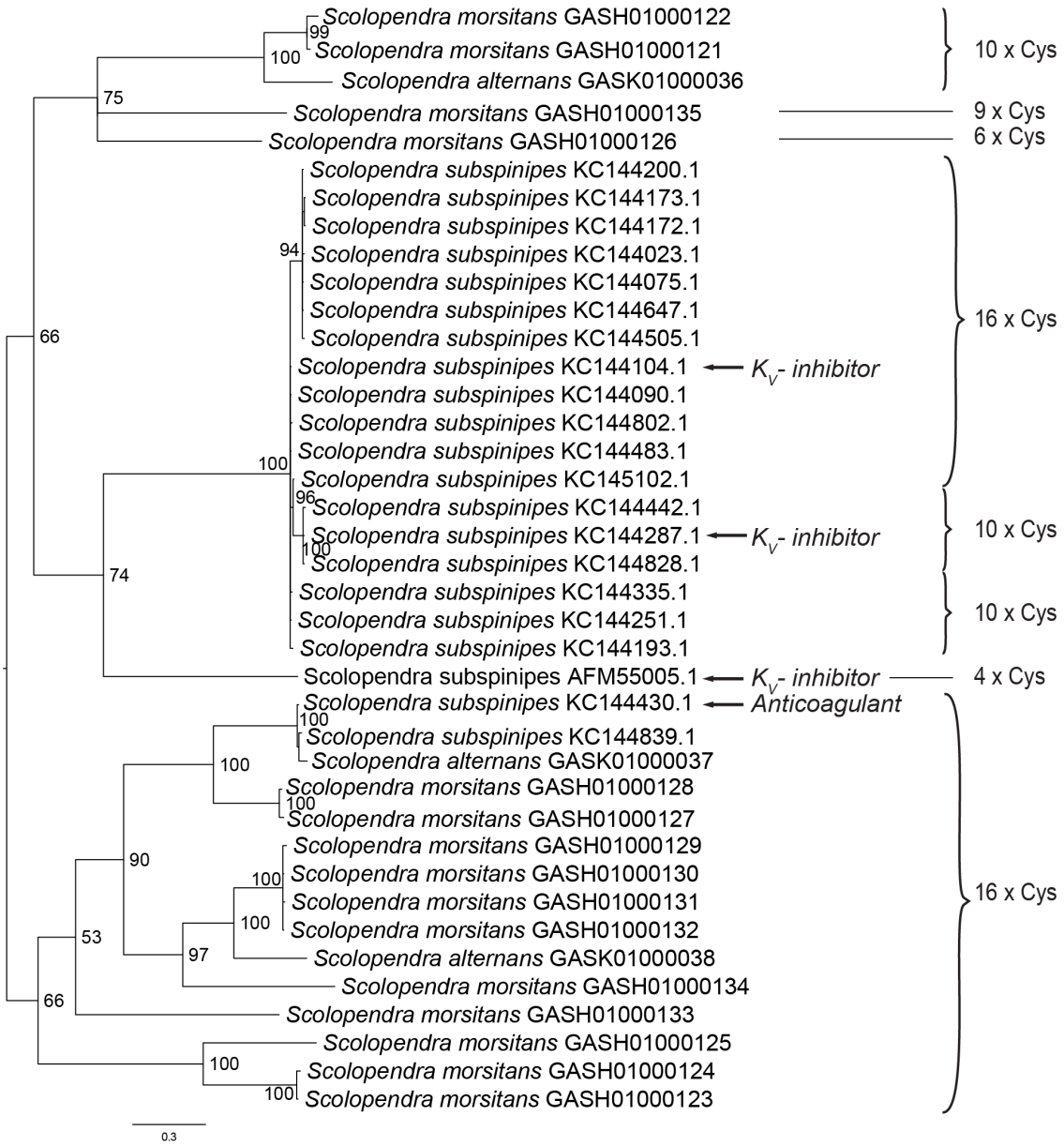

FIGURE 11