ORIGINAL ARTICLE

Tentacle Transcriptomes of the Speckled Anemone (Actiniaria: Actiniidae: *Oulactis* sp.): Venom-Related Components and Their Domain Structure

Michela L. Mitchell[1,2,3,4] • Gerry Q. Tonkin-Hill[4] • Rodrigo A. V. Morales[1,5] • Anthony W. Purcell[6] • Anthony T. Papenfuss[4,7,8,9,10] • Raymond S. Norton[1,11]

Abstract
Cnidarians are one of the oldest known animal lineages (ca. 700 million years), with a unique envenomation apparatus to deliver a potent mixture of peptides and proteins. Some peptide toxins from cnidarian venom have proven therapeutic potential. Here, we use a transcriptomic/proteomic strategy to identify sequences with similarity to known venom protein families in the tentacles of the endemic Australian 'speckled anemone' (*Oulactis* sp.). Illumina RNASeq data were assembled de novo. Annotated sequences in the library were verified by cross-referencing individuals' transcriptomes or protein expression evidence from LC-MS/MS data. Sequences include pore-forming toxins, phospholipases, peptidases, neurotoxins (sodium and potassium channel modulators), cysteine-rich secretory proteins and defensins (antimicrobial peptides). Fewer than 4% of the sequences in the library occurred across the three individuals examined, demonstrating high sequence variability of an individual's arsenal. We searched for actinoporins in *Oulactis* sp. to assess sequence similarity to the only described toxins (OR-A and -G) for this genus and examined the domain architecture of venom-related peptides and proteins. The novel putative actinoporin of *Oulactis* sp. has a greater similarity to other species in the Actiniidae family than to *O. orientalis*. Venom-related sequences have an architecture that occurs in single, repeat or multi-domain combinations of venom-related (e.g. ShK-like) and non-venom (e.g. whey acid protein) domains. This study has produced the first transcriptomes for an endemic Australian sea anemone species and the genus *Oulactis*, while identifying nearly 400 novel venom-related peptides and proteins for future structural and functional analyses and venom evolution studies.

✉ Michela L. Mitchell
michela.mitchell@qm.qld.gov.au

1 Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville, Victoria 3052, Australia

2 Sciences Department, Museum Victoria, G.P.O. Box 666, Melbourne, Victoria 3001, Australia

3 Queensland Museum, P.O. Box 3000, South Brisbane, Queensland 4101, Australia

4 Bioinformatics Division, Walter & Eliza Hall Institute of Research, 1G Royal Parade, Parkville, Victoria 3052, Australia

5 Present address: CSL Limited, 30 Flemington Road, Parkville, Victoria 3010, Australia

6 Department of Biochemistry and Molecular Biology and Infection and Immunity Program, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia

7 Peter MacCallum Cancer Centre, Melbourne, Victoria 3010, Australia

8 Department of Medical Biology, University of Melbourne, Melbourne, Victoria 3010, Australia

9 Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Victoria 3010, Australia

10 School of Mathematics and Statistics, University of Melbourne, Melbourne, Victoria 3010, Australia

11 ARC Centre for Fragment-Based Design, Monash University, Parkville, Victoria 3052, Australia

Introduction

Cnidarians (sea anemones, jellyfish, corals, and hydroids) have an ancient and unique envenomation system, with microscopic organelles (cnidae) containing venom distributed throughout the ectodermal tissue of the animal. Upon stimulation, cnidae evert an internally coiled thread with enough force to penetrate shell, scales or skin (Kass-Simon and Scappaticci Jr. 2002). The venom can induce a range of physiological reactions in specific biological targets (prey/competitors) and non-specific targets (e.g. humans). The venom contained within the cnidae can cause immobilization, death or tissue necrosis of prey and competitors during predation, aggression and defence (Frances 1973; Shick 1991). Human physiological reactions to the venom include pain, swelling, rashes, anaphylactic shock and in extreme cases renal or hepatic failure resulting in death (Garcia et al. 1994; Williamson et al. 1996; Coleman 1999; Nagata et al. 2006).

Cnidarian venom components have been isolated and assessed for their bioactivity in order to facilitate the development of anti-venom (e.g. box jellyfish) (Andreosso et al. 2014) or for their therapeutic potential (Suput 2011; Pennington et al. 2017; Prentis et al. 2018). The latter is exemplified by the peptide ShK, a potent potassium $K_V1.3$ channel inhibitor sourced from *Stichodactyla helianthus* (giant sun anemone) (Castañeda et al. 1995). ShK has already shown considerable promise as a therapeutic to treat autoimmune disease. An analogue of this peptide (ShK-186) has been developed into the first-in-class clinical candidate dalazatide and phase 1 clinical trials completed for the treatment of psoriasis (Chandy and Norton 2017; Tarcha et al. 2017). The discovery of novel peptides (< 10 kDa) from sea anemone venom is therefore of great interest for their therapeutic potential.

Characterized sea anemone venom components include cytolysins such as actinoporins and MACPF/perforins (Anderluh and Maček 2002), enzymes (e.g. phospholipase A2 (PLA2)) (Romero et al. 2010) and ion channel modulators (targeting acid-sensing, potassium and sodium channels) (Castañeda et al. 1995; Honma and Shiomi 2006; Rodríguez et al. 2014; Cristofori-Armstrong and Rash 2017). Sea anemone venom also contains protease and peptidase inhibitors such as serine peptidase inhibitors (e.g. Kunitz-type) (Beress 1982; Diochot et al. 1998; Monastyrnaya et al. 2016). The Tox-Prot database (Jungo and Bairoch 2005) contains sea anemone venom proteins and toxins sourced from fewer than 50 of the > 1100 known sea anemone species, of which none are endemic to Australia. Moreover, there has been scant research into toxins of the genus *Oulactis* (type locality Australia). Only two cytolytic toxins (OR-A and OR-G) have been isolated from *O. orientalis*, which occurs in Japanese

waters (Il'ina et al. 2005). As members of the genus *Oulactis* feed on small crustaceans (Mitchell 2010), their venom is likely to contain bioactive peptides used in prey capture, making them a potential source of novel peptides for pharmaceutical or agrochemical applications.

The use of venomic and transcriptome-only studies has seen a proliferation of toxin-like sequences deposited in public databases (Macrander and Daly 2016; Logashina et al. 2017; Madio et al. 2017; Ramírez-Carreto et al. 2019). This proliferation of data presents us with an opportunity to examine venom components of an Australian species in the broader context of their heavily studied Northern Hemisphere counterparts and a little studied genus of sea anemone (*Oulactis*). In addition, relationships among domains in venom proteins have not been examined and may yield new protein sequences for functional assessment. For example, the polypeptide EQ3, an isoform of equistatin (a protease inhibitor from *Actinia equina*), has three distinct structural and functional domains (Lenarcic et al. 1997; Štrukelj et al. 2000). The first domain of EQ3 shows activity against the cysteine proteinase papain whilst the following two domains have activity against cathepsin D (Štrukelj et al. 2000). Understanding the association of venom-related domains and function variation according to sequence similarity may aid in the selection of sequences for assessment as potential therapeutic leads.

Here, we describe the first transcriptome for a species from the genus *Oulactis*, the 'speckled sea anemone' *Oulactis* sp. (yet to be formally named and described; Fig. 1) which is endemic to Australia (Mitchell 2010). We assess transcriptome assembly pipelines to compile a library of putative venom-related peptides and proteins and examine the domain architecture of those sequences. Novel peptides and proteins identified in this study will be the subject of future structural, functional and venom evolution studies.

