Genomic, functional and structural analyses elucidate evolutionary innovation within the sea anemone 8 toxin family

Lauren M. Ashwood[1,2*†] , Khaled A. Elnahriry[3†], Zachary K. Stewart[4], Thomas Shafee[5,6], Muhammad Umair Naseem[7], Tibor G. Szanto[7], Chloé A. van der Burg[1,8], Hayden L. Smith[1], Joachim M. Surm[9], Eivind A. B. Undheim[10,11], Bruno Madio[12], Brett R. Hamilton[11,13], Shaodong Guo[12], Dorothy C. C. Wai[3], Victoria L. Coyne[14], Matthew J. Phillips[1], Kevin J. Dudley[1,14], David A. Hurwood[1,4], Gyorgy Panyi[7], Glenn F. King[12,15], Ana Pavasovic[16], Raymond S. Norton[3,17] and Peter J. Prentis[1,4]

Abstract

Background The ShK toxin from *Stichodactyla helianthus* has established the therapeutic potential of sea anemone venom peptides, but many lineage-specific toxin families in Actiniarians remain uncharacterised. One such peptide family, sea anemone 8 (SA8), is present in all five sea anemone superfamilies. We explored the genomic arrangement and evolution of the SA8 gene family in *Actinia tenebrosa* and *Telmatactis stephensoni*, characterised the expression patterns of SA8 sequences, and examined the structure and function of SA8 from the venom of *T. stephensoni*.

Results We identified ten SA8-family genes in two clusters and six SA8-family genes in five clusters for *T. stephensoni* and *A. tenebrosa*, respectively. Nine SA8 *T. stephensoni* genes were found in a single cluster, and an SA8 peptide encoded by an inverted SA8 gene from this cluster was recruited to venom. We show that SA8 genes in both species are expressed in a tissue-specific manner and the inverted SA8 gene has a unique tissue distribution. While the functional activity of the SA8 putative toxin encoded by the inverted gene was inconclusive, its tissue localisation is similar to toxins used for predator deterrence. We demonstrate that, although mature SA8 putative toxins have similar cysteine spacing to ShK, SA8 peptides are distinct from ShK peptides based on structure and disulfide connectivity.

Conclusions Our results provide the first demonstration that SA8 is a unique gene family in Actiniarians, evolving through a variety of structural changes including tandem and proximal gene duplication and an inversion event that together allowed SA8 to be recruited into the venom of *T. stephensoni*.

Keywords Sea anemone, Toxin evolution, Genome, Neofunctionalization, Peptide synthesis, Disulfide connectivity

†Lauren M. Ashwood and Khaled A. Elnahriry contributed equally to this work.

*Correspondence:
Lauren M. Ashwood
lauren.ashwood@qimrberghofer.edu.au
Full list of author information is available at the end of the article

Ashwood *et al. BMC Biology*   (2023) 21:121

Page 2 of 25

Background

Sea anemones (order Actiniaria) rely on venom for their survival, deploying this chemical weapon during encounters with prey, predators and competitors [1]. In contrast to other studied cnidarian groups, however, the venom of Actiniarians is dominated by peptide neurotoxins [2, 3]. These neurotoxins confer substantial clinical utility and have been utilised as molecular probes in studies of the nervous system for more than 40 years [4]. Characterised sea anemone neurotoxins can be classified into one of 12 pharmacological toxin families, based on which receptors and ion channels they target [5–8]. Sea anemone neurotoxins, however, can also be classified according to their structural scaffold [9]. For example, the six sea anemone potassium channel toxin families are characterised by different structural scaffolds [9]. Furthermore, these structural scaffolds are not always associated with a single toxin family. The inhibitor cystine knot (ICK) fold is among the most widely recruited motifs in animal venoms [10, 11] and cysteine spacing characteristic of the ICK motif has been identified in both sea anemone type 5 potassium channel toxins and acid-sensing ion channel toxins [9, 12]. Similarly, the β-defensin fold is observed in multiple sea anemone neurotoxin families [9]. Thus, much of the pharmacological diversity of venoms, including those of sea anemones, is often accounted for by a relatively small number of peptide scaffolds. Reflecting their amenability to functional innovation, these scaffolds—sometimes referred to as *privileged scaffolds*—have been recruited to venoms on numerous occasions, including within the same lineage.

ShK represents another privileged scaffold, with 7,329 proteins containing 13,829 ShK-like (ShKT) domains spanning multiple kingdoms documented in the Simple Modular Architecture Research Tool (SMART) database [6, 13]. In Actiniarians, however, the ShK fold has only been observed within toxins belonging to the type 1 potassium channel toxin family [6, 9]. This motif is named after ShK, a 35-residue toxin isolated from the sea anemone, *Stichodactyla helianthus*, which potently blocks $K_V1.3$ as well as other voltage-gated potassium channels [14–18]. An analogue of ShK (dalazatide, Shk-186) has shown remarkable efficacy in preclinical animal models of autoimmune disease [15, 19–21] and entered phase 1 clinical trials for the treatment of plaque psoriasis [22]. Because of the therapeutic potential of ShK, sea anemone proteins containing ShK-like domains have received substantial research attention [23–27], but not all ShK-like peptides block ion channels. More recently, it was reported that multiple toxin-like peptides with the cysteine spacing characteristic of ShK are localised to neurons rather than venom-related structures in the starlet sea anemone, *Nematostella vectensis* [28]. One of

these neuropeptides was a homologue of U-actitoxin-Avd8e, a member of the sea anemone 8 (SA8) toxin family [28].

SA8 putative toxins were first identified computationally in *Anemonia viridis* [29] and have since been noted in several other taxa [30–33]. The SA8 gene family appears to be restricted to the order Actiniaria and thus far it has been identified in 14 sea anemone species, spanning the Actinioidea, Metridioidea and Edwardsioidea superfamilies [33]. Five peptides with a conserved motif and relatively low similarity to known actiniarian toxins were identified in the *A. viridis* expressed sequence tag (EST) database [29], and together form the curated sea anemone 8 toxin family [5]. One of these SA8 precursors (P0DMZ3) was associated with 103 identical sequences, placing it as the most abundant novel toxin-like transcript in *A. viridis* [29]. Since then it has been demonstrated that SA8 transcripts are highly expressed in *Stichodactyla haddoni* [30] and upregulated in the acrorhagi and tentacles of *Actinia tenebrosa* [33]. Despite this, peptides from the SA8 putative toxin family have not been functionally validated as toxins and the tertiary structure of this family has not been resolved.

Given the recent discovery of the ShKT cysteine motif in a *N. vectensis* SA8-like peptide, as well as the ongoing recruitment of neuropeptides into sea anemone venom [28], SA8 putative toxins could potentially function as toxins and/or neuropeptides in different species. Here, we characterise the genomic arrangement, evolution and expression patterns of SA8 genes in two species, *A. tenebrosa* (superfamily Actinioidea) and *Telmatactis stephensoni* (superfamily Metridioidea) using a combined proteomic, transcriptomic and genomic approach. Furthermore, we investigated the structure and function of the first reported SA8 peptide recruited to the venom of *T. stephensoni*. Finally, we examined whether the SA8 gene family can be distinguished from the ShK gene family or whether it represents an extension of this gene family.

Results and discussion

Genome assemblies of *Actinia tenebrosa* and *Telmatactis stephensoni*

The genome assembly for *A. tenebrosa* generated a 312-Mb assembly scaffolded into 18 pseudomolecules with a N50 of 16.5 Mb, with 38 contigs not assigned to scaffolds (Table 1). The genome assembly for *T. stephensoni* resulted in a 484-Mb assembly with a contig N50 of 602 Kb. Completeness scores (BUSCO: Benchmarking Universal Single-Copy Orthologs) for *A. tenebrosa* and *T. stephensoni* assemblies are on par with other long-read actiniarian genomes [34–37]. In contrast, contig N50 values for the current *A. tenebrosa* and *T. stephensoni*

Table 1 Comparison of assembly metrics among actiniarian genomes

Annotation metrics	TELST	ACTTE	ACTEQ	ACTTE[a]	NEMVE	EXADI
Assembly size (Mbp)	484	312	409	238	356	256
Contig N50 (kbp)	602	716	492	188	472	442
Scaffold N50 (kbp)	-	16,458	-	-	-	-
Number of chromosomes	-	18	-	-	-	-
Percent repetitive DNA	27.3	32.3	-	19.57	26	26
BUSCO (%)	93.8	93.5	94.0	89.6	91.6	87.3

Abbreviations: TELST T. stephensoni, ACTTE A. tenebrosa; ACTTE[a], draft A. tenebrosa [38], ACTEQ Actinia equina [34], NEMVE N. vectensis [36], EXADI Exaiptasia diaphana [35]

genomes are higher than both previously assembled short- and long-read genomes.

Repeat content analysis of the *A. tenebrosa,* and *T. stephensoni* genomes found that 32.3% (101 Mb) and 27.3% (122 Mb), respectively, of these assemblies are repetitive. Similar proportions of repetitive DNA were observed in the *N. vectensis* and *Exaiptasia diaphana* genome assemblies [35, 36]. Long terminal repeats (LTRs), miniature inverted-repeat terminal elements (MITEs) and unknown repeats were the most observed repeat classes found in both sea anemone species, but the relative abundance of each class varied between species (Additional file 1: Table S1). LTRs were the most abundant repeat type found in the *A. tenebrosa* assembly, followed by MITEs and unknown repeats. In contrast, Surm et al. [38] reported that MITEs were the most common repeat type in the draft genome of *A. tenebrosa*, followed by unknown repeats. MITEs were the most abundant repeat class observed in the *T. stephensoni* assembly followed by LTRs and unknown repeats.

For *A. tenebrosa*, we predicted 53,560 protein-coding genes (59,465 with isoforms included) with a BUSCO genome completeness score of 97.3%. We predicted 50,712 protein-coding genes (54,933 with isoforms included) in *T. stephensoni* with a BUSCO genome completeness score of 96.6%. Based on these BUSCO scores, the completeness of our assemblies is comparable to that of the *A. equina* genome (97%) [34]. We predicted twice as many genes compared to previous sea anemone genome assemblies [34–38]. Furthermore, we predicted more protein-coding genes in *A. tenebrosa* compared to the other assembly for this species as Surm et al. [38] produced a draft genome using short read sequences only. This resulted in a significant underestimation of gene number in the previous *A. tenebrosa* genome assembly.

Expansion of gene families containing putative toxins within *T. stephensoni*

The phylogenetic tree generated for the five species is consistent with previous phylogenies [39], with *N. vectensis* sister to a clade containing *A. tenebrosa* and the three Metridioidean species. Computational Analysis of gene Family Evolution (CAFE) was then used to model the evolution of gene family sizes across five actiniarian species spanning the superfamilies Actinioidea, Edwardsioidea and Metridioidea (Fig. 1). This analysis revealed *T. stephensoni* had the largest number of expansions and the fewest contractions of gene families. In fact, all three Metridioidean species had a greater number of gene family expansions compared to *A. tenebrosa* and *N. vectensis.* Gene families that encode putative toxins were common in the expanded gene families, with 21 expanded toxin gene families in *T. stephensoni* compared to 15, 10, 4 and 2 for *E. diaphana*, *Paraphelliactis xishaensis*, *A. tenebrosa* and *N. vectensis*, respectively (Additional file 2: Table S2)*.* Orthogroups containing ShK-like, peptidase M12A, phospholipase A2 (PLA$_2$) and other putative toxins were expanded in *T. stephensoni*, while orthogroups containing insulin-like growth factor binding protein (IGFBP) like and Factor V-like putative toxins were expanded in *A. tenebrosa.* Furthermore, CAFE results detected an expansion of the SA8 family only in *Telmatactis,* a pattern also observed by Surm et al. [33], with a contraction of this gene family occurring in *N. vectensis*. Thus, the SA8 family appears to have a unique expansion event in *Telmatactis* and the potential mechanisms for this expansion should be examined.

Size and sequence divergence within the SA8 family coincides with phylogeny of actiniarian superfamilies

Multiple SA8 genes were identified in the genomes of *A. tenebrosa* and *T. stephensoni*, but only a single SA8 gene was identified and manually annotated in the most recent *N. vectensis* genome [37]. This single *N. vectensis* SA8 gene was comprised of three regular exons, ranging from 29 to > 120 bp. The number of genes and their arrangement also differed significantly between *A. tenebrosa* and *T. stephensoni* (Fig. 2A). In *T. stephensoni*, 10 SA8 genes were identified, nine of which were tandemly clustered on a single contig. In contrast, only six SA8 genes were identified in the *A. tenebrosa* genome on multiple pseudomolecules. Additionally, we noted distinct differences
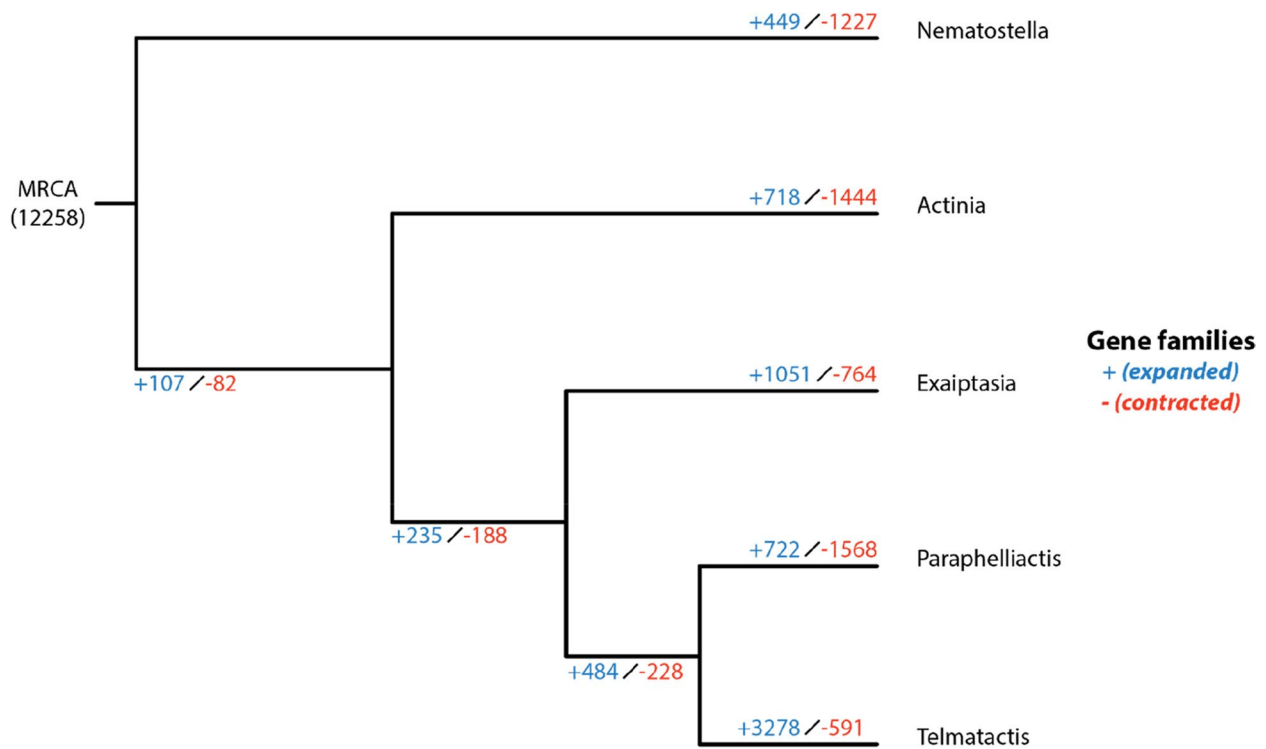
Fig. 1 CAFE expansion/contraction tree highlights increased number of expanded gene families *in T. stephensoni* relative to other sea anemone species. There were 12,258 gene families found in more than one of the five sea anemone species that CAFE predicts were present in their most recent common ancestor (MRCA). Species: Nematostella, *N. vectensis*; Actinia, *A. tenebrosa*; Exaiptasia, *E. diaphana*; Paraphelliactis, *P. xishaensis*; Telmatactis, *T. stephensoni*
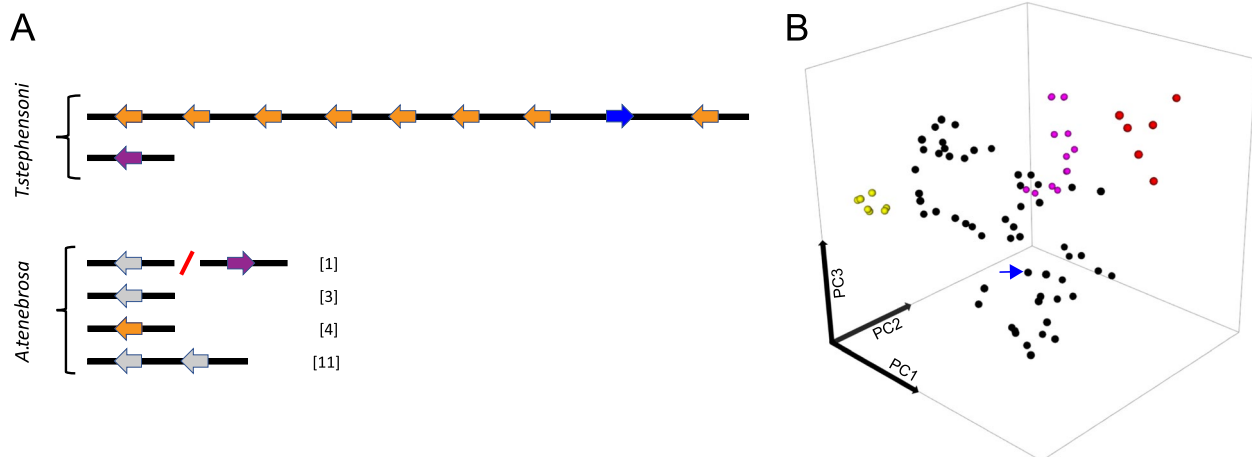
Fig. 2 A Genomic arrangement and orientation of SA8 genes in *T. stephensoni* and *A. tenebrosa*. The inverted SA8 gene identified in the genome of *T. stephensoni* is depicted in blue. The first collinear block is depicted in orange while the second collinear block is shown in purple. Numbers in square brackets indicate the scaffold number for *A. tenebrosa* SA8 genes. **B** Sequence space of the SA8 family using 102 SA8 sequences from 14 actiniarian species spanning three superfamilies. Sequence space analysis of mature SA8 peptides reveals that they fall into four clusters (yellow, black, pink, red), with the *T. stephensoni* venom SA8 peptide found in the largest cluster (indicated by blue arrow)

