The Transcriptome of the Zoanthid *Protopalythoa variabilis* (Cnidaria, Anthozoa) Predicts a Basal Repertoire of Toxin-like and Venom-Auxiliary Polypeptides

Chen Huang[1,†], Jean-Étienne RL Morlighem[2,3,†], Hefeng Zhou[1], Érica P Lima[4], Paula B Gomes[5], Jing Cai[1], Inchio Lou[6], Carlos D Pérez[4,*], Simon Ming Lee[1,*], and Gandhi Rádis-Baptista[3,*]

[1]State Key Laboratory of Quality Research in Chinese Medicine and Institute of Chinese Medical Sciences, University of Macau, Macau, China

[2]Northeast Biotechnology Network (RENORBIO), Post-graduation program in Biotechnology, Federal University of Ceará, Fortaleza, Brazil

[3]Laboratory of Biochemistry and Biotechnology, Institute for Marine Sciences, Federal University of Ceará, Fortaleza, Brazil

[4]Centro Acadêmico de Vitoria, Universidade Federal de Pernambuco, Vitória de Santo Antão, Brazil

[5]Departamento de Biologia, Universidade Federal Rural de Pernambuco, Recife, Brazil

[6]Faculty of Science and Technology, Department of Civil and Environmental Engineering, University of Macau, Macau, China

*Corresponding author: E-mail: gandhi.radis@ufc.br; simonlee@umac.mo; cdperez@ufpe.br.
[†]These authors contributed equally to this work.

Accepted: August 20, 2016

Data deposition: This Transcriptome Shotgun Assembly (TSA) project was deposited at DDBJ/EMBL/GenBank under the accession GCVI00000000, associated with the BioProject PRJNA279783 and biosample SAMN03450566. The version described in this paper is the first version, GCVI01000000.

Abstract

Protopalythoa is a zoanthid that, together with thousands of predominantly marine species, such as hydra, jellyfish, and sea anemones, composes the oldest eumetazoan phylum, i.e., the Cnidaria. Some of these species, such as sea wasps and sea anemones, are highly venomous organisms that can produce deadly toxins for preying, for defense or for territorial disputes. Despite the fact that hundreds of organic and polypeptide toxins have been characterized from sea anemones and jellyfish, practically nothing is known about the toxin repertoire in zoanthids. Here, based on a transcriptome analysis of the zoanthid *Protopalythoa variabilis*, numerous predicted polypeptides with canonical venom protein features are identified. These polypeptides comprise putative proteins from different toxin families: neurotoxic peptides, hemostatic and hemorrhagic toxins, membrane-active (pore-forming) proteins, protease inhibitors, mixed-function venom enzymes, and venom auxiliary proteins. The synthesis and functional analysis of two of these predicted toxin products, one related to the ShK/Aurelin family and the other to a recently discovered anthozoan toxin, displayed potent *in vivo* neurotoxicity that impaired swimming in larval zebrafish. Altogether, the complex array of venom-related transcripts that are identified in *P. variabilis*, some of which are first reported in Cnidaria, provides novel insight into the toxin distribution among species and might contribute to the understanding of composition and evolution of venom polypeptides in toxiferous animals.

Key words: Zoanthidea, Hexacorallia, transcriptome, venomics, peptide toxin, RNA-seq, molecular toxinology.

Introduction

The phylum Cnidaria, which diverged from the Bilateria more than 600 million years ago, is among the first branches of life and encompasses thousands of species that predominantly inhabit marine environments. Traditionally, the cnidarians are divided into five classes (Ryland and Muirhead 1993; Marques and Collins 2004): Hydrozoa (*Hydra*, Portuguese Man-O-War, fire corals and hydromedusae); Scyphozoa (jellyfishes); Cubozoa (sea wasps); Staurozoa (sessile medusae); and Anthozoa (true corals, sea anemones, zoanthids and sea fans). At the tissue level, cnidarians are structurally simple, being composed only of two epithelia: ectoderm and endoderm. Organelle-like toxin-containing capsules with eversible tubules, called nematocysts, are integrated in one or

both of these epithelia, depending on the taxon (Fautin 2009), and impart to the organism the abilities to self-defend against predators and to capture prey (Yoffe et al. 2012). Despite the exceedingly simple basic body organization of cnidarians, Mackie (Mackie 2002) attributed the evolutionary success of the group to nematocysts, "the cnidarians' secret weapon," while the mixture of toxic substances that these nematocysts contain was reported by Mariscal (Mariscal 1974) as "among the largest and most complex intracellular secretion products known." Although there are more than 11,000 cnidarian species (Daly et al. 2007), the vast majority of the currently reported toxins (approximately 300) were collected from a small number of well-studied species, including sea anemones, jellyfishes, sea wasps, and hydra (Brinkman and Burnell 2009; Sher and Zlotkin 2009; Frazao et al. 2012), while the remaining toxic peptides and proteins of the anthozoans (e.g., *Palythoa* and *Protopalythoa*) remain understudied. Presently, the genetic backgrounds of several cnidarian species have been investigated using next-generation sequencing technology. For instance, transcriptome analysis identified novel genes from the coral *Acropora palmate* (Polato et al. 2011) and characterized a complex transcriptional response to environmental CO_2-driven acidification in the coral *Acropora millepora* (Moya et al. 2012). Additionally, a large population of novel transcripts was identified in hydra by combining data from RNA sequencing with those from a whole-genome sequencing analysis (Wenger and Galliot 2013). Likewise, the completion of the whole-genome sequencing of three cnidarians, namely, *Nematostella vectensis*, *Hydra magnipapillata*, and *Acropora digitifera*, provided a deeper understanding of the molecular genetics of these representative cnidarian species (Putnam et al. 2007; Chapman et al. 2010; Polato et al. 2011; Shinzato et al. 2011).

In the present work, we studied the transcriptome of the colonial cnidarian *Protopalythoa variabilis* (class Anthozoa, subclass Hexacorallia, order Zoantharia, family Sphenopidae) in order to investigate the presence of peptide toxin-related components in its tissue. This species of zoanthid is found in the shallow waters of coral reefs along the Gulf of Mexico and the Caribbean up to the southernmost point of the Brazilian coast (Acosta et al. 2005; Boscolo and Silveira 2005; Fautin and Daly 2009; Rabelo et al. 2013). In Brazil, *Protopalythoa* species preferentially inhabit the beach rocks and reefs of the Northeast region and commonly cohabit with other zoanthid species of the genera *Zoanthus* and *Palythoa* (Rohlfs De Macedo and Belém 1994; Barradas et al. 2010). A few studies in the field of pharmacology and toxicology have been conducted using the species *P. variabilis*, including the characterization of lipidic α-amino acids with pro-apoptotic activity (Wilke et al. 2009, 2010) and of sulfonylated ceramides, which represent a new class of sulfur-containing lipids with an uncharacterized biological function (Almeida et al. 2012). It is known that the symbiotic dinoflagellate species of *Palythoa*, which is a sister group of *Protopalythoa*, produces the potent non-nematocystic cytolytic and cytotoxic palytoxin, one of the most toxic non-peptide substances ever discovered (Shimizu 1983; Wu 2009; Bellocci et al. 2011). Additionally, from *Palythoa*, a structurally unusual α-helical V-shaped neurotoxic peptide that inactivates the voltage-gated sodium (Na_v) channels and interferes with neurotransmission was recently elucidated (Lazcano-Perez et al. 2014). Considering that (1) bioactive and toxic polypeptides have been isolated and characterized from other cnidarians (Lazcano-Perez et al. 2012) and (2) so little is known about the supposed venom-related peptide precursors in zoanthids, we applied Illumina sequencing technology to cover the full transcriptome of *P. variabilis* to identify transcripts encoding animal peptide toxins. Transcriptome annotation revealed a stunning repertoire of predicted venom-related precursors that have structural relationships with various known toxin families in venomous and poisonous animals. In support of these findings, a functional validation of two novel *P. variabilis* peptides, one belonging to the ShK/Aurelin family of K^+-ion channel blockers and the other to the novel anthozoan neurotoxin family, demonstrated their lethality to zebrafish, in which they caused neurotoxicity and disturbed locomotion. To the best of our knowledge, this is the first study of a zoanthid transcriptome to report a presumable basal core of venom-related polypeptides in a zoanthid.

Materials and Methods

Sample Collection

Samples of *P. variabilis* (fig. 1A) were collected from colonies in the beach-rock bands of Porto de Galinhas, Pernambuco, Brazil (8°30′20″S, 35°00′34″W) during low tide. *P. variabilis* specimens were quickly washed in distilled water, chopped with scissors and pooled together immediately in 10 volumes of RNA*later* (Life Technologies, USA) for RNA preservation. After storage at 4 °C for 48 h, the RNA-preserving solution was drained, and the tissue was kept at −80 °C until processing. The minced *P. variabilis* tissue was powdered with a porcelain mortar and pestle under liquid nitrogen, and the total RNA was purified using TRIzol reagent (Life Technologies, USA) according to the manufacturer's protocol.

Library Preparation and RNA Deep Sequencing Using Illumina Technology

The *Protopalythoa variabilis* library for deep RNA sequencing was prepared according to a standard protocol established by the Beijing Genomic Institute—BGI (Shenzhen, China). Initially, polyadenylated RNA sequences were isolated using oligo(dT). Single-stranded 5′-RNA adaptors were ligated to mRNA fragments using T4 RNA ligase (Ambion, Austin, TX, USA) and then reverse transcribed into cDNA using Superscript III reverse transcriptase (Invitrogen, Carlsbad, CA,

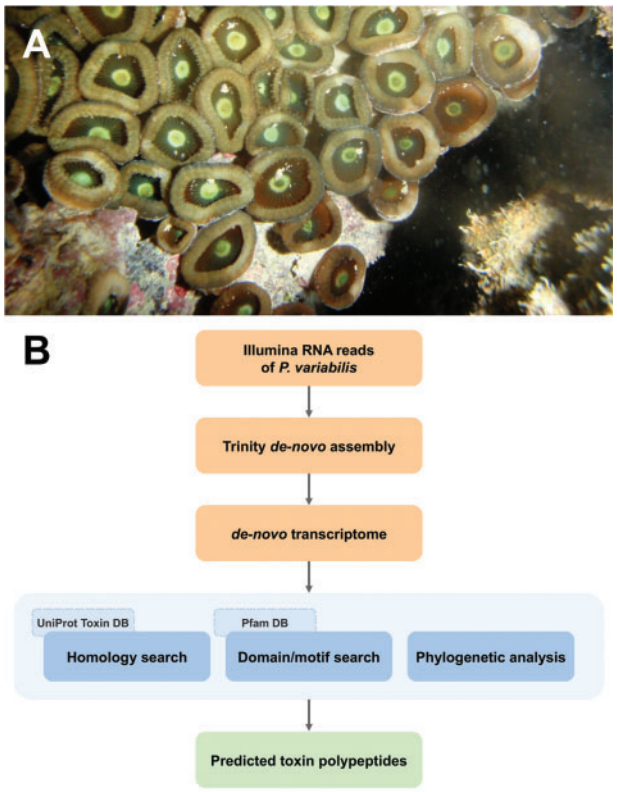

Fig. 1.—*Protopalythoa variabilis* in nature and the transcriptomic analysis pipeline. (*A*) A submerged colony of *Protopalythoa variabilis* (Duerden, 1898) on reefs of Porto de Galinhas Beach, PE, Brazil, with their polyps partially opened. (Courtesy of Liany Melo. B, M.Sc.). (*B*) Diagram of the pipeline that was used in this research to mine the toxin polypeptides of *P. variabilis* transcriptome.

USA). Then, a 3′-DNA adaptor was ligated to the digested DNA fragments after digestion with MmeI, and the products were amplified using PCR. Finally, 90 bp pair-end RNA deep sequencing (RNA-seq) was performed on a HiSeq 2500 automatic sequencing platform at BGI, China.

Data Processing, Assembly, and Assessment

The paired-end reads that were generated using RNA-sequencing on Illumina technology were filtered by an in-house C ++ script with the aim of excluding low-quality (bases below Q20 quality and/or with unknown nucleotides) and duplicate reads. The Illumina software was used to eliminate the adaptor sequences. The trimmed and cleaned-up reads were assembled using the Trinity program pipeline (Haas et al. 2013). Briefly, all of the sequenced reads were mounted *in silico* into contigs, and paired-end information was used to construct scaffolds. The unknown bases were filled with Ns, and non-redundant transcripts were then acquired using TGICL (Pertea et al. 2003). This Transcriptome Shotgun Assembly (TSA) project was deposited at DDBJ/EMBL/GenBank under the accession GCVI00000000, associated with the BioProject PRJNA279783 and biosample SAMN03450566. The version described in this paper is the first version, GCVI01000000.

