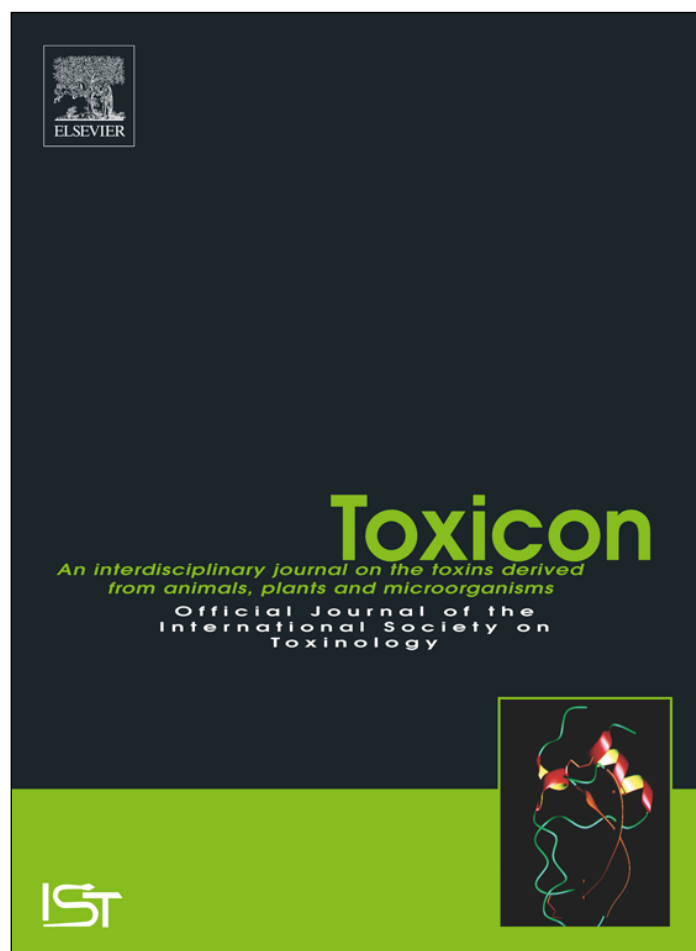


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the author's institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>



Contents lists available at ScienceDirect

Toxicon

journal homepage: www.elsevier.com/locate/toxicon

Stichodactyla helianthus' de novo transcriptome assembly: Discovery of a new actinoporin isoform



Esperanza Rivera-de-Torre^{a,b,*}, Álvaro Martínez-del-Pozo^b, Jessica E. Garb^a

^a Department of Biological Sciences, University of Massachusetts Lowell, Lowell, MA, USA

^b Departamento de Bioquímica y Biología Molecular, Facultad de CC. Químicas, Universidad Complutense de Madrid, 28040, Madrid, Spain

ARTICLE INFO

Keywords:

Transcriptome
Anemone
Actinoporin
Toxin
Isoform

ABSTRACT

Transcriptomic profiling of venom producing tissues from different animals is an effective approach for discovering new toxins useful in biotechnological and pharmaceutical applications, as well in evolutionary comparative studies of venomous animals. *Stichodactyla helianthus* is a Caribbean sea anemone which produces actinoporins as part of its toxic venom. This family of pore forming toxins is multigenic and at least two different isoforms, encoded by separate genes, are produced by *S. helianthus*. These isoforms, sticholysins I and II, share 93% amino acid identity but differ in their pore forming activity and act synergistically. This observation suggests that other actinoporin isoforms, if present in the venomous mixture, could offer an advantageous strategy to modulate whole venom activity. Using high-throughput sequencing we generated a *de novo* transcriptome of *S. helianthus* and determined the relative expression of assembled transcripts using RNA-Seq to better characterize components of this species' venom, focusing on actinoporin diversity. Applying this approach, we have discovered at least one new actinoporin variant from *S. helianthus* in addition to several other putative venom components.

1. Introduction

Sea anemones belong to one of the oldest venomous lineages, the phylum *Cnidaria*. Since these animals usually live fixed to a substrate, their predatory and defensive abilities rely on the production of potent toxins. The predominant toxins in sea anemone venoms studied so far include phospholipases A2s (PLA2s), cytolytins and neurotoxins, like potassium and sodium gated channels inhibitors (KTxs and NaTxS) (Madio et al., 2017; Macrander et al., 2015a, 2016). Other minor components like metalloproteases and cysteine rich peptides are also found in anemone venoms (Macrander et al., 2015a). This toxic cocktail is usually stored in a specialized organelle called the nematocyst, a stinging capsule that injects the venom upon activation by contact. Anemone venom components work together in remodeling target membranes (e.g., PLA2s), producing prey cell death by osmotic shock (cytolytins) and provoking pain and paralysis by altering neurotransmission secretion (neurotoxins) (Suput, 2009). In combination, these toxins cause paralysis and death in small fishes and crustaceans (Salgado and Kem, 1992). Past work showed that nematocysts were only present in the tentacles, which surround the anemone oral disc, and this was expected given the role of tentacles in attracting and catching prey (Basulto et al., 2006). Recent studies question this

observation, suggesting that the toxic secretion can be detected across most of the sea anemone tissues (Rojko et al., 2016).

Among the sea anemone cytolytins (cell-lysing toxins), actinoporins are pore-forming proteins. They constitute multigenic families (Wang et al., 2008), where a single individual expresses different genes coding for several isotoxins with distinct activities (Wang et al., 2008; Monastyrnaya et al., 2010). *Stichodactyla helianthus* is a Caribbean sea anemone which produces large amounts of two easily detectable actinoporin variants: sticholysin I (StnI) and sticholysin II (StnII) (Lanio et al., 2001). StnII is the most potent actinoporin of those that have been studied with molecular detail (García-Linares et al., 2016a). StnI is approximately four times less active than StnII in terms of hemolytic activity (Monastyrnaya et al., 2010; García-Linares et al., 2016a). Both isoforms interact with each other to produce a synergistic effect; just trace amount of StnII enhances StnI binding to lipid model vesicles driving a dramatic improvement of hemolytic activity (Rivera-de-Torre et al., 2016). These observations extend the complexity and regulatory capacity of toxin function, similar to a rudimentary immune system (Wang et al., 2008), potentially increasing the range of species *S. helianthus* can capture or defend itself from.

As mentioned above, only StnI and StnII have been detected in *S. helianthus* lysates thus far (Lanio et al., 2001; Valle et al., 2015). Both

* Corresponding author. Departamento de Bioquímica y Biología Molecular, Facultad de CC. Químicas, Universidad Complutense de Madrid, 28040, Madrid, Spain.
E-mail address: esperanza.rivera.detorre@gmail.com (E. Rivera-de-Torre).

are produced in high concentration, although it remains unclear which one is more abundant (Lanio et al., 2001). However, considering their synergistic interaction (Rivera-de-Torre et al., 2016) and the multigene character of this protein family (Wang et al., 2008), it has been hypothesized (Rivera-de-Torre et al., 2016) that small amounts of other actinoporins, not yet detected in the lysate, might be produced to further modulate whole venom activity.

Next-generation RNA-sequencing is a powerful technology which is the emerging method of choice in the analysis of venoms because of the massive amounts of sequence data that can be rapidly obtained. Specifically, a single Illumina instrument run can capture enough data to fully characterize a venom transcriptome, including all its constituent transcripts and their encoded proteins, and enabling estimation of the relative abundance of assembled transcripts through read mapping (Macrander et al., 2015a, 2015b, 2016). Further analysis can be helpful in reconstructing the evolutionary relationships of toxins within and across different species, which can also illuminate how sea anemone species have diversified their venom composition (Haney et al., 2016; Garb et al., 2004; Garb and Hayashi, 2005; Kitchen et al., 2015; Macrander and Daly, 2016).

In this study we utilized Illumina RNA-seq methods to sequence and assemble a first draft of the *Stichodactyla helianthus* transcriptome. We used this new database to identify and quantify the relative expression of candidate venom toxins, focusing on characterizing novel putatively active actinoporins, in addition to other toxic proteins.