in the genomic arrangement of the SA8 gene family between these two species. The six *A. tenebrosa* genes were located across four scaffolds, with multiple SA8 genes present on both scaffolds 1 and 11. The two SA8 genes on scaffold 11 were arranged tandemly, but in scaffold 1, the SA8 genes were 1.5 Mb apart and had different

orientations (non-tandem). MCScanX analysis indicated that the cluster of nine SA8 genes in *T. stephensoni* originated from two episodes of tandem duplication involving eight genes collectively, with the remaining clustered SA8 gene originating from a proximal duplication event. The non-clustered *T. stephensoni* SA8 gene appears to have originated from a dispersed duplication event. Similarly, four of the *A. tenebrosa* SA8 genes originated from dispersed duplication events. However, there was no evidence of tandem duplication of SA8 genes in this species, with the two clustered SA8 genes on scaffold 11 arising from proximal duplication. Instead, a genome doubling event in the genus *Actinia* [34] may have resulted in the distribution of SA8 genes across double the number of chromosomes compared to *T. stephensoni*.

Two separate instances of collinearity were detected between *T. stephensoni* and *A. tenebrosa* SA8 genes. The first of these collinear gene blocks included the *A. tenebrosa* SA8 gene and surrounding genes on scaffold 4 and the tandemly clustered *T. stephensoni* SA8 genes (orange arrows in Fig. 2A). The second SA8 gene on scaffold 1 of *A. tenebrosa* and the non-clustered *T. stephensoni* SA8 gene were the second instance of SA8 and other homologous genes evolving in a conserved order across both species (purple arrows in Fig. 2A). This indicates that collinearity can be observed in a subset of SA8 loci and suggests whole genome and/or segmental duplication has played a role in the evolution of the SA8 gene family in *A. tenebrosa*, with genes located on double the number of chromosomes.

Consistent with a recent expansion in *T. stephensoni*, a high degree of sequence similarity and conserved gene structure present can be observed in *T. stephensoni*, but not *A. tenebrosa*, SA8 genes. While SA8 genes are comprised of 3–5 microexons in both species, high levels of variability are observed in the second exon in *A. tenebrosa* only. In *T. stephensoni*, there is high sequence similarity between the clustered SA8 genes, with the greatest sequence divergence associated with the non-clustered SA8 gene. Additionally, sequence space analysis revealed that SA8 peptides cluster into four groups across sea anemones (Fig. 2B). SA8 peptides from *A. tenebrosa* were found in all four groups while most SA8 peptides from *T. stephensoni* (including a venom peptide that mapped to the inverted SA8 gene; referred to hereafter as venom SA8 peptide) were confined to a single group. The clustering patterns observed for *A. tenebrosa* and *T. stephensoni* sequences are reinforced in other species of Actinioidea and Metridioidea. Actinioidea species *Anthopleura dowii*, *Anthopleura buddemeieri*, *S. haddoni* and *Aulactinia veratra* have SA8 peptides distributed across all four clusters. Conversely, SA8 peptides from Metridioidea species *E. diaphana*, *Nemanthus*

annamensis and *Calliactis polypus* are restricted to one cluster, as are SA8 peptides identified in *Edwardsia carnea* and *N. vectensis* of superfamily Edwardsioidea. Therefore, despite the high copy numbers of SA8 genes in *T. stephensoni*, levels of sequence variation are lower in Metridioidea relative to Actinioidea. This finding favours the hypothesis that clustered SA8 genes observed in the *T. stephensoni* genome can be explained by a recent tandem duplication event and provides support for the hypothesis that tandem duplication may have played a key role in the expansion of the SA8 family across the superfamily Metridioidea; however, further investigation using other actiniarian genomes will be needed to verify this.

Gene duplication has been shown to play a major role in the expansion of protein families in most venomous phyla. The snake venom metalloproteinase (SVMP) family in rattlesnakes originated from a single ancestral gene via a series of gene duplication events, resulting in 31 tandem SVMP genes in *Crotalus atrox* [40]. Similarly, venom PLA_2 genes cluster together in the genomes of multiple rattlesnake species, however, the number of genes in the cluster differs among *C. scutulatus*, *C. atrox*, and *C. adamanteus* [41]. A comparable pattern may emerge as SA8 genes are annotated in the genome of other Metridioidea species. As with SA8 clustered genes in *T. stephensoni*, the exons of clustered venom PLA_2 genes in rattlesnakes are characterised by a conserved length and low levels of intron sequence variability [41]. Furthermore, tandem duplications are responsible for the evolution of defensin-like peptides in platypus venom [42–45]. Within order Actiniaria, a cluster of twelve genes encoding the Nv_1 toxin were identified in the genome of *N. vectensis* [46]. These clustered Nv_1 genes were proposed to have evolved via a birth and death mechanism [46, 47], a pattern consistent with the levels of nucleotide diversity found in the *T. stephensoni* SA8 genes. Furthermore, expansion events were reported to be more common in actiniarian-specific gene families — such as SA8 — compared to shared gene families [38], and tandem duplication was recently implicated in the evolution of the actiniarian 'dominant toxin' families [48]. Thus, there is a strong precedence for the expansion of the toxin gene families via duplication — both within sea anemones and other venomous lineages — and our data support the involvement of duplication events in the expansion and evolution of the SA8 family in *T. stephensoni*.

SA8 peptides are associated with one of two expression patterns in *T. stephensoni*

The tissue distribution patterns of SA8 genes in *T. stephensoni* and *A. tenebrosa* were examined using several approaches. Differential gene expression analysis utilising

Ashwood *et al. BMC Biology*   (2023) 21:121

Page 6 of 25

previously generated tissue-specific reads for three envenomating structures of *A. tenebrosa* and three functional regions of *T. stephensoni* revealed distinct expression patterns of SA8 genes in each species. The tentacles and club-tips facilitate prey capture and immobilisation prior to transport through the actinopharynx to the mesenterial filaments for prey killing and digestion [49, 50]. In contrast, the epidermis (body column and pedal disc) likely have a key role in predator deterrence, with the pedal disc also mediating substrate attachment [32]. Acrorrhagi are inflatable organs used during intraspecific competition and are only observed in Actiniidae species [51].

Only four of the six *A. tenebrosa* SA8 genes were differentially expressed (false discovery rate (FDR) ≤ 0.05, four-fold) across the structures examined. Of these four differentially expressed *A. tenebrosa* SA8 genes (Fig. 3A), two were upregulated in the acrorrhagi (SA8_scaffold1_1 and SA8_scaffold3), one was upregulated in the mesenterial filaments (SA8_scaffold11_1), and one was upregulated in the tentacles (SA8_scaffold1_2). Conversely, in *T. stephensoni*, the nine transcripts corresponding to the 10 SA8 genes are differentially expressed across the tentacles (club-tips and tentacles), gastrodermis (actinopharynx and mesenterial filaments) and epidermis (body column and pedal disc) (Fig. 3B). The majority of *T. stephensoni* SA8 genes are upregulated in the epidermis (four transcripts, five genes; SA8 clustered genes 1, 2, 6, 7 and 8) or the tentacles (three transcripts/genes; SA8 clustered genes 3, 4 and 5), although the non-clustered SA8 gene is upregulated in both the epidermis and tentacles.

Co-expression toxin module membership of all transcripts with significant BLASTp matches to SA8 peptides was assessed using weighted gene co-expression network analysis (WGCNA). In *T. stephensoni*, 11 transcripts (from eight genes) of the 15 transcripts (from ten SA8 genes) are assigned to two modules, each with a characteristic tissue-specific expression pattern (Fig. 4A); the turquoise module is characterised by upregulated expression in the tentacle tissue while the brown module is composed of putative toxins upregulated in column and pedal disc tissue. In *A. tenebrosa* SA8 transcripts are found in all four of the network modules identified (blue, brown, turquoise and yellow; Fig. 4B). Thus, the co-expression toxin modules associated with SA8 are defined by a specific expression pattern, and SA8 transcripts from *A.*

tenebrosa are characterised by a greater range of expression profiles than those from *T. stephensoni*.

Following gene duplication, duplicates can be translocated, pseudogenised or retained, with expression divergence often occurring in those duplicates that are retained [47, 52–54]. One of the two major models to explain the different functional roles of duplicated genes is subfunctionalisation, where the roles of the ancestral gene are partitioned between the duplicates [52, 55]. In venomous snakes, subfunctionalisation coincides with changes in the tissue distribution of toxin duplicates derived from salivary proteins [56]. By examining the tissue distribution of duplicated toxin gene families across venomous and non-venomous lineages, it was demonstrated that while toxin duplicates are restricted to the venom gland, their non-toxic counterparts continue to be expressed in the salivary glands of non-venomous reptiles [56]. Likewise, in the current study, we identified two major tissue-specific expression patterns for the SA8 family of *T. stephensoni*. Furthermore, evidence of subfunctionalisation was observed in *A. tenebrosa* with distinct tissue-specific expression patterns of duplicate SA8 genes, with subfunctionalisation of globin and toxin genes previously documented in this species [38, 57].

A gene inversion event leads to recruitment of a SA8 peptide into the venom of *T. stephensoni*

An inversion event was also identified within the tandemly duplicated SA8 gene cluster of *T. stephensoni*. This inverted SA8 gene contains an N-terminal addition (YGKGR) not observed in other clustered SA8 genes. Like other *T. stephensoni* SA8 genes, the inverted SA8 gene is upregulated in the epidermis; however, it is unique in that elevated expression can also be observed in gastrodermis structures, specifically the mesenterial filaments. Furthermore, the transcript corresponding to the inverted SA8 gene is found in a separate co-expression module (black) to other *T. stephensoni* SA8 peptides. In addition, two peptides (TR19056) from the milked venom proteome of *T. stephensoni* [58] — with similarity to previously identified *A. viridis* SA8 peptides [29] — were mapped back to the inverted SA8 gene (Additional file 4: Figure S1A). These peptides represent alternate splice isoforms of the inverted SA8 gene, with different transcriptional start sites that maintain the open reading frame and have identical mature peptides.

(See figure on next page.)
Fig. 3 Expression of SA8 genes across functional regions in **A** *A. tenebrosa* (*n* = 3) and **B** *T. stephensoni* (*n* = 3). Each graph shows the tissue-specific expression pattern of a transcript that corresponds to a SA8 gene. One transcript corresponds to two SA8 genes (SA8_clustered_2&7) in *T. stephensoni*. Bars represent the mean trimmed mean of the M-values (TMM) normalised expression value of the SA8 transcript in each anatomical structure, with error bars representing the standard deviation. Supporting data values are available in Additional File 3: File S1. Abbreviations: A, acrorrhagi; C, club-tips; T, tentacle; P, actinopharynx; M, mesenterial filaments; B, body column; PD, pedal disc

