UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

GENOMIC INVESTIGATION OF CNIDARIAN TOXIN EVOLUTION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

ANUJ GURUACHARYA
Norman, Oklahoma
2017

GENOMIC INVESTIGATION OF CNIDARIAN TOXIN EVOLUTION

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF BIOLOGY

BY

_____
Dr. Richard Broughton, Chair

_____
Dr. Lawrence Weider

_____
Dr. JP Masly

_____
Dr. Michael Markham

_____
Dr. Cecil Lewis

Dedication

To my family and friends.

Acknowledgement

I would like to thank Richard Broughton for his helpful advice throughout my education. He has also helped me through one year of research assistantship and provided a space and motivation to work. I would like to thank Christian Lemon, who has been my friend and helped me better understand statistics in a biological concept. I would also like to thank David Durica, who has listened to some of my work and been supportive. I would like to thank Larry Weider for his very helpful comments on the early draft of the first two chapters and his moral support. I would also like to thank JP Masly, who has provided me some useful advice, which I unfortunately, forgot to heed. I would also like to thank my advisors Michael Markham and Cecil Lewis, who agreed to serve in my committee, despite their busy schedule.

I would also like to thank Dahang Yu and Si Wu, who helped me a lot of the proteomics portions of the project and also helped me in writing portions of the manuscript for proteomics.

I would like to thank my friend Binita Rajbanshi for discussing most of chapter 3 with me and helping me in the formatting of the document. I would like to thank Sharmistha Shyamal for listening to me create my initial hypothesis and showing me wet lab procedures. I have benefitted from conversations with her as equally as she has benefitted from me. I would like to thank Krithi Sankanarayan for discussing some of the procedures for my initial studies.

Table of Contents

List of Tables

Chapter 1

Chapter 2

Chapter 3

List of Figures

Chapter 1

Chapter 2

Chapter 3

Abstract

The phylum Cnidaria (jellyfish, hydras and anemones) was one of the earliest diverging animal groups and member species have simple, diffuse nervous systems. A trait unique to Cnidaria is specialized stinging cells (nematocysts) that are considered as part of the nervous system. Nematocysts inject complex venoms which include a diverse set of neurotoxins that bind to and block voltage-gated ion channel genes. Perhaps not coincidentally, cnidarians also possess more voltage-gated ion channel genes than any other animal group. Like other cnidarians, sea anemones use their nematocysts to secure small animal prey, yet a group of potential prey fishes (clownfishes) have evolved symbiotic relationships with anemones and live and breed within anemone tentacles. This symbiotic relationship is not well understood.

I used genomic, transcriptomic, and proteomic approaches in three studies to investigate the evolution of cnidarian toxins and potential mechanisms of anemone-clownfish symbiosis. The first study investigates the potential co-evolution of neurotoxins and ion channels in cnidarians. The second study explores toxin gene and protein diversity in clownfish-hosting and non-hosting sea anemones. The third study examines tissue-specific expression of toxin genes in clownfish-hosting and non-hosting sea anemones. Among the 36 cnidarian species investigated in the first study, neurotoxin and ion channel proteins showed phylogenetic evidence of co-evolution. In the second study, toxin diversity was found to be higher in the anemone that hosted clownfish than the one that did not host clownfish; however, in third study, it was found that the overall

expression level of the toxins was found to be lower (both in the tentacles and column) of the clownfish hosting anemone than in anemone that did not host a clownfish.

These results suggest that cnidarian neurotoxins co-evolved with their target ion channels; and that anemone-clownfish symbiosis maybe related to lowered expression levels of certain toxins in the anemone.

Introduction

The phylum Cnidaria is composed of animals such as jellyfish, box jellies, hydra, sea anemones, and corals. Cnidarians were one of the first phylum to evolve a primitive nervous system. How the diversity of neural proteins evolved in early animal forms is still unexplored. Cnidarians are also the only phyla that have stinging cells (nematocysts) that are considered as part of the nervous system. These stinging cells are specialized for toxin injection; neurotoxins being one of the major toxins used by this animal group. The broad objective of this dissertation is to study the evolution of such neurotoxins.

In chapter 1, I have explored the evolutionary diversity of ion channels in cnidarians. There was a burst of evolutionary innovation in the genes responsible for nervous system function in the earliest animal ancestors. Even though the cnidarian lineage wasn't the first to evolve ion channels, this group has a large diversity of ion channels that are used in its nervous tissue. The factors that led to this great diversity remain unknown. In an effort to answer this question, I have tested the hypothesis that ion channels and neurotoxins coevolved in cnidarians.

Despite having a potent mix of toxins, sea anemones, which are a major group of cnidarians, serve as hosts for various animals which are able to live among their stinging tentacles. In chapter 2, I have investigated the toxin content of two sea anemone species: one that is symbiotic with clownfish, another that is not. The

mechanisms that allow some anemones to host many species of fish while others host a few, or only one, are not clearly understood. In an effort to understand such mechanism, I have tried to identify potential toxins in two different sea anemone transcriptomes: one that hosts clownfish (*Entamaea quadricolor)* and one that does not (*Condylactis gigantea).*

In chapter 3, I have investigated the toxin content and expression level in different tissues of two sea anemones: *Entamaea quadricolor* and *Condylactis gigantea*. Cnidarians are the simplest animals to have tissue level differentiation. They however do not have a centralized nervous system as bilaterans nor do they have a centralized venom delivery system like snakes and spiders. Even though tentacles of sea anemones have been the most widely used tissue from which toxins have been isolated, their toxin composition relative to other tissues is not well known. In an effort to understand such tissue level differences, I have identified differentially expressed toxins and proteins in the tentacle tissue and column tissue of the two sea anemones that had been previously examined in chapter 2.

All the code and data that is generated from these projects has been deposited on www.github.com/anuj2054. Raw genomic data generated from this dissertation has been deposited in the European Bioinformatics Institute (EBI) database that is also cross-referenced with the National Center for Biotechnology Institute (NCBI) Short Read Archive (SRA) database.

Chapter 1: Co-evolution of ion channels and neurotoxins in cnidarians

Abstract

Understanding the diversity of ion channels in cnidarians may shed light on the origin and evolution of early nervous systems. It is hypothesized that variation in cnidarian neurotoxins led to the evolution of diverse ion channel proteins in the same animal phylum. I tested this hypothesis by investigating several evolutionary factors of both cnidarian neurotoxins and their target ion channels. I examined homologs of 250 cnidarian toxins, 75 ion channel genes, and 70 housekeeping genes from 36 transcriptomes/genomes of cnidarian species. Correlation analysis based on annotation, selection analysis, evolutionary rate analysis, and gene tree – species tree reconciliation analysis were performed on the homologs of neurotoxin and ion channel proteins. Results indicated 1) evidence of positive selection; 2) correlation between the number of homologous toxin genes and ion channel genes; and 3) difference in the evolutionary rates between toxin genes and ion channel genes. I show for the first time that neurotoxins are likely to have co-evolved with the ion channels in cnidarians. This is consistent with an evolutionary arms race between ion channels and neurotoxins leading to the extensive diversity of ion channel genes found in cnidarians.

Keywords: Phylogenetics, Early animals, Cnidaria, Evolution, Nervous systems

Introduction

Cnidarians may provide important clues to the evolution of nervous systems because of their position as one of the earliest diverging lineages of animals exhibiting a rudimentary nervous system (Bucher and Anderson, 2015). Nervous systems allow animals to integrate sensory information and translate this information into behavior. It

has been suggested that neurons could have provided early animals with the ability to control a hydrostatic skeleton such as elongating or contracting the body or the ability to open or close feeding appendages according to sensory cues (Anderson, 2015). Neurons are also likely to have played an integral role in the evolution of muscle tissue. Yet the origin and evolution of early nervous systems remain obscure (Kelava et al., 2015). Taxa in the phylum Cnidaria (e.g. sea anemones, hydras, and jellyfishes) have a simple diffuse nervous system unlike the centralized nervous system virtually all other animals (the Bilateria). Comparative anatomical studies spanning more than 150 years point to the common origin of the nervous systems in the ancestor of Bilateria and Cnidaria, with centralization evolving in the bilaterian lineage. Despite the relative simplicity of their nervous systems, cnidarians have undergone a lineage specific expansion of genes for voltage-gated ion channels as shown in Figure 1 (Moran et al., 2014; Li et al., 2015; Liebeskind et al., 2015). Voltage-gated ion channels are the primary regulators of ion movement across the membranes of neurons and other excitable cells and are therefore fundamental to action potential formation and signal specificity. Because the expansion of sodium ion channel subtypes in vertebrates appears to be correlated with increased neuronal complexity (Barzilai et al., 2012), it has been suggested by Moran et al. in 2015 that the expansion of ion channel types in Cnidaria might also correlate with increased neuronal complexity (Moran et al., 2015). But given the simple nervous systems of cnidarians, the nature of such possible neuronal complexity is unclear. The evolutionary factors that have driven ion channel diversification in cnidarians, thus, remain poorly understood.

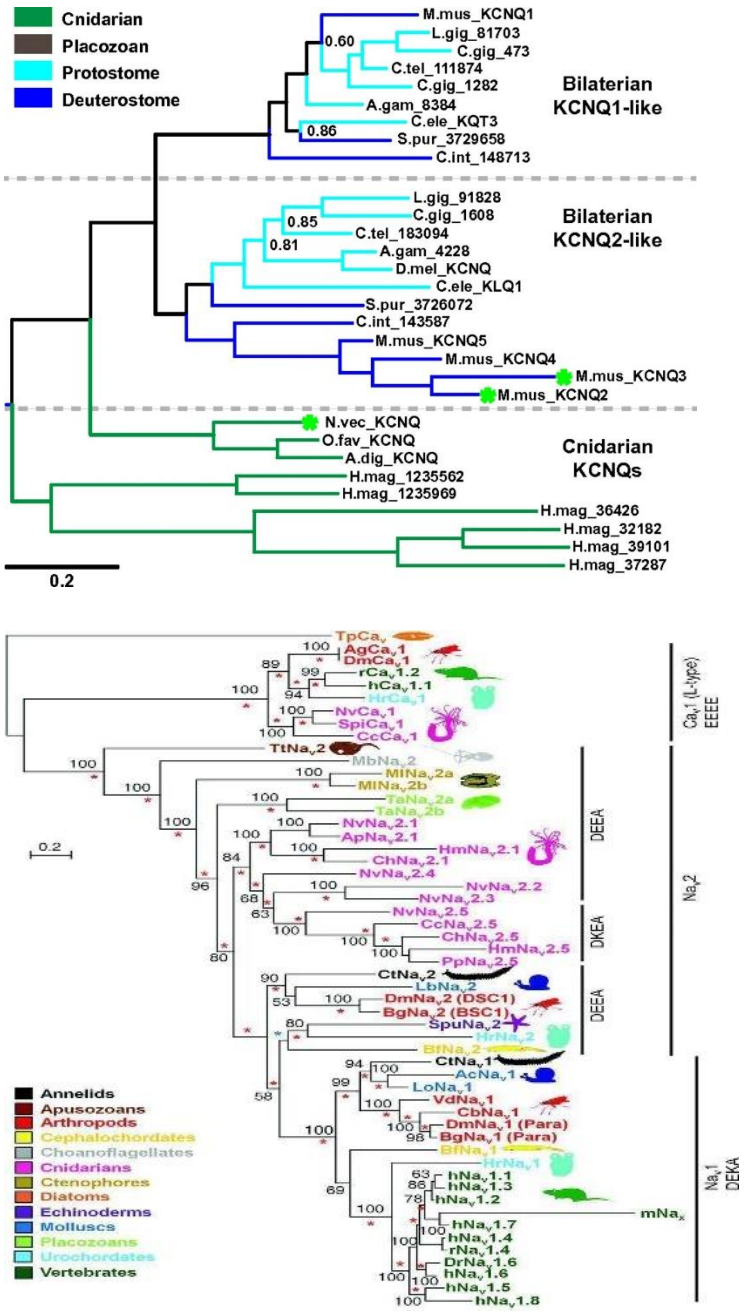


Figure 1. Gene family phylogeny of (top) KCNQ (K_v7) voltage-gated potassium channels and (bottom) sodium ion channels in metazoans. Cnidarians are represented by green color in potassium ion channel gene phylogeny (top) and by pink color in sodium ion channel gene phylogeny (bottom). These figures were reproduced with permission from Barzilai et al., Cell reports, 2012 and Li et. al., PNAS, 2015.

The diversity of ion channels may be related to the broad diversity of toxins that cnidarians produce. Cnidarians are the only phyla possessing nematocysts. These are

stinging cells specialized for toxin injection and are considered as part of the nervous tissue (Jouiaei et al., 2015; Frazao et al., 2012). Nematocysts inject a cocktail of various peptide and non-peptide toxins, with neurotoxins being major components of the mix (Torres-Ramos et al., 2003). Cnidarian peptide neurotoxins specifically bind to voltage-gated ion channels (Messerli et al., 2006), thereby inhibiting nervous system function (Lazcano-Perez et al., 2016). Sodium channel neurotoxins and potassium channel neurotoxins are the two best characterized toxin groups in these animals (Moran et al., 2009; Castaneda and Harvey, 2009).

Previous studies have explored phylogenetic analyses of either cnidarian nervous system subunits or cnidarian neurotoxins (Sunagar et al., 2016; Sunagar and Moran, 2015; Jouiaei, 2016; Rachamim et al., 2014; Jouiaei et al., 2015), but a systematic investigation of the evolution of both channels and toxins has not yet been reported. With the rise of genomics and the increasing number of cnidarian nucleotide sequences available (Technau and Schwaiger, 2015), it has been possible to address questions about ion channel and neurotoxin diversity. Only recently have sufficient cnidarian genome and transcriptome sequences become available to test different hypotheses by examining the homologs of toxins and neural proteins. It has been suggested that evolutionary arms races (Van Valen, 1977; Dawkins and Krebs, 1979) between predators and prey lead to increasingly potent peptide toxins as well as repeated compensatory changes to defense against those toxins, causing the evolutionary diversification of both members of interacting protein pairs (Harris and Arbuckle, 2016). Well known examples include predator-prey pairs such as grasshopper mice and

scorpions (Rowe, 2004), and garter snakes and newts (Brodie et al., 2002). Like the co-evolving proteins in these predator-prey species, I predicted that ion channels in ancestral cnidarians could have diversified because of natural selection favoring resistance to specific neurotoxins. Diversification of these ion channels may also have been induced by competitive encounters with other cnidarians having a different neurotoxin cocktail than their own.

Thus, I hypothesized that selective pressure to resist the deleterious effects of neurotoxins may have led to diversification of ion channels in early cnidarians. Such diversification could have involved both nucleotide substitutions as well as gene duplication to expand gene families. If gene duplications have occurred in both toxins and ion channels due to their interactions, then I might predict a correlation between the number of ion channel and neurotoxin genes in different cnidarian species. In addition, if changes in one member of the interacting pair result in compensatory changes in the other, I would expect their evolutionary rates to be correlated along the various lineages. To explore this hypothesis, I acquired genome or transcriptome sequences from 39 diverse cnidarian species. These data were used to construct a species phylogeny of cnidarians which was used to perform gene tree analysis, character analysis on homologous genes, tests of positive selection, and analyses of evolutionary rates of ion channels and neurotoxins.

Methods

Data collection

Genome and transcriptome sequences were acquired for 36 different cnidarian species from three different sources (Supplementary Table 1) : NCBI genome database, NCBI (Transcriptome Shotgun Assembly) TSA nucleotide database, and unpublished transcriptome assemblies (people.oregonstate.edu/~meyere/data.html). Genomes of a sponge (*Amphimedon queenslandica*) and fruit fly (*Drosophila melanogaster*) were also obtained for use as outgroups from the NCBI genome database. A set of 70 vertebrate housekeeping genes (Warrington et al., 2000) (Supplementary Table 2) were entered into Uniprot and clustered using a percent similarity identity of 50% to other species via the Uniref50 tool. Only the representative sequence from each cluster was used as the query sequence for blast searches of genomes and transcriptomes. Amino acid sequences for cnidarian venom proteins were collected from venomzone.expasy.org (Supplementary Table 3). Ion channel genes were collected from Uniprot with the GO terms: “sodium ion channel”,” potassium ion channel”,” calcium ion channel”. These channels were then clustered with 50% similarity and only their representative sequences were used for further analysis (Supplementary Table 4). The gene families used were of voltage-gated sodium, potassium, and calcium channels; actinoporin toxins; Small Cysteine Rich Protines (SCRIPS) toxins; jellyfish toxins; sodium channel neurotoxin type I and type II; and potassium channel neurotoxin type 1, 2, 3, and 5.

Species tree

A phylogenetic tree was constructed for all species using the amino-acid sequences of 70 housekeeping gene sequences. The 70 housekeeping genes were reciprocally blasted against the genome and transcriptome collection using tblastn and blastx to obtain homologous sequences of each species for each housekeeping gene. The E value used as cut-off for both the blast searches was -10. These genes were then aligned using ClustalO (Sievers et al., 2012), then manually trimmed and aligned again. The genes were then concatenated. Partition Finder (Lanfear et al., 2016) was used to partition the concatenated data into one partition per gene and to find the appropriate evolutionary model for each partition. The LGX model (Le and Gascuel, 2008) was found to be best model using Akaike information criterion for all the housekeeping genes. The sponge and fruit fly were included as outgroups. Maximum likelihood was used to find the best phylogenetic tree with RAxML (Stamatakis, 2014). Rapid bootstrapping of 100 replicates was performed using the -a option in RAxML. Mr. Bayes (Ronquist and Huelsenbeck, 2003) was used to construct the best supported species tree using a Bayesian phylogenetic approach. The rate variation parameter was gamma with 4 rate categories. The chain length for MCMC was 1,100,000 with a subsampling frequency of 200 and a burn in length of 100,000. Treegraph (Stöver and Müller, 2010) was used to visualize and annotate the species tree with branch lengths and support values.

Gene tree-species tree reconciliation

In order to assess gene family history, individual gene trees were compared with the species tree. The protein data set was queried against the genomes/transcriptomes with an E value of -10 using tblastn. They were then reciprocally searched using blastx

against the hits obtained from tblastn. The blast hits for the proteins were grouped according to the gene family they belonged to. The amino-acid sequences of the resulting hits were aligned using ClustalW for each protein resulting in 252 multiple sequence alignments. For each gene family, gene alignments were grouped into single file. The best protein evolution model was found using jModelTest (Darriba et al., 2012) for each of the alignments to be Gamma WAG (Whelan and Goldman, 2001). Maximum likelihood gene trees were estimated using RAXML with 100 bootstrap replicates on the potassium channel, sodium channel, calcium channel, potassium toxin and scrip toxin families. Reconciliation analysis used maximum parsimony to estimate the minimum number of gene duplications and losses using Notung (Chen et al., 2000). An edge weight threshold of 1.4255 was used. Costs/weights were set to 1.5 for duplication, 0 for co-divergence, and 1.0 for losses.

Gene presence/absence analysis

Correlation analysis was performed based on homologous gene presence or absence data among different species. A character matrix was created using the proteins that were obtained from reciprocal blast for each species. Correlations in the character state matrix were investigated using Pearson's correlation method in R (R Core Team, 2000). A linear regression was performed between the total number of channels and the total number of neurotoxins in each species.

Selection analysis

Gene homologs of the ion channels and neurotoxins were input into Codeml from the PAML package (Yang, 2007) for selection analysis based on non-synonymous

substitution rate by synonymous substitution rate (dN/dS) ratios. Several of the genes were present in so few species that selection analyses were not possible. For genes that were present in many but not all species, species trees that were pruned to match the species present were used as the input topologies. A custom script was used to prune the species tree to fit the number of animals in each gene alignment. The alignments were manually curated to verify that they were in frame. I ran random sites models M0, M1, M2, M3, M7, and M8 found in PAML. Selection was inferred from likelihood ratio tests comparing M1 vs. M2 and M7 vs. M8. Likelihood Ratio Tests (LRTs) were performed on M0 vs M3, M1 vs M2, and M7 vs M8. Using a chi square table for one degree of freedom, a cutoff of 3.841 was used to predict statistical significance of positive selection.

Analysis based on evolutionary rates of gene trees

The dN/dS of the genes that were present in more than four animals were estimated and used to examine evolutionary rates. The mean of the dN/dS of genes in the potassium channel, sodium channel, and potassium toxin families were examined. T-tests were performed to determine significant differences in evolutionary rates between groups.

Data access

A custom script was built to automate most of the processes mentioned. The custom scripts and related data are deposited in GitHub. (www.github.com/anuj2054/perseus).

Results

Species tree

Not all of the species I examined have been included in recent phylogenetic analyses (Ehsan et al., 2013; Zapata et al., 2015), thus I conducted a phylogenetic analysis of the currently included species. I used both the Bayesian and maximum likelihood approaches to phylogenetic reconstruction. The topologies produced by both methods were identical (Figure 2), and there were only minor differences in the support values (posterior probabilities and bootstrap values, respectively). Each of the morphologically distinct groups (e.g. anthozoans, myxozoans, hydrozoans) formed highly supported monophyletic groups in the species tree. The myxozoan lineage exhibited substantially longer branch lengths than the rest of the cnidarians which indicates that there is a much higher rate of evolution in that clade.

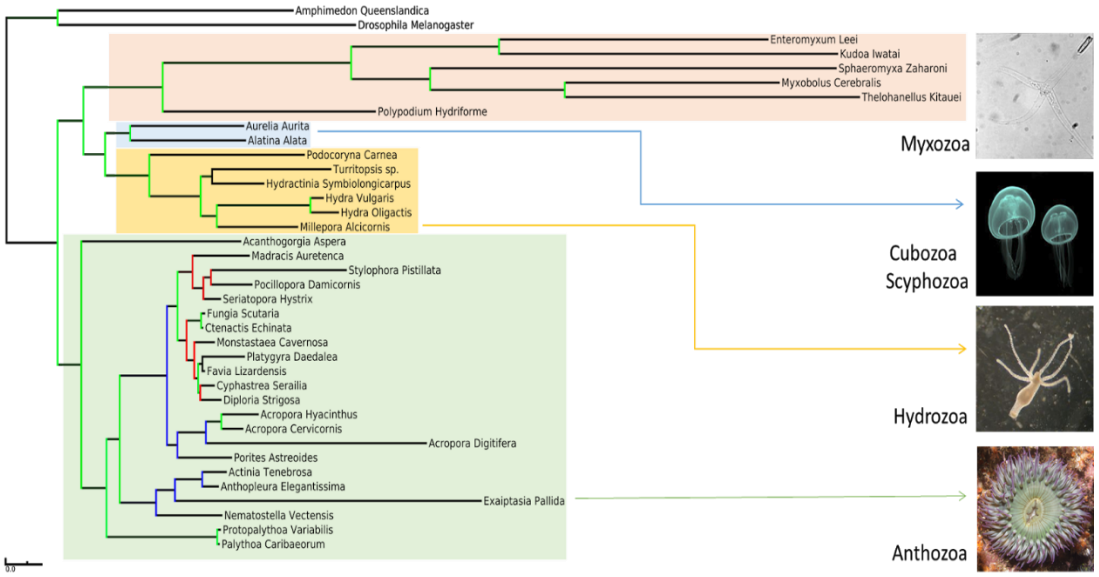


Figure 2. Species tree with branch lengths proportional to substitution rates. The color of node lines indicates the level of support. Nodes with green lines represent a posterior probability value higher than 0.95 and a bootstrap value higher than 95. Nodes with blue lines represent cases where the posterior probability and bootstrap values

were at least 0.90 and 90 and at least one value was less than 0.95 or 95. The nodes with red lines represent nodes where at least one value was less than 0.90 or 90. Each taxonomic class is indicated by a different background color and illustrated by a photo of a representative species.

Occurrence of toxin and ion channel genes among species

The presence or absence of each gene is shown in Supplementary Table 5. Most toxins were found in a restricted number of species. Although the toxin VKTC occurs in 30 of 38 species examined, most other toxins were found in only a few species. Among ion channels, potassium ion channels were found widely in cnidarians, with the major diversity arising from sea anemones. Because of the limited presence of each gene across the whole species tree, many evolutionary comparisons or analyses were limited.

Gene tree-species tree reconciliation

The gene trees constructed for each of the reciprocal blast hits of the gene families were grouped together into separate alignments. These alignments were used to reconstruct the history of gene duplications and losses. Horizontal gene transfer was assumed not to occur. The results for each of the gene families are shown in Table 1. There was higher number of duplications and losses observed in the potassium channel family compared to the other ion channel families. Similarly, there was a higher number of duplications and losses in the potassium toxin family compared to the other toxin families.

Table 1. The number of duplications and losses for different gene families in 39 cnidarian species.