Fig. 1 The Australian endemic, cold temperate, speckled sea anemone *Oulactis* sp. (Brighton, Port Phillip Bay, Victoria, Australia) Photo: Museum Victoria-J. Finn

Materials and Methods

Sea Anemone Collection, Aquaria and Preservation

Specimens (Actiniaria: Actiniidae: *Oulactis* sp.) were collected by hand from the intertidal zone in Brighton, Victoria, Australia (under Museum Victoria collection permit: RP699) and identified by M.L.M. Artificial seawater was prepared using Kirby's premium SPS sea salt and mixed according to manufacturer's instructions. Animals were housed in aquaria using artificial seawater (water temperature 18 °C) and fed prawn meat every 3–4 days. After tissue sample collection (described below), specimens were narcotised using menthol crystals and preserved in 10% formalin: seawater. Preserved specimens of *Oulactis* sp. were deposited in the Museum Victoria invertebrate collection (NMV F223094).

Tissue Collection and Total RNA Extraction

Animals were kept in aquaria for a minimum of 1 week prior to sampling and not fed for 3 days before tissue collection to ensure high venom expression (Schmidt 1982; Madio et al. 2017). Three to four tentacles were removed (~ 40 mg of tissue) from each animal and immediately transferred to individual sterile Eppendorf tubes containing 600 μL of RNA*later*®. Samples were stored at 4 °C for 2 days prior to performing Total RNA extractions.

Tissue samples were lysed on a QIAGEN® Tissue II Lyser (3 mm tungsten carbide beads). Total RNA extractions were performed using the QIAGEN RNeasy Mini Kit® quick-start protocol, incorporating the on-column DNase digestion step (QIAGEN RNase-Free DNase). Extracts were stored at − 20 °C prior to quality testing (Agilent 2100 Bioanalyzer). Extracts meeting a minimum quality criterion (> 9 out of 10) for intact RNA (RIN) were sent for Illumina sequencing at the Monash Health Translation Precint (MHTP)-Medical Genomics Facility.

Total RNA Sequencing and Assessment

cDNA libraries were constructed using Truseq PolyA RNA-seq Illumina stranded sample preparation kit and sequenced on Illumina HiSeq1500 (Chemistry Hi Seq v4; Illumina protocol 15,035,788 Rev. D, Apr 2014)—Rapid Mode, 100 bp Paired End reads and individuals (1, 2 and 3) barcoded. Returned libraries were quality assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

De Novo Assembly

Sequence reads for individual 1 were de novo assembled via three pipelines: Oases/Velvet (v0.2.09/v1.2.10) (Zerbino and Birney 2008; Schulz et al. 2012),

SOAPdenovo-Trans (v1.03) (Yang and Smith 2013; Xie et al. 2014) and trinityrnaseq (version r20140413p1) (Grabherr et al. 2011). Reads were cleaned with Trimgalore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) prior to running the Oases and SOAPdenovo-Trans pipelines, whilst Trimmomatic (v0.36) (Bolger et al. 2014) was used prior to processing with Trinity. For the Oases and SOAPdenovo-Trans pipelines, the ranges of kmer values considered included 21, 31, 41, 51 and 61. The resulting assemblies were then merged as described in Yang and Smith (2013). We used the default kmer value (25) for the Trinity pipeline. Six de novo pipelines were run for assessment, with and without in silico read normalization. Raw reads were assembled individually; in addition, the raw reads of all individuals were pooled and assembled to produce an overarching transcriptome.

Assembly Metrics

By default, the Trinity pipeline removes contigs < 200 nucleotides in length; to ensure consistency, contigs < 200 nucleotides were removed from SOAPdenovo-Trans and Oases assemblies for statistical comparison. Metrics used to compare the assembly pipelines included the number of contigs assembled, maximum contig length, the contig N50, GC content (%) (calculated with the EMBOSS geecee tool (Bruskiewich 1999)), and the percentage of contigs mapping back to reads (Salmon v0.8.0 (Patro et al. 2017) quasi-mapping method). Finally, the gene content of each assembly was assessed using the Benchmarking Universal Single Copy Orthologs program (BUSCO v2.0.1) (Simão et al. 2015) against the Eukaryote lineage database (*E*-value 1×10^{-5}). Versions of software utilized by BUSCO were NCBI BLAST+ (Altschul et al. 1990), AUGUSTUS (v3.2.1) (Stanke et al. 2004) and HMMER (v3.1b2) (Finn et al. 2011).

Functional Annotation of Transcriptomes

Transcriptomes were annotated using Trinotate: Transcriptome Functional Annotation and Analysis Suite (v3.0) (Bryant et al. 2017) (https://trinotate.github.io/). TransDecoder's longest open reading frame (ORFS) length was lowered to 50 aa (amino acids) from the default length of 100 aa to capture short toxin-like peptides. Predicted peptides were annotated with the suite of database searches and prediction tools listed below, prior to loading annotations into the Trinotate prepopulated SQLite database. Sequence similarity searches against published venom peptides and proteins were conducted using NCBI BLAST+ (v2.3.0): the NCBI non-redundant/nucleotide database and Tox-Prot (Jungo and Bairoch

2005) as a custom database search. Signal peptide and transmembrane prediction were obtained using SignalP (4.1 server) (Petersen et al. 2011) and TMHMM (v2.0) (Krogh et al. 2001) respectively. Protein family domain identification was performed using PFAM/HMMER (v3. 1b2) (Finn et al. 2011, 2014) and RNAMMER (server 1. 2) (Lagesen et al. 2007). Molecular function was annotated using the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa et al. 2012) and gene ontology (GO) (Ashburner et al. 2000) databases. All BLAST searches were performed with an *E*-value cut-off of 1×10^{-5} and Trinotate annotation reports run with an *E*-value cut-off of 1×10^{-10}.

Library of Putative Venom-Related Components

The library consists of full-length sequences (i.e. consisting of a start codon, predicted signal peptide and a stop codon) with a Pfam annotation to known venom domains or to Tox-Prot hits. The translated sequences were cross-matched between individuals and proteomic data.

Actinoporin Phylogeny

Data for tree construction included sea anemone actinoporins downloaded from UniProt, the putative actinoporin identified in this study and the *Stichodactyla helianthus* actinoporins StnI and Stn III identified by Rivera-de-Torre et al. (2018). The putative *Oulactis* sp. actinoporin nucleotide sequences have been deposited in EMBL-ENA (European Nucleotide Archive) under PREJEB 278993; Acc. No. LR743519-LR743521. The protein sequences were aligned using M-Coffee (T-Coffee v11.00.d625267 (http://tcoffee.crg.cat/) (Notredame et al. 2000; Wallace et al. 2006; Moretti et al. 2007; Di Tommaso et al. 2011) in a PHYLIP format. The alignment may be found in supplementary data file 8.

The maximum-likelihood tree was constructed using PhyML 3.0 (Guindon et al. 2010) online at http://www. atgc-montpellier.fr/phyml/. Model of best fit was calculated using SMS: automatic model selection v1.8.1 (Lefort et al. 2017) testing both the AIC (Akaike Information) and BIC (Bayesian Information) criteria. For AIC, the model of best fit was WAG +I + G (WAG matrix, proportion of invariable sites allowed and a discrete Gamma model) (Gu et al. 1995; Whelan and Goldman 2001) and for BIC, WAG+I. Starting trees calculated using BioNJ (Gascuel 1997) and the Nearest Neighbour Interchange (NNI) and Subtree-Pruning-Regrafting (SPR) (Hordijk and Gascuel 2005) algorithms. Branch support was calculated using the approximate likelihood ratio test (aLRT SH-likelihood) (Anisimova and Gascuel 2006; Guindon et al. 2010) versus 1000 bootstrap replicates. Log-likelihoods were compared and

the final tree selected with best log-likelihood. Trees were visualized with phylogenetic tree visualization (PRESTO) (http://www.atgc-montpellier.fr/presto/#).

