

## RESEARCH ARTICLE

## Open Access

# Diversification of a single ancestral gene into a successful toxin superfamily in highly venomous Australian funnel-web spiders

Sandy S Pineda<sup>1†</sup>, Brianna L Sollod<sup>2,7†</sup>, David Wilson<sup>1,3,8†</sup>, Aaron Darling<sup>1,9</sup>, Kartik Sunagar<sup>4,5</sup>, Eivind A B Undheim<sup>1,6</sup>, Laurence Kely<sup>6</sup>, Agostinho Antunes<sup>4,5</sup>, Bryan G Fry<sup>1,6\*</sup> and Glenn F King<sup>1\*</sup>

## Abstract

**Background:** Spiders have evolved pharmacologically complex venoms that serve to rapidly subdue prey and deter predators. The major toxic factors in most spider venoms are small, disulfide-rich peptides. While there is abundant evidence that snake venoms evolved by recruitment of genes encoding normal body proteins followed by extensive gene duplication accompanied by explosive structural and functional diversification, the evolutionary trajectory of spider-venom peptides is less clear.

**Results:** Here we present evidence of a spider-toxin superfamily encoding a high degree of sequence and functional diversity that has evolved via accelerated duplication and diversification of a single ancestral gene. The peptides within this toxin superfamily are translated as prepropeptides that are posttranslationally processed to yield the mature toxin. The N-terminal signal sequence, as well as the protease recognition site at the junction of the propeptide and mature toxin are conserved, whereas the remainder of the propeptide and mature toxin sequences are variable. All toxin transcripts within this superfamily exhibit a striking cysteine codon bias. We show that different pharmacological classes of toxins within this peptide superfamily evolved under different evolutionary selection pressures.

**Conclusions:** Overall, this study reinforces the hypothesis that spiders use a combinatorial peptide library strategy to evolve a complex cocktail of peptide toxins that target neuronal receptors and ion channels in prey and predators. We show that the  $\omega$ -hexatoxins that target insect voltage-gated calcium channels evolved under the influence of positive Darwinian selection in an episodic fashion, whereas the  $\kappa$ -hexatoxins that target insect calcium-activated potassium channels appear to be under negative selection. A majority of the diversifying sites in the  $\omega$ -hexatoxins are concentrated on the molecular surface of the toxins, thereby facilitating neofunctionalisation leading to new toxin pharmacology.

**Keywords:** Spider toxin, Spider venom, Hexatoxin,  $\omega$ -hexatoxin,  $\kappa$ -hexatoxin, Australian funnel-web spider, Molecular evolution, Gene duplication, Positive selection, Negative selection

\* Correspondence: [bgfry@uq.edu.au](mailto:bgfry@uq.edu.au); [glenn.king@imb.uq.edu.au](mailto:glenn.king@imb.uq.edu.au)

†Equal contributors

<sup>1</sup>Institute for Molecular Bioscience, The University of Queensland, 306 Carmody Road, St Lucia, QLD 4072, Australia

<sup>6</sup>Venom Evolution Lab, School of Biological Sciences, The University of Queensland, St Lucia, QLD 4072, Australia

Full list of author information is available at the end of the article

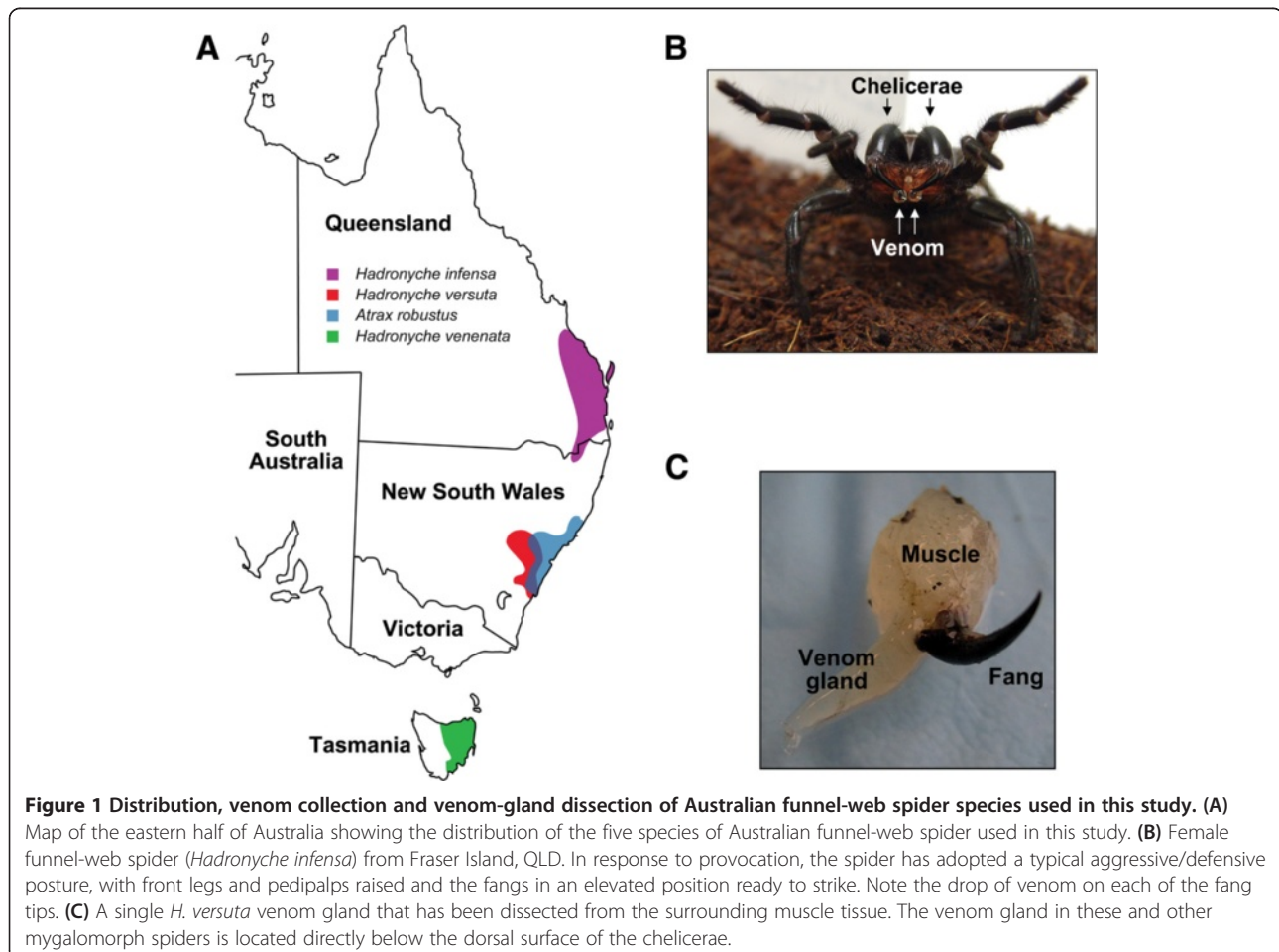
## Background

Venoms have proven to be key evolutionary innovations for many divergent animal lineages [1,2]. Although the most extensively studied venoms are from the medically important scorpions, snakes, and spiders, venom systems are present in many other lineages including cnidarians, echinoderms, molluscs, fish, lizards, and mammals [1,2]. These venoms have evolved to serve a variety of purposes, including prey capture, competitor deterrence, and defense against predators. There has been considerable innovation both in the chemical composition of these venoms as well as the method of venom delivery, which includes barbs, beaks, fangs, harpoons, nematocysts, pinchers, proboscises, spines, spurs, and stingers [1,2].

From a molecular evolutionary perspective, the venoms of snakes are the best understood. There is now abundant evidence that snake venoms evolved by recruitment of genes encoding normal body proteins followed by extensive duplication, neofunctionalization, and in some instances relegation to the status of pseudogene [1,3-5]. In many cases, these genes have been explosively replicated to

produce large multigene families. This process is analogous to the birth-and-death model of evolution proposed for multigene families involved in adaptive immunity, such as the major histocompatibility complex and immunoglobulin  $V_H$  genes [6]. However, the evolutionary trajectory is less clear for the venoms of spiders, scorpions, and molluscs, which are dominated by disulfide-rich peptides of mass 2–9 kDa [7-12]. These peptides typically possess high affinity and often-exquisite specificity for particular classes of ion channels and other nervous system targets [13-15]. These neurotoxic functions are perhaps not surprising given that the primary role of these venoms is to paralyze or kill envenomated prey [11,16,17].

In this study, we analysed toxin-encoding transcripts from five species of Australian funnel-web spider (Aranae: Mygalomorphae: Hexathelidae: Atracinae) from the genera *Atrax* and *Hadronyche*, representing a geographic spread of more than 2000 km (Figure 1), in order to provide insight into the evolutionary trajectory of the  $\omega$ -hexatoxin-1 ( $\omega$ -HXTX-1) family.  $\omega$ -Hexatoxins (formerly known as  $\omega$ -atracotoxins) are peptides comprising ~37 residues that were first isolated from the venom of the lethal Blue



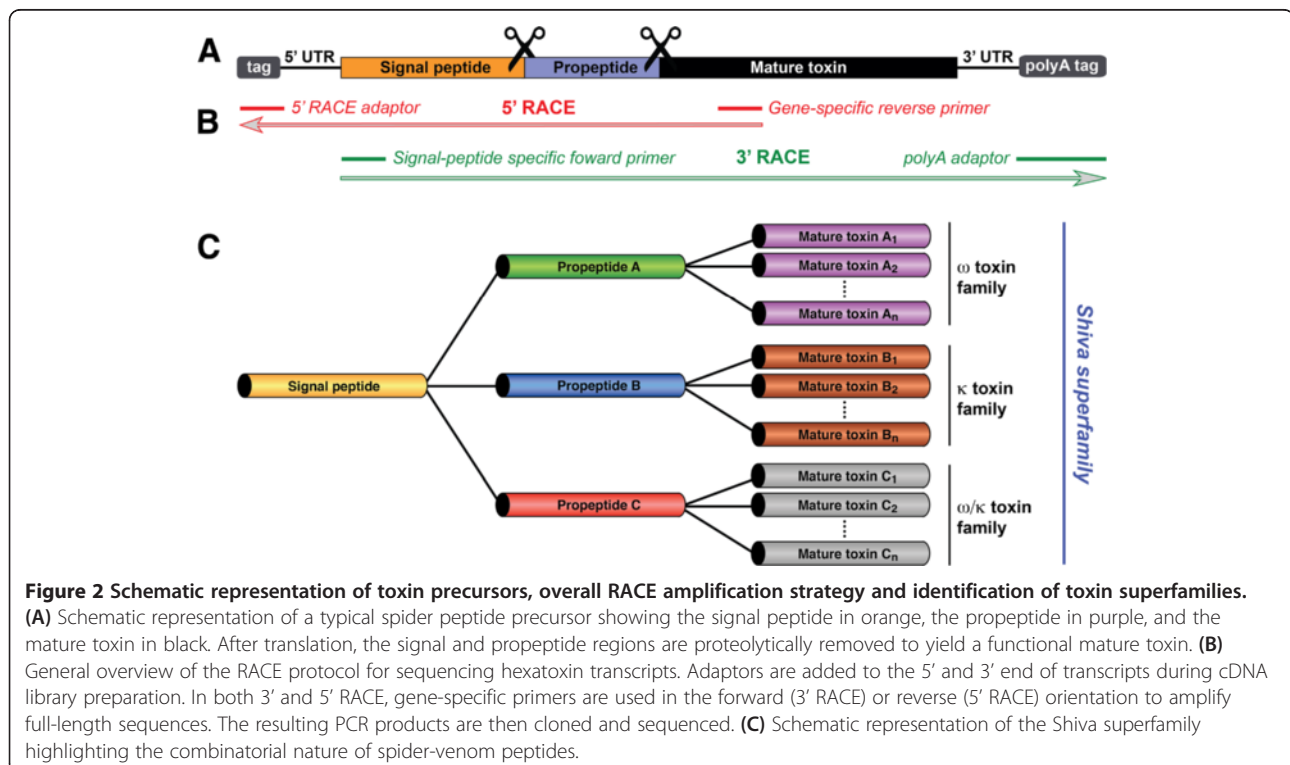
Mountains funnel-web spider *Hadronyche versuta* [18]. The  $\omega$ -hexatoxins are major components in the venom of Australian funnel-web spiders [18-20] and they contribute significantly to prey immobilization by virtue of their ability to specifically block insect, but not vertebrate, voltage-gated calcium ( $Ca_v$ ) channels [17,18,20-22]. Their potent insecticidal activity has engendered interest in these peptides as bioinsecticides [11,17,23]. Proteomic analysis of *H. versuta* venom revealed a number of  $\omega$ -HXTX-Hv1a paralogs [19], suggesting that this peptide toxin might belong to a multigene family. However, because the venom used in this previous study was pooled from several spiders, it was unclear whether these apparent paralogs are simply polymorphisms resulting from allelic variation. By using cDNA libraries obtained from a single spider, we demonstrate here that  $\omega$ -HXTX-Hv1a is indeed part of a large multigene family that appears to have arisen from explosive gene duplication followed by extensive sequence divergence and neofunctionalization. Within this *superfamily* of toxins, we show that pharmacologically distinct toxin classes are evolving under starkly different selection pressures, with some toxin classes accumulating variation under episodic bursts of adaptation, while others remain constrained by negative selection. This work reinforces the idea that the chemical and pharmacological diversity present in spider venoms may have evolved from a relatively small number of ancestral genes.