Gene family	Duplication	Loss
Sodium channels	5	22
Potassium channels	22	56
Calcium channels	16	39
Potassium toxins	25	46
SCRIP toxins	8	15
Actinotoxins	2	21

Selection analysis

Random site selection analysis was performed on the genes obtained from reciprocal blast. LRTs were performed on the log likelihoods of each gene on the tree under different models of evolution. The LRTs were performed on M0 vs M3, M1 vs M2, and M7 vs M8. A positive result for the comparison of M0 vs M3, indicates significant variation in the dN/dS ratio among sites. I would not expect to detect positive selection in any case where there is no significant variation in the dN/dS ratio among sites. A positive result for M1 vs M2 and/or M7 vs M8 provides evidence of positive selection. Ten out of 30 (33%) of toxin or ion channel genes on which the test could be performed exhibited evidence for positive selection. This includes significant evidence of positive selection on four potassium ion channels and three potassium channel toxins. This is the pattern I expect to see if there are evolutionary interactions between toxins and the ion channels that they bind. In addition, one calcium ion channel and two other miscellaneous toxins exhibited evidence of positive selection.

Table 2. Results of the selection analyses. LRTs between the respective models for sodium, calcium, potassium, and toxin gene families. The random sites model from CodeML was used to calculate the log likelihoods of individual models. M0, M1, M2, M3, M7, and M8 respectively have higher number of categories of dN/dS values that are tested out. The genes that are positively selected are *CAC1C*, *KCNA2*, *KCNA1*, *KCNH5*, *KCND3*, *MCTX1*, *PA2*, *VKT1*, *VKT3C*, and *VKT3*.

Gene family	Gene name	M0 vs M3	M1 vs M2	M7 vs M8
Sodium channel	*SCN1A*	-46.4	0	-125.2
Sodium channel	*NALCN*	37.3	0	0
Calcium channel	*CAC1H*	0	-75.1	0
Calcium channel	*CAC1B*	131.4	0	-116.1
Calcium channel	*CACB2*	146.8	0	0

Calcium channel	*CAC1C*	355.1	0	3.9
Calcium channel	*CA2D4*	219.9	0.1	1.3
Calcium channel	*TPC2*	184.3	0	3.6
Calcium channel	*CAC1I*	-5491.9	0	0
Calcium channel	*TPC1*	0	0	1
Potassium channel	*KCNC1*	580.6	0	3.3
Potassium channel	*KCNA2*	391.4	0	14.6
Potassium channel	*KCNA1*	351.3	0	14.4
Potassium channel	*KCAB2*	91.6	0	0
Potassium channel	*KCAB1*	267	0	1
Potassium channel	*KCNH5*	581	0	53.9
Potassium channel	*KCNH6*	0	-5850.2	0
Potassium channel	*KCNQ5*	139.3	0	0
Potassium channel	*KCND3*	229.8	0	11.3
Potassium channel	*CSEN*	71.5	0	0
Cnidaria small cysteine-rich protein	*SCR1*	121.6	0	0.9
Dermatopontin	*MCTX1*	72.3	1.9	8.8
Phospholipase A2	*PA2*	67.1	81.8	110.3
Venom Kunitz-type, TypeII Potassium channel toxin	*VKT1*	43.5	40.4	29.1
Venom Kunitz-type, TypeII Potassium channel toxin	*VKT3A*	39.8	0	0
Venom Kunitz-type, TypeII Potassium channel toxin	*VKT3B*	92.9	0	0
Venom Kunitz-type, TypeII Potassium channel toxin	*VKT3C*	46.3	0	7.9
Venom Kunitz-type, TypeII Potassium channel toxin	*VKT3*	9.2	5.6	7.1
Venom Kunitz-type, TypeII Potassium channel toxin	*VKT3*	74.9	0	0.4
Venom Kunitz-type, TypeII Potassium channel toxin	*VKT6*	168.2	0	0

Correlation analysis of gene presence/absence

A table was constructed for the number of homologs of the different types of toxin genes and ion channel genes for each species. The relationship between the number of different ion channel genes and the number of different toxin genes per species was investigated using correlation analyses.

The correlation between neurotoxin and ion channel using the Pearson method was 0.49 with a p-value of 0.0046 (not shown). The correlation between potassium toxin and potassium channel using the Pearson method was 0.50 with a p-value of 0.0015 (Figure 3). There were too few toxin/channel genes of other types to do correlation analyses by themselves.

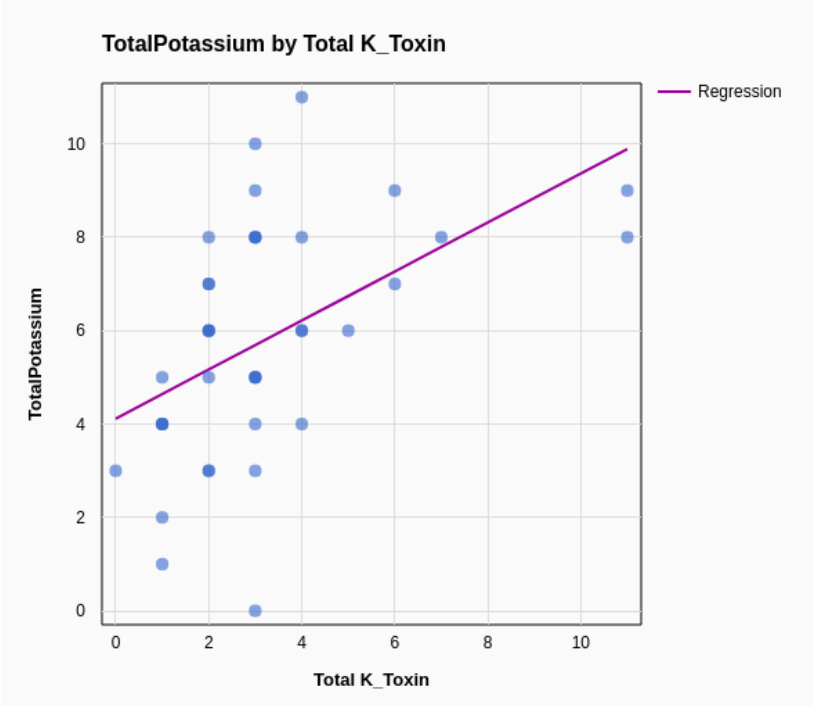


Figure 3. Comparison of potassium channel and toxin homologs. Shown is a scatter plot comparing the number of homologs of potassium channels to the number of homologs of potassium toxins in each of the animal studied. The equation for the best fit line was *Total_K_Channels = 0.525*Total_K_Toxins + 4.117.* An R squared value of 0.246 was obtained. A significant correlation of 0.5 was found using Pearson's correlation method between the potassium ion channel genes and potassium toxins shown in the figure.

Analysis of evolutionary rates

The dN/dS of the genes that were present in more than four animals were estimated and used to examine evolutionary rates. The mean of the dN/dS (shown in Figure 5) of the gene related to potassium channels, sodium channels, and potassium toxins were examined as these were the only groups containing specific genes found in at least four taxa. T-tests were performed to determine significant differences in evolutionary rates between groups. The p-value of calcium channels vs. potassium channels was 0.3626. The p-value of potassium channels vs. potassium toxins was 0.07561. The p-value of calcium vs. toxin was 0.0731. Thus, there were no significant rate differences between any class of genes but the evolutionary rates of toxin genes were appreciably higher than rates of the ion-channel genes.

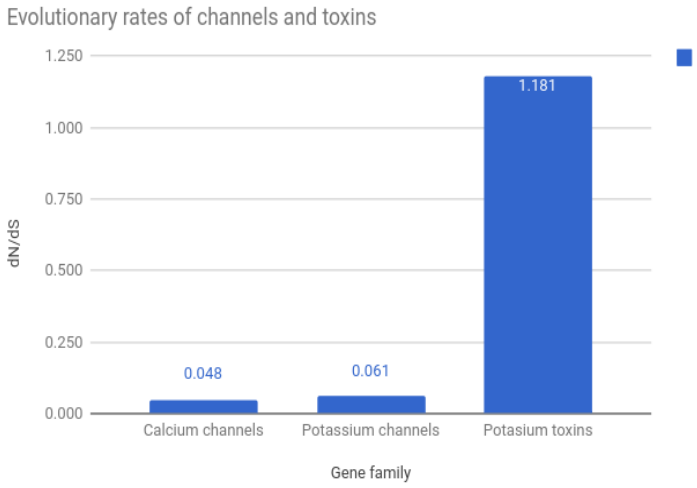


Figure 4. Evolutionary rates of channels and toxins. The mean of the dN/dS evolutionary rates of the calcium channels, potassium channels and potassium toxins. These values were obtained from using Model 8 in CODEML.

Discussion

My results provide a novel perspective on the evolution of neurotoxins and their target voltage-gated ion channels in Cnidaria. The level of ion channel diversity observed in

cnidarians, at both the gene family and nucleotide levels, is not observed in any other animal group. The co-occurrence of this diversity with the diverse set of toxins unique to cnidarians suggests a causal relationship. Moreover, the greatest diversity of ion channels at the intraspecific level occurs in sea anemones, with a correspondingly high diversity of toxins, while the lowest diversity of ion channels and neurotoxins occurs in the parasitic cnidarians. This pattern also holds for comparisons among whole clades, e.g. taxonomic classes (although it is possible this could be a function of the level of taxonomic sampling). These observations reinforce the evolutionary relationship between neurotoxins and their ion channel targets in Cnidaria and because voltage-gated ion channels are integral components of neurons, this has important implications for nervous system evolution in early animals.

Conclusions of this study require a few caveats. One limitation is in the datasets themselves. In many cases I used transcriptome sequences. A transcriptome only contains sequences of those genes that have been expressed at the time the tissue is sampled. To reduce this problem, only transcriptomes whose source tissue contained tentacle tissue were used in this study. Another limitation involves the sensitivity of reciprocal blast to identify short sequences (as in the case of toxin genes). Shorter sequences have a lower sensitivity of being detected than longer sequences (Ward and Moreno-Hagelsieb, 2014). For short sequences, proteins having related functions may not show overall high similarity yet contain a few short amino-acid motifs or residues that are highly conserved. Alternatively, for long sequences, proteins may show overall high percentage similarity but can contain a few differences in important functional

domains that change their function (Chen et al., 2007). Hidden Markov model profiles were not an appropriate approach since they require a protein sequence database of each animal. For proteins with short amino-acid sequences, like the neurotoxins, translating them from assembled nucleotide contigs in a transcriptome to proteins sequences leads to a high number of false positives. Finally, many other possible analyses could not be performed due to the limited presence of many genes across taxa. For example, the sparse distribution of specific homologs among taxa meant that in many cases there was not a sufficient number of sequences to perform tests of positive natural selection. In addition, I had hoped to perform analyses of evolutionary rate correlations among genes, yet because many homologs were absent from most taxa, there were not sufficient data to perform correlation analyses. In the end, I believe the analyses that I were able to perform provide sufficient support for an evolutionary relationship between neurotoxins and ion channels.

Rapid evolution of parasitic cnidarians

Myxozoa present a unique situation so I will discuss them separately. The general evolutionary pattern observed for parasitic cnidarians is one of rapid evolution with a small number of ion channels and neurotoxins. Two common reasons for rapid evolution are shorter generation times or a change in environment (Gaylord, 1944). It appears that the environment of parasitic cnidarians would have changed substantially with a change in lifestyle. This dramatic change in addition to adaptation to a range of hosts (e.g. from jellyfish to salmon) would have required rapid and extensive changes at a genomic scale.

A low number of neurotoxins and their target ion channels in parasitic cnidarians could be due to a lack of necessity to attack prey, defend against predators, or compete for space. As parasites acquire energy from their hosts, killing or paralyzing prey using neurotoxins is not necessary. Even though these animals retain nematocysts their utility remains unclear. For example, whirling disease is a common disease found in salmonid fishes that are caused by parasitic cnidarian, *Myxobolus cerebralis* (Hoffman, 1990). In this case the animal lives in the cartilage and bone of the fishes and are known to cause neurological damage (Langdon, 1990; Rose et al., 2000). The role of nematocysts and their neurotoxins as part of the infection process or during larval stages remains unknown. However, myxozoans have been reported to have undergone an extreme reduction in genome size and gene content (Chang et al., 2015), thus many neurotoxin and ion channel genes may have been lost since their divergence from free living ancestors.

Ion channels and neurotoxins may have co-evolved in an evolutionary arms race

Evolution of ion channels and their toxins is a dynamic process with the hallmarks of an evolutionary arms race. This study provided several lines of evidence that are consistent with this model. One is that there clearly have been numerous gene duplications along with two to four times as many gene losses. Potassium ion channels and potassium channel toxins are the most diverse systems with the highest number of duplications and losses. If ion channels are evolving in response to the toxins that bind to them, a diversity of channels might arise as a result of an increasing number of toxins. Thus duplicated and diverged (Lynch, 2003) channels could provide an escape mechanism to

allow species to evade the toxins of related taxa. Toxin gene expansion has been observed on various branches of the cnidarian tree. The very short (mostly close to 100 amino-acid) length of common toxin genes suggests there may be a limited number of effective forms, increasing the chance of convergent toxin evolution. Nonetheless, the loss of toxin genes appears to be far more common, suggesting that many toxin genes may be lost after alternative forms arise. The independent gain and loss of the toxins as indicated by my gene tree species tree reconciliation studies suggests a very dynamic positive feedback system where toxin genes evolve in response to corresponding evasive changes in their target ion channel genes.

The second line of evidence supporting the hypothesis is that I observed a significant correlation between the number of toxin genes and the number of ion channel genes within taxa. This is consistent with a scenario where multiple alternative ion channels arise in response to diversifying toxins. This correlation suggests that neurotoxins may have been an important driving force for the evolution of neural and muscular systems in which ion channels play an important role. While correlation does not indicate causation, this is additional evidence that is consistent with a toxin-channel evolutionary arms race.

The third line of evidence involves the results obtained from analysis of nucleotide sequence (codons) of specific genes. There is evidence for positive natural selection on several ion channel and toxin proteins (among those that were present in a sufficient number of taxa to perform tests of selection). This is particularly true for potassium

channels and the toxins that target them. This is exactly the result I would expect if toxins and their target ion channels are co-evolving. There may be selection on other genes in this study, as the method I used is known to be conservative in cases where dN/dS ratios are elevated but still less than 1.0 (Pond and Frost, 2005). I note that recent reports have concluded that venoms from evolutionarily younger lineages such as snakes and cone snails were under positive selection, while more ancient lineages such as cnidarians, spiders, centipedes and scorpions tended to be more constrained under negative selection (Jouiaei et al., 2015; Sunagar and Moran, 2015). Yet they suggested that episodic bursts of adaptive selection could occur on most toxin types with shifts in ecological parameters. I suggest that many cnidarian toxins have undergone positive selection due to recent arms race interactions with their associated ion channels.

In many arthropods venom is used as a weapon in predation as well as for intraspecific competitive interactions (Ligabue-Braun et al., 2012). Cnidarians similarly use venom for predation and defense (Talvinen and Nevalainen, 2002; Nevalainen et al., 2004). Many anemones as well as scleractinian corals are known to use venom to attack other individuals (Nelsen et al., 2014; Williams, 1991). They attack conspecifics or related taxa in competitive interactions to protect or expand their territories (Honma et al., 2005; Macrander and Brugler, 2015). Thus competition could have led to the diversification of the neurotoxin arsenal and the need for protection against the arsenals of conspecifics and close relatives.

Conclusion

I provide the first integrated analyses of cnidarian neurotoxins and their voltage-gated ion channel targets. My results provide multiple lines of evidence that consistently

support the coevolution of neurotoxins and ion channels in cnidarians. The evolutionary arms race scenario I have described provides a compelling explanation for the unique diversity of ion channels and toxins found in cnidarians. This study has important implications in studies investigating evolution of early nervous systems because voltage-gated ion channels form an integral part of nervous systems. In addition, the coevolution of toxins and ion channels provides a foundation for further studies of nervous systems if more taxa and genomes are available.

References

Anderson, Peter AV. "On the origins of that most transformative of biological systems–the nervous system." (2015): 504-505.

Barzilai, Maya Gur, et al. "Convergent evolution of sodium ion selectivity in metazoan neuronal signaling." *Cell reports* 2.2 (2012): 242-248.

Brodie Jr, Edmund D., B. J. Ridenhour, and E. D. Brodie III. "The evolutionary response of predators to dangerous prey: hotspots and coldspots in the geographic mosaic of coevolution between garter snakes and newts." *Evolution* 56.10 (2002): 2067-2082.

Bucher, Dirk, and Peter AV Anderson. "Evolution of the first nervous systems–what can I surmise?" *The Journal of Experimental Biology* 218.4 (2015): 501-503.

Castaneda, Olga, and Alan L. Harvey. "Discovery and characterization of cnidarian peptide toxins that affect neuronal potassium ion channels."*Toxicon* 54.8 (2009): 1119-1124.

Chang, E. Sally, et al. "Genomic insights into the evolutionary origin of Myxozoa within Cnidaria." *Proceedings of the National Academy of Sciences* 112.48 (2015): 14912-14917.

Chen, Feng, et al. "Assessing performance of orthology detection strategies applied to eukaryotic genomes." *PloS one* 2.4 (2007): e383.

Chen, Kevin, Dannie Durand, and Martin Farach-Colton. "NOTUNG: a program for dating gene duplications and optimizing gene family trees." *Journal of Computational Biology* 7.3-4 (2000): 429-447.

Darriba, Diego, et al. "jModelTest 2: more models, new heuristics and parallel computing." *Nature methods* 9.8 (2012): 772-772.

Dawkins, Richard, and John R. Krebs. "Arms races between and within species." *Proceedings of the Royal Society of London B: Biological Sciences*205.1161 (1979): 489-511.

Frazão, Bárbara, Vitor Vasconcelos, and Agostinho Antunes. "Sea anemone (Cnidaria, Anthozoa, Actiniaria) toxins: an overview." *Marine drugs* 10.8 (2012): 1812-1851.

Harris, Richard J., and Kevin Arbuckle. "Tempo and Mode of the Evolution of Venom and Poison in Tetrapods." *Toxins* 8.7 (2016): 193.

Hoffman, G. L. "Myxobolus cerebralis, a worldwide cause of salmonid whirling disease." *Journal of Aquatic Animal Health*2.1 (1990): 30-37.

Honma, Tomohiro, et al. "Novel peptide toxins from acrorhagi, aggressive organs of the sea anemone Actinia equina." *Toxicon* 46.7 (2005): 768-774.

Jouiaei, Mahdokht, et al. "Ancient venom systems: a review on Cnidaria toxins." *Toxins* 7.6 (2015): 2251-2271.

Jouiaei, Mahdokht, et al. "Evolution of an ancient venom: recognition of a novel family of cnidarian toxins and the common evolutionary origin of sodium and potassium neurotoxins in sea anemone." *Molecular biology and evolution* 32.6 (2015): 1598-1610.

Jouiaei, Mahdokht. "Evolution and diversification of the cnidarian venom system." (2016).

Kayal, Ehsan, et al. "Cnidarian phylogenetic relationships as revealed by mitogenomics." *BMC Evolutionary Biology* 13.1 (2013): 1.

Kelava, Iva, Fabian Rentzsch, and Ulrich Technau. "Evolution of eumetazoan nervous systems: insights from cnidarians." *Phil. Trans. R. Soc. B* 370.1684 (2015): 20150065.

Kosakovsky Pond, Sergei L., and Simon DW Frost. "Not so different after all: a comparison of methods for detecting amino-acid sites under selection." *Molecular biology and evolution* 22.5 (2005): 1208-1222.

Lanfear, Robert, et al. "PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses." *Molecular biology and evolution* 34.3 (2016): 772-773.

Langdon, J. S. "Observations on new Myxobolus species and Kudoa species infecting the nervous system of Australian fishes." *Journal of Applied Ichthyology* 6.2 (1990): 107-116.

Lazcano-Pérez, Fernando, et al. "Cnidarian neurotoxic peptides affecting central nervous system targets." *Central Nervous System Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Central Nervous System Agents)* 16.3 (2016): 173-182.

Le, Si Quang, and Olivier Gascuel. "An improved general amino-acid replacement matrix." *Molecular biology and evolution* 25.7 (2008): 1307-1320.

Li, Xiaofan, et al. "Major diversification of voltage-gated K+ channels occurred in ancestral parahoxozoans." *Proceedings of the National Academy of Sciences* 112.9 (2015): E1010-E1019.

Liebeskind, Benjamin J., David M. Hillis, and Harold H. Zakon. "Convergence of ion channel genome content in early animal evolution." *Proceedings of the National Academy of Sciences* 112.8 (2015): E846-E851.

Ligabue-Braun, Rodrigo, Hugo Verli, and Célia Regina Carlini. "Venomous mammals: a review." *Toxicon* 59.7 (2012): 680-695.

Lynch, Michael, and John S. Conery. "The origins of genome complexity." *Science* 302.5649 (2003): 1401-1404.

Macrander, Jason, Mercer R. Brugler, and Marymegan Daly. "A RNA-seq approach to identify putative toxins from acrorhagi in aggressive and non-aggressive Anthopleura elegantissima polyps." *BMC genomics* 16.1 (2015): 221.

Messerli, Shanta M., and Robert M. Greenberg. "Cnidarian toxins acting on voltage-gated ion channels." *Marine Drugs* 4.3 (2006): 70-81.

Moran, Yehu, and Harold H. Zakon. "The evolution of the four subunits of voltage-gated calcium channels: ancient roots, increasing complexity, and multiple losses." *Genome biology and evolution* 6.9 (2014): 2210-2217.

Moran, Yehu, Dalia Gordon, and Michael Gurevitz. "Sea anemone toxins affecting voltage-gated sodium channels–molecular and evolutionary features." *Toxicon* 54.8 (2009): 1089-1101.

Moran, Yehu, et al. "Evolution of voltage-gated ion channels at the emergence of Metazoa." *The Journal of experimental biology* 218.4 (2015): 515-525.

Nelsen, David R., et al. "Poisons, toxungens, and venoms: redefining and classifying toxic biological secretions and the organisms that employ them." *Biological Reviews* 89.2 (2014): 450-465.

Nevalainen, Timo J., et al. "Phospholipase A2 in cnidaria." *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*139.4 (2004): 731-735.

Rachamim, Tamar, et al. "The dynamically evolving nematocyst content of an anthozoan, a scyphozoan, and a hydrozoan." *Molecular biology and evolution* 32.3 (2014): 740-753.

Ronquist, Fredrik, and John P. Huelsenbeck. "MrBayes 3: Bayesian phylogenetic inference under mixed models." *Bioinformatics* 19.12 (2003): 1572-1574.

Rose, James D., et al. "Whirling disease behavior and its relation to pathology of brain stem and spinal cord in rainbow trout." *Journal of Aquatic Animal Health* 12.2 (2000): 107-118.

Rowe, Ashlee Hedgecock. "Coevolution between grasshopper mice (Onychomys spp.) and bark and striped scorpions (Centruroides spp.)." (2004).

Sievers, Fabian, et al. "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." *Molecular systems biology* 7.1 (2011): 539.

Simpson, George Gaylord. *Tempo and mode in evolution*. No. 15. Columbia University Press, 1944.

Stamatakis, Alexandros. "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies." *Bioinformatics* 30.9 (2014): 1312-1313.

Stöver, Ben C., and Kai F. Müller. "TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses." *BMC bioinformatics* 11.1 (2010): 7.

Sunagar, Kartik, and Yehu Moran. "The rise and fall of an evolutionary innovation: contrasting strategies of venom evolution in ancient and young animals." *PLoS Genetics* 11.10 (2015): e1005596.

Sunagar, Kartik, et al. "Ecological venomics: How genomics, transcriptomics and proteomics can shed new light on the ecology and evolution of venom." *Journal of proteomics* 135 (2016): 62-72.

Talvinen, Kati A., and Timo J. Nevalainen. "Cloning of a novel phospholipase A2 from the cnidarian Adamsia carciniopados." *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 132.3 (2002): 571-578.

Team, R. Core. "R language definition." Vienna, Austria: R foundation for statistical computing (2000).

Technau, Ulrich, and Michaela Schwaiger. "Recent advances in genomics and transcriptomics of cnidarians." *Marine genomics* 24 (2015): 131-138.

Torres-Ramos, Monica A., and Manuel B. Aguilar. "Recent advances in cnidarian neurotoxin research." *Comments on Toxicology* 9.2 (2003): 161-174.

Van Valen, Leigh. "The red queen." *The American Naturalist* 111.980 (1977): 809-810.

Ward, Natalie, and Gabriel Moreno-Hagelsieb. "Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: how much do I miss?."*PLoS One* 9.7 (2014): e101850.