Mass Spectrometry—Tissue Preparation

Owing to tank mortality of individuals 1 and 3 prior to tissue collection for mass spectrometry experiments, tissue could only be collected from individual 2. To compensate for the loss of animals, an additional *Oulactis* sp. specimen was collected. The animal was narcotised prior to being transferred to 70% ethanol for 30 min for preservation. Tentacle tissue (~ 50 mg) was removed, placed in a minimal volume of 70% ethanol and frozen at − 80 °C. Frozen tentacle tissue preparations were crushed with dry ice in a mortar and pestle, washed and dissolved in 20 mL 80% acetonitrile acid (ACN): 0.1% trifluoroacetic acid (TFA) and made up to 30 mL total volume with 80% ACN: 0.1% TFA.

Frozen tissue samples were sonicated for 3×30-s bursts then centrifuged for 25 min at 14,000 rpm at 4 °C, the supernatant was filtered and stored at − 80 °C. Thawed supernatant sample volume was adjusted to 30 mL with MilliQ water. To isolate peptides < 10 kDa, samples were passed through a Vydac® Bioselect SPE C18 Column (218SPE3000 13 μm, 3 mL) and eluted with 80% ACN: 0.1% TFA and the eluate lyophilised.

Mass Spectrometry—Sample Preparation and Data Acquisition

Lyophilised eluates were reconstituted with 15 μL 1 M ammonium bicarbonate and protein content quantified in a Thermo Scientific™ NanoDrop™ Spectrophotometer. Aliquots were prepared with 15 μL 1 M ammonium bicarbonate and a minimum protein content of 50 μg/20 μL of reconstituted sample for trypsin digestion. Samples were prepared for LC-MS/MS analysis using the filter-aided sample preparation (FASP) protocol (Wisniewski et al. 2009), reduced and alkylated with 5 μL 140 mM iodoacetamide, followed by overnight trypsin digestion at 30 °C. Digested samples were dried and resuspended in 0.1% formic acid (FA), prior to clean-up with Omnix-mini Bead 96 C18 ziptips. Samples were dried and reconstituted with 0.1% FA prior to LC-MS/MS of tryptic peptides performed on a Q-Exactive™ Orbitrap (Thermo Scientific™) mass spectrometer to derive mass spectra of individual peptides, and peptide fragment ions were identified with tandem mass spectrometry (MS/MS).

Trypsin digested protein samples were injected onto a Thermo RSLC pepmap100, 75 μm id, 100 Å pore size, 50 cm reversed-phase nano column equilibrated with 95% buffer A (0.1% FA) at a flow rate of 300 nL/min. The peptides were eluted using a linear gradient to 40% buffer

B (80% acetonitrile, 0.1% FA) over 60 min. Peptides were selected for MS/MS analysis using Xcalibur software (Thermo Finnigan) in full MS/dd-MS^2 (TopN) mode. Spectral data files obtained from the Xcalibur software (Thermo Finnigan) were converted to .mgf format using the MSConvert tool. Instrument-specific settings were set at MS tolerance 0.05 Da, MS/MS tolerance 0.1 Da, charge state + 2–+ 5, with biological modification probabilistic features turned on. Peptide identities were mapped according to strict bioinformatic criteria and confidence values assigned to each identified with the use of a decoy database to calculate the false discovery rate (FDR). A FDR cut-off of 1% was applied.

Transcriptome and LC-MS/MS Data Matching

Custom search databases were created from each *Oulactis* sp. transcriptome for matching against LC-MS/MS data. Custom databases were prepared by removing redundant contigs (> 99% similarity) from each assembly prior to translation of transcripts into six ORFS using Mungo (https://github.com/PapenfussLab/Mungo). Converted .mgf files were analysed using ProteinPilot v5 (AB SCIEX) software. ProteinPilot search parameters included thorough identification mode, the Paragon algorithm for protein and post-translation modification identification, trypsin digestion and cysteine alkylation (iodoacetamide) constraints turned on. Peptide identifications were considered a match with coverage confidence of ≥ 95% for the N and C termini and peptide confidence of 99%. Peptide fragments between the N and C termini of the full-length sequence were scrutinized individually if they fell below the > 95% confidence range. A 1% FDR cut-off was applied for the analysis.

Results

Sequencing and Assembly

Illumina sequencing of Total RNA extractions from the tentacles of three individual specimens of *Oulactis* sp. (individuals 1, 2 and 3) yielded 137.6 million read pairs in total, and libraries averaged 45.9 million reads each. Illumina data (raw reads) for individuals have been deposited in ENA (European Nucleotide Archive) under project accession number PRJEB27893. ENA run and sample accessions for individuals are as follows: 1—ERR2710214, ERS2223530; 2—ERR2710215, ERS222353; 3—ERR2710216, ERS2223532.

Data for individual 1 were assembled using six different pipelines to determine the most robust assembly method. Metrics were based on the number of contigs assembled, longest contig, N50, %GC content, reads mapping back to contigs (results for all pipelines found in Supplementary Table S1), and BUSCO assessment (Supplementary Fig. S1). Trinity gave the best assembly metrics overall compared to SOAPdenovo-Trans & Oases, which were previously shown to perform better than Trinity for NGS (high-throughput) short reads (Yang and Smith 2013). The small improvement found running Trinity in a non-normalized mode was not offset by the considerable difference in processing time (days vs weeks). In addition, the normalized assembly possessed a higher number of complete conserved genes. Based on the assembly metrics and BUSCO results for individual 1, the data for individuals 2 and 3 plus the pooled raw read data were assembled via the normalized data Trinity pipeline. Table 1 summarizes the Trinity assembly metrics for individuals and pooled data.

Analysis of the conserved orthologues assembled (Fig. 2) shows that the pooled data had the highest number of missing and fragmented genes, thereby becoming the lowest-quality assembly compared to that of an individual.

To further assess the quality of the pooled data assembly, a transcript of interest from individual 1 (c12169_g1_i1) with similarity to a potassium channel modulator was selected. This contig was used to identify sequence assembly variation across transcriptomes. The translated amino acid sequence found in individual 1 was identical in individual 3 (c6040_g1_i1). A similarly translated sequence was present in individual 2 (c17236_g1_i1) but with a single amino acid residue difference; this contig also had an associated isoform (c17236_g2_i1). Examination of the contigs from individuals 1 and 2 revealed a single-point nucleotide variation which caused a translational change at the penultimate residue, Thr to Ala (Supplementary Fig. S2).

As the translated sequence from individual 2 could not be verified through transcriptomic data cross-matching, we were able to verify via proteomics that the translational change was valid at the penultimate residue and an Ala (Supplementary Fig. S3). From these data, we could surmise an allelic variation between the individuals and not a sequencing error or artefact of assembly. Analysis of the pooled data assembly showed that the translated sequence found in individuals 1 and 3 (with no proteomic verification) was preferentially assembled, and the proteomically verified sequence found in individual 2 (and its associated isoform) had been lost. The loss of a sequence of interest that had been verified as a bioactive peptide (Sunanda et al. 2018) represented a clear disadvantage of the pooled assembly. For this reason, the pooled data assembly was abandoned in favour of individual assemblies.

Salmon quantification files (quant.sf) for each transcriptome (individuals 1, 2 and 3) can be found online in the supplementary data set (files 1, 2 and 3, respectively, https://doi.org/10.5281/zenodo.2561554). Each quant.sf file contains

Table 1 Summary statistics for *Oulactis* sp. individual de novo assemblies and pooled raw reads assembly using the Trinity pipeline (normalized data)

Metric	Individual 1	Individual 2	Individual 3	Pooled Raw Reads
No. of contigs[a]	76,304	79,359	79,299	62,914
Longest contig	26,209	23,850	30,222	14,459
Contig N50	1819	1631	1750	1569
% of reads mapping to contigs	90.4	90.9	91.2	90.6

[a] Based on contigs > 200 nucleotides in length

the transcript identifier, transcript length, transcripts per million (TPM) and the number of reads mapping back to individual transcripts of the assemblies.