Fig. 3 (See legend on previous page.)
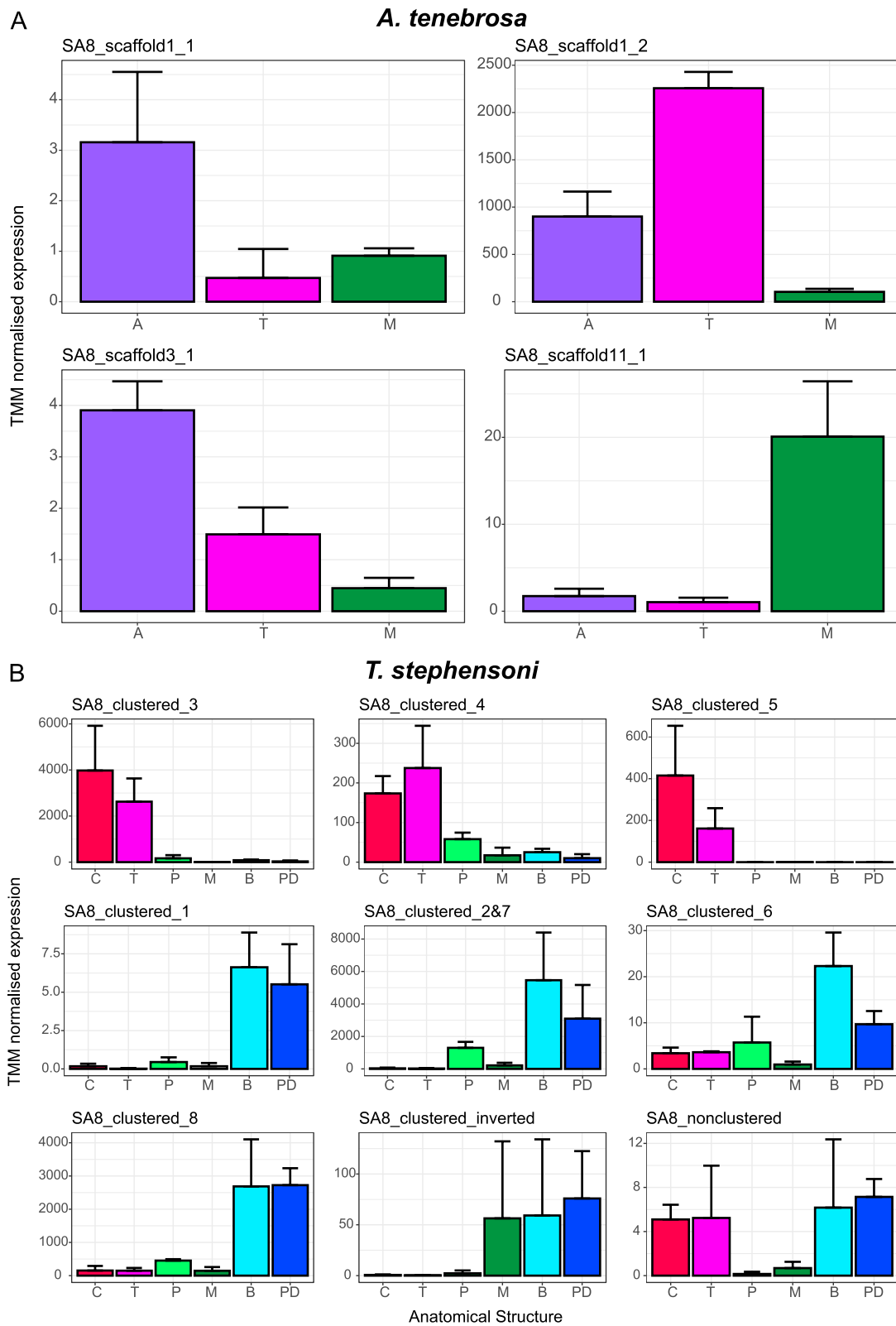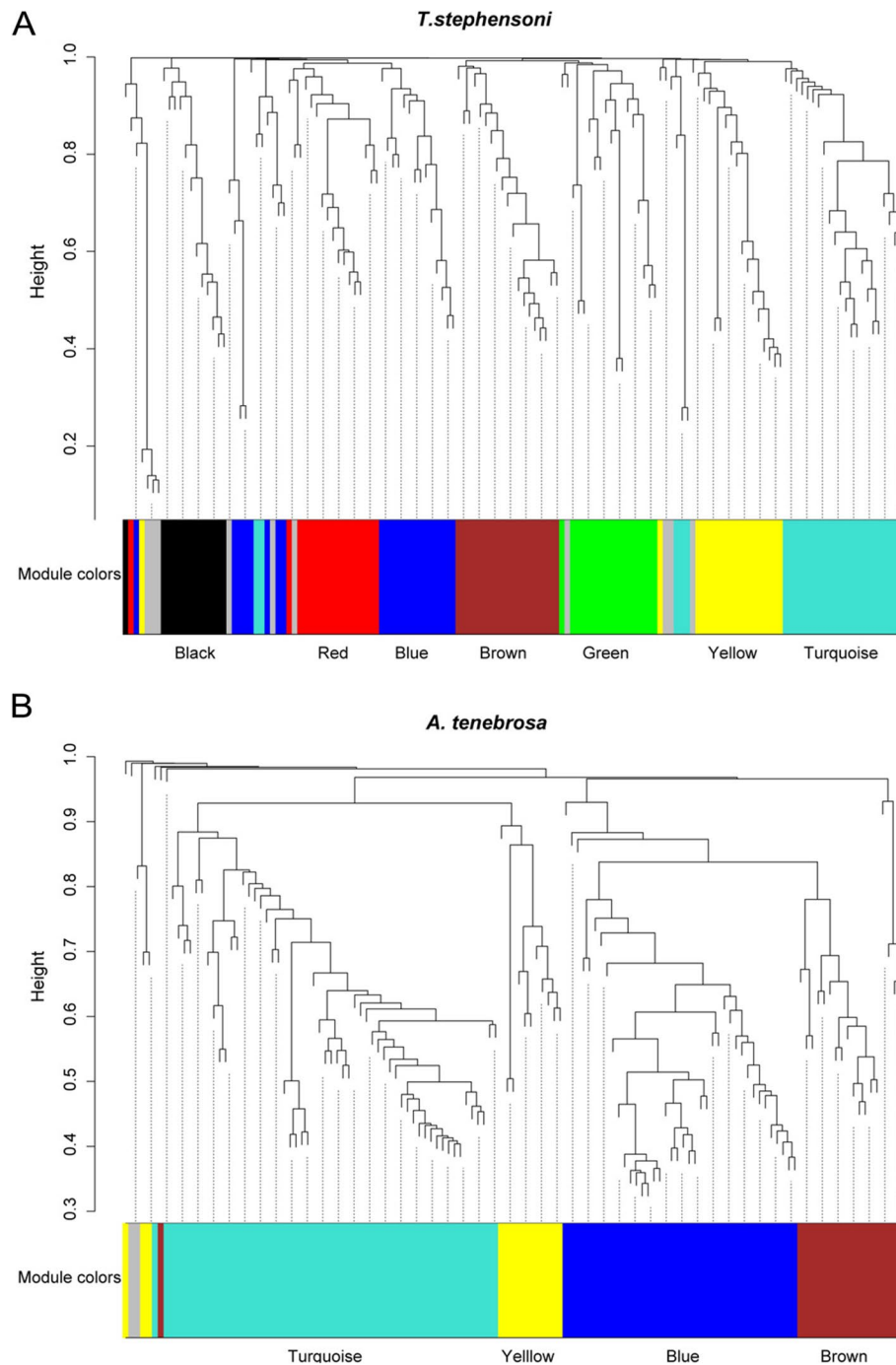
Fig. 4 Cluster dendrogram of putative toxin transcripts from **A** *T. stephensoni* and **B** *A. tenebrosa* constructed using WGCNA. Each colour represents one module. **A** Most *T. stephensoni* SA8 sequences were assigned to the turquoise and brown modules: turquoise = higher expression in the tentacles; brown = higher expression in the epidermis. The inverted *T. stephensoni* SA8 sequence was assigned to the black module, which was not associated with a specific expression pattern. Other modules: blue = higher expression in the mesenterial filaments; green = higher expression in the epidermis and tentacles; red = higher expression in the actinopharynx; yellow = higher expression in the tentacles and mesenterial filaments. **B** *A. tenebrosa* SA8 sequences were assigned to all four modules: blue = higher expression in the acrorhagi; brown = higher expression in the acrorhagi and tentacles; yellow = higher expression in the tentacles; turquoise = no distinct expression pattern. Putative toxin transcripts not assigned to a module are shown in grey
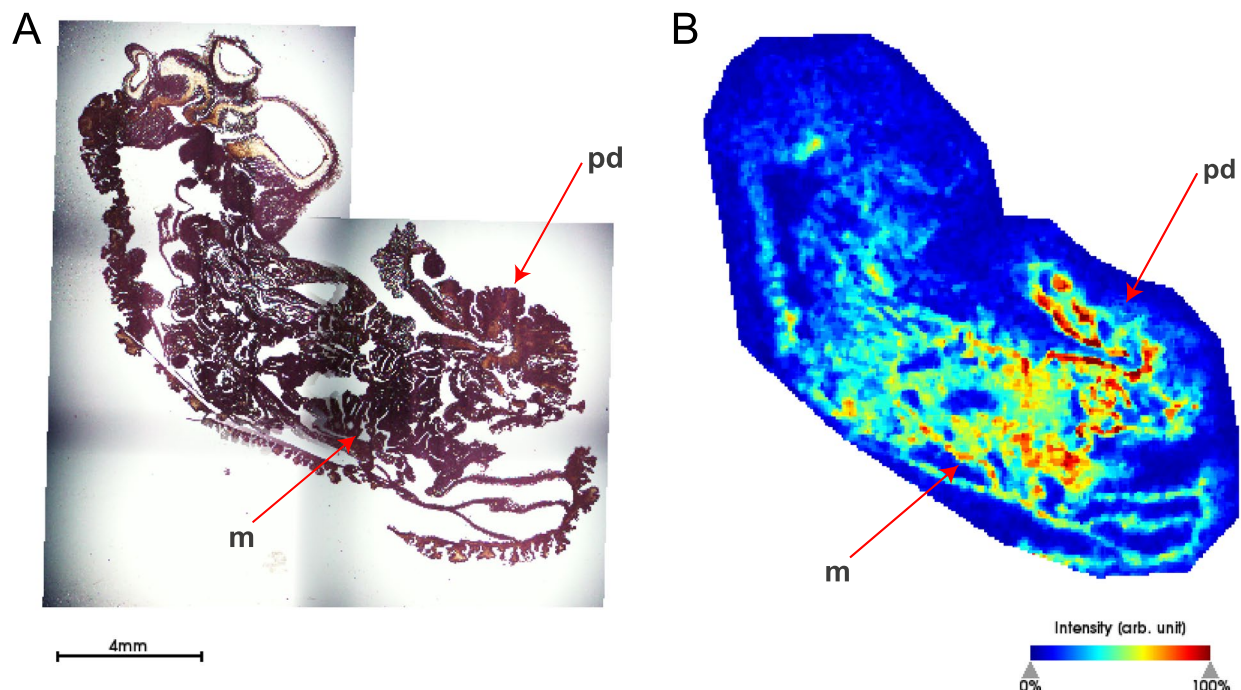
Fig. 5 Peptide corresponding to the calculated mass of the *T. stephensoni* venom SA8 peptide is abundant in acontia, gastrodermis and epidermis. **A** Haematoxylin and eosin stain of *T. stephensoni* section. **B** Putative peptide observed at 5409 m/*z* with the highest abundance in internal structures adjacent to the mesenterial filaments (m) and pedal disc (pd), which corresponds to the location of acontial filaments

Mass spectrometry imaging (MSI) indicates that the peptide corresponding to the calculated mass of the mature *T. stephensoni* venom SA8 peptide is abundant within internal structures located adjacent to the pedal disc (Fig. 5). While this may reflect the localisation of this peptide within the pedal disc and distal mesenterial filaments, the spatial distribution pattern of the venom SA8 peptide is consistent with the principal component corresponding to peptides with higher intensity in the acontia (defensive thread-like structures formed by the free end of mesenterial filaments) [58–60]. Previous research provided evidence for minimal transcriptional activity within *T. stephensoni* acontia, with maturation of nematocysts occurring in proximity to the acontia-mesenterial filament junction [58, 59, 61], which may explain why the venom SA8 peptide is associated with high gene expression in the mesenterial filaments. Furthermore, a SA8 peptide was recently discovered in the acontia proteome of *C. polypus* [62]. Overall, MSI data provide evidence for acontial, mesenterial and pedal disc localisation of the venom SA8 peptide in *T. stephensoni* and are consistent with the higher expression of the inverted SA8 gene in the gastrodermis and epidermis.

As none of the peptides identified in the milked venom of *A. tenebrosa* had significant similarity to SA8 sequences, the SA8 sequences in this species are probably non-venom peptides. This suggests that the inversion event, in concert with tandem duplication, may have facilitated the recruitment of this SA8 peptide into the venom of *T. stephensoni*, with SA8 only previously found in the acontial proteome of C. polypus [62]. Furthermore, the inverted gene is found in a well-supported clade on a branch sister to a Metridioidea-specific clade (Fig. 6), indicating that the recruitment of this orthologue into the venom may have only occurred in the *Telmatactis* genus or related acontiarian sea anemones.

Neofunctionalisation occurs when the duplicated gene assumes a role distinct from that of the original gene, and it is the second major model used to account for functional divergence following gene duplication [54, 55]. Although considered a rare event [56], neofunctionalisation has occurred in multiple snake venom gene families [41, 63–67]. For example, in the snake-venom PLA$_2$ gene family, duplication and subsequent inversion of a basic PLA$_2$ gene resulted in the first acidic-PLA$_2$ gene, and thus this combination of events can represent a scenario that enables neofunctionalisation [41]. We propose an inversion event and subsequent escape from a gene regulatory network that has allowed the venom *T. stephensoni* SA8 gene to undergo neofunctionalization and a novel gene expression
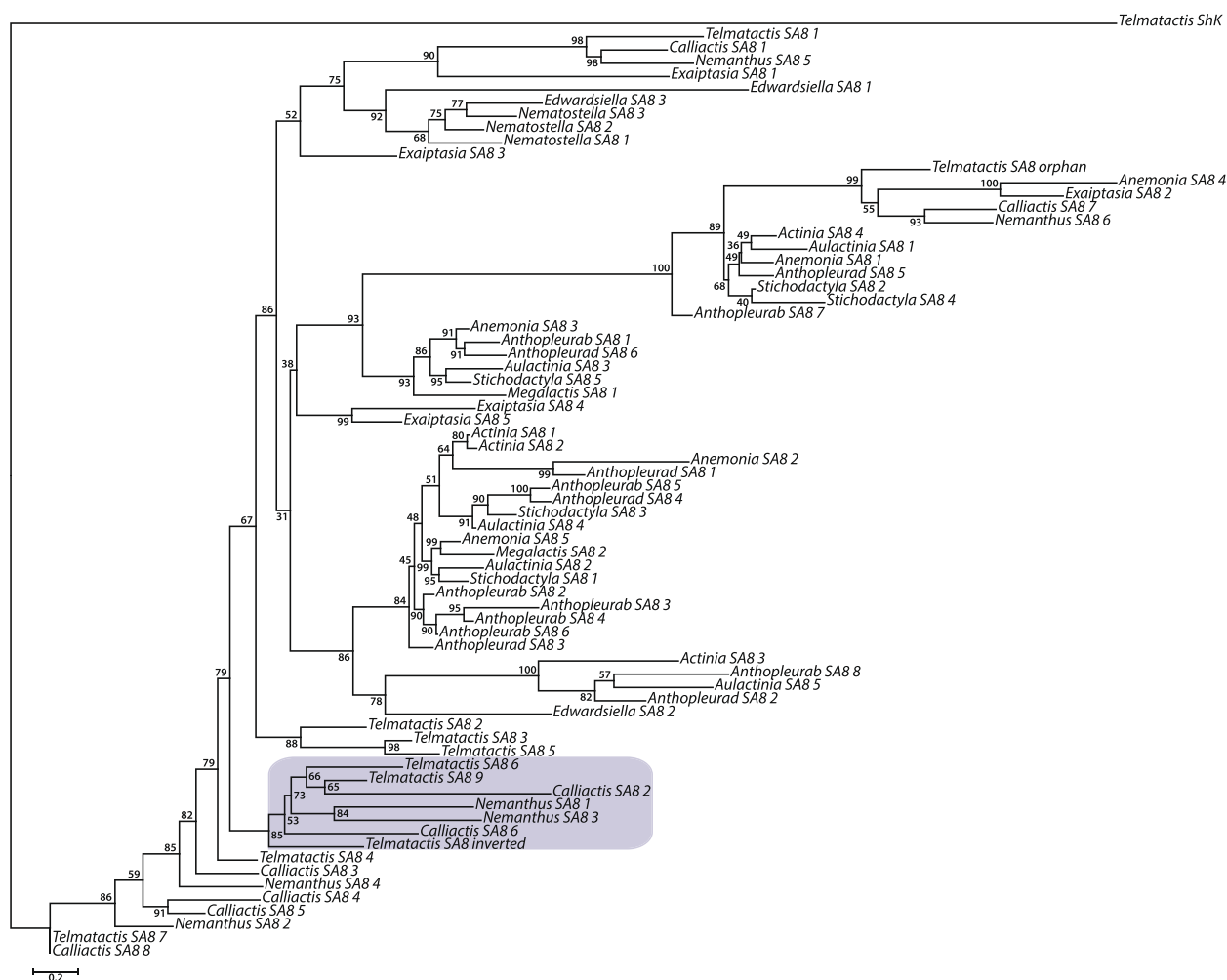
Fig. 6 Phylogenetic relationships among 71 sea anemone sequences from the SA8 gene family were inferred under maximum likelihood in IQ-TREE. The inverted SA8 gene from *T. stephensoni* is found in a well-supported clade (support value > 80) on a branch sister to a Metridioidea-specific clade (purple). Support values are from 1000 ultrafast bootstrap replicates
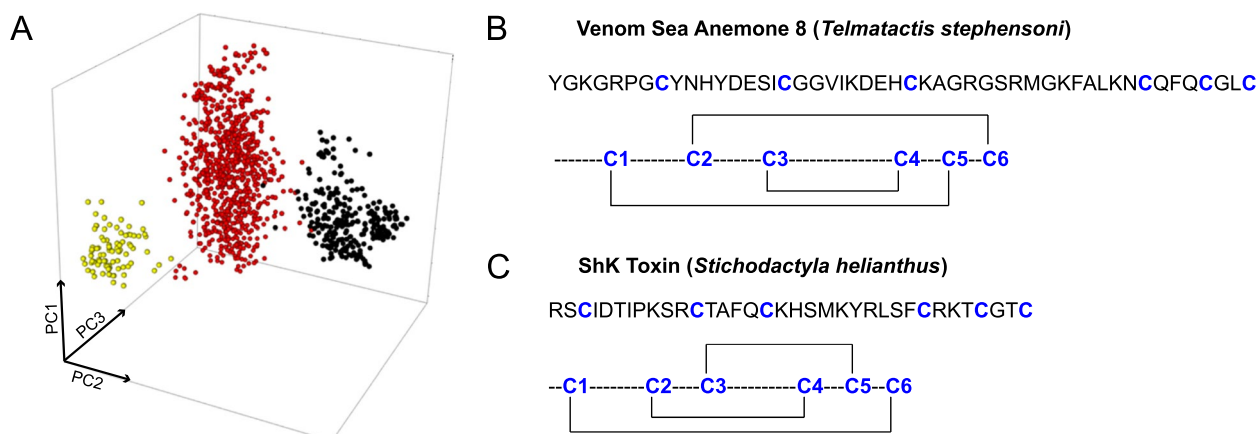
program. This gene also possesses a unique N-terminus, not observed in the other clustered SA8 genes in *T. stephensoni*.

The SA8 family is distinct from the ShK and possesses a unique disulphide network

All SA8 putative toxins in *T. stephensoni* had a cysteine spacing consistent with motif 3, described by Kozlov and Grishin [29], which has previously been observed in SA8 putative toxins from *A. viridis* (P0DMZ3, P0DMZ4, P0DMZ5, P0DMZ6, P0DMZ7), sea anemone type 1 potassium channel toxins with a ShKT domain (P29187, P29186, P81897, Q9TW91) [29] and cysteine-rich domain (CRD) of snake venom cysteine-rich secretory proteins (CRISPs) [68]. Sequence space analysis [27] revealed that ShK-like proteins, CRISPs and SA8 putative toxins are clearly delineated into three groups (Fig. 7A),

reinforcing the unique sequence composition of SA8 putative toxins and indicating that they represent a single distinct protein family.

