

RESEARCH ARTICLE

Open Access

A RNA-seq approach to identify putative toxins from acrorhagi in aggressive and non-aggressive *Anthopleura elegantissima* polyps

Jason Macrander^{1*}, Mercer R Brugler^{2,3} and Marymegan Daly¹

Abstract

Background: The use of venom in intraspecific aggression is uncommon and venom-transmitting structures specifically used for intraspecific competition are found in few lineages of venomous taxa. Next-generation transcriptome sequencing allows robust characterization of venom diversity and exploration of functionally unique tissues. Using a tissue-specific RNA-seq approach, we investigate the venom composition and gene ontology diversity of acrorhagi, specialized structures used in intraspecific competition, in aggressive and non-aggressive polyps of the aggregating sea anemone *Anthopleura elegantissima* (Cnidaria: Anthozoa: Hexacorallia: Actiniaria: Actiniidae).

Results: Collectively, we generated approximately 450,000 transcripts from acrorhagi of aggressive and non-aggressive polyps. For both transcriptomes we identified 65 candidate sea anemone toxin genes, representing phospholipase A2s, cytolytins, neurotoxins, and acrorhagins. When compared to previously characterized sea anemone toxin assemblages, each transcriptome revealed greater within-species sequence divergence across all toxin types. The transcriptome of the aggressive polyp had a higher abundance of type II voltage gated potassium channel toxins/Kunitz-type protease inhibitors and type II acrorhagins. Using toxin-like proteins from other venomous taxa, we also identified 612 candidate toxin-like transcripts with signaling regions, potentially unidentified secretory toxin-like proteins. Among these, metalloproteinases and cysteine rich (CRISP) candidate transcripts were in high abundance. Furthermore, our gene ontology analyses identified a high prevalence of genes associated with “blood coagulation” and “positive regulation of apoptosis”, as well as “nucleoside: sodium symporter activity” and “ion channel binding”. The resulting assemblage of expressed genes may represent synergistic proteins associated with toxins or proteins related to the morphology and behavior exhibited by the aggressive polyp.

Conclusion: We implement a multifaceted approach to investigate the assemblage of expressed genes specifically within acrorhagi, specialized structures used only for intraspecific competition. By combining differential expression, phylogenetic, and gene ontology analyses, we identify several candidate toxins and other potentially important proteins in acrorhagi of *A. elegantissima*. Although not all of the toxins identified are used in intraspecific competition, our analysis highlights some candidates that may play a vital role in intraspecific competition. Our findings provide a framework for further investigation into components of venom used exclusively for intraspecific competition in acrorhagi-bearing sea anemones and potentially other venomous animals.

Keywords: Phospholipase A2, Cytolysin, Sodium channel toxin, Potassium channel toxin, Acrorhagins, Venom, Cysteine, Neurotoxins, Intraspecific Competition

* Correspondence: macrander.1@osu.edu

¹The Ohio State University, Evolution, Ecology, and Organismal Biology, 318 W. 12th Avenue, Columbus, OH 43210-1293, USA

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

Background

Venomous animals use specialized structures to transmit a cocktail of noxious peptides into other organisms for defense or predation. Although the use of venom for intraspecific competition has been implied [1,2], specific examples of venom used on conspecifics via an intraspecific venom delivery system are rare [3]. Sea anemones are thus unique among venomous animals, as many species participate in intraspecific aggressive encounters using specialized structures to transmit venom [4-7]. Within the family Actiniidae, several species engage in aggressive intraspecific encounters using structures called acrorhagi [5,7]. Acrorhagi are inflatable structures at the tentacle-column margin that bear holotrichous nematocysts (reviewed in [8]). During an aggressive encounter, the acrorhagi inflate and adhere to the opponent (Figure 1), leaving an acrorhagial peel on the opponent [9]. The whitish peel, aggressor's acrorhagi adhering to the victim (Figure 1), is surrounded by necrotic tissue [10].

In the sea anemone *Anthopleura elegantissima* (Actiniaria: Actiniidae), fierce competition for space in the coastal intertidal zone may have selected for strategies and behaviors that provide an advantage in intraspecific aggressive encounters [11-13]. These animals form dense clonal aggregations of asexually produced polyps that are physically distinct but closely spaced. Those polyps at the boundary of a clonal aggregation have a high number of acrorhagi proportionate to body size and often show signs of localized necrosis from acrorhagial peels of nearby non-clonemate anemones [13]. Acrorhagi-induced necrosis in

A. elegantissima may be the result of an autoimmune process by which the allogeneic acrorhagial peel is isolated and expelled or may be caused by acrorhagi-specific toxins and necrosis-inducing compounds. The frequency of acrorhagial application is greater in intraspecific interactions than in interspecific interactions [5], highlighting their importance in intraspecific competition. The ectoderm of an acrorhagus generally does not adhere to the body of its bearer and the structure is not activated during prey capture, suggesting that the stimulus for the reaction and the discharge of nematocysts is "not-self" chemical signals. The mechanism behind the localized necrosis at the molecular level remains unknown; however, acrorhagi have been shown to transmit venom [14] and other bioactive components [15].

Toxins that have been well characterized within sea anemones fall into three major classes: phospholipase A2s (PLA2s), cytolytins, and neurotoxins. Within each class, several types (or groups) have been described based on sequence similarity and pharmacological target [16-19]. PLA2 genes belong to a large gene family whose members play varied roles in membrane remodeling, localized inflammation, and cell membrane, lipid, and amino acid metabolism [20-23]. The functional role of PLA2s has been studied in several cnidarians [17,24,25]; in some of these cases, PLA2 activity is associated with skin irritation in humans (eg. *Millepora* sp., see [25]). Group I and II PLA2s have been labeled as functionally toxic; along with an unknown venom component, they hydrolyze phospholipids and disrupt the cell membrane [26,27].

Although classified into four paralogous groups, all cytolytins form pores in the cellular membrane, creating an ionic imbalance that results in cytolysis [18,28-31]. Unlike other classes of toxins discussed here, cytolytins do not have disulfide bonds, relying instead on several amino acid residues for proper folding [18,32]. In terms of function, cytolytins are ideal candidate agents for the localized necrosis observed in the victim of an intraspecific aggressive encounter, however, they cannot form pores in cnidarian cells because cnidarians lack the target lipid sphingomyelin in their cell membranes [18,33,34].

Neurotoxins, specifically voltage gated potassium channel (VGPC) and voltage gated sodium channel (VGSC) toxins, target residues on voltage gated ion channels, disrupting the normal flux of ions in to or out of the cell [35,36]. Diverse animal toxin genes target components within the VGPC and VGSC, including the elements that filter, activate, and close these channels [37-40]. VGPC are the most diverse type of ion channel [41], regulating a variety of cellular processes and functions by permitting the efflux of potassium ions across the membrane in response to cellular depolarization [16]. VGSC are transmembrane complexes consisting of four homologous domains, each of which comprises six subunits that span

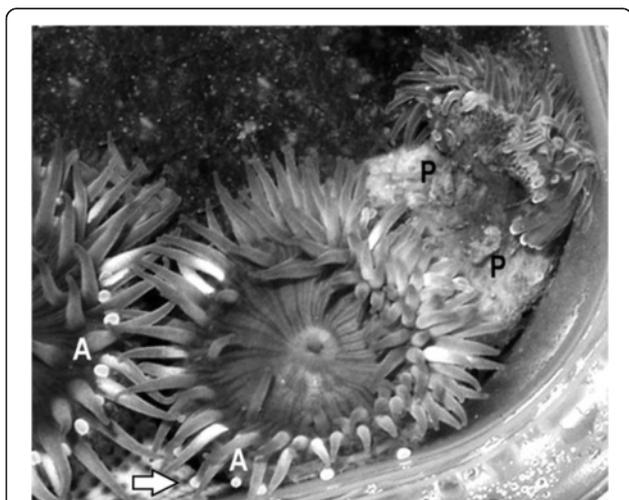


Figure 1 Intraspecific aggression in *Anthopleura elegantissima* polyps. Acrorhagi (A) are visible on the polyp in the center and the polyp to the left. A highly extended acrorhagus of the central polyp (arrow) is being applied to the column of the polyp on the left. Unlike the filiform tentacles, acrorhagi are opaque and rounded at the tip, even in extension. The polyp on the right has contracted in response to its encounter with the polyp in the center; its column is covered with mucus and several acrorhagial peels (P) from the central polyp.

the cellular membrane [42]. The VGSC enables the initiation and propagation of action potentials through a rapid release of sodium ions across the cell membrane [43].

Unlike other venomous animals, cnidarians lack a centralized venom gland. This may permit the evolution of specialized venom cocktails in association with specific tissues or structures. *Adamsia carciniopados* (Actiniaria: Hormathiidae) is the only sea anemone in which toxin activity was analyzed in multiple tissues; in this species, acontia showed higher phospholipase A2 activity than tentacles or whole body extracts [44]. Unfortunately, knowledge about the occurrence of functionally important venom in specific tissue types is rare: the majority of sea anemone toxins have been characterized from either whole animal or tentacle extracts (see Table six in reference [45]).

We characterize the diversity and abundance of toxins and potentially important peptides within the acrorhagi of *A. elegantissima*. Acrorhagi-specific toxins involved in intraspecific competition have been explored previously in *Actinia equina* (Actiniaria: Actiniidae) through a combined protein sequencing and RT-PCR approach [14], resulting in the identification of two candidate peptide toxins (acrorhagins). We sequenced RNA from acrorhagi of a single aggressive polyp and the (pooled) RNA from acrorhagi of several non-aggressive polyps. We screened each transcriptome for toxin genes using structural bioinformatics and phylogenetics. Gene networks of candidate toxin genes were used to investigate evolutionary patterns of gene diversity. We annotated candidate genes to highlight differences between the transcriptomes of acrorhagi from

aggressive and non-aggressive polyps and provide insight into the putative function of acrorhagi.

Results and discussion

Next-Gen sequencing

Our approach provides a comprehensive view of acrorhagi-specific venom toxins within aggressive and non-aggressive polyps of *A. elegantissima*. Each transcriptome was subjected to a suite of analyses to investigate the diversity of toxin-like sequences and overall gene ontology within acrorhagi of aggressive and non-aggressive polyps (Figure 2). Contrary to previous toxin specific studies using mass spectrometry, an RNA-seq approach permits the rapid identification of multiple toxins and their relative expression levels. We retrieve entire toxin transcripts, including the signal and propeptide region, which are cleaved in post translation modification. Similar approaches have been used for studies of the venom gland or duct of non-cnidarian venomous taxa [40,46,47]. Because cnidarians do not have a specialized venom gland, the majority of genes in our transcriptomes are presumably not involved in envenomation; the “acrorhagi” transcriptomes include transcripts from holotrichous nematocysts and adjacent cells that perform other functions. Although gland cells found within tentacles excrete components of sea anemone venom [48], their role in acrorhagi venom excretion has not been explored.