In order to assess transcriptome assembly quality, all reads were initially aligned to the assembled transcripts using Burrows–Wheeler Aligner (BWA) (Version: 0.7.7-r441) (Li and Durbin 2010). Then samtools (Version: 0.1.19-44428cd) (Li et al. 2009) and bedtools (Version: 2.17.0) (Quinlan and Hall 2010) were applied to evaluate the depth and coverage of alignment. The transcripts which were mapped with at least ten reads and covered more than 80% were retained for the transcriptome analysis.

Functional Annotation of *P. variabilis* Transcriptome

All of the transcripts that were generated from the assembly process were annotated using BLASTx (Mount 2007) based on similarity with entries that have been structurally and functionally characterized and are maintained in the National Center for Biotechnological Information (NCBI), the non-redundant protein sequence UniProtKB/Swiss-Prot, the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the Cluster of Orthologous Groups (COG) databases. The *E*-value threshold was 1*E*−5.

Identification of Toxin-Like Polypeptides in the *P. variabilis* Transcriptome

To identify potential peptide toxins, toxin-like and venom-auxiliary polypeptides from *P. variabilis*, a comprehensive three-steps selection pipeline was established, initiating with a pre-screening, followed by a domain/motif sequence comparison, and then a phylogenetic analysis (fig. 1B). Briefly, all of the transcripts were compared using BLASTx to an exclusive database for toxins and venom proteins in the UniProt Animal Toxin Annotation Program (http://www.uniprot.org/program/Toxins, last accessed February 2015), restricting the *E*-value to 1*E*−5. The resulting putative polypeptides were considered positive if the presence of the toxin motif or family domain (partial or complete) could be confirmed in their predicted sequence. This validation step relied on a BLASTp (version 2.2.30+) comparison using the toxin family domain HMM profile consensus sequence, as well as a profile search using the toxin family domain HMM profile with HMMER v3.1b2 (http://hmmer.org/) (Choo et al. 2004). Exceptions are described in the supplementary table S4, Supplementary Material online. The last identification step consisted of a Maximum Likelihood method based tree inference of the selected toxin domain/motif sequence, with a bootstrap test of 500 replicates using MEGA (Kumar et al. 2016). The resulting phylogenetic trees were visualized and produced using the FigTree v1.4.2 software (http://tree.bio.ed.ac.uk/software/figtree/) and Inkscape 0.91 (https://inkscape.org).

Sequence Alignment

Multiple sequence alignments of *in silico*-translated toxin-related sequences were performed using MUSCLE v3.8.31 (Edgar 2004) or Kalign v2.04 (Lassmann et al. 2009) whether the toxin domain/motif was a single or repeated structural units, respectively. The amino acid sequence identities and similarities were highlighted by Jalview 2.9.0b2 (Waterhouse et al. 2009).

Modeling and ProBiS Analysis of *P. variabilis* Toxin-Related Polypeptides

Two different methods, MODELLER (Eswar et al. 2006) and PEP-FOLD (Thevenet et al. 2012), were used to predict the 3D models of putative toxin-related polypeptides from translated transcript sequences of *P. variabilis* based on the available coordinates of the already-known structures of similar toxins. In almost all cases, the structures of same-family toxins were suitable for use as templates for homology modeling using the MODELLER software. For the predicted novel anthozoan neuropeptide toxin from *P. variabilis*, a short polypeptide for which no structural information is currently available, the PEP-FOLD *de novo* structure prediction server was used. The visualization of known proteins structures and structural models of *P. variabilis* putative toxin-related polypeptides was achieved using the VMD program v1.9.2 (Humphrey et al. 1996). To predict the binding sites of the 3D structures of toxin-related proteins, the ProBiS server was used (Konc and Janezic 2012).

Signal Peptide Prediction

The presence of signal peptide motif in the protein precursor sequences of *P. variabilis* was predicted and compared using both the Signal-BLAST with the detection mode set as "Balanced Prediction" (Frank and Sippl 2008) and SignalP with the default *D*-cutoff value (Petersen et al. 2011).

Synthesis of the *P. variabilis* ShK/Aurelin-like Peptide and the Novel Anthozoan Neurotoxin-Like Peptide

Sequences of the novel mature ShK/Aurelin-like peptide (i.e., TGGSCVDRNQKCRNWTRYCSYHPYVRANCKKTCRLC) and the novel anthozoan neurotoxin-like peptide (KYWILNVPASVCDEYCWSQMLLLGISSVSISV) were retrieved from the *P. variabilis* transcriptome and synthesized by solid phase chemistry; a purity grade of greater than 95% was obtained, confirmed by the presence of a single peak in analytical reverse-phase HPLC and mass spectrometry analyses (Cellmano Biotech Limited, Hefei, China). Complete deprotection and cleavage were performed with trifluoroacetic acid in water. The crude peptides were precipitated out by adding chilled ether. Then, the final peptides were purified by HPLC and freeze-dried for storage and further study.

Zebrafish Maintenance

Wild-type zebrafish (AB strain) were used in this study. Embryos were collected after natural spawning and maintained at 28.5 °C in embryo medium (13.7 mM NaCl, 540 μM KCl, 25 μM Na_2HPO_4, 44 μM KH_2PO_4, 300 μM $CaCl_2$, 100 μM $MgSO_4$, and 420 μM $NaHCO_3$, pH 7.4). The Animal Research Ethics Committee, University of Macau, granted ethical approval for the animal experiments.

Toxicity of *P. variabilis* ShK/Aurelin-like Peptide and Novel Anthozoan Neurotoxin-Like Peptide toward Zebrafish Larvae

Zebrafish larvae at 3 days post-fertilization (dpf) were separated in 24-well plates (12–15 larvae per well). For the toxicity test of ShK/Aurelin-like peptide, 3-dpf fish larvae were exposed to 1-mL solutions with increasing concentrations of *P. variabilis* ShK/Aurelin-like peptide (ranging from 5 to 20 μM, i.e., 0, 5, 10, 15, and 20 μM). For the toxicity test of novel anthozoan neurotoxin-like peptide, 3-dpf fish were exposed to 1-mL solutions with increasing concentrations (ranging from 0.3 to 30 μM, i.e., 0, 0.3, 1, 3, 10, 30) of *P. variabilis* novel anthozoan neurotoxin-like peptide for three different durations (5 min, 20 min, and 24 h). The acute toxicity and mortality of zebrafish were determined by monitoring the absence of a heartbeat under a light microscope. The survival rates were calculated as the percentage of surviving larvae in relation to the total number of treated larvae.

Zebrafish Locomotion Test

Zebrafish embryos at 6 dpf were incubated with 10 μM *P. variabilis* ShK/Aurelin-like peptide for the indicated durations. Zebrafish larvae at 6 dpf were transferred into 96-well plates (one larva per well and 12 larvae per group). Zebrafish showing signs of excessive stress upon handling (such as rapid and disorganized swimming or immobility for 2 min) were discarded. The experiments were performed in a calm, sealed area. The swimming behavior was monitored by an automated video tracking system (Viewpoint, ZebraLab, LifeSciences, France). The 96-well plates and camera were housed inside a Zebrabox, and the swimming pattern of each fish was recorded in three sessions of 15 min each. The total distance moved was defined as the distance (in millimeters) that a zebrafish moved during one session (15 min). A statistical analysis of the total distance traveled by each zebrafish larva in the different treatment groups was performed using an ANOVA and Dunnett's test.

Results and Discussion

The majority of our knowledge on cnidarian toxic components is mainly derived from sea anemones (Frazao et al. 2012), hydra (Sher and Zlotkin 2009), and the deadly jellyfishes (Brinkman and Burnell 2009; Brinkman et al. 2015).

The exotic appearance of cnidarians, particularly from the class Anthozoa, attracts the attention and curiosity of intruders and collectors that invade their habitat and make contact with these exquisite marine organisms. Eventually, undesirable health problems occur as a consequence of improper handling. For example, aquarists that commercialize these wild marine animals and remain in close contact with them suffer from intoxication with palytoxin (Deeds et al. 2011) or with a number of yet uncharacterized toxins. Because *Cnidaria* species have thrived successfully in oceans for millions of years, we aimed to characterize the presumable primordial toxic polypeptide precursors that are potentially expressed in *P. variabilis* by analyzing its transcriptome. The present study has allowed us to identify a repertoire of transcripts related to toxic polypeptides and other venom-related proteins that share molecular and structural characteristics with toxins of other species of cnidarians and animals of distinct phyla as well.

General Features of *P. variabilis* Transcriptome

All *P. variabilis* samples, collected from a single location, were pooled together as we aimed to report the global venom composition found in this zoanthid species rather than comparing variations among individuals. The transcriptome sequencing yielded a total of 67,549,914 raw paired-end reads with a length of 90 nucleotides (nt). After eliminating adaptor sequences, duplicated sequences and low-quality reads, the remaining 60,891,368 clean reads were assembled using two assembly runs into 126,441 transcripts (supplementary fig. S1 and Table S1, Supplementary Material online). Sequence alignment of clean reads against assembled transcripts using BWA indicated that 58,916,865 (87.22%) reads could be mapped to the assembled transcripts. Furthermore, the depth and coverage assessment of alignment using samtools and bedtools showed that 121,877 transcripts were mapped at least ten reads and in which more than 80% of sequences were covered by reads. For 53,950 transcripts (~42%), BLASTx identified significant similarity with known proteins available from the Swiss-Prot, COG, KEGG, and NCBI nr databases (supplementary fig. S2 and Table S2, Supplementary Material online). In total, approximately half (i.e., 48%) of the annotated transcripts shared 14-40% similarity with the protein sequences of distinct species (supplementary fig. S2*A*, Supplementary Material online). Based on fixed *E*-value distributions, 41% of the annotated transcripts fit between 1*E*−5 and 1*E*−15, and 21% between 1*E*−15 and 1*E*−30 (supplementary fig. S2*B*, Supplementary Material online). In terms of the distribution of species with similar transcriptomes, nearly 25% of these annotated transcripts have protein sequence counterparts in the starlet sea anemone *Nematostella vectensis* (supplementary fig. S2*C*, Supplementary Material online), one of the most ancient lineages of cnidarians (Rodriguez et al. 2014).

Concerning the general functions of *P. variabilis* transcripts, a total of 18,762 transcripts shared significant sequence similarity with the annotated proteins available in the COG database, being grouped into 25 functional categories (supplementary fig. S3, Supplementary Material online). A significant proportion of translated transcripts code for precursors associated with DNA replication, recombination, and repair; ribosomal structure and biogenesis; and translation. From the KEGG database, 333 pathways were mapped and retrieved, containing 34,468 *P. variabilis* transcripts. From these pathways, more than 300 metabolic routes fit in the top 20 KEGG pathways (supplementary Table S3, Supplementary Material online), including almost 3,000 unigene-encoding proteins involved in the biosynthesis of secondary metabolites.

Identification of Toxin-Like Polypeptides in *P. variabilis* Transcriptome

Initially, 1556 transcripts whose translated sequences are similar to the protein sequences of toxins were found in *P. variabilis* during the pre-screening step (supplementary table S4, Supplementary Material online). Using the bioinformatics pipeline depicted in figure 1B, 1554 transcripts were first assigned to a repertoire of sequences in the UniProt animal toxin annotation project database. In complement to the UniProt animal toxin database, the sequence of a novel neurotoxin peptide from *P. caribaeorum* species (Lazcano-Perez et al. 2014) was included and used to retrieve one additional transcript using BLAST against our *P. variabilis* sequences. Moreover, the Pfam annotation, based on the protein domain motif, of the whole *P. variabilis* transcript revealed one more transcript with a *Stichodactyla* toxin (ShK) domain. In order to confirm the number and quality of the *P. variabilis* toxin-like sequences with higher confidence, the pre-selected sequences were filtered to remove false positives through an additional BLAST and HMMER analysis against the toxin family Pfam. Out of the initially retrieved 1556 contigs, 690 were confirmed with the presence of total or partial toxin family domain(s) in their respective translated products. Finally, the last stringent step consisted of a maximum likelihood inference of the toxin domain/motif in order to retain only transcripts whose products were related to a toxin clade. Whenever applicable, the similarity of protein structures was validated with ProBiS (supplementary table S4 and supplementary file S1, Supplementary Material online).