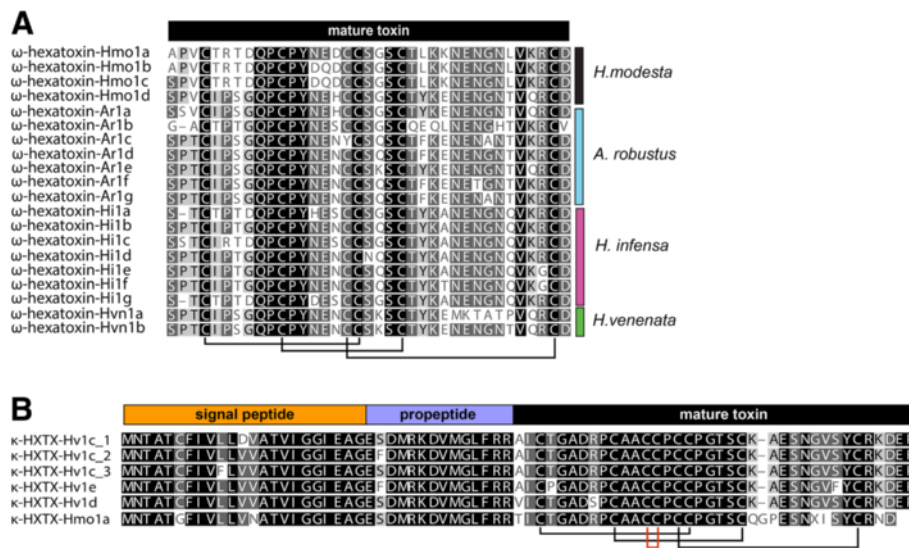
## Results and discussion

### The $\omega$ -hexatoxins are expressed as prepropeptide precursors

RACE analysis was used to amplify transcripts encoding orthologs of  $\omega$ -HXTX-Hv1a from four species of Australian funnel-web spider: *Atrax robustus*, *H. infensa*, *H. venenata*, and *H. versuta* (Figure 2). Multiple  $\omega$ -HXTX-Hv1a orthologs were identified in each species (i.e., 24 paralogs encoding seven distinct mature toxins were identified in *H. infensa*, 18 paralogs encoding six mature toxins were identified in the Sydney funnel-web spider *A. robustus*, and eight paralogs encoding two mature toxins identified in the Tasmanian funnel-web spider *H. venenata*) (Figure 3A). A further eight paralogs encoding four distinct mature toxins were identified in the venom-gland transcriptome of *H. modesta* (Figure 3A). Thus, the amino acid sequence diversity previously reported for  $\omega$ -HXTX-1 based on analysis of pooled venom samples [19] is due to expression of multiple related transcripts in a single spider rather than allelic variation. The almost complete conservation of the signal sequence, as well as the pattern of conserved cysteines in the mature toxin (Figure 3A), indicates that these  $\omega$ -HXTX-Hv1a homologs arose by duplication and sequence divergence of the original toxin-encoding gene.

All of the  $\omega$ -HXTXs are expressed as prepropeptide precursors that are posttranslationally processed to yield the mature toxin sequence (Figure 2A). The highly hydrophobic 22-residue signal sequence is of similar length to





**Figure 3** Sequence alignments of ω and κ-hexatoxins. **(A)** Sequence alignment of ω-HXTX-Hv1a paralogs from each species: *Hadronyche modesta* (Hmo1a–Hmo1d), *Atrax robustus* (Ar1a–Ar1g), *Hadronyche infensa* (Hi1a–Hi1g), and *Hadronyche venenata* (Hvn1a and Hvn1b). The level of residue conservation is graded from black (fully conserved across all paralogs) to dark grey (conserved in most toxins) to light grey (conserved in a majority of orthologs). The lines below the sequence alignment indicate the disulfide-bond connectivity of ω-HXTX-Hv1a. **(B)** Alignment of κ-HXTX-Hv1a paralogs from *Hadronyche versuta* and one ortholog from *Hadronyche modesta*. The level of residue conservation is graded as in panel (A) and the signal peptide, propeptide, and mature toxin regions are highlighted. The lines below the sequence alignment indicate the disulfide-bond connectivity of κ-HXTX-Hv1c, with the vicinal disulfide bond highlighted in red. Note that ω-HXTX-Ar1a (UniProt PF06357), ω-HXTX-Hi1a (UniProt P0C2L5), ω-HXTX-Hi1b (UniProt P0C2L6), ω-HXTX-Hi1c (UniProt P0C2L7), and κ-HXTX-Hv1c (UniProt P82228) have been previously isolated directly from venom.

that reported for peptide-toxin precursors from spiders, scorpions, and cone snails [8]. The 20-residue propeptide sequence is highly acidic, with a net charge of -4, a feature that has been noted for numerous spider-toxin propeptide sequences [24-28] but which is not characteristic of toxin precursors from other venomous animals. Moreover, the presence of a propeptide region contrasts with most scorpion-toxin precursors in which the signal sequence is fused directly to the mature toxin without an intervening propeptide [8]. The reason for the highly acidic propeptide region in spider toxin precursors remains to be determined, but it may be related to specific interactions between the toxin precursor and components of the secretory and/or protein folding pathway in spider venom glands.

The propeptide sequence terminates with a dibasic Arg-Arg signature; dibasic sequences are common recognition sites for proteolytic removal of propeptide segments in neuropeptide precursors from both vertebrates [29] and invertebrates [30]. While Arg is the terminal residue in virtually all known spider-toxin propeptide sequences, the penultimate residue is variable, though it is commonly Asp, Glu, or Lys [25,26,31-33].

**ω-hexatoxins belong to a large toxin-gene superfamily**

Orthologs of ω-HXTX-Hv1a were identified in all five species of Australian funnel-web spider examined in this

study. These species are distributed along the eastern seaboard of Australia with a geographic spread of more than 2000 km (Figure 1). The hexathelids are a group of approximately 40 species divided into three genera: *Atrax*, *Hadronyche* and *Illawarra* [34-36]. They are adapted to forest environments but can also be found in habitats that range from montane herblands and open woodland to closed forest [36]. Conservation of the ω-hexatoxin family of toxins over this wide range of environments and differing prey distributions implies that there has been strong evolutionary pressure to maintain these peptides as part of the venom arsenal, which is perhaps not surprising given that they are broadly active against many different arthropods [11,17,37].

In addition to obvious homologs of ω-HXTX-Hv1a, the RACE and transcriptomic analyses revealed additional families of toxins that had almost identical signal sequences to the ω-HXTX-1 transcripts, but divergent propeptide and mature toxin sequences. We named one of these families the ω/κ-HXTX family (described as U-ACTX in [38]). The ω/κ-HXTX peptides appear to be distributed in two of the species examined (*A. robustus* and *H. versuta*); this reinforces the idea that these toxins most likely arose ancestrally by duplication of a ω-HXTX-1 gene followed by hypermutation of the propeptide and mature-toxin regions in order to create a new function (neofunctionalization). The conservation and radiation of



these toxins across this family of spiders implies that they are not nonfunctional relics of an explosive radiation of this toxin-gene superfamily, and we confirmed this by showing that recombinant  $\omega/\kappa$ -HXTX-Hv1a is highly insecticidal [38]. The high insecticidal potency of this family of peptides is believed to result from a synergistic effect on insect voltage-gated calcium ( $Ca_v$ ) channels and calcium-activated potassium ( $K_{Ca}$ ) channels [38].

RACE analysis of the venom-gland cDNA library from *H. versuta* also led to amplification of transcripts encoding the insecticidal toxin  $\kappa$ -HXTX-Hv1c [39], and sequencing of the venom-gland transcriptome from *H. modesta* also uncovered an ortholog of this toxin (Figure 3B). This was entirely unexpected since this toxin has a vastly different primary structure to  $\omega$ -HXTX-Hv1a [39]. Moreover, in addition to the six conserved cysteine residues in  $\omega$ -HXTX-Hv1a that form an inhibitor cystine knot (ICK) motif [40,41],  $\kappa$ -HXTX-Hv1c contains two additional cysteine residues that form an extremely rare vicinal disulfide bond [42-44]. Furthermore, in contrast to  $\omega$ -HXTX-Hv1a, which blocks insect  $Ca_v$  channels,  $\kappa$ -HXTX-Hv1c is a potent and specific blocker of  $K_{Ca}$  channels [45]. Nevertheless, the near identity of the signal sequence in these two toxin families and the conservation of cysteine residues in the mature toxin indicate that they evolved from the same ancestral toxin gene and are members of the same gene superfamily.

We did not find orthologs of  $\kappa$ -HXTX-Hv1c in any of the other three species of Australian funnel-web spider (*H. infensa*, *A. robustus*, and *H. venenata*). However,  $\kappa$ -HXTX-Hv1a,  $\kappa$ -HXTX-Hv1b, and  $\kappa$ -HXTX-Hv1c are expressed at very low levels in *H. versuta* venom [42], and consequently we cannot rule out the possibility that these toxins are present in the venom of the other three spiders but the transcript levels are too low to be detected using the methods employed here.