2. Materials and methods

2.1. Specimens

Two specimens of *Stichodactyla helianthus* were purchased from Dynasty Marine Associates, Inc (Marathon, Florida), which were collected from Big Pine Key, Florida (Biosample PRJNA464282). The specimens were cut into thin radial pieces comprising all the structures in the sea anemone, principally the tentacles and column. This tissue was dissolved in TRIzol (Ambion, Life Technologies) up to a final concentration of 100 µg/mL by homogenization with a rotar-stator homogenizer when it was still fresh, and they were frozen at -80°C for less than a week before RNA extraction.

2.2. RNA extraction

Total RNA was extracted through the standard protocol recommended for TRIzol (Chomczynski, 1993). Briefly, homogenized tissue was incubated at room temperature for 5 min and then chloroform was added to the samples. After vigorous shaking, the mixtures were centrifuged at $12000 \times g$ for 15 min at 4°C . The aqueous phase was recovered in a new tube. RNA was precipitated with isopropanol recovered by centrifugation followed by several washing steps with 75% ethanol. Dried RNA samples were dissolved in RNase-free water. Samples were DNase treated with the TURBO DNA-free kit (Ambion, Life technologies). Finally, RNA was stored at -80°C until utilization. The RNA concentration was quantified with a Qubit 3.0 (Thermo Scientific) and its integrity was confirmed through 2100 BioAnalyzer (Agilent Technologies) analysis.

2.3. Library preparation and high-throughput sequencing

Using the TruSeq stranded mRNA LT Sample prep kit (Illumina), double-stranded cDNA libraries were obtained from the two total RNA samples. To test the quality of the libraries a BioAnalyzer chip (Agilent Technologies) was run. The relative concentration of each bar-coded library was determined through qPCR and they were immediately pooled in a single sample considering that equivalent amounts of each library were present in the mixture. Libraries were sequenced on a single NextSeq[®] 500/550 High Output Kit v2 (300 cycles) (Illumina), on

a NextSeq 500 instrument as a paired-end 150 base sequencing.

2.4. Cleanup and de novo assembly

Raw Illumina reads from different libraries were de-multiplexed according to their barcoding label. Illumina adaptors were trimmed using the Trim Galore! tool (Krueger, 2012). As a modification to the default setting, the unpaired reads were retained and 5 extra base pairs were cleaved at the 3' edge of the reverse reads to remove low-quality bases. The quality of these cleaned raw reads was verified with the FastQC reports in the Trim Galore! suite. Cleaned reads were deposited in NCBI's SRA database (Accession: SRR7126073) within Bioproject PRJNA464282.

S. helianthus RNA-seq cleaned data from the two libraries were put together in a *de novo* transcriptome assembly using Trinity v2.4.0 or Trinity v2.0.6 (Grabherr et al., 2011) with default parameters. Two different versions of Trinity (v2.4.0 and v2.0.6) were compared to analyze potential differences in transcriptome fragmentation (Bryant et al., 2017). Transcriptome completeness was determined using BUSCO v3 (Benchmarking Universal Single-Copy Orthologs (Simao et al., 2015)) with the Metazoa data set. This program assayed the presence of 978 proteins conserved from 65 different species within this group. Transcripts were deposited in the TSA database (GGNY00000000).

2.5. Data analysis, candidate toxin gene identification and transcript quantification

The resulting transcriptome was annotated by identification of alignments through BLASTx with proteins in the National Center for Biotechnology Information (NCBI) with each transcript. The restrictions applied were an E-value $< 1 \times 10^{-5}$ and allowing for a maximum of ten target sequences per query sequence.

Starting with the Trinity assembly file, we identified the open-reading frame (ORF) for each transcript with Transdecoder (Haas et al., 2013). The default settings were modified to retain the longest ORF with the best BLASTx hit showing an E-value < 10 or in absence of a BLASTx alignment, the longest ORF. In this way, we obtained a comprehensive protein prediction database if transcripts lacked a BLAST hit.

Relative expression levels for each transcript were calculated using the program RSEM (Li and Dewey, 2011). This program mapped back the cleaned sequence reads to the assembly, estimating the number of transcripts as TPM (Transcripts Per Million).

A nucleotide database for BLAST searches was built with both Trinity assemblies from *S. helianthus* (obtained with Trinity v2.4.0 and v2.0.6 respectively). From the ToxProt data set (<http://www.uniprot.org/program/Toxins> last accession 2018/02/14), venom proteins and toxins sequences from different categories like KTX-I, KTX-III, NaTx, Metalloproteases, PLA2s and Cytolysins were pooled and extracted. These protein families are known to be present in sea anemone venoms (Madio et al., 2017; Macrander et al., 2015a, 2015b, 2016; Macrander and Daly, 2016). A tBLASTn search of these toxin sets against the *S. helianthus* trinity assembly (E-value < 10) was conducted to identify toxin sequence candidates. After eliminating the duplicated hits, the number of sequences per toxin category and the average expression within each category were analyzed.

StnI and StnII protein sequences were used as queries in a search against the *S. helianthus* transcriptome nucleotide sequences database using tBLASTn (E-value $< 1e-5$, matching length $> 60\%$). We used tBLASTn because protein sequences are more conserved than nucleotides and the search was consequently less restrictive than the BLASTn query. The assembly and the subsequent analysis was performed on the University of California Riverside Biocluster (Intel Broadwell 32 cores and 512 GB memory).

A phylogenetic analysis of Stn transcripts was conducted along with

other closely related actinoporin sequences. The purpose was to assess relationships of previously known and novel Stn toxins, rather than to reconstruct a comprehensive phylogenetic tree for all actinoporins as in Macrander and Daly (2016). Closely related sequences were obtained by performing a BLASTx search of Uniprot using StnI and II as query sequences, taking the top 15 hits, and limiting analyses to hits with corresponding nucleotide sequences. Nucleotide sequences were trimmed to their coding sequence and aligned by their translated amino acids using MUSCLE implemented in geneious 9.1.8, keeping translated codons in frame. MrModelTest2.3 (Nylander, 2004) was used to determine an appropriate nucleotide substitution model (GTR + I + G), which was used in a Bayesian phylogenetic analysis with MrBayes3.2.6 (Ronquist et al., 2012). The Bayesian analysis was run for 5×10^6 generations, sampling trees every 1000 generations and discarding the first 25% of sampled trees as burn-in. A 50% majority-rule Bayesian consensus tree was computed from post burn-in trees to determine clade posterior probabilities and the consensus tree was mid-point rooted.

2.6. Novel actinoporin complete sequence determination

Two potential new actinoporin sequences discovered through the BLAST search against *S. helianthus*' de novo transcriptome assembly database (termed StnIII and IV) were PCR-amplified. First cDNA was synthesized with the SuperScript III First-Strand Synthesis SuperMix kit (Thermo Fisher) using RNA extracted from the same anemone individual used for the high-throughput sequencing described above. In order to obtain the sequence of mature StnIII, a complementary primer was designed against its 5' edge (5'-ATGAATCCTTTAGCAGTCGCTGG TGC-3'). Taking into account the lack of differences between StnI and II at their 3' end and minor variation in putative StnIII and IV, a degenerate 3' primer was designed to match these sequences: (C/T)TA(A/G) (C/T)G(A/T)GA(A/G)ATCT(C/T)AAATTTGCA(G/T)TTTTGCTTC GCC-3'). A PCR program was run in two phases: five initial cycles at a low annealing temperature (40 °C) followed by 30 cycles at 55 °C as the annealing temperature. The resulting amplified product was purified and Sanger sequenced.