Essentially, the identified putative toxin-related polypeptides were categorized into six main groups (table 1). The neurotoxic peptides, with five transcripts related to four toxin families, is the most represented group, followed by the membrane-active peptides, the hemostatic and hemorrhagic toxins, the protease inhibitors, the venom-associated enzymes with mixed functions, and the venom auxiliary proteins. The transcripts encoding these predicted toxin-related

Table 1
Repartition of the Predicted Toxin-Like Polypeptides in the Anthozoan *P. variabilis*

Toxin families	Function	Organism(s) where toxin majoritarily reported	Number of contigs in *P. variabilis*
Neurotoxin			
ShK/Aurelin	Blocks voltage-dependent potassium, olfactory and retinal cng channels	Sea anemone, jellyfish, snake	1
Anthozoan neurotoxin	Inactives voltage-gated sodium channel	Coral	1
Three finger toxin	Inhibits the nicotinic and muscarinic acetylcholine receptor, with some exhibiting cytolytic activity	Snake	1
Turripeptide	Ion-channel blocker	Sea anemone, marine gastropod	2
Hemostatic & hemorrhagic toxins			
Snake venom VEGF	Induces capillary permeability and angiogenesis	Snake	1
C-type lectin (Snaclec)	Interferes with hemostasis	Snake	2
Protease inhibitors			
Kunitz-type	Serine protease inhibitor, with some exhibiting an ion channel blocker activity	Sea anemone, sea snail, snake, arachnid, insect	1
Thyroglobulin type-1	Probable cysteinyl proteinase inhibitor	Spider	1
Membrane-active peptides			
Actinoporin	Forms cations-selective pores and causes hemolysis	Sea anemone, sea snail, coral	1
Helofensin	Lethal toxicity	Lizard	1
MACPF	Lethal toxicity and hemolytic activity	Sea anemone, sea snail, coral	1
Mixed function enzymes			
Phospholipase A2	Induce necrosis, cause inflammation, inhibitor of blood coagulation, block neuromuscular transmission, ...	Reptile, sea anemone, sea snail, coral, bee, scorpion	3
Venom auxiliary proteins			
Venom protein 302	-	Sea anemone, scorpion, ant	2

polypeptides according to the aforementioned toxin categories are listed in supplementary table S5, Supplementary Material online.

Neurotoxic Peptides and Proteins

Putative peptides of transcripts similar to four neurotoxin families were identified in the *P. variabilis* transcriptome, namely, turripeptide, ShK, three-finger toxin, and a novel anthozoan neurotoxin-like peptide (Table 1).

Turripeptide-Like

Originally discovered in turrid gastropods (Duda et al. 2001), a turripeptide-like peptide was also reported recently in marine annelids (von Reumont et al. 2014) and box jellyfish (Brinkman et al. 2015). Two sequences exhibiting similarity to turripeptide-like neurotoxin were found in the *P. variabilis* transcriptome with a greater than 30% identity. Similar to conotoxins, turripeptides are variable in size and structure and are classified with respect to their cysteine framework. The putative turripeptide type found in *P. variabilis* has the cysteine framework IX (C-C-C-C-C-C), structurally similar to members of the conopeptide P-like superfamily. Phylogenetic analysis of the Kazal domains shows that the relationship of the putative *P. variabilis* turripeptide to two different turripeptide clades, which suggests we could be in presence of isoforms divergence (fig. 2A) with 56% identity between both domain sequences (supplementary fig. S4, Supplementary Material online).

ShK/Aurelin-Like Peptide Toxin

ShK is a 35-residue peptide sequence that was first discovered in the sea anemone *Stichodactyla helianthus* and can block voltage-gated potassium channels (Castaneda et al. 1995). Through Pfam domain/motif screening, we identified in the transcriptome of *P. variabilis* one transcript containing the *Stichodactyla* toxin (ShK) domain (Pfam ID: PF01549). BLASTp and HMMER validations resulted in 29% identities and an expected accuracy of 0.97 using the HMM profile of the family domain. The predicted *P. variabilis* toxin-like is located in the clade of the toxic ShK peptides (fig. 2B) and the ProBiS result of the modeled structure of the putative *P. variabilis* ShK-like peptide showed that the protein aurelin (PDB ID: 2LG4) was the second best match based on Z scores (supplementary file S1, Supplementary Material online). Aurelin is a known antimicrobial peptide that was identified from the mesoglea of the scyphoid jellyfish *Aurelia aurita* (Ovchinnikova

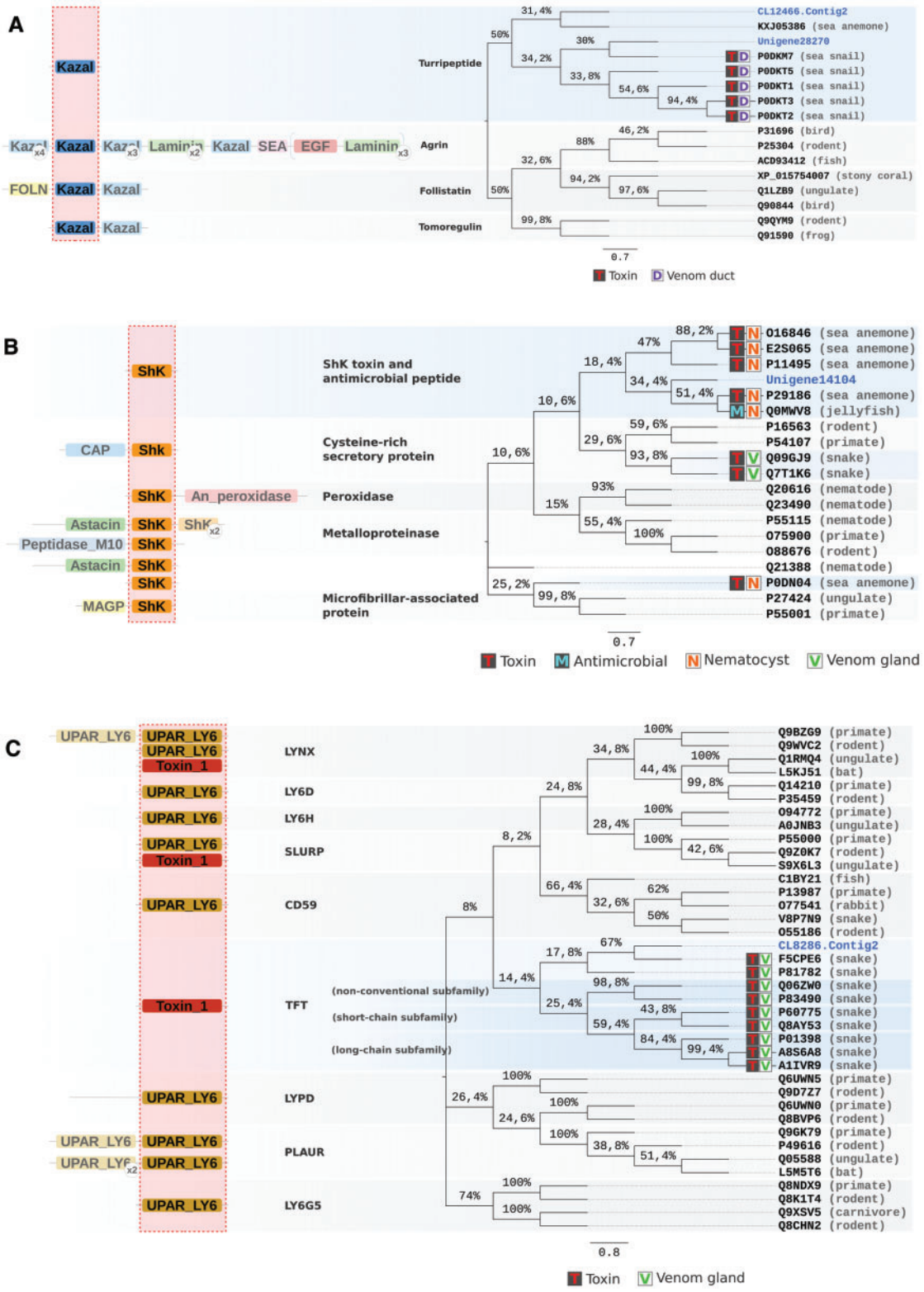

Fig. 2.—Phylogenetic analysis of the predicted neurotoxin-like polypeptides. For each analysis, on the left panel is a depiction of the protein domain organization of the respective protein families used in the phylogenetic inference. The region used for the analysis is highlighted with a red box. On the right, the resulting consensus phylogenetic tree was inferred by using the Maximum Likelihood (ML) method from 500 bootstrap replicates. The bootstrap percentage values are shown next to the branches. Toxin and non-toxic clades are colored in blue and gray respectively, with the *P. variabilis* IDs written in

(continued)

et al. 2006), with structural and spatial conformation similar to that of ShK. The domain alignment of the predicted *Protopalythoa* ShK-like peptide with Aurelin revealed that both amino acid sequences have a percentage identity of 38%, with the conservation of the six cysteine residues, which stabilizes the structure of the ShK domain by forming three disulfide bridges. These structural attributes and the predicted structural model of *P. variabilis* ShK-like peptide is illustrated below (fig. 8).

Three-Finger Toxin-Like

One translated *P. variabilis* transcript shared proper sequence similarity with the snake venom TFT Toxin_1 domain region (Pfam ID: PF00087). This domain displays the closest similarity with the snake toxin sequence TFT MALT0070C (UniProt ID: F5CPE6) from the venom gland of *Micrurus altirostris* (Correa-Netto et al. 2011). The TFTs encompass a large family of toxic polypeptides from snake venom; these polypeptides are mainly monomeric and are characterized by the archetypical three β-strand loops constrained by four disulfide bonds, spatially resembling fingers protruding from a core. Despite minor but critical structural variations, TFTs evolved to produce an array of pharmacological activities and targets, such as neurotoxins (α-neurotoxins, β-cardiotoxins, muscarinic toxins, and Ca^{2+}-channel blockers), acetylcholinesterase inhibitors, cytotoxins (cardiotoxins), platelet aggregation inhibitors, coagulation factor inhibitors, heparin binders, and K^{+}-channel, and integral-receptor ligands (Kini and Doley 2010; Kini 2011; Utkin 2013). The multiple sequence alignment and phylogenetic relationship of the putative *P. variabilis* TFT-like Toxin_1 domain with related sequences indicate structural similarity with MALT0070C and Bucandin (fig. 2C and supplementary file S1, Supplementary Material online). Using ProBiS for binding site prediction indicated similarity between the structural conformation of binding sites of the *P. variabilis* TFT-like sequence and that of Bucandin, a TFT from the venom of the elapid *Bungarus candidus* (PDB ID: 1F94) (supplementary file S1, Supplementary Material online).

Anthozoan Neurotoxin-Like

This peptide has neurotoxic activity, exerting its effect by inactivating the voltage-gated sodium channel subtype 1.7, and was purified and characterized from the tissue of the *P. caribaeorum* (Lazcano-Perez et al. 2014). This recently isolated neurotoxin is a 32-residue-long sequence with two cysteine residues that are involved in the formation of a single intrachain disulfide bridge and an unusual V-shape α-helical structure. A direct BLAST search against all of our contigs identified a similar predicted neurotoxin sequence in the *P. variabilis* transcriptome. The overall structural topologies of the main α-helix of both neurotoxic peptide sequences are perfectly superposed and are depicted below (fig. 9).

Hemostatic and Hemorrhagic Toxins

This group comprises three *P. variabilis* predicted homologs of animal venom components that, separately or in combination, are known to interfere with hemostasis and the associated vasculature physiology, such as angiogenesis, capillary permeability, blood pressure changes, and bleeding (table 1).

C-Type Lectin-Like

Two contigs with a putative protein product with a C-type lectin domain (Pfam ID: PF00059) were found in the *P. variabilis* transcriptome. C-type lectins in venom are associated with hemagglutination, inflammatory response, and hemostasis. Two main toxin groups containing C-type lectin domain comprise the true venom lectin and the snake C-type lectin (snaclec) families. Accordingly, both putative *P. variabilis* protein products are phylogenetically related to the snaclec toxins (fig. 3A).

Snake Venom VEGF-Like

The *P. variabilis* transcriptome contains one predicted sequence similar to that of the snake venom VEGF domain (VEGF-F) (Pfam ID: PF00341) of the snake toxin barietin (UniProt ID: C0K3N1) (Yamazaki et al. 2009). Although distinct, the predicted *P. variabilis* toxin-like polypeptide displays close phylogenetic relationship with sequences in snake venom VEGF toxin clade, compared to other non-toxic VEGF domains (fig. 3B). The multiple sequence alignment (supplementary fig. S7, Supplementary Material online) showed that the putative *P. variabilis* VEGF-like peptide shares common features with known snake venom VEGF sequences but

Fig. 2.—Continued

light blue. Legend of the labels used to describe toxin sequences is available below each tree. (*A*) Turripeptide-like. Multiple sequence alignment of the Kazal domains was achieved with Kalign. ML tree inferred based on a WAG + Gamma substitution model. Multiple sequence alignment of the Kazal domain from the predicted *P. variabilis* protein sequence with selected toxin is available in the supplementary figure S4, Supplementary Material online. (*B*) ShK-like. Multiple sequence alignment of the ShK domains was achieved with Muscle. ML tree inferred based on a WAG + I+Gamma substitution model. Multiple sequence alignment of the ShK domain from the predicted *P. variabilis* protein sequence with selected toxin is available in the figure 8*A*. (*C*) TFT-like. Multiple sequence alignment of the UPAR_LY6/Toxin_1 domains was achieved with Kalign. TFT and Ly6/uPAR protein families were already reported as structurally similar (Tirosh et al, 2013; Tsetlin, 2015). ML tree inferred based on a WAG + I+Gamma substitution model. Multiple sequence alignment of the Toxin_1 domain from the predicted *P. variabilis* protein sequence with selected toxin is available in the supplementary figure S5, Supplementary Material online.