All SA8 peptides are characterised by a conserved glycine residue and FA dyad (Additional file 4: Figure S1A). This dyad has not been reported in other toxin groups and further research is required to determine its functional significance. Few SA8 peptides contain the KY dyad characteristic of ShK, although most contain a KY dyad alternative (QY, KW, TY, RF); however, the two protein families possess distinct sequence patterns and exonstructure. SA8 peptides contain more residues between the pro-cleavage site and final cysteine, resulting in larger proteins, but only ShK-like sequences have an extension at the C terminus. SA8 genes are consistently composed of 3–5 microexons (Additional file 4: Figure S1B), while ShK genes are composed of 2–3 exons. Additionally, large

A

B   Venom Sea Anemone 8 (*Telmatactis stephensoni*)

YGKGRPG**C**YNHYDESI**C**GGVIKDEH**C**KAGRGSRMGKFALKN**C**QFQ**C**GL**C**

-------**C1**--------**C2**-------**C3**----------------**C4**---**C5**--**C6**

C   ShK Toxin (*Stichodactyla helianthus*)

RS**C**IDTIPKSR**C**TAFQ**C**KHSMKYRLSF**C**RKT**C**GT**C**

--**C1**--------**C2**----**C3**----------**C4**---**C5**--**C6**

Fig. 7 Sequence space of the SA8 family (**A**) and disulphide connectivities of venom SA8 (**B**) and ShK (**C**). **A** The SA8 family (yellow) forms a distinct cluster from the ShKT domains of ShK-like (red) and CRISP (black) peptides [27]. **B** The C1–C5, C2–C6, C3–C4 disulphide connectivity of the mature venom SA8 peptide from *T. stephensoni*. **C** The C1–C6, C2–C4, C3–C5 disulphide connectivity of ShK toxin from *S. helianthus*

introns are present in all SA8 genes, but particularly so in *A. tenebrosa*, with a 17-kb intron identified in this species. Therefore, although ShK-like and SA8 peptides have a similar cysteine spacing pattern, these two families are resolved into discrete groups when examining their other characteristics.

To examine the structure of the SA8 peptide present in milked venom and determine if it is similar to that of ShK, we expressed this venom peptide and characterised it using liquid chromatography-tandem mass spectrometry (LC–MS/MS) and nuclear magnetic resonance (NMR) spectroscopy. The His$_6$-MBP-SA8 (polyhistidine; maltose binding protein) fusion protein was the dominant protein expressed in transformed *E. coli* cells following Isopropyl β-D-1-thiogalactopyranoside (IPTG) induction. SA8 was liberated from the 50.5 kDa fusion protein using tobacco etch virus (TEV) protease (Additional file 5: Figure S2), then purified by affinity chromatography followed by reversed-phase high-performance liquid chromatography (RP-HPLC) (Additional file 6: Figure S3A). The liquid chromatography-mass spectrometry (LC–MS) profile of the fraction corresponding to SA8 showed a pure peak whose molecular mass (MH$^+$ $m/z = 5438$ Da) was in good agreement with the expected mass (theoretical MH$^+$ $m/z = 5438.23$ Da) (Additional file 6: Figure S3B). The recombinant SA8 co-elutes on RP-HPLC with native SA8, with only a very slight difference in elution time (11.44 min and 11.38 min, respectively) due to the presence of an additional C-terminal glycine on the recombinant peptide (Additional file 7: Figure S4). Furthermore, the observed mass of the recombinant SA8 was 6 Da less than that of the reduced form of the peptide (MH$^+$ $m/z = 5444.23$ Da), indicating formation of three disulphide bonds. LC–MS/MS analysis

of pepsin-digested SA8 revealed the disulphide-bond connectivity to be C1–C5, C2–C6, and C3–C4 (Fig. 7B, Additional file 8: Table S3). This pattern differs from the ShK disulphide framework (C1–C6, C2–C4, C3–C5) (Fig. 7B), supporting the hypothesis that SA8 is not a simple extension of the ShK fold.

A one-dimensional (1D) ^{1}H NMR spectrum of recombinant SA8 acquired at pH 3.5 and 25 °C revealed poor dispersion of amide-proton resonances region (Additional file 9: Figure S5). Subsequently, several 1D spectra were recorded at temperatures over the range 5–30 °C in 5 °C increments and pH values 2–7 (Additional file 10: Figure S6 and Additional file 11: Figure S7), but no significant improvement in peak dispersion was observed. Moreover, a sample of the recombinant peptide that was reduced and unfolded, then oxidatively refolded in vitro gave the same spectral dispersion as the recombinant peptide, indicating that we are dealing with the most stable fold of this sequence (Additional file 12: Figure S8). Thus, the NMR data indicate that the *T. stephensoni* venom SA8 peptide adopts a partially disordered structure in solution, unlike ShK which has a well-defined structure [18]. Overall, we report the SA8 family is associated with unique disulphide-bond connectivities, exon–intron structure, and sequence spaces that distinguish it from ShK.

Functional activity of SA8 peptides remains unknown

ShK is a 35-residue peptide isolated from the sea anemone *S. helianthus* [14] that potently inhibits K_V1 channels through a conserved KY functional dyad [69–71]. The presence of KY dyad alternatives in multiple SA8 sequences suggests this putative toxin family may have some activity against K^+ channels [72]. First, we tested

the effects of 100 nM recombinant SA8 on the following human K_V channels: $K_V1.1$, $K_V1.2$, and $K_V1.3$. Whole-cell $K_V1.2$ currents were recorded sequentially on the same transiently transfected Chinese hamster ovary (CHO) cell, before (control trace, black) and after perfusion of the cell with 100 nM SA8 (red trace) (Fig. 8A). At the equilibrium block, SA8 caused ~10% reduction in current amplitude. This weak block was almost fully reversible upon washing the perfusion chamber with inhibitor-free solution. Both the association and dissociation rates of SA8 were rapid; development of equilibrium block and recovery up to ~85% of control current took ~1 min. In contrast, 100 nM SA8 did not inhibit $K_V1.1$ or $K_V1.3$ (Additional file 13: Figure S9A and B). We performed a concentration–response experiment for the inhibition of $K_V1.2$ channels by SA8 (Fig. 8B) and determined the reciprocal of the slope of the fitted line yielded an IC_{50} of 40.8 ± 4.9 μM ($n=4$).

To reveal the selectivity profile of SA8 we assayed its effect on other K^+ channels as well. We tested its activity on $K_V10.1$ and $K_V11.1$ due to their slightly different structure compared to $K_V1.x$ channels [73–75]. We also tested the effect of SA8 on $K_{Ca}1.1$, the large conductance voltage- and Ca^{2+}-activated channel, and $K_{Ca}3.1$, a Ca^{2+}-activated K^+ channel. Moreover, as SA8 is suggested to be a defensive peptide and possibly involved in painful stings, channels involved in pain recognition were also tested including the voltage-gated sodium

channel $Na_V1.7$, and the transient receptor potential channels TRPA1 and TRPV1. We used positive controls (2 μM astemizole for $K_V10.1$, 10 mM tetraethylammonium (TEA^+) for $K_{Ca}1.1$, 20 nM TRAM-34 for $K_{Ca}3.1$, 50 μm HC-030031 for TRPA1, and 50 μm capsazepine for TRPV1) diluted freshly in extracellular solution (ECS) in order to confirm ion channel expression and proper operation of the perfusion system. We found that most of the channels shown in Fig. 8 and Additional file 13: Figure S9 are not inhibited significantly by SA8 at the indicated concentrations except for the modest block of the hKv1.2 current.

Injection of *D. melanogaster* with recombinant SA8 at a dose of 0.5 μg per fly did not result in paralysis or death within 24 h. In contrast, neurotoxic spider-venom peptides with insecticidal activity cause paralysis and death at dose of 0.005 μg per fly in *D. melanogaster* [76]. Thus, the recombinant SA8 peptide may not exert any toxic effects in insects and only causes a weak inhibition of human Kv1.2 channels. However, the presence of this peptide in the venom of *T. stephensoni* indicates that it probably has at least an auxiliary function in venom, and the abundance of the venom SA8 peptide within structures with defensive roles (acontia and epidermis), suggests that one potential role of this putative peptide may be targeting the neuronal channels of vertebrate predators in order to fulfil the ecological function of predator
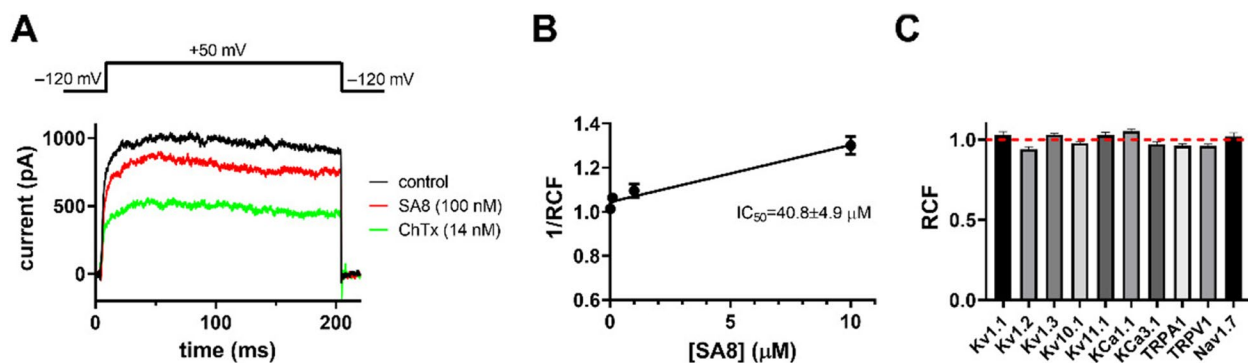
Fig. 8 SA8-induced inhibition of h$K_V1.2$ and selectivity profile of SA8. **A** Representative whole-cell h$K_V1.2$ current traces were recorded using the voltage protocol shown above the raw current traces. Depolarizing pulses were repeated every 15 s, selected traces are shown in the absence of blockers (black, control) and upon reaching equilibrium block in the presence of 100 nM SA8 (red) or 14 nM charybdotoxin (ChTx, green, positive control). **B** Low affinity, concentration-dependent block of h$K_V1.2$ channels by SA8 was determined by fitting a straight line to the reciprocal of the remaining current fraction (1/RCF) plotted as a function of SA8 concentration. The remaining current fraction (RCF) was calculated as I/I_0, where I_0 is the peak current in the absence and I is the peak current at the equilibrium block in the presence of SA8 at concentrations of 0.01, 0.1, 1, and 10 μM (filled circles). Points on the linear concentration–response curve represent the mean of four independent measurements where the error bars represent the standard error of mean (SEM). The line was drawn using linear least squares fit and the reciprocal of the slope of the best fit yielded an IC_{50} of 40.8 ± 4.9 μM. **C** The effect of SA8 (100 nM, except for $K_V10.1$ and $Na_V1.7$ which were tested at 1 μM) on the peak currents was reported as the RCF. Bars represent the mean of 3–6 independent measurements; error bars indicate the SEM. Data are shown for the following channels: H$k_V1.1$ ($n=4$), h$K_V1.2$ ($n=4$), h$K_V1.3$ ($n=3$), h$K_V10.1$ ($n=4$), h$K_V11.1$ ($n=5$), m$K_{Ca}1.1$ ($n=4$), h$K_{Ca}3.1$ ($n=6$), hTRPA1 ($n=5$), hTRPV1 ($n=5$), and h$Na_V1.7$ ($n=5$) (for details of the expression systems, solutions, and voltage protocols, see Materials and Methods, and for raw current traces see Additional File 13: Figure S9). SA8 did not inhibit any of the investigated channels at the applied concentrations. Supporting data values are available in Additional File 3: File S1

Ashwood *et al. BMC Biology*   (2023) 21:121

Page 13 of 25

deterrence [58]. Further functional testing is required to determine the biological activity of this venom.

Conclusions

Based on sequence, structural and genomic data, we find that SA8 putative toxins are a distinct family that has only superficial similarities to ShK. Thus, characterising the SA8 family offers an opportunity to explore a novel group of putative toxins that may possess pharmacological properties that make them useful as pharmacological tools or therapeutic leads.

We report that the SA8 family is expanded in *T. stephensoni* but not *A. tenebrosa* relative to other actiniarian species. The two genomic locations of SA8 genes in *T. stephensoni* form two distinct collinear blocks with two of the four loci of SA8 genes in *A. tenebrosa*, indicating that segmental duplication may have played a role in evolution of the SA8 family. However, the expansion of clustered SA8 genes in *T. stephensoni* appears to be the consequence of proximal and tandem duplication events.

While members of this family are expressed in the neuronal cells of *N. vectensis*, a SA8 peptide has been recruited to the venom of *T. stephensoni* following an inversion event. This inverted SA8 gene is distinct from other members of the SA8 family across multiple levels and represents a potential neofunctionalisation event in the SA8 gene family. In contrast, the absence of non-inverted SA8 peptides from the venom of *T. stephensoni* and *A. tenebrosa*, as well as the previously reported neuronal localisation of SA8-like peptides suggests that other SA8 genes may encode non-venom neuropeptides.

Furthermore, while we have not yet determined the functional activity of venom SA8 peptides, we have shown that they are structurally distinct from ShK toxins and possess a unique disulphide bond pattern. We propose that the inverted *T. stephensoni* SA8 peptide may have a role in defence against predators, as either a toxin or an auxiliary venom protein, based on its distribution across the body plan.

Additional testing will be required to determine the exact function of the inverted SA8 and the SA8 family in general, but given the findings of the current study, the genetic origins and diversification of this putative toxin family should be investigated in other sea anemones. Future research should determine whether the patterns we report are conserved across Actinioidea and Metridioidea, particularly whether venom SA8 peptides are only found when there is an inversion within a cluster of tandemly duplicated genes.

Methods

Animal care
T. stephensoni were sourced by Cairns Marine Pty Ltd from the Great Barrier Reef (QLD, Australia) while *A. tenebrosa* were collected from the intertidal zone at Point Cartwright (QLD, Australia). Both species were housed in holding tanks at the Queensland University of Technology until experimental use. Tank salinity and temperature were maintained at 33–37 ppt and 20–28 °C, respectively.

Genome sequencing
Short-read DNA sequencing
DNA was extracted from *A. tenebrosa* using the E.Z.N.A Mollusc DNA Kit (Omega Bio-Tek). Tentacle tissue was homogenised in a tube containing ML1 buffer (Omega Bio-Tek) and a stainless-steel ball bearing (Qiagen) using a Tissuelyser II (Qiagen), before the remaining steps were performed following the manufacturer's protocol. To extract DNA from *T. stephensoni*, tentacle tissue was homogenised in liquid nitrogen with a mortar and pestle. The sample was then processed following the QIAGEN Genomic-tip procedure. DNA quantity was determined using a Qubit 4 fluorometer. Genomic DNA for both *A. tenebrosa* and *T. stephensoni* was sequenced using 150-bp paired-end chemistry on an Illumina HiSeq 2500 and X10, respectively.

PacBio long-read sequencing
For *A. tenebrosa*, high molecular weight (HMW) DNA was extracted using two rounds of the method described for short-read DNA-seq from the same individual. HMW DNA was extracted from *T. stephensoni* according to the 10X Genomics salting-out protocol for DNA extraction from single insects [77]. In order to extract HMW DNA, minor modifications were made to this protocol, including the use of wide-bore pipette tips, low bind tubes and a gentle bead clean-up. DNA molecular weight was assessed using PippinPulse pulsed-field gel electrophoresis, while purity and quantity were determined using a Qubit fluorometer. Samples of sufficient purity (260:280 ratio of 1.8–1.9 and 260:230 ratio of 1.6–2.0) and quantity were sequenced.

Five micrograms of genomic DNA from each species was prepared using needle shearing and the BluePippin (SageScience) size-selection system. The SMRTBell® Template Prep Kit 1.0 (PacBio) was then used to prepare 14-kilobase libraries. Sequencing was undertaken on four sequencing chips using a 16-h run time on a PacBio Sequel®.

Ashwood *et al. BMC Biology*   (2023) 21:121

Page 14 of 25

Hi-C sequencing

Hi-C sequencing was performed to generate a chromosomal-level assembly for *A. tenebrosa*. Tentacle tissue from *A. tenebrosa* was crosslinked using 1% formaldehyde. Chromatin extraction, library preparation and sequencing were performed by Phase Genomics (Seattle, WA).

Genome assembly and polishing

PacBio reads from *A. tenebrosa* and *T. stephensoni* were assembled using SMART *denovo* [78] and HGAP4 (Pacific Biosciences, SMART Link), respectively. Assemblies were polished using raw PacBio reads and a custom Arrow pipeline [79]. Four iterations of Arrow polishing were carried out for *A. tenebrosa*, while three iterations were conducted for *T. stephensoni*. Subsequently, both assemblies underwent two rounds of polishing with a custom Pilon pipeline [80] using Illumina short reads. Indel errors were removed from the polished *A. tenebrosa* genome assembly with deGRIT [81].