### The Shiva superfamily of peptide toxins

It has previously been suggested that superfamilies of spider-venom peptides evolved from a single ancestral gene via explosive gene duplication [8]; the work described here further supports this idea as it is clear that the  $\omega$ -HXTXs,  $\omega/\kappa$ -HXTXs, and  $\kappa$ -HXTXs belong to a large superfamily of toxins that arose via gene duplication (Figure 2C). We have chosen to name spider-toxin gene superfamilies after deities of death and destruction since the major biological role of these toxins is to paralyze and/or kill envenomated prey. Accordingly, we have named the  $\omega/\kappa$ -HXTX/ $\omega$ -HXTX/ $\kappa$ -HXTX gene superfamily after the Hindu deity Shiva, commonly known as the “destroyer”.

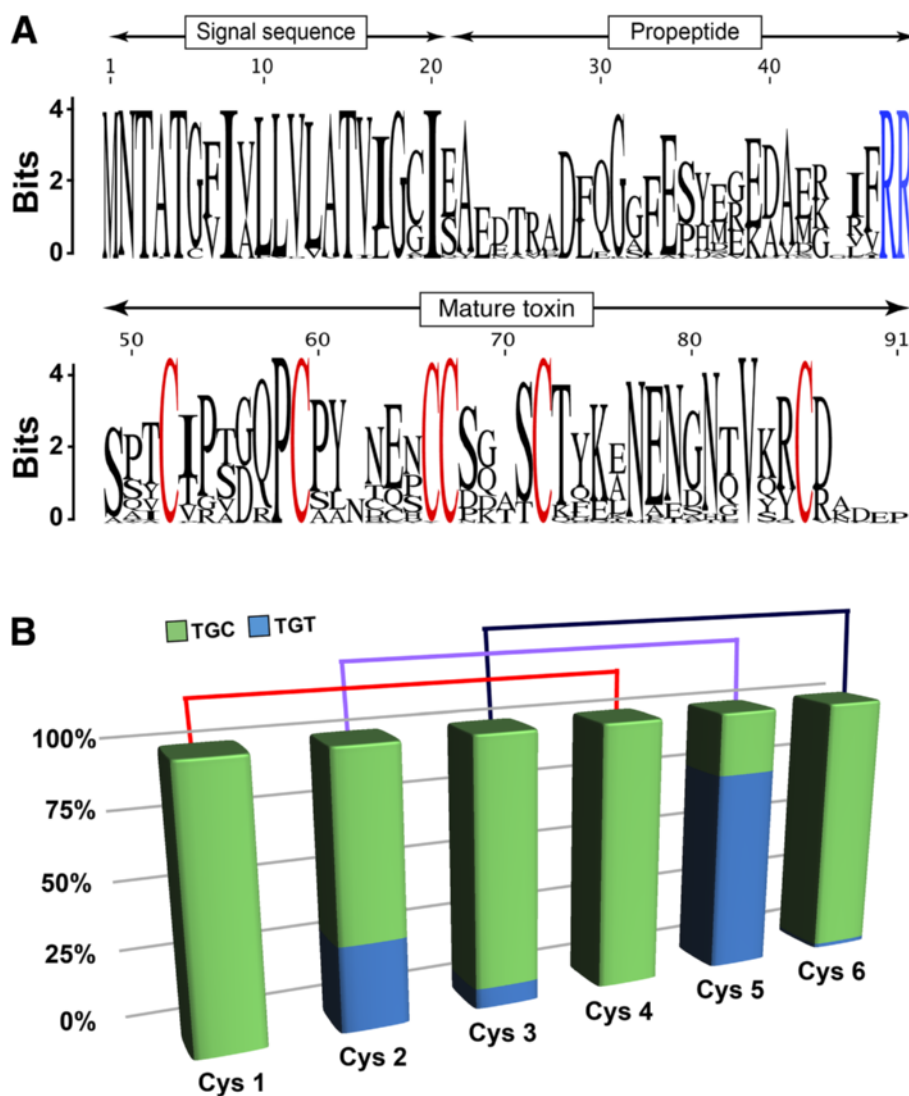
Sequence logos were previously used to analyse differences in the level of sequence conservation between the three parts of the  $\omega$ -HXTX toxin precursor, namely the

signal peptide, propeptide, and mature toxin [8]. A revised logo analysis of the Shiva superfamily (Figure 4A) that incorporated all of the new sequences and species reported here reinforced the dichotomy in evolutionary forces affecting various elements of the toxin precursor. The signal peptide has clearly been highly conserved throughout the evolution of this toxin superfamily and it is presumably under negative selection in order to ensure that these toxins are directed to the appropriate secretory pathway. In contrast, there is significant sequence variation in both the propeptide and mature toxin sequences, with two notable exceptions. First, in contrast to the highly variable upstream region of the propeptide sequence, the C-terminal proteolytic recognition signal (Arg-Arg) is completely preserved (Figure 4A). Presumably there has been strong selection pressure to ensure processing of the propeptide by a specific protease. Second, in contrast to the overall low level of conservation of the mature toxin sequence, the cysteine residues, which direct the three-dimensional (3D) fold of the toxins, are completely conserved (Figure 4A). The marked variation in levels of sequence conservation between the spider-toxin signal sequence and the propeptide and mature toxin regions is reminiscent of that observed for superfamilies of cone snail toxins [46-51].

There are two striking differences between the Shiva superfamily precursors and transcripts encoding human neuropeptides and other secreted proteins. First, whereas precursors of human neuropeptides often encode multiple mature neuropeptide sequences [29,52], we and others have not found any examples of spider-toxin transcripts that encode more than a single mature toxin sequence. Secondly, in direct contrast to the toxin precursors, the sequence of the mature human neuropeptide(s) is usually strongly conserved whereas there is significantly more variability in the signal sequence. This is perhaps not surprising given that human neuropeptides usually act on a single well-defined molecular target whereas spider toxins typically target a specific subtype of receptor or ion channel that nevertheless might vary significantly in primary structure between prey taxa (Note that most spiders are generalist predators that target a phylogenetically diverse range of prey). Thus, expressing a family of related toxins in the venom (essentially a mini-combinatorial peptide library) might ensure that the desired receptor/ion channel is targeted, regardless of prey taxa.

### Position-specific cysteine codon bias

Mature  $\omega$ -HXTXs contain three disulfide bonds with 1-4, 2-5, 3-6 connectivity. These disulfides form an ICK motif that provides these toxins with a high degree of chemical, thermal and biological stability [53]. Although it is clear from a protein structure viewpoint why these six cysteine residues need to be strictly maintained in



**Figure 4 Sequence logo and codon usage analysis from the Shiva superfamily. (A)** Sequence logo [54] based on alignment of prepropeptides from the Shiva superfamily. There is a much higher level of sequence conservation within the signal peptide than within the propeptide and mature toxin regions. Note, however, that the cysteine residues that form the cystine-knot motif in the mature toxin and the Arg-Arg protease recognition site that terminates the propeptide region are both completely conserved (highlighted in blue and red, respectively). **(B)** Codon usage for the six-cysteine residues that form the cystine-knot motif. Note the strong bias for TGC at cysteine positions 1, 3, 4, and 6. Shown above the histogram is the disulfide bridge arrangement for the six cysteines as inferred from the 3D structures of  $\omega$ -HXTX-HV1a and  $\kappa$ -HXTX-HV1c.

order to preserve the toxin's 3D scaffold, one would not expect to find a preference for either one of the two possible cysteine codons (TGT and TGC). Intriguingly, however, previous analysis of  $\omega$ -HXTX precursors revealed a strong bias for TGC at four of the six cysteine positions in the mature toxin region [8]. An extended logo analysis [54] incorporating all of the newly discovered sequences reported in this study corroborated the previously observed codon bias (Figure 4B). We found an extreme TGC codon bias for the four cysteine residues that form the 1–4 and 3–6 disulfide bridges in the  $\omega$ -HXTX family but not for the two cysteines that form the 2–5

disulfide bond (Figure 4B). The observed position-specific codon bias is not simply a manifestation of global codon bias in these spiders as we have observed a preference for TGT as opposed to TGC for cysteine residues in other hexatoxin superfamilies (data not shown). Moreover, we did not observe extreme codon bias for any other conserved residue in the mature hexatoxins.

Position-specific cysteine-codon bias has also been observed in superfamilies of cone snail toxins and it has been proposed that these codons might serve as attractants for a mutator complex that includes a poorly processive and highly mutagenic polymerase (e.g., DNA Pol

V) that promotes radiation of the toxin superfamily by facilitating hypermutation of the mature toxin region [49,50]. However, there is currently no direct evidence that cysteine-codon bias plays a part in directing the evolution of spider or cone snail toxins.

### Molecular evolution analyses

We utilized various state-of-art molecular evolutionary assessment methods to determine the influence of natural selection on the evolution of genes encoding Shiva superfamily toxins (see *Methods* section for full details of the selection analyses). The one-ratio model, the simplest of the codon-specific models, estimated the non-synonymous-to-synonymous nucleotide-substitution rate ratio ( $\omega$ ) to be 0.64, 1.06 and 0.69 for the  $\omega$ -HXTXs,  $\kappa$ -HXTXs, and combined Shiva superfamily dataset, respectively (Additional file 1: Table S1–3). This highly conservative model can only detect positive selection when  $\omega$ , averaged over all sites along the lineages in a phylogenetic tree, is significantly greater than one. As lineage-specific models of PAML, such as the one-ratio model, often fail to detect positive-Darwinian selection that only affects certain sites in proteins, we also employed

site-specific models (Table 1: codon numbers based on  $\kappa$ -HXTX-Hv1c\_2 and  $\omega$ -HXTX-Ar1a\_1; Additional file 1: Table S1–3). Model 8 estimated  $\omega$  of 0.69, 1.06 and 0.78 for the  $\omega$ -HXTXs,  $\kappa$ -HXTXs, and the combined Shiva superfamily dataset, respectively (Table 2 and Additional file 1: Table S1–3). Although the computed  $\omega$  for the  $\kappa$ -HXTXs was  $>1$ , the assessment was not statistically significant ( $p > 0.05$ ) in comparison with the null model ( $M7 \beta$ ). The Bayes Empirical Bayes (BEB) approach implemented in M8 was only able to identify one positively selected site in the combined toxin dataset (Table 2 and Additional file 1: Table S3). Thus, the site-specific models failed to detect the influence of adaptive selection pressures in shaping evolution of the Shiva superfamily. In contrast, the more advanced Fast, Unconstrained Bayesian Approximation (FUBAR) [55,56] implemented in HyPhy detected a handful of positively selected sites in both the  $\omega$ -HXTXs and the combined dataset (Table 1).