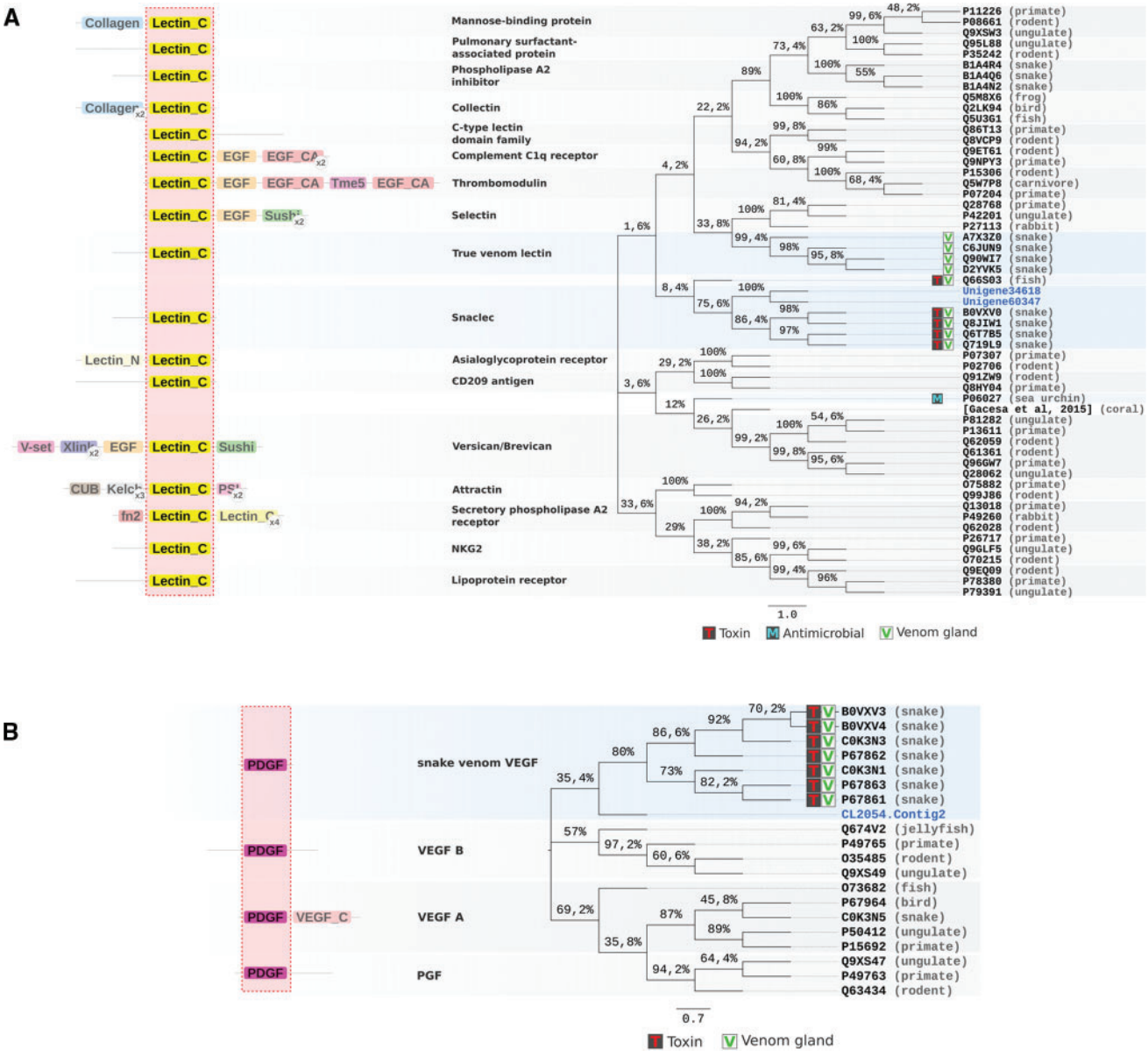

Fig. 3.—Phylogenetic analysis of the predicted hemostatic and hemorrhagic toxin-like polypeptides. Same legend information like detailed in figure 2. (*A*) C-type lectin-like. Multiple sequence alignment of the Lectin_C domains was achieved with Kalign. ML tree inferred based on a JTT + I+Gamma substitution model. Multiple sequence alignment of the Lectin_C domain from the predicted *P. variabilis* protein sequence with selected toxins is available in the supplementary figure S6, Supplementary Material online. (*B*) Snake venom VEGF-like. Multiple sequence alignment of the PDGF domains was achieved with Muscle. ML tree inferred based on a WAG + Gamma substitution model. Multiple sequence alignment of the PDGF domain from the predicted *P. variabilis* protein sequence with selected toxins is available in the supplementary figure S7, Supplementary Material online.

with an identity that is low (28%) in the VEGF domain compared to that of toxin barietin, but close to that expected (commonly 29–64%) for VEGF-Fs (Yamazaki et al. 2009). The structural model was submitted to the ProBiS server and indicated that this anthozoan VEGF homolog possesses the typical binding site of VEGF for the cognate receptor (supplementary file S1, Supplementary Material online). Interestingly, we also discovered transcripts coding for the VEGF receptor in *Protopalythoa*, as well as the putative non-venom VEGF-C (data not shown). This finding is intriguing in the light of the physiological role of VEGF-like polypeptides in *Protopalythoa* and their relation with organismal venomics.

Protease Inhibitors

Another class of venom-related components identified in the *P. variabilis* transcriptome that might contribute to a presumable polypeptide toxin repertoire comprises precursors coding

for protease inhibitors of Kunitz-type and thyroglobulin type-1 (table 1). Venoms commonly contain several kinds of protease inhibitors that, with the exception of cystatin, which is mostly non-toxic, display diverse pharmacological effects, including neurotoxicity.

Kunitz-Type-Like

The retrieval of one transcript predicting a polypeptide similar to Kunitz-type protease inhibitor, demonstrates the potential presence of protease inhibitor in the pharmacological cocktail of *P. variabilis*. Several members of the Kunitz-domain family have been characterized from marine and terrestrial organisms (Mourao and Schwartz 2013). These peptides display dual functions (protease inhibition and/or ion channel blockade) depending on their structural characteristics. In sea anemones, homologs of Kunitz-type protease inhibitors are classified as type II toxins that are used for defense by protecting organisms from endogenous proteases and causing paralysis in prey by blocking several voltage-dependent potassium ion channel subtypes. Despite the fact that all venom Kunitz-type protease inhibitors are constituted of a single Kunitz/Bovine pancreatic trypsin inhibitor (BPTI) domain (Pfam ID: PF00014), they together form several distinct clades. The putative *P. variabilis* Kunitz-domain precursor splits in the clade of the sea snail Kunitz-type serine protease inhibitor conotoxins (fig. 4A). The amino acid sequences shares 45% identity with a conserved cysteine framework of type IX (C-C-C-C-C-C) (supplementary fig. S8, Supplementary Material online), also seen in the turripeptide precursor above described.

Thyroglobulin Type-1 Protease Inhibitor-Like

One *P. variabilis* polypeptide was found to be similar to ctenitoxin (fig. 4B, supplementary fig. S9, Supplementary Material online). Ctenitoxin is a thyroglobulin type-1 protease inhibitor that was isolated from the venom gland of the Brazilian spider *Phoneutria nigriventer*. Generally inhibiting cysteine proteases, the thyroglobulin type-1 protease inhibitor can also exhibit inhibitory activity against aspartic proteases and metalloproteases (Mihelic and Turk 2007).

Membrane-Active Polypeptides

A proportion of the predictable toxic components from the transcriptome of *P. variabilis* encompass proteins that target cell membranes. We detected putative precursors that might be able to damage cell membrane via different mechanisms (table 1).

Actinoporin-Like

These toxins represent a major class of pore-forming toxins (cytolysins) that were discovered in sea anemones (Anderluh and Macek 2002) and are categorized as type II cnidarian cytolysins or β-pore forming proteins (β-PFPs) that induce hemolysis. In sea anemones, several isoforms have been identified in distinct species (Anderluh and Macek 2002; Frazao et al. 2012). With a molecular size of approximately 20 kDa, they are cysteine-free, compact β-sandwich structures, flanked by α-helices at N- and C-termini, that specifically interact with sphingomyelin, olygomerize on cell surfaces and perforate eukaryotic membranes causing cell lysis and death (Bakrac et al. 2008). Actinoporin family is basically composed of two subfamilies: the conoidea subfamily that groups toxins from sea snails, and the sea anemone-like subfamily that combines toxins from different anthozoan species, including our *P. variabilis* precursor (fig. 5A). The putative *P. variabilis* actinoporin precursor exhibits a highly conserved sequence when compared to sea anemone actinoporins and display all of the structural attributes that are required for biological activity (supplementary fig. S10, Supplementary Material online): the conserved sphingomyelin recognition and membrane-binding domain (the so-called phosphocoline binding site, POC-domain) comprising the tryptophan-rich stretch (P-[WYF]-[DR] pattern, position 106–108 in the alignment), the RGD-motif preserved in the form of the triad KGD (position 147–149), a short N-terminal α-helix (position 12–26). Generally, actinoporins from cnidarians and actinoporin-like proteins (ALP) from mollusks are detrimental to erythrocyte membranes and cause severe hemolysis (Anderluh and Macek 2002; Kawashima et al. 2003; Gunji et al. 2010). The actinoporin-like predicted sequence of *P. variabilis* represents a newly added member of the ALP family of β-pore-forming toxins that was previously reported in sea anemones (actiniarians) and hydra (hydrozoans) but is now identified in zoanthids.

Helofensin-Like

Toxins of the helofensin subfamily are members of the β-defensin family, which includes antimicrobial peptides that are present in the tissue of diverse organisms, including mostly mammals but also the anthozoan *Nematostella* (Putnam et al. 2007). One contig was found in the *P. variabilis* transcriptome and was validated by BLASTp but not by HMMER (supplementary table S4, Supplementary Material online). A possible explanation for this discrepancy is that the HMM profile uses the whole β-defensin family and not the more specific helofensin subfamily. Phylogenetic analysis shows the relationship of the *P. variabilis* predicted toxin-like sequences with helofensins (fig. 5B).