3. Results and discussion

3.1. Sequencing and assembly

The obtained transcriptome was assembled from 42,678,357 paired-end reads. The assembly quality data is summarized in Table 1. The two assemblies generated with different versions of Trinity yielded a number of genes (177,655–v2.4.0– and 140,745–v2.0.6) and isoforms (226,240–v2.4.0– and 194,773–v2.0.6) comparable with other previously described sea anemone's transcriptomes (Macrander et al., 2015a, 2015b, 2016; Ayala-Sumuano et al., 2017; Tulin et al., 2013). N50 parameters were high enough to ensure a sufficiently good assembly with large contigs (N50 1544–v2.4.0– and 1590–v2.0.6) and again, comparable to other sea anemone's assemblies obtained with the same program (N50 392–1832) (Macrander et al., 2015a, 2015b, 2016; Ayala-Sumuano et al., 2017). Although both Trinity versions yielded

Table 1
Statistics of the *S. helianthus* assembly with two different Trinity versions.

Trinity version	2.4.0	2.0.6
No. of paired-end reads	42,678,357	
No. of total assembled bases	183,455,014	171,292,869
No. of assembled genes	171,655	140,745
No. of assembled transcripts	226,240	194,773
Assembly GC percent	39.74	39.72
Contig N50	1544	1590
Average contig	810.89	879.45

Table 2
BUSCO statistics for both assemblies with different Trinity versions.

Trinity version	2.4.0	2.0.6
Total BUSCO groups searched	978	
Complete BUSCOs	918/93.9%	921/94.2%
Fragmented BUSCOs	45/4.6%	43/4.4%
Missing BUSCOs	15/1.5%	14/1.4%

similar results, the absolute isoform and gene numbers were higher when using Trinity v2.4.0, with a higher degree of fragmentation of the contigs. The N50 parameter and average contig length comparison between both methods suggested that Trinity v2.0.6 rendered a less fragmented assembly presenting larger contigs and a reduced number of total predicted genes and isoforms.

Completeness scores calculated with BUSCO software are shown in Table 2. BUSCO's completeness scores were high for both assembly's attempts, performed with v2.4.0 and v2.0.6 (93.9% and 94.2% expected conserved full-length proteins were found respectively), and the missing BUSCOs were low (1.5% and 1.4%), showing that both Trinity assemblies contained nearly all expected proteins in an unfragmented state.

3.2. Toxin protein venom profile

Sea anemone venoms are composed of a complex mixture of toxins. Some of the most common/abundant toxins are the neurotoxins KTxs-I, KTxs-III and NaTxs, the cytolysins, which include actinoporins and other pore forming toxins, and enzymes capable of modifying membranes and extracellular matrix like phospholipases and metalloproteases. All these toxins were considered for the profiling analysis performed. In addition, the main venomous components whose proteomic evidence has been recently described for *Stichodactyla haddoni* (Madio et al., 2017), an anemone species belonging to the same genus as *S. helianthus*, were included, for example Ankyrin-repeat proteins, S1 peptidases and FBG-domain proteins. Considering the slightly improved, less fragmented assembly obtained with Trinity v2.0.6 all further analyses were continued with these transcripts.

All sequences present in the ToxProt data set, whose description contained the keyword for each kind of anemone toxin described above, were pooled in a single multiFASTA file. These sequences were used as a query against the database built with the *S. helianthus* assembly in tBLASTn search. The results showed phospholipases A2 (PLA2s) represented the most numerous category regarding toxin isoforms number, while cytolysins corresponded to one of the smallest categories (Fig. 1A). However, in terms of relative expression, the highest average value was by far found for cytolysins, while the lowest one was represented by metalloproteases (M12A and M13 subgroups) (Fig. 1B). Overall, this distribution would be consistent with the fact that actinoporins are not enzymes and therefore they do not show catalytic effects as phospholipases and metalloproteases do. StnII is the most potent actinoporin known from cytolytic assays against red blood cells (García-Linares et al., 2016a). Our present results show it is also the most highly expressed Stn (Table 3) in *S. helianthus*, potentially explaining the potency of its venom in causing cell lysis.

We found that the total number and abundance (in terms of TPM) of the toxin categories differed substantially between *S. helianthus* and that reported for *S. haddoni* (Madio et al., 2017). But this pattern of high variability of toxin cocktails among closely related species has been noted for other anemones (Macrander et al., 2016). Considering that the condition of the anemones in our study prior to dissection may have influenced expression levels, and that our transcriptome represents a whole-body transcriptome, rather than one specific to venom-secreting tissues as produced for *S. haddoni* (Madio et al., 2017), our estimates of sequence diversity and expression levels may not be directly

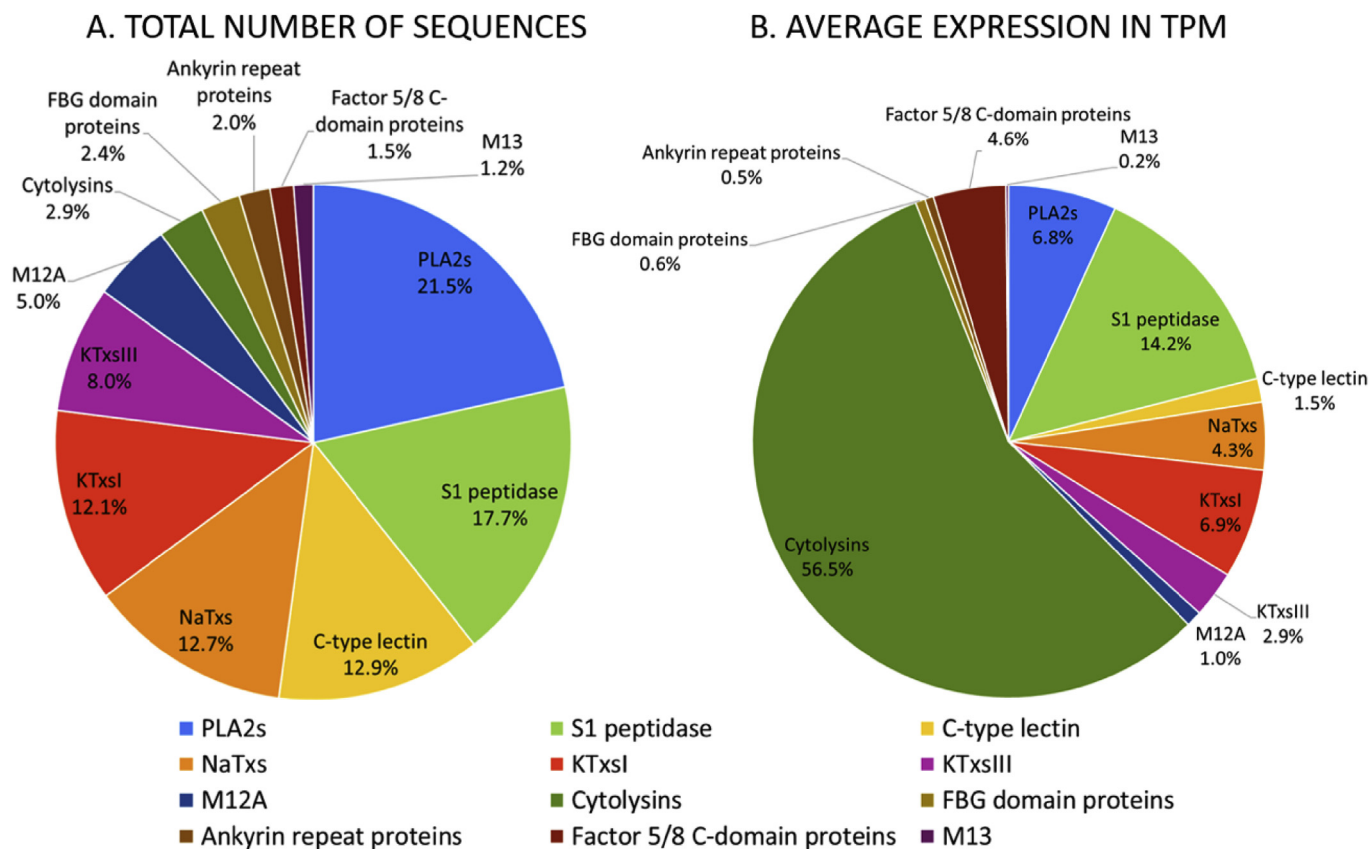


Fig. 1. Summary of the putative components of the *S. helianthus* venom secretion. (A) Total number of sequences found for each toxin type. (B) Average expression of the transcripts recovered in TPM for each toxin category. Selected toxins types in this study are indicated in the legend.