We found four unexpected outcomes in our comparison of the transcriptomes of acrorhagi from aggressive

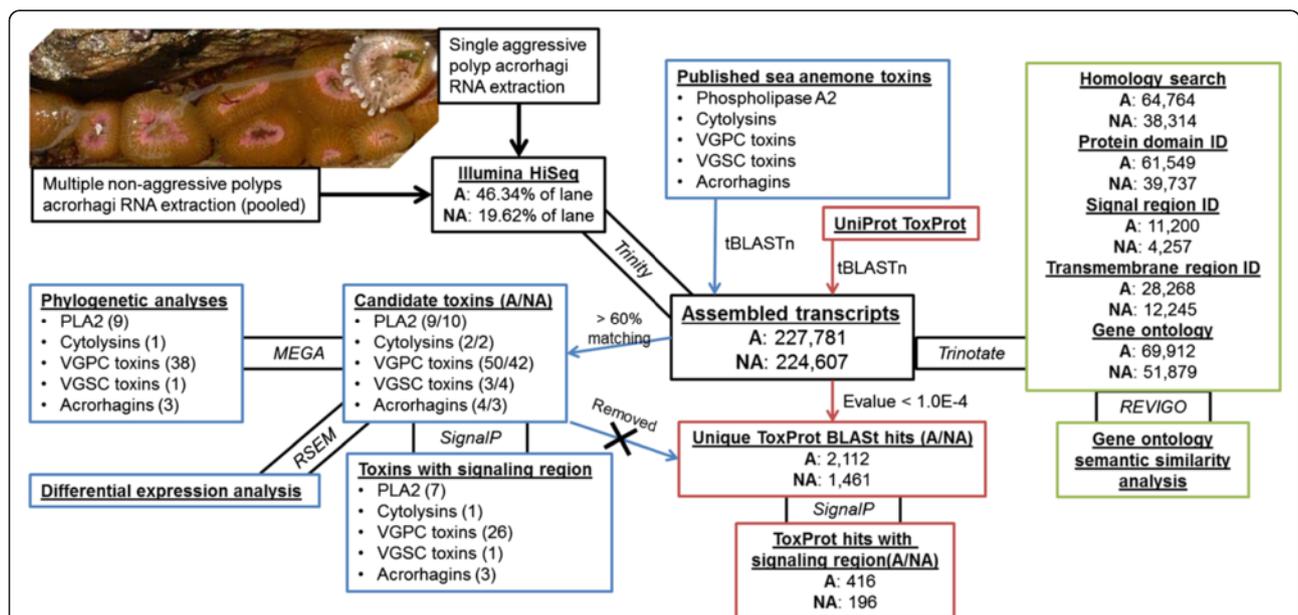


Figure 2 Analytical pipeline for acrorhagi transcriptomes. Colored boxes correspond to subsets of analyses with results reported in the text specific to sea anemone venom (blue), the UniProt animal toxin annotation, ToxProt (red), or transcriptome annotation, Trinotate (green). Text connecting these boxes indicate the analytical program used. Arrows with text indicates BLAST search strategies or thresholds used in initial screening.

and non-aggressive polyps. First, the size of the transcriptome did not influence our ability to retrieve candidate toxin genes. The aggressive polyp transcriptome was more than double the size of the non-aggressive polyp transcriptome in terms of raw sequence number (Table 1); nonetheless, we recover similar numbers of candidate toxin genes in each (Table 2). Second, there were no toxin genes expressed exclusively at high levels in either the aggressive or non-aggressive polyp transcriptome, which suggests that toxin components used in intraspecific aggression may always be expressed at some baseline level or that they cannot be identified based on homology to known toxin sequences. Third, acrorhagi contain a tremendous diversity of toxins. It is unlikely that every toxin gene is necessary for intraspecific aggression (discussed below); however, the array of toxin genes in acrorhagi of aggressive and non-aggressive polyps suggests that toxin genes may be expressed at some baseline level in most nematocysts or in glandular cells regardless of tissue, that there are alternative tissue specific forms for each class of toxin, or that acrorhagi retain some genes that reflect their “pre-acrorhagi” origin. Finally, despite having high sequence diversity and abundance among toxin genes, the frequency of multiple alleles or isoforms among our transcripts is relatively low compared to previous EST approaches in sea anemones [49-51]. It is likely that similar isoforms being expressed at low levels were incorporated into a single transcript [52].

Candidate toxin genes

By combining BLAST [53] searches with structural bioinformatics and toxin gene networks, we identified 65 candidate toxin genes that belong to five classes (Table 2; Additional file 1). We differentiate the relative levels of expression for each of these in each transcriptome, highlighting higher levels of expression in the acrorhagi of the aggressive polyp for the type II VGPC toxins/

Table 1 Summary of sequencing, cleanup, and assembly

Sequencing	Aggressive	Non-aggressive
Total reads (% of Lane)	189,025,968 (46.34%)	83,914,158 (19.62%)
Trimmomatic cleanup		
PE recovered (% of reads)	90,945,167 (96.23%)	41,957,079 (96.33%)
Forward only (% of reads)	2,993,563 (3.17%)	1,257,747 (3.00%)
Reverse only (% of reads)	263,778 (0.27%)	192,526 (0.46%)
Dropped (% of reads)	310,476 (0.33%)	90,639 (0.22%)
Trinity Assembly		
Transcripts	227,781	224,607
Components/genes'	150,270	184,129
N50	1832	961

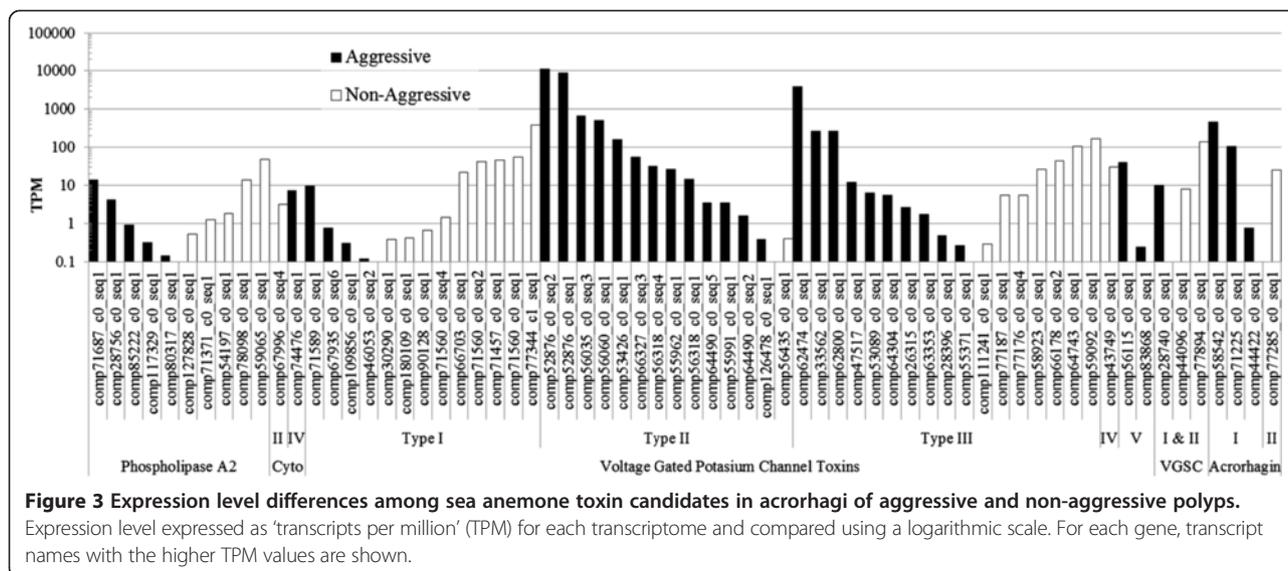
PE = Paired End. Modified sequences used in analyses can be found in Additional file 1.

Table 2 Summary of candidate sea anemone toxin sequences

	A	NA	N	C	S
Phospholipase A2	9	10	10	9	7
Cytolysins					
Type II	1	1	1	1	0
Type IV	1	1	1	-	1
VGPC Toxins					
Type I	9	10	13	11	8
Type II	19	12	14	13	4
Type III	19	17	17	14	12
Type IV	1	1	1	-	0
Type V	2	2	2	-	2
VGSC Toxins					
Type I, II, & III	3	4	3	3	1
Acrorhagins					
Type I	3	2	2	-	3
Type II	1	1	1	-	0
Total	69	59	65	42	38

A = aggressive polyp acrorhagi, NA = non-aggressive polyp acrorhagi, N = number of new toxin candidates (>90% similarity), C = number of gene clusters (>50 bootstrap support), S = genes with a signaling region; - = no phylogenetic reconstruction done.

Kunitz-type protease inhibitors and type I acrorhagins (Figure 3). The newly-identified candidate toxin genes include 10 PLA2s, two cytolysins, 47 VGPCs, three VGSCs, and three acrorhagins. Of the 65 toxin genes identified, 38 included the start codon and signaling region (Table 2). Due to low levels of sequence variation, we could not differentiate among type I, II or III VGSC toxins based on sequence similarity alone. For many toxin types, our data contributed a large (>25%) proportion of sequences in each respective toxin gene network (PLA2 and types I – III VGPC toxins). This likely reflects our RNA-seq approach, rather than any intrinsic property of the focal tissue or taxon. Alignment lengths of the mature toxin residues varied considerably between the candidate toxins (41 – 467; Additional file 2: Table S1). In cases where taxonomic diversity was limited for toxin genes (type IV cytolysins, acrorhagins, types IV and V VGPC toxins), we did not conduct a gene network analysis. For the remaining toxin types (PLA2s, type II cytolysins, type I-III VGPC toxins, and VGSC toxins), we conducted gene network reconstructions with our new candidate toxin genes in combination with previously described toxins. Gene network reconstructions permitted the grouping of isoforms into different genes based on high levels of sequence similarity. Even though the isoforms clustered, their high sequence divergence suggests several new toxin gene candidates for PLA2s (9), type I (11), type II (13), and type III (14) VGPC toxins (Table 2).



Phospholipases A2

Toxin gene networks were reconstructed for the mature PLA2 gene sequences from acrorhagi, previously characterized sea anemone PLA2s, and PLA2s from venomous and non-venomous vertebrate and non-vertebrate taxa (Figure 4). The resulting gene network suggests that *A. elegantissima* has at least six kinds of PLA2 in the acrorhagi, with some having multiple genes clustered together (Figure 4). At least three of these are part of a cluster that includes only genes from sea anemones (A_1 , B_{1a} , B_{1b} ; Figure 4). The remaining genes are part of a complex network of genes from vertebrates and non-vertebrate species (including other sea anemones) that are associated with venoms and not known to have toxic properties (A_2 ; Figure 4).

Within this complex network, PLA2 genes from the model organism *Nematostella vectensis* are associated with genes from the urochordate *C. intestinalis* and the placozoan *T. adhaerens* (see B_2 ; Figure 4). No sequences from our transcriptomes belong to this clade. The remaining sea anemone PLA2 genes nest within a cluster of genes from vertebrates (B_3 ; Figure 4). This gene cluster forms a large polytomy with a sea anemone exclusive cluster (C_1 ; Figure 4); a non-venomous vertebrate gene cluster that also contains a candidate sea anemone PLA2 gene (C_3 ; Figure 4); a mixed venomous and nonvenomous PLA2 gene cluster of vertebrates (C_2 ; Figure 4); and a cluster that contains genes with both venomous and non-venomous function with a single incomplete *A. elegantissima* candidate PLA2 toxin gene (C_4 ; Figure 4). The polytomies and relatively low bootstrap support for groups within this network likely reflect incomplete taxon sampling [54]. Nonetheless, the placement of the sea anemone PLA2 genes throughout the network is worth noting; sea anemone PLA2 genes (or clusters of genes) lie outside the

split between vertebrates and non-vertebrates, as would be predicted based on organismal phylogeny and also lie within a cluster of genes from vertebrates, suggesting an ancestral gene duplication event with subsequent gene loss. The network of PLA2s is complex in terms of the phylogenetic affinities of various members of the gene family and the nature of these in terms of their known use as venoms. Denser sampling of taxa and of copies within these taxa is necessary to understand the evolution and diversification of this gene family.