MACPF-Like

Another important type of membrane-active polypeptide predicted from the *P. variabilis* transcriptome includes a MACPF-like protein. Phylogenetic inference of the MACPF domain shows that the predicted *P. variabilis* product branches into the toxin clade (fig. 5C). MACPF domain-containing proteins

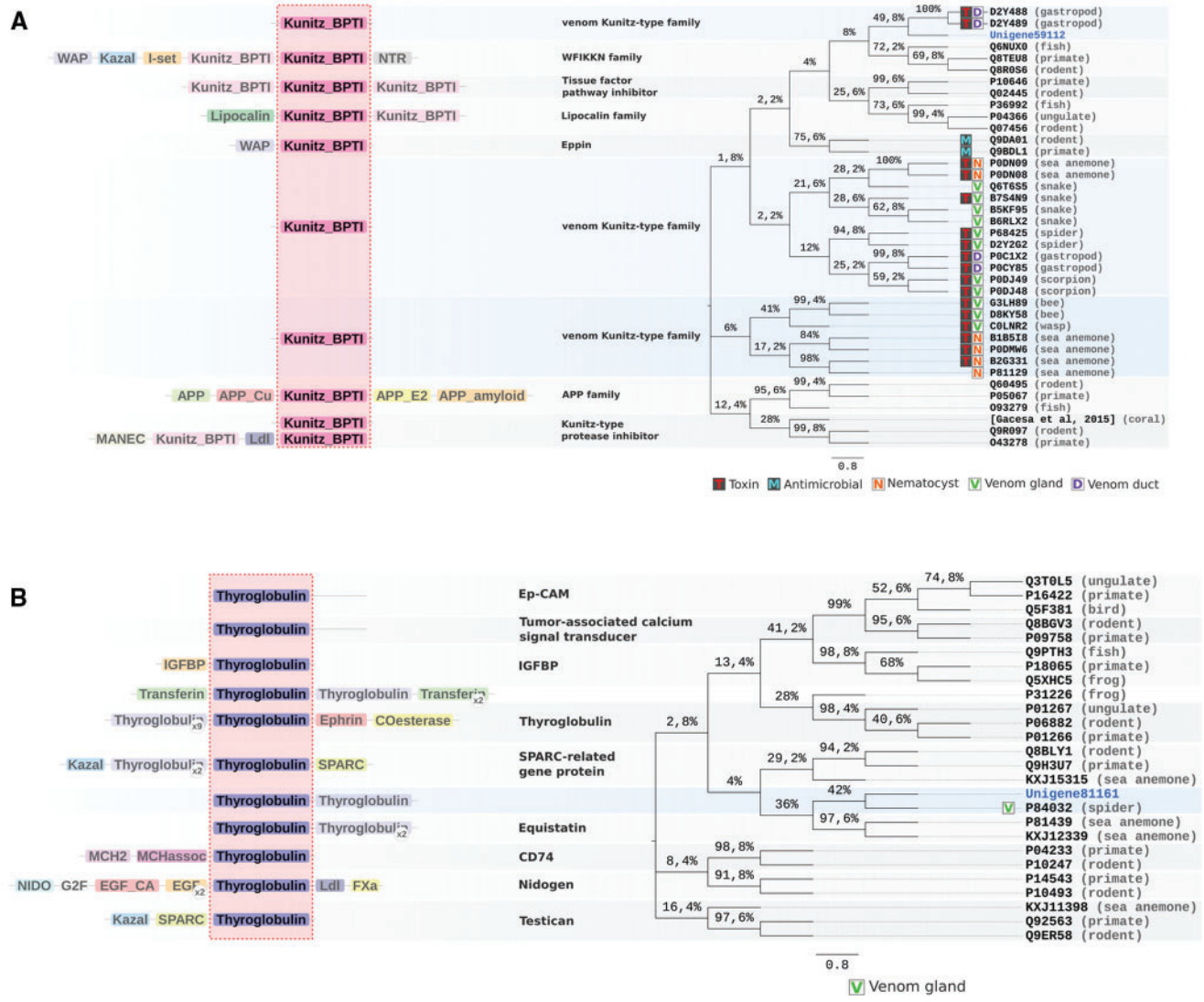

Fig. 4.—Phylogenetic analysis of the predicted protease inhibitors-like. Same legend information like described in figure 2. (*A*) Kunitz-type-like. Multiple sequence alignment of the Kunitz_BPTI domains was achieved with Kalign. ML tree inferred based on a WAG + I+Gamma substitution model. Multiple sequence alignment of the Kunitz_BPTI domain from the predicted *P. variabilis* protein sequence with selected toxins is available in the supplementary figure S8, Supplementary Material online. (*B*) Thyroglobulin type-1-like. Multiple sequence alignment of the Thyroglobulin domains was achieved with Kalign. ML tree inferred based on a WAG + Gamma substitution model. Multiple sequence alignment of the Thyroglobulin domain from the predicted *P. variabilis* protein sequence with selected toxins is available in the supplementary figure S9, Supplementary Material online.

are pore-forming polypeptides and toxins that are found in a diverse range of organisms. They are components of the host immune system and are hemolytic agents of cnidarian venom. The putative *P. variabilis* MACPF-like precursor bears a high structural resemblance to AvTx-60, a lethal MACPF toxin that was isolated from the venom of the sea anemone *Actineria villosa* (Oshiro et al. 2004). The retrieved putative *P. variabilis* MACPF-like nucleotide sequence is not complete, but it covers the second half of the mature AvTX-60A protein, including the last half of the MACPF domain, as well as the EGF domain (supplementary fig. S12, Supplementary Material online). This predicted *P. variabilis* MACPF-like toxin appears here as the first example of a MACPF pore-forming polypeptide described in an anthozoan other than sea anemones. These results reinforce our findings that membrane-active cytolysins and perforins might be significant contributors to the supposed toxic molecular weaponry of *P. variabilis* tissue that induces damage to the cytoplasmic membranes of target cells of predators or unwary victims.

Mixed Function Enzymes

Transcripts of *P. variabilis* were found that encode phospholipases A2 (table 1). Phospholipases A2 (PLA2) are enzymes that

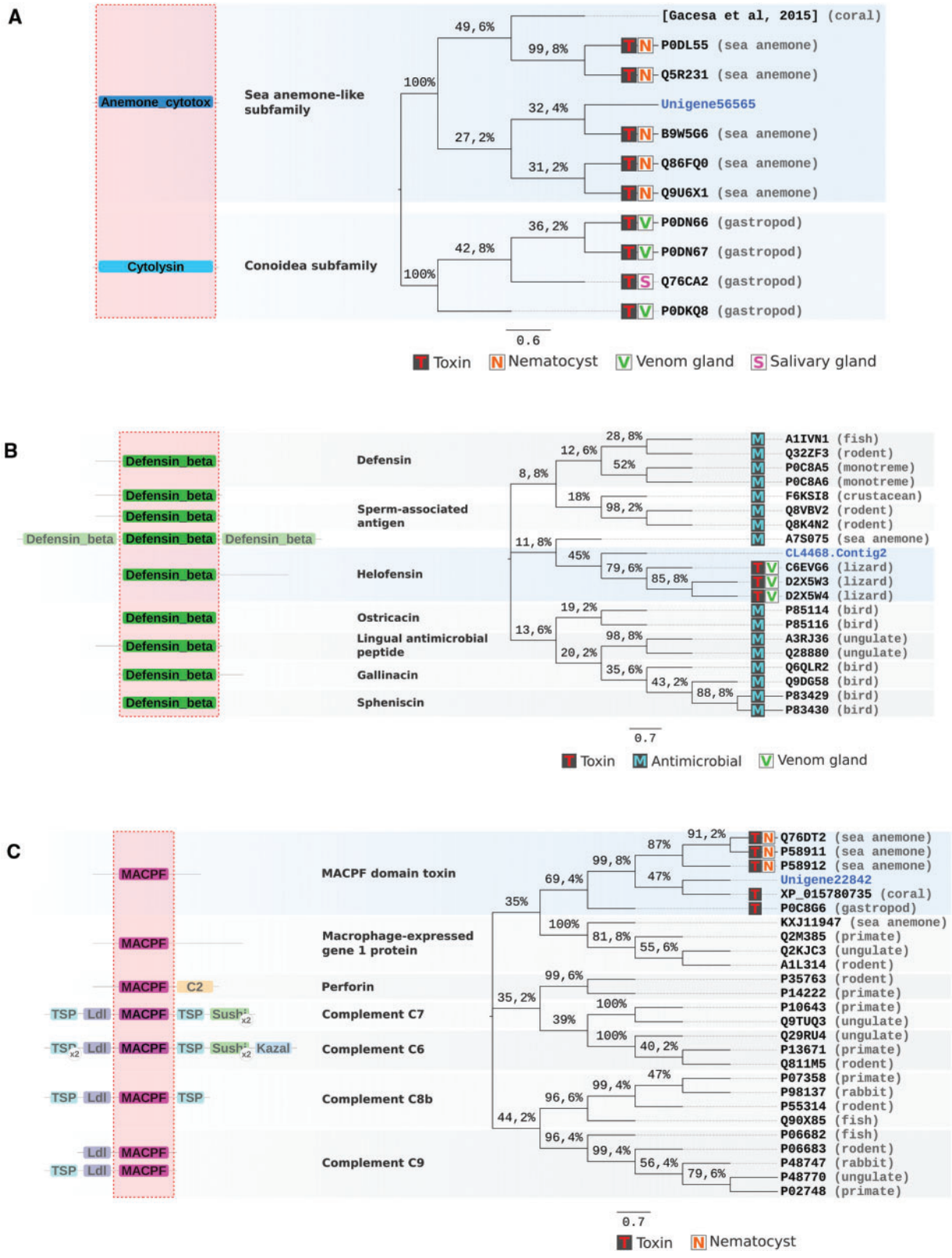

Fig. 5.—Phylogenetic analysis of the predicted membrane-active-like polypeptides. Same legend information like described in figure 2. (*A*) Actinoporin-like. Multiple sequence alignment of the Cytolysin domains was achieved with Muscle. ML tree inferred based on a WAG + Gamma substitution model. Multiple sequence alignment of the Cytolysin domain from the predicted *P. variabilis* protein sequence with selected toxins is available in

(continued)

can hydrolyze phospholipids into fatty acids and other lipophilic substances, and they play key roles in regulating signaling during numerous cellular events. Members of the PLA2 family are disseminated in the venom of numerous animals that exert multiple biological and pharmacological effects, including neurotoxicity, cytotoxicity, and platelet aggregation as well. In the *P. variabilis* transcriptome, three transcripts were found that encode putative PLA2s. The superfamily of PLA2s is complex and, in mammals, comprises eleven groups of secretory PLA2s (Murakami and Kudo 2002). Secretory PLA2s are also highly disseminated venom components among animals of different kingdoms, from type-III cytolysins in sea anemones to acidic, basic, and neutral PLA2s in snake venoms (Anderluh and Macek 2002; Kini 2003). Of the three predicted PLA2-like proteins from the *P. variabilis* transcriptome, one of them forms a distinct clade with other anthozoan sequences and the cone snail's Conodipine-M, this latter an unique representative until now of the PLA2 group IX subfamily (fig. 6 and supplementary fig. S13, Supplementary Material online). Conodipine-M, as a novel phospholipase A2, was first isolated from the venom of the marine snail *Conus magus* (McIntosh et al. 1995) and can inhibit the binding of isradipine to L-type calcium channels. In the basis of multiple alignments, the Conodipine-M-like peptide of *P. variabilis* shared a relatively high similarity with Conodipine-M of *Conus magus*. The functional domain scanning of Pfam and multiple sequence alignment revealed that the key domain region of this peptide was quite shorter than that of other common PLA2, while our *P. variabilis*' sequence has extra residues in the C-terminal region (supplementary fig. S13, Supplementary Material online). The two other predicted sequences were found more closely related to sequences from centipede and sea anemone toxic PLA2s (fig. 6). In hydra and sea anemones, PLA2s were characterized by structure and potency

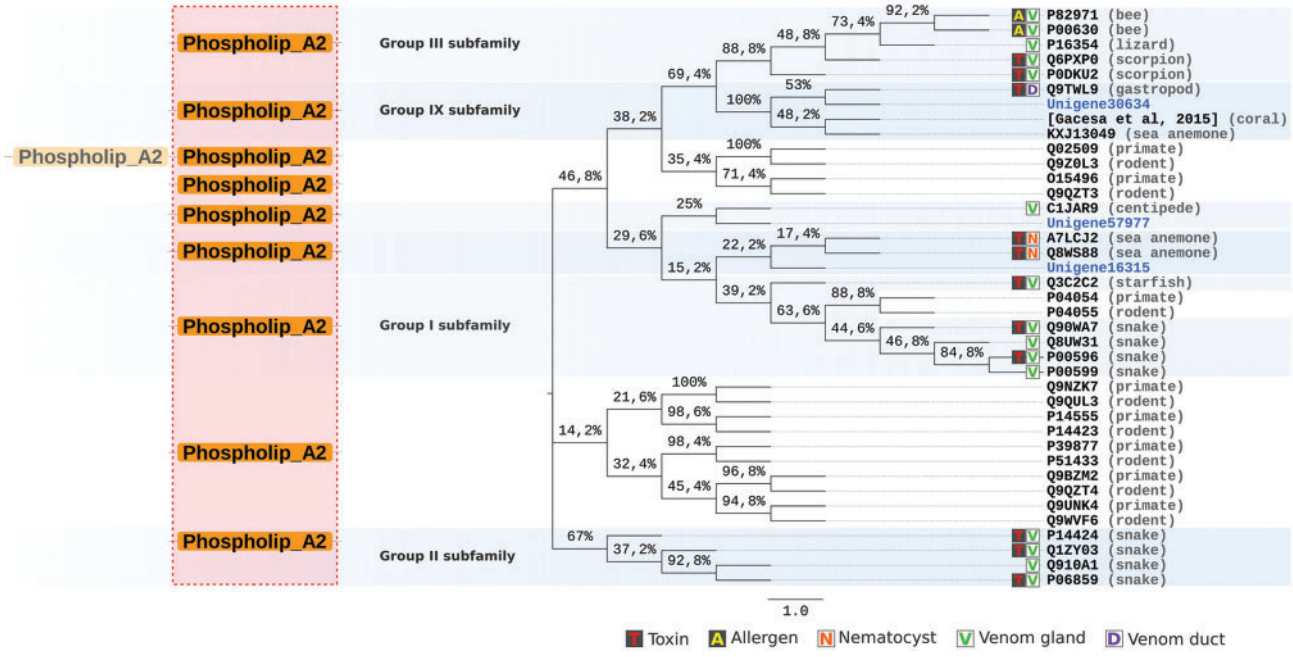

Fig. 6.—Phylogenetic analysis of the predicted PA2-like and PA2 Conodipine M-like. Same legend information like described in figure 2. Multiple sequence alignment of the Phospholip_A2 domains was achieved with Muscle. ML tree inferred based on a WAG + Gamma substitution model. Multiple sequence alignment of the Phospholip_A2 domain from the predicted *P. variabilis* protein sequence with selected toxins is available in the supplementary figure 13, Supplementary Material online.