Site-specific models for detecting positive selection work best when detecting pervasive selection pressures. However, the majority of positively selected sites are often subjected to transient or episodic adaptations. When the majority of lineages evolve under the influence of

**Table 1 Nucleotide and complementary protein analyses for  $\omega$  toxins**

Codon	Site <sup>a</sup> Amino Acid	CodeML		Tree SAAP		Accessible surface area <sup>f</sup>
		M2a <sup>b</sup>	M8 <sup>c</sup>	Property <sup>d</sup>	Magnitude <sup>e</sup>	
21	E	0.99 ± 0.38 (0.201)	0.87 ± 0.42 (0.257)	-	-	-
40	V	1.41 ± 0.49 (0.589)	1.47 ± 0.37 (0.842)	$M_W, M_V, V^O, \mu$	8, 8, 8, 8	42.0 Partially exposed
44	S	1.25 ± 0.37 (0.420)	1.28 ± 0.41 (0.647)	$M^W, M_V, V^O, \mu$	8, 8, 8, 8	82.1 Exposed
53	H	1.62 ± 0.63 (0.749)	1.55 ± 0.36 (0.929)	-	-	0.0 Buried
57	G	1.39 ± 0.48 (0.570)	1.45 ± 0.38 (0.825)	-	-	57.3 Exposed
60	T	0.80 ± 0.44 (0.122)	0.71 ± 0.41 (0.154)	-	-	49.7 Exposed
64	N	0.53 ± 0.40 (0.031)	0.51 ± 0.30 (0.036)	-	-	100.0 Exposed
69	T	1.34 ± 0.45 (0.507)	1.38 ± 0.40 (0.751)	-	-	59.8 Exposed
72	R	1.10 ± 0.33 (0.26)	1.01 ± 0.42 (0.374)	-	-	0.0 Buried

<sup>a</sup>Sites detected as positively selected using the integrative approach.

<sup>b</sup>M2a Bayes empirical Bayes (BEB) posterior probability and post-mean  $\omega$  indicated in parentheses.

<sup>c</sup>M8 Bayes empirical Bayes (BEB) posterior probability and post-mean  $\omega$  indicated in parentheses.

<sup>d</sup>Amino acid property under selection ( $M_W$ : molecular weight;  $M_V$ : molecular volume;  $V^O$ : partial specific volume;  $\mu$ : Refractive index).

<sup>e</sup>Magnitude if selection on the amino acid property.

<sup>f</sup>Accessible surface area of 10–20% corresponds to buried residues, 40–50% indicates partially exposed amino acid residues, and  $\geq 50\%$  indicates solvent exposed residues.

**Table 2 Molecular evolution of  $\omega$  and  $\kappa$  toxins from Australian funnel-web spiders**

	FUBAR <sup>a</sup>	MEME <sup>b</sup>	PAML <sup>c</sup>	
			M8	M2a
$\omega$	$\omega > 1^d$ : 3	7	0	0
toxins	$\omega < 1^e$ : 5		0	0
			0.69	0.73
$\kappa$	$\omega > 1^d$ : 1	0	0	0
toxins	$\omega < 1^e$ : 0		0	0
			1.06 <sup>NS</sup>	1.06 <sup>NS</sup>
ALL	$\omega > 1^d$ : 3	8	1	0
toxins	$\omega < 1^e$ : 7		(0 + 1)	0
			0.78	0.83

<sup>a</sup>Fast, Unconstrained Bayesian Approximation (FUBAR).

<sup>b</sup>Sites detected as experiencing episodic diversifying selection (0.05 significance) by mixed effects model evolution (MEME).

<sup>c</sup>Positively selected sites detected using the Bayes empirical approach implemented in the site models M8 and M2a. Number of positively selected sites detected at the posterior probability  $\geq 0.99$  and 0.95 are indicated in parenthesis.  $\omega$  computed using M8 and M2a are also presented.

<sup>d</sup>Number of sites evolving under the influence of pervasive diversifying selection, detected by FUBAR at 0.9 posterior probability.

<sup>e</sup>Number of sites evolving under the influence of pervasive purifying selection, detected by FUBAR at 0.9 posterior probability.

$\omega$  = mean dN/dS.

**NS** = not significant at 0.05 compared to the null model (M7: beta).

negative selection, they mask the signal of positive selection that influences only a small number of lineages. In such scenarios, the aforementioned analyses may fail to detect the influence of positive selection. To address the shortcomings of the aforementioned approaches, we employed the advanced Mixed Effects Model Evolution (MEME) [57], which uses fixed effects likelihood (FEL) along the sites and random effects likelihood (REL) across the branches to detect episodic diversifying selection. MEME is capable of identifying both pervasive and episodic adaptations. MEME identified 7 and 8 episodically diversifying sites in the  $\omega$ -HXTXs and combined toxin dataset, respectively (Table 2), highlighting the vital role of episodic diversifying selection in shaping the evolution of these spider toxins. Six out of eight episodically diversifying sites (75%) were located on the molecular surface of the toxins (Table 1 and Figure 5B) with their side chains completely or partially exposed to solvent, suggesting that they could act as pharmacological sites and participate in prey envenomation; these findings are also in agreement with the selection forces found on the surface of the SGTx toxin family from the venom of the African Baboon spider *Scodra griseipies* [58]. Rapid Accumulation of Variations in Exposed Residues (RAVER), where the toxin molecular chemistry undergoes hypervariations under the influence of positive Darwinian selection and focal mutagenesis [59], has been documented in a plethora of venom-components from a wide diversity of venomous animal lineages

[59-64]. Since the synthesis and secretion of venom proteins is energetically expensive [65-67], mutations that disrupt the structure/function of proteins are filtered out of the population by negative selection over time, favoring the conservation of catalytic and structurally important residues. RAVER not only aids in generation of a rapidly variable toxin molecular surface biochemistry, but it also ensures the conservation of structurally and functionally important residues. Accumulation of variations on the molecular surface of the toxin is advantageous as the altered surface chemistry might lead to new toxin functions (neofunctionalisation).

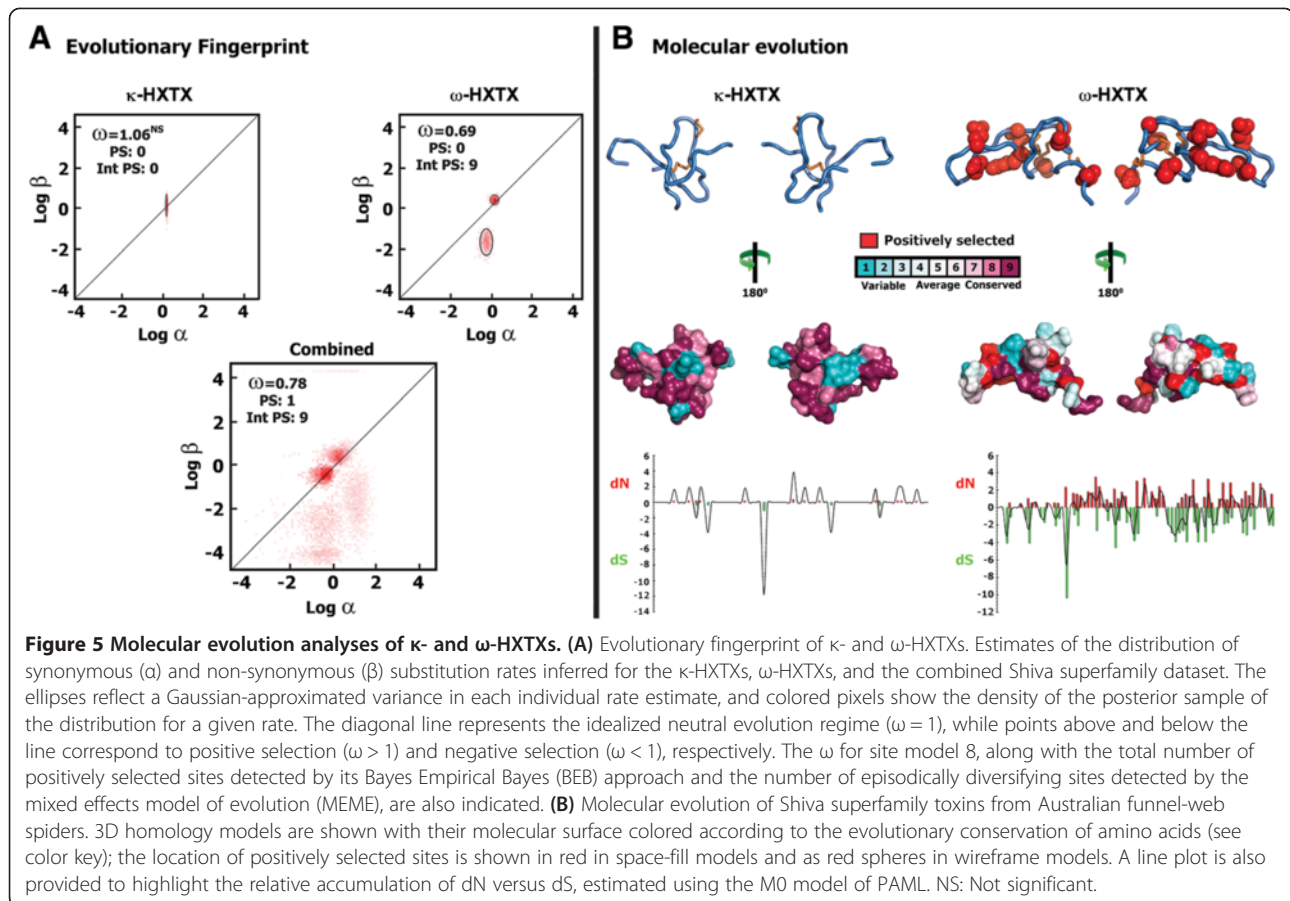
To derive further support for the positively selected sites detected by nucleotide analyses, we employed a complementary protein-level approach implemented in TreeSAAP (Table 1). TreeSAAP identified two positively selected sites in the  $\omega$ -HXTXs that were in common with the sites identified by site-model 8 of PAML (Table 1). Evolutionary fingerprint analyses (Figure 5A) clearly revealed several residues in the  $\omega$ -HXTXs and the combined toxin dataset that evolve under the influence of positive selection, while a majority of residues in the  $\kappa$ -HXTXs remained under evolutionary constraint (Figure 5A,B). Thus, evolution of the  $\omega$ -HXTXs has been significantly influenced by short bursts of episodic adaptations, while the  $\kappa$ -HXTXs appear to be under negative selection.

Phylogenetic analysis revealed that the  $\kappa$ -HXTXs form a separate clade to the  $\omega$ -HXTXs, rendering the Shiva superfamily non-monophyletic (Figure 6). There are also significant variations within the  $\omega$ -HXTXs suggestive of functional diversification (Figure 6). The "hybrid"  $\omega/\kappa$ -HXTXs exhibit functional characteristics of both the  $\omega$ -HXTXs and  $\kappa$ -HXTXs as they block  $Ca_v$  channels (like the  $\omega$ -HXTXs) as well as  $K_{Ca}$  channels (like the  $\kappa$ -HXTXs). The functional activity of the  $\omega/\kappa$ -HXTXs combined with their relative phylogenetic placement and cysteine pattern indicates that they are structurally and functionally intermediate between the  $\omega$ - and  $\kappa$ -HXTXs. The evolution of new cysteine residues to create the vicinal disulfide bond in the  $\kappa$ -HXTXs potentiated toxin activity on  $K_{Ca}$  channels, since mutagenesis and analogue studies indicate that this vicinal disulfide bond is the most critical part of the  $K_{Ca}$  pharmacophore [43,45,68].

#### Constraints on mutation of the mature toxin sequence

It is generally considered that conservation of the cysteine scaffold in toxin-gene superfamilies is critical for conserving the toxin's 3D fold [35]. However, the incredible disparity in the amino acid sequence between  $\omega$ -HXTX-Hv1a and  $\kappa$ -HXTX-Hv1c (Figure 7A) begs the question of whether this is reflected in a significant difference in their 3D structures, despite their common cysteine-knot scaffold. The 3D structure of both toxins has been determined previously using homonuclear NMR





spectroscopy [18,42] and their pharmacophores elucidated using alanine scanning mutagenesis [17,21,43,69].