Table 3
Average RSEM calculated Transcripts per Million (TPM) for the Stn predicted transcripts.

Trinity version	2.4.0	2.0.6
StnI	389	455
StnII	704	733
StnIII	129	125
StnIV	113	62

comparable. Thus, our results should be considered a first approximation for *S. helianthus* in terms of relative expression values, given that the putative toxins identified in this study could also expressed in non-venomous tissues and have alternative functions. Further, it is unlikely that all the transcripts we identified as toxins are functionally venomous because the criteria for inclusion was based solely on BLAST search e-scores and the prediction of a function does not guarantee actual function.

A tBLASTn search was also performed with the complete ToxProt database against the *S. helianthus* transcriptome assembly to find putative toxin sequences corresponding to other proteins not necessarily specific to sea anemones venom. Highly similar hits (e-value < 1e⁻⁵ and an overlapping length > 60%) were the subject of the search. After merging the list with the expression data obtained with RSEM, the ten most highly expressed sequences were examined for further analysis (Table 4). Some of the most highly expressed predicted toxic proteins did not have their top BLAST hit to sea anemone toxins. Instead, their top BLAST hits to ToxProt proteins were to toxins in families such as calmodulin, PLA2s and cysteine-rich proteins from snakes, scorpions and corals. However, PLA2s and cysteine-rich proteins are some of the most well studied toxin families in sea anemones, and these results may simply be due to the absence of these anemone toxin proteins in the

ToxProt database. Nevertheless, the most highly expressed toxins were identified as known anemone venom compounds such as the most highly expressed hit U-actitoxin-Avd8b (PODMZ4) from *Anemonia viridis*, the specific function of which is still unknown. In this analysis, StnI and StnII did not appear as hits because they are not included in the ToxProt database. The contigs corresponding to StnI and StnII found however a highly similar protein in actinoporins from *A. villosa* (Q5R231) and *A. equina* (P6191), which were the second and third most highly expressed transcripts, respectively (Table 4). Indeed, the expression values for these proteins match the data recorded in Table 3 regarding expression levels for StnI and StnII. The fourth sequence on the list is a voltage-gated sodium channel inhibitor and the fifth on the list includes the first PLA2 (Table 4).

3.3. New sticholysin sequence prediction and evolution

Though the older Trinity version (v2.0.6) resulted in lower fragmentation, having longer contigs and slightly better BUSCO scores, we kept both program assemblies when looking for new potential actinoporins because these toxins are extremely similar at the nucleotide level (StnI and StnII share 93% of their amino acid residues). Trinity can collapse fragment reads from different isoforms into the same transcript because some fragment reads from different isoforms may be nearly identical in certain regions. Thus, we compared the results from both assemblies since the one obtained with Trinity v2.4.0 is more fragmented so this version may have assembled fewer fragments into the same transcript. Trinity v2.4.0 offers a more conservative assembly approach, although it returns more incomplete transcripts. On the other hand, while the 2.0.6 version gives back less fragmented transcripts, it is combining reads to build a single transcript so it could return more complete sequences.

A tBLASTn search using Stn I and II against a database built with

Table 4

Most highly expressed transcripts among all the ToxProt data tBLASTn hits. Actinoporins are marked with an asterisk.

Transcript ID	TPM	UniProtKB	Animal group	Gene name	E-value
TR74943 c0_g1_i1	1291	PODMZ4	Anemone	U-actitoxin-Avd8b	2.00E-25
*TR72237 c1_g3_i1	734	Q5R231	Anemone	DELTA-thalatoxin-Av11a	2.00E-62
*TR72237 c1_g3_i2	455	P61914	Anemone	DELTA-actitoxin-Aeq1a	4.00E-93
TR22445 c1_g4_i1	421	E2S064	Anemone	Kappa-thalatoxin-Cad2a	7.00E-42
TR56102 c1_g1_i5	285	Q9PUG8	Snake	Acidic phospholipase A2 S16-19	3.00E-18
TR59776 c1_g1_i7	201	Q8AY75	Snake	Calglandulin	5.00E-35
TR51365 c5_g3_i1	190	POCJ14	Scorpion	Venom protein 302	2.00E-09
TR62101 c0_g8_i4	136	PODL54	Anemone	Delta-actitoxin-Avd1e 3	6.00E-06
TR52902 c1_g1_i4	80	COH691	Coral	Small cysteine-rich protein 2	1.00E-07
TR2041 c1_g2_i2	74	PODMZ6	Anemone	U-actitoxin-Avd8d	5.00E-41

both Trinity assemblies was performed which found the following result: 25 or 21 different actinoporin candidates using v2.4.0 or v2.0.6, respectively. After pulling out the nucleotide sequences and the Transdecoder predicted ORF for each candidate transcript, alignments with StnI and StnII were performed through Clustal Omega (Sievers et al., 2011). Most of these potential actinoporins showed a universally conserved motif present in all previously characterized actinoporins: “**Y**/FDYNWY**S**N**W**W” (Garcia-Linares et al., 2016a). Residues in bold are part of a cluster of aromatic residues involved in membrane binding (Garcia-Linares et al., 2016b; Garcia-Ortega et al., 2011; Malovrh et al., 2000), while the underlined tyrosine belongs to the sphingomyelin-specific phosphorylcholine binding site (Mancheno et al., 2003). These are two of the most significant structural signatures regarding actinoporin activity. Closer inspection of each alignment revealed that not all the potentially new toxins shared this conserved domain. Thus, they were discarded as new actinoporins. In addition, considering that actinoporins are generally cysteine-less and their average size is around 175 amino acids, these two criteria were also used to identify potential new actinoporins. Small fragments with an overlapping identity score over 80% were kept. After this strict selection, only four transcripts for each version (v2.4.0 and v2.0.6) were finally retrieved. The resulting selected actinoporins were used as a query for a new search in the *S. helianthus* transcriptome but it did not render any new hits different from the ones obtained in the first search.

In both transcriptomes, two of the four retrieved actinoporin sequences were the already known StnI and StnII, having 100% identity with StnI and 100% identity with StnII reported by Huerta & col. in 2001 (Huerta et al., 2001). Only the older Trinity version (2.0.6) retrieved complete sequences, while the newest version showed StnI and II with a truncated C-termini. Both assemblies yielded a similar, though not identical sequence, here termed StnIII (Fig. 2). This sequence was almost complete in both cases and only lacked a small fragment at the C-terminal edge. However, this missing fragment occurs in a sequence stretch which is also conserved across all the most-well studied actinoporins (Garcia-Linares et al., 2016a).

A fourth transcriptomic sequence, we initially named StnIV, was also found in both assemblies, though containing only the C-terminal stretch (Fig. 2). This feature precluded its complete analysis given that major differences in these proteins primarily accumulate at the N-terminal end (Garcia-Linares et al., 2016a; Ros et al., 2015). However, the sequences recovered as StnIV in the two assemblies were not identical but differed at some positions as well as to StnI and II. One possible explanation for this observation was that StnIV represented the protein of a different paralog to those encoding StnI-III. However, PCR amplification and sequencing of Stn III (see below) shows it is nearly identical at its overlap with Stn IV, suggesting they may represent allelic variants of the same locus.

All actinoporin sequences starting from a methionine codon (StnI, II and III) present a signal peptide, reported before for other actinoporins such as equinatoxin (Anderluh et al., 1999, 2000; Pungercar et al., 1997) but never before described for sticholysins (Fig. 3). This signal

peptide comprises the first 34 amino acids of the deduced protein sequence and it shows the typical arrangement of a eukaryotic signal peptide (Kapp et al., 2009). This region can be divided into two functionally well-defined domains: a signal peptide, followed by a pro-sequence region, both of them leading actinoporins to their presumed final localization in nematocysts. The signal peptide region is constituted by a standard distribution made of a hydrophilic positively charged N-terminal, followed by a hydrophobic stretch. Overall, this signal sequence constitutes a motif which is conserved in other known actinoporins (Fig. 3). The pro-part region comprises the last 15–16 amino acids. This region is mainly polar and charged, highly conserved among the actinoporins within the same species, but highly variable in terms of length and net charge among different species. The predicted cleavage site is a conserved Lys/Arg pair (Kex2-like and subtilisin-like proteases).