Fig. 5.—Continued

the supplementary figure S10, Supplementary Material online. (*B*) Helofensin-like. Multiple sequence alignment of the Defensin_beta domains was achieved with Kalign. ML tree inferred based on a JTT + Gamma substitution model. Multiple sequence alignment of the Defensin_beta domain from the predicted *P. variabilis* protein sequence with selected toxins is available in the supplementary figure S11, Supplementary Material online. (*C*) MACPF-like. Multiple sequence alignment of the MACPF domains was achieved with Kalign. ML tree inferred based on a WAG + Gamma substitution model. Multiple sequence alignment of the MACPF domain from the predicted *P. variabilis* protein sequence with selected toxins is available in the supplementary figure S12, Supplementary Material online.

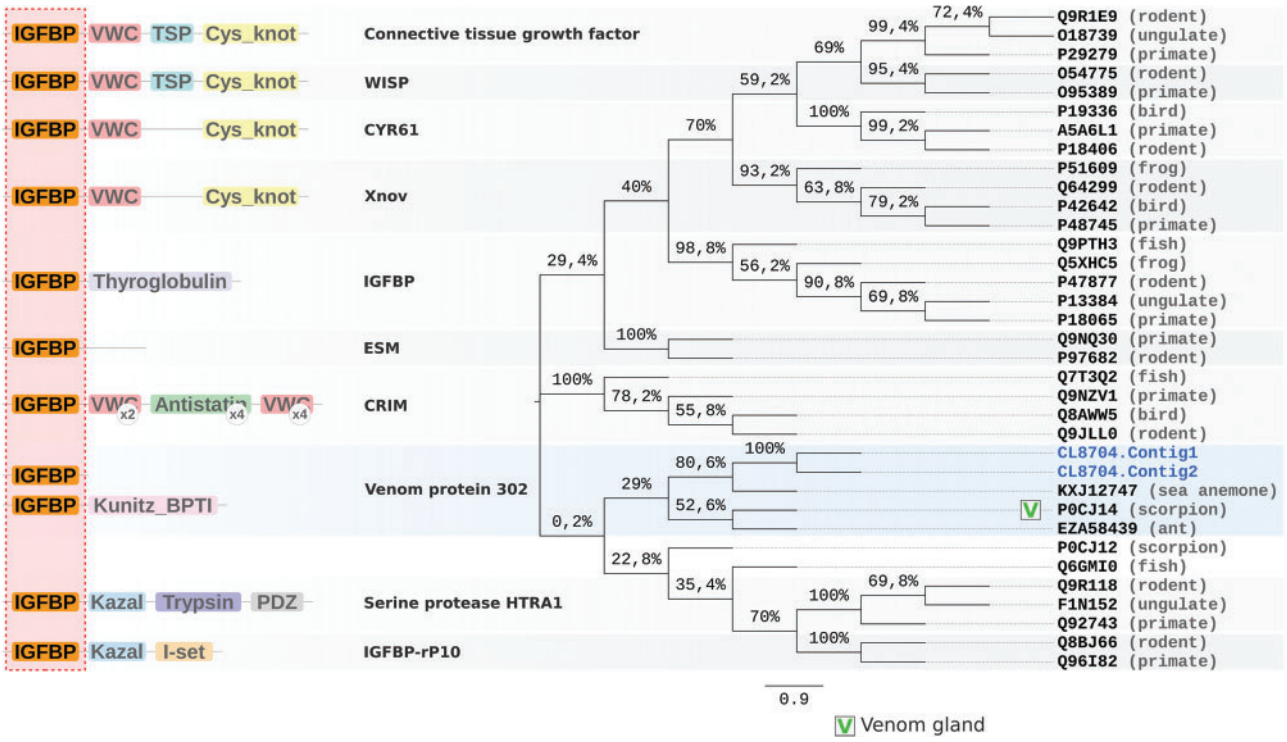

Fig. 7.—Phylogenetic analysis of the predicted venom protein 302-like. Same legend information like explained in figure 2. Multiple sequence alignment of the IGFBP domains was achieved with Muscle. ML tree inferred based on a WAG + I+Gamma substitution model. Multiple sequence alignment of the IGFBP domain from the predicted *P. variabilis* protein sequence with selected toxins is available in the supplementary figure 14, Supplementary Material online.

comparable to those of snake venom PLA2s (Sher and Zlotkin 2009; Frazao et al. 2012). In hydrozoa, PLA2 corresponds to type IV cytolysins, which were detected in nematocysts and tentacle extracts (Brinkman and Burnell 2009). Differential and high levels of PLA2 activity were also observed in the tissue extracts of four classes of cnidarians, which appear related to innate immunity against microbes and to the capture and digestion of prey (Nevalainen et al. 2004). Thus, due to the known broad spectrum of pharmacological activities of snake venom PLA2s (e.g., neurotoxic, myotoxic, necrotizing, hemolytic, anticoagulant, hypotension, and edematous), a presumed contribution of cnidarian PLA2s to a toxic weaponry of *P. variabilis* against predators should not be underestimated.

Venom Auxiliary Proteins

The putative protein sequences of venom protein 302 were herein identified and classified into a group of venom-auxiliary proteins from *P. variabilis* transcriptome (table 1). Venom-auxiliary proteins are assumed to be transcribed and compartmented in the venom glands or secreted in the venom to facilitate the secretion, processing, stabilization, or spreading of toxic peptide components. Venom protein 302 was initially identified from *Lychas mucronatus* (Ruiming et al. 2010). Expressed in the venom gland, it might be a kind of insulin-like growth factor binding protein. The predicted venom protein 302-like peptide of *P. variabilis* has a binding site similar to that of serine protease and to that of the single insulin-like growth factor binding domain protein. The two *P. variabilis* precursors, with another sea anemone sequence, form a new phylogenetic clade closely related to the venom protein 302 clade (fig. 7).

Toxicity Validation Assays Using Zebrafish Model

To evaluate the effectiveness of our toxin-mining strategy in elucidating the *P. variabilis* transcriptome and venom-like components, an in-house established *in vivo* model with zebrafish larvae was used to confirm the toxicity of putative *P. variabilis* ShK/Aurelin-like and novel anthozoan neurotoxin-like peptides. We first elected to validate ShK/Aurelin-like peptide (fig. 8) because it is a known potent neurotoxin from animals, such as sea anemones, that specifically blocks neural ion channels. The novel anthozoan neurotoxin-like peptide (fig. 9) contains a segment structurally identical to a V-shapped neurotoxin that was recently characterized from *Palythoa*, a genetically close relative to *P. variabilis* species.

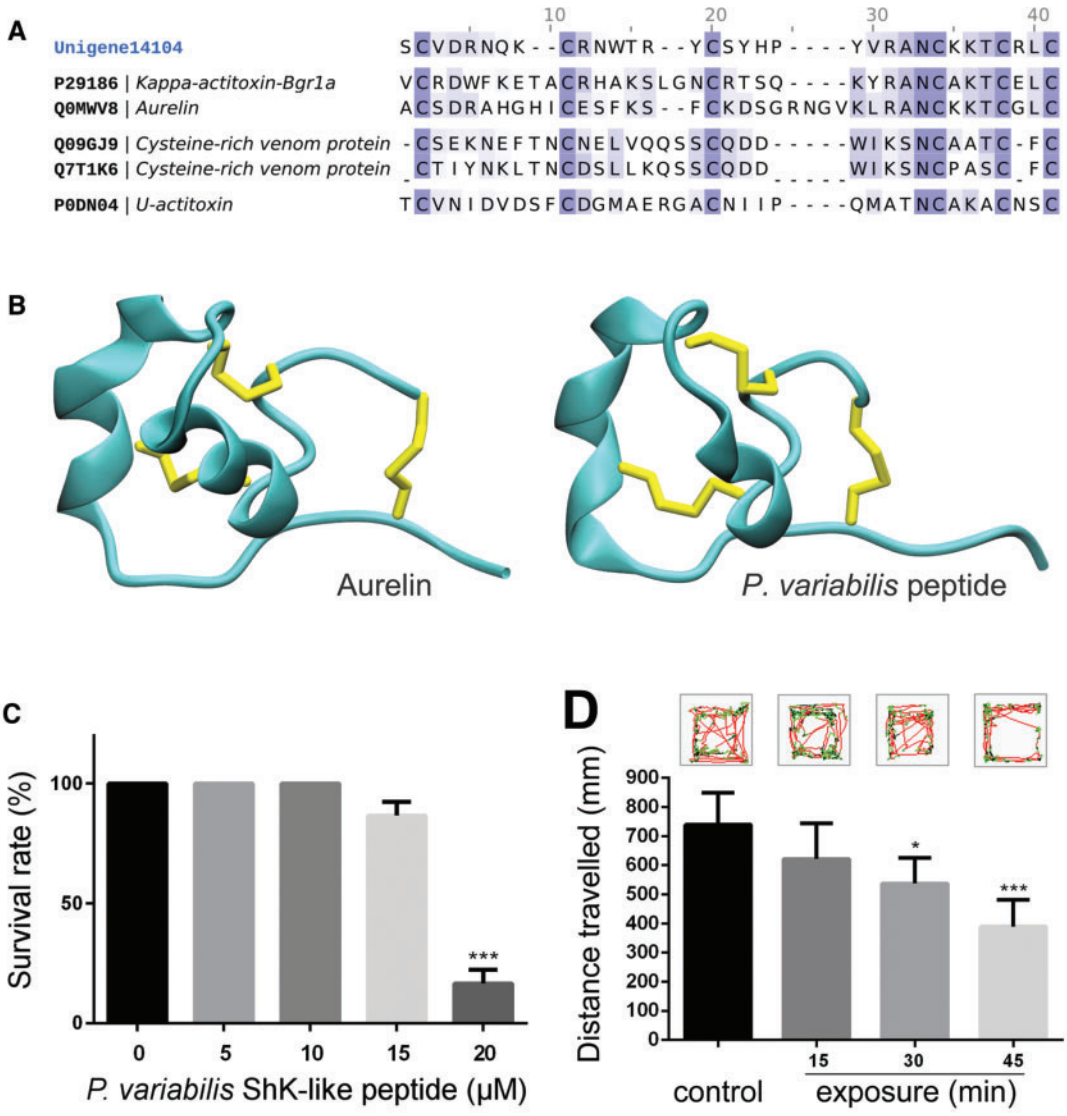

Fig. 8.—Toxicity test of ShK/Aurelin-like peptide from *P. variabilis* transcriptome. (*A*) Multiple sequence alignment of the ShK domain of the predicted *P. variabilis* poplypeptide with ShK toxins. (*B*) Display of the Aurelin toxin structure retrieved from PDB (PDB ID: 2LG4) and the model structure of the *P. variabilis* peptide predicted using MODELLER with the 2MCR PDB template. Disulfide bridges are in yellow. (*C*) Survival rates of zebrafish larvae (3 dpf) exposed to various concentrations of the *P. variabilis* ShK/Aurelin-like peptide for 12 h. Values are expressed as the means ± SEM. $***P < 0.001$ vs. control. The control group (0 μM) value was set to 100% and used to adjust the values from the other concentrations. (*D*) The *P. variabilis* ShK/Aurelin-like peptide attenuates the locomotion of zebrafish larvae. Zebrafish larvae at 6 dpf were incubated with 10 μM *P. variabilis* ShK/Aurelin-like peptide during for indicated durations. After treatment, the larvae were collected to carry out locomotion tests using the Viewpoint Zebrabox system, and the total distance moved in 10 min was computed (locomotion readout above each exposure time). Each treatment group contained 12 larvae, and three independent trials were performed for each larva. The results represent the mean distance traveled by the larvae. The values are expressed as the means ± SEM. $*P < 0.05$ and $***P < 0.001$ versus control (no exposure to the peptide) were considered statistically significant.