Warrington, Janet A., et al. "Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes." *Physiological genomics* 2.3 (2000): 143-147.

Whelan, Simon, and Nick Goldman. "A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach." *Molecular biology and evolution* 18.5 (2001): 691-699.

Williams, R. B. "Acrorhagi, catch tentacles and sweeper tentacles: a synopsis of ‘aggression of actiniarian and scleractinian Cnidaria." *Hydrobiologia*. Vol. 216. No. 1. Kluwer Academic Publishers, 1991.

Yang, Ziheng. "PAML 4: phylogenetic analysis by maximum likelihood." *Molecular biology and evolution* 24.8 (2007): 1586-1591.

Zapata, Felipe, et al. "Phylogenomic analyses support traditional relationships within Cnidaria." *PloS One* 10.10 (2015): e0139068.

List of Tables

List of Figures

List of Supplementary Information

Chapter 2: Transcriptomic and proteomic analyses of toxin composition in clownfish-host and non-host sea anemones

Abstract

Despite having a potent mix of toxins, sea anemones serve as hosts for various animals which are able to live among their stinging tentacles. The mechanisms that allow some anemones to host many species while others host a few, or only one, are not clearly understood. Here, I describe an investigation of the toxin content of two sea anemone species: *Entacmaea quadricolor* (Bubble tip anemone) and *Condylactis gigantea* (Condy anemone). *Entacmaea quadricolor* is known to host 13 different species of clownfish, while *Condylactis gigantea* has not been reported to host any clownfish. I performed RNA-seq based transcriptome analysis and tandem mass spectrometry (MS/MS) based bottom-up proteomic analysis to identify toxins present in these animals. Both transcriptomic and proteomic analyses provide independent perspectives on toxin expression, allowing a comparison of their results. In the results, I found that two highly represented toxin families present in the sea anemones I studied were snaclec family, which are major constituents of snake venoms, and venom Kunitz type (VKT) family, which are unique to cnidarians. Seventy-nine toxins in *Condylactis gigantea* and 56 toxins in *Entacmaea quadricolor* were identified using a RNA-seq approach. Only 3 toxins in *Condylactis gigantea* and 7 toxins in *Entacmaea quadricolor* were identified using a MS/MS study. Only one toxin, NA1_CONGI, was identified using both a RNA-seq approach and a MS/MS approach. As *Entacmaea quadricolor* is an anemone that hosts clownfish, the higher number of toxins relative to the non-host *Condylactis gigantea* was unexpected. It is possible that the quantity of each toxin in

Entacmaea quadricolor is lower than *Condylactis gigantea*. Thus, despite there being a larger diversity of toxins in *Entacmaea quadricolor*, lower toxin expression levels could provide a more habitable environment for clownfish that use it as a host.

Keywords: Sea anemones, *Entacmaea*, *Condylactis*, Venom, Transcriptome, Proteome, Toxin, Nematocyst.

Introduction

Marine organisms are known to be outstanding sources for biologically active compounds (Munro et al., 1999). They include some of the most venomous species on earth, such as cone snails, stone fishes, stingrays and octopuses. In addition, the phylum Cnidaria (jellyfish, sea anemones and hydras), includes the box jellyfish, considered the most venomous organisms in the sea. Sea anemones, members of the class Anthozoa within Cnidaria, possess a diverse array of toxins used in venoms (Honma and Shiomi, 2006). These toxins include proteins that bind voltage-gated Na^+ and K^+ ion channels, form pores in membranes (actinoporins), and act as protease inhibitors (Bosmans and Tytgat, 2007; Standkera et al., 2006). In addition, they may also act through several cytolytic, hemolytic, immunomodulating mechanisms (Pento et al., 2011). Despite having a potent mix of toxins, sea anemones serve as hosts for various animals, which are able to live among their stinging tentacles. In these symbiotic relationships, resident species derive protection and nesting sites from anemones, but possible benefits to the host anemones remain unclear (Porat and Chadwick, 2005; Mebs, 1994; Nedosyko et al., 2014). Of about 1,000 species of anemones that have been investigated, 10 taxonomically disparate species host anemone fishes, commonly known as clownfish

(Dunn, 1981; Debelius and Baensch, 1997). Twenty-nine species of clownfish from the genera *Amphiprion* and *Premnas* inhabit 10 different sea anemone host species (Dunn, 1984; Dunn et al., 1992; Fautin and Allen, 1992; Froese and Pauly, 2016; Table 1). These belong to three different families in the order Actiniaria, including Actiniidae, Stichodactylidae, and Thalassianthidae (Debelius, 1994). The Actiniidae is the largest family and includes the most well-known anemones such as the snakelocks anemone which is consumed as a delicacy in Spain. The family Stichodactylidae includes large species known as carpet anemones which are found only in the tropics. The Thalassianthidae contains four genera but only one species in this family is known to host clownfish. Thus, the resident clownfish species belong to two closely- related genera, whereas the anemone hosts are from disparate families.

Table 1. Anemones known to host clownfish, and the number of clownfish species they host.

Family	Common name	Scientific name	Number of clownfish species hosted
Actiniidae	Bubble tip anemone	*Entacmaea quadricolor*	13
Actiniidae	Sebae anemone	*Heteractis crispa*	14
Actiniidae	Magnificent anemone	*Heteractis magnifica*	12
Actiniidae	Delicate anemone	*Heteractis malu*	1
Actiniidae	Long tentacle anemone	*Macrodactyla doreensis*	5
Actiniidae	Beaded anemone	*Heteractis aurora*	7
Stichodactylidae	Giant carpet anemone	*Stichodactyla gigantean*	7
Stichodactylidae	Saddle anemone	*Stichodactyla haddoni*	6
Stichodactylidae	Merten's carpet anemone	*Stichodactyla mertensii*	13

Thalassianthidae	Adhesive anemone	*Cryptodendrum adhaesivum*	1

The genomic mechanisms that allow some anemones to host many clownfish species, and others host only a few, remain unknown (Mebs, 1994). It has been suggested that resident clownfish may be tolerant to anemone venom due to a protective mucus layer (Fautin, 1991), but presence of other resistance mechanisms in them is not clear. If clownfish have evolved mechanisms that confer resistance to specific toxins, then there may be a relationship between the complexity of venoms produced by different anemone species and the number of fishes that have evolved resistance to these venoms. Specifically, for anemones that produce a complex venom with a variety of evolutionarily derived toxin activities such as channel blockers (Dauplais et al., 1997) and proteases (Putnam et al., 2007), it is likely that few species of fishes will have evolved resistance to this complex venom. Alternatively, more species of fishes may have evolved resistance to anemones with simple, relatively unspecialized venoms (Fry et al., 2009; Miller et al., 2007). Thus, I predict that there may be a relationship between the number of different species of clownfish that are hosted by anemones and the number of toxins that the anemones produce.

Here, I investigated the toxin content of two sea anemone species: *Entacmaea quadricolor* (Bubble tip anemone, Figure 1) and *Condylactis gigantea* (Condy anemone, Figure 1). *Entacmaea quadricolor* is known to host 13 different species of clownfish (Dunn, 1981; Dunn, 1984; Scott and Harrison, 2009), whereas *Condylactis gigantea* has not been reported to host any clownfish (Hanlon and Hixton, 1986;

Sheridan et al., 2015; Porat and Furman, 2004). I chose these two anemone species for my study because of their ease of availability and maintenance, and because of their difference in the number of fish species they host (Debelius, 1994). I performed RNA-seq transcriptome analysis and tandem mass spectrometry (MS/MS) bottom-up proteomic analysis to identify toxins and ion channels present in these animals. Transcriptomic and proteomic analyses provide independent perspectives on toxin expression (Ponce et al., 2016; Brinkman et al., 2015; and Li et al., 2014) allowing a more thorough investigation of cnidarian toxins and their molecular targets.

Figure 1. Images of *Entacmaea quadricolor* (left) and *Condylactis gigantea* (right) used for the study.

Methods

Sample collection and dissection

A single individual of *Entacmaea quadricolor* and *Condylactis gigantea* were purchased from a local pet store, transported to the laboratory in marine water, and immediately sacrificed. Multiple samples of approximately 100 mg from the tips of all the tentacles and approximately 100 mg from the epidermal layer of the column were slowly dissected to avoid the anemone from retracting its tentacles into its oral disk. Some of the samples were flash frozen in liquid nitrogen and stored at -80^{o}C before

protein extraction. Some of the samples were stored in RNA*later* (Invitrogen) solution before RNA extraction. Some samples were also stored in ethanol for DNA extraction.

DNA extraction and COX-1 sequencing

Genomic DNA was purified from the anemone column tissue using a DNeasy™ Blood & Tissue kit (Qiagen) following the manufacturer's protocol. COX-1 sequencing was performed to verify species identity using DNA barcodes. The COX-1 primers were designed using OligoArchitect™, an online tool provided by Sigma-Aldrich. The forward primers were designed to be GGTATGATAGGCACAGCT and the reverse primers were GAAAGTTGTATTAAARTTCCTATCTG. The same primers were used for both *Entacmaea quadricolor* and *Condylactis gigantea*. A gradient PCR using Q5 High Fidelity DNA Polymerase (New England Biolabs) was performed to find the best annealing temperature. The best annealing temperature for both *Entacmaea quadricolor* and *Condylactis gigantea* was found to be 56.1^0C. PCR was performed using Q5 High Fidelity DNA Polymerase (New England Biolabs) with an annealing temperature of 56.1^0C. The amplified COX-1 genes for *Condylactis gigantea* was 432 base pairs (bp) long and for *Entacmaea quadricolor* was 357 bp long. They were sequenced on an ABI 3130xl Genetic Analyzer using the BigDye Terminator v3.1 cycle sequencing mix. The COX1 sequences were blasted (blastn) against NCBI nucleotide database to verify the animal.

mRNA extraction and sequencing

Anemone tentacle tip and anemone column were used for RNA extraction using TRIzol™ reagent (Thermo Fisher Scientific) using the manufacturer's protocol. RNA

was quantified using a Qubit™ (Thermo Fisher Scientific) fluorometer. The mRNA purification and cDNA synthesis was carried out with a TruSeq™ Stranded mRNA Library Prep Kit (Illumina). Library integrity was assessed using a Tape Station 2000 (Agilent Technologies). cDNA libraries from the three samples of anemone tentacle tissue and anemone column tissue for both of the species were sequenced in one lane on a HiSeq 3000 (Illumina) instrument.

Transcriptome assembly and annotation

After Illumina sequencing, the quality of raw sequence data was assessed using FastQC (version 0.9.2; Andrews, 2010). Illumina adapter sequences and low quality bases (Phred score > 32) were then removed from the sequence reads using Trimmomatic (Bolger et al., 2014). Reads shorter than 36 base pairs were discarded and the quality of filtered data was re-evaluated using FastQC. After quality control, paired-end sequences from six fastq files of each species were *de novo* assembled into contigs using the default parameters in Trinity (version 2.2; Haas et al., 2013). After assembling with Trinity, the coding regions and proteins were predicted using Transdecoder (Haas and Papanicolaou, 2016). This provided an additional protein database for each anemone species, which could be used for later homology analysis.

To annotate the predicted functions of the predicted proteins, transcripts of both animals were aligned to sequences available in a set of public databases using the tBLASTn and BLASTx algorithms (E-value cutoff of 1E-5) (Altschul et al., 1990). Searches were conducted against public and custom-made databases including (a) the Swiss-Prot database (as of 1 October, 2017); (b) the complete proteomic data sets from Uniprot of

two cnidarians, which have high quality genomes (*Hydra vulgaris* and *Nematostella vectensis*); and (c) the UniProt animal toxin and venom database (Jungo et al., 2012). Since anemones are symbiotic with zooxanthellae algae, only metazoan protein homologs were retained and all non-metazoan homologs were filtered out by using a search term for "Metazoa". The homologs of toxin sequences were visually verified to include start and stop codons using Geneious (Kearse et al., 2012).

Protein extraction from sea anemone tissue samples

A bead-beating method was performed to extract proteins from the tissue samples as described previously in von der Haar (2007). Briefly, 500 µL lysis buffer (i.e., 1 mM PMSF in 25mM NH_4HCO_3) was mixed with 0.1 g tissue sample and 100 µL of 0.1 mm zirconia/silica beads (Polyscience, Inc.). After bead-beating, the supernatant was centrifuged at 10,000 x G at 4°C for 10 minutes to remove unbroken cells and debris. The supernatant was collected, and the protein concentration was measured with a Pierce BCA protein assay kit. For protein extraction quality evaluation, all the extracted protein samples were analyzed on 4%-12% polyacrylamide SDS-PAGE gels. Gel electrophoresis was performed with a Mini-Protean Tetra System (Bio-Rad Laboratories, Inc.), at 150V for 75min. The gel was stained with Gelcode Blue Safe Gel Stain (Thermo Fisher Scientific), and the image was collected using the ChemiDoc Imaging System (Bio-Rad).

Bottom-up MS/MS analysis

The extracted protein samples from the sea anemone tissues were firstly denatured in 6M urea. Denatured samples were further reduced with 200mM DTT for 1 hour at 37°C,

and alkylated with 200mM IAA for 30 minutes. After reduction and alkylation, protein samples were diluted to 10-fold volume with 25mM NH_4HCO_3, and tryptic-digested overnight with a 30:1 trypsin to protein mass ratio at 37°C (Zhang et al., 2014). The digested peptides were desalted and loaded onto an in-house packed C18 column (5 μm, 75 μm × 15 cm) for the bottom-up MS/MS study. The mobile phases were 0.1% formic acid in water, which acted as Mobile Phase A (MPA), and 0.1% formic acid in acetonitrile, which acted as Mobile Phase B (MPB). The gradient was from 8% MPB to 35% MPB over 90 minutes following a 15-minute sample loading step. The column was regenerated with 90% MPB for 10 minutes and equilibrated to 3% MPB for 30 minutes. The eluent was injected into an online-coupled LTQ Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific) with a customized nano-ESI interface and a home-made HF etched tip. The ion-spray voltage was set to 2.4kV, and the temperature of the capillary was 300°C. Higher Collisional Induced Dissociation (HCD) MS/MS with a normalized collision energy 30, in a data-dependent mode, top 10 highest abundant parent ions were analyzed.

In bottom-up experiments, peptides were identified using MSGF+ (Kim et al., 2008; Kim and Pevzner, 2014) to search the mass spectra from Liquid Chromatography MS/MS (LC-MS/MS) analysis against the annotated RNA-seq database and its decoy database. Peptide identifications were filtered based on the calculated FDR <1% at the unique peptide level.

Results

COX-1 barcode sequencing

The consensus of the forward and reverse reads of COX1 sequences for *Condylactis gigantea* and *Entacmaea quadricolor* were obtained and deposited in NCBI with accession numbers MG132209 (*Condylactis gigantea*) and MG132210 (*Entacmaea quadricolor*). The sequences were blasted against Genbank using blastn. *Entacmaea quadricolor* samples were matched with an E value of 3×10^{-67} and a coverage of 89% to a protein sequence with NCBI accession JQ839204.1, which is also a COX1 gene from *Entacmaea quadricolor*. However, *Condylactis gigantea* did not have any COX1 sequences deposited in NCBI, so its species identity was verified by morphological features such as the length of tentacles, the color of its column and its large size.

Transcriptome analysis

All raw sequence data were deposited in European Bioinformatics Institute databases with study accession numbers as PRJEB21970. Total RNA, purified from whole *Condylactis gigantea* and *Entacmaea quadricolor* tentacle and column tissues, was used to generate 83,039,342 and 87,802,748 paired reads, respectively. These reads were then assembled *de novo*, using Trinity into 194,413 and 341,370 transcripts that are summarized in Table 2. The average length of assembled transcripts was 1,108.30 bases for *Condylactis gigantea* and 730 bases for *Entacmaea quadricolor* (Table 2) with an N50 of 2281 bases for *Condylactis gigantea* and 1245 bases for *Entacmaea quadricolor*. Due to the limited number of *Condylactis gigantea* or *Entacmaea quadricolor* sequences available in protein and gene databases, transcripts were queried

against three databases using blastx and blastp — SwissProt, predicted protein sets from the *Hydra magnipapillata* genome and predicted protein sets from the *Nematostella vectensis* genome. Approximately 54.4% of *Condylactis gigantea* and 18.2% of *Entacmaea quadricolor* sequences returned a high-scoring (e-value < = 10e-5) match to SwissProt using Blastp (Table 3). The assembled transcripts were used to perform protein prediction using Transdecoder, which predicted 20,562 protein sequences for *Condylactis gigantea*, of which 65,601 were Open Reading Frames (ORFs) that contained both start and stop codons, and 140,594 protein sequences for *Entacmaea quadricolor*, of which 55,339 were ORFs that contained both start and stop codons. The complete annotation table for *Entacmaea quadricolor* and *Condylactis gigantea* is provided in Supplementary File 1.

Table 2. Quality control, assembly and structural annotation summary for *Condylactis gigantea* and *Entacmaea quadricolor* transcriptomes.

Parameters	***Condylactis gigantea (all tissue)***	***Entacmaea quadricolor* (all tissue)**
# PE Reads (2 × 150 bp)	83,039,342	87,802,748
Forward and Reverse recovered after trimming	71,866,086	75,555,907
Total trinity 'genes'	194,413	341,370
Total trinity transcripts	250,270	424,748
Median contig length based on contigs (bp)	502	381
Median length (N50) (bp) based on contigs	2,281	1,245
Average contig (bp)	1,108.30	730.27
Number of predicted peptides (Transdecoder)	142,654	140,594
Complete (ORF with both start and stop codons)	65,601	55,339
5' partial (ORF with start codon missing)	36,023	39,213

3' partial (ORF with stop codon missing)	9,810	12,430
Internal (ORF with no start or stop codons)	31,220	33,612

Table 3: Functional annotation summary for *Entacmaea quadricolor* (EQ) and *Condylactis gigantea* (CG) transcriptome.

Homologs Found Using:	Number of Unique Homologs for CG	Hits (%) for CG	Number of Unique Homologs for EQ	Hits (%) for EQ
BLASTx against Swissprot	24,775	9.80	91,745	21.5
BLASTx against Swissprot, Metazoans Only	16,167	6.45	76,575	18.02
BLASTx against Swissprot, Cnidarians Only	199	0.07	1,898	0.4
BLASTX against Swissprot, Toxins	56	0.02	79	0.01
BLASTX against Swissprot, Cnidarian Toxins	36	0.014	50	0.008
BLASTX against ion channels in Swissprot	79	0.05	98	0.06
BLASTP against Swissprot	77,238	54.14	25,616	18.2
BLASTP against Swissprot, Metazoans Only	42,407	29.72	18,539	13.18
BLASTP against Swissprot, Cnidarians Only	801	5.6	200	0.14
BLASTP against Swissprot, Toxins Only	43	0.03	52	0.04
BLASTP against Swissprot, Cnidarian Toxins	19	0.01	28	0.02
BLASTP against Swissprot, ion channels	86	0.06	89	0.06

Identification of Toxin proteins

To identify potential toxins in *Entacmaea quadricolor* and *Condylactis gigantea*, assembled transcripts of each animal were compared to the Uniprot animal toxin database using blastx and blastp. Using blastx, 56 transcripts (0.02%) in *Condylactis*

gigantea and 79 transcripts (0.01%) in *Entacmaea quadricolor* provided high-scoring BLAST hits (bit score > 50). Among these transcripts, 36 transcripts in *Condylactis gigantea* and 50 transcripts in *Entacmaea quadricolor* were found to be associated with the word "Cnidaria" using GO term filtering. Using blastp, 43 transcripts (0.03%) in *Condylactis gigantea* and 52 transcripts (0.04%) in *Entacmaea quadricolor* provided high-scoring BLAST hits (bit score > 50). Out of these transcripts, 19 transcripts in *Condylactis gigantea* and 28 transcripts in *Entacmaea quadricolor* were found to associated with the word "Cnidaria" using GO term filtering. The blastx and blastp results were combined to represent the toxins in *Entacmaea quadricolor* and *Condylactis gigantea*. The toxins were classified according to the Uniprot Animal Toxin classification system, which resulted in the representation of 17 venom protein families (Table 4). Two highly represented toxin families present in the sea anemones studied are the snaclec family, which are major constituents of snake venoms (Clemetson et al., 2009), and the venom Kunitz type (VKT) family, which are unique to cnidarians. Ten different types of snacelc proteins were found for *Condylactis gigantea*, and eight different types were found for *Entacmaea quadricolor*. Seven different types of VKT proteins were found for *Condylactis gigantea* and four different types of VKT proteins were found for *Entacmaea quadricolor*. Representatives of other toxin families included Cnidaria small cysteine-rich protein (SCRiP), Complement C3 homolog, Cysteine Rich Secretory Protein (CRISP), Flavin monoamine oxidase (L-amino-acid oxidase), Metalloproteinase (M12A), Multicopper oxidase, Natriuretic peptide, Nucleotide pyrophosphatase/phosphodiesterase, Peptidase S1 (serine protease), Phospholipase A2 (PA2), Phospholipase B-like (PB), Snaclec, Cubozoan protein toxin

(CPT) family, True venom lectin (C-type lectin), Type-B carboxylesterase/lipase, VKT, and Venom metalloproteinase (M12B).

Table 4. For each gene family, the number and identity of candidate toxins in *Condylactis gigantea* (CG) and *Entacmaea quadricolor* (EQ) identified using blastx. Only proteins that could be assigned to a particular family using the Uniprot toxin database have been tabulated. Snaclec, VKT, and Phospholipase A are the most abundant toxin homologs in both the animals.

Toxin Protein Family	Number of Homologs in EQ	Homologs in EQ	Number of Homologs in CG	Homologs in CG
Cnidaria small cysteine-rich protein (SCRiP)	1	SCR1_ACRMI	0	
Complement C3 homolog	2	VCO3_NAJKA, VCO31_AUSSU	2	VCO3_NAJKA, VCO31_AUSSU
CRISP	3	CRVP_CERRY, CRVP_LATSE, CRVP_PSEPO	0	
Flavin monoamine oxidase (L-amino-acid oxidase)	1	OXLA_BUNMU	1	OXLA_OPHHA
Metalloproteinase (M12A)	3	VMP_NEMVE, VMPA_LOXIN, VMPA3_LOXIN	1	VMP_NEMVE
Multicopper oxidase	3	FA5V_OXYMI, FA5V_OXYSU, FA5V_PSETE	3	FA5V_OXYMI, FA5V_OXYSU, FA5V_PSETE
Natriuretic peptide	0		1	SVMI1_CERCE
Nucleotide pyrophosphatase/phosphodiesterase	1	PDE2_CROAD	2	PDE1_CROAD, PDE2_CROAD
Peptidase S1 (serine protease)	1	FAXD2_DEMVE	0	
Phospholipase A2 *	6	COMA_CONMA, PA2_ADAPA, PA2_CONGI, PA2_URTCR, PA2A5_AUSSU, PA2NB_NAJSP	6	COMA_CONMA, PA2_ADAPA, PA2_CONGI, PA2_URTCR, PA2A_NAJAT, PA2BA_PSEPO
Phospholipase B-like	0		1	PLB_DRYCN
Snaclec *	10	SL1_CRODU, SL3_SISCA, SLA_TRIST, SLAA_TRIST, SLB_BOTIN,	8	SLA_BITAR, SLA_TRIST, SLA6_MACLB, SLAA_TRIST, SLB_BITAR,

		SLB_CRODU, SLB_GLOHA, SLB1_DEIAC, SLEB_BOTJA, SLLC1_DABSI		SLEA_CALRH, SLUA_DEIAC, SLUB_DEIAC
Cubozoan protein toxin	1	CTXA_CARAL	0	
True venom lectin (C-type lectin)	1	LECG_THANI	1	LECG_THANI
Type-B carboxylesterase/lipase	1	ACES_BUNFA	1	ACES_BUNFA
Venom Kunitz-type *	7	VKT1A_ANEVI, VKT3A_ACTEQ, VKT5_ANEVI, VKT6_ANEVI, VKTA_ANEVI, VKTB_ANEVI, VKTC_ANEVI	4	VKT3_WALAE, VKT3A_ACTEQ, VKT6_ANEVI, VKTC_ANEVI
Venom metalloproteinase (M12B)	3	VM2P1_PROMU, VM38_DRYCN, VM3DK_DABRR	2	VM3_NAJKA, VM3AA_CROAT
Total toxins assignable to a family	**44**		**33**	
Total toxins not assignable to a family	**49**		**34**	

Proteome analysis

The peptides identified by searching the MS/MS spectral data against the animal toxin database are given in Table 5 for *Condylactis gigantea* and in Table 6 for *Entacmaea quadricolor*. Two unique toxin peptides were identified from *Entacmaea quadricolor*. One unique toxin peptide was identified from *Condylactis gigantea*. Seven toxin peptides that were also shared with peptides from other proteins used in the database search were identified from *Condylactis gigantea,* and similarly again seven shared toxin peptides were identified from *Entacmaea quadricolor*. The unique peptides identified in *Entacmaea quadricolor* were delta-actitoxin (ACTP_ENTQU) and snake venom serine protease (VSP_ECHOC). The unique peptide identified in *Condylactis gigantea* was Delta-actitoxin-Cgg1a (NA1_CONGI). The shared peptides identified in *Entacmaea quadricolor* were delta-actitoxin (ACTP_ENTQU), snake venom serine

protease (VSP_ECHOC), Fasciculin-2 (3SE2_DENAN), Fasciculin-1 (3SE1_DENAN), Delta-actitoxin-Cps1a (NA11_CONPS), Delta-actitoxin-Cgg1a (NA1_CONGI), and Delta-actitoxin-Cps1b (NA1P2_CONPS). The shared peptides identified in *Condylactis gigantea* were Delta-actitoxin-Cps1a (NA11_CONPS), Delta-actitoxin-Cgg1a (NA1_CONGI), and Delta-actitoxin-Cps1b (NA1P2_CONPS). The results of the database search of the raw spectral data against the cnidarian Uniprot database is given in Supplementary Table 3. The results of the database search of the spectral raw data against the animal toxin database is provided in Supplementary File 2. Out of these toxins, only NA1_CONGI was found to be present in the table of toxins identified using the RNA-seq approach. The transcripts per million (TPM) value for the homolog NA1_CONGI, which was identified in both the column and tentacle tissue of the *Condylactis gigantea* was 1433 in the column tissue and 39 in the tentacle tissue. When I use TPM as a variable to compare expression levels against, it normalizes for gene length first and then it normalizes for sequencing depth second. The TPM value for the homolog NA1_CONGI in *Entacmaea quadricolor* was however in the range of 0.5 to 1.