Transcriptome Annotation

The complete Trinotate annotation report for each transcriptome may be found in the online supplementary data set (files 4, 5 and 6, https://doi.org/10.5281/zenodo.2561554). Table 2 summarizes the percentage of transcripts (predicted proteins) annotated for each *Oulactis* sp. individual, derived from the annotation reports. The percentage of unannotated transcripts refers to those with no known similarity to any genes in the databases searched, i.e. NCBI or biological functions (e.g. GO ontology). On average, half of each individual's transcripts were unannotated, i.e. transcripts did not bear any resemblance to known genes in the databases, indicating how much there is yet to discover of this species' genes. Approximately 1% of each individual's transcripts showed similarity to known venom-related components in Tox-Prot.

Venom-Related Component Library

The library (Supplementary data file 7) contains 398 sequences with similarity to known venom components, identified from the transcriptomes of three individuals. All sequences within the library are novel, i.e. they do not share 100% identity to any other known venom components. Only 15 identical translated sequences (3.8% of the library) occurred across all three individuals, and a further 16 were common to two individuals. The remaining translated sequences were unique to each transcriptome (115, 115 and 141 sequences, respectively), illustrating a high level of variability among individuals of the species.

The graphical summary of the venom-related component library (Fig. 3) shows the proportion of sequences contained in the library and their predicted biological function, based on sequence similarity. Owing to the diverse range of functions and processes documented for ShK-like and CAP (cysteine-rich secretory protein family) domains, sequences retain their respective domain names and are not assigned to a specific function or process group (e.g. ion channel modulation).

The *Oulactis* sp. library contains protease and peptidases (including metallopeptidases), peptidase inhibitors, ion channel modulators (neurotoxins and defensins), antimicrobial peptides, phospholipases, and pore-forming toxins. These findings are consistent with other cnidarian studies (Li et al. 2014; Ponce et al. 2016; Madio et al. 2017; Ramírez-Carreto et al. 2019). Transcripts with an ShK-like domain dominate the library, closely followed by carbohydrate-binding domains including EGF-like and C-type lectins. Transcripts annotated as 'predicted toxin' and 'toxin' have similarity to sequences found in Tox-Prot, although these 'predicted' toxins in Tox-Prot lack functional characterization (Jungo and Bairoch 2005).

Of the 398 sequences in the venom-related library, eight sequences were proteomically matched; this low number could be a reflection of several factors. Firstly, many of the sequences in the library are rich in Arg and Lys so using the standard trypsin digestion meant that some peptides were unmatched owing to the low molecular weight of tryptic fragments produced and the lack of proteotypes. Secondly, tissue samples were subjected to size exclusion (M_r > 10 kDa) meaning that larger proteins were no longer present in the samples for mass spectrometry analysis, and larger proteins formed the greater proportion of the venom-related library. Lastly, the unanticipated variability of an individuals' transcriptome may have contributed to a mismatch between the custom transcriptome database and mass spectrometry data searches, resulting in fewer full-length sequence matches. For example, searching the mass spectrometry data of individual 2 against the custom transcriptome databases for individual 1 or 3 may result in fewer full-length sequence matches due to an individual's venom expression profile, i.e. allelic variation in genes.

Oulactis Actinoporins

One identical full-length translated sequence with similarity to the protein family Anemone_cytotox (PF06369.9) occurred in all three *Oulactis* sp. transcriptomes. Figure 4 shows the alignment of the predicted mature peptide of three actinoporin sequences (the characterized cytolysins of *O. orientalis* and putative cytolysin of *Oulactis* sp.). Alignments show the putative actinoporin to be more similar to Or-G (UniProt Q5I2B1) (77.2% similarity) than to Or-A (UniProt Q5I4B8) (72.8% similarity) (pairwise alignments of individual sequences are shown in Supplementary Fig. S4).
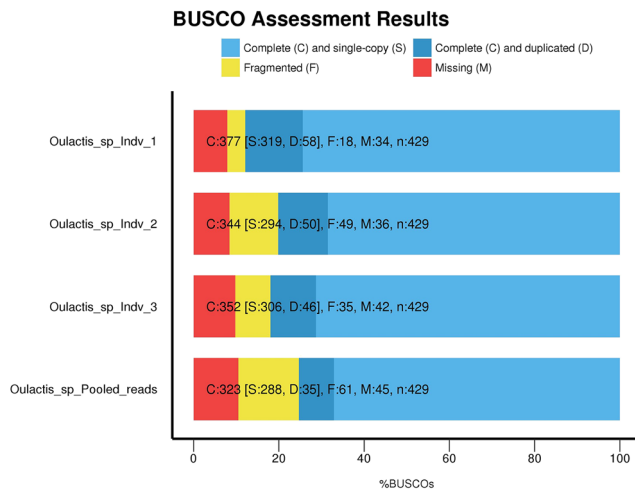
Fig. 2 BUSCO graphical summary for the Trinity-assembled transcriptomes (normalized data) for individuals 1, 2 and 3 of *Oulactis* sp. and pooled raw reads. $n = 429$ is the number of conserved eukaryote orthologues in the search database

According to the top five NCBI BLAST and Tox-Prot hits, the *Oulactis* sp. putative actinoporin has a higher sequence similarity to other sea anemone genera than to *O. orientalis*. These species include *Anthopleura asiatica* (C5NSL2), *Actinia equina* (Q93109, Q9Y1U9, P61914) and *A. tenebrosa* (P61915). Supplementary Table S2 contains the complete results of sequence BLAST hits found. Actinoporins with a higher % sequence identity to the *Oulactis* sp. sequence occurred within the same taxonomic family (Actiniidae) as *Oulactis* sp.

Actinoporins have been isolated from four other sea anemone families in addition to the Actiniidae: Stichodactylidae, Thalassianthidae, Aliciidae and Sagartiidae. We compared the actinoporins of these families to those from the Actiniidae family to ascertain sequence variation between families and the putative actinoporin of *Oulactis* sp. The logo plot (Supplementary Fig. S5) illustrates the 19 key conserved residues of actinoporin sequences and the highly variable N-termini. Supplementary data file 8 contains the alignment of the actinoporin protein sequences used to produce the maximum-likelihood tree. The maximum-likelihood tree (Fig. 5) (best model generated using AIC/SPR and 1000 bootstraps, LogLK = − 3116) shows that the actinoporins cluster according to their taxonomic family, except for the *O. orientalis* actinoporins (Or-A and Or-G), which appear in a separate clade to the Actiniidae actinoporins and more similar to those in Stichodactylidae.

Venom-Related Sequence Domain Architecture

Sequences in the venom-related component library fall into three distinct domain architecture types: single, repeat (i.e. containing only the one protein domain repeated) or multi-domain proteins. The protein families that occur with more

than one domain structure are summarized in Fig. 6. All annotated protein families, their putative biological function, and associated domain architecture (single, repeat or multi) are summarized in Supplementary Table S3.

Oulactis sp. single-domain sequences have similarity to characterized cnidarian toxins such as the defensins, pore-forming toxins, phospholipases and antimicrobial peptides. Nearly half of the library (47%) consists of single-domain peptides or proteins. Repeat domain sequences comprise a small proportion of the library (approx. 10%) and occurred across seven protein families including Kazal-type 1 and 2 (serine protease inhibitors), ShK-like, Kunitz_BPTI and four metallopeptidase families (M13, M14, M15 and M 16). The multi-domain proteins comprise ∼ 42% of the sequences in the library. At the other end of the spectrum, protein families such as ShK-like (PF01549) and Kunitz/BPTI (PF00014) occur in all three architectures: single, repeat and multi-domain.

Of all the protein families, the most common and widely distributed is ShK-like, occurring in equal numbers across the three domain groupings (Fig. 6). Lectin_C (PF00059) is the second most commonly found protein domain but rarely occurred as a repeated motif. The Kazal-type 1 and 2 (serine protease inhibitor families PF00050, PF07648) never occurred as a single-domain, but only in a combination repeat, e.g. Kazal-type 1 with Kazal-type (Fig. 7a) or in a multi-domain protein (Fig. 7b).