As described above, the predicted transcripts for StnI and StnII showed the highest expression levels while the new potential variants had a much-reduced representation in comparison. For example, the StnII transcript was around seven times more abundant than presumed StnIII and StnIV isotoxins. Both assemblies gave similar expression data, again confirming that the assembly merged into a similar number of fragments.

The main difference between both assembly methods was also reflected in the relative expression values obtained for the sequence provisionally designated as StnIV. This sequence was lowly expressed which likely accounts for it being fragmented and incomplete. This ambiguity might be solved through more deeply sequencing the transcriptome. It is easier to assemble sequences when the transcripts are more abundant in the mixture and therefore, considering that actinoporins share high identity, the Trinity program might be collapsing similar but not identical sequences, reducing the potential to retrieve variant sequences. This would explain why StnI and StnII were obtained as full-length sequences while the less highly expressed sequences were fragmented and missing pieces (Table 3). The reduced expression of StnIII could be also an artifact if its expression was suppressed under laboratory conditions and would be more accurately assessed in transcriptomes derived from venom-producing tissues. Nevertheless, if indeed lowly expressed, StnIII could still induce dramatic effects in the activity of *S. helianthus* venom given that small amounts of StnII can enhance Stn I binding to lipid model vesicles and hemolysis.

Our phylogenetic analysis of *S. helianthus* sticholysins show StnI and II as most closely related to each other, but that StnIII is divergent and instead is placed as more closely related to other species' actinoporins (Fig. 5). As actinoporins appear to be encoded by a multi-genic family in other closely related anemone species (Macrander and Daly, 2016), this suggests that StnI and II potentially arose following a species-specific gene duplication event. In contrast, the phylogenetic position of StnIII suggests the duplication event giving rise to this toxin gene predates the origin of *S. helianthus* and the duplication event that generated Stn I and II. The greater divergence of StnIII relative to Stn I and II reinforces the need to functionally evaluate the role of StnIII in *S.*

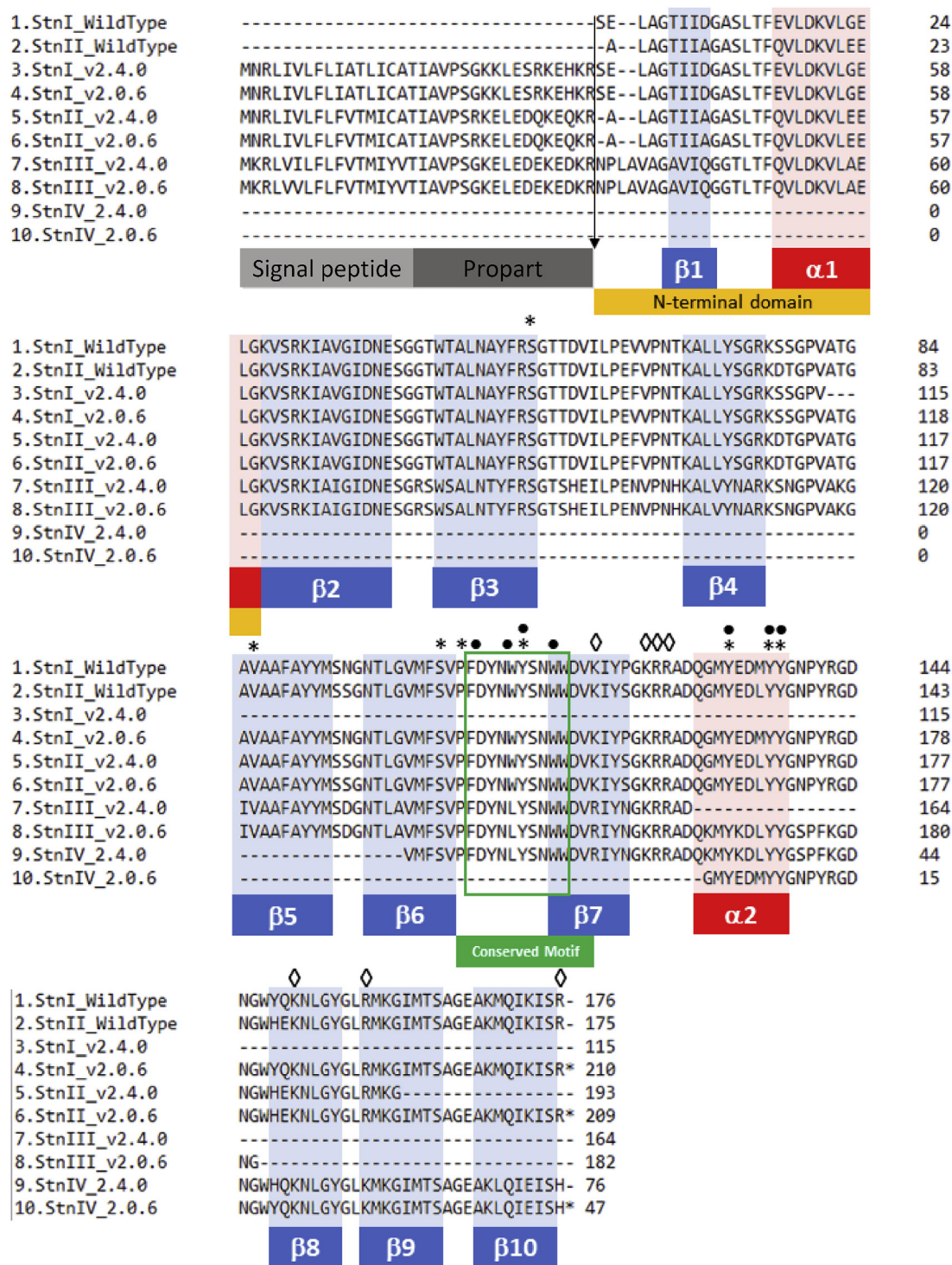


Fig. 2. Predicted protein sequence alignments for the tBLASTn selected hits for the wild type toxins StnI and StnII (sequences 1 and 2). Two Trinity versions were assayed for the *de novo* assembly: Trinity v2.4.0 and Trinity v2.0.6. The transcripts are sorted by decreasing identity percentage when compared with wild type StnI and StnII. Alignment performed with Clustal Omega. In light grey the signal peptide sequence followed by the propeptide in dark grey. A dashed arrow marks the excision position. Highlighted in blue the β -sheets and in red the α -helices experimentally determined for StnI and StnII wild type. In orange the N-terminal domain. Squared in green is the conserved motif cited in the text. Residues part of the phosphorylcholine binding site are marked with an asterisk. Black dots marks residues belonging to the cluster of aromatic residues. Diamonds point out the array of basic amino acids. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

helianthus venom, given its potential to modulate the toxicity of StnI and II in a synergistic fashion.