Table 5. Unique peptide and shared peptide coverage for *Condylactis gigantea.*

Protein Name	Unique Peptide Count	Shared Peptide Count	Sequence Coverage
NA11_CONPS	0	14	59.57%
NA1_CONGI	0	18	89.36%
NA1P2_CONPS	4	18	89.36%

Table 6. Unique peptide and shared peptide coverage for *Entacmaea quadricolor.*

Protein Name	Unique Peptide Count	Shared Peptide Count	Sequence Coverage
ACTP_ENTQU	6	6	37.50%
VSP_ECHOC	1	1	5.04%

3SE2_DENAN	0	1	34.43%
3SE1_DENAN	0	1	34.43%
NA11_CONPS	0	3	59.57%
NA1_CONGI	0	3	59.57%
NA1P2_CONPS	0	3	59.57%

Discussion

The primary aim of this study was to compare the putative protein toxins produced by a clownfish-hosting anemone with an anemone that does not host fishes, using both a RNA-seq approach and a MS/MS approach. I used sequence homology against manually curated animal toxins from the Uniprot animal toxin database, as a strategy for the identification of potential toxins from the transcriptomes of *Condylactis gigantea* and *Entacmaea quadricolor*. I further confirmed the existence of proteins in the venom proteome using MS/MS.

NA1_CONGI was the only toxin protein that was verified to be present in both species from both a RNA-seq and MS/MS approach. In the absence of a reference genome, I generated a *de novo* assembly using a methodology designed to maximize reference coverage, while minimizing redundancy and chimera rate (Yang and Smith, 2013). These transcripts were used for the identification of major toxin families present in the transcriptome and for the generation of a set of predicted proteins suitable for use in proteomics searches. These searches revealed the presence of proteins with homology to seven known toxin families including metalloproteinase, protease inhibitors, alpha-macroglobulin and CfTX proteins.

Although screening for known toxin families using a transcriptomic analysis will not reveal the presence of novel anemone-specific toxins or their functions, identification of multiple toxins belonging to Snaclec, Venom Kunitz type (VKT), and Phospholipase A in *Condylactis gigantea* and *Entacmaea quadricolor* suggest that their venom cocktail act through similar mechanisms as the major toxin families identified. Snaclec refers to C-type lectin-like proteins (CLPs) and is derived from the word: SNAke C-type LECtinS (snaclec). It is Lectin-like because of its similarity to classic C-type lectins, but different because of its hetero- or oligomeric structure. Their functions include anticoagulation, binding to prothrombin, and agonists of platelet aggregation (Drickamer, 1999). VKT-type proteins are mostly serine protease inhibitors (Masci et al., 2000; Choo et al., 2012), or they block potassium channels (Harvey, 2001), or both (Schweitz et al., 1995; Chen et al., 2012). Phospholipases A2 (PLA2s) toxins degrade phospholipids found in cell membranes. They are found in intracellular and extracellular forms in animals (Murakami, 2002). They may have different functions such as neurotoxicity, anticoagulant property, myonecrosis, cardiotoxicity, inhibition/activation of platelet aggregation, hemorrhage, and hemolysis (Kini, 2003).

UniProt release 2017_09 of Sep-27, 2017 contains 555,594 reviewed entries including 5,826 animal toxin entries. In this set, there are 274 cnidarian toxin proteins. Out of these 274 cnidarian toxins, 52 toxins were found in *Entacmaea quadricolor* and 47 were found in *Condylactis gigantea*. Using a proteomic approach only three proteins were found in *Condylactis gigantea* and seven proteins were found in *Entacmaea quadricolor*. The one toxin protein found in both species using MS/MS and RNA-seq

was NA1_CONGI, which slows the inactivation process of TTX-sensitive voltage-gated sodium channels. Standker et al. in 2006 has shown that this toxin has strong paralysis activity in crabs with a Lethal Dose 50 (LD50) of 1 µg/kg when the toxin is injected in between the body and walking legs of the crab. They also determined the molecular mass of this protein using MALDI to be 5043 Daltons from amino-acid positions 1 to 47. I observed that NA1_CONGI and NA1P2_CONPS had high sequence similarity as would be expected by evolution from a recent common ancestor. Both proteins were identified using a MS/MS study, but only NA1_CONGI was identified using my RNA-seq study. Although gene duplications have been found to be extensive in some species of sea anemones (Gacesa et al., 2015), conservation of these sequences suggest they are functional and not pseudo-mRNAs (Firth et al., 2006).

The small number of toxins identified via MS/MS relative to RNA-seq is puzzling. Brinkman et al. in 2015 also found that out of 455 toxins identified in box jellyfish (*Chironex fleckeri)* using RNA-seq, only 26 proteins were identified as putative toxins using MS/MS analysis. The 26 proteins they identified were also identified in the transcriptome. However, in my results, out of 79 toxins in *Condylactis gigantea* and 56 toxins in *Entacmaea quadricolor* identified using a RNA-seq approach, only three toxins in *Condylactis gigantea* and seven toxins in *Entacmaea quadricolor* were identified using a MS/MS study. Only one toxin, NA1_CONGI, was identified using both a RNA-seq approach and a MS/MS approach. This toxin had a high Transcript Per Million (TPM) value for *Condylactis gigantea* but a low TPM value for *Entacmaea quadricolor*. When I use TPM as a variable to compare expression levels, it normalizes

for gene length first and for sequencing depth second. The lower number of toxins identified in MS/MS compared to RNA-seq could be because of post-transcriptional modifications that prevented or reduced translation into protein.

From my results, I have found there is no relationship between number of fishes hosted and anemone toxicity. Previous studies of cnidarian toxins using an RNA-seq approach, by Brinkman et al. in 2015 reported that the extremely toxic box jellyfish, *Chironex fleckeri*, had 455 toxin homologies identified using RNA-seq data. The anemones in the present study had considerably fewer, with *Entacmaea quadricolor* having 79 toxins and *Condylactis gigantea* having 56 toxin homologies. These differences may be expected, given the highly toxic nature of box jellyfish venom compared to clownfish hosting anemones (Suput, 2009). Because *Entacmaea quadricolor* is an anemone that hosts clownfish, the higher number of toxin homologs in *Entacmaea quadricolor* relative to the non-host anemone *Condylactis gigantea* was unexpected. It is possible that the quantity of each toxin in *Entacmaea quadricolor* is lower than *Condylactis gigantea*. Thus, despite the possibility of more specific toxins in *Entacmaea quadricolor*, lower toxin expression levels could provide a more habitable environment for clownfish.

The presence of nematocysts distinguishes cnidarians from other phyla (Beckmann and Ozbek, 2012), but the roles of these stinging cells and their venoms in cnidarians are not completely understood. Specifically, my understanding of the mechanisms that allow anemones to use nematocysts to capture prey, including fishes, while simultaneously

not harming other fishes that live with them in symbiotic association, remain elusive (Rachamim et al., 2014). In my study, I have taken a step forward in investigating such behavior by exploring the toxin diversity in clownfish hosting and non-hosting sea anemones.

I note that investigation of differential expression using RNA-seq, complemented by a quantitative proteomics study, may provide further insights on toxin expression relative to anemone-clownfish symbiosis. Future studies might include biochemical investigations of the structural and functional properties of each putative toxin identified, examining novel toxins in other host anemone species, and investigating the activity of specific toxins on target molecules and potential adaptation in clownfish.

My results also provide insights on the diversity of toxins in different anemone groups, which is an important first step in understanding the biology of anemone toxins. A better understanding of anemone toxin resistance by clownfish may shed light on the biomedical utility of these toxins. Sea anemone toxin are currently being investigated for biological activity and drug development (Weston et al., 2013; Wilson et al., 2014; Cheng-Yi et al., 2014; Irina et al., 2016; Daly et al., 2014), including possible anti-carcinogenic properties (Marino et al., 2004; Ramezanpour et al., 2012). Anemone toxins are a cocktail of peptides, proteins, and non-peptide small molecule toxins. All cnidarians possess such toxins, yet their toxicity varies greatly among different groups within the phylum. Although these venoms have long been the subject of research (Irina et al., 2016), they remain less-well characterized than those of terrestrial taxa such as

snakes and scorpions (Lewis and Garcia, 2003). More research need to be performed on biochemical or bioinformatics characterization of anemone toxins. The current study elucidates the diversity of toxins in two sea anemones that differ in the presence or absence of their symbiotic relationship with clownfish.

References

Altschul, Stephen F., et al. "Basic local alignment search tool." *Journal of molecular biology* 215.3 (1990): 403-410.

Andrews, Simon. "FastQC: a quality control tool for high throughput sequence data." (2010): 175-176.

Beckmann, Anna, and Suat Özbek. "The nematocyst: a molecular map of the cnidarian stinging organelle." *International Journal of Developmental Biology* 56.6-7-8 (2012): 577-582.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. "Trimmomatic: a flexible trimmer for Illumina sequence data." *Bioinformatics* 30.15 (2014): 2114-2120.

Bosmans, Frank, and Jan Tytgat. "Sea anemone venom as a source of insecticidal peptides acting on voltage-gated Na+ channels." *Toxicon* 49.4 (2007): 550-560.

Brinkman, Diane L., et al. "Transcriptome and venom proteome of the box jellyfish Chironex fleckeri." *BMC genomics* 16.1 (2015): 407.

Chen, Zong-Yun, et al. "Hg1, novel peptide inhibitor specific for Kv1. 3 channels from first scorpion Kunitz-type potassium channel toxin family." *Journal of Biological Chemistry* 287.17 (2012): 13813-13821.

Chiang, Cheng-Yi, Yi-Lin Chen, and Huai-Jen Tsai. "Different visible colors and green fluorescence were obtained from the mutated purple chromoprotein isolated from sea anemone." *Marine biotechnology* 16.4 (2014): 436-446.

Choo, Young Moo, et al. "Antifibrinolytic role of a bee venom serine protease inhibitor that acts as a plasmin inhibitor." *PLoS One* 7.2 (2012): e32269.

Clemetson, K. J., T. Morita, and R. Manjunatha Kini. "Classification and nomenclature of snake venom C-type lectins and related proteins." *Toxicon* 54.1 (2009): 83.

Daly, Norelle L., Jamie Seymour, and David Wilson. "Exploring the therapeutic potential of jellyfish venom." *Future* 6.15 (2014): 1715-1724.

Dauplais, Marc, et al. "On the convergent evolution of animal toxins conservation of a diad of functional residues in potassium channel-blocking toxins with unrelated structures." *Journal of Biological Chemistry* 272.7 (1997): 4302-4309.

Debelius, Helmut. *Marine atlas: The joint aquarium care of invertebrates and tropical marine fishes*. Steven Simpson Books, 1994.

Drickamer, Kurt. "C-type lectin-like domains." *Current opinion in structural biology* 9.5 (1999): 585-590.

Dunn, Daphne Fautin. "The clownfish sea anemones: Stichodactylidae (Coelenterata: Actiniaria) and other sea anemones symbiotic with pomacentrid fishes." *Transactions of the American Philosophical Society* 71.1 (1981): 3-115.

Fautin, Daphne Gail, et al. "Field guide to anemonefishes and their host sea anemones." (1992).

Fautin, Daphne G. "The anemonefish symbiosis: what is known and what is not." (1991).

Frith, Martin C., et al. "Pseudo–messenger RNA: Phantoms of the transcriptome." *PLoS genetics* 2.4 (2006): e23.

Fry, Bryan G., et al. "The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms." *Annual review of genomics and human genetics* 10 (2009): 483-511.

Gacesa, Ranko, et al. "Gene duplications are extensive and contribute significantly to the toxic proteome of nematocysts isolated from Acropora digitifera (Cnidaria: Anthozoa: Scleractinia)." *BMC genomics* 16.1 (2015): 774.

Gladkikh, Irina, et al. "New Kunitz-type HCRG polypeptides from the sea anemone Heteractis crispa." *Marine drugs* 13.10 (2015): 6038-6063.

Haas, B. J., and A. Papanicolaou. "TransDecoder (find coding regions within transcripts)." (2016).

Haas, Brian J., et al. "De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis." *Nature protocols* 8.8 (2013): 1494-1512.

Hanlon, Roger T., and Raymond F. Hixon. "Behavioral associanons of coral reef fishes with the sea anemone Condylactis gigantea in the Dry Tortugas, Florida." *Bulletin of marine science* 39.1 (1986): 130-134.

Harvey, Alan L. "Twenty years of dendrotoxins." *Toxicon* 39.1 (2001): 15-26.

Baensch, Hans A., Helmut Debelius, and Helmut Debelius. "Marine Atlas." (1994).

Honma, Tomohiro, and Kazuo Shiomi. "Peptide toxins in sea anemones: structural and functional aspects." *Marine biotechnology* 8.1 (2006): 1-10.

Jouiaei, Mahdokht, et al. "Ancient venom systems: a review on cnidaria toxins." *Toxins* 7.6 (2015): 2251-2271.

Jungo, Florence, et al. "The UniProtKB/Swiss-Prot Tox-Prot program: a central hub of integrated venom protein data." *Toxicon* 60.4 (2012): 551-557.

Kearse, Matthew, et al. "Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data." *Bioinformatics* 28.12 (2012): 1647-1649.

Kim, Sangtae, and Pavel A. Pevzner. "Universal database search tool for proteomics." *Nature communications* 5 (2014): 5277.

Kim, Sangtae, Nitin Gupta, and Pavel A. Pevzner. "Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases." *Journal of proteome research* 7.8 (2008): 3354-3363.

Kini, R. Manjunatha. "Excitement ahead: structure, function and mechanism of snake venom phospholipase A 2 enzymes." *Toxicon* 42.8 (2003): 827-840.

Lewis, Richard J., and Maria L. Garcia. "Therapeutic potential of venom peptides." *Nature Reviews Drug Discovery* 2.10 (2003): 790-802.

Li, Rongfeng, et al. "Jellyfish venomics and venom gland transcriptomics analysis of Stomolophus meleagris to reveal the toxins associated with sting." *Journal of Proteomics* 106 (2014): 17-29.

Marino, Angela, et al. "Cytotoxicity of the nematocyst venom from the sea anemone Aiptasia mutabilis." *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* 139.4 (2004): 295-301.

Masci, P. P., et al. "Textilinins from Pseudonaja textilis textilis. Characterization of two plasmin inhibitors that reduce bleeding in an animal model." *Blood coagulation & fibrinolysis* 11.4 (2000): 385-393.

Mebs, D. "Anemonefish symbiosis: vulnerability and resistance of fish to the toxin of the sea anemone." *Toxicon*32.9 (1994): 1059-1068.

Miller, David J., et al. "The innate immune repertoire in Cnidaria-ancestral complexity and stochastic gene loss." *Genome biology* 8.4 (2007): R59.

Murakami, Makoto, and Ichiro Kudo. "Phospholipase A2." *The Journal of Biochemistry* 131.3 (2002): 285-292.

Nedosyko, Anita M., et al. "Searching for a toxic key to unlock the mystery of anemonefish and anemone symbiosis." *PloS one* 9.5 (2014): e98449.

Pentón, David, et al. "Validation of a mutant of the pore-forming toxin sticholysin-I for the construction of proteinase-activated immunotoxins." *Protein Engineering, Design & Selection* 24.6 (2011): 485-493.

Ponce, Dalia, et al. "Tentacle transcriptome and venom proteome of the pacific sea nettle, Chrysaora fuscescens (Cnidaria: Scyphozoa)." *Toxins* 8.4 (2016): 102.

Porat, D., and N. E. Chadwick-Furman. "Effects of anemonefish on giant sea anemones: ammonium uptake, zooxanthella content and tissue regeneration." *Marine and Freshwater Behaviour and Physiology* 38.1 (2005): 43-51.

Porat, D., and N. E. Chadwick-Furman. "Effects of anemonefish on giant sea anemones: expansion behavior, growth, and survival." *Hydrobiologia* 530.1-3 (2004): 513-520.

Putnam, Nicholas H., et al. "Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization." *Science* 317.5834 (2007): 86-94.

Rachamim, Tamar, et al. "The dynamically evolving nematocyst content of an anthozoan, a scyphozoan, and a hydrozoan." *Molecular biology and evolution* 32.3 (2014): 740-753.

Rachamim, Tamar, et al. "The dynamically evolving nematocyst content of an anthozoan, a scyphozoan, and a hydrozoan." *Molecular biology and evolution* 32.3 (2014): 740-753.

Schweitz, H., et al. "Kalicludines and kaliseptine: two different classes of sea anemone toxins for voltage-sensitive K+ channels." *Toxicon* 10.34 (1996): 1095.

Scott, Anna, and Peter Lynton Harrison. "Gametogenic and reproductive cycles of the sea anemone, Entacmaea quadricolor." *Marine Biology* 156.8 (2009): 1659.

Sheridan, Nancy E., Daphne Gail Fautin, and Matt J. Garrett. "Gametogenesis and reproductive periodicity of the biologically vulnerable giant Caribbean Sea anemone, Condylactis gigantea, in Florida." *Invertebrate biology* 134.2 (2015): 116-128.

Ständker, Ludger, et al. "A new toxin from the sea anemone Condylactis gigantea with effect on sodium channel inactivation." *Toxicon* 48.2 (2006): 211-220.

Šuput, Dušan. "In vivo effects of cnidarian toxins and venoms." *Toxicon* 54.8 (2009): 1190-1200.

Von Der Haar, Tobias. "Optimized protein extraction for quantitative proteomics of yeasts." *PloS one* 2.10 (2007): e1078.

Weston, Andrew J., et al. "Proteomic characterisation of toxins isolated from nematocysts of the South Atlantic jellyfish Olindias sambaquiensis." *Toxicon* 71 (2013): 11-17.

Yang, Ya, and Stephen A. Smith. "Optimizing de novo assembly of short-read RNA-seq data for phylogenomics." *BMC genomics* 14.1 (2013): 328.

Zhang, Zhaorui, et al. "High-throughput proteomics." *Annual review of analytical chemistry* 7 (2014): 427-454.

List of Tables

List of Figures

List of Supplementary Information

Chapter 3: Tissue-specific differential toxin gene expression analysis of toxins in two sea anemones

Abstract

Cnidarians are the simplest animals to have tissue level differentiation. They however do not have a centralized nervous system nor do they have a centralized venom delivery system. Sea anemones are a well-studied group within Cnidaria in regards to venom systems and neural systems. Tentacles of sea anemones have been the most widely used tissue from which toxins have been isolated. In this study, I use a RNA-seq approach to characterize the expression patterns and composition of venom and ion channels across different tissues (tentacles and column) in two species of sea anemone: *Entacmaea quadricolor* (common name: bubble tip anemone, family: actiniidae), and *Condylactis gigantea* (common name: condy anemone, family: actiniidae). These species vary in their symbiotic associations with clownfish and in their morphology. My results led to two conclusions: 1) Toxin diversity and expression levels vary across tissues of the sea anemone. The column tissue of sea anemones has more upregulated toxins. 2) A sea anemone with a higher number of clownfish symbiont had a lower expression level of toxins compared to another sea anemone with no host clownfish. By investigating such differences, I hope to lay the foundations for further work on clownfish-anemone symbiosis and anemone toxin composition, and the evolutionary relations between toxins and other proteins.

Keywords: Neurotoxins, Ion channels, Differential expression, Cnidaria, Transcriptomics.

Introduction

Cnidarians are the simplest animals to have tissue level differentiation (Holstein et al., 2013). They, however, do not have a centralized nervous system nor do they have a centralized venom delivery system. Snakes and spiders are known to have a centralized venom gland which eject peptide and small molecule toxins into its prey (Smith and Wheeler, 2006; Fry et al., 2009; Castelin et al., 2012; Casewell et al., 2013). Cnidarians however lack a centralized venom delivery system instead having stinging cells called nematocysts which is used to deliver venom. The composition of the nematocysts varies across different tissues of the animal (Mariscal, 1974; Kass-Simon and Scappaticci, 2002). The venom composition and expression of proteins in the nematocysts in different tissues of cnidarians is not well known (Casewell et al., 2013). All metazoans with the exception of poriferans, cnidarians, and ctenophores have a centralized nervous system. Cnidarians have a diffuse nerve net with each neuron holding as much importance as any other neuron.

Sea anemones are a well-studied group within cnidaria in regards to venom systems and neural systems. (Mariscal, 1974; Frazao et al., 2012; Reft and Daly, 2012; Jouiaei et al., 2015). Tentacles of sea anemones have been the most widely used tissue from which toxins have been isolated (Oliveira et al., 2012). The toxins from tentacles are reported to be used to prey capture, defense from predators, and territorial control (Frazao et al., 2012). The column of sea anemones also contains toxins although there have not been detailed studies characterizing the toxins from the column tissue (Frazao et al., 2012). Toxins in the nematocysts from this region of the animal could be used for competition

with non-clonal mates, predation or defense (Frazao et al., 2012). Although ion channel diversity has been known to be higher in cnidarians than bilaterans, its tissue level diversity and expression is not studied yet.

Previous tissue-specific investigations of cnidarians (esp. sea anemones) that undertook biochemical assays of venom extracts from tentacles have found that tissues thought to be originally devoid of nematocysts also have high concentrations of venom (Mathias et al., 1960). Tissue-specific transcriptome annotation provides us a list of candidate toxin genes and other target proteins such as ion channels in a comparative context. Differential expression analysis of tissue-specific venom composition has been able to identify the expression level and diversity for various sea anemones (Macrander et al., 2016). However, a tissue-specific transcriptome analysis of both venom compositions with a special focus on neurotoxins and its target proteins, ion channels, have not been undertaken.

In this study, I use a RNA-seq approach to characterize the expression patterns and composition of venom and ion channels across different tissues (tentacles and column) in two species of sea anemone: *Entacmaea quadricolor* (EQ, common name: bubble tip anemone, family: actiniidae), and *Condylactis gigantea* (CG, common name: condy anemone, family: actiniidae). These species vary in their symbiotic associations with clownfish and in their morphology. Envenomation from either of these anemones is not painful to humans. *Entacmaea quadricolor* is known for hosting a wide range of fishes while *Condylactis gigantea* is not known to host as many symbionts (Fautin, 2006).

These two specific species were also chosen for their ease of availability from aquarium trade shops. I predicted that I would identify differentially expressed toxins and ion channels that would correlate with unique tissue-specific functions and with variety of symbionts hosted by each species.