Cytolysins

We found one of each type II and type IV candidate cytolysin toxins in the transcriptomes of acrorhagi from aggressive and non-aggressive polyps. Due to the small number of type IV cytolysins previously identified, we did not construct a gene network for this toxin. In the gene network for type II cytolysin toxins, our new candidate gene from the acrorhagi of *A. elegantissima* groups as the sister to a cytolysin from *Sagartia rosea* (Figure 5), a finding that was supported whether we considered protein or nucleotide sequences. The relationship inferred from the cytolysin genes is at odds with the phylogeny of sea anemones [55]: the acrorhagi-bearing species *Ac. equina* (and *Anthopleura asiatica*) are more closely related to *A. elegantissima* than is the metridioidean *S. rosea*. The divergence observed in our candidate acrorhagi cytolysins is worth noting, as previously described cytolysins of *Ac. equina* were extracted from whole specimens, not acrorhagi [56,57]. Strong selective pressures or potentially loss of function could explain the observed divergence between the *A. elegantissima* cytolysin and *Ac. equina* cytolysin isolated from whole specimens. The high level of sequence variation may be indicative of a highly divergent or paralogous type II cytolysin in the acrorhagi

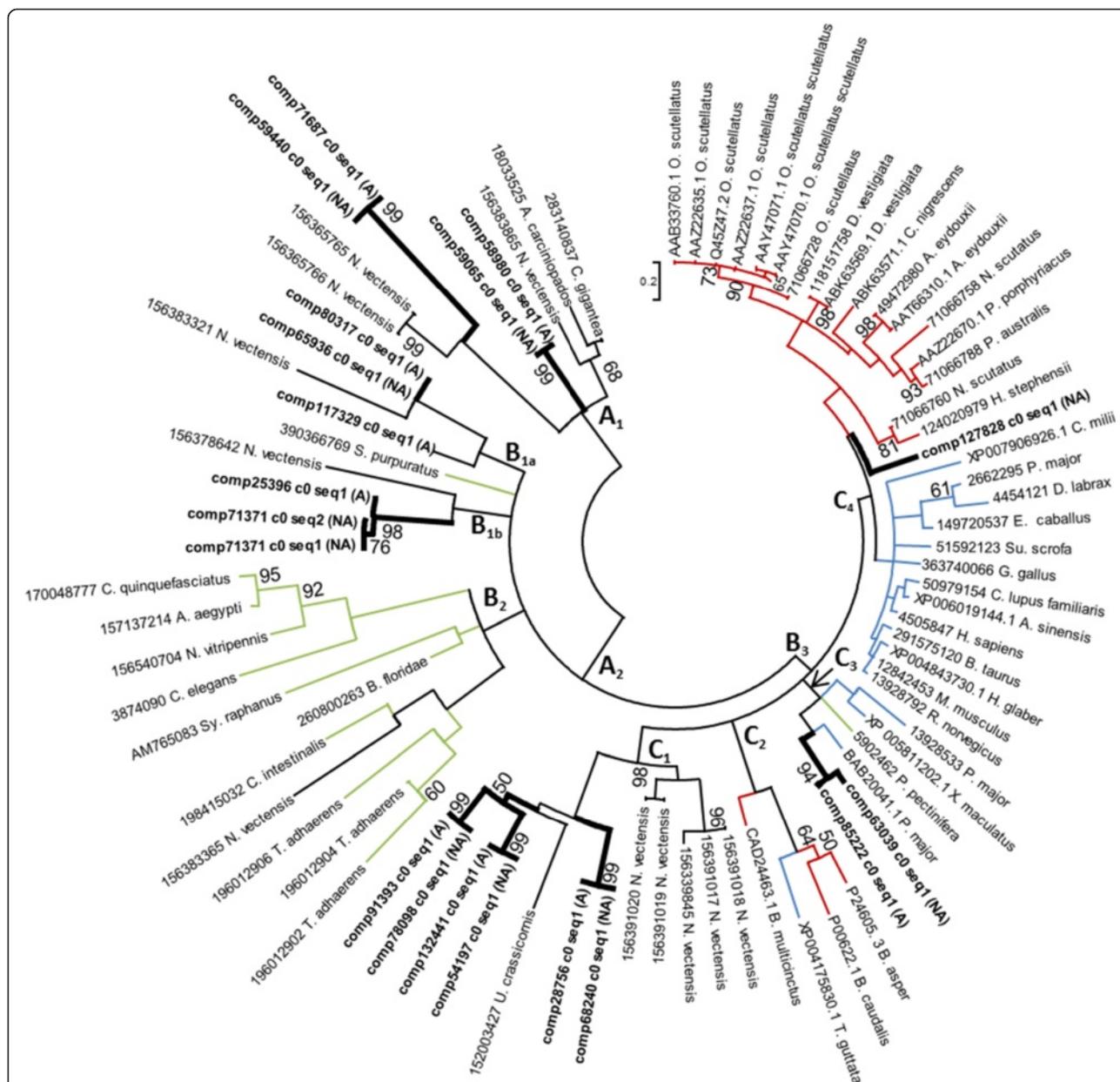


Figure 4 Maximum likelihood gene network of PLA2 genes. Color of the branches indicate function and origin: non-venomous PLA2 genes from non-cnidarian invertebrates (green, B₁ [*S. purpuratus*: sea urchin] and B₂ [*A. aegypti* & *C. quinquefasciatus*: mosquito; *B. floridae*: lancelet; *C. elegans*: nematode; *C. intestinalis*: tunicate; *N. vitripennis*: parasitoid wasp; *Sy. raphans*: sponge; *T. adhaerens*: placozoan]); PLA2 toxins found in vertebrates (red, C₂ [*B. asper*, *B. caudalis*, & *B. multicinctus*: snakes] and C₄ [*A. eydouxi*, *C. nigrescens*, *D. vestigiata*, *H. stephensii*, *N. scutatus*, *O. scutellatus*, *P. porphyriacus*, & *P. australis*: snakes]); non-venomous PLA2 genes found in vertebrates (blue, C₁ [*T. guttata*: zebra finch] C₃ [*P. major* & *X. maculatus*: fish] and C₄ [*A. sinensis*: alligator; *B. taurus*: cattle; *C. millii*: elephant fish; *C. lupus familiaris*: dog; *D. labrax* & *P. major*: fish; *E. caballus*: horse; *G. gallus*: chicken; *H. glaber*, *M. musculus*, & *R. norvegicus*: rodents; *H. sapien*: human; *Su. scrofa*: pig]) and an invertebrate (green C₃ [*P. pectiniifera*: sea star]); and PLA2 genes from sea anemones which may or may not be venomous (black, A₁ and A₂). Newly-identified candidate toxin genes are in bold with thick branches and the source is indicated (acrorhagi from A: aggressive or NA: non-aggressive polyps). Labels at the terminal tips indicate GenBank accession number and species identity. For full species names refer to S. Table 2. Bootstrap support values greater than 50 are shown.

of *A. elegantissima* that has undergone neofunctionalization or alternatively is losing its cytolytic potency. Whether this toxin still targets sphingomyelin, another lipid, or has completely lost its cytolytic function cannot be determined based on sequence data alone.

Voltage gated potassium channel toxins

In sea anemones, toxins that target components of the voltage gated potassium channel (VGPC) are interpreted to be diverse in origin and function, compared to those targeting the voltage gated sodium channel (VGSC) [45].

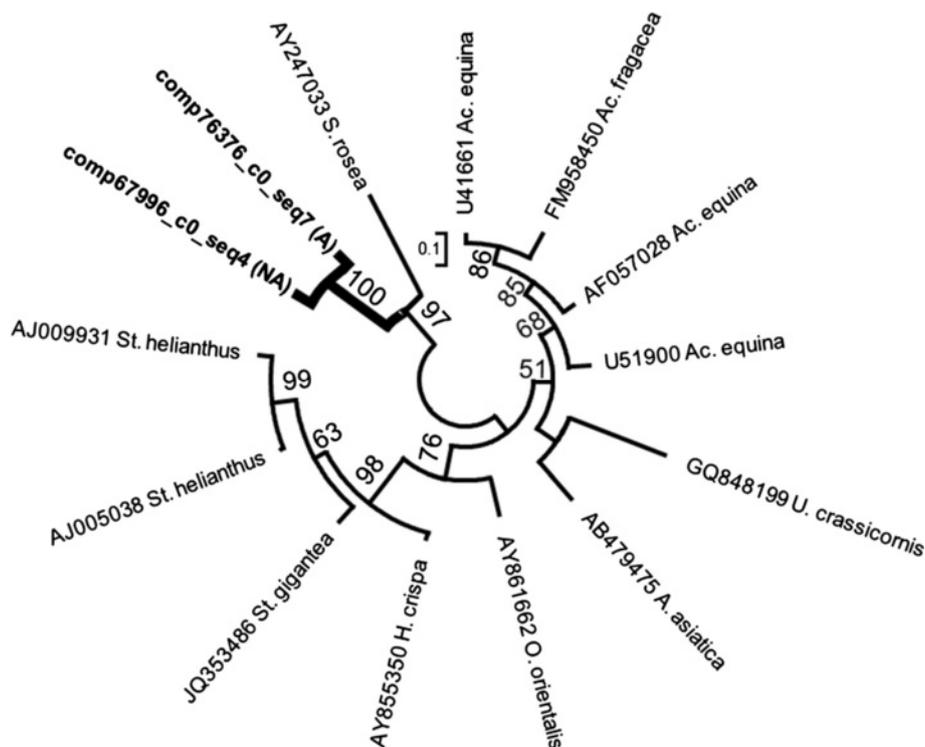


Figure 5 Maximum likelihood gene network of the type II cytolytic genes. Newly-identified candidate toxins are in bold with thick branches. The transcriptome source is indicated (acrorhagi from A: aggressive or NA: non-aggressive polyps). Labels at the terminal tips indicate GenBank accession number and species identity. For full species names refer to S. Table 2. Bootstrap support values greater than 50 are shown.

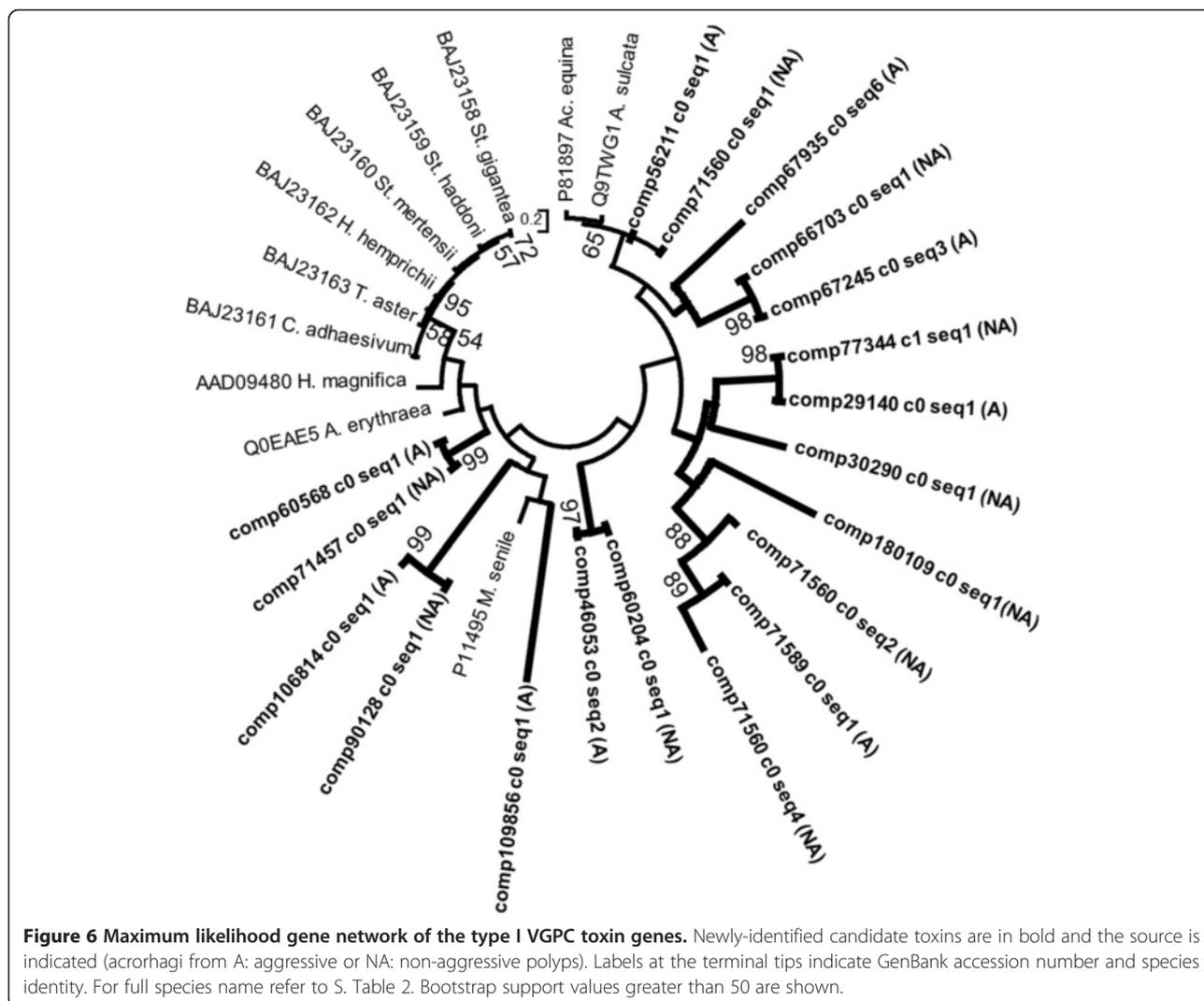
Sea anemone VGPC toxins have been classified into five types (I-V) based on amino acid composition, folding pattern, and target site [19,45,50,58]. We find candidate genes belonging to each type in the transcriptomes of acrorhagi from aggressive and non-aggressive polyps.