ShK-like domains appear in up to nine repeats, but in multi-domain sequences can appear up to four times in combination with peptidases such as CAP (cysteine-rich secretory protein family PF00188.23) and Kunitz_BPTI. Proteases and peptidases only occur as multi-domains, except peptidase_M28 (PF04389), which occurred mostly as a single-domain. Exemplified in Fig. 7 are representative multi-domain structures found in *Oulactis* sp. (amino acid sequences for domain examples may be found in Supplementary data file 9).

Discussion

This study has produced the first transcriptomes of an endemic Australian species, the speckled anemone (*Oulactis* sp.) and the first for the genus *Oulactis*. From these data, we compiled a venom-related component library of 398 putative venom-related peptides and proteins, including one putative actinoporin. Venom-related peptides and proteins in the library appear in a variety of domain architectures including single, repeat and multi-domain co-occurrences.

As this is the first transcriptome produced for this genus, we attempted to produce one overarching assembly to represent the *Oulactis* sp. tentacles. However, pooling the raw reads resulted not only in a statistically inferior assembly, but also the loss of venom-related sequences. Trinity assemblies are known to hide transcripts of multicopy genes (Macrander

Table 2 Annotation summary for transcripts of *Oulactis* sp. individuals from reports (*E*-value 1×10^{-10}) generated using the Trinotate suite

	Individual 1	Individual 2	Individual 3
Total number of transcripts	65,535	86,948	87,704
Number of unannotated transcripts	27,400 (41.8%)	50,090 (57.6%)	49,658 (56.6%)
Annotated transcripts			
GO terms Blast	18,171 (27.7%)	22,419 (25.8%)	23,071 (26.3%)
GO terms Pfam	14,079 (21.5%)	17,381 (20.0%)	18,052 (20.6%)
Pfam	11,009 (16.8%)	10,808 (12.4%)	10,938 (12.5%)
KEGG	15,938 (24.3%)	18,946 (21.8%)	19,559 (22.3%)
EGGNOG	15,670 (23.9%)	19,436 (22.4%)	20,018 (22.8%)
signalP	3592 (5.5%)	3357 (3.9%)	3802 (4.3%)
TmHMM	11,030 (16.8%)	9358 (10.8%)	9699 (11.1%)
RNAmmer	4 (0.006%)	6 (0.006%)	10 (0.01%)
BLAST hits			
Blastp	11,039 (16.8%)	10,776 (12.4%)	10,975 (13.7%)
Blastx	15,101 (23.0%)	11,833 (13.6%)	11,898 (13.6%)
Tox-Prot (Blastp)	824 (1.3%)	1000 (1.2%)	1097 (1.3%)

et al. 2015), and it appears that the pooling of data may compound this problem even further with the loss of isoforms of potential biomedical interest.

As demonstrated in this study, proteomic validation of sequences is invaluable in genomic surveys of bioactive proteins where transcriptome data may exist for only one individual. Single-residue variations and discrepancies can be problematic when identifying peptides for characterization. A single change in a key residue of a venom peptide may affect its bioactivity by altering function, binding affinity or resulting in the loss of activity (Won et al. 2011; Peigneur et al. 2012; Wang et al. 2016; Pennington et al. 2017). Using individual assemblies instead of pooled data is a means to identify potential sequencing errors or highlight allelic variations in venom genes.

The design of the mass spectrometry experiments is also key in the discovery of peptides and proteins of sea anemones. The limited number of sequences matched proteomically to the transcriptomes in this study is a reflection of several factors, including tissue sample preparation, size exclusion, digestion enzyme selection and the unforeseen variability of venom expression between individuals. The use of a whole tissue crush sample preparation was the preferred protocol as cnidae are embedded in the tissue, and bioactive molecules are also contained within the mucus (Stabili et al. 2015; Sintsova et al. 2018) and gland cells (Moran et al. 2012) of sea anemones. The drawback of the tissue crush method is that all cells are analysed, potentially masking the small quantities of venom present and making it impossible to determine which novel proteins and peptides originate exclusively from the venom.

Future venomic experimental design for sea anemones should consider using a combination of tissue sample preparations, such as electrical stimulation (Malpezzi et al. 1993) or cnidae isolation (Bloom et al. 1998), in conjunction with whole tissue analysis to garner a comprehensive venom inventory. Using alternative enzymes (i.e. Asp or GluC) should cleave peptides at alternate sites, resulting in larger, proteotypic peptide fragments for matching. In addition, utilizing venom-only isolation methods, as exemplified by Madio et al. (2017) for *Stichodactyla haddoni* venom, may reveal additional novel toxins and toxin scaffolds in *Oulactis* sp.

The 398 venom-related transcripts identified in this study represent just 10% of the translational genes (i.e. those with a signal peptide) for this species, from one tissue type. The Tox-Prot database holds over 200 toxin peptides and proteins from sea anemones (Prentis et al. 2018), and the library created here would almost double that.

Strikingly, the *Oulactis* sp. venom-related library contains all novel sequences, with none showing 100% identity to any known sequences. Moreover, each individual transcriptome had over 100 unique putative venom-related peptides and proteins. The number of unique venom-related sequences in the library highlights the variability of venom gene expression among individuals of *Oulactis sp.*

The results here indicate that venomic study design should consider using several biological samples, three at a minimum, in order to capture the chemical diversity among individuals of sea anemone species. Venom chemical diversity has been documented among individuals in snakes (Menezes et al. 2006; Gibbs et al. 2009) and now in sea anemones, as seen here. Venom composition in sea anemones varies not only between individuals but also throughout life cycle stages as observed in *Nematostella vectensis* (Columbus-Shenkar et al. 2018). Chemical diversity correlating with the biological use of the venom (e.g. defence vs. feeding) has also been observed (Dutertre et al. 2014; Macrander et al. 2015). In addition, studies have documented venom diversity across tissue-

Fig. 3 The total number of translated sequences found for each putative biological function in the venom-related component library, compiled from the three *Oulactis* sp. transcriptomes

specific regions of sea anemones (Macrander et al. 2016; Surm et al. 2019). Experimental design should take into account that venom expression is also variable according to the treatment of cnidarians prior to tissue sampling (Schmidt 1982; Madio et al. 2017).

Among the venom-related components, ShK-like domains were by far the most prevalent in the transcriptome, representing a quarter of the library. The dominance of this domain is consistent with other transcriptome studies of cnidarian venom, where they form a substantial percentage of the toxin-like components (Li et al. 2014; Madio et al. 2017). Contrary to the abundance of ShK-like sequences, only one translated actinoporin sequence was annotated in the tentacles of *Oulactis* sp. and was seemingly conserved within the species, but not within the genus or family. Examination of other species in the genus will determine the extent of this conserved gene.

The lack of similarity between the *Oulactis* actinoporins was not wholly unexpected, as the taxonomic placement of *O. orientalis* in the genus requires revision (Sanamyan and Sanamyan 2009; Mitchell 2010). Moreover, the findings here are consistent with those of Macrander and Daly (2016) in that *O. orientalis* actinoporins are less like other Actiniidae

species. These results highlight the need for correct taxonomic identification and designation of species into genera. Without the correct taxonomic assignment, it is difficult to make inferences regarding venom evolution or to identify key residues that may be required for activity in closely related species.

Two peptides in the venom component library, U-AITX-Oulsp1 and U-AITX-Oulsp2 (Sunanda et al. 2018; Krishnarjuna et al. 2018b), were predicted to be potassium channel blockers based on similarity to ShK/BgK sequence and structures (Castañeda et al. 1995; Tudor et al. 1996). Oulsp1 (Sunanda et al. 2018) showed activity against $K_V1.2$ and $K_V1.6$ channels, while Oulsp2 (Krishnarjuna et al. 2018b) showed no activity against the mammalian K_V channels tested, even though both peptides share high sequence and structure similarity with ShK and BgK; these peptides may be active on invertebrate channels but this is yet to be tested.