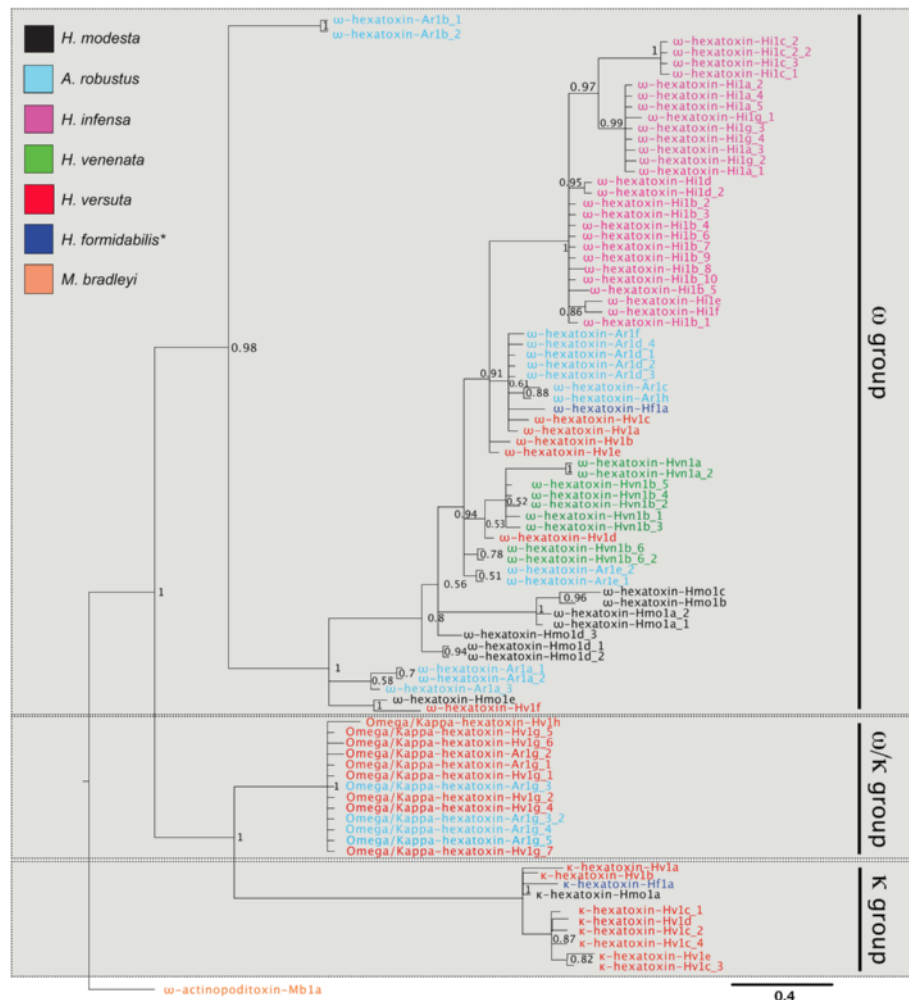
Figure 7B and C show schematic representations of the 3D structure of κ-HXTX-Hv1c and ω-HXTX-Hv1a, respectively. The two toxins can be considered to comprise four inter-cystine loops, which are labelled 1–4 from N- to C-terminus. Although there is an obvious similarity in the disposition of the three centrally located disulfide bridges that form the cystine-knot motif in each toxin, the overall topology of the toxins, as well as the size and relative orientation of the four inter-cystine loops, appears quite different. However, the structural overlay in Figure 7D, which was generated automatically by the DaliLite structural alignment program [70], reveals that the two structures are in fact remarkably similar.

The DaliLite alignment yields a root mean square deviation of 2.4 Å over the backbone atoms of the 28 aligned residues, indicating that the two toxins are indeed structural homologs. The three central disulfide bridges and loop 1 align remarkably well. Loop 4, which encompasses the β-hairpin present in both toxins, also aligns well except for the four-residue insertion in ω-HXTX-Hv1a (Figure 7A), which increases the size of the hairpin loop at the tip of Loop 4. The major

structural differences between the two toxins are the very different orientations of loops 2 and 3. However, these structural variations cannot disguise the fact that the two toxins essentially conform to the same 3D scaffold despite their extraordinary sequence divergence (16% identity if the cysteine framework is excluded). This ability to maintain a consistent molecular architecture despite massive variation in the inter-cystine loop sequences has important implications for the mechanism by which this superfamily of peptide toxins has evolved.

## Conclusions

Spiders and other venomous animals rely on the production of pharmacologically complex venoms for defense, prey capture, and competitor deterrence. The major components of most spider venoms are disulfide-rich peptides that have evolved to target a wide range of receptors and ion channels in the insect nervous system. The ω-HXTX and κ-HXTX families were the first peptides isolated from Australian funnel-web spiders that were shown to be insecticidal [17]. Analysis of all transcripts encoding these peptides showed that they are initially expressed as pre-peptides that are proteolytically processed to yield a 36–37 residue mature peptide that contains three disulfide



**Figure 6 Bayesian phylogenetic tree representing the molecular evolutionary history of the Shiva superfamily toxins.** The tree shows the split between the three main toxin classes ( $\omega$ ,  $\kappa$ , and  $\omega/\kappa$ ).  $\omega$ -Actinopoditoxin-Mb1a from the Eastern mouse spider *Missulena bradleyi* was used as the outgroup. Toxins belonging to each species are highlighted in the following colours: *H. versuta*, red; *H. modesta*, black; *H. venenata*, green; *H. infensa*, magenta; *A. robustus*, pale blue; *H. formidabilis*, dark blue;  $\omega$ -actinopoditoxin-Mb1a from *M. bradleyi*, orange. \*denotes a species not sequenced as part of this study; the sequence was downloaded from UniProt under accession number P83588.

bridges that form an ICK motif plus a non-canonical vicinal disulfide bond in the  $\kappa$ -HXTXs.

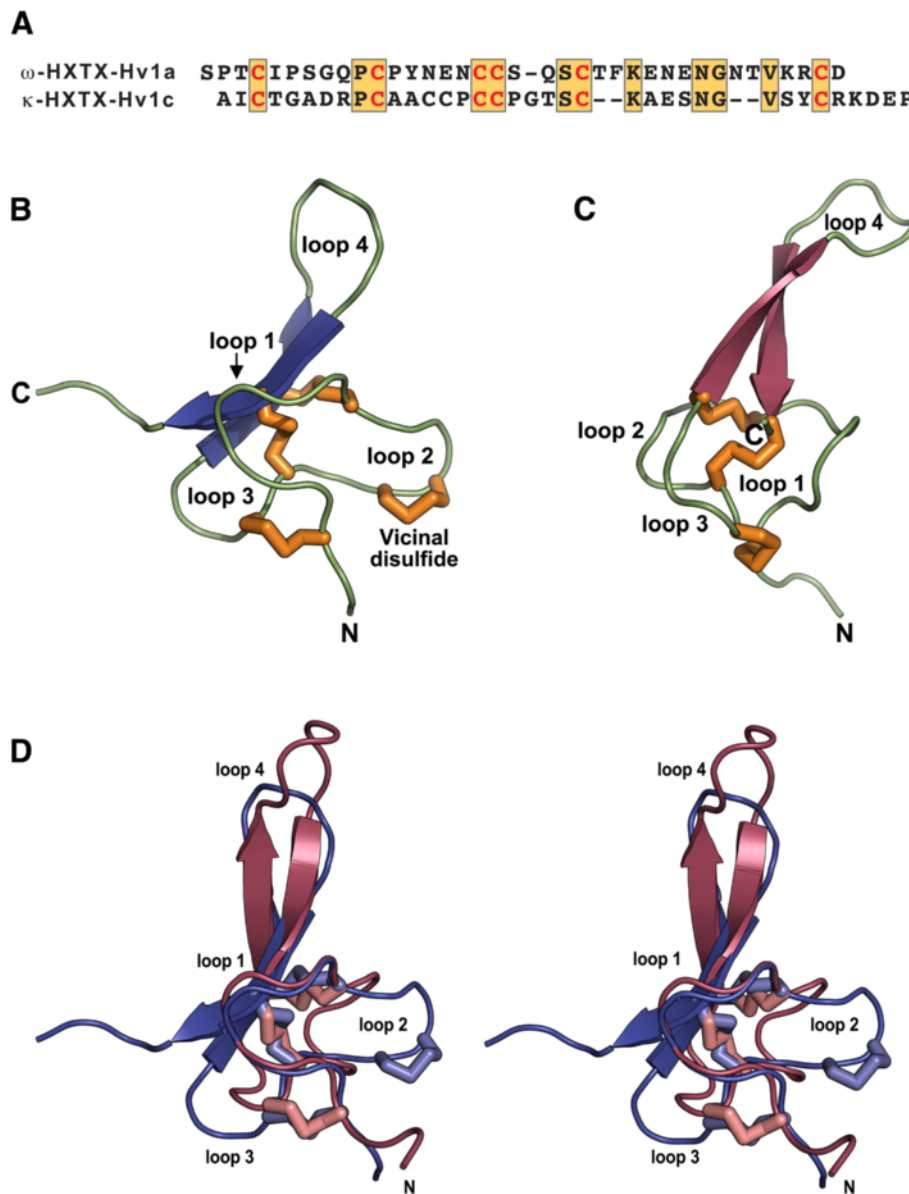
The extreme diversity of primary structure within the Shiva toxin superfamily suggests that there have been few evolutionary restraints on sequence diversification outside of the disulfide bridges that direct the 3D fold of these peptides. The  $\omega$ -HXTXs, in particular, seem to have evolved under the influence of positive Darwinian selection in an episodic fashion, whereas the  $\kappa$ -HXTXs appear to be constrained by negative selection pressures. Functional assessments of these toxins should shed further light on why they have adopted quite contrasting molecular evolutionary regimes.  $\omega$ -HXTXs were also found to have adopted RAVER, where a large number of the episodically diversifying sites are concentrated on the molecular surface, facilitating the generation of novel

pharmacological sites. These toxins may therefore be good candidates for *in vitro* evolution studies designed to produce modified peptides with desired therapeutic [14] or agrochemical [11] properties. Most importantly, this study reinforces the idea that the remarkable chemical and pharmacological complexity of spider venoms may be derived from a relatively small number of ancestral genes.