We identify 13 candidate genes that correspond to type I VGPC toxin genes based on shared sequence identity, conserved cysteine residues, and toxin gene network reconstruction (Table 2). The network for type I VGPC genes had relatively low support values throughout, with 11 candidate gene clusters having >50% bootstrap support values (Figure 6). Although the cysteine residues are maintained across our toxin candidates, the rate of amino acid substitution throughout the candidate toxin genes exceeds what has been reported previously (Additional file 2: Figure S1). In addition to these 13 candidate VGPC genes, we found several transcripts containing multiple (rather than single) type I VGPC toxin-like domains. These transcripts may also be type I VGPC toxins, as other toxins have been shown to go through post translational modifications [59].

Type II VGPC toxins are dual functioning, acting as VGPC blockers and Kunitz-type protease inhibitors (KPI). The role of KPI in sea anemone venom is not completely understood. They are inferred to inhibit endogenous proteases in their targets or to protect the

toxins against proteases following target injection [36]. KPI may have adopted VGPC blocking activity following slight changes in amino acid residues [60]. KPI have been recruited into venom in diverse lineages through convergent evolution [1]. From the sequence alignments, we were unable to distinguish between type II VGPC toxins and KPI and included both in our type II VGPC toxin network reconstruction. In this network, many KPI genes form well-supported groups (Figure 7); still other KPI genes outside of these clusters were found alongside known type II VGPC toxins. KPI and type II VGPC toxins show close affinity and this evidence of acquisition of VGPC blocking; sequences of known venom function from *Anemonia sulcata* are sister to genes that have KPI-type function (Figure 7). The lack of separation between type II VGPC toxins and KPI genes could reflect strong selection pressures needed to maintain KPI function, rather than convergence towards VGPC blocking activity. Regardless of cause, this close affinity between KPI and type II VGPC genes makes predicting the function of our candidate sequences difficult: based on the tree, it is impossible to determine whether the diversification represents a series of candidate type II toxins or a diversification of KPI genes or a combination of the two (Figure 7).

The majority of the type II VGPC/KPI candidates were expressed at higher levels in the acrorhagi of aggressive

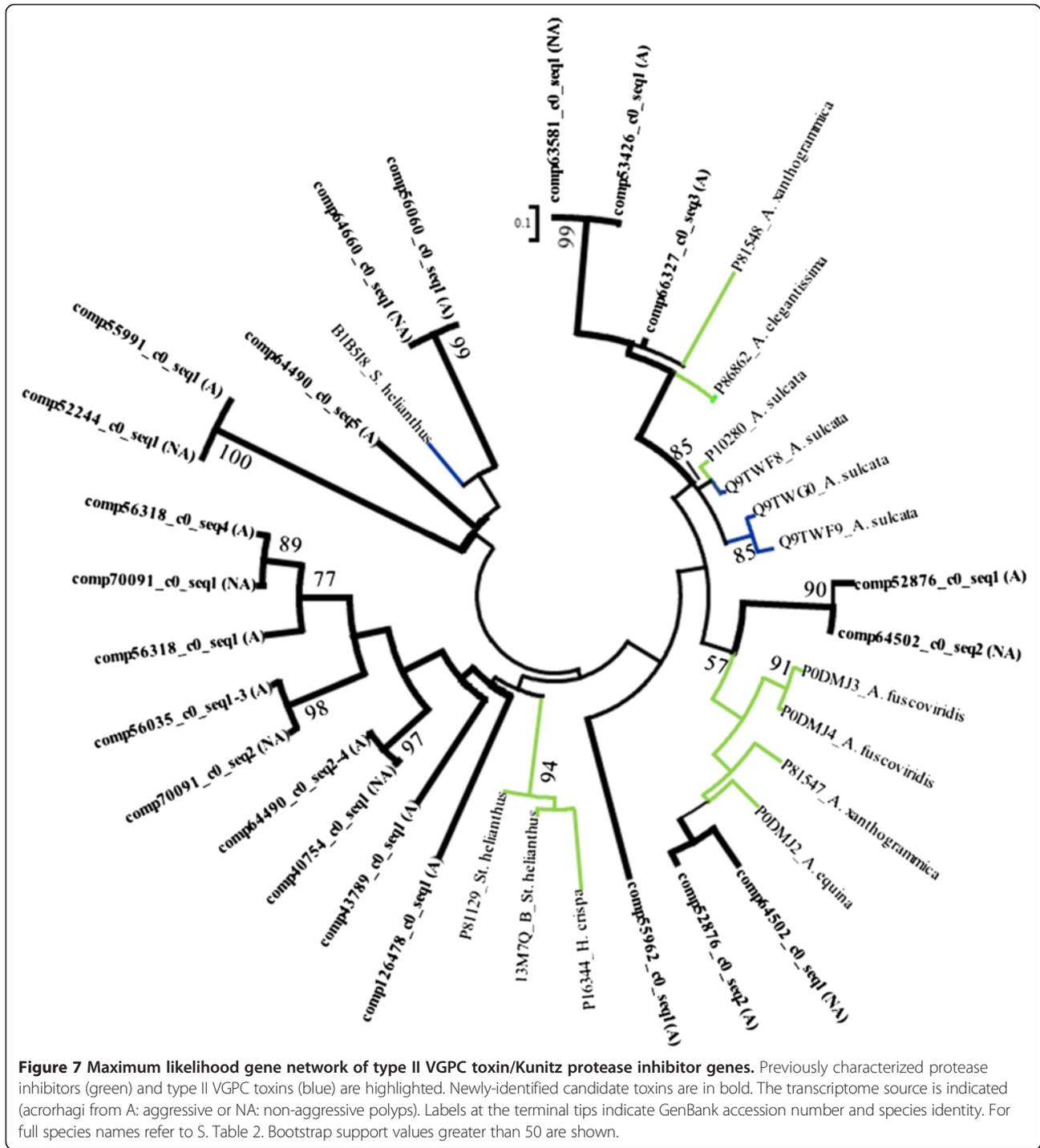


polyps (Figure 3). The two type II VGPC/KPI candidate toxins with the highest TPM values (comp64502_c0_seq1 & comp52876_c0_seq2) were part of a clade that contained genes that lack KPI function (Figure 7), which likely indicate that they are functionally KPI genes rather than both KPI and VGPC toxins. Although type II VGPC toxins are also functionally KPI genes [61], gene expression levels do not necessarily predict functional importance [62]. The high expression candidate KPI genes identified here may behave synergistically with functionally important toxins following target injection (i.e. the type II cytolysins and acrorhagins discussed below), rather than targeting the VGPC. The role of type II VGPC toxins as synergistic KPI proteins needs to be explored with regards to intraspecific envenomation, as well as their use in predation or defense.

Type III VGPC toxins are similar to VGSC toxins based on the placement of structurally important cysteine residues, but they show no effect on VGSC [36]. Type III VGPC toxins in sea anemones vary considerably in their

function and have been described as having multiple target sites [39,63,64]; this could provide plasticity among targets [62] and allow these toxins to modify more than just VGPC [16]. Our toxin gene network for type III VGPC toxins includes the acid sensing channel toxin (APETx2), as there were high levels of sequence similarity shared between this and other type III VGPC toxins. The toxin gene network has several new candidate type III VGPC toxins clustered with toxin genes previously characterized in *A. elegantissima* and in *Bunodosoma granuliferum*, *B. cangicum*, and *B. cassarum* (Figure 8). The type III VGPC toxin genes from *Anemonia viridis* and *An. sulcata* form a cluster apart from those from *Bunodosoma* and *Anthopleura* (Figure 8). The clustering of type III VGPC genes from the acrorhagi of *A. elegantissima* suggests that there are at least 13 forms of this toxin, including APETx2 genes (Figure 8).

Type IV VGPC toxins are relatively short, containing only two disulfide bonds and have been identified in only



two species of sea anemones [65,66]. We recovered a single type IV toxin candidate gene that was represented in both transcriptomes (Accession: GBXJ01030381); however, the transcript was incomplete and did not have a signaling region (Table 2).

The type V VGPC toxins appear to be relatively conserved and have been found in distantly related species

of Actiniaria [58]. We identified two candidate genes for type V VGPC toxins (Accession: GBXJ011150398 & GBXJ01058837); each of these has a unique signaling region (Table 2). The candidate toxin genes from each transcriptome had different amino acid sequences, which could be due to multiple alleles incorporated into a single transcript [52].

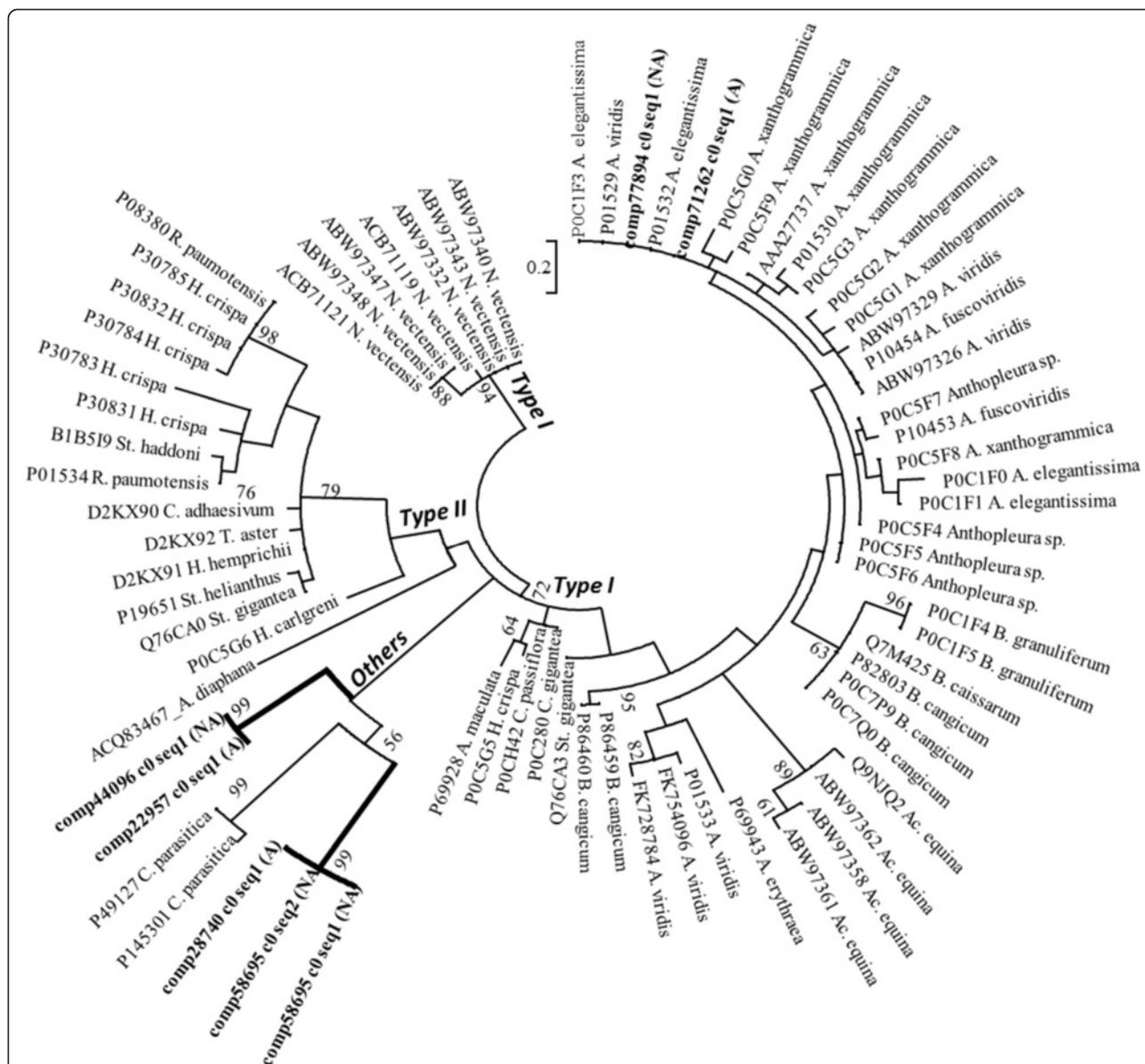


Figure 9 Maximum likelihood gene network of the VGSC toxin genes. The different toxin types (Type I, Type II, and "others") are noted. Newly-identified candidate toxins are in bold and the source is indicated (acrorhagi from A: aggressive or NA: non-aggressive polyps). Labels at the terminal tips indicate GenBank accession number and species identity. For full species name refer to S. Table 2. Bootstrap support values greater than 50 are shown.

two unique VGSC candidate toxin genes. The type I genes we identified cluster with sequences previously identified from *A. elegantissima* (Figure 9), as well as other sequences from other species of *Anthopleura* and its allies (members of family Actiniidae). The unique VGSC candidates grouped outside the type I and type II VGSC genes, clustering instead with the "other" VGSC toxin genes from *Calliactis parasitica* (Figure 9). This is the first report of candidate VGSC toxin genes for endomyarian sea anemones clustering outside the type I and type II groupings [45].