3.4. Novel actinoporin complete sequence and structural prediction

Heterologous production and functional and structural characterization of newly discovered actinoporins will be extremely helpful in improving our understanding of the role of these proteins within sea

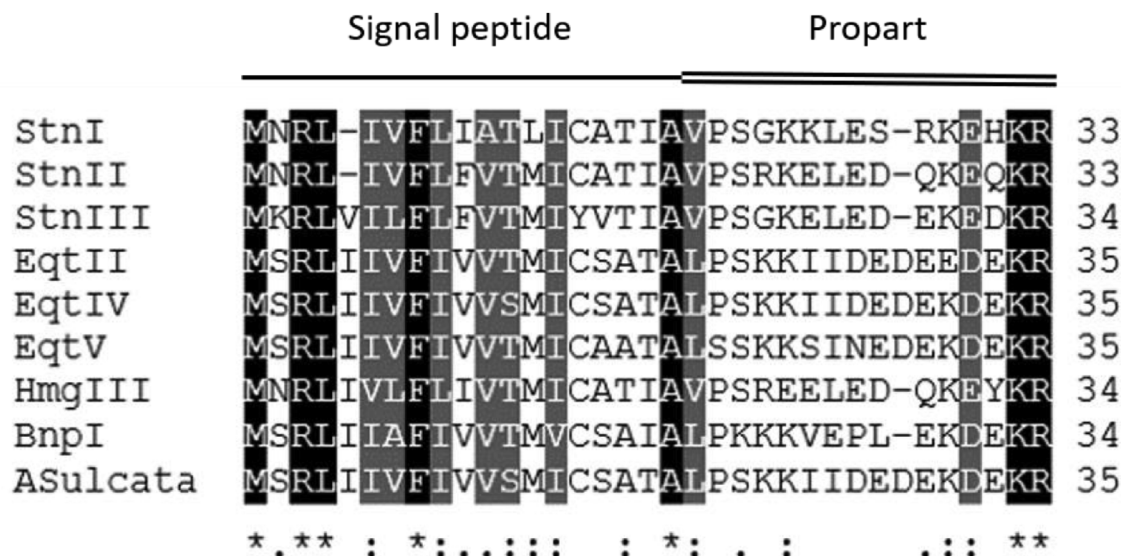


Fig. 3. Aligned signal peptides (single line) and propeptide (double-line) from different actinoporins. StnI, StnII and StnIII signal peptides were obtained in the Trinity assembly from this study. Other sequences were obtained from UniprotKB. Accession codes are the following: EqtII-P61914, EqtIV-Q9Y1U9, EqtV-Q93109, HmgIII-Q9U6X1, BnpI-C5NSL2, *A. sulcata*-A0A0S1M135. Black shade and asterisks mark identical residues, grey shade and colons (:) highlight conserved substitutions between groups of strongly similar properties, and dots (.) indicate conservation between groups of weakly similar properties.

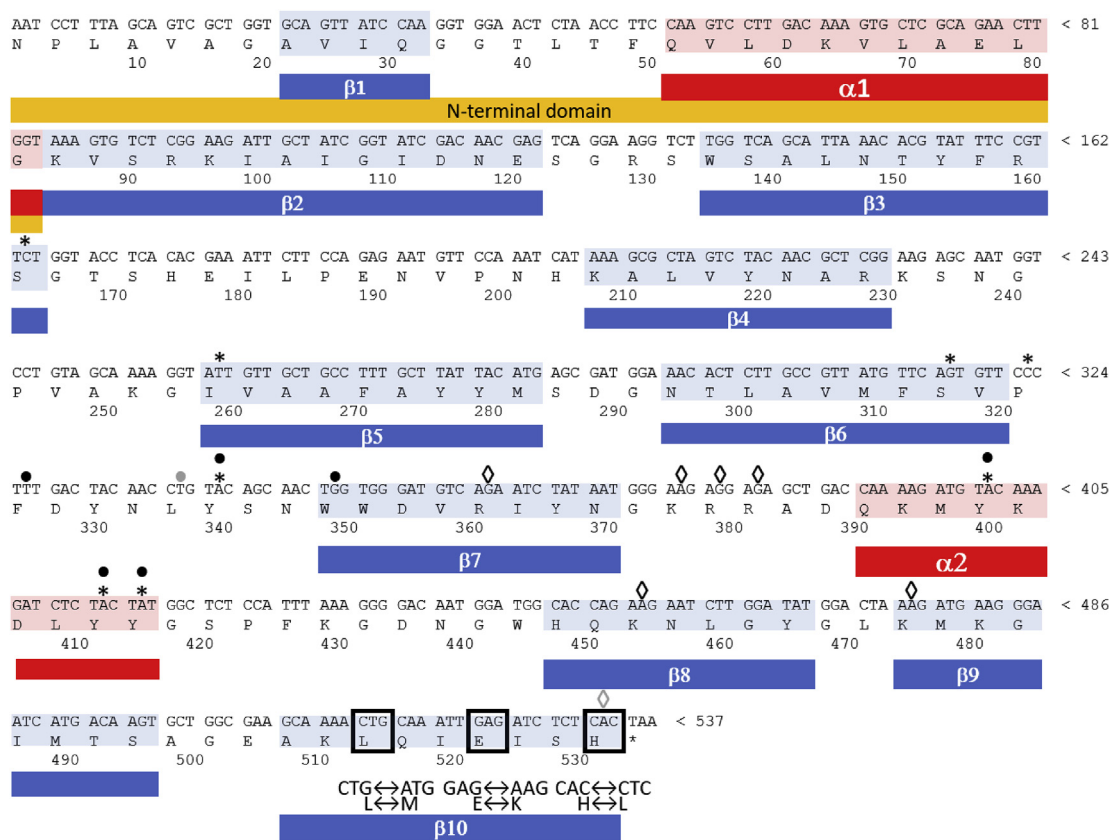


Fig. 4. StnIII mature sequence translated from sequenced PCR product (MH327769). Highlighted in blue the β -sheets and in red the α -helices experimentally determined for StnI and StnII wild type. In orange the N-terminal domain. Residues part of the phosphorylcholine binding site are marked with an asterisk. Black dots mark residues belonging to the cluster of aromatic residues. Diamonds indicate the array of basic amino acids. Grey dot (CTG-337) and grey diamond (CAC-533) indicate a residue part of the cluster of aromatic residues and the array of basic amino acids not conserved with respect to StnI and StnII. Squared in black are differences in the StnIII sequence relative to StnIV at their overlap (see Fig. 2). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

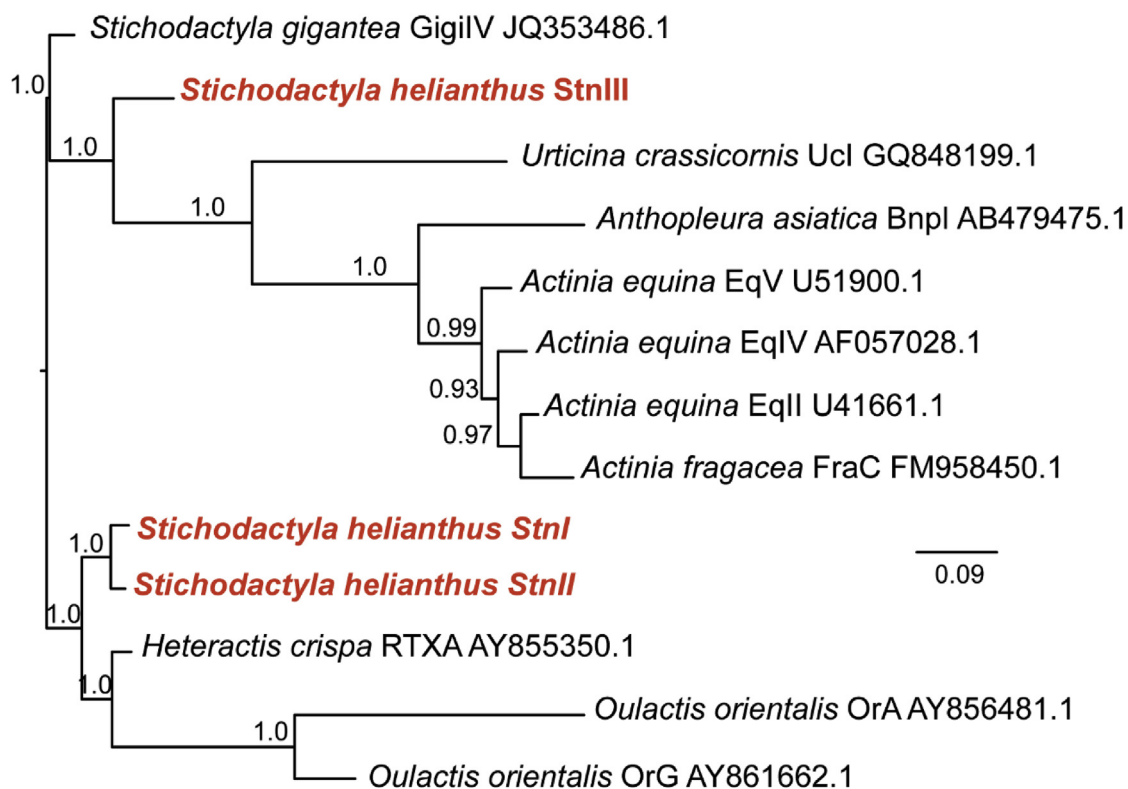


Fig. 5. Phylogenetic tree (Bayesian consensus) depicting relationships of *S. helianthus* sticholysins (StnI-III – in bold red text) in relation to closely related actinoporin sequences. Numbers at nodes indicate clade posterior probabilities computed from post burn-in trees. Tree is 50% majority rule of post burn-in trees, and is mid-point rooted. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