Redundancy removal for *A. tenebrosa* was performed using the Purge Haplotigs pipeline [82]. This pipeline involved the alignment of raw reads to the genome using minimap2 [83] and sorting using samtools [84]. Repeat regions were accounted for by providing Purge Haplotigs the results of our repeat masking (see the "Repeat library generation" section) converted to BED file format with a custom script. Program parameters were set by manual inspection of the histogram produced by the pipeline (low: 20, mid: 75, high: 170).

Finally, proximity-guided assembly was carried out on our *A. tenebrosa* assembly by Phase Genomics (Seattle, WA) using the Proximo platform. The quality of genome assemblies was assessed using BUSCO [85] and a custom script [86].

Repeat library generation

A custom repeat library (CRL) was generated using homology and structure-based prediction in addition to de novo repeat prediction. MITEs were predicted using MITE-HUNTER v.11–2011 [87] and detectMITE v.20170425 [88]. The output MITE models of both programs were clustered using CD-HIT v.4.6.4 [89] with the same parameters employed by the detectMITE program to produce a non-redundant set of MITE models (cd-hit-est -c 0.8 -s 0.8 -aL 0.99 -n 5). In addition, a structure-based prediction of long terminal repeat retrotransposons (LTR-RTs) was performed using LTRharvest (GT 1.5.10) [90] and LTR_FINDER v.1.06 [91] following the recommended protocol indicated by LTR_retriever commit 8180c24 [92] to identify canonical and non-canonical (i.e. non-TGCA motif) LTR-RTs. The MITE and LTR-RT libraries were used to mask the genome

assembly using RepeatMasker open-4.0.7 [93] with settings '-e ncbi -nolow -no_is -norna'. After this homology and structure-based modelling, de novo repeat prediction was performed using RepeatModeler open-1.0.11 [94] with the masked genome as input.

All repeat models predicted by the aforementioned programs were then curated to remove models that were potentially part of genuine protein-coding genic regions. This process first involved the removal of any models that were confidently annotated by either LTR_retriever or by RepeatModeler (i.e. were not classified as "Unknown") from consideration as these were assumed to be true positives. The remaining nucleotide repeat models were translated into six reading frames and were searched using HMMER 3.1b2 [95] for a list of domain models associated with transposable elements (TEs). This list was generated by adding Pfam [96] and NCBI CDD [97] domains to a list of domains identified by a previous study [98]. The added Pfam domains were based upon visual inspection of domain prediction results for putative transposable elements and comparison to the "Domain organisation" graphics provided by the Pfam website to find models that were not likely to appear in non-TEs. The CDD domains were part of the cl02808 RT_like superfamily. Any repeat models that obtained a TE-associated domain prediction were assumed to be true positives and were removed from consideration.

To remove repeat models that may be part of genuine protein-coding genes, we generated a database of known genes. This database consisted of UniProtKB/Swiss-Prot proteins (version 10/26/17) [99] in conjunction with the gene models of *N. vectensis* (v.2.0) [100], *E. diaphana* (v.1.1) [35], *Acropora digitifera* (v.0.9) [101], and *Hydra vulgaris* (NCBI annotation release 102, June 8, 2015) [102]. Putative transposons were removed in this database by the same process detailed above using HMMER 3.1b2 and the list of TE-associated domains (i.e. any sequences with TE-associated domain hits were removed). Repeat models that remained after this curation process were removed from the initial CRL if they obtained a BLASTx result with E-value more significant than $1E^{-2}$ when submitted as a query against the gene model database. This process resulted in a high-quality CRL which was used to soft-mask the *A. tenebrosa* genome using RepeatMasker (-e ncbi -s -nolow -no_is -norna -xsmall) for subsequent gene model prediction. Scripts were produced to automate this process [103].

Gene model prediction

Gene model prediction was performed using two complementary approaches: transcriptome-based gene model creation and *ab initio* gene model prediction.

RNA-seq read quality control and mapping

RNA-seq reads generated from previous studies involving *A. tenebrosa* [104] and *T. stephensoni* [58] were used to assist in gene model prediction. Raw reads were quality trimmed using Trimmomatic [105] with parameters "ILLUMINACLIP:$TRIMDIR/adapters/TruSeq3-PE.fa:2:30:10 SLIDINGWINDOW:4:5 LEADING:5 TRAILING:5 MINLEN:25" based on those used by the Trinity de novo assembler [106, 107]. Trimmed sequences were aligned to each genome file using STAR 2.5.4b (commit 5dbd58c) using the 2-pass procedure to assist in the identification of intron splice sites. The resulting SAM files were converted into sorted BAM files using samtools [84].

Transcriptome-based gene model prediction

Multiple transcriptome building programs were used to create a 'master' transcriptome file for subsequent curation using the EvidentialGene tr2aacds pipeline [108] following the parameters and methodology of Visser et al. [109]. This involved de novo transcriptome assembly using SOAPdenovo-Trans v.1.03[110] and Velvet/Oases [111, 112] with multiple k-mer lengths (23, 25, 31, 39, 47, 55, and 63 for both plus 71 for SOAPdenovo-Trans only) alongside Trinity v.2.5.1 [106]. We differed from Visser et al. [109] from here on. For Trinity, we modified the parameters of Visser et al. [109] to not specify min_contig_length and we set min_kmer_cov = 2 and SS_lib_type = RF. All resultant de novo assemblies had sequences shorter than length of 350 bp removed with a custom script. Genome-guided transcriptomes were constructed for each genome assembly using Trinity and Scallop v.0.10.2 [113] with the results of the STAR spliced alignment. For Trinity, we specified genome_guided_max_intron = 21,000 to reduce false positives. This number was based on preliminary analysis of an Illumina-based *A. tenebrosa* assembly [38] and the *E. diaphana* genome for which, of 152,518 total introns, only 51 are > 21 kb (0.03%) in the v.1.1 genome.gff. Scallop was run with default settings except where library_type = first. No minimum size cut-off was enforced for genome-guided transcriptomes as we assumed any short predictions would likely be more accurate than those built from de novo assembly. All transcriptomes were then concatenated and subjected to the EvidentialGene tr2aacds pipeline which rendered a master transcriptome consisting of non-redundant genes including alternative isoforms.

PASA v.2.2.0 [114] (commit af03820) was used to align the master transcriptome against the assembled genomes (using BLAT v.35 [115] and GMAP [116]) and to generate gene model predictions using default recommended parameters (i.e. transcriptome files were 'cleaned' using the seqclean tool, and validate_alignments_in_db.dbi were provided arguments MIN_PERCENT_ALIGNED = 75, MIN_AVG_PER_ID = 95, and NUM_BP_PERFECT_SPLICE_BOUNDARY = 0).

BRAKER1 ab initio gene model prediction

Gene models were predicted by BRAKER v.2.0.6 [117] using the repeat library soft-masked genome assembly (–softmasking) and the BAM file produced by STAR for use as 'hints' during Augustus v.3.2.3 [118] gene prediction.

Combined gene model prediction

EvidenceModeler [119] was used to combine transcriptome-based gene model predictions with BRAKER1-derived ab initio gene models with parameters set to favour transcriptome models above ab initio models (Augustus model weight = 1, Transcript-based model weight = 10). Following this, a custom pipeline to curate gene models was generated [120]. This pipeline finds additional gene copies missed by prior annotation steps and removes spurious models derived from transposable elements and from ribosomal RNA.

Phylogenetic and comparative genomic analysis
Genomic data resources

In addition to the two genomes reported herein, publicly available genomic resources for sea anemones were obtained for comparative genomic analyses. These included a chromosome-level *N.vectensis* genome (Nvec200) [37], the *E. diaphana* genome (GCF_001417965.1) [35], and the *P. xishaensis* genome (version 3) [121]. The annotation of *N. vectensis* was modified by manually annotating a putative SA8 gene into its genome prior to further analysis.

Gene family expansion and contraction

Gene families were predicted for the five sea anemone species using OrthoFinder (version 2.5.4) [122, 123]. Single copy genes were aligned with MAFFT (version 7.311) [124] and used as input to IQ-TREE 2 [125] with an estimated divergence date of 500 mya for *Actinia* and *Nematostella*. Gene family counts and our time-calibrated tree were used as input to CAFE 5 [126] to identify gene families with expansion/contracted predicted to a P-value threshold of 0.05 (fixed orthogroups).

Enrichment analysis

Gene ontologies (GOs) [127] and PFAMs [128] were predicted for each of the five sea anemone species' representative gene models using custom scripts [129]. In short, Mmseqs2 [130] was used to query gene models against the UniRef90 database [131] and sequence

matches had associated GO terms obtained from the id_mapping.tab file provided by the UniProt Knowledgebase [99]. PFAM 35.0 domain models were downloaded, and sequences were searched for these domains using HMMER3 [95]. Chi-squared tests were used to identify annotation terms enriched in expanded gene families; these tests were performed using a custom script [132]

Phylogenetic analysis

Relationships among 71 sea anemone full-length sequences (alignment length 121 amino acid residues) from the SA8 gene family were inferred under maximum likelihood (ML) in IQ-TREE v2.1.3[125]. The best fitting model of amino acid evolution was identified by ModelFinder within IQ-TREE as LG [133], with rate variation among sites accommodated by a proportion of invariant sites and a gamma distribution. This LG + I + G model was favoured by both BIC and corrected AIC (AICc). Default tree search parameters were employed within IQ-TREE and following reconstruction of the ML tree, clade support values were evaluated with 1000 ultrafast bootstrap samples.

Analysis of genome collinearity and classification of gene duplications

To identify similar proteins in inter-genome and self-genome comparisons, homology of amino acid sequences between the two species and within a single genome was examined using BLASTp ($1xE^{-5}$). These BLASTp alignments were leveraged by MCScanX (default parameters) to identify collinear blocks. Additionally, the duplicate_gene_classifier program from the MCScanX package was used to classify the origins of duplicate genes in the genomes of *A. tenebrosa* and *T. stephensoni*, independently.

Expression analysis of SA8 genes and identification of toxin-like transcripts

To examine the expression patterns of SA8 genes, we made use of tissue-specific reads for three individuals of *A. tenebrosa* [33] and *T. stephensoni* [58]. Transcriptome assembly and differential gene analysis were conducted using the Trinity short-read de novo assembler v2.5.1 [134] and edgeR Bioconductor package v3.3.1 [135], respectively, as described in Ashwood et al. [58]. Similarly, Trinotate v3.1.1 was used for functional annotation as per Ashwood et al. [58].

Transcripts with homology to known animal toxins were identified from BLASTp searches against the UniProt database by filtering for transcripts annotated with the "toxin activity" GO term (GO:0,090,729), excluding

bacterial toxins. Consequently, WGCNA one-step network construction and module detection (power = 7, mergeCutHeight = 0.25, minModuleSize = 10, deep-Split = 3) was used to identify correlation networks of toxin-like transcripts present in *T. stephensoni* and *A. tenebrosa* [136].

Milked venom peptide identification and mass spectrometry imaging

Peptides present in the milked venom of *T. stephensoni* and *A. tenebrosa* were previously identified by Ashwood et al. [58] and Surm et al. [33], respectively. Using these signal peptide-containing milked venom peptides as queries, homology to transcriptomic sequence and ToxProt entries was determined using tBLASTn and BLASTp, respectively. Significant BLAST hits were defined by a threshold *e*-value of $1 \times E^{-5}$. Similarly, the mass spectrometry imaging (MSI) data for *T. stephensoni* utilised in the current study were generated as detailed previously [58]. The spatial distribution of peptides of interest was determined from the MSI dataset using SciLS Lab.

Chemical/sequence space analysis

A database of 102 SA8 peptides was generated using SA8 sequences identified from 14 actiniarian species spanning three superfamilies [33], as well as the SA8 translated ORFs of *T. stephensoni* from the current study. We analysed these SA8 sequences by position-specific biophysical property distance analyses, or sequence-space analyses [137]. Briefly, the multiple sequence alignment was converted to a matrix based on the physiochemical properties of the amino acids at each position in the alignment (gaps assigned column average values for each property), before dimensionality reduction (by principal component analysis) and group assignment (by Bayesian model based clustering) [137]. Additionally, SA8 sequences were aligned against ShK-like proteins (PF01549) and the homologous C-terminal regions of CRISPs (PF08562) available in the Pfam database [96], allowing SA8 putative toxins to be visualised in a sequence space with ShKT-domain containing sequences. Data were visualised with custom [R] scripts based on rgl [138]. WGCNA modules were mapped to the clusters identified in the sequence space of mature SA8 peptides, and the correlation between the two categorical variables was assessed using the GoodmanKruskal package [139].

Peptide synthesis and structure characterisation
Design of vector used for expressing SA8

A nucleotide sequence encoding the mature peptide sequence of SA8, containing codons optimised

Fig. 9 Schematic of the expression vector used for expression of SA8 in the periplasm of *E. coli*. Abbreviations: RBS, ribosome binding site; MalE$_{SS}$, MalE signal sequence; His$_6$, poly-histidine affinity tag; TEV, TEV protease recognition site; SA8, sea anemone 8

for expression in *Escherichia coli*, was synthesised by GenScript (Piscataway, NJ, USA). The sequence was subcloned into a modified pET21a vector (Novagen, Madison, WI, USA) using *Bam*HI and *Eco*R1 restriction sites (Fig. 9). This modified vector encodes a MalE signal sequence (for targeting the fusion protein to the *E. coli* periplasm), a His$_6$ affinity tag (for nickel affinity chromatography used for peptide purification), MBP fusion tag (to enhance the solubility of the expressed peptide), and a TEV protease recognition site (ENLYFQ) immediately preceding the coding region for SA8.

Recombinant expression of SA8

The pET21a-SA8 plasmid was transformed into *E. coli* strain BL21 (DE3; Thermo Fisher) using a standard heat shock protocol [140]. Single colonies were inoculated into a 5-mL culture medium (Luria–Bertani (LB) plus 100 µg/mL of ampicillin) and grown overnight at 37 °C with shaking at 150 rpm. The overnight starter culture was used to inoculate LB medium containing 100 µg/mL of ampicillin, then the culture was grown until it attained an optical density at 600 nm (OD$_{600}$) of 0.6–0.8. IPTG was added to a final concentration of 0.5 mM to induce expression, and the culture was grown overnight at room temperature with shaking at 180 rpm. Induced cells were harvested by centrifugation for 15 min at 2408 g and 4 °C, and the cell pellets were stored at −80 °C.

Peptide purification

The His$_6$-MBP-SA8 fusion protein was recovered from the bacterial periplasm by osmotic shock [141]. Cell pellets were thawed on ice to avoid lysis, resuspended in 30 mL 100 mM Tris, 30% sucrose, 2 mM EDTA pH 8.0, and incubated for 10 min at 4 °C. The suspension was centrifuged for 15 min at 4 °C and 5251 g. The pellets were resuspended in 30 mL of ice-cold water and 5 mM MgCl$_2$, then incubated for 10 min at 4 °C before centrifugation for 15 min at 4 °C and 21,002 g. The supernatant containing soluble His$_6$-MBP-SA8 fusion protein was diluted in 20 mM Tris, 150 mM NaCl, 5% glycerol

pH 7.5 (TNG buffer) and incubated for 30 min with 5 mL Ni-nitrilotriacetic acid (Ni–NTA) Fast Flow resin (GE Healthcare, Uppsala, Sweden) in a gravity-fed column to capture the fusion protein via its affinity to Ni–NTA. The resin was washed with 25 mM imidazole to remove weakly bound proteins. The His$_6$-MBP-SA8 fusion protein was then eluted with 200 mM imidazole, and then it was concentrated by centrifugal filtration (Amicon® Ultra, Millipore) and desalted using a PD-10 column (GE Healthcare, Pittsburgh, PA, USA) to remove imidazole. The fusion protein was incubated with His$_6$-TEV protease (produced in-house; 40 µg per mg of fusion protein) for 16 h with shaking at 100 rpm and 30 °C. The cleavage mixture was then passed over Ni–NTA Fast Flow resin to remove His$_6$-MBP and His$_6$-TEV protease, and the eluate containing the liberated SA8 was collected for further purification using RP-HPLC. RP-HPLC was performed on a Vydac C18 column (250 × 10 mm) using a flow rate of 1 mL/min and a gradient of 30–50% solvent B (0.1% trifluoroacetic acid (TFA) in 90% acetonitrile) in solvent A (0.1% TFA in water) over 40 min with absorbance monitored at 214 nm. Fractions containing SA8 were collected and assessed by LC–MS. Fractions containing SA8 (> 99%) were lyophilised for NMR studies, mapping disulphide connectivity, and functional assays.