Acrorhagins

Acrorhagins were first described from acrorhagi of *Ac. equina* and were originally thought to contribute to the phenotypic response observed in targets of intraspecific aggression [14]. Acrorhagins have also been isolated from acrorhagi of *Anthopleura xanthogrammica* and *A. fuscoviridis* [70]. Unlike those of *Ac. equina*, the acrorhagins of *A. xanthogrammica* and *A. fuscoviridis* exhibit no lethality to crabs [70]. The candidate acrorhagin I sequences from *A. elegantissima* are slightly longer (by 15 – 20 AA) than those from *Ac. equina* [14] but they have a similar signal

peptide (Additional file 2: Figure S2). The candidate acrorhagin II sequences from *A. elegantissima* lacked any signaling region (Additional file 2: Figure S2). In contrast to the assumption that acrorhagins are unique to acrorhagi, we identify via reciprocal BLAST searches a candidate acrorhagin I toxin gene from an EST library from the sea anemone *Metridium senile* (Additional file 2: Figure S2). Although *M. senile* is distantly related and lacks acrorhagi, it does engage in aggressive intraspecific encounters with specialized structures containing holotrichous nematocysts [7].

Other candidate toxin sequences

Through convergent evolution, selective processes have shaped the venom repertoire in many distantly related venomous taxa, often converting non-venomous proteins into venomous counterparts [reviewed in 1]. As a result, many distantly related taxa have toxins with similar functional residues, and these can be used to identify toxin gene candidates from transcriptome data [71]. We used BLAST to compare the transcriptomes of acrorhagi from aggressive and non-aggressive polyps to 5,938 annotated toxin protein sequences from the UniProt ToxProt dataset. We found 2,112 (aggressive polyp) and 1,461 (non-aggressive polyp) unique transcripts unique to the ToxProt sequences and would not have been identified by the searches that focused on sea anemone toxins alone. Of these, there were 416 (aggressive polyp) and 196 (non-aggressive polyp) with identified signaling regions (Additional file 3). The majority of the UniProt candidate toxin genes are from snakes and spiders (Figure 10).

Overall, the aggressive and non-aggressive polyp transcriptomes are similar in the genes and gene families recovered by BLAST searches against the UniProt ToxProt dataset (Figure 10A, B; Table 3), although the taxonomic association of the toxin genes shifted when ranked by

level of expression (Figure 10C, D). Acrorhagi of an aggressive polyp have proportionally more expressed transcripts with significant hits from spider toxin genes (Figure 10C), specifically due to the TPM of the toxin Neprilysin-1, a metalloprotease. In contrast, those from the acrorhagi of non-aggressive polyps appear to have more transcripts associated with a snail toxin (Figure 10D), specifically Augerpeptide hhe53. Among the most highly expressed transcripts recovered in our BLAST analysis are the metalloprotease, CRISP, and peptidase proteins, with the majority of the highly expressed toxin-like gene types found in both the aggressive and non-aggressive polyp acrorhagi transcriptome BLAST results (Table 3). Although most transcripts identified here are likely non-venomous proteins of large gene families, the presence of a signaling region and high sequence similarity to other toxins indicates that there may be tremendous uncharacterized toxin diversity found in sea anemones.

Transcriptome characteristics and gene ontology

CEGMA [72] was used to assess the completeness of the acrorhagi from aggressive and non-aggressive polyps transcriptome [73,74]. Of the 248 core eukaryotic proteins, 245 (99%) in the aggressive polyp and 219 (88%) of the non-aggressive polyp acrorhagi transcriptomes were identified and considered complete (>70% alignment length with core proteins). There was an average of ~3.5 (aggressive polyp) and ~2.5 (non-aggressive polyp) orthologs per core protein. Despite having a large difference in the number of raw sequences for each transcriptome, both were deemed relatively complete by CEGMA.

The homology search identified 64,764 (aggressive polyp) and 38,314 (non-aggressive polyp) transcripts (Table 4, Additional file 4). Each of the BLASTx searches identified a greater proportion of transcripts when compared to the BLASTp searches for the transcriptome of acrorhagi from an aggressive (81,171 vs. 67,648) and non-aggressive polyp (55,555 vs. 39,539) (Table 4, Additional file 4). The lower number of BLASTp matches for each transcriptome could be due to the incorrect ORF being translated in the Trinotate pipeline or reflect alternative search strategies between the two methods. The protein domain identification steps matched to the largest portion of the transcriptome, with the program HMMER using the Pfam database [75] identifying 53.37% (aggressive polyp) and 28.89% (non-aggressive polyp) of the transcripts as containing protein domains (Table 4, Additional file 4), with the smallest portions of the transcriptomes having signaling regions or transmembrane helices (Table 4, Additional file 4). Overall, Trinotate characterized only a small portion of the entire transcriptome (Table 4, Additional file 4), likely due to *A. elegantissima* being distantly related to the taxa populating the comparative databases.

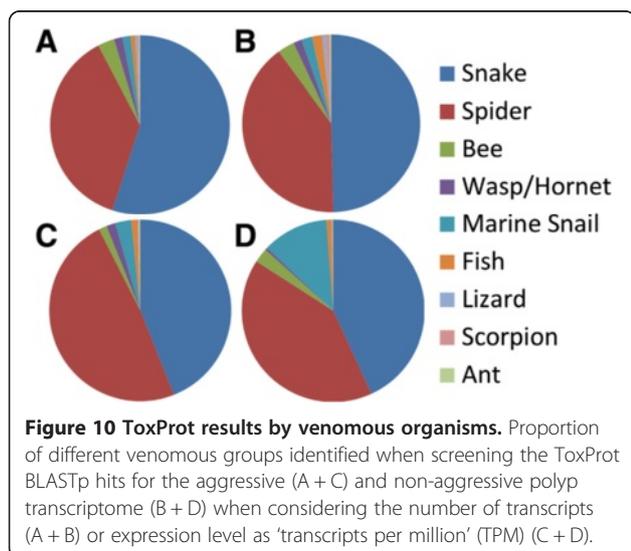


Table 3 Most highly expressed transcripts with significant ToxProt data BLAST hits

Aggressive polyp transcriptome					
transcript_id	TPM	ToxPROT	Group	Gene name (gene family)	E-value
comp71163_c0_seq1	2734	G3LU44	Spider	Translationally-controlled tumor protein homolog (TCPT)	3.00E-08
comp29480_c0_seq1	1521	W4VS99	Spider	Neprilysin-1 (Metalloprotease)*	6.00E-64
comp68333_c3_seq1	1129	Q8AY75	Snake	Cysteine-rich venom protein pseudochetoxin (CRISP)	2.00E-34
comp71203_c0_seq1	903	P24541	Snake	Kunitz-type serine protease inhibitor (Kunitz-type)	3.00E-13
comp61501_c0_seq2	251	P35786	Wasp	Venom allergen 5 (CRISP)	1.00E-06
comp61420_c0_seq1	227	J3S9D9	Snake	Ophiophagus venom factor (NONE)*	4.00E-40
comp67467_c0_seq2	175	P33589	Snake	Thrombin-like enzyme gyroxin analog (Peptidase S1)	2.00E-12
comp62144_c0_seq1	169	P0DM62	Spider	Astacin-like metalloprotease toxin 5 (Metalloprotease)	3.00E-28
comp67061_c0_seq19	164	P0DM61	Spider	Astacin-like metalloprotease toxin 4 (Metalloprotease)	3.00E-12
comp62592_c0_seq1	163	Q4PRD2	Snake	Snaclec coagulation factor X-activating enzyme light chain 2 (Snaclec)*	0.0001
Non aggressive polyp transcriptome					
transcript_id	TPM	ToxPROT	Group	Gene name (gene family)	E-value
comp54814_c0_seq1	3051	G3LU44	Spider	Translationally-controlled tumor protein homolog (TCPT)	2.00E-03
comp71606_c0_seq1	1961	Q8AY75	Snake	Cysteine-rich venom protein pseudochetoxin (CRISP)	8.00E-35
comp43780_c0_seq1	1413	P0CI21	Snail	Augerpeptide hhe53 (NONE)*	0.00003
comp73391_c0_seq1	329	P24541	Snake	Kunitz-type serine protease inhibitor (venom Kunitz-type)	1.00E-13
comp73577_c0_seq1	315	P0DM62	Spider	Astacin-like metalloprotease toxin 5 (Metalloprotease)	2.00E-28
comp43784_c1_seq1	217	Q2XXR2	Lizard	Cysteine-rich venom protein VAR4 (Metalloprotease)*	0.00005
comp43791_c0_seq1	202	P0C8M2	Snake	Disintegrin ocellatin (Metalloprotease)*	0.0001
comp77542_c0_seq1	177	P43685	Wasp	Venom allergen 5 (CRISP)	2.00E-14
comp73785_c0_seq2	170	P33589	Snake	Thrombin-like enzyme gyroxin analog (Peptidase S1)	3.00E-13
comp66731_c0_seq1	136	Q8AY75	Snake	Cysteine-rich venom protein pseudochetoxin (CRISP)	1.00E-14

Transcripts are expressed as 'transcripts per million' (TPM); bold BLAST hits have an identified signaling region.

The gene ontology annotation assigned 69,912 (aggressive polyp) and 51,879 (non-aggressive polyp) transcripts to at least one gene ontology group. There was a large discrepancy in the number of associated GO terms between the two transcriptomes: 469,339 (aggressive) and 359,316 (non-aggressive). This was due to the majority of sequences belonging to more than one gene ontology group (Figure 11). Although the CEGMA analysis revealed that the non-aggressive polyp acrorhagi transcriptome was less complete (99% vs 88%), the difference observed in these transcriptomes may be attributed to acrorhagal tissues

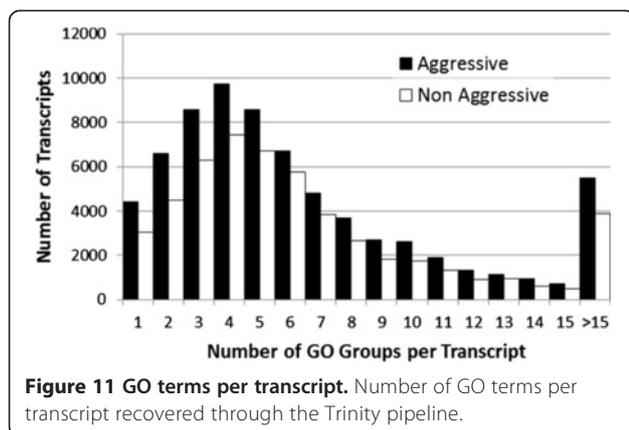
being more transcriptionally active during an aggressive encounter. During the aggressive encounter, the acrorhagi of an aggressive polyp inflate, move, and respond to the target polyp, which likely involves a complex array of cellular signaling and metabolic processes not engaged in a non-aggressive polyp. Regardless, the transcripts of each transcriptome can be attributed to approximately 7900 different GO terms.