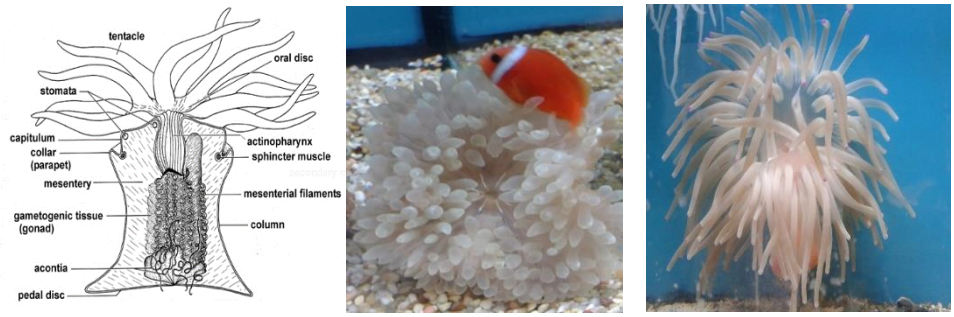


Figure 1. (left) Anatomy of a sea anemone (reproduced with permission from Brian McCloskey); (middle) a photograph of *Entacmaea quadricolor*, living with its clownfish symbiont; (right) *Condylactis gigantean*.

Methods

Sample collection and dissection

An individual each of *Entacmaea quadricolor* and *Condylactis gigantea* were purchased from a local pet store and were brought to the laboratory in marine water and sacrificed immediately. Parts of the tentacle region and column region were carefully dissected out. DNA extractions, COX1 sequencing, and morphological comparisons verified them to be of the species identified, as described in Chapter 2. Three samples of tentacle tissue and 3 samples of column tissue from the same individual of each species was used for RNA extraction, mRNA library preparation and 150bp paired end sequencing in an Illumina HiSeq 3000.

Transcriptome assembly and annotation

After Illumina sequencing, the quality of raw sequence data was assessed using FastQC (version 0.9.2) (Andrews, 2010). Illumina adapter sequences and low quality bases (Phred score > 32) were then removed from the sequence reads using Trimmomatic (Bolger et al., 2014). Reads shorter than 36 base pairs were discarded and the quality of filtered data was re-evaluated using FastQC. After quality control, paired-end sequences from six fastq files of each species were de novo assembled into contigs using the default parameters in Trinity (version 2.2) (Haas et al., 2013). After assembling with Trinity, the coding regions and proteins were predicted using Transdecoder (Haas and Papanicolaou, 2016). Thus, two protein databases for each animal were obtained which could later use for homology analysis.

In order to identify homologous proteins, transcripts of both animals were aligned to sequences available in a set of public databases using the tBLASTx and BLASTx algorithms (E-value cutoff of 1×10^{-5}) (Altschul et al., 1990). Searches were conducted against public and custom-made databases including (a) the Swiss-Prot database (as at 1 October, 2017) ;(b) the complete proteomic data sets of *Hydra vulgaris* and *Nematostella vectensis* from Uniprot; and (c) the UniProt animal toxin and venom database (Jungo et al., 2012). The sequences were visually verified to include start and stop codons using Geneious (Kearse et al., 2012). The transcriptome was functionally annotated using Trinotate.

Differential gene expression analysis of toxins

The raw reads were mapped back to the generated transcriptome assembly using Bowtie2 (Langmead and Salzberg, 2012). The expression levels in Counts Per Million (CPM) for each sample was calculated using eXpress (Roberts and Pachter, 2013) by inputting the raw Fasta files and the Sam files from Bowtie2. The annotation matrix obtained from blast results and the count matrix obtained from eXpress was merged into a single table. CPM was used because it can be used to perform differential expression analysis in EdgeR. The output from eXpress was fed into EdgeR available from the Bioconductor repository in R. This approach inputs tissue-specific CPM values directly into EdgeR without any pre-normalization. EdgeR was used to perform Likelihood Ratio tests between the tissue samples. EdgeR was preferred over DeSeq2 because of an excellent documentation and user manual. Transcripts per Million (TPM) was used as the comparison variable against the different tissues. When I use TPM as a variable to compare expression levels against, it normalizes for gene length first and then it normalizes for sequencing depth second. TPM was preferred over Reads per Kilobase per Millions (RPKM) or Fragments per Kilobase per Million (FPKM) because TPM eases the comparison of proportion of reads that map to a gene in each sample. This is done by making the sum of all TPMs in each sample to be the same. RPKM and FPKM are harder to use to compare samples directly because the sum of normalized read in each sample may be different. It is to be noted that TPM and FPKM are within-sample normalizations, which allows us to compare the quantitative levels of expressions of different genes from the same sample. The code used for analysis is available online at https://github.com/anuj2054/Differential_Expression

Results

Sequencing, assembly, and annotation

The raw reads from the three samples of tentacles of *Condylactis gigantea* were 14,088,619, then 13,829,038, and 13,464,882. This totaled 41,382,539 reads. The raw reads from the three samples of column of *Condylactis gigantea* was 13,400,056, then 15,777,480, and 12,479,267. This totaled 41,656,803 reads. The raw reads from the three samples of tentacles of *Entacmaea quadricolor* was 15,194,270, then 13,825,050, and 14,403,799. This totaled 43,423,119 reads. The raw reads from the three samples of column of *Entacmaea quadricolor* was 13,529,326, then 15,672,981, and 15,177,322. This totaled 44,379,629 reads. So the total number of reads for *Condylactis gigantea* was 83,039,342 and the total number of reads for *Entacmaea quadricolor* was 87,802,748. All triplicates for the column and tentacle tissue had similar quality reads, thus none of the triplicates had to be discarded.

Table 1: Statistics of transcriptome assembly and protein prediction.

Parameters	***Condylactis gigantea* (all tissue)**	***Entacmaea quadricolor* (all tissue)**	***Condylactis gigantea* Tentacles**	***Condylactis gigantea* Column**	***Entacmaea quadricolor* Tentacles**	***Entacmaea quadricolor* Column**
# PE Reads (2 × 150 bp)	83,039,342	87,802,748	41,382,539	41,656,803	43,423,119	44,379,629
Forward and Reverse recovered after trimming	71,866,086	75,555,907	36,103,260	35,762,826	37,640,851	37,640,851
Total trinity 'genes'	194,413	341,370	167,252	99,745	249,943	256,042
Total trinity transcripts	250,270	424,748	195,938	125,704	299,918	305,127
Median contig length based on contigs	502	381	507	468	404	415

Median length (N50) (bp) based on contigs	2,281	1,245	1,710	2,260	1,022	1,100
Average contig	1,108.30	730.27	976.95	1,072.58	701.10	727.55
Number of predicted peptides (Transdecoder)	142,654	140,594	109,566	54,527	115,986	104,359
Complete (ORF with both start and stop codons)	65,601	55,339	39,136	32,002	35,753	32,928
5' partial (ORF with start codon missing)	36,023	39,213	30,911	11,853	33,738	27,614
3' partial (ORF with stop codon missing)	9,810	12,430	8,254	3,482	10,069	9,047
Internal (ORF with no start or stop codons)	31,220	33,612	31,265	7,190	36,426	34,770

The number of homologs were combined from both the blastp and blastx results, regardless of abundance level of any particular homolog. For *Condylactis gigantea* the number of combined homologs were 67, while for *Entacmaea quadricolor* it was 93. For *Condylactis gigantea*, the column tissue had 61 homologs while the tentacle tissue had 65 homologs. For *Entacmaea quadricolor*, the column tissue had 83 homologs while the tentacle tissue had 86 homologs.

Table 2: Homology analysis of transcripts using blastx and blastp. The percentage of the total hits against the number of predicted proteins are indicated in brackets besides the raw number.

Homologs Found Using:	**Number of unique homologs for *Condylactis gigantea***	**Number of unique homologs for *Entacmaea quadricolor***	**Number of unique homologs for *Condylactis gigantea* Column**	**Number of unique homologs for *Condylactis gigantea* Tentacles**	**Number of unique homologs for Column tissue in *Entacmaea quadricolor***	**Number of unique homologs for tentacle tissue in *Entacmaea quadricolor***
BLASTx against Swissprot	24,775 (9.80%)	91,745 (21.5%)	15,210	23,093	24,028	26,219

BLASTx against Swissprot, Metazoans Only	16,167 (6.45%)	76,575 (18.02%)	13,214	15,805	17,113	18,863
BLASTx against Swissprot, Cnidarians Only	199 (0.07%)	1,898 (0.4%)	190	202	219	222
BLASTX against SwissProt, Toxins	56 (0.02%)	79 (0.01%)	41	53	66	67
BLASTX against SwissProt, Cnidarian Toxins	36 (0.014%)	50 (0.008%)	31	34	41	43
BLASTX against ion channels in Swissprot	79	98	54	74	88	86
BLASTP against Swissprot	77,238 (54.14%)	25,616 (18.2%)	14,314	22,172	21,723	23,867
BLASTP against Swissprot, Metazoans Only	42,407 (29.72%)	18,539 (13.18%)	12,770	15,442	15,699	17,094
BLASTP against Swissprot, Cnidarians Only	801 (5.6%)	200 (0.14%)	174	174	180	188
BLASTP against SwissProt, Toxins Only	43 (0.03%)	52 (0.04%)	30	37	38	46
BLASTP against SwissProt, Cnidarian Toxins	19 (0.01%)	28 (0.02%)	20	18	17	24
BLASTP against Swissprot, ion channels	86	89	57	83	79	83

Mapping, abundance, and Gene Ontology (GO) analysis

A bowtie read mapping and eXpress transcript abundance was performed. In column tissue of *Entacmaea quadricolor*, 3791,5056 (100.00%) were paired; of these:

7,198,348 (18.99%) aligned concordantly 0 times, 11,323,639 (29.87%) aligned concordantly exactly 1 time, 19,393,069 (51.15%) aligned concordantly >1 times. This gave an 96.41% overall alignment rate. In tentacle tissue of *Entacmaea quadricolor*, 37,640,851 raw reads (100%) were paired; of these: 7,356,672 (19.54%) aligned concordantly 0 times, 11,458,659 (30.44%) aligned concordantly exactly 1 time, 18,825,520 (50.01%) aligned concordantly >1 times. This gave an 96.19% overall alignment rate. In *Condylactis gigantea* tentacles, 36,103,260 reads; of these: 36,103,260 (100.00%) were paired; of these: 5,235,652 (14.50%) aligned concordantly 0 times, 9,428,559 (26.12%) aligned concordantly exactly 1 time, 21,439,049 (59.38%) aligned concordantly >1 times. This gave an 97.71% overall alignment rate. In *Condylactis gigantea* column, 35,762,826 reads; of these: 35,762,826 (100.00%) were paired; of these: 4,698,581 (13.14%) aligned concordantly 0 times, 9,105,415 (25.46%) aligned concordantly exactly 1 time, 21,958,830 (61.40%) aligned concordantly >1 times. This gave an 98.12% overall alignment rate.

Table 3: Abundance levels and number of homologs. The table shows the mean expression level in TPM and the number of homologs using Blastx found for each tissue type in the two animals according to the toxin protein family, this includes all homologs regardless of the TPM value. The percentage of the total hits against the number of predicted proteins are indicated in brackets besides the raw number.

Toxin family	**Number and Average TPM for all candidate toxin genes in *Condylactis gigantea***	**Number and Average TPM for all candidate toxin genes in *Condylactis gigantea* column**	**Number and Average TPM for all candidate toxin genes in *Condylactis gigantea* tentacles**	**Number and Average TPM for all candidate toxin genes in *Entacmaea quadricolor***	**Number and Average TPM for all candidate toxin genes in *Entacmaea quadricolor* column**	**Number and Average TPM for all candidate toxin genes in *Entacmaea quadricolor* tentacles**
All toxins	67 (22.50)	61 (25.72)	65 (19.27)	93 (13.66)	83 (16.14)	86 (11.17)

SCRiP	0	0	0	1 (80.04)	1 (111.59)	1(48.49)
Complement C3 homolog	2 (22.49)	2 (24.6)	2(20.32)	2 (2.23)	2 (1.8)	2 (2.67)
CRISP	0	0	0	3 (0.31)	2 (0.18)	3 (0.45)
Flavin monoamine oxidase	1 (0.88)	0	1(1.75)	1 (0.76)	1 (0.68)	1 (0.84)
M12A	1 (1.10)	1 (0.479)	1 (1.72)	3 (0.44)	3 (0.53)	2 (0.35)
Multicopper oxidase	3 (6.90)	3 (4.3)	3(9.44)	3 (5.63)	3 (3.93)	3 (7.34)
Natriuretic peptide	1 (0.02)	0	1(0.046)	0	0	0
Nucleotide pyrophosphatase/phosphodiesterase	2 (0.93)	2 (0.76)	2(1.101)	1 (1.33)	1 (0.97)	1 (1.70)
serine protease	0	0	0	1 (2.76)	1(0.50)	1 (5.02)
PA2	6 (28.90)	6 (39.09)	6(18.7)	6 (2.39)	6 (2.17)	6 (2.62)
PB	1 (0.94)	1 (1.33)	1(0.54)	0	0	0
Snaclec	8 (1.77)	6 (1.799)	7(1.73)	10 (0.52)	8 (0.47)	9 (0.57)
CPT.	0	0	0	1 (0.39)	1 (0.6)	1 (0.10)
True venom lectin (C-type lectin)	1 (0.75)	1 (0.173)	1(1.33)	1 (0.22)	1 (0.27)	1 (0.17)
Type-B carboxylesterase/lipase	1 (4.27)	1 (4.452)	1(4.08)	1 (0.88)	1 (0.99)	1 (0.76)
VKT	4 (13.75)	4 (19.31)	4(8.17)	7 (5.18)	6 (5.29)	6 (5.08)
M12B	2 (0.47)	1 (0.015)	2(0.93)	3 (0.24)	2 (0.17)	3 (0.30)
Unassigned to a family	34 (231.78)	33 (263. 67)	33(199.89)	49 (115.16)	44 (128.01)	45 (102.31)

From the toxins above, I then tabulated the highly expressed toxins by filtering only the toxins that had TPM greater than 10. A table similar to Table 3 was created. It was found that DELTA-actitoxin-Aeq1b (ACTP5), Delta-stichotoxin-Hcr4a (ACTPA), and U-actitoxin-Avd8a (TX8A) had the highest TPM values in the tentacles of *Condylactis gigantea*. It was found that U-actitoxin-Avd8a (TX8A), U-actitoxin-Avd8d (TX8D), Delta-actitoxin-Cgg1a (NA1), and U-actitoxin-Bcs2a (BDS2A) had the highest TPM values in the column of *Condylactis gigantea*. It was found that DELTA-thalatoxin-Avl2a (TX60A), U-actitoxins (TX8B and TX8D) had the highest TPM values in the tentacles of Entacmaea quadricolor. It was found that U-actitoxins (TX8B and TX8D) and Delta-thalatoxin-Avl2a (TX60A) had the highest TPM values in the column of *Entacmaea quadricolor* (as shown in Figure 2).

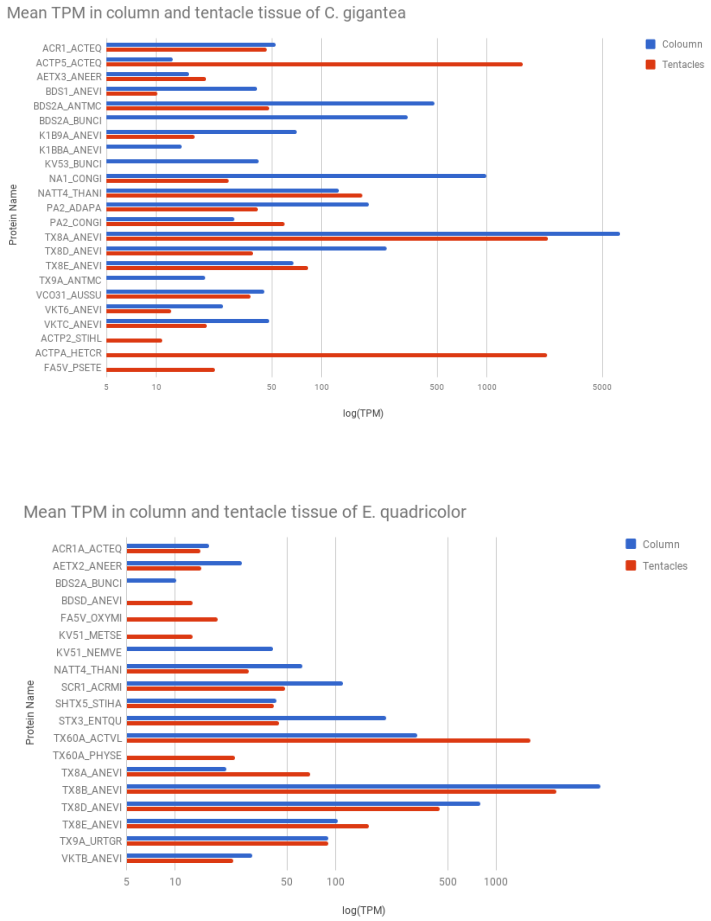


Figure 2. (Top) Highly expressed toxins in *Condylactis gigantea*. TPM values of toxins that have TPM>10 in either the tentacles or column of *Condylactis gigantea*. **(Bottom)Highly expressed toxins in *Entacmaea quadricolor*.** TPM values of toxins that have TPM>10 in either the tentacles or column of *Entacmaea quadricolor*.

GO analysis for highly expressed toxins in *Condylactis gigantea*

In total, 23 different types of toxins were highly expressed in column and tentacles of *Condylactis gigantea*. 20 proteins were highly expressed in column of *Condylactis gigantea* while 19 proteins were highly expressed in tentacles of *Condylactis gigantea*. I then did a GO analysis, where 11 proteins in column of Condylactis gigantea and 12 proteins in tentacles of *Condylactis gigantea* had the GO term “molecular function.” GO term “biological function” were shown by 4 and 10 proteins in column of

Condylactis gigantea and tentacles of *Condylactis gigantea* respectively. Three dominant biological processes were selected: cellular (3 in column of *Condylactis gigantea* and 5 in tentacles of *Condylactis gigantea*), metabolic (3 in column of *Condylactis gigantea*, 4 in tentacles of *Condylactis gigantea*) and biological regulation (4 in column of *Condylactis gigantea*, 4 in tentacles of *Condylactis gigantea*) processes. The mean TPM value were more than 10 for the following toxins in column of *Condylactis gigantea*: BDS2A_BUNCI (330.042), BDS2A_ANTMC (480.95), NA1_CONGI (992.81) and TX8A_ANEVI (6,422.29). In tentacles of *Condylactis gigantea*, a high mean TPM value was observed in NATT4_THANI (176.255), ACTP5_ACTEQ (1,644.27), ACTPA_HETCR (2,309.68) and TX8A_ANEVI (2,326.60). In both column of *Condylactis gigantea* and tentacles of *Condylactis gigantea*, TX8A_ANEVI (6,422.29 in column of *Condylactis gigantea*, and 2,326.60 in tentacles of *Condylactis gigantea*) was common and reported the highest TPM value.

GO analysis for highly expressed toxins in *Entacmaea quadricolor*

In total, 19 different types of toxins were highly expressed in column and tentacles of *Entacmaea quadricolor*. In the column tissue of *Entacmaea quadricolor*, 15 proteins were highly expressed while 17 proteins were highly expressed in tentacles. In GO analysis of the highly expressed toxins for *Entacmaea quadricolor*, 4 proteins in column and 5 proteins in tentacles showed the GO term “molecular function.” GO term “cellular component” were shown by 15 and 17 proteins in column and tentacles of *Entacmaea quadricolor* respectively. Three dominant cellular components were selected: intracellular (14 in column of *Entacmaea quadricolor*, 15 in tentacles of *Entacmaea quadricolor*), extracellular (15 in column of *Entacmaea quadricolor*, 17 in

its tentacles) and nematocysts (14 in column of *Entacmaea quadricolor*, 15 in its tentacles). The mean TPM value was more than 10 for the following toxins in column of *Entacmaea quadricolor*: STX3_ENTQU (205.682) TX60A_ACTVL (321.396), TX8D_ANEVI (794.68) and TX8B_ANEVI (4481). In tentacles of *Entacmaea quadricolor*, a high TPM value was observed in TX8E_ANEVI (160.677), TX8D_ANEVI (444.489), TX60A_ACTVL (1637.299) and TX8B_ANEVI (2366.929). TX8D_ANEVI, TX60A_ACTVL and TX8B_ANEVI were the most abundant proteins common in both column of *Entacmaea quadricolor* and its tentacles. The highest TPM value was reported by TX8B_ANEVI (4481 in column of *Entacmaea quadricolor*, 2366.929 in its tentacles). In column of *Entacmaea quadricolor*, AETX2_ANEER, ACR1A_ACTEQ, BDSD_ANEVI, KV51_METSE, SCR1_ACRMI, SHTX5_STIHA, STX3_ENTQU, TX8E_ANEVI, TX8A_ANEVI, TX8D_ANEVI, TX60A_PHYSE, TX8B_ANEVI, TX60A_ACTVL, TX9A_URTGR and VKTB_ANEVI were found to have associated with GO term "nematocyst." In tentacles of *Entacmaea quadricolor*, AETX2_ANEER, ACR1A_ACTEQ, BDS2A_BUNCI, KV51_NEMVE, SCR1_ACRMI, SHTX5_STIHA, STX3_ENTQU, TX8E_ANEVI, TX8A_ANEVI, TX8D_ANEVI, TX8B_ANEVI, TX60A_ACTVL, TX9A_URTGR and VKTB_ANEVI were found to have associated with GO term "nematocyst."

Differential expression and GO analysis

For *Entacmaea quadricolor*, no differentially expressed toxins were found for a P value of 0.001, but using a P value of 0.01, I were able to find 3 differentially expressed toxins which were CTXA_CARAL, ACTP5_ACTEQ, and X60A_PHYSE. For *Condylactis gigantea*, using a P value of 0.001, seven toxins were found to be

differentially expressed which were BDS2A_BUNCI, NA1_CONGI, KV53_BUNCI, LECG_THANI, CTX1_CARRA, ACTP2_STIHL, and CTP2_ACTEQ (as shown in Table 4). Three toxins (BDS2A_BUNCI, NA1_CONGI, and KV53_BUNCI) were upregulated in column and four toxins (LECG_THANI, CTX1_CARRA, ACTP2_STIHL, and CTP2_ACTEQ) were upregulated in tentacles for *Condylactis gigantea*. Two toxins (CTXA_CARAL, and ACTP5_ACTEQ) were upregulated in column and one toxin (X60A_PHYSE) was upregulated in tentacles for *Entacmaea quadricolor* (as shown in Figure 2 and Figure 3).

Table 4. Highly expressed and differentially expressed toxins

Parameters	**Number of Toxins identified with TPM > 10**	**Number of Toxins differentially expressed**
Entacmaea quadricolor Column	12	3
Entacmaea quadricolor Tentacle	15	3
Condylactis gigantea Column	20	7
Condylactis gigantea Tentacle	19	7

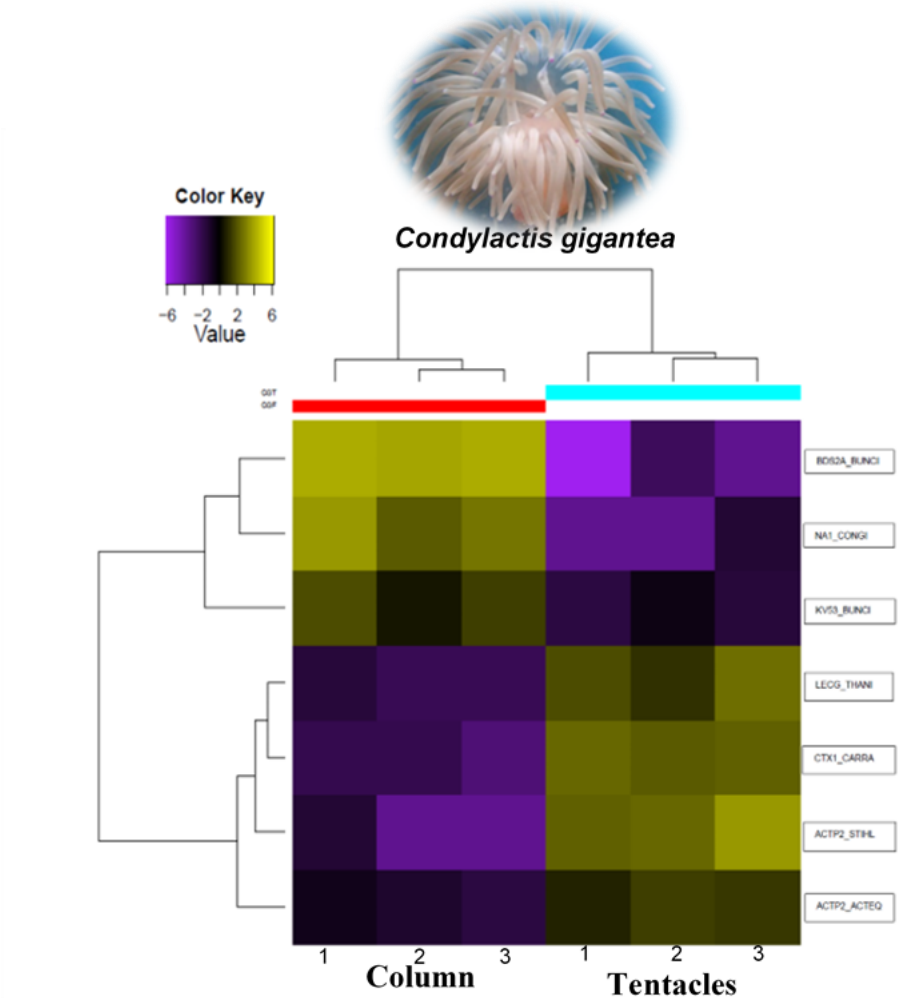


Figure 3. Differentially expressed toxins in *Condylactis gigantea*. A p value of 0.001 was used and the log of the CPM was taken to construct the heatmap for the toxins in tentacles and column of *Condylactis gigantea*. The dendrogram at the left of the heatmap represents genes that are clustered together by their expression levels. BDS2A_BUNCI, NA1_CONGI, and KV53_BUNCI are upregulated in the column while LECG_THANI, CTX1_CARRA, ACTP2_STIHL, and ACTP2_ACTEQ are upregulated in the tentacles of the animals.