NMR spectroscopy and disulphide connectivity

A SA8 sample for NMR experiments was prepared by dissolving lyophilised peptide in 90% H$_2$O and 10% ^{2}H$_2$O and adjusting the pH by addition of 0.1 N HCl and 0.1 M NaOH. NMR spectra were acquired on a 600-MHz Bruker NMR spectrometer (Billerica, MA) equipped with a cryogenically cooled TCI probe. One-dimensional ^{1}H NMR spectra were acquired over the temperature range 5–30 °C in increments of 5 °C to investigate the effects of temperature on the spectrum before selecting conditions for structure determination. Bruker TopSpin, USA, version 3.6.1, was used for processing all spectra.

The disulphide connectivity was determined by analysing pepsin-digested recombinant, oxidised SA8 peptide

using LC–MS/MS. 5 μg SA8 was dissolved in 10 μL 10 mM hydrochloric acid and digested with 20 ng/μL pepsin (Merck, cat. No. 10108057001) at 37 °C for 6 h. The digested sample was diluted with formic acid (FA) to yield a final concentration of 0.3 ng/μL in 15 μL 0.5% FA, from which 1 μL was analysed by LC–MS/MS. The sample was separated with a nanoElute nano-HPLC (Bruker, Germany) across a Bruker TEN C18 nano separation column (100 mm length, 75 μm inner diameter, 1.9 μm particle size, 120 Å pore size) heated to 50 °C, using a gradient of 5–30% solvent B (90% acetonitrile [ACN] in 0.1% formic acid [FA]) in solvent A (0.1% FA) over 18 min before a step gradient to 95% B for 18 min at a flow rate of 0.5 μL/min. The eluting peptides were analysed with an in-line Bruker timsTOF Pro operated in positive, data-dependent analysis, parallel accumulation-serial fragmentation (DDA-PASEF) mode with a cycle time of 0.5 s and an m/z range of 350–2200. The resulting MS and MS2 spectra were manually compared to a list of all possible peptic fragment precursors and their fragment ions estimated from all theoretical disulphide permutations. The disulphide bond connectivities identified in the recombinant peptide were confirmed with a sample of a natural peptide from *T. stephensoni*.

Co-elution of native and recombinant SA8

Crude extract containing the native SA8 peptide was isolated using a protocol based on that of Honma et al. [142]. Briefly, tissue from the body column of *T. stephensoni* was placed into tubes containing a single stainless-steel ball bearing (Qiagen) and homogenised using a Tissuelyser II (Qiagen). Subsequently, Milli-Q H_2O was added to the tubes containing the macerate, which were then centrifuged for 15 min at 18,000 g, allowing the supernatant to be recovered. A second round of homogenisation and centrifugation was conducted after additional H_2O was added to the tubes containing pelleted material. The supernatant from both rounds of extraction was combined, then flash-frozen in liquid nitrogen and lyophilized.

The toxin extract sample was diluted with 500 μL of Milli-Q water and centrifuged for 1 min at 17,680 g. The sample was loaded onto Sep-Pak C18 Vac cartridge/5000 mg (Waters) that was previously equilibrated with 1 mL methanol and 1 mL Milli-Q water. The sample was then washed with 500 μl of Milli-Q water and eluted with 750 μl of TFA/Milli-Q/acetonitrile (1:70:29). The eluate was evaporated under a constant stream of N_2. The residue was reconstituted in Milli-Q water (80 μL) and analysed using LC–MS. An analytical Luna C8 LC column (100 mm×2 mm) was utilised for co-elution of native and the recombinant peptides with a gradient of 0–60% buffer B (90% ACN, 0.1% trifluoroacetic acid

[TFA]) in buffer A (0.1% TFA) and constant flow rate of 0.2 mL/min over 10 min.

In vitro refolding of reduced recombinant SA8 peptide

The recombinant SA8 peptide was reduced and unfolded in 10 mM dithiothreitol (DTT), 20 mM Tris at pH 8. Complete reduction occurred after approximately 30 min, as confirmed using LC–MS. Following reduction, a PD-10 column was used to desalt the sample and remove DTT. The reduced SA8 was then oxidised using 0.1 M NH_4HCO_3 buffer at pH 8, with complete oxidation occurring after 8 h, as verified using LC–MS. 1D ^1H NMR spectra were acquired at pH 3.5 and 25 °C for the major re-oxidised product to check folding.

In vitro and in vivo functional testing
Patch clamp electrophysiology

CHO and human embryonic kidney (HEK 293) cells were grown in DMEM-high glucose supplemented with 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin-g, and 100 μg/mL streptomycin (Invitrogen) at 37 °C in a 5% CO_2 and 95% air humidified atmosphere. Cells were passaged twice per week following a 5-min incubation in PBS containing 0.2 g/L EDTA (Invitrogen, Carlsbad, CA).

CHO cells (ATCC) were transiently transfected with plasmids encoding $hK_V1.1$, $hK_V1.2$, $hK_V10.1$, hKCa3.1, hTRPA1, or $hNa_V1.7$ using Lipofectamine 2000 (Invitrogen, Carlsbad, CA) following the manufacturer's protocol, then cultured under standard conditions. The cells were co-transfected with a plasmid encoding green fluorescent protein (GFP) at a molar ratio of 10:1 except for $hK_V1.1$, $hK_V1.2$, and hKCa3.1 for which GFP-tagged ion channel vectors were used. Currents were recorded 24–36 h after transfection. GFP-positive transfectants were identified with a Nikon Eclipse TS100 fluorescence microscope (Nikon, Tokyo, Japan) using bandpass filters of 455–495 nm and 515–555 nm for excitation and emission, respectively, and these cells were used for current recordings (>70% success rate for transfection). HEK 293 cells stably expressing the hKv11.1 channel (hERG, hKCNH2 gene, a kind gift from H. Wulff, University of California, Davis, CA) and mKCa1.1 channel (BKCa, hKCNMA1, a kind gift from C. Beeton, Baylor College of Medicine, Houston, TX) were used. hTRPV1 channels were expressed in a stable manner in CHO cells. Transfected cells were washed twice with 2 mL of ECS (see below) and re-plated onto 35-mm polystyrene cell culture dishes (Cellstar, Greiner Bio-One). $hK_V1.3$ currents were recorded from activated peripheral blood mononuclear cells (PBMCs, see below) 3–4 days after activation. Human veinous blood was obtained from anonymised healthy donors. PBMCs were isolated using Histopaque1077 (Sigma-Aldrich Hungary, Budapest, Hungary)

density gradient centrifugation. Cells obtained were resuspended in RPMI 1640 medium containing 10% foetal calf serum (Sigma-Aldrich Hungary, Budapest, Hungary), 100 μg/mL penicillin, 100 μg/mL streptomycin, and 2 mM L-glutamine, seeded in a 24-well culture plate at a density of $5–6 \times 10^5$ cells/mL, and grown in a 5% CO_2 incubator at 37 °C for 3–5 days. Phytohemagglutinin A (PHA, Sigma-Aldrich Hungary, Budapest, Hungary) was added to the medium at 10 μg/mL to amplify the $K_V 1.3$ expression. Cells were washed gently twice with 2 mL of ECS for the patch-clamp experiments.

Conventional whole-cell patch-clamp electrophysiology [143] was used to record ionic currents. Micropettes were pulled in four stages using a Flaming Brown automatic pipette puller (Sutter Instruments, San Rafael, CA) from Borosilicate Standard Wall with Filament aluminium–silicate glass (Harvard Apparatus Co., Holliston, MA) with tip diameters between 0.5 and 1 μm and heat polished to a tip resistance of 2–8 MΩ. All measurements were carried out by using Axopatch 200B amplifier connected to a personal computer using Axon Digidata 1550A data acquisition hardware (Molecular Devices, Sunnyvale, CA). In general, the holding potential was -120 mV. Records were discarded when a leak at the holding potential was $> 10\%$ of the peak current at the test potential. Series resistance compensation up to 70% was used to minimise voltage errors and achieve proper voltage-clamp conditions. Experiments were performed at room temperature (20–24 °C). Data were analysed using Prism 8 (GraphPad, CA, USA) and ClampFit 10.5 software (Molecular Devices Inc., Sunnyvale, CA). Before analysis, whole-cell current traces were corrected for ohmic leakage and digitally filtered with a three-point boxcar smoothing filter. For hKCa3.1, hTRPA1 and hTRPV1, the reversal potential for K^+ was determined and only those currents were analysed for which the reversal potential fell into the range of the theoretical reversal potential ± 5 mV (-86.5 ± 5 mV for KCa3.1, and 0 ± 5 mV for hTRPA1 and hTRPV1).

Solutions: All salts and components of the solutions were purchased from Sigma-Aldrich, Budapest, Hungary. For $hK_V 1.1$, $hK_V 1.2$, $hK_V 1.3$, $hK_V 10.1$, mKCa1.1, and $hNa_V 1.7$, the ECS contained 145 mM NaCl, 5 mM KCl, 2.5 mM $CaCl_2$, 1 mM $MgCl_2$, 10 mM HEPES, and 5.5 mM glucose (pH 7.35 with NaOH), while the intracellular (pipette) solution (ICS) contained 140 mM KF, 2 mM $MgCl_2$, 1 mM $CaCl_2$, 11 mM EGTA, and 10 mM HEPES (pH 7.22 with KOH). For recordings of hNav1.7 currents, the composition of ICS was 105 mM CsF, 10 mM NaCl, 10 mM EGTA, and 10 mM HEPES, (pH 7.2 with CsOH). For recordings of hKv11.1 currents, the ECS contained 140 mM choline-chloride, 5 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 0.1 mM $CdCl_2$, 20 mM glucose, and

10 mM HEPES (pH 7.35 with NaOH) and the ICS contained 140 mM KCl, 10 mM EGTA, 2 mM $MgCl_2$, and 10 mM HEPES (pH 7.3 with KOH). hKCa3.1 currents were recorded with an ECS of the following composition: 160 mM L-aspartic acid sodium salt, 5 mM KCl, 2.5 mM $CaCl_2$, 1.0 mM $MgCl_2$, 5,5 mM glucose and 10 mM HEPES (pH 7.4 with NaOH), while the composition of ICS was 150 mM L-aspartic acid potassium salt, 5 mM HEPES, 10 mM EGTA, 8.7 mM $CaCl_2$, and 2 mM $MgCl_2$ (pH 7.22 with KOH) giving ~ 1.2 μM free Ca^{2+} to fully activate the KCa3.1 current [144]. For recordings of TRPA1 and TRPV1 currents, the ECS and ICS contained 150 mM NaCl, 2 mM Na-EDTA, and 10 mM HEPES (pH 7.35 with NaOH). hTRPA1 and hTRPV1 currents were activated by 100 μM allyl-isothiocyanate (AITC) and 1 μM capsaicin, respectively, and SA8 was added after complete activation of the current. SA8, activators and positive controls were dissolved in ECS, except for TEA^+ that was stored in 100 mM stock at 4 °C. The osmolarity of the ECS and ICS were 302–308 mOsM and ~ 295 mOsM, respectively. SA8 and positive controls were dissolved in ECS supplemented with 0.1 mg/mL bovine serum albumin (BSA; Sigma-Aldrich Hungary, Budapest, Hungary). Bath perfusion around the measured cell with different extracellular solutions was achieved using a gravity flow microperfusion system using a rate of 0.5 mL/min. Excess fluid was removed continuously.

Voltage protocols: For measurement of $hK_V 1.1–1.3$ and $hK_V 10.1$, currents, voltage steps to $+50$ mV, were applied from a holding potential of -120 mV every 15 or 30 s and the peak current was measured. Channels were activated by a series of depolarisation pulses to $+50$ mV from a holding potential of -120 mV. For $K_V 1.1$, 50-ms-long activating stimuli were used every 15 s. Due to the highly variable activation kinetics of $K_V 1.2$ [145] 200-ms-long pulses were applied every 15 s to maximise the open probability of the channel. Due to the slow inactivation kinetics of $K_V 1.2$, the currents did not inactivate even at 200-ms-long depolarisation pulses. $K_V 1.3$ currents were evoked by 15-ms-long depolarisation pulses to $+50$ mV from -120 mV every 15 s. The use of such short pulses every 15 s was sufficient to fully activate the channels and ensured that there was no cumulative inactivation of $K_V 1.3$. Positive controls were applied at a concentration equivalent to their K_d values (14 nM ChTx for $K_V 1.2$, 0.3 mM TEA^+ for $K_V 1.1$, 2 μM Astemizole for $K_V 10.1$ and 10 mM TEA^+ for $K_V 1.3$ and $K_{Ca} 1.1$). The approximate 50% reduction in the current amplitude in the presence of these compounds was an indicator of both the ion channel and the proper operation of the perfusion system. For mKCa1.1 channels, a voltage step to $+100$ mV from a holding potential of -100 mV was used. Pulses were delivered every 10 s. hKCa3.1 currents were elicited

every 15 s with voltage ramps to +50 mV from a test potential of −120 mV at a rate of 0.85 mV/ms. The holding potential was set to −85 mV. For hK$_V$11.1 channels, currents were evoked with a voltage step to +20 mV followed by a step to −40 mV, during which the peak current was measured. The holding potential was −80 mV, and pulses were delivered every 30 s. hTRPA1 and hTRPV1 currents were elicited every 5 s with 200-ms-long voltage ramps to +50 mV from a test potential of −50 mV. The cells were held at 0 mV during the subsequent pulses. The TRPA1 and TRPV1 currents were activated by 100 µM AITC and 1 µM capsaicin, respectively. Currents through hNa$_V$1.7 channels were evoked every 15 s with a voltage step to +50 mV from a holding potential of −120 mV. The RCF at a given molar concentration was calculated as I/I_0, where I_0 is the peak current in the absence, and I is the peak current at equilibrium block or in the absence of inhibition after ~2 min perfusion by SA8. Data points on concentration–response curves represent the mean of four individual measurements. These curves were fitted via simple linear regression according to the equation $Y = 1 + \text{slope} \times [\text{toxin}]$, where Y is reciprocal of the RCF and [toxin] is the molar concentration of SA8. The reciprocal of the slope yielded the IC$_{50}$ value.

Injection activity assay in Drosophila melanogaster

The activity of the recombinant SA8 peptide in *Drosophila melanogaster* was assessed using an injection activity assay [76]. The mass of female *D. melanogaster* used for this experiment ranged from 0.8 to 1.1 mg. Eight *D. melanogaster* were used as an injection control group and administered 50 nL of water. Five doses of SA8 (0.00005, 0.0005, 0.005, 0.05 and 0.5 µg per fly; 50 nL) were each administered to eight *D. melanogaster* (*n* total = 40), with observations of paralysis and mortality made at 2 and 24 h post-injection (Additional file 14: File S6). Dose–response data were analysed as described previously [76].

Abbreviations

SA8	Sea anemone 8
ICK	Inhibitor cystine knot
SMART	Simple Modular Architecture Research Tool
BUSCO	Benchmarking Universal Single-Copy Orthologs
LTR	Long terminal repeat
MITE	Miniature Inverted-repeat Terminal Element
CAFE	Computational Analysis of gene Family Evolution
PLA$_2$	Phospholipase A$_2$
SVMP	Snake venom metalloproteinase
WGCNA	Weighted gene co-expression network analysis
MSI	Mass spectrometry imaging
CRISP	Cysteine-rich secretory protein
LS-MS	Liquid chromatography-mass spectrometry
LC–MS/MS	Liquid chromatography-tandem mass spectrometry
NMR	Nuclear magnetic resonance
His$_6$	Polyhistidine
MBP	Maltose binding protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
TEV	Tobacco etch virus
RP-HPLC	Reversed-phase high-performance liquid chromatography
1D	One-dimensional
CHO	Chinese hamster ovary
TEA$^+$	Tetraethylammonium
ECS	Extracellular solution
ChTx	Charybdotoxin
SEM	Standard error of mean
RCF	Remaining current fraction

Supplementary Information

Additional file 1: Table S1. Repeats masked in the genomes of *A. tenebrosa* and *T. stephensoni*.

Additional file 2: Table S2. CAFE was used to identify expanded gene families encoding putative toxinsacross actiniarian superfamilies Actinioidea, Edwardsioidea and Metridioidea. Species: Telmatactis = *T. stephensoni*; Paraphelliactis = *P. xishaensis*; Exaiptasia = *E. diaphana*; Actinia = *A. tenebrosa*; Nematostella = *N. vectensis*. Uncharacterised toxin gene families have been designated U# or Z#.