Overall, the patterns of association for transcripts were similar in the transcriptomes of acrorhagi from aggressive and non-aggressive polyps: most transcripts are associated

Table 4 Summary of Trinotate results

	Homology search		Protein domain ID				Gene ontology		
	BLASTx	BLASTp	HMMER	Pfam	SignalP	tmHMM	BP	MF	CC
Aggressive	81,171	67,648	134,103	72,554	11,200	28,268	51,143	52,699	55,211
% of Assembly	32.31%	26.92%	53.37%	28.88%	4.46%	11.25%			
Both	64,764		61,549				69,912		
Non-aggressive	55,555	39,539	66,294	39,737	4,257	12,245	42,941	43,177	43,387
% of assembly	24.21%	17.23%	28.89%	17.32%	1.86%	5.34%			
Both	38,314		39,737				51,879		

Values represent number of transcripts and percent assembly during analysis. BP = Biological Process; MF = Molecular Function; CC = Cellular Components.



with cellular components, followed by molecular function and biological process domains (Table 4). The REVIGO clustering algorithm identified semantic similarities among transcripts between groups by compressing the >7900 GO terms into representative lists of fewer than 350 for both the biological process and molecular function domains. All but 12 of the GO terms for the biological process domain grouped with cell adhesion, proteolysis, and protein transport (Figure 12), with the largest REVIGO value differences observed in the “proteolysis” (GO:0006508) GO group for the aggressive polyp transcriptome, and “RNA-dependent DNA replication” (GO:0006278) for the non-aggressive polyp transcriptome (Figure 12). Some GO groups with high REVIGO values were exclusively associated with the aggressive polyp transcriptome (“blood coagulation” GO:0007596 and “positive regulation of apoptosis” GO:0043065) and non-aggressive polyp transcriptome (“induction of apoptosis” GO:0006917 and “platelet activation” GO:0030168). These GO groups may contain toxin-like transcripts similar to synergistic components of snake venom which induce blood coagulation and affect platelet function [76].

Transcripts associated with GO groups in the molecular function domain were grouped into microtubule motor activity, symporter activity, miscellaneous binding, protein serine/threonine kinase activity, and protein homodimerization activity; the remaining 29 GO groups were highlighted as “others” (Figure 13). The largest differences in REVIGO values were observed in the “ATP binding” (GO:0005524) and “RNA-directed DNA polymerase activity” (GO:0003964) for the transcriptome of acrorhagi from aggressive and non-aggressive polyps, respectively (Figure 13). REVIGO groups with the highest values found exclusively in the aggressive polyp transcriptome were associated with the transfer of solutes across a cell membrane (“nucleoside:sodium symporter activity” GO:0005415 and “nucleoside: hydrogen

symporter activity” GO:0015506) as well as ion channel function (“ion channel binding” GO:0044325, “sodium channel activity” GO:0005272, and “acetylcholine-activated cation-selective channel activity” GO:0004889) (Figure 13). These GO groups may be involved in disrupting homeostasis in target cells, thus behaving similarly to cytotoxins or neurotoxins or gene products acting on the ion channels necessary for acrorhagi motility and movement.

Conclusion

Venoms are rarely used exclusively in intraspecific competition, being more commonly employed in defense or predation against other species. We used transcriptomes of the acrorhagi from aggressive and non-aggressive polyps of the sea anemone *A. elegantissima* to investigate the venom proteins and peptides in a tissue specifically used in intraspecific competition. We found a diversity of genes associated with types I – III VGPC/KPI toxins and PLA2s; cytotoxins and VGSC toxins were comparatively less diverse. The high number of candidate toxin genes we found is likely not specific to *A. elegantissima* or to acrorhagi, but reflects our next-generation sequencing approach and relatively sparse prior knowledge of genetic diversity of toxins in sea anemones. We found high sequence divergence among these toxin genes and hypothesize that some toxin alleles with low divergence were incorporated into a single transcript. Our study of cytotoxins, type II VGPC/KPI toxins, VGSC toxins, and type I acrorhagins all produced unexpected results in terms of the inferred pattern of sequence diversity, placement within the gene network, and/or levels of gene expression. Whether or not these toxins play an active role in intraspecific competition remains unknown but merits further investigation. Transcriptome annotation highlighted toxin gene abundance and identified new metalloproteases and CRISP candidate toxin genes based on toxin sequences from other venomous taxa. The functional GO groupings identified transcripts that may be more abundant during acrorhagi inflation and expansion, which occurs during an aggressive encounter. Additionally, semantic similarities of some GO groupings identified transcripts which may behave synergistically with other toxin peptides. Our results provide a baseline for future RNAseq analyses to investigate the role that various peptides may play in aggressive intraspecific encounters.

Methods

Library prep, sequencing, cleanup, assembly, annotation

Polyps of *A. elegantissima* were collected from the intertidal zone in Coos Bay, Oregon in 2012 and kept alive in the lab in flow-through aquaria filled with artificial sea water. The anemones were observed daily to watch for

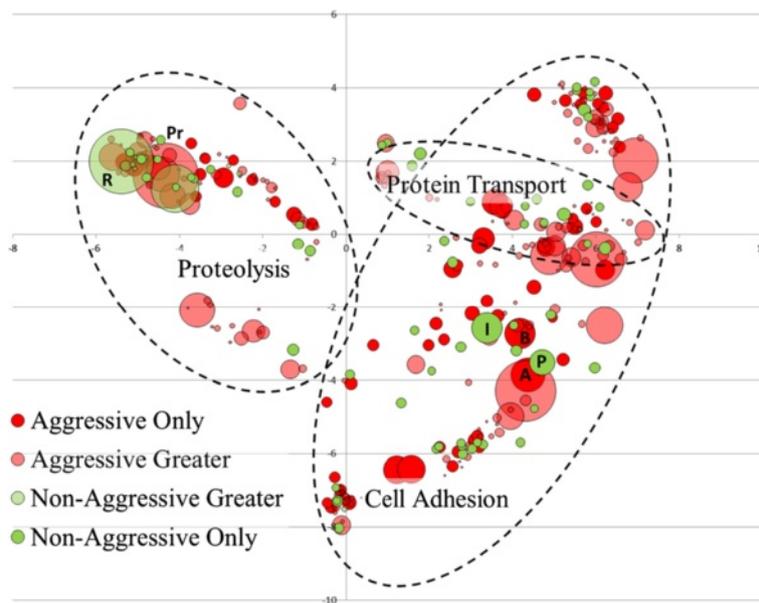


Figure 12 Biological Process domain GO groupings for acrorhagi from aggressive and non-aggressive polyps. Scatterplot clustering of transcripts representing Biological Process domain groupings in REVIGO. The x and y axis represent semantic coordinates in REVIGO; functionally similar gene ontology groups are closer together. Each colored circle represents a GO term, with its size proportionate to its assigned REVIGO value. Specific GO terms referenced in the text are noted: Pr: "proteolysis" (GO:0006508); R: "RNA-dependent DNA replication" (GO:0006278); B: "blood coagulation" (GO:0007596); P: "positive regulation of apoptosis" (GO:0043065); I: "induction of apoptosis" (GO:0006917); A: "platelet activation" (GO:0030168).

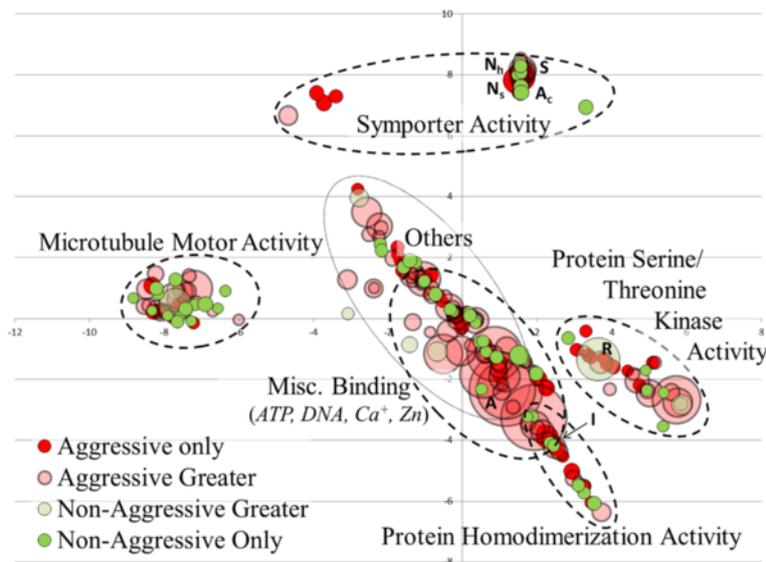


Figure 13 Molecular Function domain GO groupings for acrorhagi from aggressive and non-aggressive polyps. Scatterplot clustering of transcripts representing Molecular Function domain groupings in REVIGO. The x and y axis represent semantic coordinates in REVIGO; functionally similar gene ontology groups are closer together. Each colored circle represents a GO term, with its size proportional to the assigned REVIGO value. Specific GO terms referenced in the text are noted: A:ATP binding (GO:0005524); R: RNA-directed DNA polymerase activity (GO:0003964); N_s: nucleoside: sodium symporter activity (GO:0005415); N_h: nucleoside:hydrogen symporter activity (GO:0015506); I: ion channel binding; (GO:0044325); S: sodium channel activity; GO:0005272; A_c: acetylcholine-activated cation-selective channel activity; GO:0004889.

any indications of aggressive behavior [11]. Once a polyp exhibited aggressive behavior, the polyp initiating the aggressive behavior was removed and placed in a smaller acclimation tank. Following a 15 minute acclimation, five dilated acrorhagi were removed with tweezers from the aggressive polyp and combined into a single 1.5 mL microcentrifuge tube for total RNA extraction. Four polyps not engaging in aggressive encounters were moved to a separate acclimation tank. From each of these polyps, 5–10 non-dilated acrorhagi were removed and combined in a single tube to represent the acrorhagi of non-aggressive polyps. The acrorhagi were flash frozen and 600 μ L of Buffer RLT (QIAGEN) were added to the sample, along with several small (1.5 - 2 mm) ceramic beads (BioExpress). The tissue was macerated in the tube using a Mini-Beadbeater-8 (BioSpec Products) for 30 seconds on the “Homogenize” setting. After 30 seconds, the tube was placed on ice and visually inspected to verify that the tissue was homogenized. Total RNA was extracted following the RNeasy Mini Kit (QIAGEN) protocol.

RNA samples were quantified using the Qubit RNA BR Assay kit (Life Technologies) on a Qubit 2.0 Fluorometer (Life Technologies) and RIN scores were calculated using the Agilent RNA 6000 Nano kit (Agilent Technologies) on the BioAnalyzer (Agilent Technologies). First strand synthesis, library construction, and subsequent paired-end 100 base sequencing was conducted at the Nucleic Acid Shared Resource – Illumina Core, The Ohio State University, Columbus, OH, USA. For each transcriptome, reads were filtered using the program Trimmomatic [77] to remove adapters, leading and trailing low quality bases (using a sliding window greater than 3 bases), reads shorter than 36 bases in length, and reads that fell below an average quality score of 15 using a four base sliding window. The cleaned data were checked using the program FastQC [78] to ensure that low quality reads and regions were removed.

Raw data were assembled *de novo* Trinity [79] using default parameters. Transcripts were analyzed to determine the expected read count and proportion of each transcript relative to the gene (IsoPct). Expression levels for aggressive and non-aggressive polyp transcripts were determined in RSEM [80]. To adjust for differences in library size and skewed expression of transcripts, we used the metric ‘transcripts per million’ (TPM) for each transcriptome [81]. TPM is preferred over other expression metrics as it is independent of mean expressed transcript length, thus allowing for cross comparison between multiple samples [74,80,82]. Our TPM values do not permit a statistical analysis of differential gene expression, but instead provide metrics that can be compared between aggressive and non-aggressive polyp transcriptomes, while controlling for transcript size and discrepancies between sample preparation [74,80].