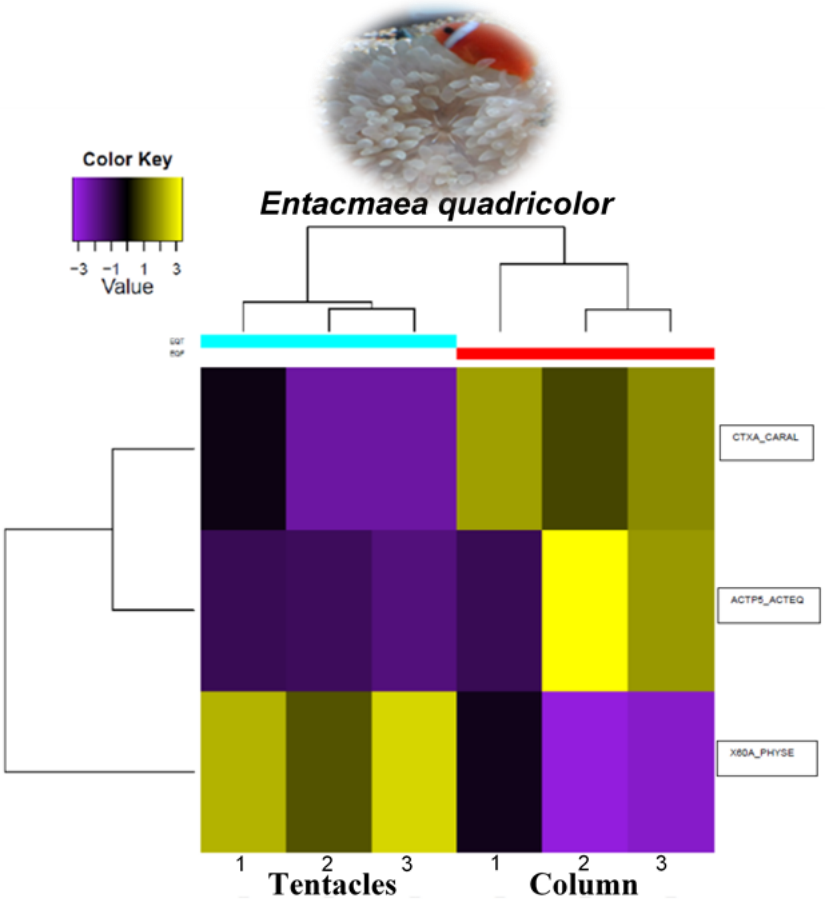


Figure 4. Differentially expressed toxins in *Entacmaea quadricolor*. A p value of 0.01 was used and the log of the CPM was taken to construct the heatmap for the toxins in tentacles and column of *Entacmaea quadricolor*. The dendrogram at the left of the heatmap represents genes that are clustered together by their expression levels. CTXA_CARAL and ACTP5_ACTEQ are upregulated in the column while X60A_PHYSE is upregulated in the tentacles of the animals.

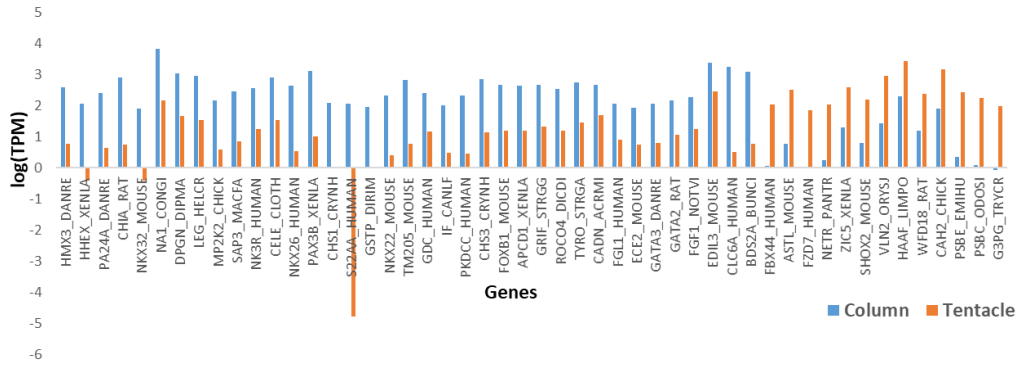


Figure 5. Top 50 differential expressed genes in *Condylactis gigantea.* A p value of 0.001 was used and the log of the CPM was taken to construct the graph for the all proteins in tentacles and column of *Condylactis gigantea.* Genes in which the blue bars are higher than the red bars are the genes which are upregulated in the column of the animal. Genes in which the red bars are higher than the blue bars are the genes which are upregulated in the tentacles of the animal.

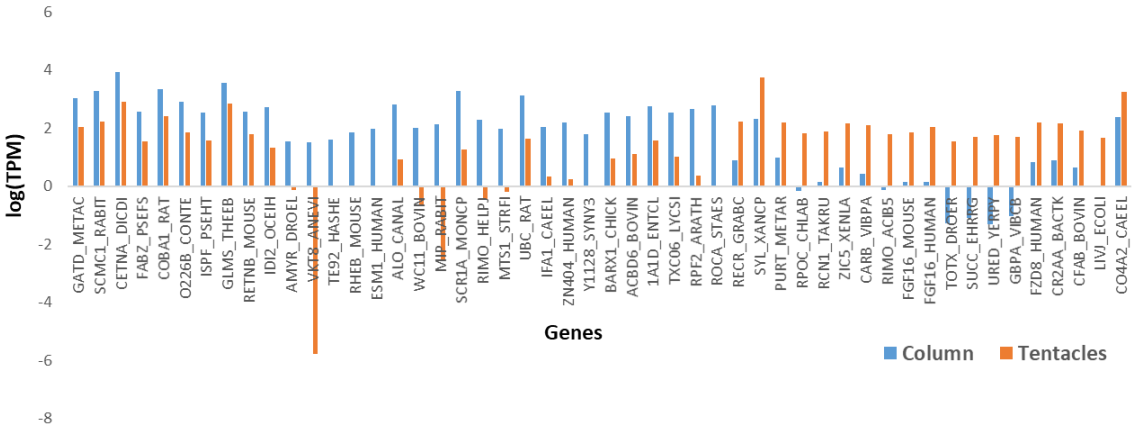


Figure 6. Top 50 differentially expressed genes in *Entacmaea quadricolor*. A p value of 0.001 was used and the log of the CPM was taken for the genes in tentacles and column of *Entacmaea quadricolor*. Genes in which the blue bars are higher than the red bars are the genes which are upregulated in the column of the animal. Genes in which the red bars are higher than the blue bars are the genes which are upregulated in the tentacles of the animal.

GO analysis for differentially expressed genes in *Condylactis gigantea*

Out of the 50 top differentially expressed proteins, 13 proteins were upregulated in tentacles of *Condylactis gigantea* while 37 proteins were upregulated in column of *Condylactis gigantea*. However, when I did a GO analysis, proteins that were found to have GO term "molecular function" were 33 (89%) in column of *Condylactis gigantea* and 12 (92%) in tentacles of *Condylactis gigantea*. 24 (72%) proteins had the GO term "binding" in column of *Condylactis gigantea* while only 9 (75%) proteins in tentacles of *Condylactis gigantea*. 12 (50%) proteins had GO term "ion binding" in column of *Condylactis gigantea* while 6 (66%) proteins in tentacles of *Condylactis gigantea*. GO term "biological function" were shown by 33 (89%) and 11 (91%) proteins in column of *Condylactis gigantea* and tentacles of *Condylactis gigantea* respectively. Three dominant biological processes were selected: cellular (26 in column of *Condylactis gigantea* and 10 in tentacles of *Condylactis gigantea*), metabolic (18 in column of *Condylactis gigantea,* 7 in tentacles of *Condylactis gigantea*) and developmental (16 in column of *Condylactis gigantea*, 3 in tentacles of *Condylactis gigantea*) processes. In Condylactis gigantea, two proteins, NA1_CONGI and BDS2A_BUNCI, were among the top 50 differentially expressed were found to have GO term "nematocyst" (with ID 42151) associated with it.

GO analysis for differentially expressed genes in *Entacmaea quadricolor*

Out of the top 50 differentially expressed genes, 19 proteins were upregulated in tentacles of *Entacmaea quadricolor* while 31 proteins were upregulated in its column. In GO analysis, proteins that were found to have GO term "molecular function" were 26 (42%) in column of *Entacmaea quadricolor* and 17 (89%) in its tentacles. 20 (77%)

proteins had the GO term “binding” in column of *Entacmaea quadricolor* while only 15 (88%) proteins in its tentacles. 13 (65%) proteins had GO term “ion binding” in column of *Entacmaea quadricolor* while 9 (60%) proteins in its tentacles. GO term “biological function” were shown by 22 (71%) and 17 (89%) proteins in column of *Entacmaea quadricolor* and its tentacles respectively. Four dominant biological processes were selected: cellular (17 in column of *Entacmaea quadricolor* and 14 in its tentacles), metabolic (15 in column of *Entacmaea quadricolor,* 10 in tentacles of *Entacmaea quadricolor*), biological regulation (7 in column of *Entacmaea quadricolor,* 5 in tentacles of *Entacmaea quadricolor*) and developmental (4 in column of *Entacmaea quadricolor,* 5 in tentacles of *Entacmaea quadricolor*) processes. In *Entacmaea quadricolor,* two proteins, VKT8_ANEVI, and SCR1A_MONCP, that were among the top 50 differentially expressed were found to have GO term “nematocyst” (with GO identity 42151) associated with it.

Discussion

The goal of this study was to compare the expression of toxins and other proteins in tentacles of a sea anemone against an another tissue in the same animal. The other tissue used in my study was the column tissue of the same sea anemone. No difference was seen in the number and expression levels of toxins and proteins identified in the tentacles vs. the column of sea anemones investigated. However, a difference was seen in the types of toxins and proteins that were differentially expressed in the tissues of the two animals. A difference was also seen in the overall expression levels of toxins between the two animals using data from all tissue.

Tissue-specific diversity and quantitative expression of toxins

More snaclec, VKT, and PL2 was found in all tissues of the two sea anemone species. A large difference was not found in the number of candidate toxin genes or homologs between the two tissues sampled. Almost same number of toxin homologs in tentacle and column was discovered. When comparing tissue-specific toxin gene expression, as shown in Figure 2 and Figure 3, no difference in number or quantitative expression was seen between the tentacles and the column tissue of either of the animals. My studies are consistent with previous studies by Macrander et al. in 2016, where they have found almost the same quantitative expression of candidate toxin gene families in different tissues of animals such as *Anemonia sulcata.*

As shown in Figure 2 and Figure 3, more quantitatively expressed toxins were found in *Condylactis gigantea* compared to *Entacmaea quadricolor*. In total, 23 toxins in *Condylactis gigantea* and 19 toxins in *Entacmaea quadricolor* were quantitatively expressed.

In *Condylactis gigantea*, I report that a higher quantity of column dominant neurotoxins (BDS2A_BUNCI, BDS2A_ANTMC, NA1_CONGI and TX8A_ANEVI) belonging to sea anemone type family. Neurotoxins are responsible for the excitation of cells affecting Na+ or K+ permeability. Neurotoxins in column inhibit the voltage-gated Na+ (NA1_CONGI) channels which causes electrical imbalance leading to the paralysis (e.g. BDS2A_ANTMC has PD50: 420 µg/kg) of the crab (Orts D.J et al., 2013). Other *Condylactis gigantea* toxins such as: NATT4_THANI (natterin family),

ACTP5_ACTEQ (actinoporin family), ACTPA_HETCR (actinoporin family) and TX8A_ANEVI (sea anemone family) were abundant in tentacles. Toxin NATT4_THANI has nociceptive, edema-inducing and kinogenase activity. ACTP5_ACTEQ and ACTPA_HETCR are the pore-forming proteins that have channeling activity and hemolysis activity (Monastyrnaia M.M et al., 2002). These toxins disrupt the ion concentration gradients in cells leading to hemolysis. ACTPA_HETCR is lethal to mice (LD50:50 µg/kg) and injected intraperitoneally. Tentacle abundant toxins in *Condylactis gigantea* might be useful for prey acquisition. Highest TPM value was expressed by toxin TX8A_ANEVI in both column and tentacles of *Condylactis gigantea.* This toxin was expressed in quantitative analysis but was not reported in differential analysis.

Toxins STX3_ENTQU, TX60A_ACTVL, TX8D_ANEVI and TX8B_ANEVI belonging to sea anemone family reported high mean TPM value in column of *Entacmaea quadricolor*. In tentacles of *Entacmaea quadricolor,* a high mean TPM value was observed in TX8E_ANEVI, TX8D_ANEVI, TX60A_ACTVL and TX8B_ANEVI. Toxins TX8D_ANEVI, TX60A_ACTVL and TX8B_ANEVI were common in both column and tentacles of *Entacmaea quadricolor*. Toxin STX3_ENTQU (LD50:10 pg/kg) is lethal when injected into crabs (Ishikawa Y et al., 1979). Toxin TX60A_ACTVL is lethal to mice (Oshiro N et al., 2004). Most abundant toxins in column of *Entacmaea quadricolor* and tentacles of *Entacmaea quadricolor* are reported to disrupt action potential (Ishikawa Y et al., 1979) and have hemolysis activity. As in tentacles of *Condylactis gigantea*, the abundance of toxin in tentacles of

Entacmaea quadricolor might be useful for defense and prey acquisition while the toxins in column of *Entacmaea quadricolor* might be useful for digestion and other biological processes. Toxin TX8B_ANEVI reported the highest TPM value in both column and tentacles of *Entacmaea quadricolor*. This toxin was expressed in quantitative analysis but was not reported in differential analysis. Although toxins TX8A_ANEVI and TX8B_ANEVI (in both column and tentacles of *Condylactis gigantea*) were found to be highly expressed during abundance analysis, neither of them were found to be differentially expressed. It would be of further interest to perform a qPCR analysis of the toxins identified for the confirmation of results.

It is also to be noted that *Entacmaea quadricolor* had lower TPM for toxins while *Condylactis gigantea* have higher TPM for toxins. The two animals were chosen because of the ability of *Entacmaea quadricolor* to host more clownfish while *Condylactis gigantea* does not host clownfish in the wild. It could be that the higher expression level of toxins in *Condylactis gigantea* could have made *Condylactis gigantea* an uninhabitable host for symbiosis for the clownfish.

Tissue-specific differential expression of toxins

As shown in Figure 3 and Figure 4, there was found to be more differentially expressed toxins in *Condylactis gigantea* compared to *Entacmaea quadricolor*. Four toxins were differentially expressed in *Condylactis gigantea* and seven toxins were differentially expressed in *Entacmaea quadricolor*.

This experimental study showed that CTXA_CARAL (jellyfish toxin family) and ACTP5_ACTEQ (actinoporin family) was reported higher in column of *Entacmaea quadricolor* while toxin X60A_PHYSE was abundant in tentacles. The abundance of toxin X60A_PHYSE might be useful for defense and prey acquisition while the other two toxins (i.e. CTXA_CARAL, ACTP5_ACTEQ) abundant in the column form a membrane channel in the prey leading to disruption of ion concentration gradients (Oliveira JS et al., 2012). Both toxins have hemolytic activity, with CTXA_CARAL (LD50:5-25 µg/kg) being lethally toxic to crayfish through intraperitoneal route of administration (Nagai H et al., 2000).

In *Condylactis gigantea*, I report a higher quantity of column dominant neurotoxins (BDS2A_BUNCI, NA1_CONGI and KV53_BUNCI) belonging to sea anemone type family. As neurotoxins play a vital role in excitation of cells affecting Na+ or K+ permeability, output secretory activities affecting the release of neurotransmitter or input generator activities affecting the receptor molecules for transmitter themselves. Neurotoxins in column inhibit the voltage-gated Na+ (NA1_CONGI, LD50:1µg/kg) or K+ (KV53_BUNCI) channels which disrupt the electrical impulses ultimately paralyze crabs (*Xenopus laevis*) that the toxin has been tested on (Orts et al., 2013). Other *Condylactis gigantea* toxins such as: LECG_THANI (true venom lectin family), CTX1_CARRA (jellyfish toxin family), ACTP2_STIHL (actinoporin family) and CTP2_ACTEQ were abundant in tentacles. LECG_THANI is a galactose specific lectin toxin that has hema-glutination activity and pro-inflammatory activity which induce neutrophil mobilization (Lopes-Ferreira M et al., 2011). CTX1_CARRA is lethal to

mice (LD50:20 µg/kg injected intravenously) and (LD50:5µg/kg injected intraperitoneally) crayfish (Nagai H et al., 2000). CTX1_CARRA and ACTP2_STIHL are the pore-forming toxins which disrupt the ion concentration gradients in cells leading to hemolysis. As in *Entacmaea quadricolor*, tentacle abundant toxins in *Condylactis gigantea* might also be useful for prey acquisition.

In both *Condylactis gigantea* and *Entacmaea quadricolor*, as shown in Figure 5 and Figure 6, a larger percentage of upregulated proteins in tentacles were found to be involved in molecular functions and biological processes compared to the upregulated proteins in column. The main basis of this finding may be because tentacles act as the primary tissue for prey capture. Most of the toxin peptides in anemones are derived from tentacles (Oliveira et al., 2012). Tentacles are also multifunctional to immobilize prey (e.g. neurotoxins) and repel predators (e.g. acid sensing ion channel targeting toxins). Past studies have shown that one of the proteins that is exclusively expressed in the column of a hydra (*Hydra vulgaris)* is Translationally-controlled tumor protein homolog (TCTP), which is involved in cell proliferation (Yan et al., 2000). However, in my studies, I did not find TCTP homologs.

Among the top 50 differentially expressed genes in *Condylactis gigantea*, the only toxins were NA1_CONGI and BDS2A_BUNCI, which were both upregulated in the column. NA1_CONGI acts by delaying the inactivation process of voltage-gated sodium channels (Standker et al., 2006). BDS2A_BUNCI has been shown to have ion channel inhibitory activity (Oliveira et al., 2006). Both these toxins are thus neurotoxic

in nature and have been shown to cause paralysis in crabs. Among the top 50 differentially expressed genes in *Entacmaea quadricolor*, the only toxins were VKT8_ANEVI and SCR1A_MONCP, which were both upregulated in the column. There have been no published physiological studies on VKT8_ANEVI, but by sequence similarity to VKT2_ANESU it may be postulated to have similar functions; that is, it can inhibit the serine protease trypsin or inhibit voltage-gated potassium channels (Kv1.2/KCNA2). SCR1A_MONCP has been suggested to be involved in calcification (Sunagawa, et al., 2009), but sequence similarity to SCR2_ACRMI and SCR3_ACRMI, both of which have caused neurotoxic symptoms in zebrafish, suggests that SCR1A_MONCP may also have neurotoxic activity. Thus, all the top differentially expressed toxins were found to have neurotoxic functions, but were surprisingly upregulated in the column tissue instead of the tentacle tissue, like I had predicted.

Differential expression studies on toxins give more meaningful insights when tissues are exposed to different environments (Lopez-Maury et al., 2008) over different time periods. In the future, such studies could be performed by exposing anemones to their host inhabitants such as clownfish vs not exposing them to the host inhabitant. They could also be performed by exposing the anemones to stimulants that could incite the animal to release its toxins or not exposing them to such stimulants. Future analysis on tissue-specific toxin expression could also benefit by combining transcriptomic data with proteomic data. A large number of transcripts from the transcriptome assembly did not return any Blast or Hmmer hits. This points to the growing need for better genomic resources on cnidarians.

Conclusion

Unlike other venomous animals, sea anemones do not have specialized venom glands where most of the venom is localized. This characteristic of these animals leads to the question of which tissue regions of the animal have what types of toxins and what their expression levels are. Such questions are important because of the potential pharmaceutical applications of the toxins from these animals (Frazao et al., 2012). My approach tried to answer these questions by selecting two contrasting anemones: one that hosts clownfish and another that does not. My results led to three conclusions: 1) Toxin diversity and expression levels vary across tissues of the sea anemone. The column tissue of sea anemones tended to have more upregulated toxins. 2) Since toxin sequences are short, most of the toxins identified in this study have to be verified using either a proteomics approach or a thorough biochemical approach. 3) A sea anemone with a higher number of clownfish symbiont had a lower expression level of toxins compared to another sea anemone with no host clownfish. By answering such questions, I hope to lay the foundations for further work on clownfish-anemone symbiosis and anemone toxin composition, and the evolutionary relations between toxins and other proteins.

References

Altschul, Stephen F., et al. "Basic local alignment search tool." *Journal of molecular biology* 215.3 (1990): 403-410.

Andrews, Simon. "FastQC: a quality control tool for high throughput sequence data." (2010): 175-176.

Botstein, David, et al. "Gene Ontology: tool for the unification of biology." *Nat Genet* 25.1 (2000): 25-9.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. "Trimmomatic: a flexible trimmer for Illumina sequence data." *Bioinformatics* 30.15 (2014): 2114-2120.

Casewell, Nicholas R., et al. "Complex cocktails: the evolutionary novelty of venoms." *Trends in ecology & evolution*28.4 (2013): 219-229.

Castelin, M., et al. "Macroevolution of venom apparatus innovations in auger snails (Gastropoda; Conoidea; Terebridae)." *Molecular phylogenetics and evolution* 64.1 (2012): 21-44.

Fautin, Daphne Gail. "Structural diversity, systematics, and evolution of cnidae." *Toxicon* 54.8 (2009): 1054-1064.

Frazão, Bárbara, Vitor Vasconcelos, and Agostinho Antunes. "Sea anemone (Cnidaria, Anthozoa, Actiniaria) toxins: an overview." *Marine drugs* 10.8 (2012): 1812-1851.

Fry, Bryan G., et al. "The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms." *Annual review of genomics and human genetics* 10 (2009): 483-511.

Haas, B. J., and A. Papanicolaou. "TransDecoder (find coding regions within transcripts)." (2016).

Haas, Brian J., et al. "De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis." *Nature protocols* 8.8 (2013): 1494-1512.

Holstein, T. W., E. Hobmayer, and Ulrich Technau. "Cnidarians: an evolutionarily conserved model system for regeneration?" *Developmental Dynamics* 226.2 (2003): 257-267.

Ishikawa, Y., K. Onodera, and A. Takeuchi. "Purification and effect of the neurotoxin from the sea anemone Parasicyonis actinostoloides." *Journal of neurochemistry* 33.1 (1979): 69-73.

Jouiaei, Mahdokht, et al. "Evolution of an ancient venom: Recognition of a novel family of cnidarian toxins and the common evolutionary origin of sodium and potassium neurotoxins in sea anemone." *Molecular biology and evolution*32.6 (2015): 1598-1610.

Jungo, Florence, et al. "The UniProtKB/Swiss-Prot Tox-Prot program: a central hub of integrated venom protein data." *Toxicon* 60.4 (2012): 551-557.

Kass-Simon, G. S. A. A., and A. A. Scappaticci, Jr. "The behavioral and developmental physiology of nematocysts." *Canadian Journal of Zoology* 80.10 (2002): 1772-1794.

Kearse, Matthew, et al. "Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data." *Bioinformatics* 28.12 (2012): 1647-1649.

Langmead, Ben, and Steven L. Salzberg. "Fast gapped-read alignment with Bowtie 2." *Nature methods* 9.4 (2012): 357-359.

Lopes-Ferreira, Mônica, et al. "Structural and biological characterization of Nattectin, a new C-type lectin from the venomous fish Thalassophryne nattereri." *Biochimie* 93.6 (2011): 971-980.