## Methods

### Identification of $\omega$ -HXTX-1 homologs via rapid amplification of cDNA ends

Venom-gland cDNA libraries were prepared from individual specimens of the following species of Australian funnel-web spider (Arthropoda: Chelicerata: Arachnida: Araneae: Opisthothelae: Mygalomorphae: Hexathelidae): *Hadronyche infensa*, *H. versuta*, *H. venenata*, and *Atrax*



**Figure 7 Structural comparisons of  $\kappa$ -HXTX-Hv1c and  $\omega$ -HXTX-Hv1a.** (A) Comparison of the primary structures of  $\kappa$ -HXTX-Hv1c and  $\omega$ -HXTX-Hv1a. Identities are boxed and shaded orange while the conserved cysteines that form the cystine-knot motif in each toxin are coloured red. (B) Richardson representation of the 3D structure of  $\kappa$ -HXTX-Hv1c (PDB accession code 1DL0; [42]) and (C)  $\omega$ -HXTX-Hv1a (PDB accession code 1AXH; [18]). Disulfide bonds are shown as orange tubes. The inter-cystine loops and the N- and C-termini are labelled. (D) Stereo-view of an overlay of the 3D structures of  $\kappa$ -HXTX-Hv1c (dark blue with light-blue disulfide bonds) and  $\omega$ -HXTX-Hv1a (raspberry with light-salmon disulfide bonds). The structural alignment, which was automatically generated by DaliLite [70], yielded a root mean square deviation of 2.4 Å over the backbone atoms of the 28 aligned residues (Ala1–Ala11, Pro15–Pro18, Ser21–Asn27, and Gly28–Arg33 in  $\kappa$ -HXTX-Hv1c versus Pro2–Pro12, Asn16–Ser19, Ser21–Asn27, and Thr32–Asp37). The inter-cystine loops and the N-terminus are labelled. Figures generated using MacPyMOL [91].

*robustus*, which were collected from geographically distinct regions of Australia (Figure 1A). Spiders were cooled to  $-20^{\circ}\text{C}$  for 40–60 min, then paired venom glands were carefully dissected from each specimen (Figure 1C). Each pair of venom glands was combined, then polyA + mRNA was extracted using a QuickPrep Micro mRNA Purification Kit (Amersham Pharmacia Biotech/GE Healthcare

Life Sciences, Rydalmere, NSW, Australia) and stored at  $-20^{\circ}\text{C}$  until further use. cDNA libraries were constructed using a Marathon cDNA Amplification Kit (Clontech Laboratories, Mount View, CA, USA). From the mRNA template, single-stranded cDNA was constructed using Superscript III reverse transcriptase (Life Technologies, Grand Island, NY, USA) and a poly(dT) anchor primer

(Echoclonanch-2, GGGCAGGT<sub>17</sub>). Second-strand synthesis was performed according to the kit specifications except the cDNA was purified using a Concert Rapid PCR Purification kit (Gibco/Life Technologies) instead of a phenol/chloroform extraction. A Marathon cDNA Amplification adaptor (Clontech Laboratories) was then ligated to the double-stranded cDNA. After overnight ligation, the sample was precipitated using 10 µl of 5% w/w glycogen, 10 µl of 3 M sodium acetate pH 5.2, and 100 µl of 100% ethanol at -20°C. The sample was subsequently washed with 80% ethanol and dried for 10 min prior to re-suspension in Tris-EDTA buffer.

Transcripts encoding ω-HXTX-Hv1a and paralogs/orthologs thereof were subsequently obtained via rapid amplification of cDNA ends (RACE) (Figure 2B) [71]. A redundant 3' PCR primer based on residues 24–31 of ω-HXTX-Hv1a (ω-HV1 5' RTTNC CRTTYTCRTTYT CYTTRAA 3') was used in conjunction with a 5' universal adaptor primer in a 5' RACE experiment designed to extract information about the upstream region of the ω-HXTX-Hv1a transcript (EchoAPI: 5' CACCCCTAA TACGACTCACTATAGG 3'). A gene-specific primer for 3' RACE was then designed based on the leader sequence obtained from the 5' RACE experiment (3' RACE primer: 5' TGCTGCAATATGAATACCGC 3'). This primer was used in combination with the Echoclonanch-2 oligo(dT) primer (5' GGGCAGGTTTTTTTTTTTTTTT TTT 3') to generate transcripts that encode a signal sequence homologous to that of ω-HXTX-Hv1a.

PCR products were extracted from agarose gels using a Gibco gel purification kit, precipitated using Pellet Paint Co-Precipitant kit (Novagen/EMD Millipore, Billerica, MA, USA), then phosphorylated with kinase in preparation for cloning. PCR products were then ligated into pSMART and transformed into *E. coli* cells (Lucigen, Middleton, WI, USA) using the Lucigen CloneSmart Blunt Cloning kit. Transformed clones were cultured for one hour in Recovery Medium, then plated with 50 µg/mL ampicillin to allow for overnight growth. PCR screening was then used to select colonies with the expected insert size for DNA sequencing. DNA sequences (and the corresponding protein sequences) were collated and analysed using Geneious Pro, version 3.8.5 [72] and signal sequence cleavage sites were predicted using SignalP, version 3.0 [73].

#### Identification of ω-HXTX-1 homologs via transcriptomics

Paired venom glands from *Hadronyche modesta* were dissected out and pooled. Total RNA was extracted using the standard TRIzol<sup>®</sup> Plus method (Invitrogen/Life Technologies) according to the manufacturer's protocol. One microgram of Total RNA was used to construct a cDNA library using the CREATOR<sup>™</sup> SMART<sup>™</sup> cDNA library construction kit (Clontech Laboratories) following

the manufacturer's protocol. Briefly, total RNA was reverse transcribed using the SMART<sup>™</sup> Moloney Murine Leukemia virus (MMLV) reverse transcriptase. Second-strand synthesis was completed using long distance polymerase chain reaction (PCR) as follows: 1 min at 95°C followed by 20 cycles of 1 min at 95°C and 6 min at 68°C. Products were then digested and size fractionated using a CHROMA SPIN-400 DEPC-H<sub>2</sub>O column (Clontech Laboratories) and then ligated into the pDNR-lib donor vector. Recombinant plasmids were electroporated into E-shot<sup>™</sup> DHB10<sup>™</sup>-T1<sup>R</sup> electro competent cells (Invitrogen/Life Technologies). 384 clones were randomly selected and sequenced by capillary electrophoresis on an Applied Biosystems 3730xl DNA analyzer (Applied Biosystems/Life Technologies) at the Brisbane node of the Australian Genome Research Facility (AGRF). Sequences were processed so vector and polyA + tails were clipped using CLC Main Work Bench (CLC-Bio), and the Blast2GO bioinformatic suite [74,75] was used to provide Gene Ontology, BLAST and domain/Interpro annotation. Signal sequence cleavage sites were predicted using SignalP, version 3.0 [73].

#### Nomenclature

In accordance with the recently introduced systematic nomenclature for naming peptide toxins from venomous animals [76], ω-ACTX-Hv1a [18] and J-ACTX-Hv1c [39] have been renamed ω-HXTX-Hv1a and κ-HXTX-Hv1a, respectively, and the various paralogs and orthologs uncovered in this study have been named accordingly. Briefly, the Greek letter denotes the molecular target of the peptide, followed by the generic name indicating the family from which the toxin is derived; in this case the abbreviation is HXTX for hexatoxin. After the generic family name, a two-letter abbreviation is used to denote the genus and species, indicated by upper and lowercase letters respectively (i.e., Hv for *H. versuta*, Hi for *H. infensa*, etc.). The name of the species is immediately followed by a numeral that helps to distinguish different toxins with similar pharmacology and this number is followed by letter that denotes the paralog number (this is based on the number of different encoded mature toxin sequences).

#### Molecular evolution analyses

A total of 73 nucleotide and 90 peptide sequences were aligned using the default settings in Geneious Pro, version 3.8.5 [72] then manually adjusted for optimal alignment prior to the following molecular evolution analyses (see Additional file 1: Figure S1 and S2).

#### Test for recombination

To overcome the effects of recombination on the phylogenetic and evolutionary interpretations [77], we employed Single Breakpoint algorithms implemented in the HyPhy package and assessed the effect of recombination on all the



toxin forms examined in this study [78,79]. When potential breakpoints were detected using the small sample Akaike information criterion (AIC), the sequences were compartmentalized or partitioned before conducting selection analyses to allow the recombining units to have distinct phylogenies (as described in [80,81]).

### Selection analyses

We evaluated the influence of natural selection on the toxins using maximum-likelihood models [82,83] implemented in CODEML of the PAML software [84]. We employed site-specific models that estimate positive selection statistically as an  $\omega$  value significantly greater than 1. We compared likelihood values for three pairs of models with different assumed  $\omega$  distributions as no *a priori* expectation exists for  $\omega$ : M0 (constant  $\omega$  rates across all sites) versus M3 (allows  $\omega$  to vary across sites within  $n$  discrete categories, where  $n \geq 3$ ); M1a (a model of neutral evolution) where all sites are assumed to be either under negative ( $\omega < 1$ ) or neutral selection ( $\omega = 1$ ) versus M2a (a model of positive selection) which in addition to the site classes mentioned for M1a, assumes a third category of sites; sites with  $\omega > 1$  (positive selection) and M7 ( $\beta$ ) versus M8 ( $\beta$  and  $\omega$ ), and models that mirror the evolutionary constraints of M1 and M2 but assume that  $\omega$  values are drawn from a  $\beta$  distribution [85]. Only if the alternative models (M3, M2a and M8: allow sites with  $\omega > 1$ ) show a better fit in a Likelihood Ratio Test (LRT) relative to their null models (M0, M1a and M7: do not allow sites  $\omega > 1$ ), are their results considered significant. LRT is estimated as twice the difference in maximum likelihood values between nested models and compared with the  $\chi^2$  distribution with the appropriate degree of freedom—the difference in the number of parameters between the two models. The Bayes empirical Bayes (BEB) approach [86] was used to identify amino acids under positive selection by calculating the posterior probabilities that a particular amino acid belongs to a given selection class (neutral, conserved, or highly variable). Sites with greater posterior probability (PP  $\geq 95\%$ ) of belonging to the ' $\omega > 1$  class' were inferred to be positively selected.

FUBAR [55-57] implemented in HyPhy [78,79] was employed to detect sites evolving under positive and negative selection. MEME [57], which is designed to overcome the drawbacks of site-specific assessments, was used to detect episodic diversifying selection. Mutations were also assessed via a complementary protein-level approach implemented in TreeSAAP [87]. An evolutionary fingerprint analysis was carried out using the ESD algorithm implemented in datamonkey [78,79,88] in order to clearly depict the proportion of sites under different regimes of selection.

Logo plots [54] showing cysteine codon bias were constructed using Geneious software, version 5.4.

### Shiva superfamily phylogenetic tree

The molecular evolutionary history of the Shiva superfamily toxins was reconstructed using Bayesian inference as implemented in MrBayes version 3.2.1 [89], using *lset rates = invgamma* with the *prset aamodelpr = mixed* command, which enables the program to optimize between the nine different amino acid substitution matrices implemented in MrBayes. WAG [90] was chosen as the best substitution matrix by the program. Tree searches were run using four Markov chains for a minimum of 10 million generations, sampling every 100th tree. 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 of the analyses reached their asymptote. 25% of the total trees sampled were discarded as burnin. The posterior probabilities for clades were established by constructing a majority rule consensus tree for all trees generated after completion of the burnin. The tree was rooted using the sequence of  $\omega$ -actinopodotoxin-Mb1a (also known as  $\omega$ -missulenatoxin-Mb1a) from the Eastern mouse spider *Missulena bradleyi* as an outgroup; this toxin also blocks  $Ca_v$  channels and it has sequence homology to the  $\omega$ -HXTXs, including conservation of the  $\omega$ -HXTX-Hv1a pharmacophore. However, although *M. bradleyi* and Australian funnel-web spiders (family Hexathelidae) both belong to the infraorder Mygalomorphae, *M. bradleyi* is a member of the Actinopodidae family.