anemones venom. Thus, we fully sequenced the mature form of StnIII from a single band that was PCR-amplified from cDNA from one of the same individuals used for the transcriptome. The C-terminal domain of StnIII overlapped with the sequence obtained for StnIV and exhibited only three non-synonymous differences (Fig. 4). We therefore conclude that StnIII and IV may correspond to allelic variants of the same locus. However, given the C-terminal region is relatively invariant, we cannot unequivocally reject the possibility that the StnIV fragment is encoded by a different locus and may be connected to a more variable upstream region we were unable to obtain. The complete sequence determined for StnIII was deposited in NCBI's GenBank database (MH327769).

StnI and II share 93% amino acid identity, whereas StnIII is 76% identical to StnI and 77% to StnII at the protein level. This sequence conservation allowed the prediction of StnIII's three-dimensional structure by comparing with the already known 3D structures of StnI and StnII (Mancheno et al., 2003; Castrillo et al., 2009). This prediction was achieved using the Phyre2 web portal software (Kelley et al., 2015) which compares the primary sequence of the query with the amino acid sequence of already known structures. As expected, the model for StnIII's structural prediction displays the typical actinoporin folding assembly: an antiparallel β -sandwich flanked by two α -helices (Fig. 6).

Amino acid sequence comparison also shows that the N-terminal stretch of StnIII is 2 and 3 residues longer than the corresponding region in StnI and StnII. The hydrophobicity index of its 18 first residues (calculated according to (Eisenberg et al., 1984)), a critical feature for its cytolytic properties (Rivera-de-Torre et al., 2017), was very similar to StnI (Rivera-de-Torre et al., 2017). The hydrophobic moment, a magnitude that measures the affinity for the membrane core, is lower than StnI and StnII (0.209), suggesting a lower tendency to cross the membrane.

All the residues comprising important domains for the actinoporins function (García-Ortega et al., 2011) are basically conserved or substituted by residues with similar physicochemical properties in StnIII.

Within this context, Leu-113 in StnIII, corresponding to Trp-111 in StnI and Trp-110 in StnII stands out because this residue belongs to the so called cluster of aromatic residues, responsible for driving membrane binding and involved in maintaining the required hydrophobicity to achieve that goal. In agreement with this observation, StnII Trp-110 has been recently described as penetrating the hydrophobic membrane core (García-Linares et al., 2016b). Accordingly, substitution by another highly hydrophobic amino acid, as is the case with Leu-113, is expected to have only minor influence on StnIII membrane interaction ability.

A more specific analysis of the different structural features and functionalities among the three sticholysins will still have to wait for the heterologous production of StnIII and its corresponding structural and functional analysis.

4. Conclusions

Sea anemone venom has been studied from different points of view and for different reasons, such as elucidating the molecular evolution of cnidarians (Macrander et al., 2016; Valle et al., 2015; Macrander and Daly, 2016) or obtaining biotechnological tools (Deng et al., 2016). Among the most well studied toxins in the sea anemone venom, PLA2s represent the largest family, with at least 50% more members than the next more representative group, NaTxs. Cytolysins are a much less diverse family but they show much higher average levels of expression. That is to say, they constitute a very important component of the toxic mixture in quantitative terms. This would also agree with the non-enzymatic character of their toxic effect. Interestingly, we found that actinoporins are some of the most highly expressed toxins considering all the possible venom and toxins hits gathered in the ToxProt data set. Therefore, actinoporins may be one of the main components in many sea anemone venoms. In the specific case of *S. helianthus* two actinoporin isoforms had been previously studied in deep detail (García-Linares et al., 2013, 2016a, 2016b, 2016c; Maula et al., 2013). StnII is

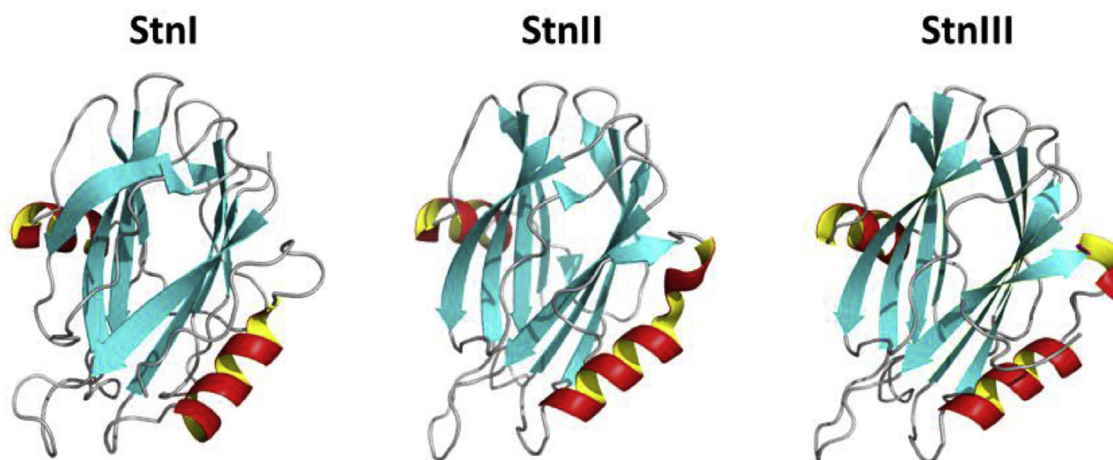


Fig. 6. Three-dimensional soluble structures of sticholysins I (PDB 2KS4), and II (PDB 1GWY) in comparison to the StnIII 3D structural prediction. All figures were made using Pymol 1.3 modeling software (The PyMOL). β -sheets are shown in blue, α -helices in red-yellow and disordered loops are grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the most potent actinoporin among the most well studied ones (García-Linares et al., 2016a) and displays a synergistic effect in combination with StnI (Rivera-de-Torre et al., 2016). After our transcriptome analysis, we can conclude that there is at least one new actinoporin isoform which had not been described before even though there was evidence of the existence of other actinoporin variants in this anemone species (de los Ríos et al., 2000). Transcriptomic approaches have become common to study the composition and diversity of venom, as well as the evolutionary genomics of venom production, with special emphasis on the evolution of multigene families. While transcriptome analysis is based on predictions around sequence similarity, it is necessary to perform functional studies of the main toxins to determine toxin activity, and to understand the ecological role of venom toxins. Thus, StnIII from *S. helianthus* venom must be the subject of further study given that it could be a key factor in modulating the toxic effect and range of prey targeted by sea anemones.

Acknowledgements

We thank Robert Haney, Erin Sullivan, Fitz Anthony Forsyth and Evelyn Schwager for their support in experimental design and data analysis. This work was supported by a Banco Santander-UCM fellow and an Erasmus + short fellowship to E.R.-d.-T., as well as a Major Research Instrumentation Grant from the US National Science Foundation (DBI-1429212) to UMass Lowell and Santander-UCM Project PR41_17_2012 to AMP. We also thank Jack Lepine and the UMass Lowell CRF for support in the Illumina sequencing.

References

Anderluh, G., et al., 1999. Equinatoxins, pore-forming proteins from the sea anemone *Actinia equina*, belong to a multigene family. *Toxicon* 37 (10), 1391–1401.

Anderluh, G., Podlessek, Z., Macek, P., 2000. A common motif in parts of *Cnidarian* toxins and nematocyst collagens and its putative role. *Biochim Biophys Acta*. 1476 (2), 372–376.