López-Maury, Luis, Samuel Marguerat, and Jürg Bähler. "Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation." *Nature Reviews Genetics* 9.8 (2008): 583-593.

Macrander, Jason, Michael Broe, and Marymegan Daly. "Tissue-specific venom composition and differential gene expression in sea anemones." *Genome biology and evolution*8.8 (2016): 2358-2375.

Mariscal, R. N. "Nematocysts. Pp. 129â€” 178 in Coelenterate Biology. Reviews and New Perspectives. L. Muscatine and HM Lenhoff, eds." (1974).

Mathias, A. P., D. M. Ross, and M. Schachter. "The distribution of 5-hydroxytryptamine, tetramethylammonium, homarine, and other substances in sea anemones." *The Journal of physiology* 151.2 (1960): 296-311.

Monastyrnaya, Margarita M., et al. "Biologically active polypeptides from the tropical sea anemone Radianthus macrodactylus." *Toxicon* 40.8 (2002): 1197-1217.

Nagai, Hiroshi, et al. "Isolation and characterization of a novel protein toxin from the Hawaiian box jellyfish (sea wasp) Carybdea alata." *Biochemical and biophysical research communications* 275.2 (2000): 589-594.

Oliveira, Joacir Stolarz, Deyanira Fuentes-Silva, and Glenn F. King. "Development of a rational nomenclature for naming peptide and protein toxins from sea anemones." *Toxicon* 60.4 (2012): 539-550.

Oliveira, Joacir Stolarz, Deyanira Fuentes-Silva, and Glenn F. King. "Development of a rational nomenclature for naming peptide and protein toxins from sea anemones." *Toxicon* 60.4 (2012): 539-550.

Oliveira, Joacir Stolarz, et al. "BcIV, a new paralyzing peptide obtained from the venom of the sea anemone Bunodosoma caissarum. A comparison with the Na+ channel toxin BcIII." *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1764.10 (2006): 1592-1600.

Orts, Diego JB, et al. "BcsTx3 is a founder of a novel sea anemone toxin family of potassium channel blocker." *The FEBS journal* 280.19 (2013): 4839-4852.

Oshiro, Naomasa, et al. "A new membrane-attack complex/perforin (MACPF) domain lethal toxin from the nematocyst venom of the Okinawan sea anemone Actineria villosa." *Toxicon* 43.2 (2004): 225-228.

Reft, Abigail J., and Marymegan Daly. "Morphology, distribution, and evolution of apical structure of nematocysts in Hexacorallia." *Journal of morphology* 273.2 (2012): 121-136.

Roberts, Adam, and Lior Pachter. "Streaming fragment assignment for real-time analysis of sequencing experiments." *Nature methods* 10.1 (2013): 71-73.

Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. "edgeR: A Bioconductor package for differential expression analysis of digital gene expression data." *Bioinformatics* 26.1 (2010): 139-140.

Smith, William Leo, and Ward C. Wheeler. "Venom evolution widespread in fishes: a phylogenetic road map for the bioprospecting of piscine venoms." *Journal of Heredity* 97.3 (2006): 206-217.

Ständker, Ludger, et al. "A new toxin from the sea anemone Condylactis gigantea with effect on sodium channel inactivation." *Toxicon* 48.2 (2006): 211-220.

Sunagawa, Shinichi, et al. "Identification and gene expression analysis of a taxonomically restricted cysteine-rich protein family in reef-building corals." *PLoS One* 4.3 (2009): e4865.

Team, R. Core. "R language definition." *Vienna, Austria: R foundation for statistical computing* (2000).

Yan, Li, et al. "A cnidarian homologue of translationally controlled tumor protein (P23/TCTP)." *Development genes and evolution* 210.10 (2000): 507-511.

List of Tables

List of Figures

List of Supplementary Information

Supplementary File 4. Abundance values in TMM for all transcripts in *Entacmaea quadricolor* for triplicates of tentacles and column

Supplementary File 5. Abundance values in CPM for toxins in *Condylactis gigantea* for triplicates of tentacles and column

Supplementary File 6. Abundance values in CPM for toxins in *Entacmaea quadricolor* for triplicates of tentacles and column

Supplementary File 7. Abundance values in CPM for Swissprot IDs in *Condylactis gigantea* for triplicates of tentacles and column

Supplementary File 8. Abundance values in CPM for Swissprot IDs in *Entacmaea quadricolor* for triplicates of tentacles and column

Supplementary Figure 1. Correlation map for toxins of *Condylactis gigantea* between biological replicates

Supplementary Figure 2. MA plot and Volcano plot for toxins of *Condylactis gigantea* tentacles Vs column

Supplementary Figure 3. Correlation map for toxins of *Entacmaea quadricolor* between biological replicates

Supplementary Figure 4. MA plot and Volcano plot for toxins of *Entacmaea quadricolor* Tentacles Vs Column

Supplementary Figure 5. Correlation map for uniprot IDs of *Condylactis gigantea* between biological replicates

Supplementary Figure 6. Correlation map for Uniprot IDs of *Entacmaea quadricolor* between biological replicates

Appendix I: Supplementary information for Chapter 1

Supplementary Table 1. Animals used in the study and the source of their genome or transcriptome data

Class	Species	Accession number/Source
Anthozoa	*Acanthogorgia aspera*	GETB00000000.1,GEXC00000000.1
Anthozoa	*Acropora cervicornis*	GASU00000000.1
Anthozoa	*Acropora digitifera*	BACK00000000.2
Anthozoa	*Acropora hyacinthus*	GDIF00000000.1
Anthozoa	*Actinia tenebrosa*	GEVE00000000.1
Anthozoa	*Anthopleura elegantissima*	GBYC00000000.1
Anthozoa	*Ctenactis echinata*	GDZV00000000.1
Anthozoa	*Cyphastrea serailia*	GETH00000000.1
Anthozoa	*Diploria strigosa*	CCMS00000000.1
Anthozoa	*Exaiptasia pallida*	LJWW00000000.1
Anthozoa	*Favia lizardensis*	GDZU00000000.1
Anthozoa	*Fungia scutaria*	people.oregonstate.edu/~meyere/data.html
Anthozoa	*Madracis auretenca*	people.oregonstate.edu/~meyere/data.html
Anthozoa	*Monstastaea cavernosa*	people.oregonstate.edu/~meyere/data.html
Anthozoa	*Nematostella vectensis*	ABAV00000000.1
Anthozoa	*Palythoa caribaeorum*	GESO00000000.1
Anthozoa	*Platygyra daedalea*	people.oregonstate.edu/~meyere/data.html
Anthozoa	*Pocillopora damicornis*	GEFF00000000.1
Anthozoa	*Porites astreoides*	GEHP00000000.1
Anthozoa	*Protopalythoa variabilis*	GCVI00000000.1
Anthozoa	*Seriatopora hystrix*	people.oregonstate.edu/~meyere/data.html
Anthozoa	*Stylophora pistillata*	GARY00000000.1
Cubozoa	*Alatina alata*	GEUJ00000000.1
Hydrozoa	*Hydra oligactis*	GBFD00000000.1
Hydrozoa	*Hydra vulgaris*	ACZU00000000.1
Hydrozoa	*Hydractinia polyclina*	GANA00000000.1
Hydrozoa	*Hydractinia symbiolongicarpus*	GCHW00000000.1
Hydrozoa	*Millepora alcicornis*	GFAS00000000.1

Hydrozoa	*Podocoryna carnea*	GCHV00000000.1,GBEH00000000.1
Hydrozoa	*Turritopsis sp.*	IAAF00000000.1
Myxozoa	*Enteromyxum leei*	LDNA00000000.1
Myxozoa	*Kudoa iwatai*	GBGI00000000.1,JRUX00000000.1
Myxozoa	*Myxobolus cerebralis*	GBKL00000000.1
Myxozoa	*Sphaeromyxa zaharoni*	LDMZ00000000.1
Myxozoa	*Thelohanellus kitauei*	JWZT00000000.1
Polypodiozoa	*Polypodium hydriforme*	GBGH00000000.1
Scyphozoa	*Aurelia aurita*	GBRG00000000.1
Porifera	*Amphimedon queenslandica*	GCA_000090795.1
Arthropoda	*Drosophila melanogaster*	GCA_000001215.4

Supplementary Table 2. Housekeeping genes used in the study to construct a species tree

Index number	Uniref ID	Representative Gene	Description
HKG1	UniRef50_O00186	STXB3_HUMAN	Syntaxin-binding protein 3
HKG2	UniRef50_O00567	NOP56_HUMAN	Nucleolar protein 56
HKG3	UniRef50_O14777	NDC80_HUMAN	Kinetochore protein NDC80 homolog
HKG4	UniRef50_O15379	HDAC3_HUMAN	Histone deacetylase 3
HKG5	UniRef50_O15514	RPB4_HUMAN	DNA-directed RNA polymerase II subunit RPB4
HKG6	UniRef50_O43395	PRPF3_HUMAN	U4/U6 small nuclear ribonucleoprotein Prp3
HKG7	UniRef50_O43660	PLRG1_HUMAN	Pleiotropic regulator 1
HKG8	UniRef50_O95602	RPA1_HUMAN	DNA-directed RNA polymerase I subunit RPA1
HKG9	UniRef50_O95639	CPSF4_HUMAN	Cleavage and polyadenylation specificity factor subunit 4
HKG10	UniRef50_P00558	PGK1_HUMAN	Phosphoglycerate kinase 1
HKG11	UniRef50_P01215	GLHA_HUMAN	Glycoprotein hormones alpha chain
HKG12	UniRef50_P04406	G3P_HUMAN	Glyceraldehyde-3-phosphate dehydrogenase
HKG13	UniRef50_P07195	LDHB_HUMAN	L-lactate dehydrogenase B chain
HKG14	UniRef50_P07339	CATD_HUMAN	Cathepsin D
HKG15	UniRef50_P07355	ANXA2_HUMAN	Annexin A2
HKG16	UniRef50_P08238	HS90B_HUMAN	Heat shock protein HSP 90-beta
HKG17	UniRef50_P11413	G6PD_HUMAN	Glucose-6-phosphate 1-dehydrogenase
HKG18	UniRef50_P11926	DCOR_HUMAN	Ornithine decarboxylase
HKG19	UniRef50_P12235	ADT1_HUMAN	ADP/ATP translocase 1
HKG20	UniRef50_P18124	RL7_HUMAN	60S ribosomal protein L7
HKG21	UniRef50_P20226	TBP_HUMAN	TATA-box-binding protein
HKG22	UniRef50_P26373	RL13_HUMAN	60S ribosomal protein L13
HKG23	UniRef50_P26640	SYVC_HUMAN	Valine--tRNA ligase
HKG24	UniRef50_P30291	WEE1_HUMAN	Wee1-like protein kinase
HKG25	UniRef50_P30405	PPIF_HUMAN	Peptidyl-prolyl cis-trans isomerase F, mitochondrial
HKG26	UniRef50_P33993	MCM7_HUMAN	DNA replication licensing factor MCM7
HKG27	UniRef50_P39023	RL3_HUMAN	60S ribosomal protein L3
HKG28	UniRef50_P39656	OST48_HUMAN	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit
HKG29	UniRef50_P40937	RFC5_HUMAN	Replication factor C subunit 5

HKG30	UniRef50_P41091	IF2G_HUMAN	Eukaryotic translation initiation factor 2 subunit 3
HKG31	UniRef50_P42766	RL35_HUMAN	60S ribosomal protein L35
HKG32	UniRef50_P46779	RL28_HUMAN	60S ribosomal protein L28
HKG33	UniRef50_P46782	RS5_HUMAN	40S ribosomal protein S5
HKG34	UniRef50_P49368	TCPG_HUMAN	T-complex protein 1 subunit gamma
HKG35	UniRef50_P49736	MCM2_HUMAN	DNA replication licensing factor MCM2
HKG36	UniRef50_P53621	COPA_HUMAN	Coatomer subunit alpha
HKG37	UniRef50_P62269	RS18_HUMAN	40S ribosomal protein S18
HKG38	UniRef50_P62280	RS11_HUMAN	40S ribosomal protein S11
HKG39	UniRef50_P62333	PRS10_HUMAN	26S protease regulatory subunit 10B
HKG40	UniRef50_P62841	RS15_HUMAN	40S ribosomal protein S15
HKG41	UniRef50_P62917	RL8_HUMAN	60S ribosomal protein L8
HKG42	UniRef50_P63208	SKP1_HUMAN	S-phase kinase-associated protein 1
HKG43	UniRef50_P67809	YBOX1_HUMAN	Nuclease-sensitive element-binding protein 1
HKG44	UniRef50_P68133	ACTS_HUMAN	Actin, alpha skeletal muscle
HKG45	UniRef50_P78406	RAE1L_HUMAN	mRNA export factor
HKG46	UniRef50_P83731	RL24_HUMAN	60S ribosomal protein L24
HKG47	UniRef50_Q01081	U2AF1_HUMAN	Splicing factor U2AF 35 kDa subunit
HKG48	UniRef50_Q10570	CPSF1_HUMAN	Cleavage and polyadenylation specificity factor subunit 1
HKG49	UniRef50_Q12996	CSTF3_HUMAN	Cleavage stimulation factor subunit 3
HKG50	UniRef50_Q13547	HDAC1_HUMAN	Histone deacetylase 1
HKG51	UniRef50_Q13563	PKD2_HUMAN	Polycystin-2
HKG52	UniRef50_Q13952	NFYC_HUMAN	Nuclear transcription factor Y subunit gamma
HKG53	UniRef50_Q15038	DAZP2_HUMAN	DAZ-associated protein 2
HKG54	UniRef50_Q15046	SYK_HUMAN	Lysine--tRNA ligase
HKG55	UniRef50_Q15233	NONO_HUMAN	Non-POU domain-containing octamer-binding protein
HKG56	UniRef50_Q53GS7	GLE1_HUMAN	Nucleoporin GLE1
HKG57	UniRef50_Q8NHQ9	DDX55_HUMAN	ATP-dependent RNA helicase DDX55
HKG58	UniRef50_Q92890	UFD1_HUMAN	Ubiquitin fusion degradation protein 1 homolog
HKG59	UniRef50_Q969P0	IGSF8_HUMAN	Immunoglobulin superfamily member 8
HKG60	UniRef50_Q96GQ7	DDX27_HUMAN	Probable ATP-dependent RNA helicase DDX27
HKG61	UniRef50_Q99832	TCPH_HUMAN	T-complex protein 1 subunit eta
HKG62	UniRef50_Q9BZL6	KPCD2_HUMAN	Serine/threonine-protein kinase D2

HKG63	UniRef50_Q9H8V3	ECT2_HUMAN	Protein ECT2
HKG64	UniRef50_Q9NRG1	PRDC1_HUMAN	Phosphoribosyltransferase domain-containing protein 1
HKG65	UniRef50_Q9NTN3	S35D1_HUMAN	UDP-glucuronic acid/UDP-N-acetylgalactosamine transporter
HKG66	UniRef50_Q9P253	VPS18_HUMAN	Vacuolar protein sorting-associated protein 18 homolog
HKG67	UniRef50_Q9UBR2	CATZ_HUMAN	Cathepsin Z
HKG68	UniRef50_Q9UHB9	SRP68_HUMAN	Signal recognition particle subunit SRP68
HKG69	UniRef50_Q9UMR2	DD19B_HUMAN	ATP-dependent RNA helicase DDX19B
HKG70	UniRef50_Q9UNP9	PPIE_HUMAN	Peptidyl-prolyl cis-trans isomerase E

Supplementary Table 3. Toxin genes used in the study

Toxin index	**UniProt Entry**	**Gene name**	**Toxin Family**
toxin1	P0C1H0	ACTP1_ACTEQ	Actinoporin family, Sea anemone subfamily
toxin2	Q5R231	ACTP1_ACTVL	Actinoporin family, Sea anemone subfamily
toxin3	C5NSL2	ACTP1_ANTAS	Actinoporin family, Sea anemone subfamily
toxin4	P58689	ACTP1_HETMG	Actinoporin family, Sea anemone subfamily
toxin5	P0DL55	ACTP1_PHYSE	Actinoporin family, Sea anemone subfamily
toxin6	Q86FQ0	ACTP1_SAGRO	Actinoporin family, Sea anemone subfamily
toxin7	P81662	ACTP1_STIHL	Actinoporin family, Sea anemone subfamily
toxin8	C9EIC7	ACTP1_URTCR	Actinoporin family, Sea anemone subfamily
toxin9	P61914	ACTP2_ACTEQ	Actinoporin family, Sea anemone subfamily
toxin10	D2YZQ3	ACTP2_ACTVL	Actinoporin family, Sea anemone subfamily
toxin11	P58690	ACTP2_HETMG	Actinoporin family, Sea anemone subfamily
toxin12	P0DL56	ACTP2_PHYSE	Actinoporin family, Sea anemone subfamily
toxin13	P07845	ACTP2_STIHL	Actinoporin family, Sea anemone subfamily
toxin14	P0C1H2	ACTP3_ACTEQ	Actinoporin family, Sea anemone subfamily
toxin15	Q9U6X1	ACTP3_HETMG	Actinoporin family, Sea anemone subfamily
toxin16	Q9Y1U9	ACTP4_ACTEQ	Actinoporin family, Sea anemone subfamily
toxin17	P0DMX2	ACTP4_HETMG	Actinoporin family, Sea anemone subfamily
toxin18	Q93109	ACTP5_ACTEQ	Actinoporin family, Sea anemone subfamily
toxin19	P30833	ACTPA_ACTTE	Actinoporin family, Sea anemone subfamily
toxin20	P58691	ACTPA_HETCR	Actinoporin family, Sea anemone subfamily
toxin21	Q5I4B8	ACTPA_OULOR	Actinoporin family, Sea anemone subfamily
toxin22	P30834	ACTPB_ACTTE	Actinoporin family, Sea anemone subfamily
toxin23	B9W5G6	ACTPC_ACTFR	Actinoporin family, Sea anemone subfamily
toxin24	P61915	ACTPC_ACTTE	Actinoporin family, Sea anemone subfamily
toxin25	Q5I2B1	ACTPG_OULOR	Actinoporin family, Sea anemone subfamily
toxin26	P39088	ACTPH_HETMG	Actinoporin family, Sea anemone subfamily
toxin27	P0C1H1	ACTPP_ACTEQ	Actinoporin family, Sea anemone subfamily
toxin28	P0DMX3	ACTP_ENTQU	Actinoporin family, Sea anemone subfamily
toxin29	P0DMX4	ACTP_STIME	Actinoporin family, Sea anemone subfamily
toxin30	P0C1F8	ACTS2_HETCR	Actinoporin family, Sea anemone subfamily
toxin31	C0H693	SCR1A_MONCP	Cnidaria small cysteine-rich protein (SCRiP) family
toxin32	C0H694	SCR1B_MONCP	Cnidaria small cysteine-rich protein (SCRiP) family
toxin33	C0H690	SCR1_ACRMI	Cnidaria small cysteine-rich protein (SCRiP) family
toxin34	C1KIY9	SCR1_ORBFA	Cnidaria small cysteine-rich protein (SCRiP) family
toxin35	C0H691	SCR2_ACRMI	Cnidaria small cysteine-rich protein (SCRiP) family
toxin36	C1KIZ0	SCR2_ORBFA	Cnidaria small cysteine-rich protein (SCRiP) family

toxin37	C0H692	SCR3_ACRMI	Cnidaria small cysteine-rich protein (SCRiP) family
toxin38	C1KIZ3	SCR4_ORBFA	Cnidaria small cysteine-rich protein (SCRiP) family
toxin39	C1KIZ4	SCR5_ORBFA	Cnidaria small cysteine-rich protein (SCRiP) family
toxin40	C1KIZ5	SCR6_ORBFA	Cnidaria small cysteine-rich protein (SCRiP) family
toxin41	B2ZG38	SCR8_ORBFA	Cnidaria small cysteine-rich protein (SCRiP) family
toxin42	P0DL61	SCR_ANEVI	Cnidaria small cysteine-rich protein (SCRiP) family
toxin43	P0DL60	SCR_METSE	Cnidaria small cysteine-rich protein (SCRiP) family
toxin44	Q331K1	CYT_CYACP	Cystatin family
toxin45	A8QZJ5	MCTX1_MILDI	Dermatopontin family
toxin46	Q9GV16	EGCSE_CYANO	Glycosyl hydrolase 5 (cellulase A) family
toxin47	Q9GV72	CTX1_CARRA	Jellyfish toxin family
toxin48	A7L035	CTX1_CHIFL	Jellyfish toxin family
toxin49	A7L036	CTX2_CHIFL	Jellyfish toxin family
toxin50	Q9GNN8	CTXA_CARAL	Jellyfish toxin family
toxin51	P58762	CTXA_CHIQU	Jellyfish toxin family
toxin52	K7Z9Q9	VMP_NEMVE	Peptidase M12A family
toxin53	Q8WS88	PA2_ADAPA	Phospholipase A2 family
toxin54	P86780	PA2_BUNCI	Phospholipase A2 family
toxin55	D2X8K2	PA2_CONGI	Phospholipase A2 family
toxin56	A7LCJ2	PA2_URTCR	Phospholipase A2 family
toxin57	P43318	PA2_RHONO	Phospholipase A2 family, Group III subfamily
toxin58	P0DMZ3	TX8A_ANEVI	Sea anemone 8 toxin family
toxin59	P0DMZ4	TX8B_ANEVI	Sea anemone 8 toxin family
toxin60	P0DMZ5	TX8C_ANEVI	Sea anemone 8 toxin family
toxin61	P0DMZ6	TX8D_ANEVI	Sea anemone 8 toxin family
toxin62	P0DMZ7	TX8E_ANEVI	Sea anemone 8 toxin family
toxin63	P01535	STX3_ANESU	Sea anemone short toxin (type III) family
toxin64	P09949	STX3_ENTQU	Sea anemone short toxin (type III) family
toxin65	C3TS08	STX71_ANEVI	Sea anemone short toxin (type III) family
toxin66	C3TS04	STX72_ANEVI	Sea anemone short toxin (type III) family
toxin67	C3TS10	STX73_ANEVI	Sea anemone short toxin (type III) family
toxin68	C3TS06	STX74_ANEVI	Sea anemone short toxin (type III) family
toxin69	C3TS07	STXA_ANEVI	Sea anemone short toxin (type III) family
toxin70	P0DMZ2	STX_DOFAR	Sea anemone short toxin (type III) family
toxin71	E3P6S4	NAU1A_AIPDI	Sea anemone sodium channel inhibitory toxin family
toxin72	P14531	TXCL1_CALPA	Sea anemone sodium channel inhibitory toxin family
toxin73	P49127	TXCL2_CALPA	Sea anemone sodium channel inhibitory toxin family
toxin74	P0DMZ1	NA111_ANEVI	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin75	P0C5F7	NA116_ANTS7	Sea anemone sodium channel inhibitory toxin family, Type I subfamily

toxin76	Q9NJQ2	NA11_ACTEQ	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin77	P69943	NA11_ANEER	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin78	P01533	NA11_ANESU	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin79	P0C1F0	NA11_ANTEL	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin80	P10453	NA11_ANTFU	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin81	P0C5F8	NA11_ANTXA	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin82	P0CH42	NA11_CONPS	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin83	P0C5G5	NA11_HETCR	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin84	B1NWU2	NA121_ACTEQ	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin85	B1NWU3	NA122_ACTEQ	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin86	P0DL50	NA122_ANEVI	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin87	P0C1F3	NA122_ANTEL	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin88	B1NWU4	NA123_ACTEQ	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin89	P0DL51	NA123_ANEVI	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin90	P0DL52	NA124_ANEVI	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin91	P0DL53	NA125_ANEVI	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin92	P0DL54	NA126_ANEVI	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin93	P0C5F4	NA12A_ANTS7	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin94	B1NWT7	NA12D_ANEVI	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin95	P01528	NA12_ANESU	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin96	P0DL49	NA12_ANEVI	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin97	P10454	NA12_ANTFU	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin98	P0C5F9	NA12_ANTXA	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin99	P0C1F4	NA12_BUNGR	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin100	B1NWU5	NA131_ACTEQ	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin101	P0C1F1	NA13_ANTEL	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin102	P69928	NA13_ANTMC	Sea anemone sodium channel inhibitory toxin family, Type I subfamily