The results from the structure and function studies of Oulsp1 and 2, as well as AsK132958 from *Anemonia sulcata* (Krishnarjuna et al. 2018a), emphasize that sequence and structural similarity are not necessarily a predictor of function and no longer form a reliable criterion for progressing novel candidates through to functional studies. This, coupled with the exponential growth in databases of functionally

Fig. 4 Alignment of the predicted *Oulactis* sp. actinoporin mature peptide amino acid sequence against characterized cytolysins found in *Oulactis orientalis* Or-A (UniProt Q5I4B8) and Or-G (UniProt Q5I2B1) (Il'ina et al. 2005)

```
Oulactis_sp.    SPLAVAGAVIEGGNLAMSVLEKVLKAIGDVSRKIAIGVDNESGKPWTAINTYFRSGTSDI   60
OR-G|Q5I2B1|    ------GAIIAGAALGFNVHQTVLKALGQVSRKIAIGVDNESGGTWTALNAYFRSGTTDV   54
OR-A|Q5I4B8|    --------------ATFRVLAKVLAELGKVSRKIAVGVDNESGGSWTALNAYFRSGTTDV   46
                      :  *   .**  :*.******:******* ****:*.:******:*:

Oulactis_sp.    ILPHEVKSGKALLYNSQKNRGPVATGVVGVLTYAMSDGNTLAVLFSVPYDYNLYSNWWNV  120
OR-G|Q5I2B1|    ILPEFVPNQKALLYSGQKDTGPVATGAVGVLAYYMSDGNTLGVMFSVPFDYNLYSNWWDV  114
OR-A|Q5I4B8|    ILPDLVPNQKALLYRGGKDTGPVATGVVGVLAYAMSDGNTLAILFSVPYDYNLYSNWWNV  106
                ***. * .  ***** . *: ****** .****:* ******* .::*****:*********.*

Oulactis_sp.    RVYSGKQRATQYMYEKLYYDYSPFKGDNGWHSRNLGYGLKCRGFMNSSGKAVLEIQVSQA  180
OR-G|Q5I2B1|    KVYRGRRRADQAMYEGLLYG-IPYGGDNGWHARKLGYGLKGRGFMKSSAQSILEIHVTKA  173
OR-A|Q5I4B8|    KVYSGKRRADQGMSEDLSYG-NPYGGDNGWHARKLAYGLKERGFMKSSAQSILEIHATKA  165
                :** *::** * * * * *  *: ******:*:*.**** ****:**::::.***:..:*
```

Fig. 5 Maximum-likelihood tree showing taxonomic family groupings of sea anemone actinoporins and the relevative placement of *O. orientalis* (Q514B8 & Q512B1) and the putative *Oulactis* sp. (Oul_sp) identified in this study. Rootpath highlighted in red for *Oulactis* spp. actinoporins. Sequences aligned in M-Coffee (Supplementary data file 8) and tree constructed in PhyML (v. 3.0). Bootstrapping values shown at the nodes. Tip labels include UniProt Acc. No. and genus abbreviation, except StnI & StnIII (*Stichodactyla helianthus*) labelled as sourced from Rivera-de-Torre et al. (2018)
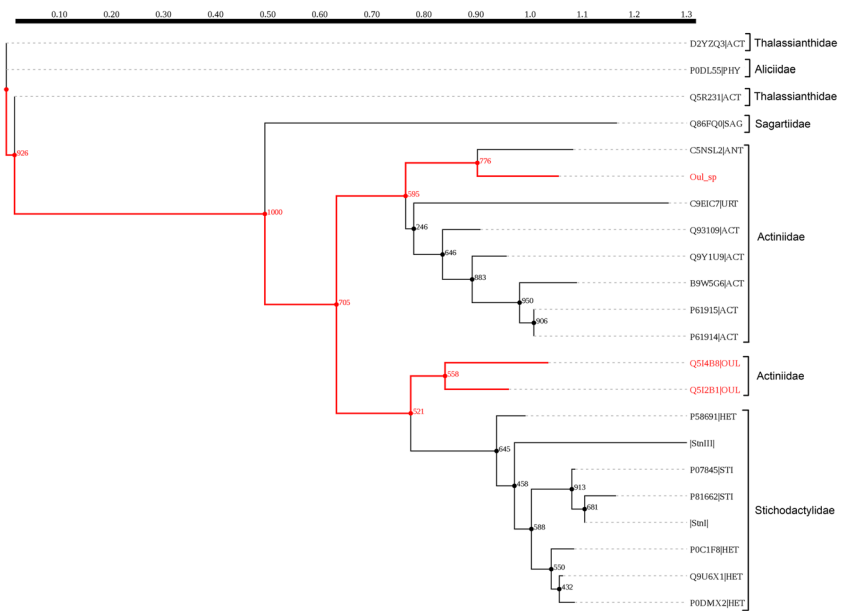
uncharacterised sequences, indicates the need for caution when selecting sequences for functional studies based on annotation alone.

It also highlights the importance of characterizing novel sequences and conducting venom evolution studies to understand the mechanisms that underlie potent toxins, such as ShK, for potential biomedical applications. Evolution studies using more complex bioinformatic analyses may help guide the selection of sequences on which to conduct functional studies (Shafee et al. 2019; Mitchell et al. 2019).

It is possible that some of the multi-domain sequences containing the ShK-like domain may not be toxins or venom-related genes. The ShK-like domain is widespread throughout the natural world and is documented to have various functions other than as a toxin (Prentis et al. 2018; Shafee et al. 2019). It
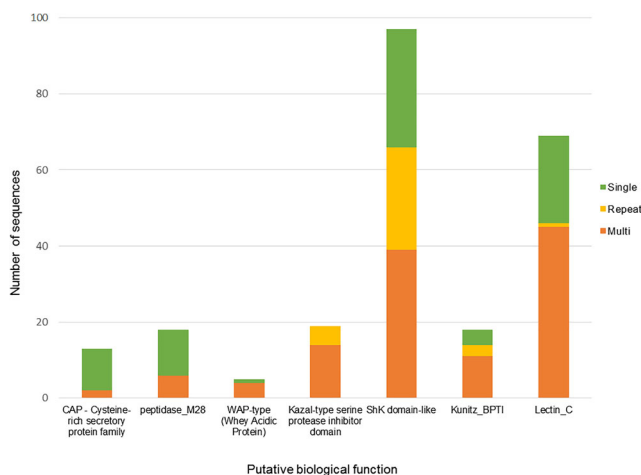
Fig. 6 Protein families and their domain structure, illustrating the proportion of sequences and venom families that occur in single, repeat (i.e. composed of the one domain repeated) or multi-domain architectures in the *Oulactis* sp. venom-related component library