Ayala-Sumano, J.T., et al., 2017. Sequencing and de novo transcriptome assembly of *Anthopleura dowii* Verrill (1869), from Mexico. *Genom Data*. 11, pp. 92–94.

Basulto, A., et al., 2006. Immunohistochemical targeting of sea anemone cytolytins on tentacles, mesenteric filaments and isolated nematocysts of *Stichodactyla helianthus*. *J. Exp. Zool. Comp. Exp. Biol.* 305 (3), 253–258.

Bryant, D.M., et al., 2017. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* 18 (3), 762–776.

Castrillo, I., et al., 2009. 1H, 13C, and 15N NMR assignments of the actinoporin Sticholysin I. *Biomol NMR Assign* 3 (1), 5–7.

Chomczynski, P., 1993. A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques* 15 (3), 532–534 536–7.

de los Ríos, V., et al., 2000. Overproduction in *Escherichia coli* and purification of the hemolytic protein sticholysin II from the sea anemone *Stichodactyla helianthus*. *Protein Expr. Purif.* 18 (1), 71–76.

Deng, Y., et al., 2016. Sphingomyelin is sorted at the trans Golgi network into a distinct class of secretory vesicle. *Proc. Natl. Acad. Sci. U. S. A.* 113 (24), 6677–6682.

Eisenberg, D., et al., 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* 179 (1), 125–142.

Garb, J.E., Hayashi, C.Y., 2005. Modular evolution of egg case silk genes across orb-weaving spider superfamilies. *Proc. Natl. Acad. Sci. U. S. A.* 102 (32), 11379–11384.

Garb, J.E., Gonzalez, A., Gillespie, R.G., 2004. The black widow spider genus *Latrodectus* (Araneae: theridiidae): phylogeny, biogeography, and invasion history. *Mol. Phylogenet. Evol.* 31 (3), 1127–1142.

García-Linares, S., et al., 2013. Three-dimensional structure of the actinoporin sticholysin I. Influence of long-distance effects on protein function. *Arch. Biochem. Biophys.* 532 (1), 39–45.

García-Linares, S., et al., 2016a. Differential effect of membrane composition on the pore-forming ability of four different sea anemone actinoporins. *Biochemistry* 55 (48), 6630–6641.

García-Linares, S., et al., 2016b. Role of the tryptophan residues in the specific interaction of the sea anemone *Stichodactyla helianthus*'s actinoporin sticholysin II with biological membranes. *Biochemistry* 55 (46), 6406–6420.

García-Linares, S., et al., 2016c. Toxin-induced pore formation is hindered by intermolecular hydrogen bonding in sphingomyelin bilayers. *Biochim. Biophys. Acta* 1858 (6), 1189–1195.

García-Ortega, L., et al., 2011. The behavior of sea anemone actinoporins at the water-membrane interface. *Biochim. Biophys. Acta* 1808 (9), 2275–2288.

Grabherr, M.G., et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29 (7), 644–652.

Haas, B.J., et al., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8 (8), 1494–1512.

Haney, R.A., et al., 2016. Effects of gene duplication, positive selection, and shifts in gene expression on the evolution of the venom gland transcriptome in widow spiders. *Genome Biol Evol* 8 (1), 228–242.

Huerta, V., et al., 2001. Primary structure of two cytolytic isoforms from *Stichodactyla helianthus* differing in their hemolytic activity. *Toxicon* 39 (8), 1253–1256.

Kapp, K.S.S., Lemberg, M.K., Dobberstein, B., 2009. Post-targeting Functions of Signal Peptides, in Protein Transport into the Endoplasmic Reticulum. (Landes Bioscience).

Kelley, L.A., et al., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10 (6), 845–858.

Kitchen, S.A., et al., 2015. De novo assembly and characterization of four anthozoan (phylum Cnidaria) transcriptomes. *G3 (Bethesda)* 5 (11), 2441–2452.

Krueger, F., 2012. Trim Galore! Version 0.4.3. Available from: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

Lanio, M.E., et al., 2001. Purification and characterization of two hemolysins from *Stichodactyla helianthus*. *Toxicon* 39 (2–3), 187–194.

Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* 12, 323.

Macrander, J., Daly, M., 2016. Evolution of the cytolytic pore-forming proteins (actinoporins) in sea anemones. *Toxins* 8 (12).

Macrander, J., Brugler, M.R., Daly, M., 2015a. A RNA-seq approach to identify putative toxins from acrorhagi in aggressive and non-aggressive *Anthopleura elegantissima* polyps. *BMC Genom.* 16, 221.

Macrander, J., Broe, M., Daly, M., 2015b. Multi-copy venom genes hidden in de novo transcriptome assemblies, a cautionary tale with the snakelocks sea anemone *Anemonia sulcata* (Pennant, 1977). *Toxicon* 108, 184–188.

Macrander, J., Broe, M., Daly, M., 2016. Tissue-specific venom composition and differential gene expression in sea anemones. *Genome Biol Evol* 8 (8), 2358–2375.

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

- Malovrh, P., et al., 2000. Structure-function studies of tryptophan mutants of equinatoxin II, a sea anemone pore-forming protein. *Biochem. J.* 223–232 346 Pt 1.
- Mancheno, J.M., et al., 2003. Crystal and electron microscopy structures of sticholysin II actinoporin reveal insights into the mechanism of membrane pore formation. *Structure* 11 (11), 1319–1328.
- Maula, T., et al., 2013. 2NH and 3OH are crucial structural requirements in sphingomyelin for sticholysin II binding and pore formation in bilayer membranes. *Biochim. Biophys. Acta* 1828 (5), 1390–1395.
- Monastyrnaya, M., et al., 2010. Actinoporins from the sea anemones, tropical *Radianthus macrodactylus* and northern *Oulactis orientalis*: comparative analysis of structure-function relationships. *Toxicon* 56 (8), 1299–1314.
- Nylander, J.A.A., 2004. MrModeltest V2. Program Distributed by the Author. Evolutionary Biology Centre, Uppsala University.
- Pungercar, J., et al., 1997. Sequence analysis of the cDNA encoding the precursor of equinatoxin V, a newly discovered hemolysin from the sea anemone *Actinia equina*. *Biochim. Biophys. Acta* 1341 (2), 105–107.
- Rivera-de-Torre, E., et al., 2016. Synergistic action of actinoporin isoforms from the same sea anemone species assembled into functionally active heteropores. *J. Biol. Chem.* 291 (27), 14109–14119.
- Rivera-de-Torre, E., et al., 2017. One single salt bridge explains the different cytolytic activities shown by actinoporins sticholysin I and II from the venom of *Stichodactyla helianthus*. *Arch. Biochem. Biophys.* 636, 79–89.
- Rojko, N., et al., 2016. Pore formation by actinoporins, cytolysins from sea anemones. *Biochim. Biophys. Acta* 1858 (3), 446–456.
- Ronquist, F., et al., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61 (3), 539–542.
- Ros, U., et al., 2015. Differences in activity of actinoporins are related with the hydrophobicity of their N-terminus. *Biochimie* 116, 70–78.
- Salgado, V.L., Kem, W.R., 1992. Actions of three structurally distinct sea anemone toxins on crustacean and insect sodium channels. *Toxicon* 30 (11), 1365–1381.
- Sievers, F., et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539.
- Simao, F.A., et al., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31 (19), 3210–3212.
- Suput, D., 2009. In vivo effects of cnidarian toxins and venoms. *Toxicon* 54 (8), 1190–1200.
- The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.**
- Tulin, S., et al., 2013. A quantitative reference transcriptome for *Nematostella vectensis* early embryonic development: a pipeline for de novo assembly in emerging model systems. *EvoDevo* 4, 16.
- Valle, A., et al., 2015. The multigene families of actinoporins (part I): isoforms and genetic structure. *Toxicon* 103, 176–187.
- Wang, Y., et al., 2008. A multigene family of *Heteractis magnificalis* (HMgs). *Toxicon* 51 (8), 1374–1382.