Identification of candidate toxin genes

Bioinformatic processing of the transcriptomes used a combination of custom PERL scripts and sequence annotation programs. To identify candidate toxin genes, taxonomically diverse sea anemone toxin sequences were downloaded from GenBank (Additional file 2: Table S2). Candidate toxin genes from both transcriptomes were identified via the tBLASTn search algorithm (E value cutoff of 10). Matches from the BLAST searches were visually inspected for the key cysteine amino acid residues that are characteristic of the different venom types [50] and then screened for premature stop codons. We conducted a BLASTx search against the NCBI-nr protein database and BLASTn searches against the NCBI-nr nucleotide database and EST database to confirm that these sequences would retrieve toxin sequences. Protein sequences of candidate toxin genes were visualized in BioEdit [83] and aligned to known sea anemone toxin genes using ClustalW [84]. Protein sequences were screened for the placement of key cysteine amino acid residues [50]. Sequences that did not introduce large gaps (greater than variation observed in previously described proteins) into the alignments were retained for further processing (Additional file 5). The signaling region for each transcript (if present) was determined using SignalP [85]. Sequences with sequence identity of 90% or greater were considered isoforms of the same gene; however, sequences within gene networks with greater than 50% bootstrap support were also explored as potential divergent alleles.

To broaden our search for additional toxin genes beyond those previously identified in sea anemones, we used BLAST to search the transcriptomes of acrorhagi from aggressive and non-aggressive polyps against the UniProt ToxProt dataset [86], with sequences identified in the sea anemone toxin query removed. The candidate toxin-like transcripts were further processed by removing any redundant protein sequences and screened based on the following criteria: E value $<1e-04$ and identity percentage $>20\%$. In addition to characterizing the overall taxonomic diversity associated with candidate toxin genes, we also looked at relative abundance of each of these genes and the presence of a signaling region which is characteristic of many toxin coding genes [67,71].

Alignment and gene network reconstruction

Phylogenetic reconstruction provides a robust tool for identifying toxin genes [87]. We reconstructed unrooted gene networks for toxins identified as PLA2s, cytolysins, VGPC, and VGSC. Nucleotide sequences of candidate toxin genes and previously described sea anemone toxins were translated into protein sequences and aligned using default parameters in MAFFT (L-INS-i) [88]. Nucleotide alignments were created in BioEdit [83] using the protein alignment as a guide. Due to significant length

variation and absence of the signaling/propeptide region in other taxa, gene networks were reconstructed with and without this region for the protein and nucleotide alignments (Additional file 5). As the majority of sea anemone toxins were acquired via mass spectrometry and thus lack the signaling/propeptide region, mature protein sequences of previously characterized toxin genes were incorporated into the final gene network reconstructions (Additional file 2: Table S1). MEGA [89] was used to select appropriate evolutionary models and reconstruct gene networks for both protein and nucleotide sequences (Additional file 2: Table S1). Protein and nucleotide datasets were subjected to 1000 rounds of bootstrap resampling in a maximum likelihood framework to determine branch support values for all toxin gene families.

Transcriptome annotation and gene ontology

To assess the completeness of the transcriptome of acrorhagi from aggressive and non-aggressive polyps we used CEGMA (Core Eukaryotic Genes Mapping Approach) to identify the presence of 248 highly conserved proteins found across eukaryotes [72]. Annotation and homology searches of assembled transcripts from both transcriptomes were conducted using a suite of sequence annotation tools in the Trinotate pipeline [90]. Contrary to the hierarchical GO assignment approach of the popular program Blast2GO [91], Trinotate assigns transcripts among varying hierarchical levels within gene ontology (GO) networks. Free from hierarchical groupings, Trinotate permits cross comparisons across potentially functionally important genes that are found at different hierarchical levels within the gene ontology. Assembled transcripts were translated into their longest open reading frame peptide in TransDecoder [92]. Homology searches were conducted against SwissProt [93] using BLASTx searches of raw transcripts and BLASTp for the translated protein sequences. Conserved protein domains were identified using the program HMMER [75] against the Pfam domain database [94]. SignalP [85] was used to predict the signaling region of each transcript and tmHMM [95] was used to identify transmembrane regions. Functional annotation was performed by comparing BLAST hits with the annotated GO Pathways databases [96]. REVIGO [97] was used to visualize the GO group frequency among the transcripts with 0.9 similarity and all other parameters default.

Availability of supporting data

Candidate toxin sequences used in our analyses are included as Additional file 1 and unmodified transcriptome assemblies on our data website (<http://u.osu.edu/anemonedata/data/>). All of the raw sequencing reads used to construct each transcriptome is available under BioProject accession

PRJNA266623. The raw Illumina sequence reads have been deposited on NCBI's SRA archive (Aggressive: SRR1645256, Non-Aggressive Polyp: SRR1646677) and the assembled transcriptomes on NCBI's TSA database (Aggressive: GBXJ00000000, Non-Aggressive Polyp: GBYC00000000) that have been subjected to NCBI's contamination screen.

Additional files

Additional file 1: Nucleotide sequences of candidate toxin sequences identified in this study, includes both signaling and mature toxin when applicable.

Additional file 2: Table S1. Parameters for toxin gene network reconstruction for various toxin candidates. **Table S2.** Genbank accession numbers for known anemone toxin sequences used in BLAST searches and subsequent analyses. **Figure S1.** Protein sequence alignment of candidate type I VGPC toxins alongside previously characterized toxin proteins. **Figure S2.** Sequence alignments of Acrorhagin I and Acrorhagin II along with candidate toxin sequences identified from each transcriptome.

Additional file 3: ToxProt BLAST hits and subsequent analysis for the aggressive polyp acrorhagi and the non-aggressive polyp acrorhagi transcriptomes.

Additional file 4: Trinotate output for the aggressive polyp acrorhagi and the non-aggressive polyp acrorhagi transcriptomes.

Additional file 5: Protein sequence alignments for all the toxin genes used in the toxin gene network reconstructions.

Abbreviations

RNA-seq: RNA sequencing; CRISP: Cysteine rich secretory protein; PLA2s: Phospholipase A2s; VGPC: Voltage gated potassium channel; VGSC: Voltage gated sodium channel; APETx2: Acid sensing channel toxin; BLAST: Basic local alignment search tool; GO: Gene ontology; CEGMA: Core eukaryotic genes mapping approach; ORF: Open reading frame.

Competing interest

The authors declare that they have no competing interests.

Authors' contributions

JM conceived experiments, generated data, analyzed data, and drafted the manuscript. MB conceived experiments. MD conceived experiments and drafted the manuscript. All authors edited and approved the final manuscript.

Acknowledgments

We thank Jenna Valley for collecting polyps of *A. elegantissima* for this study and Joe Cora for assisting with installing components of this bioinformatic pipeline on the Ohio Biodiversity Conservation Partnership (OBCP) cluster. We thank Dr. Anthony D'Orazio for the use of Figure 1 and Figure 2 insert. We thank the editor and the anonymous reviewers for their feedback, which helped us improve this manuscript. This work was supported by US National Science Foundation DEB-1257796 to M. Daly.

Author details

¹The Ohio State University, Evolution, Ecology, and Organismal Biology, 318 W. 12th Avenue, Columbus, OH 43210-1293, USA. ²Sackler Institute for Comparative Genomics, Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA. ³Biological Sciences Department, NYC College of Technology (CUNY), 300 Jay Street, Brooklyn, NY 11201, USA.

Received: 31 October 2014 Accepted: 28 February 2015

Published online: 21 March 2015

References

1. Fry BG, Roelants K, Champagne DE, Tyndall JDA, King GF, Nevalainen TJ, et al. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu Rev Genomics Hum Genet.* 2009;10:483–511.