remains to be determined what the functions of venom and non-venom component multi-domain proteins are, and to assess their potential value as pharmacological tools. These results highlight the vast potential for the discovery of novel venom-related sequences from this, and the over 1000 unexamined, sea anemone species. A major challenge going
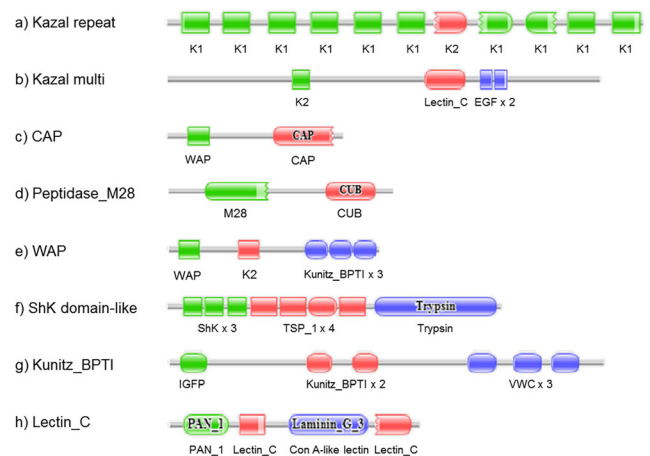
Fig. 7 Examples of multi-domain architectures with similarity to venom-related components in *Oulactis* sp. tentacle transcriptomes. The grey bar represents the base sequence. Full-length domain matches represented by curved ends and squares represent a fragment match. Domain architecture examples illustrated using the Pfam domain graphic generator code (https://pfam.xfam.org/generate_graphic). **a** Kazal serine protease inhibitor repeat. **b** Kazal serine protease inhibitor. **c** CAP. **d** Peptidase_ M28. **e** WAP. **f** ShK-like. **g** Kunitz_BPTI. **h** Lectin_C. Abbreviations: K1 Kazal_1 serine protease inhibitor domain; K2 Kazal_2 serine protease inhibitor domain; WAP whey acidic protein; CAP cysteine-rich secretory protein family; CUB CUB domain; TSP thrombospondin type 1 domain; Kuntiz_BPTI Kunitz/Bovine pancreatic trypsin inhibitor; IGFP insulin-like growth factor binding protein; VWC von Willebrand factor type C domain; PAN_1 PAN domain; Con A-like lectin (Laminin_ G_3) Concanavalin A-like lectin/glucanase (Supplementary data file 9: sequences of domain occurrence examples)

forward is the judicious selection of sequences to progress through to structural and functional studies.

Author Contributions All authors contributed to the study conception and design. Material collection was performed by M.L.M. and R.A.V.M. and material preparation for transcriptome studies by M.L.M. Material preparation for mass spectrometry experiments was performed by M.L.M. and R.A.V.M.. Data collection was performed by M.L.M. and G.Q.T. and data analysis by M.L.M. All authors read and approved the final manuscript.

Compliance with Ethical Standards

All animal studies were carried out in accordance with the Victorian Fisheries Authority collection permit and Monash University Animal Ethics Regulations, Guidelines and Codes.

Conflict of Interest The authors declare that they have no conflict of interest.

References

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

Anderluh G, Maček P (2002) Cytolytic peptide and protein toxins from sea anemones (Anthozoa: Actiniaria). Toxicon 40:111–124

Andreosso A, Smout MJ, Seymour JE (2014) Dose and time dependence of box jellyfish antivenom. J Venomous Anim Toxins Incl Trop Dis 20:1–5

Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol 55: 539–552

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. Nat Genet 25:25–29

Beress L (1982) Biologically active compounds from coelenterates. Pure Appl Chem 54:1981–1994

Bloom DA, Burnett JW, Alderslade P (1998) Partial purification of box jellyfish (*Chironex fleckeri*) nematocyst venom isolated at the beachside. Toxicon 36:1075–1085

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120

Bruskiewich R (1999) geecee tool. http://www.bioinformatics.nl/cgi-bin/emboss/geecee. Accessed 21 Aug 2017

Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D, Lee TJ, Leigh ND, Kuo TH, Davis FG, Bateman J, Bryant S, Guzikowski AR, Tsai SL, Coyne S, Ye WW, Freeman RM Jr, Peshkin L, Tabin CJ, Regev A, Haas BJ, Whited JL (2017) A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. Cell Rep 18:762–776

Castañeda O, Sotolongo V, Amor AM, Stöcklin R, Anderson AJ, Harvey AL, Engström A, Wernstedt C, Karlsson E (1995) Characterization of a potassium channel toxin from the Caribbean Sea anemone *Stichodactyla helianthus*. Toxicon 33:603–613

Chandy KG, Norton RS (2017) Peptide blockers of $K_V1.3$ channels in T cells as therapeutics for autoimmune disease. Curr Opin Chem Biol 38:97–107

Coleman N (1999) Dangerous sea creatures - aquatic survival guide. Neville Coleman's underwater. Geographic Pty Ltd, Springwood

Columbus-Shenkar YY, Sachkova MY, Macrander J et al (2018) Dynamics of venom composition across a complex life cycle. Elife 7:e35014

Cristofori-Armstrong B, Rash LD (2017) Acid-sensing ion channel (ASIC) structure and function: insights from spider, snake and sea anemone venoms. Neuropharmacology 127:173–184

Di Tommaso P, Moretti S, Xenarios I et al (2011) T-coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res 39:W13–W17

Diochot S, Schweitz H, Béress L, Lazdunski M (1998) Sea anemone peptides with a specific blocking activity against the fast inactivating potassium channel $K_V3.4$. J Biol Chem 273:6744–6749

Dutertre S, Jin A-H, Vetter I et al (2014) Evolution of separate predation- and defence-evoked venoms in carnivorous cone snails. Nat Commun 5:3521

Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 39:W29–W37

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M (2014) Pfam: the protein families database. Nucleic Acids Res 42:D222–D230

Frances L (1973) Intraspecific aggression and its effect on the distribution of *Anthopleura elegantissima* and some related sea anemones. Biol Bull 144:73–92

Garcia PJ, Schein RMH, Burnett JW (1994) Fulminant hepatic failure from a sea anemone sting. Ann Intern Med 120:665–666

Gascuel O (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol 14:685–695

Gibbs HL, Sanz L, Calvete JJ (2009) Snake population venomics: proteomics-based analyses of individual variation reveals significant gene regulation effects on venom protein expression in *Sistrurus* rattlesnakes. J Mol Evol 68:113–125

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652

Gu X, Fu YX, Li WH (1995) Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. Mol Biol Evol 12:546–557

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321

Honma T, Shiomi K (2006) Peptide toxins in sea anemones: structural and functional aspects. Mar Biotechnol (NY) 8:1–10

Hordijk W, Gascuel O (2005) Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood. Bioinformatics 21:4338–4347

Il'ina AP, Monastyrnaya MM, Sokotun IN et al (2005) Actinoporins from the sea of Japan anemone *Oulactis orientalis*: isolation and partial characterization. Russ J Bioorg Chem 31:34–42

Jungo F, Bairoch A (2005) Tox-Prot, the toxin protein annotation program of the Swiss-Prot protein knowledgebase. Toxicon 45:293–301

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40:D109–D114

Kass-Simon G, Scappaticci AA Jr (2002) The behavioral and development physiology of nematocysts. Can J Zool 80:1772–1794

Krishnarjuna B, MacRaild CA, Sunanda P, Morales RAV, Peigneur S, Macrander J, Yu HH, Daly M, Raghothama S, Dhawan V, Chauhan S, Tytgat J, Pennington MW, Norton RS (2018a) Structure, folding and stability of a minimal homologue from *Anemonia sulcata* of the sea anemone potassium channel blocker ShK. Peptides 99:169–178

Krishnarjuna B, Villegas-Moreno J, Mitchell ML, Csoti A, Peigneur S, Amero C, Pennington MW, Tytgat J, Panyi G, Norton RS (2018b) Synthesis, folding, structure and activity of a predicted peptide from the sea anemone *Oulactis* sp. with an ShKT fold. Toxicon 150:50–59

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580

Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108

Lefort V, Longueville J-E, Gascuel O (2017) SMS: smart model selection in PhyML. Mol Biol Evol 34:2422–2424

Lenarcic B, Ritonja A, Strukelj B, Turk B, Turk V (1997) Equistatin, a new inhibitor of cysteine proteinases from *Actinia equina*, is structurally related to thyroglobulin type-1 domain. J Biol Chem 272:13899–13903

Li R, Yu H, Xue W, Yue Y, Liu S, Xing R, Li P (2014) Jellyfish venomics and venom gland transcriptomics analysis of *Stomolophus meleagris* to reveal the toxins associated with sting. J Proteome 106:17–29

Logashina YA, Solstad RG, Mineev KS et al (2017) New disulfide-stabilized fold provides sea anemone peptide to exhibit both antimicrobial and TRPA1 potentiating properties. Toxins (Basel) 9:1–23