Additional file 3: File S1. Supporting data values for figures. **Fig. 2A_I.** MCScanX duplication classification. **Fig. 2A_II.** MCScanX collinear blocks. **Fig. 3A.** *A. tenebrosa* SA8 expression matrix. **Fig. 3B.** *T. stephensoni* SA8 expression matrix. **Fig. 4A.** *T. stephensoni* WGCNA module membership. **Fig. 4B.** *A. tenebrosa* WGCNA module membership. **Fig. 6.** SA8 phylogenetic distance matrix. **Fig. 8_Fig. S9.** patch clamp electrophysiology measurements.

Additional file 4: Figure S1. Sequence and structure of *T. stephensoni* SA8 sequences. Alignment of SA8 gene and peptide sequences from *T. stephensoni,* with conserved cysteine framework, glycine residue, and FA dyad highlighted. The nine clustered genes of *T. stephensoni* are composed of 3–5 microexons and large introns. The inverted *T. stephensoni* SA8 gene is indicated in red.

Additional file 5: Figure S2. SDS-PAGE gel stained with Coomassie Blue showing samples of the His6-MBP-SA8 fusion protein obtained during different steps of expression and purification. Lane 1: molecular mass standards; Lanes 2 and 3: *E. coli* cells pre- and post-induction with IPTG; Lane 4: sucrose extract; Lane 5: periplasmic extract; Lanes 6 and 7: purified MBP fusion protein before and after cleavage with TEV protease.

Additional file 6: Figure S3. RP-HPLC purification of recombinant SA8. The peptide was separated on a Vydac C18 columnusing a flow rate of 1 mL/min and a 40 min linear gradient of 30–50% solvent B, as indicated by the red line. LC-MS profile of oxidised SA8 after purification using RP-HPLC. The observed mass is consistent with formation of three disulfide bonds.

Additional file 7: Figure S4. Comparison of LC-MS profiles for co-eluted native and recombinant SA8. LC-MS profile of the native SA8 peptide. LC-MS profile of the recombinant SA8 peptide.

Additional file 8: Table S3. Direct mass spectrometric characterisation of disulfide linkages in the recombinant SA8 peptide. The main diagnostic pepsin fragments appeared as MS2 precursors with high intensity.

Additional file 9: Figure S5. Full 1D ^{1}H NMR spectrum of recombinant SA8 recorded on Bruker 600 MHz NMR spectrometer at pH 3.5 and 298 K and 256 scans. Expanded amide/aromatic region.

Additional file 10: Figure S6. 1D ^{1}H NMR spectra of SA8 recorded on Bruker 600 MHz NMR spectrometer. Spectra were acquired at pH 3.5 and different temperatures over the range 5–30 °C. Expanded amide/aromatic region.

Additional file 11: Figure S7. 1D ^{1}H NMR spectra of SA8 recorded on Bruker 600 MHz NMR spectrometer. Spectra were acquired at 298 K using 256 scans, over the pH range 2-7. Expanded amide/aromatic region.

Ashwood *et al. BMC Biology*   (2023) 21:121

Page 21 of 25

Additional file 12: Figure S8. 1D ^{1}H NMR spectra of SA8 recorded on Bruker 600 MHz NMR spectrometer. Spectra were acquired at pH 3.5 and 298 K using 256 scans. Expanded amide/aromatic region. N.B. The sharp strong peaks in the spectra are from low molecular weight molecules.

Additional file 13: Figure S9. SA8 has no effect on several voltage-gated and Ca^{2+}-activated ion channels. Current traces recorded before application of SA8, after 1-2 min perfusion with SA8, and after perfusing the recording chamber with control solutions. Data are shown for the following channels: $HK_V1.1$, $hK_V1.3$, $hK_V10.1$, $hK_V11.1$, $mKCa1.1$, $hNa_V1.7$, $hKCa3.1$, hTRPA1, and hTRPV1. For details on the expression systems, solutions, and voltage protocols, see Materials and Methods. For hKCa3.1, hTRPA1, and hTRPV1, the currents were recorded in response to a voltage ramp, corrected for ohmic leakage and then displayed as a function of test potential. The horizonal dashed line shows the zero current level and the vertical dashed line indicates the expected reversal potential for K^+.

Additional file 14: File S6. Injection activity assay in *Drosophila melanogaster*.

Additional file 15: Table S4. Accession numbers for tissue samples of *A. tenebrosa* and *T. stephensoni*.

Acknowledgements

Authors' contributions

Conceptualization, L.M.A., K.A.E., R.S.N. and P.J.P; transcriptomic data analysis, L.M.A.; genomic data analysis, L.M.A., Z.K.S., C.A.V.B., H.L.S., J.M.S., V.L.C., M.J.P., K.J.D. and P.J.P; sequence space analysis: T.S.; structural and functional analysis: K.A.E., M.U.N., T.G.S., S.G. and D.C.C.W.; proteomic data analysis, E.A.B.U., B.M. and B.R.H.; original draft preparation, L.M.A., K.A.E., Z.K.S., T.G.S., R.S.N. and P.J.P; draft review and editing, L.M.A., K.A.E., Z.K.S., T.S., M.U.N., T.G.S., C.A.V.B., H.L.S., J.M.S., E.A.B.U., B.M., B.R.H., S.G., D.C.C.W., V.L.C., M.J.P., K.J.D., D.A.H., G.P., G.F.K., A.P., R.S.N. and P.J.P; supervision, E.A.B.U., D.C.C.W., D.A.H., G.P. G.F.K., A.P., R.S.N. and P.J.P. All authors have read and agreed to the published version of the manuscript.

Funding

Availability of data and materials

All data generated or analysed during this study are included in this published article, its supplementary information files, and publicly available repositories. Raw RNA reads for *T. stephensoni* and *A. tenebrosa* are available at the NCBI Sequence Read Archive (SRA) under BioProjects PRJNA728752 [58, 146], PRJNA350366 [33, 147] and PRJNA313244 [104, 148]. Accession numbers for each tissue sample are given in Additional file 15: Table S4. Assembled genomes for *T. stephensoni* and *A. tenebrosa* have been uploaded to NCBI Gen-Bank under accession numbers JALIZF000000000 [149] and JALIZE000000000 [150], respectively. The MS proteomics data have been uploaded to the ProteomeXchange Consortium via the PRIDE [151] partner repository with the dataset identifier PXD029717 [58, 152]. Custom pipelines and scripts can be accessed through the zkstewart GitHub repository [79–81, 86, 103, 120, 129, 132]. Genome annotations and alignment files used in this study have also been uploaded as GFF3 and FASTA files to figshare (https://doi.org/10.6084/m9.figshare.22777052).

Declarations

Ethics approval and consent to participate

This study was confirmed as Outside the Scope and exempt from the need for the University Animal Ethics Committee (UAEC) review, approval, and monitoring in conformity with the Australian code for the care and use of animals for scientific purposes (approval number 1800000229).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]School of Biology and Environmental Science, Faculty of Science, Queensland University of Technology, Brisbane, QLD 4000, Australia. [2]Cancer Program, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia. [3]Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, VIC 3052, Australia. [4]Centre for Agriculture and the Bioeconomy, Queensland University of Technology, Brisbane, QLD 4000, Australia. [5]Department of Animal Plant & Soil Sciences, La Trobe University, Melbourne, Australia. [6]Swinburne University of Technology, Melbourne, VIC, Australia. [7]Department of Biophysics and Cell Biology, Faculty of Medicine, University of Debrecen, 4032 Debrecen, Hungary. [8]Department of Anatomy, School of Biomedical Sciences, University of Otago, Dunedin 9016, New Zealand. [9]Department of Ecology, Evolution and Behavior, Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, 9190401 Jerusalem, Israel. [10]Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Blindern, PO Box 1066, 0316 Oslo, Norway. [11]Centre for Advanced Imaging, The University of Queensland, St Lucia, QLD 4072, Australia. [12]Institute for Molecular Bioscience, The University of Queensland, St Lucia, QLD 4072, Australia. [13]Centre for Microscopy and Microanalysis, The University of Queensland, St Lucia, QLD 4072, Australia. [14]Research Infrastructure, Central Analytical Research Facility, Queensland University of Technology, Brisbane, QLD 4000, Australia. [15]ARC Centre for Innovations in Peptide and Protein Science, The University of Queensland, St Lucia, QLD 4072, Australia. [16]School of Biomedical Sciences, Faculty of Health, Queensland University of Technology, Brisbane, QLD 4000, Australia. [17]ARC Centre for Fragment-Based Design, Monash University, Parkville, VIC 3052, Australia.

References

1. Schendel V, Rash LD, Jenner RA, Undheim EAB. The diversity of venom: the importance of behavior and venom system morphology in understanding its ecology and evolution. Toxins. 2019;11(11):666.
2. Frazão B, Vasconcelos V, Antunes A. Sea anemone (Cnidaria, Anthozoa, Actiniaria) toxins: an overview. Mar Drugs. 2012;10(8):1812–51.
3. Rachamim T, Morgenstern D, Aharonovich D, Brekhman V, Lotan T, Sher D. The dynamically evolving nematocyst content of an anthozoan, a scyphozoan, and a hydrozoan. Mol Biol Evol. 2015;32(3):740–53.
4. Romey G, Abita JP, Schweitz H, Wunderer G. Sea anemone toxin: a tool to study molecular mechanisms of nerve conduction and excitation-secretion coupling. Proc Natl Acad Sci U S A. 1976;73(11):4055–9.
5. Jungo F, Bairoch A. Tox-Prot, the toxin protein annotation program of the Swiss-Prot protein knowledgebase. Toxicon. 2005;45(3):293–301.
6. Prentis PJ, Pavasovic A, Norton RS. Sea anemones: quiet achievers in the field of peptide toxins. Toxins. 2018;10(1):36.
7. Madio B, Peigneur S, Chin YK, Hamilton BR, Henriques ST, Smith JJ, et al. PHAB toxins: a unique family of predatory sea anemone toxins evolving via intra-gene concerted evolution defines a new peptide fold. Cell Mol Life Sci. 2018;75(24):4511–24.

8. Orts DJB, Peigneur S, Silva-Gonçalves LC, Arcisio-Miranda M, Bicudo JEPW, Tytgat J. AbeTx1 is a novel sea anemone toxin with a dual mechanism of action on shaker-type K$^+$ channels activation. Mar Drugs. 2018;16(10):360.

9. Madio B, King GF, Undheim EAB. Sea anemone toxins: a structural overview. Mar Drugs. 2019;17(6):325.

10. Undheim EAB, Mobli M, King GF. Toxin structures as evolutionary tools: using conserved 3D folds to study the evolution of rapidly evolving peptides. BioEssays. 2016;38(6):539–48.

11. 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(10):1833–9.

12. Rodríguez AA, Salceda E, Garateix AG, Zaharenko AJ, Peigneur S, López O, et al. A novel sea anemone peptide that inhibits acid-sensing ion channels. Peptides. 2014;53:3–12.

13. Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A. 1998;95(11):5857–64.

14. Castañeda O, Sotolongo V, Amor AM, Stöcklin R, Anderson AJ, Harvey AL, et al. Characterization of a potassium channel toxin from the Caribbean Sea anemone *Stichodactyla helianthus*. Toxicon. 1995;33(5):603–13.

15. Beeton C, Pennington MW, Norton RS. Analogs of the sea anemone potassium channel blocker ShK for the treatment of autoimmune diseases. Inflamm Allergy Drug Targets. 2011;10(5):313–21.

16. Kalman K, Pennington MW, Lanigan MD, Nguyen A, Rauer H, Mahnir VM, et al. ShK-Dap22, a potent Kv1.3-specific immunosuppressive polypeptide. J Biol Chem. 1998;273(49):32697–707.

17. Yan L, Herrington J, Goldberg E, Dulski PM, Bugianesi RM, Slaughter RS, et al. *Stichodactyla helianthus* peptide, a pharmacological tool for studying Kv3.2 channels. Mol Pharmacol. 2005;67(5):1513–21.

18. Tudor JE, Pallaghy PK, Pennington MW, Norton RS. Solution structure of ShK toxin, a novel potassium channel inhibitor from a sea anemone. Nat Struct Biol. 1996;3(4):317–20.

19. Pennington MW, Beeton C, Galea CA, Smith BJ, Chi V, Monaghan KP, et al. Engineering a stable and selective peptide blocker of the Kv1.3 channel in T lymphocytes. Mol Pharmacol. 2009;75(4):762–73.

20. Chi V, Pennington MW, Norton RS, Tarcha EJ, Londono LM, Sims-Fahey B, et al. Development of a sea anemone toxin as an immunomodulator for therapy of autoimmune diseases. Toxicon. 2012;59(4):529–46.

21. Tajti G, Wai DCC, Panyi G, Norton RS. The voltage-gated potassium channel K(V)1.3 as a therapeutic target for venom-derived peptides. Biochem Pharmacol. 2020;181:114146.

22. Tarcha EJ, Olsen CM, Probst P, Peckham D, Munoz-Elias EJ, Kruger JG, et al. Safety and pharmacodynamics of dalazatide, a Kv1.3 channel inhibitor, in the treatment of plaque psoriasis: A randomized phase 1b trial. PLoS One. 2017;12(7):e0180762.

23. Krishnarjuna B, Villegas-Moreno J, Mitchell ML, Csoti A, Peigneur S, Amero C, et al. Synthesis, folding, structure and activity of a predicted peptide from the sea anemone *Oulactis* sp. with an ShKT fold. Toxicon. 2018;150:50–9.

24. Sunanda P, Krishnarjuna B, Peigneur S, Mitchell ML, Estrada R, Villegas-Moreno J, et al. Identification, chemical synthesis, structure, and function of a new KV1 channel blocking peptide from *Oulactis* sp. Pept Sci. 2018;110(4):e24073.

25. Mitchell ML, Tonkin-Hill GQ, Morales RA, Purcell AW, Papenfuss AT, Norton RS. Tentacle transcriptomes of the speckled anemone (Actiniaria: Actiniidae: *Oulactis* sp.): venom-related components and their domain structure. Mar Biotechnol. 2020;22:207–2019.

26. Krishnarjuna B, MacRaild CA, Sunanda P, Morales RAV, Peigneur S, Macrander J, et al. Structure, folding and stability of a minimal homologue from *Anemonia sulcata* of the sea anemone potassium channel blocker ShK. Peptides. 2018;99:169–78.

27. Shafee T, Mitchell ML, Norton RS. Mapping the chemical and sequence space of the ShKT superfamily. Toxicon. 2019;165:95–102.

28. Sachkova MY, Landau M, Surm JM, Macrander J, Singer SA, Reitzel AM, et al. Toxin-like neuropeptides in the sea anemone *Nematostella* unravel recruitment from the nervous system to venom. Proc Natl Acad Sci U S A. 2020;117(44):27481–92.

29. Kozlov S, Grishin E. The mining of toxin-like polypeptides from EST database by single residue distribution analysis. BMC Genomics. 2011;12:88-.

30. Madio B, Undheim EAB, King GF. Revisiting venom of the sea anemone *Stichodactyla haddoni*: omics techniques reveal the complete toxin arsenal of a well-studied sea anemone genus. J Proteomics. 2017;166:83–92.

31. Oliveira JS, Fuentes-Silva D, King GF. Development of a rational nomenclature for naming peptide and protein toxins from sea anemones. Toxicon. 2012;60(4):539–50.

32. Macrander J, Broe M, Daly M. Tissue-specific venom composition and differential gene expression in sea anemones. Genome Biol Evol. 2016;8(8):2358–75.

33. Surm JM, Smith HL, Madio B, Undheim EAB, King GF, Hamilton BR, et al. A process of convergent amplification and tissue-specific expression dominates the evolution of toxin and toxin-like genes in sea anemones. Mol Ecol. 2019;28(9):2272–89.

34. Wilding CS, Fletcher N, Smith EK, Prentis PJ, Weedall GD, Stewart ZK. The genome of the sea anemone *Actinia equina* (L.): meiotic toolkit genes and the question of sexual reproduction. Mar Genomics. 2020;53:100753.

35. Baumgarten S, Simakov O, Esherick LY, Liew YJ, Lehnert EM, Michell CT, et al. The genome of *Aiptasia*, a sea anemone model for coral symbiosis. Proc Natl Acad Sci U S A. 2015;112(38):11893–8.

36. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, et al. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science. 2007;317(5834):86–94.

37. Zimmermann B, Robb SMC, Genikhovich G, Fropf WJ, Weilguny L, He S, et al. Sea anemone genomes reveal ancestral metazoan chromosomal macrosynteny. bioRxiv. 2020;2020:359448.

38. Surm JM, Stewart ZK, Papanicolaou A, Pavasovic A, Prentis PJ. The draft genome of *Actinia tenebrosa* reveals insights into toxin evolution. Ecol Evol. 2019;9(19):11314–28.

39. McFadden CS, Quattrini AM, Brugler MR, Cowman PF, Dueñas LF, Kitahara MV, et al. Phylogenomics, Origin, and Diversification of Anthozoans (Phylum Cnidaria). Syst Biol. 2021;70(4):635–47.