P. variabilis ShK/Aurelin-Like Peptide

The zebrafish toxicity test indicated that the mortality of the zebrafish increased linearly when they were exposed to augmented concentrations of synthetic *P. variabilis* ShK/Aurelin-like peptide (fig. 8C). No significant effect on the survival rate was observed in zebrafish groups that were exposed to peptides at 5 and 10 μM. However, treatment of the zebrafish with 15 and 20 μM of *P. variabilis* ShK/Aurelin-like peptide resulted in survival rates of 77% and 38%, respectively. None of the zebrafish survived when the peptide concentrations was 30 μM or higher. Therefore, the lethal dose of the *P. variabilis* ShK/Aurelin-like peptide LD50 that was calculated for zebrafish was between 15 and 20 μM. Furthermore, the

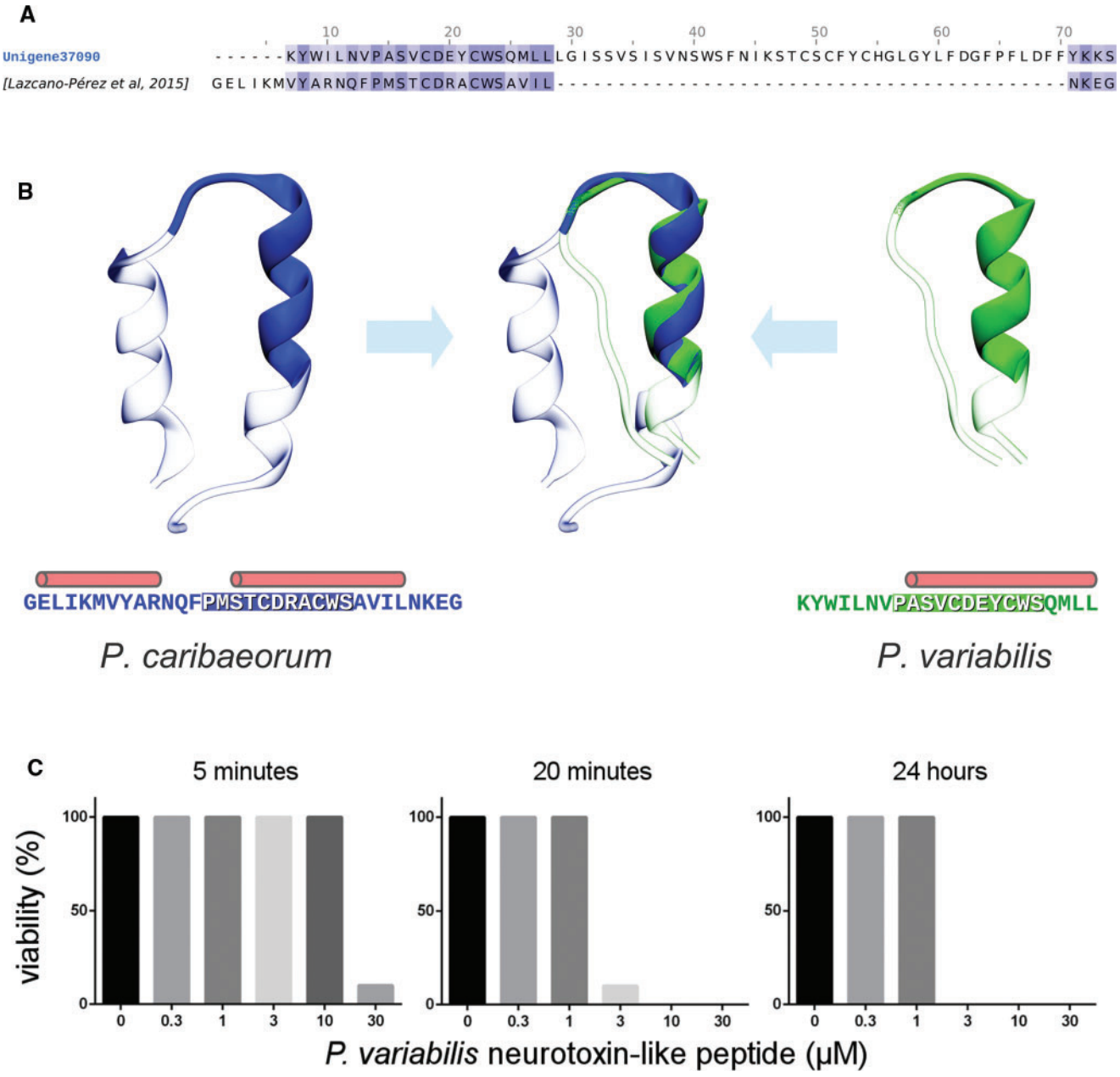

Fig. 9.—Toxicity test of the novel anthozoan neurotoxin-like peptide from *P. variabilis* transcriptome. (*A*) Alignment of the predicted *P. variabilis* poplypeptide similar to the recently reported Anthozoan neurotoxin. (*B*) Structure predictions for the full *P. caribaeorum* sequence and the corresponding aligned sequence of *P. variabilis*. Alpha helices are depicted above the sequences that were used for the structure predictions based on the PSIPRED (McGuffin, 2000) link results. Images of tertiary structure prediction with PEP-FOLD are shown, with the highly conserved region in full color. Structures superposition for both predicted structures are shown in the center. (*C*) Zebrafish larvae at 3 dpf were exposed to different concentrations of the novel anthozoan neurotoxin-like peptide (0, 0.3, 1, 3, 10, and 30 μM) for three different times (5 min, 20 min, and 24 h), respectively.

zebrafish locomotion test showed that the total distance that was travelled by the zebrafish larvae decreased significantly after exposure to 10 μM of peptide (figs. 8D and 9). These data indicate that the ShK/Aurelin-like peptide of *P. variabilis* is a potent neurotoxin that severely disturbs the locomotion (swimming) of zebrafish at concentrations as low as 10 μM.

P. variabilis Anthozoan Neurotoxin-Like Peptide

The toxicity test of 3-dpf fish for 5 min indicated no significant effect on the survival rate when the zebrafish groups were exposed to synthetic novel anthozoan neurotoxin-like peptides at 0.3, 1, 3, and 10 μM. However, treatment of the

zebrafish with 30 μM peptide resulted in a sharp decrease in the survival rate (fig. 9C "5 min"). The test of 3-dpf fish that were treated for 20 min showed that the fish were not affected by the peptides at 0.3 or 1 μM. The viability dramatically decreased when the 3-dpf fish were treated at 3 μM, and the lethality rate reached 100% when the 3-dpf fish were exposed to the peptides at 10 and 30 μM (fig. 9C "20 min"). Furthermore, the 3-dpf fish that were treated for 24 h at 0.3 and 1 μM were not affected, but none could survive at 3, 10, or 30 μM (fig. 9C "24 h").

Venom and Zoantharians

The majority of the venom-related transcripts that were retrieved from the *P. variabilis* transcriptome predicted precursors with detectable putative signal peptides (supplementary table S4, Supplementary Material online). The presence of leader sequences in these putative *P. variabilis* toxin precursors indicates that the presumed venom-like components are "tagged" for secretion. The identification of transcripts that code for venom auxiliary proteins, such as calglandulin, astacin-like metalloprotease and cystatin, suggests some level of specialization of *P. variabilis* tissue for the biogenesis of "venom" proteins. Interestingly, Anderluh et al. (2000) reported what they named the "cnidocyte signal": a consensus protein motif (*pp*X[DE]*h*EAKR, where *p* = polar; *h* = hydrophobic; and X = any amino acid residues) that would drive cnidarian toxins to cnidocytes in cnidoblast cells. We have tried to identify such a sorting signal within the full-length sequences of the *P. variabilis* putative toxin precursors, but no similar consensus motif was observed. However, the presence of a predicted signal peptide indicates that potential toxic polypeptide components might be compartmentalized in *P. variabilis* cnidocytes or might accumulate in specific organelles or groups of cells in anthozoan tissue. In fact, in sea anemones, toxins are not only located and released from nematocysts but also are maintained in other stores from which they are delivered, such as ectodermal gland cells (Anderluh and Macek 2002; Frazao et al. 2012). In cnidarians, as exemplified by hydrozoans, non-nematocystic toxins that cause cytolysis and paralysis are normally expressed in other tissues of the organism (Sher and Zlotkin 2009).

From a chemical ecology standpoint, the array of putative toxins that were undisclosed in the *P. variabilis* transcriptome are highly likely to be employed as an anti-predatory armamentarium rather than toxic formulations for active prey capture. Several species of fishes and turtles are reportedly potential predators of zoanthids, including reef fishes from the Abrolhos Bank in eastern Brazil (Francini-Filho and de Moura 2010) and the hawksbill turtle that forages on colonies of *P. caribaeorum* in southern Brazil (Lundin and Lindén 2007). Consequently, *P. variabilis* may rely on neurotoxic, hemorrhagic, and membranolytic toxins as molecular weapons to fend off predators. Moreover, the trophic ecology of *P. variabilis* demonstrated a suspensivorous feeding behavior (personal communication) that is similar to that of *Palythoa caribaeorum* along the Pernambuco Coast (Northeast Brazil) (de Santana et al. 2015). In contrast, the marine annelids of the genus *Glycera* that were once believed to be detritivorous were recently repositioned as an effective predator after the transcriptomic description of the complex cocktail of toxin precursors that are present in the *Glycera* venom gland (von Reumont et al. 2014). Although marine annelids and cnidarians belong obviously to relatively distant taxa and have, at first glance, distinct feeding behaviors, they encode several common toxic polypeptides, such as actinoporins and ShK-domain neurotoxins.

At present, there is a debate regarding how exactly toxins that originated from widespread body proteins duplicate and become restrictively expressed in venom gland tissues; there is a discussion comparing the duplication and recruitment hypothesis (convergent duplication) with the duplication and restriction explanation (Hargreaves et al. 2014). Considering this debate and the data from the present study, it appears plausible that the duplication of modular toxin-related domains, followed by restrictive functionalization to specialized venom cells and tissues, occurs within the realm of toxin evolution. Hence, the analysis of the repertoire of venom-related transcripts and their putative encoded polypeptides in basal marine organisms, such as the anthozoan *Protopalythoa*, offers important clues concerning how the early toxin arsenal evolved, how it develops within the body "compartments" of metazoan species, and how modern venomous animals restricted deadly toxins and venom proteins to specialized tissues. Last but not least, the identification of putative venom-related polypeptides and toxins in the *P. variabilis* transcriptome, such as actinoporin, MACPF, helofensin-like, also suggests that these putative polypeptide components might have derived from effectors of the innate immunity system with primordial roles of helping these marine organisms to thrive in a harsh aquatic environment.

Conclusion

The diversity of putative toxic polypeptide sequences in *Protopalythoa variabilis* based on a transcriptome analysis was greater than initially expected. The presumed toxic armamentarium of *P. variabilis* appears to have a complex representation of toxin families like seen in hydrozoans, jellyfishes and sea anemones. The functional validation of transcribed *P. variabilis* neurotoxin precursors (i.e., the novel ShK/Aurelin-like peptide and the anthozoan neurotoxin-like peptide) provides insight into the existence of a zoanthid pharmacological peptide arsenal that might function as deterrent of marine vertebrates (e.g., fishes and turtles). Evidently, a standardized comparative study using a similar analytical procedure would

provide a more complete understanding of the composition of such potential anthozoan toxic weapons and known toxin cocktails from other cnidarians. In summary, the putative *P. variabilis* toxin-related repertoire includes potentially neurotoxic components, hemorrhagic toxins and membrane-disrupting/pore-forming proteins that, in combination, might function as an efficacious anti-predatory formulation, inducing severe dysfunctions in the body system, such as neurotoxicity, hemorrhage, and generalized cytotoxicity. These effects are also potentially harmful to vertebrate victims that inadvertently come into contact with this apparently inoffensive marine organism that harbors a secret, ancient recipe of encoded peptide toxins and toxin-related protein precursors.

Supplementary Material

Supplementary figures S1–S14, tables S1–S5 and file S1 are available at *Genome Biology* and *Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

Research at the University of Macau was supported by grants from the Science and Technology Development Fund (FDCT) of Macau SAR (Ref. No. 134/2014/A3, FDCT078/2011/A3 and 069/2015/A2,) and Research Committee of University of Macau (MYRG2015-00214-ICMS-QRCM and MYRG2016-00133-ICMS-QRCM). Research work at the Institute for Marine Sciences, Federal University of Ceará, was supported by the Program on Toxinology (Issue 2010), the Coordination for the Improvement of Higher Education Personnel (CAPES), the Ministry of Education of the Federal Government of Brazil, and the Brazilian National Council for Scientific and Technological Development, CNPq (Marine Biotechnology Network—Proc. 408835/2013-3 and 408934/2013-1), the Ministry of Science, Technology and Innovation (MCTI). At the University Federal of Pernambuco, financial support was obtained from the Brazilian National Council for Scientific and Technological Development (CNPq/MCTI-PROTAX 2010) and FACEPE (APQ-2012). We are very thankful to Dr Álvaro R. B. Prieto-da-Silva and Dr André J. Zaharenko for their critical comments about cnidarian toxins and essential facts on molecular toxinology.