### Structural alignment of omega and kappa hexatoxins

Atomic coordinates for  $\omega$ -HXTX-Hv1a [18] and  $\kappa$ -HXTX-Hv1c [42] were downloaded from the Protein DataBank (PDB accession codes 1AXH and 1DL0, respectively). A structural alignment of the toxins was automatically generated using DaliLite [70]. All structure figures were generated using MacPyMOL [91]. The Consurf webserver [92] was used for mapping evolutionary selection pressures on 3D homology models.

### Availability of supporting data

Nucleic acid and protein sequence alignments and their respective accession numbers can be accessed from the supplementary material along with tables relevant to the molecular evolution analyses.

### Additional file

**Additional file 1: Diversification of a single ancestral gene into a successful toxin superfamily in highly venomous Australian funnel-web spiders.**

### Abbreviations

cDNA: Complementary DNA; PolyA+: Polyadenylated RNA; mRNA: Messenger RNA; RACE: Rapid amplification of cDNA ends; PTP: Pico titre plate; AGRF: Australian genome research facility; HXTX: Omega hexatoxins;

k-HXTX: Kappa hexatoxins; *H. infensa*: *Hadrorynche infensa*; *H. modesta*: *Hadrorynche modesta*; *H. versuta*: *Hadrorynche versuta*; *H. venenata*: *Hadrorynche venenata*; *A. robustus*: *Atrax robustus*; MEME: Mixed effects model evolution; SLAC: Single likelihood ancestor counting; REL: Random effects likelihood; FEL: Fixed-effects likelihood; FUBAR: Fast unbiased approximate bayesian; PAML: Phylogenetic analysis by maximum likelihood; BEB: Bayes empirical Bayes; ASA: Accessible surface area; RAVER: Rapid Accumulation of Variations in Exposed Residues.

#### Competing interests

None of the authors have competing interests.

#### Authors' contributions

GFK and BGF proposed the research plan. DW collected spiders. DW and BLS performed venom-gland isolation, cDNA library construction, and RACE analyses. SSP analysed toxin sequences, performed nucleotide and peptide alignments, and submitted sequences to EMBL. BGF, KS, LK and AA performed the molecular evolution analyses. EABU prepared the venom-gland cDNA library from *H. modesta*. All authors were involved in data analysis. GFK, BGF, SSP, and KS wrote the manuscript. All authors read and approved the final manuscript.

#### Acknowledgements

This work was supported by the Australian Research Council (Discovery Grants DP0774245 and Grant DP1095728 to GFK) and the U.S. National Science Foundation (MCB9983242 to GFK). BGF gratefully acknowledges financial support from the Australian Research Council (ARC) and The University of Queensland. KS was funded by a PhD grant (SFRH/BD/61959/2009) from F.C.T (Fundação para a Ciência e a Tecnologia). AA was funded by the project PTDC/AAC-AMB/121301/2010 (FCOMP-01-0124-FEDER-019490) and PesT-C/MAR/LA0015/2011 from F.C.T. SSP was supported by an International Postgraduate Research Scholarship (IPRS) from The University of Queensland. EABU was supported by The University of Queensland (IPRS, UQ Centennial Scholarship, and UQ Advantage Top-Up Scholarship) and the Norwegian State Education Loans Fund. The authors thank Mr Glen Gregg for collection of *Hadrorynche venenata* specimens.

#### Author details

<sup>1</sup>Institute for Molecular Bioscience, The University of Queensland, 306 Carmody Road, St Lucia, QLD 4072, Australia. <sup>2</sup>Department of Molecular, Microbial & Structural Biology, School of Medicine, University of Connecticut, Farmington, CT 06030, USA. <sup>3</sup>Xenome, P.O. Box 1024, Indooroopilly Centre, QLD 4068, Australia. <sup>4</sup>CIMAR/CLIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas, 177, 4050-123 Porto, Portugal. <sup>5</sup>Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal. <sup>6</sup>Venom Evolution Lab, School of Biological Sciences, The University of Queensland, St Lucia, QLD 4072, Australia. <sup>7</sup>Current address: Monsanto Company, 800 N. Lindbergh Blvd, St. Louis, MO 63167, USA. <sup>8</sup>Current address: Faculty of Medicine, Health and Molecular Science-Queensland, Tropical Health Alliance, James Cook University, Cairns, Queensland, Australia. <sup>9</sup>Current address: The i3 Institute, University of Technology Sydney, Ultimo, NSW 2007, Australia.

Received: 3 December 2013 Accepted: 26 February 2014

Published: 5 March 2014

#### References

1. Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, Nevalainen TJ, Norman JA, Lewis RJ, Norton RS, Renjifo C, de la Vega RC R: **The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms.** *Annu Rev Genom Hum Genet* 2009, **10**:483–511.
2. Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG: **Complex cocktails: the evolutionary novelty of venoms.** *Trends Ecol Evol* 2013, **28**:219–229.
3. Fry BG, Wüster W, Kini RM, Brusic V, Khan A, Venkataraman D, Rooney AP: **Molecular evolution and phylogeny of elapid snake venom three-finger toxins.** *J Mol Evol* 2003, **57**:110–129.
4. Fry BG, Wüster W: **Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences.** *Mol Biol Evol* 2004, **21**:870–883.
5. Fry BG: **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* 2005, **15**:403–420.
6. Nei M, Gu X, Sitnikova T: **Evolution by the birth-and-death process in multigene families of the vertebrate immune system.** *Proc Natl Acad Sci U S A* 1997, **94**:7799–7806.
7. Olivera BM, Hillyard DR, Marsh M, Yoshikami D: **Combinatorial peptide libraries in drug design: lessons from venomous cone snails.** *Trends Biotechnol* 1995, **13**:422–426.
8. Sollod BL, Wilson D, Zhaxybayeva O, Gogarten JP, Drinkwater R, King GF: **Were arachnids the first to use combinatorial peptide libraries?** *Peptides* 2005, **26**:131–139.
9. Escoubas P: **Molecular diversification in spider venoms: a web of combinatorial peptide libraries.** *Mol Divers* 2006, **10**:545–554.
10. de la Vega RC R, Schwartz EF, Possani LD: **Mining on scorpion venom biodiversity.** *Toxicon* 2010, **56**:1155–1161.
11. King GF, Hardy MC: **Spider-venom peptides: structure, pharmacology, and potential for control of insect pests.** *Annu Rev Entomol* 2013, **58**:475–496.
12. Smith JJ, Herzig V, King GF, Alewood PF: **The insecticidal potential of venom peptides.** *Cell Mol Life Sci* 2013, **70**:3665–3693.
13. King GF, Escoubas P, Nicholson GM: **Peptide toxins that selectively target insect  $N_{av}$  and  $Ca_v$  channels.** *Channels* 2008, **2**:100–116.
14. King GF: **Venoms as a platform for human drugs: translating toxins into therapeutics.** *Expert Opin Biol Ther* 2011, **11**:1469–1484.
15. Lewis RJ, Dutertre S, Vetter I, Christie MJ: **Conus venom peptide pharmacology.** *Pharmacol Rev* 2012, **64**:259–298.
16. Escoubas P, Rash L: **Tarantulas: eight-legged pharmacists and combinatorial chemists.** *Toxicon* 2004, **43**:555–574.
17. Tedford HW, Gilles N, Menez A, Doering CJ, Zamponi GW, King GF: **Scanning mutagenesis of  $\omega$ -atracotoxin-Hv1a reveals a spatially restricted epitope that confers selective activity against insect calcium channels.** *J Biol Chem* 2004, **279**:44133–44140.
18. Fletcher JI, Smith R, O'Donoghue SI, Nilges M, Connor M, Howden MEH, Christie MJ, King GF: **The structure of a novel insecticidal neurotoxin,  $\omega$ -atracotoxin-HV1, from the venom of an Australian funnel web spider.** *Nat Struct Biol* 1997, **4**:559–566.
19. Wang X, Smith R, Fletcher JI, Wilson H, Wood CJ, Howden ME, King GF: **Structure-function studies of  $\omega$ -atracotoxin, a potent antagonist of insect voltage-gated calcium channels.** *Eur J Biochem* 1999, **264**:488–494.
20. Chong Y, Hayes JL, Sollod B, Wen S, Wilson DT, Hains PG, Hodgson WC, Broady KW, King GF, Nicholson GM: **The  $\omega$ -atracotoxins: selective blockers of Insect M-LVA and HVA calcium channels.** *Biochem Pharmacol* 2007, **74**:623–638.
21. Tedford HW, Fletcher JI, King GF: **Functional significance of the  $\beta$ -hairpin in the insecticidal neurotoxin  $\omega$ -atracotoxin-Hv1a.** *J Biol Chem* 2001, **276**:26568–26576.
22. Tedford HW, Maggio F, Reenan RA, King GF: **A model genetic system for testing the in vivo function of peptide toxins.** *Peptides* 2007, **28**:51–56.
23. Khan SA, Zafar Y, Briddon RW, Malik KA, Mukhtar Z: **Spider venom toxin protects plants from insect attack.** *Transgenic Res* 2006, **15**:349–357.
24. Diniz MR, Paine MJ, Diniz CR, Theakston RD, Crampton JM: **Sequence of the cDNA coding for the lethal neurotoxin Tx1 from the Brazilian "armed" spider *Phoneutria nigriventer* predicts the synthesis and processing of a preprotoxin.** *J Biol Chem* 1993, **268**:15340–15342.
25. Krapcho KJ, Kral RM, Vanwagenen BC, Eppler KG, Morgan TK: **Characterization and cloning of insecticidal peptides from the primitive weaving spider *Diguetia canities*.** *Insect Biochem Mol Biol* 1995, **25**:991–1000.
26. Wang X-H, Connor M, Wilson D, Wilson HI, Nicholson GM, Smith R, Shaw D, Mackay JP, Alewood PF, Christie MJ, King GF: **Discovery and structure of a potent and highly specific blocker of insect calcium channels.** *J Biol Chem* 2001, **276**:40806–40812.
27. Ostrow KL, Mammoser A, Suchyna T, Sachs F, Oswald R, Kubo S, Chino N, Gottlieb PA: **cDNA sequence and in vitro folding of GsMTx4, a specific peptide inhibitor of mechanosensitive channels.** *Toxicon* 2003, **42**:263–274.
28. Liang S: **An overview of peptide toxins from the venom of the Chinese bird spider *Selenocosmia huwena* Wang [*Ornithoctonus huwena* (Wang)].** *Toxicon* 2004, **43**:575–585.
29. Hook V, Funkelstein L, Lu D, Bark S, Wegrzyn J, Hwang S-R: **Proteases for processing proneuropeptides into peptide neurotransmitters and hormones.** *Annu Rev Pharmacol Toxicol* 2008, **48**:393–423.