40. Giorgianni MW, Dowell NL, Griffin S, Kassner VA, Selegue JE, Carroll SB. The origin and diversification of a novel protein family in venomous snakes. Proc Natl Acad Sci U S A. 2020;117(20):10911–20.

41. Dowell Noah L, Giorgianni Matt W, Kassner Victoria A, Selegue Jane E, Sanchez Elda E, Carroll SB. The deep origin and recent loss of venom toxin genes in rattlesnakes. Curr Biol. 2016;26(18):2434–45.

42. Torres AM, de Plater GM, Doverskog M, Birinyi-Strachan LC, Nicholson GM, Gallagher CH, et al. Defensin-like peptide-2 from platypus venom: member of a class of peptides with a distinct structural fold. Biochem J. 2000;348 Pt 3(Pt 3):649–56.

43. Torres AM, Wang X, Fletcher JI, Alewood D, Alewood PF, Smith R, et al. Solution structure of a defensin-like peptide from platypus venom. Biochem J. 1999;341(Pt 3):785–94.

44. Whittington CM, Papenfuss AT, Kuchel PW, Belov K. Expression patterns of platypus defensin and related venom genes across a range of tissue types reveal the possibility of broader functions for OvDLPs than previously suspected. Toxicon. 2008;52(4):559–65.

45. Whittington CM, Papenfuss AT, Bansal P, Torres AM, Wong ESW, Deakin JE, et al. Defensins and the convergent evolution of platypus and reptile venom genes. Genome Res. 2008;18(6):986.

46. Moran Y, Weinberger H, Sullivan JC, Reitzel AM, Finnerty JR, Gurevitz M. Concerted evolution of sea anemone neurotoxin genes is revealed through analysis of the *Nematostella vectensis* genome. Mol Biol Evol. 2008;25(4):737–47.

47. Sachkova MY, Singer SA, Macrander J, Reitzel AM, Peigneur S, Tytgat J, et al. The birth and death of toxins with distinct functions: a case study in the sea anemone *Nematostella*. Mol Biol Evol. 2019;36(9):2001–12.

48. Smith EG, Surm JM, Macrander J, Simhi A, Amir G, Sachkova MY, et al. Dominant toxin hypothesis: unravelling the venom phenotype across micro and macroevolution. bioRxiv. 2022;2022.06.22.497252.

49. Fautin DG, Mariscal RN. Cnidaria: Anthozoa. In: Harrison FW, A WJ, editors. Microscopic anatomy of invertebrates 2. New York: Wiley-Liss; 1991. p. 267–358.

50. Wallace CC, Crowther AL. Hexacorals 1: Sea anemones and anemone-like animals (Actiniaria, Zoantharia, Corallimorpharia, Ceriantharia and Antipatharia). In: Hutchings P, Kingsford M, Hoegh-Guldberg O, editors.

Ashwood *et al. BMC Biology*   (2023) 21:121

Page 23 of 25

The Great Barrier Reef: Biology. Environment and Management: CSIRO; 2019. p. 257–66.

51. Daly M. The anatomy, terminology, and homology of acrorhagi and pseudoacrorhagi in sea anemones. Zoologische Verhandelingen. 2003;345:89–101.

52. Wong ESW, Belov K. Venom evolution through gene duplications. Gene. 2012;496(1):1–7.

53. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290(5494):1151–5.

54. Rodin SN, Parkhomchuk DV, Rodin AS, Holmquist GP, Riggs AD. Repositioning-dependent fate of duplicate genes. DNA Cell Biol. 2005;24(9):529–42.

55. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 1999;151(4):1531–45.

56. Hargreaves A, Swain M, Hegarty M, Logan D, Mulley J. Restriction and recruitment — gene duplication and the origin and evolution of snake venom toxin. Genome Biol Evol. 2014;6(8):2088–95.

57. Smith HL, Pavasovic A, Surm JM, Phillips MJ, Prentis PJ. Evidence for a Large Expansion and Subfunctionalization of Globin Genes in Sea Anemones. Genome Biol Evol. 2018;10(8):1892–901.

58. Ashwood LM, Undheim EAB, Madio B, Hamilton BR, Daly M, Hurwood DA, et al. Venoms for all occasions: the functional toxin profiles of different anatomical regions in sea anemones are related to their ecological function. Mol Ecol. 2021;31(3):866–83.

59. Östman C, Kultima JR, Roat C, Rundblom K. Acontia and mesentery nematocysts of the sea anemone *Metridium senile* (Linnaeus, 1761) (Cnidaria: Anthozoa). Sci Mar. 2010;74(3):483–97.

60. Lam J, Cheng Y-W, Chen W-NU, Li H-H, Chen C-S, Peng S-E. A detailed observation of the ejection and retraction of defense tissue acontia in sea anemone (Exaiptasia pallida). PeerJ. 2017;5:e2996.

61. Sunagar K, Columbus-Shenkar YY, Fridrich A, Gutkovich N, Aharoni R, Moran Y. Cell type-specific expression profiling unravels the development and evolution of stinging cells in sea anemone. BMC Biol. 2018;16:1–16.

62. Smith HL, Prentis PJ, Bryan SE, Norton RS, Broszczak DA. Acontia, a Specialised Defensive Structure, Has Low Venom Complexity in Calliactis polypus. Toxins. 2023;15(3):218.

63. Kini RM. Molecular moulds with multiple missions: functional sites in three-finger toxins. Clin Exp Pharmacol Physiol. 2002;29(9):815–22.

64. Kini RM. Excitement ahead: structure, function and mechanism of snake venom phospholipase A2 enzymes. Toxicon. 2003;42(8):827–40.

65. Kini RM, Doley R. Structure, function and evolution of three-finger toxins: mini proteins with multiple targets. Toxicon. 2010;56(6):855–67.

66. Lynch VJ. Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. BMC Evol Biol. 2007;7:2.

67. Casewell NR, Wagstaff SC, Harrison RA, Renjifo C, Wüster W. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. Mol Biol Evol. 2011;28(9):2637–49.

68. Guo M, Teng M, Niu L, Liu Q, Huang Q, Hao Q. Crystal structure of the cysteine-rich secretory protein stecrisp reveals that the cysteine-rich domain has a K+ channel inhibitor-like fold. J Biol Chem. 2005;280(13):12405–12.

69. Pennington MW, Chang SC, Chauhan S, Huq R, Tajhya RB, Chhabra S, et al. Development of highly selective Kv1.3-blocking peptides based on the sea anemone peptide ShK. Mar Drugs. 2015;13(1):529–42.

70. Rashid MH, Heinzelmann G, Huq R, Tajhya RB, Chang SC, Chhabra S, et al. A potent and selective peptide blocker of the Kv1.3 channel: prediction from free-energy simulations and experimental confirmation. PLoS One. 2013;8(11):e78712.

71. Kalman K, Pennington MW, Lanigan MD, Nguyen A, Rauer H, Mahnir V, et al. ShK-Dap22, a potent Kv1.3-specific immunosuppressive polypeptide. J Biol Chem. 1998;273(49):32697–707.

72. Dauplais M, Lecoq A, Song J, Cotton J, Jamin N, Gilquin B, et al. On the convergent evolution of animal toxins. Conservation of a diad of functional residues in potassium channel-blocking toxins with unrelated structures. J Biol Chem. 1997;272(7):4302–9.

73. Barros F, Pardo LA, Domínguez P, Sierra LM, de la Peña P. New Structures and Gating of Voltage-Dependent Potassium (Kv) Channels and

Their Relatives: A Multi-Domain and Dynamic Question. Int J Mol Sci. 2019;20(2):248.

74. Tomczak AP, Fernández-Trillo J, Bharill S, Papp F, Panyi G, Stühmer W, et al. A new mechanism of voltage-dependent gating exposed by K(V)10.1 channels interrupted between voltage sensor and pore. J Gen Physiol. 2017;149(5):577–93.

75. Wang W, MacKinnon R. Cryo-EM Structure of the Open Human Ether-à-go-go-Related K(+) Channel hERG. Cell. 2017;169(3):422-30.e10.

76. Guo S, Herzig V, King GF. Dipteran toxicity assays for determining the oral insecticidal activity of venoms and toxins. Toxicon. 2018;150:297–303.

77. 10x Genomics. Sample Preparation Demonstrated Protocol: DNA Extraction from Single Insects. https://support.10xgenomics.com/de-novo-assembly/sample-prep/doc/demonstrated-protocol-dna-extraction-from-single-insects. Accessed 12 Jan 2018.

78. Liu H, Wu S, Li A, Ruan J. SMARTdenovo: a *de novo* assembler using long noisy reads. Preprints. 2020.

79. Stewart ZK. Arrow polishing pipeline. https://github.com/zkstewart/Genome_analysis_scripts/tree/master/pipeline_scripts/polishing_pipeline_scripts/arrow. Accessed 26 Nov 2019.

80. Stewart ZK. Pilon polishing pipeline. https://github.com/zkstewart/Genome_analysis_scripts/tree/master/pipeline_scripts/polishing_pipeline_scripts/pilon. Accessed 11 Nov 2019.

81. Stewart ZK. deGRIT. https://github.com/zkstewart/deGRIT. Accessed 9 Oct 2019.

82. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 2018;19(1):460.

83. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100.

84. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

85. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2.

86. Stewart ZK. Genome Stats. https://github.com/zkstewart/Genome_analysis_scripts/blob/master/genome_stats.py. Accessed 16 Sept 2019.

87. Han Y, Wessler SR. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res. 2010;38(22):e199.

88. Ye C, Ji G, Liang C. detectMITE: A novel approach to detect miniature inverted repeat transposable elements in genomes. Sci Rep. 2016;6:19688.

89. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150–2.

90. Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. BMC Bioinformatics. 2008;9:18.

91. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007;35:W265–8.

92. Ou S, Jiang N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. 2018;176(2):1410–22.

93. Smit A, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015. http://www.repeatmasker.org. Accessed 5 Mar 2018.

94. Smit A, Hubley R. RepeatModeler Open-1.0. 2008–2015. http://www.repeatmasker.org. Accessed 13 Dec 2018.

95. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7(10):e1002195.

96. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42:D222–30.

97. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015;43:D222–6.

98. Piriyapongsa J, Rutledge MT, Patel S, Borodovsky M, Jordan IK. Evaluating the protein coding potential of exonized transposable element sequences. Biol Direct. 2007;2:31.

99. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45(D1):D158–69.

Ashwood *et al. BMC Biology*   (2023) 21:121

Page 24 of 25

100. Fredman D, Schwaiger M, Rentzsch F, Technau U. *Nematostella vectensis* transcriptome and gene models v2.0. https://doi.org/10.6084/m9.figshare.807696.v1. Accessed 8 Jan 2021.

101. Shinzato C, Shoguchi E, Kawashima T, Hamada M, Hisata K, Tanaka M, et al. Using the *Acropora digitifera* genome to understand coral responses to environmental change. Nature. 2011;476(7360):320–3.

102. Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, Weinmaier T, et al. The dynamic genome of *Hydra*. Nature. 2010;464(7288):592–6.

103. Stewart ZK. Repeat pipeline. https://github.com/zkstewart/Genome_analysis_scripts/tree/master/pipeline_scripts/repeat_pipeline_scripts. Accessed 11 Nov 2019.

104. van der Burg CA, Prentis PJ, Surm JM, Pavasovic A. Insights into the innate immunome of actiniarians using a comparative genomic approach. BMC Genomics. 2016;17(1):850.

105. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.

106. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8(8):1494.

107. Macmanes MD. On the optimal trimming of high-throughput mRNA sequence data. Front Genet. 2014;5:13.

108. Gilbert D. Gene-omes built from mRNA seq not genome DNA. 7th annual arthropod genomics symposium; Notre Dame. 2013.

109. Visser EA, Wegrzyn JL, Steenkmap ET, Myburg AA, Naidoo S. Combined de novo and genome guided assembly and annotation of the *Pinus patula* juvenile shoot transcriptome. BMC Genomics. 2015;16:1057.

110. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, et al. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics. 2014;30(12):1660–6.

111. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18(5):821–9.

112. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics. 2012;28(8):1086–92.

113. Shao M, Kingsford C. Accurate assembly of transcripts through phase-preserving graph decomposition. Nat Biotechnol. 2017;35(12):1167–9.

114. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003;31(19):5654–66.

115. Kent WJ. BLAT — the BLAST-like alignment tool. Genome Res. 2002;12(4):656–64.

116. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics. 2005;21(9):1859–75.

117. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 2016;32(5):767–9.

118. Stanke M, Steinkamp R, Waack S, Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 2004;32(Web Server issue):W309-12.

119. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9(1):R7.

120. Stewart ZK. Gene model curation processing pipeline. https://github.com/zkstewart/Genome_analysis_scripts/blob/master/gene_annotation_pipeline/gene_model_curation/processing_pipeline.sh. Accessed 29 Nov 2019.

121. Feng C, Liu R, Xu W, Zhou Y, Zhu C, Liu J, et al. The genome of a new anemone species (Actiniaria: Hormathiidae) provides insights into deep-sea adaptation. Deep Sea Res Part I Oceanogr Res Pap. 2021;170:103492.

122. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238.

123. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16:157.

124. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013;30(4):772–80.

125. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37(5):1530–4.

126. Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics. 2020;36(22–23):5516–8.

127. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology The Gene Ontology Consortium. Nat Genet. 2000;25(1):25–9.

128. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar Gustavo A, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412–9.

129. Stewart ZK. Annotation table. https://github.com/zkstewart/Genome_analysis_scripts/tree/master/annotation_table. Accessed 27 Apr 2022.

130. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017;35(11):1026–8.

131. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, the UniProt C. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics. 2015;31(6):926–32.

132. Stewart ZK. CAFE enrichment analysis. https://github.com/zkstewart/Various_scripts/blob/master/DGE/enrichment/cafe_enrichment_analysis.py. Accessed 5 May 2022.

133. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol. 2008;25(7):1307–20.

134. Garber M, Grabherr MG, Guttman M, Trapnell C. Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Methods. 2011;8(6):469–77.

135. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.

136. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559-.

137. Mitchell ML, Shafee T, Papenfuss AT, Norton RS. Evolution of cnidarian trans-defensins: sequence, structure and exploration of chemical space. Proteins. 2019;87(7):551–60.

138. Adler D, Nenadic O, Zucchini W, editors. Rgl: A r-library for 3d visualization with opengl. Proceedings of the 35th Symposium of the Interface: Computing Science and Statistics; 2003; Salt Lake City.

139. Pearson R. The GoodmanKruskal package: measuring association between categorical variables. https://cran.r-project.org/web/packages/GoodmanKruskal/vignettes/GoodmanKruskal.html.

140. Froger A, Hall JE. Transformation of plasmid DNA into *E. coli* using the heat shock method. J Vis Exp. 2007;6:253.

141. Neu HC, Heppel LA. The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts. J Biol Chem. 1965;240(9):3685–92.

142. Honma T, Hasegawa Y, Ishida M, Nagai H, Nagashima Y, Shiomi K. Isolation and molecular cloning of novel peptide toxins from the sea anemone *Antheopsis maculata*. Toxicon. 2005;45(1):33–41.

143. Hamill OP, Marty A, Neher E, Sakmann B, Sigworth FJ. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. Pflugers Arch. 1981;391(2):85–100.

144. Grissmer S, Nguyen AN, Cahalan MD. Calcium-activated potassium channels in resting and activated human T lymphocytes. Expression levels, calcium dependence, ion selectivity, and pharmacology. J Gen Physiol. 1993;102(4):601–30.

145. Rezazadeh S, Kurata HT, Claydon TW, Kehl SJ, Fedida D. An activation gating switch in Kv1.2 is localized to a threonine residue in the S2–S3 linker. Biophys J. 2007;93(12):4173–86.

146. Tissue specific expression of toxins in *Telmatactis stephensoni*. Sequence Read Archive. 2021. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA728752

147. Toxin and toxin-like genes show dynamic gene family evolution and expression patterns in phylum Cnidaria. Sequence Read Archive. 2016. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA350366

148. Transcriptome sequencing of multiple Actiniaria species. Sequence Read Archive. 2016. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA313244

149. *Telmatactis stephensoni*. GenBank. 2023. https://www.ncbi.nlm.nih.gov/nuccore/JALIZF000000000.1

150. *Actinia tenebrosa*. GenBank. 2023. https://www.ncbi.nlm.nih.gov/nuccore/JALIZE000000000.1

Ashwood *et al. BMC Biology*   (2023) 21:121

Page 25 of 25

151. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. 2019;47(D1):D442–50.

152. Venoms for all occasions: the functional toxin profiles of different anatomical regions in sea anemones are related to their ecological function. PRIDE Database. 2022. https://www.ebi.ac.uk/pride/archive/projects/PXD029717

Publisher's Note