Literature Cited

Acosta A, Sammarco PW, Duarte LF. 2005. New fission processes in the zoanthid *Palythoa caribaeorum*: description and quantitative aspects. Bull. Mar. Sci. 76:1–26.

Almeida JG, et al. 2012. Palyosulfonoceramides A and B: unique sulfonylated ceramides from the Brazilian zoanthids Palythoa caribaeorum and Protopalythoa variabilis. Mar. Drugs 10:2846–2860.

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

Anderluh G, Podlesek Z, Macek P. 2000. A common motif in proparts of Cnidarian toxins and nematocyst collagens and its putative role. Biochim. Biophys. Acta. 1476:372–376.

Bakrac B, et al. 2008. Molecular determinants of sphingomyelin specificityof a eukaryotic pore-forming toxin. J Biol Chem. 283:18665–18677.

Barradas JI, Amaral FD, Hernández MIM, Montes MJF, Steiner AQ. 2010. Spatial distribution of benthic macroorganisms on reef flats at Porto de Galinhas Beach (northeastern Brazil), with special focus on corals and calcified hydroids. Biotemas 23:61–67.

Bellocci M, Sala GL, Prandi S. 2011. The cytolytic and cytotoxic activities of palytoxin. Toxicon 57:449–459.

Boscolo HK, Silveira FL. 2005. Reproductive biology of *Palythoa caribaeorum* and *Protopalythoa valilbilis* (Cnidaria, Anthozoa, Zoanthidea) from the southeastern coast of Brazil. Braz. J. Biol. 65:29–41.

Brinkman DL, Burnell JN. 2009. Biochemical and molecular characterisation of cubozoan protein toxins. Toxicon 54:1162–1173.

Brinkman DL, et al. 2015. Transcriptome and venom proteome of the box jellyfish *Chironex fleckeri*. BMC Genomics 16:407.

Castaneda O, et al. 1995. Characterization of a potassium channel toxin from the Caribbean Sea anemone *Stichodactyla helianthus*. Toxicon 33:603–613.

Chapman JA, et al. 2010. The dynamic genome of Hydra. Nature 464:592–596.

Choo KH, Tong JC, Zhang L. 2004. Recent applications of Hidden Markov Models in computational biology. Genomics Proteomics Bioinformatics 2:84–96.

Correa-Netto C, et al. 2011. Snake venomics and venom gland transcriptomic analysis of Brazilian coral snakes, *Micrurus altirostris* and *M. corallinus*. J. Proteomics 74:1795–1809.

Daly M, et al. 2007. The phylum Cnidaria: a review of phylogenetic patterns and diversity 300 years after Linnaeus. Zootaxa 1668: 127–182.

de Santana EFC, et al. 2015. Trophic ecology of the zoanthid Palythoa caribaeorum (Cnidaria: Anthozoa) on tropical reefs. J. Mar. Biol. Assoc. UK 95:301–309.

Deeds JR, Handy SM, White KD, Reimer JD. 2011. Palytoxin found in Palythoa sp. zoanthids (Anthozoa, Hexacorallia) sold in the home aquarium trade. PLoS One 6:e18235.

Duda TF, Kohn AJ, Palumbi SR. 2001. Origins of diverse feeding ecologies within Conus, a genus of venomous marine gastropods. Biol. J. Linn. Soc. 73:391–409.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Eswar N, et al. 2006. Comparative protein structure modeling using Modeller. Curr. Protoc. Bioinformatics 5:6.1–6.30.

Fautin DG, Daly M. 2009. Actiniaria, Corallimorpharia, and Zoanthidea (Cnidaria) of the Gulf of Mexico: Gulf of Mexico Origin, Waters, and Biota. Biodiversity 1:349.

Fautin DG. 2009. Structural diversity, systematics, and evolution of cnidae. Toxicon 54:1054–1064.

Francini-Filho RB, de Moura RL.2010. Predation on the toxic zoanthid *Palythoa caribaeorum* by reef fishes in the Abrolhos Bank, eastern Brazil. Brazil. J. Oceanogr. 58:77–79.

Frank K, Sippl MJ. 2008. High-performance signal peptide prediction based on sequence alignment techniques. Bioinformatics 24: 2172–2176.

Frazao B, Vasconcelos V, Antunes A. 2012. Sea anemone (Cnidaria, Anthozoa, Actiniaria) toxins: an overview. Mar. Drugs 10:1812–1851.

Gunji K, Ishizaki S, Shiomi K. 2010. Cloning of complementary and genomic DNAs encoding echotoxins, proteinaceous toxins from the salivary gland of marine gastropod *Monoplex echo*. Protein J. 29:487–492.

Haas BJ, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8:1494–1512.
Hargreaves AD, Swain MT, Hegarty MJ, Logan DW, Mulley JF. 2014. Restriction and recruitment—gene duplication and the origin and evolution of snake venom toxins. Genome Biol. Evol. 6:2088–2095.
Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14:33–38, 27–38.
Kawashima Y, Nagai H, Ishida M, Nagashima Y, Shiomi K. 2003. Primary structure of echotoxin 2, an actinoporin-like hemolytic toxin from the salivary gland of the marine gastropod *Monoplex echo*. Toxicon 42:491–497.
Kini RM, Doley R. 2010. Structure, function and evolution of three-finger toxins: mini proteins with multiple targets. Toxicon 56:855–867.
Kini RM. 2003. Excitement ahead: structure, function and mechanism of snake venom phospholipase A2 enzymes. Toxicon 42:827–840.
Kini RM. 2011. Evolution of three-finger toxins–a versatile mini protein scaffold. Acta Chim. Slov. 58:693–701.
Konc J, Janezic D. 2012. ProBiS-2012: web server and web services for detection of structurally similar binding sites in proteins. Nucleic Acids Res. 40:W214–W221.
Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol. 3:1870–1874.
Lassmann T, Frings O, Sonnhammer EL. 2009. Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. Nucleic Acids Res. 37:858–865.
Lazcano-Perez F, et al. 2014. A purified Palythoa venom fraction delays sodium current inactivation in sympathetic neurons. Toxicon 82:112–116.
Lazcano-Perez F, Roman-Gonzalez SA, Sanchez-Puig N, Arreguin-Espinosa R. 2012. Bioactive peptides from marine organisms: a short overview. Protein Pept. Lett. 19:700–707.
Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595.
Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.
Lundin C, Lindén O. 2007. Predation on the zoanthid Palythoa caribaeorum (Anthozoa, Cnidaria) by a hawksbill turtle (*Eretmochelys imbricata*) in Southeastern Brazil. Marine Turtle Newsletter: 1:117.
Mackie G. 2002. What's new in cnidarian biology? Can. J. Zool. 80: 1649–1653.
Mariscal RN. 1974. Nematocysts. In: Muscatine L, Lenhoff HM, editors. Coelenterate biology: reviews and new perspectives. New York: Academic Press. p. 129–178.
Marques AC, Collins AG. 2004. Cladistic analysis of Medusozoa and cnidarian evolution. Invertebr. Biol. 123:23–42.
McGuffin LJ, Bryson K, Jones DT. 2000. The PSIPRED protein structure prediction server. Bioinformatics 16:404–405.
McIntosh JM, et al. 1995. Conodipine-M, a novel phospholipase A2 isolated from the venom of the marine snail *Conus magus*. J. Biol. Chem. 270:3518–3526.
Mihelic M, Turk D. 2007. Two decades of thyroglobulin type-1 domain research. Biol. Chem. 388:1123–1130.
Mount DW. 2007. Using the Basic Local Alignment Search Tool (BLAST). CSH Protoc 2007: pdb top17.
Mourao CB, Schwartz EF. 2013. Protease inhibitors from marine venomous animals and their counterparts in terrestrial venomous animals. Mar. Drugs 11:2069–2112.
Moya A, et al. 2012. Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to CO(2)-driven acidification during the initiation of calcification. Mol. Ecol. 21: 2440–2454.
Murakami M, Kudo I. 2002. Phospholipase A2. J. Biochem. 131:285–292.
Nevalainen TJ, et al. 2004. Phospholipase A2 in cnidaria. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 139:731–735.
Oshiro N, et al. 2004. A new membrane-attack complex/perforin (MACPF) domain lethal toxin from the nematocyst venom of the Okinawan sea anemone *Actineria villosa*. Toxicon 43:225–228.
Ovchinnikova TV, et al. 2006. Aurelin, a novel antimicrobial peptide from jellyfish Aurelia aurita with structural features of defensins and channel-blocking toxins. Biochem. Biophys. Res. Commun. 348: 514–523.
Pertea G, et al. 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19:651–652.
Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8:785–786.
Polato NR, Vera JC, Baums IB. 2011. Gene discovery in the threatened elkhorn coral: 454 sequencing of the Acropora palmata transcriptome. PLoS One 6:e28634.
Putnam NH, et al. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317:86–94.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.
Rabelo EF, Soares MdO Matthews-Cascon H. 2013. Competitive interactions among zoanthids (Cnidaria: Zoanthidae) in an intertidal zone of Northeastern Brazil. Brazil. J. Oceanogr. 61:35–42.
Rodriguez E, et al. 2014. Hidden among sea anemones: the first comprehensive phylogenetic reconstruction of the order Actiniaria (Cnidaria, Anthozoa, Hexacorallia) reveals a novel group of hexacorals. PLoS One 9:e96998.
Rohlfs De Macedo C, Belém M. 1994. The genus Zoanthus in Brazil. 1. Characterization and anatomical revision of *Zoanthus sociatus* (Cnidaria, Zoanthinaria, Zoanthidae). Iheringia 77:135–144.
Ruiming Z, et al. 2010. Comparative venom gland transcriptome analysis of the scorpion *Lychas mucronatus* reveals intraspecific toxic gene diversity and new venomous components. BMC Genomics 11:452.
Ryland JS, Muirhead MA. 1993. Order Zoanthidea (Class Anthozoa, Zoantharia). In: Mather P, Bennett I, editors. A coral reef handbook. 3rd edn. Sydney: Surrey Beatty& Son Pty Ltd. p. 52–58.
Sher D, Zlotkin E. 2009. A hydra with many heads: protein and polypeptide toxins from hydra and their biological roles. Toxicon 54: 1148–1161.
Shimizu Y. 1983. Complete structure of palytoxin elucidated. Nature 302:212.
Shinzato C, et al. 2011. Using the *Acropora digitifera* genome to understand coral responses to environmental change. Nature 476:320–323.
Thevenet P, et al. 2012. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. Nucleic Acids Res. 40:W288–W293.
Tirosh Y, Ofer D, Eliyahu T, Linial M. 2013. Short toxin-like proteins attack the defense line of innate immunity. Toxins (Basel) 5(7): 1314–1331.
Tsetlin VI. 2015. Three-finger snake neurotoxins and Ly6 proteins targeting nicotinic acetylcholine receptors: pharmacological tools and endogenous modulators. Trends Pharmacol. Sci. 36(2): 109–123.
Utkin YN. 2013. Three-finger toxins, a deadly weapon of elapid venom–milestones of discovery. Toxicon 62:50–55.
von Reumont BM, et al. 2014. A Polychaete's powerful punch: venom gland transcriptomics of Glycera reveals a complex cocktail of toxin homologs. Genome Biol. Evol. 6:2406–2423.
Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191.

Wenger Y, Galliot B. 2013. RNAseq versus genome-predicted transcriptomes: a large population of novel transcripts identified in an Illumina-454 Hydra transcriptome. BMC Genomics 14:204.
Wilke DV, et al. 2009. Cytotoxic lipidic α-amino acids from the zoanthid *Protopalythoa variabilis* from the northeastern coast of Brazil. J. Brazil. Chem. Soc. 20:1455–1459.
Wilke DV, et al. 2010. Pro-apoptotic activity of lipidic alpha-amino acids isolated from *Protopalythoa variabilis*. Bioorg. Med. Chem. 18:7997–8004.
Wu CH. 2009. Palytoxin: membrane mechanisms of action. Toxicon 54:1183–1189.
Yamazaki Y, et al. 2009. Snake venom Vascular Endothelial Growth Factors (VEGF-Fs) exclusively vary their structures and functions among species. J. Biol. Chem. 284:9885–9891.
Yoffe C, Lotan T, Benayhau Y. 2012. A modified view on octocorals: *Heteroxenia fuscescens* nematocysts are diverse, featuring both an ancestral and a novel type. PLoS One 7:e31902.

Associate editor: Rebecca Zufall

Highlights editor: George Zhang