Macrander J, Daly M (2016) Evolution of the cytolytic pore-forming proteins (Actinoporins) in sea anemones. Toxins (Basel) 8:368

Macrander J, Broe M, Daly M (2015) 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 8:2358–2375

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

Malpezzi ELA, de Freitas J, Muramoto K, Kamiya H (1993) Characterization of peptides in sea anemone venom collected by a novel procedure. Toxicon 31:853–864

Menezes MC, Furtado MF, Travaglia-Cardoso SR, Camargo AC, Serrano SM (2006) Sex-based individual variation of snake venom proteome among eighteen *Bothrops jararaca* siblings. Toxicon 47:304–312

Mitchell ML (2010) Actiniaria (Cnidaria:Anthozoa) of Port Phillip Bay, Victoria: including a taxonomic case study of *Oulactis muscosa* and *Oulactis mcmurrichi*. Southern Cross University

Mitchell ML, Shafee TMA, Papenfuss AT, Norton RS (2019) Evolution of cnidarian *trans*-defensins: sequence, structure and exploration of chemical space. Proteins Struct Funct Bioinforma Prot 25679:1–10

Monastyrnaya M, Peigneur S, Zelepuga E et al (2016) Kunitz-type peptide HCRG21 from the sea anemone *Heteractis crispa* is a full antagonist of the TRPV1 receptor. Mar Drugs 14:229

Moran Y, Genikhovich G, Gordon D et al (2012) Neurotoxin localization to ectodermal gland cells uncovers an alternative mechanism of venom delivery in sea anemones. Proc R Soc B Biol Sci 279:1351–1358

Moretti S, Armougom F, Wallace IM, Higgins DG, Jongeneel CV, Notredame C (2007) The M-coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. Nucleic Acids Res 35:W645–W648

Nagata K, Hide M, Tanaka T, Ishii K, Izawa M, Sairenji T, Tomita K, Shimizu E (2006) Anaphylactic shock caused by exposure to sea anemones. Allergol Int 55:181–184

Notredame C, Higgins DG, Heringa J (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 302:205–217

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14:417–419

Peigneur S, Béress L, Möller C, Marí F, Forssmann WG, Tytgat J (2012) A natural point mutation changes both target selectivity and mechanism of action of sea anemone toxins. FASEB J 26:5141–5151

Pennington MW, Czerwinski A, Norton RS (2017) Peptide therapeutics from venom: current status and potential. Bioorg Med Chem 26:2738–2758

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786

Ponce D, Brinkman D, Potriquet J, Mulvenna J (2016) Tentacle transcriptome and venom proteome of the Pacific Sea nettle, *Chrysaora fuscescens* (Cnidaria: Scyphozoa). Toxins (Basel) 8:102

Prentis PJ, Pavasovic A, Norton RS (2018) Sea anemones: quiet achievers in the field of peptide toxins. Toxins (Basel) 10:36

Ramírez-Carreto S, Vera-Estrella R, Portillo-Bobadilla T et al (2019) Transcriptomic and proteomic analysis of the tentacles and mucus of *Anthopleura dowii* Verrill, 1869. Mar Drugs 17:436–459

Rivera-de-Torre E, Martínez-del-Pozo Á, Garb JE (2018) *Stichodactyla helianthus'* de novo transcriptome assembly: discovery of a new actinoporin isoform. Toxicon 150:105–114

Rodríguez AA, Salceda E, Garateix AG, Zaharenko AJ, Peigneur S, López O, Pons T, Richardson M, Díaz M, Hernández Y, Ständker L, Tytgat J, Soto E (2014) A novel sea anemone peptide that inhibits acid-sensing ion channels. Peptides 53:3–12

Romero L, Marcussi S, Marchi-Salvador DP, Silva FP Jr, Fuly AL, Stábeli RG, da Silva SL, González J, Monte AD, Soares AM (2010) Enzymatic and structural characterization of a basic phospholipase A2 from the sea anemone *Condylactis gigantea*. Biochimie 92:1063–1071

Sanamyan NP, Sanamyan KE (2009) Shallow water anemones (Cnidaria: Actiniaria) from south-eastern coast of Kamchatka. Invertebr Zool 5:155–172

Schmidt GH (1982) Replacement of discharged cnidae in the tentacles of *Anemonia sulcata*. J Mar Biol Assoc U K 62:685

Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086–1092

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

Shick JM (1991) A functional biology of sea anemones. Chapman & Hall, London

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212

Sintsova O, Gladkikh I, Chausova V, Monastyrnaya M, Anastyuk S, Chernikov O, Yurchenko E, Aminin D, Isaeva M, Leychenko E,

Kozlovskaya E (2018) Peptide fingerprinting of the sea anemone *Heteractis magnifica* mucus revealed neurotoxins, Kunitz-type proteinase inhibitors and a new β-defensin α-amylase inhibitor. J Proteome 173:12–21

Stabili L, Schirosi R, Parisi M, Piraino S, Cammarata M (2015) The mucus of *Actinia equina* (Anthozoa, Cnidaria): an unexplored resource for potential applicative purposes. Mar Drugs 13:5276–5296

Stanke M, Steinkamp R, Waack S, Morgenstern B (2004) AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res 32: W309–W312

Štrukelj B, Lenarčič B, Gruden K et al (2000) Equistatin, a protease inhibitor from the sea anemone *Actinia equina*, is composed of three structural and functional domains. Biochem Biophys Res Commun 269:732–736

Sunanda P, Krishnarjuna B, Peigneur S et al (2018) Identification, chemical synthesis, structure, and function of a new K_V1 channel blocking peptide from *Oulactis* sp. Pept Sci 110:1–10

Suput D (2011) Interactions of cnidarian toxins with the immune system. Inflamm Allergy Drug Targets 10:429–437

Surm JM, Smith HL, Madio B, Undheim EAB, King GF, Hamilton BR, van der Burg C, Pavasovic A, Prentis PJ (2019) A process of convergent amplification and tissue-specific expression dominates the evolution of toxin and toxin-like genes in sea anemones. Mol Ecol 28:2272–2289

Tarcha EJ, Olsen CM, Probst P et al (2017) Safety and pharmacodynamics of dalazatide, a $K_V1.3$ channel inhibitor, in the treatment of plaque psoriasis: a randomized phase 1b trial. PLoS One 12:1–19

Tudor JE, Pallaghy PK, Pennington MW, Norton RS (1996) Solution structure of ShK toxin, a novel potassium channel inhibitor from a sea anemone. Nat Struct Biol 3:317–320

Wallace IM, O'Sullivan O, Higgins DG, Notredame C (2006) M-coffee: combining multiple sequence alignment methods with T-coffee. Nucleic Acids Res 34:1692–1699

Wang X, Gao B, Zhu S (2016) A single-point mutation enhances dual functionality of a scorpion toxin. Comp Biochem Physiol C Toxicol Pharmacol 179:72–78

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18:691–699

Williamson JA, Fenner PJ, Burnett JW, Rifkin JF (eds) (1996) Venomous and poisonous marine animals: a medical and biological handbook. University of New South Wales Press, Sydney

Wisniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation method for proteome analysis. Nat Methods 6:359

Won A, Pripotnev S, Ruscito A, Ianoul A (2011) Effect of point mutations on the secondary structure and membrane interaction of antimicrobial peptide anoplin. J Phys Chem B 115:2371–2379

Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, Zhou X, Lam TW, Li Y, Xu X, Wong GK, Wang J (2014) SOAPdenovo-trans: *de novo* transcriptome assembly with short RNA-Seq reads. Bioinformatics 30:1660–1666

Yang Y, Smith SA (2013) Optimizing *de novo* assembly of short-read RNA-seq data for phylogenomics. BMC Genomics 14:328

Zerbino DR, Birney E (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res 18:821–829