2. Nekaris KA-I, Moore RS, Rode EJ, Fry BG. Mad, bad and dangerous to know: the biochemistry, ecology and evolution of slow loris venom. *J Venom Anim Toxins incl Trop Dis*. 2013;19:21.
3. 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;18:986–94.
4. Purcell JE. Aggressive function and induced development of catch tentacles in the sea anemone *Metridium senile* (Coelenterata Actiniaria). *Biol Bull*. 1977;153:355–68.
5. Bigger CH. Interspecific and intraspecific acrorhagial aggressive behavior among sea anemones: a recognition of self and not-self. *Biol Bull*. 1980;159:117–34.
6. Watson GM, Mariscal RN. The development of a sea anemone tentacle specialized for aggression: morphogenesis and regression of the catch tentacle of *Haliplanella luciae* (Cnidaria Anthozoa). *Biol Bull*. 1983;164:506–17.
7. Williams RB. Acrorhagi catch tentacles and sweeper tentacles: A synopsis of 'aggression' of actiniaria and scleractinian. *Cnidaria Hydrobiologia*. 1991;216/217:539–45.
8. Daly M. The anatomy terminology and homology of acrorhagi and pseudoacrorhagi in sea anemones. *Zool Verh Leiden*. 2003;345:89–101.
9. Bigger CH. The cellular basis of the aggressive acrorhagial response of sea anemones. *J Morphol*. 1982;173:259–78.
10. Williams RB. Some recent observations on the acrorhagi of sea anemones. *J Mar Biol Ass UK*. 1978;58:787–8.
11. Francis L. Intraspecific aggression and its effect on the same distribution of *Anthopleura elegantissima* and some related sea anemones. *Bio Bull*. 1973;144:73–92.
12. Francis L. Cloning and aggression among sea anemones (Coelenterata: Actiniaria) of the rocky shore. *Biol Bull*. 1988;174:241–53.
13. Ayre DJ, Grosberg RK. Behind anemone lines: factors affecting division of labour in the social cnidarian *Anthopleura elegantissima*. *Anim Behav*. 2005;70:97–110.
14. Honma T, Minagawa S, Nagai H, Ishida M, Nagashima Y, Shiomi K. Novel peptide toxins from acrorhagi aggressive organs of the sea anemone *Actinia equina*. *Toxicon*. 2005;46:768–74.
15. Bartosz G, Finkelshtein A, Przygodzki T, Bsoor T, Neshar N, Sher D, et al. A pharmacological solution for a conspecific conflict: ROS-mediated territorial aggression in sea anemones. *Toxicon*. 2008;51:1038–50.
16. Diochot S, Lazdunski M. Sea anemone toxins affecting potassium channels. *Prog Mol Subcell Biol*. 2009;46:99–122.
17. Razpotnik A, Krizaj I, Sribar J, Kordis D, Macek P, Frangez R, et al. A new phospholipase A2 isolated from the sea anemone *Urticina crassicornis*-its primary structure and phylogenetic classification. *FEBS J*. 2010;277:2641–53.
18. Anderluh G, Sepcic K, Turk T, Macek P. Cytolytic proteins from Cnidarians-an overview. *Acta Chim Slov*. 2011;58:724–9.
19. Oliveira JS, Fuentes-Silva D, King GF. Development of a rational nomenclature for naming peptide and protein toxins from sea anemones. *Toxicon*. 2012;60:539–50.
20. Wittcoff H. *The Phosphatides*. New York, NY: Reinhold Publishing Corp; 1951. p. 99–115.
21. Murakami M, Kudo I. Phospholipase A₂. *J Biochem*. 2002;131:285–92.
22. Marcussi S, Sant'Ana CD, Oliveria CZ, Rueda AQ, Menaldo DL, Belebony R, et al. Snake venom phospholipase A₂ inhibitors: Medicinal chemistry and therapeutic potential. *Curr Top Med Chem*. 2007;7:743–56.
23. Dennis EA, Cao J, Hsu Y-H, Magrioti V, Kokotos G. Phospholipase A₂ enzymes: physical structure, biological function, disease implication, chemical inhibition and therapeutic intervention. *Chem Rev*. 2011;111:6130–85.
24. Romero L, Marcussi S, Marchi-Salvador DP, Silva Jr FP, Fuly AL, Stabeli RG, et al. Enzymatic and structural characterization of a basic phospholipase A₂ from the sea anemone *Condylactis gigantea*. *Biochimie*. 2010;92:1063–71.
25. Nevalainen TJ, Peuravuori HJ, Quinn RJ, Llewellyn LE, Benzie JAH, Fenner PJ, et al. Phospholipase A2 in Cnidaria. *Comp Biochem Phys*. 2004;139:731–5.
26. Hessinger DA, Lenhoff HM. Mechanism of hemolysis induced by nematocyst venom: roles of phospholipase A and direct lytic factor. *Arch Biochem Biophys*. 1976;173:603–13.
27. Suput D. *In vivo* effects of cnidarian toxins and venoms. *Toxicon*. 2009;54:1190–200.
28. Ojcius DM, Young JD-E. Cytolytic pore forming proteins and peptides: Is there a common structural motif? *TIBS*. 1991;16:225–9.
29. Anderluh G, Macek P. Cytolytic peptide and protein toxins from sea anemones (Anthozoa: Actiniaria). *Toxicon*. 2002;40:111–24.
30. Sher D, Zlotkin E. A hydra with many heads: Protein and polypeptide toxins from hydra and their biological roles. *Toxicon*. 2009;54:1148–61.
31. Huerta V, Morera V, Guancho Y, Chinae G, Gonzalez LJ, Betancourt L, et al. Primary structure of two cytolysin isoforms from *Stichodactyla helianthus* differing in their hemolytic activity. *Toxin*. 2001;39:1253–6.
32. Garcia-Linares S, Richmond R, Garcia-Mayoral MF, Bustamante N, Gavilanes JG, Martinez-del Pozo A. The sea anemone actinoporin (Arg-Gly-Asp) conserved motif is involved in maintaining the competent oligomerization state of these pore-forming toxins. *FEBS J*. 2014;281:1465–78.
33. Meinardi E, Florin-Christensen M, Paratcha G, Azcurra JM, Florin-Christensen J. The molecular basis of the self-nonself selectivity of a coelenterate toxin. *Biochem Biophys Res Commun*. 1995;216:348–54.
34. Bonev BB, Lam Y-H, Anderluh G, Watts A, Norton RS, Separovic F. Effects of the eukaryotic pore-forming cytolysin equinatoxin II on lipid membranes and the role of sphingomyelin. *Biophys J*. 2003;84:2382–92.
35. Norton RS. Structure and structure-function relationships of sea anemone proteins that interact with the sodium channel. *Toxicon*. 1991;29:1051–84.
36. Honma T, Shiomi K. Peptide toxins in sea anemones: structural and functional aspects. *Mar Biotechnol*. 2006;8:1–10.
37. de la Vega RCR, Merino E, Beceril B, Possani LD. Novel interactions between K⁺ channels and scorpion toxins. *Trends in Pharmacol Sci*. 2003;24:222–7.
38. Catterall WA, Ceste'le S, Yarov-Yarovsky V, Yu FH, Konoki K, Scheuer T. Voltage-gated ion channels and gating modifier toxins. *Toxicon*. 2007;49:124–41.
39. Diochot S, Baron A, Rash LD, Deval E, Escoubas P, Scarzello S, et al. A new sea anemone peptide, APETx2, inhibits ASIC3, a major acid-sensitive channel in sensory neurons. *EMBO J*. 2004;23:1516–25.
40. Rendo'n-Anaya M, Delaye L, Possani LD, Herrera-Estrella A. Global transcriptome analysis of the scorpion *Centruroides noxius*: new toxin families and evolutionary insights from an ancestral scorpion species. *PLoS One*. 2012;7:e43331.
41. Gutman GA, Chandy KG, Grissmer S, Lazdunski M, Mckinnon D, Pardo LA, et al. International union of pharmacology. LIII. Nomenclature and molecular relationships of voltage-gated potassium channels. *Pharmacol Rev*. 2005;57:473–508.
42. Catterall WA. From ionic currents to molecular mechanisms: the structure and function of voltage-gated sodium channels. *Neuron*. 2000;26:13–25.
43. Goldin AL. Mechanisms of sodium channel inactivation. *Curr Opin in Neur*. 2003;13:284–90.
44. Talvnen KA, Nevalainen TJ. Cloning of novel phospholipase A2 from the cnidarian *Adamsia carcinopados*. *Comp Biochem Phys B*. 2002;132:571–8.
45. Frazao B, Vasconcelos V, Antunes A. Sea anemone (Cnidaria Anthozoa Actiniaria) toxins: An overview. *Mar Drugs*. 2012;10:1812–51.
46. Hu H, Bandyopadhyay PK, Olivera BM, Yandell M. Elucidation of the molecular envenomation strategy of the cone snail *Conus geographus* through transcriptome sequencing of its venom duct. *BMC Genomics*. 2012;12:284.
47. Haney RA, Ayoub NA, Clarke TH, Hayashi CY, Garb JE. Dramatic expansion of the black widow toxin arsenal uncovered by multi-tissue transcriptomics and venom proteomics. *BMC Genomics*. 2014;15:366.
48. Moran Y, Genikhovich G, Gordon D, Wienkoop S, Zenkert C, Ozbek S, et al. Neurotoxin localization to ectodermal gland cells uncovers an alternative mechanism of venom delivery in sea anemones. *Proc R Soc B*. 2011;279:1351–8.
49. Wang Y, Yap LL, Chua KL, Khoo HE. A multigene family of Heteractis magnificalytins (HMgs). *Toxicon*. 2008;51:1374–82.
50. Kozlov S, Grishin E. Convenient nomenclature of cysteine-rich polypeptide toxins from sea anemones. *Peptides*. 2012;33:240–4.
51. Isaeva MP, Chausova VE, Zelepuga EA, Guzev KV, Tabakmakher VM, Monastyrnaya MM, et al. A new multigene superfamily of Kunitz-type protease inhibitors from sea anemone *Heteractis crispata*. *Peptides*. 2012;34:88–97.
52. Lighten J, Van Oosterhout C, Bentzen P. Critical review of NGS analyses for de novo genotyping multigene families. *Mol Ecol*. 2014;23:3957–72.
53. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;251:403–10.
54. Hedtke SM, Townsend TM, Hillis DM. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Syst Biol*. 2006;55:522–9.
55. Rodriguez E, Barbeitos MS, Brugler MR, Crowley LM, Grajalas A, Gusmao L, et al. Hidden among sea anemones: the first comprehensive phylogenetic reconstruction of the order Actiniaria (Cnidaria, Anthozoa, Hexacorallia) reveals a novel group of hexacorals. *PLoS One*. 2014;9:e96998.

56. Anderlugh G, Pungercar J, Strukelj B, Macek P, Gubensek F. Cloning, sequencing, and expression of Equinatoxin II. *Biochem Biophys Res Commun.* 1996;220:437–42.
57. Pungercar J, Anderlugh G, Macek P, Franc G, Strukelj B. Sequence analysis of the cDNA encoding the precursor of equinatoxin V, a newly discovered hemolysin from the sea anemone *Actinia equina*. *Biochim Biophys Acta.* 1997;1341:105–7.
58. Orts DJB, Moran Y, Cologna CT, Peigneur S, Madio B, Praher D, et al. BcsTx3 is a founder of a novel sea anemone toxin family of potassium channel blocker. *FEBS J.* 2013;280:4839–52.
59. 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;45:33–41.
60. Mourao CBF, Schwartz EF. Protease inhibitors from the marine venomous animals and their counterparts in terrestrial venomous animals. *Mar Drugs.* 2013;11:2069–112.
61. Schweitz H, Bruhn T, Guillemare E, Moinier D, Lancelin J-M, Beress L, et al. Kaliclidines and kaliseptine. *J Biol Chem.* 1995;270:25121–26.
62. Calvete JJ, Sanz L, Angulo Y, Lomonte B, Gutierrez JM. Venoms, venomics, antivenomics. *FEBS Lett.* 2009;583:1736–43.
63. Yeung SYM, Thompson D, Wang Z, Fedida D, Robertson B. Modulation of Kv3 subfamily potassium currents by the sea anemone toxin BDS: significance for CNS and biophysical studies. *J Neurosci.* 2005;25:8735–45.
64. Castaneda O, Harvey AL. Discovery and characterization of cnidarian peptide toxins that affect neuronal potassium ion channels. *Toxicon.* 2009;54:1119–24.
65. Honma T, Kawahata S, Ishida M, Nagai H, Nagashima Y, Shiomi K. Novel peptide toxins from the sea anemone *Stichodactyla haddoni*. *Peptides.* 2008;29:536–44.
66. Zaharenko AJ, Ferreira WA, Oliveira JS, Richardson M, Pimenta DC, Konno K, et al. Proteomics of the neurotoxic fraction from the sea anemone *Bunodosoma cangicum* venom: novel peptides belonging to new classes of toxins. *Comp Biochem Phys D.* 2008;3:219–22.
67. Moran Y, Gordon D, Gurevitz M. Sea anemone toxins affecting voltage-gated sodium channels- molecular and evolutionary features. *Toxicon.* 2009;54:1089–101.
68. Norton RS, Ohizumi Y, Shibata S. Excitatory effect of a new polypeptide (Anthopluerin-B) from sea anemone on the guinea-pig vas deferens. *Br J Pharmacol.* 1981;74:23–8.
69. Bruhn T, Schallerb C, Schulzeb C, Sanchez-Rodriguez J, Dannmeiera C, Ravensd U, et al. Isolation and characterisation of five neurotoxic and cardiotoxic polypeptides from the sea anemone *Anthopleura elegantissima*. *Toxicon.* 2001;39:693–702.
70. Minagawa S, Sugiyama M, Ishida M, Nagashima Y, Shiomi K. Kunitz-type protease inhibitors from acrorhagi of three species of sea anemones. *Comp Biochem Phys D.* 2008;150:240–5.
71. von Reumont BM, Blanke A, Richter S, Alvarez F, Bleidorn C, Jenner RA. The first venomous crustacean revealed by transcriptomics and functional morphology: remipedes venom glands express a unique toxin cocktail dominated by enzymes and a neurotoxin. *Mol Biol Evol.* 2014;31:48–58.
72. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics.* 2007;23:1061–67.
73. Parra G, Bradnam K, Ning Z, Keane T, Korf I. Assessing the gene space in draft genomes. *Nucleic Acids Res.* 2009;37:289–97.
74. Nakasugi K, Crowhurst RN, Bally J, Wood CC, Hellens RP, Waterhouse PM. De novo transcriptome sequence assembly and analysis of RNA silencing genes of *Nicotiana benthamiana*. *PLoS One.* 2013;8:e59534.
75. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucl Acids Res.* 2011;39:W29–37.
76. Ouyang C, Teng C-M, Huang T-F. Characterization of snake venom components acting on blood coagulation and platelet function. *Toxicon.* 1992;30:945–66.
77. Bolger A and Giorgi F: Trimmomatic: A Flexible Read Trimming Tool for Illumina NGS Data. Available online at: <http://www.usadellab.org/cms/?page=trimmomatic>
78. Andres S: FastQC. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
79. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnol.* 2011;29:644–52.
80. Li B, Dewey CN RSEM. accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
81. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 2012;131:281–5.
82. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 2010;11:R25.
83. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 1999;41:95–8.
84. Thompson JD, Higgins DG, Gibson TJ, Clustal W. Improving sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl Acids Res.* 1994;22:4637–80.
85. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 2011;8:785–6.
86. Animal toxin annotation program [<http://www.uniprot.org/program/Toxins>]
87. Fry BG, Wuster W. Assembling an arsenal: origin and evolution of snake venom proteome inferred from phylogenetic analysis of toxin sequences. *Mol Biol Evol.* 2004;21:870–83.
88. Katoh S. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–8.
89. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 2013;30:2725–9.
90. Trinotate: Transcriptome Functional Annotation and Analysis [<http://trinotate.github.io/>]
91. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005;21:3674–6.
92. TransDecoder (Find Coding Regions Within Transcripts) [<http://transdecoder.github.io/>]
93. Boeckmann B, Bairoch A, Apweiler R, Blatter M-C, Estreicher A, Gasteiger E, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL. *Nucl Acid Res.* 2003;31:365–70.
94. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. The Pfam protein families database. *Nucleic Acids Res.* 2014;42:D222–30.
95. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001;305:567–80.
96. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 2014;42:D199–205.
97. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One.* 2011;6:e21800.

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

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

Submit your manuscript at
www.biomedcentral.com/submit