30. Veenstra JA: Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. *Arch Insect Biochem Physiol* 2000, **43**:49–63.
31. Kozlov S, Malyavka A, McCutchen B, Lu A, Schepers E, Herrmann R, Grishin E: A novel strategy for the identification of toxinlike structures in spider venom. *Proteins* 2005, **59**:131–140.
32. Chen J, Deng M, He Q, Meng E, Jiang L, Liao Z, Rong M, Liang S: Molecular diversity and evolution of cystine knot toxins of the tarantula *Chilobrachys jingzhao*. *Cell Mol Life Sci* 2008, **65**:2431–2444.
33. Wong ES, Hardy MC, Wood D, Bailey T, King GF: SVM-based prediction of propeptide cleavage sites in spider toxins identifies toxin innovation in an Australian tarantula. *PLoS One* 2013, **8**:e66279.
34. Raven RJ: The spider infraorder Mygalomorphae (Araneae): cladistics and systematics. *Bull Am Mus Nat Hist* 1986, **182**:1–180.
35. Tedford HW, Sollod BL, Maggio F, King GF: Australian funnel-web spiders: master insecticide chemists. *Toxicon* 2004, **43**:601–618.
36. Gray MR: A revision of the Australian funnel-web spiders (Hexathelidae: Atracinae). *Records Austr Mus* 2010, **62**:285–392.
37. Mukherjee AK, Sollod BL, Wikel SK, King GF: Orally active acaricidal peptide toxins from spider venom. *Toxicon* 2006, **47**:182–187.
38. King GF, Sollod-McFarland B, Nicholson GM, Gunning SJ: *Insecticidal Polypeptides and Methods of use Thereof* U.S. Patent No. 7,354,993; 2008.
39. Wang X-H, Connor M, Smith R, Maciejewski MW, Howden MEH, Nicholson GM, Christie MJ, King GF: Discovery and characterization of a family of insecticidal neurotoxins with a rare vicinal disulfide bond. *Nat Struct Biol* 2000, **7**:505–513.
40. Pallaghy PK, Nielsen KJ, Craik DJ, Norton RS: A common structural motif incorporating a cystine knot and a triple-stranded  $\beta$ -sheet in toxic and inhibitory polypeptides. *Protein Sci* 1994, **3**:1833–1839.
41. King GF, Tedford HW, Maggio F: Structure and function of insecticidal neurotoxins from Australian funnel-web spiders. *J Toxicol-Toxin Rev* 2002, **21**:359–389.
42. Wang G, Peterkofsky A, Clore GM: A novel membrane anchor function for the N-terminal amphipathic sequence of the signal-transducing protein IIA<sup>Glucose</sup> of the *Escherichia coli* phosphotransferase system. *J Biol Chem* 2000, **275**:39811–39814.
43. Maggio F, King GF: Scanning mutagenesis of a Janus-faced atracotoxin reveals a bipartite surface patch that is essential for neurotoxic function. *J Biol Chem* 2002, **277**:22806–22813.
44. Mobli M, de Araujo AD, Lambert LK, Pierens GK, Windley MJ, Nicholson GM, Alewood PF, King GF: Direct visualization of disulfide bonds through diselenide proxies using <sup>77</sup>Se NMR spectroscopy. *Angew Chem Int Ed* 2009, **48**:9312–9314.
45. Gunning SJ, Maggio F, Windley MJ, Valenzuela SM, King GF, Nicholson GM: The Janus-faced atracotoxins are specific blockers of invertebrate K<sub>Ca</sub> channels. *FEBS J* 2008, **275**:4045–4059.
46. Woodward SR, Cruz LJ, Olivera BM, Hillyard DR: Constant and hypervariable regions in conotoxin propeptides. *EMBO J* 1990, **9**:1015–1020.
47. Olivera BM, Walker C, Cartier GE, Hooper D, Santos AD, Schoenfeld R, Shetty R, Watkins M, Bandyopadhyay P, Hillyard DR: Speciation of cone snails and interspecific hyperdivergence of their venom peptides potential evolutionary significance of introns. *Ann NY Acad Sci* 1999, **870**:223–237.
48. Duda TF Jr, Palumbi SR: Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc Natl Acad Sci U S A* 1999, **96**:6820–6823.
49. Conticello SG, Pilpel Y, Glusman G, Fainzilber M: Position-specific codon conservation in hypervariable gene families. *Trends Genet* 2000, **16**:57–59.
50. Conticello SG, Gilad Y, Avidan N, Ben-Asher E, Levy Z, Fainzilber M: Mechanisms for evolving hypervariability: the case of conopeptides. *Mol Biol Evol* 2001, **18**:120–131.
51. Espiritu DJD, Watkins M, Dia-Monje V, Cartier GE, Cruz LJ, Olivera BM: Venomous cone snails: molecular phylogeny and the generation of toxin diversity. *Toxicon* 2001, **39**:1899–1916.
52. Fricker LD: Neuropeptide-processing enzymes: applications for drug discovery. *AAPS J* 2005, **7**:E449–E455.
53. Saez NJ, Senff S, Jensen JE, Er SY, Herzig V, Rash LD, King GF: Spider-venom peptides as therapeutics. *Toxins* 2010, **2**:2851–2871.
54. Schneider TD, Stephens RM: Sequence logos: a new way to display consensus sequences. *Nucl Acids Res* 1990, **18**:6097–6100.
55. Kosakovsky Pond SL, Frost SDW: Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 2005, **22**:1208–1222.
56. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K: FUBAR: a Fast, Unconstrained Bayesian Approximation for inferring selection. *Mol Biol Evol* 2013, **30**:1196–1205.
57. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL: Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 2012, **8**:e1002764.
58. Kozminsky-Atias A, Zilberberg N: Molding the business end of neurotoxins by diversifying evolution. *FASEB J* 2012, **26**:576–586.
59. Sunagar K, Jackson TNW, Undheim EAB, Ali SA, Antunes A, Fry BG: Three-fingered RAVERS: rapid accumulation of variation in exposed residues of snake venom toxins. *Toxins* 2013, **5**:2172–2208.
60. Kini RM, Chan YM: Accelerated evolution and molecular surface of venom phospholipase A<sub>2</sub> Enzymes. *J Mol Evol* 1999, **48**:125–132.
61. Brust A, Sunagar K, Undheim EAB, Vetter I, Yang DC, Casewell NR, Jackson TNW, Koludarov I, Alewood PF, Hodgson WC, Lewis RJ, King GF, Antunes A, Hendriks I, Fry BG: Differential evolution and neofunctionalization of snake venom metalloprotease domains. *Mol Cell Proteomics* 2013, **12**:651–663.
62. Ruder T, Sunagar K, Undheim EAB, Ali SA, Wai T-C, Low DHW, Jackson TNW, King GF, Antunes A, Fry BG: Molecular phylogeny and evolution of the proteins encoded by coleoid (cuttlefish, octopus, and squid) posterior venom glands. *J Mol Evol* 2013, **76**:192–204.
63. Low DHW, Sunagar K, Undheim EAB, Ali SA, Alagon AC, Ruder T, Jackson TNW, Pineda Gonzalez S, King GF, Jones Antunes A, Fry BG: Dracula's children: molecular evolution of vampire bat venom. *J Proteomics* 2013, **89**:95–111.
64. Sunagar K, Fry BG, Jackson TNW, Casewell NR, Undheim EAB, Vidal N, Ali SA, King GF, Vasudevan K, Vasconcelos V, Antunes A: Molecular evolution of vertebrate neurotoxins: co-option of the highly conserved nerve growth factor gene into the advance snake arsenal. *PLoS One* 2013, **8**:e81827.
65. Nisani Z, Dunbar SG, Hayes WK: Cost of venom regeneration in *Parabuthus transvaalicus* (Arachnida: Buthidae). *Comp Biochem Physiol A-Mol Integr Physiol* 2007, **147**:509–513.
66. Nisani Z, Boskovic DS, Dunbar SG, Kelln W, Hayes WK: Investigating the chemical profile of regenerated scorpion (*Parabuthus transvaalicus*) venom in relation to metabolic cost and toxicity. *Toxicon* 2012, **60**:315–323.
67. Morgenstern D, King GF: The venom optimization hypothesis revisited. *Toxicon* 2013, **63**:120–128.
68. de Araujo AD, Herzig V, Windley MJ, Dziemborowicz S, Mobli M, Nicholson GM, Alewood PF, King GF: Do vicinal disulfide bridges mediate functionally important redox transformations in proteins? *Antioxid Redox Signal* 2013, **19**:1976–1980.
69. Maggio F, King GF: Role of the structurally disordered N- and C-terminal residues in the Janus-faced atracotoxins. *Toxicon* 2002, **40**:1355–1361.
70. Holm L, Kääriäinen S, Rosenström P, Schenkel A: Searching protein structure databases with DALI-Lite v.3. *Bioinformatics* 2008, **24**:2780–2781.
71. Frohman MA: Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. *Methods Enzymol* 1993, **218**:340–356.
72. Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir S, Stones-Havas S, Sturrock S, Thierer T, Wilson A: Geneious v5.4. 2011 2011, 2011:2011. Available from <http://www.geneious.com>.
73. Bendtsen JD, Nielsen H, Von Heijne G, Brunak S: Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004, **340**:783–795.
74. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A: High-throughput functional annotation and data mining with the Blast2GO suite. *Nucl Acids Res* 2008, **36**:3420–3435.
75. Götz S, Arnold R, Sebastián-León P, Martín-Rodríguez S, Tischler P, Jehl M-A, Dopazo J, Rattei T, Conesa A: B2G-FAR, a species-centered GO annotation repository. *Bioinformatics* 2011, **27**:919–924.
76. King GF, Gentz MC, Escoubas P, Nicholson GM: A rational nomenclature for naming peptide toxins from spiders and other venomous animals. *Toxicon* 2008, **52**:264–276.
77. Posada D, Crandall KA: The effect of recombination on the accuracy of phylogeny estimation. *J Mol Evol* 2002, **54**:396–402.
78. Kosakovsky Pond SL, Frost SDW: Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 2005, **21**:2531–2533.

79. Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL: **Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology.** *Bioinformatics* 2010, **26**:2455–2457.
80. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW: **Automated phylogenetic detection of recombination using a genetic algorithm.** *Mol Biol Evol* 2006, **23**:1891–1901.
81. Scheffler K, Martin DP, Seoighe C: **Robust inference of positive selection from recombining coding sequences.** *Bioinformatics* 2006, **22**:2493–2499.
82. Goldman N, Yang Z: **A codon-based model of nucleotide substitution for protein-coding DNA sequences.** *Mol Biol Evol* 1994, **11**:725–736.
83. Yang Z: **Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution.** *Mol Biol Evol* 1998, **15**:568–573.
84. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood.** *Mol Biol Evol* 2007, **24**:1586–1591.
85. Nielsen R, Yang Z: **Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene.** *Genetics* 1998, **148**:929–936.
86. Yang Z, Wong WSW, Nielsen R: **Bayes Empirical Bayes inference of amino acid sites under positive selection.** *Mol Biol Evol* 2005, **22**:1107–1118.
87. Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA: **TreeSAAP: selection on amino acid properties using phylogenetic trees.** *Bioinformatics* 2003, **19**:671–672.
88. Kosakovsky Pond SL, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler K: **A random effects branch-site model for detecting episodic diversifying selection.** *Mol Biol Evol* 2011, **28**:3033–3043.
89. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP: **MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space.** *Systematic Biol* 2012, **61**:539–542.
90. Whelan S, Goldman N: **A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach.** *Mol Biol Evol* 2001, **18**:691–699.
91. DeLano WL: *The PyMOL Molecular Graphics System.* San Carlos, CA, USA: DeLano Scientific; 2002.
92. Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N: **ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information.** *Bioinformatics* 2003, **19**:163–164.

doi:10.1186/1471-2164-15-177

**Cite this article as:** Pineda et al.: Diversification of a single ancestral gene into a successful toxin superfamily in highly venomous Australian funnel-web spiders. *BMC Genomics* 2014 **15**:177.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