toxin103	P0C5G0	NA13_ANTXA	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin104	Q7M425	NA13_BUNCI	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin105	P0C1F5	NA13_BUNGR	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin106	B1NWU6	NA141_ACTEQ	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin107	P0C5G1	NA14_ANTXA	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin108	P01529	NA15_ANESU	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin109	P0C5G2	NA15_ANTXA	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin110	B1NWT3	NA16_ANEVI	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin111	P0C5G3	NA16_ANTXA	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin112	P0C5F5	NA17A_ANTS7	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin113	P0C5F6	NA18A_ANTS7	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin114	B1NWT4	NA18_ANEVI	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin115	B1NWT5	NA19_ANEVI	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin116	P01530	NA1A_ANTXA	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin117	P01531	NA1B_ANTXA	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin118	P0C7P9	NA1B_BUNCN	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin119	P82803	NA1C1_BUNCN	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin120	P0C7Q0	NA1C3_BUNCN	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin121	P01532	NA1C_ANTEL	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin122	P86459	NA1D_BUNCN	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin123	Q76CA3	NA1G2_STIGI	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin124	P0DMX1	NA1P2_CONPS	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin125	P0C280	NA1_CONGI	Sea anemone sodium channel inhibitory toxin family, Type I subfamily
toxin126	B1NWS1	NA217_NEMVE	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin127	D2KX90	NA21_CRYAD	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin128	P30831	NA21_HETCR	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin129	B1NWS4	NA21_NEMVE	Sea anemone sodium channel inhibitory toxin family, Type II subfamily

toxin130	P19651	NA21_STIHL	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin131	D2KX92	NA21_THAAS	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin132	P30783	NA22_HETCR	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin133	P01534	NA22_HETMG	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin134	P0CH90	NA237_NEMVE	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin135	P0CH46	NA239_NEMVE	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin136	P30832	NA23_HETCR	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin137	P08380	NA23_HETMG	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin138	B1NWS8	NA240_NEMVE	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin139	A7SCE5	NA241_NEMVE	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin140	B1NWR7	NA245_NEMVE	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin141	P30784	NA24_HETCR	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin142	B1B5I9	NA24_STIHA	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin143	P30785	NA25_HETCR	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin144	P0CH45	NA271_NEMVE	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin145	B1NWR6	NA281_NEMVE	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin146	Q76CA0	NA2G3_STIGI	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin147	P0C5G6	NA2H_HALCG	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin148	D2KX91	NA2X_HETHE	Sea anemone sodium channel inhibitory toxin family, Type II subfamily
toxin149	P86466	TX9A0_BUNCN	Sea anemone structural class 9a family
toxin150	P0DMZ8	TX9A1_ANEVI	Sea anemone structural class 9a family
toxin151	P0DMZ9	TX9A2_ANEVI	Sea anemone structural class 9a family
toxin152	P86465	TX9A7_BUNCN	Sea anemone structural class 9a family
toxin153	P86467	TX9AN_BUNCN	Sea anemone structural class 9a family
toxin154	P69929	TX9A_ANTMC	Sea anemone structural class 9a family
toxin155	P0C7W7	TX9A_STIHA	Sea anemone structural class 9a family
toxin156	R4ZCU1	TX9A_URTGR	Sea anemone structural class 9a family
toxin157	Q0EAE5	K1A_ANEER	Sea anemone type 1 potassium channel toxin family, Type 1a subfamily
toxin158	E2S064	K1A_CRYAD	Sea anemone type 1 potassium channel toxin family, Type 1a subfamily
toxin159	E2S065	K1A_HETHE	Sea anemone type 1 potassium channel toxin family, Type 1a subfamily

toxin160	O16846	K1A_HETMG	Sea anemone type 1 potassium channel toxin family, Type 1a subfamily
toxin161	E2S061	K1A_STIGI	Sea anemone type 1 potassium channel toxin family, Type 1a subfamily
toxin162	E2S062	K1A_STIHA	Sea anemone type 1 potassium channel toxin family, Type 1a subfamily
toxin163	P29187	K1A_STIHL	Sea anemone type 1 potassium channel toxin family, Type 1a subfamily
toxin164	E2S063	K1A_STIME	Sea anemone type 1 potassium channel toxin family, Type 1a subfamily
toxin165	E2S066	K1A_THAAS	Sea anemone type 1 potassium channel toxin family, Type 1a subfamily
toxin166	C0HJC2	K1B1_BUNCI	Sea anemone type 1 potassium channel toxin family, Type 1b subfamily
toxin167	C0HJC3	K1B2_BUNCI	Sea anemone type 1 potassium channel toxin family, Type 1b subfamily
toxin168	P0DN00	K1B9A_ANEVI	Sea anemone type 1 potassium channel toxin family, Type 1b subfamily
toxin169	P0DN01	K1B9B_ANEVI	Sea anemone type 1 potassium channel toxin family, Type 1b subfamily
toxin170	P0DN02	K1B9C_ANEVI	Sea anemone type 1 potassium channel toxin family, Type 1b subfamily
toxin171	P0DN03	K1B9D_ANEVI	Sea anemone type 1 potassium channel toxin family, Type 1b subfamily
toxin172	P0DN04	K1BAA_ANEVI	Sea anemone type 1 potassium channel toxin family, Type 1b subfamily
toxin173	P0DN05	K1BBA_ANEVI	Sea anemone type 1 potassium channel toxin family, Type 1b subfamily
toxin174	P81897	K1B_ACTEQ	Sea anemone type 1 potassium channel toxin family, Type 1b subfamily
toxin175	Q9TWG1	K1B_ANESU	Sea anemone type 1 potassium channel toxin family, Type 1b subfamily
toxin176	P29186	K1B_BUNGR	Sea anemone type 1 potassium channel toxin family, Type 1b subfamily
toxin177	P11495	K1B_METSE	Sea anemone type 1 potassium channel toxin family, Type 1b subfamily
toxin178	P86461	BDS16_BUNCN	Sea anemone type 3 (BDS) potassium channel toxin family
toxin179	P11494	BDS1_ANESU	Sea anemone type 3 (BDS) potassium channel toxin family
toxin180	P0DMX6	BDS1_ANEVI	Sea anemone type 3 (BDS) potassium channel toxin family
toxin181	P61541	BDS1_ANTEL	Sea anemone type 3 (BDS) potassium channel toxin family
toxin182	P86464	BDS21_BUNCN	Sea anemone type 3 (BDS) potassium channel toxin family
toxin183	P69930	BDS2A_ANTMC	Sea anemone type 3 (BDS) potassium channel toxin family
toxin184	P84919	BDS2A_BUNCI	Sea anemone type 3 (BDS) potassium channel toxin family
toxin185	P0DMX5	BDS2C_ANTEL	Sea anemone type 3 (BDS) potassium channel toxin family
toxin186	P59084	BDS2_ANESU	Sea anemone type 3 (BDS) potassium channel toxin family

toxin187	P61542	BDS2_ANTEL	Sea anemone type 3 (BDS) potassium channel toxin family
toxin188	G0W2H8	BDS3A_BUNGR	Sea anemone type 3 (BDS) potassium channel toxin family
toxin189	G0W2H9	BDS3B_BUNGR	Sea anemone type 3 (BDS) potassium channel toxin family
toxin190	G0W2I0	BDS3C_BUNGR	Sea anemone type 3 (BDS) potassium channel toxin family
toxin191	G0W2I1	BDS3D_BUNGR	Sea anemone type 3 (BDS) potassium channel toxin family
toxin192	P0DMX7	BDS3_ANEVI	Sea anemone type 3 (BDS) potassium channel toxin family
toxin193	B3EWF9	BDS3_ANTEL	Sea anemone type 3 (BDS) potassium channel toxin family
toxin194	G0W2H7	BDS3_BUNGR	Sea anemone type 3 (BDS) potassium channel toxin family
toxin195	P0DMX8	BDS4_ANEVI	Sea anemone type 3 (BDS) potassium channel toxin family
toxin196	P86463	BDS52_BUNCN	Sea anemone type 3 (BDS) potassium channel toxin family
toxin197	P0DMX9	BDS5_ANEVI	Sea anemone type 3 (BDS) potassium channel toxin family
toxin198	P0DMY0	BDS6_ANEVI	Sea anemone type 3 (BDS) potassium channel toxin family
toxin199	P86462	BDS78_BUNCN	Sea anemone type 3 (BDS) potassium channel toxin family
toxin200	P0DMY1	BDS7_ANEVI	Sea anemone type 3 (BDS) potassium channel toxin family
toxin201	P0DMY2	BDS8_ANEVI	Sea anemone type 3 (BDS) potassium channel toxin family
toxin202	P0DMY3	BDS9_ANEVI	Sea anemone type 3 (BDS) potassium channel toxin family
toxin203	P0DMY4	BDSA_ANEVI	Sea anemone type 3 (BDS) potassium channel toxin family
toxin204	P0DMY5	BDSB_ANEVI	Sea anemone type 3 (BDS) potassium channel toxin family
toxin205	P0DMY6	BDSC_ANEVI	Sea anemone type 3 (BDS) potassium channel toxin family
toxin206	P0DMY7	BDSD_ANEVI	Sea anemone type 3 (BDS) potassium channel toxin family
toxin207	P0DMY8	BDSE_ANEVI	Sea anemone type 3 (BDS) potassium channel toxin family
toxin208	P86470	BDSV_BUNCI	Sea anemone type 3 (BDS) potassium channel toxin family
toxin209	P0DMD7	KV51_METSE	Sea anemone type 5 potassium channel toxin family
toxin210	A7RMN1	KV51_NEMVE	Sea anemone type 5 potassium channel toxin family
toxin211	C0HJC4	KV53_BUNCI	Sea anemone type 5 potassium channel toxin family
toxin212	C0HJU6	VKT5_HETCR	Venom Kunitz-type family
toxin213	C0HJU7	VKT6_HETCR	Venom Kunitz-type family
toxin214	P0DN06	VKT1A_ANEVI	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin215	Q9TWG0	VKT1_ANESU	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily

toxin216	P81547	VKT1_ANTAF	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin217	P86862	VKT1_ANTEL	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin218	B2G331	VKT1_HETCR	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin219	P31713	VKT1_STIHL	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin220	Q9TWF9	VKT2_ANESU	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin221	P81548	VKT2_ANTAF	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin222	C0HJF4	VKT2_HETCR	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin223	P81129	VKT2_STIHL	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin224	P0DMW8	VKT33_ACTEQ	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin225	P0DMW9	VKT34_ACTEQ	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin226	P0DMW6	VKT3A_ACTEQ	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin227	P0DMW7	VKT3B_ACTEQ	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin228	P0DMJ2	VKT3C_ACTEQ	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin229	Q9TWF8	VKT3_ANESU	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin230	P0DMX0	VKT3_ANTAF	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin231	C0HJF3	VKT3_HETCR	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin232	B1B5I8	VKT3_STIHA	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin233	P0DN07	VKT4_ANEVI	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin234	P16344	VKT4_HETCR	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin235	P10280	VKT52_ANESU	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin236	P0DN08	VKT53_ANEVI	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin237	P0DN09	VKT5_ANEVI	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin238	P0DN10	VKT6_ANEVI	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin239	P0DN11	VKT7_ANEVI	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin240	P0DN12	VKT8_ANEVI	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin241	P0DN13	VKT9_ANEVI	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin242	P0DN14	VKTA_ANEVI	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily

toxin243	P0DN15	VKTB_ANEVI	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin244	P0DN16	VKTC_ANEVI	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin245	P0DN17	VKTD_ANEVI	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin246	P0DN18	VKTE_ANEVI	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin247	P0DN19	VKTF_ANEVI	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin248	P0DN20	VKTG_ANEVI	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin249	P0DMJ3	VKTI1_ANTFU	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily
toxin250	P0DMJ4	VKTI3_ANTFU	Venom Kunitz-type family, Sea anemone type 2 potassium channel toxin subfamily

Supplementary Table 4. Ion channel genes used in the study

Index number of ion channels used	**Description of the ion channels used in the study, with Uniref50, description, length of the protein, taxa of the representative protein, and the name of the representative protein**
Calcium1	UniRef50_O00305 Voltage-dependent L-type calcium channel subunit beta-4 n=127 Tax=Euteleostomi RepID=CACB4_HUMAN
Calcium2	UniRef50_O00555 Voltage-dependent P/Q-type calcium channel subunit alpha-1A n=235 Tax=Sarcopterygii RepID=CAC1A_HUMAN
Calcium3	UniRef50_O43497 Voltage-dependent T-type calcium channel subunit alpha-1G n=764 Tax=Euteleostomi RepID=CAC1G_HUMAN
Calcium4	UniRef50_O60359 Voltage-dependent calcium channel gamma-3 subunit n=102 Tax=Euteleostomi RepID=CCG3_HUMAN
Calcium5	UniRef50_O88602 Voltage-dependent calcium channel gamma-2 subunit n=94 Tax=Gnathostomata RepID=CCG2_MOUSE
Calcium6	UniRef50_O95180 Voltage-dependent T-type calcium channel subunit alpha-1H n=257 Tax=Euteleostomi RepID=CAC1H_HUMAN
Calcium7	UniRef50_P27732 Voltage-dependent L-type calcium channel subunit alpha-1D n=1115 Tax=Gnathostomata RepID=CAC1D_RAT
Calcium8	UniRef50_P54289 Voltage-dependent calcium channel subunit alpha-2/delta-1 n=523 Tax=Vertebrata RepID=CA2D1_HUMAN
Calcium9	UniRef50_P54289-3 Isoform 3 of Voltage-dependent calcium channel subunit alpha-2/delta-1 n=385 Tax=Euteleostomi RepID=P54289-3
Calcium10	UniRef50_P62955 Voltage-dependent calcium channel gamma-7 subunit n=113 Tax=Euteleostomi RepID=CCG7_HUMAN
Calcium11	UniRef50_P97445 Voltage-dependent P/Q-type calcium channel subunit alpha-1A n=426 Tax=Euteleostomi RepID=CAC1A_MOUSE
Calcium12	UniRef50_Q02294 Voltage-dependent N-type calcium channel subunit alpha-1B n=356 Tax=Gnathostomata RepID=CAC1B_RAT
Calcium13	UniRef50_Q06432 Voltage-dependent calcium channel gamma-1 subunit n=257 Tax=Gnathostomata RepID=CCG1_HUMAN
Calcium14	UniRef50_Q08289 Voltage-dependent L-type calcium channel subunit beta-2 n=974 Tax=Bilateria RepID=CACB2_HUMAN
Calcium15	UniRef50_Q08289-8 Isoform 2h of Voltage-dependent L-type calcium channel subunit beta-2 n=591 Tax=Bilateria RepID=Q08289-8
Calcium16	UniRef50_Q13698 Voltage-dependent L-type calcium channel subunit alpha-1S n=228 Tax=Bilateria RepID=CAC1S_HUMAN
Calcium17	UniRef50_Q13936 Voltage-dependent L-type calcium channel subunit alpha-1C n=93 Tax=Euteleostomi RepID=CAC1C_HUMAN
Calcium18	UniRef50_Q66L44 Voltage-dependent calcium channel beta subunit-associated regulatory protein n=37 Tax=Eutheria RepID=CBARP_MOUSE
Calcium19	UniRef50_Q7Z3S7 Voltage-dependent calcium channel subunit alpha-2/delta-4 n=631 Tax=Gnathostomata RepID=CA2D4_HUMAN
Calcium20	UniRef50_Q8NHX9 Two pore calcium channel protein 2 n=147 Tax=Amniota RepID=TPC2_HUMAN
Calcium21	UniRef50_Q8R3Z5 Voltage-dependent L-type calcium channel subunit beta-1 n=153 Tax=Euteleostomi RepID=CACB1_MOUSE
Calcium22	UniRef50_Q8VHW8 Voltage-dependent calcium channel gamma-5 subunit n=256 Tax=Gnathostomata RepID=CCG5_RAT
Calcium23	UniRef50_Q8WXS4 Voltage-dependent calcium channel gamma-like subunit n=125 Tax=Amniota RepID=CCGL_HUMAN

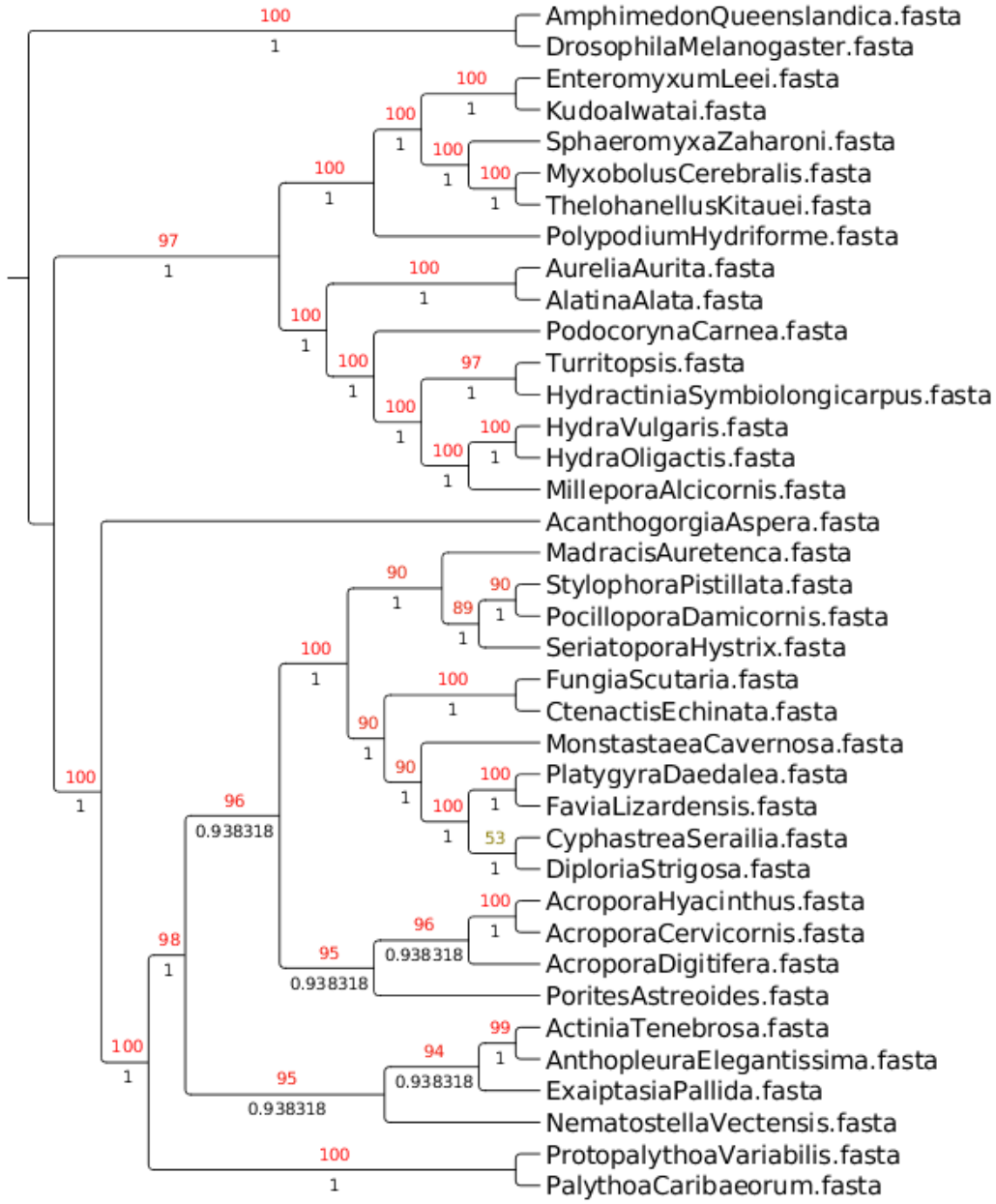
Calcium24	UniRef50_Q8WXS5 Voltage-dependent calcium channel gamma-8 subunit n=69 Tax=Euteleostomi RepID=CCG8_HUMAN
Calcium25	UniRef50_Q9BXT2 Voltage-dependent calcium channel gamma-6 subunit n=35 Tax=Boreoeutheria RepID=CCG6_HUMAN
Calcium26	UniRef50_Q9P0X4 Voltage-dependent T-type calcium channel subunit alpha-1I n=22 Tax=Eutheria RepID=CAC1I_HUMAN
Calcium27	UniRef50_Q9UBN1 Voltage-dependent calcium channel gamma-4 subunit n=219 Tax=Gnathostomata RepID=CCG4_HUMAN
Calcium28	UniRef50_Q9ULQ1 Two pore calcium channel protein 1 n=425 Tax=Gnathostomata RepID=TPC1_HUMAN
Potassium1	UniRef50_O35174 Potassium voltage-gated channel subfamily S member 2 n=149 Tax=Amniota RepID=KCNS2_MOUSE
Potassium2	UniRef50_O43525 Potassium voltage-gated channel subfamily KQT member 3 n=250 Tax=Euteleostomi RepID=KCNQ3_HUMAN
Potassium3	UniRef50_O43526 Potassium voltage-gated channel subfamily KQT member 2 n=286 Tax=Euteleostomi RepID=KCNQ2_HUMAN
Potassium4	UniRef50_O95259 Potassium voltage-gated channel subfamily H member 1 n=173 Tax=Vertebrata RepID=KCNH1_HUMAN
Potassium5	UniRef50_P15382 Potassium voltage-gated channel subfamily E member 1 n=44 Tax=Mammalia RepID=KCNE1_HUMAN
Potassium6	UniRef50_P15384 Potassium voltage-gated channel subfamily A member 3 n=142 Tax=Euteleostomi RepID=KCNA3_RAT
Potassium7	UniRef50_P15387 Potassium voltage-gated channel subfamily B member 1 n=340 Tax=Gnathostomata RepID=KCNB1_RAT
Potassium8	UniRef50_P17658 Potassium voltage-gated channel subfamily A member 6 n=119 Tax=Euteleostomi RepID=KCNA6_HUMAN
Potassium9	UniRef50_P22459 Potassium voltage-gated channel subfamily A member 4 n=185 Tax=Bilateria RepID=KCNA4_HUMAN
Potassium10	UniRef50_P22460 Potassium voltage-gated channel subfamily A member 5 n=185 Tax=Deuterostomia RepID=KCNA5_HUMAN
Potassium11	UniRef50_P22462 Potassium voltage-gated channel subfamily C member 2 n=463 Tax=Vertebrata RepID=KCNC2_RAT
Potassium12	UniRef50_P25122 Potassium voltage-gated channel subfamily C member 1 n=467 Tax=Euteleostomi RepID=KCNC1_RAT
Potassium13	UniRef50_P48547 Potassium voltage-gated channel subfamily C member 1 n=96 Tax=Gnathostomata RepID=KCNC1_HUMAN
Potassium14	UniRef50_P51787 Potassium voltage-gated channel subfamily KQT member 1 n=111 Tax=Bilateria RepID=KCNQ1_HUMAN
Potassium15	UniRef50_P56696 Potassium voltage-gated channel subfamily KQT member 4 n=113 Tax=Euteleostomi RepID=KCNQ4_HUMAN
Potassium16	UniRef50_P63141 Potassium voltage-gated channel subfamily A member 2 n=808 Tax=Bilateria RepID=KCNA2_MOUSE
Potassium17	UniRef50_Q09470 Potassium voltage-gated channel subfamily A member 1 n=267 Tax=Bilateria RepID=KCNA1_HUMAN
Potassium18	UniRef50_Q12809 Potassium voltage-gated channel subfamily H member 2 n=710 Tax=Gnathostomata RepID=KCNH2_HUMAN
Potassium19	UniRef50_Q13303 Voltage-gated potassium channel subunit beta-2 n=493 Tax=Bilateria RepID=KCAB2_HUMAN
Potassium20	UniRef50_Q14003 Potassium voltage-gated channel subfamily C member 3 n=145 Tax=Euteleostomi RepID=KCNC3_HUMAN
Potassium21	UniRef50_Q14722 Voltage-gated potassium channel subunit beta-1 n=84 Tax=Euteleostomi RepID=KCAB1_HUMAN
Potassium22	UniRef50_Q14722-3 Isoform KvB1.2 of Voltage-gated potassium channel subunit beta-1 n=543 Tax=Gnathostomata RepID=Q14722-3

Potassium23	UniRef50_Q63881 Potassium voltage-gated channel subfamily D member 2 n=168 Tax=Bilateria RepID=KCND2_RAT
Potassium24	UniRef50_Q6PIU1 Potassium voltage-gated channel subfamily V member 1 n=209 Tax=Gnathostomata RepID=KCNV1_HUMAN
Potassium25	UniRef50_Q8NCM2 Potassium voltage-gated channel subfamily H member 5 n=252 Tax=Bilateria RepID=KCNH5_HUMAN
Potassium26	UniRef50_Q8TAE7 Potassium voltage-gated channel subfamily G member 3 n=120 Tax=Amniota RepID=KCNG3_HUMAN
Potassium27	UniRef50_Q8TDN2 Potassium voltage-gated channel subfamily V member 2 n=209 Tax=Gnathostomata RepID=KCNV2_HUMAN
Potassium28	UniRef50_Q8WWG9 Potassium voltage-gated channel subfamily E member 4 n=102 Tax=Amniota RepID=KCNE4_HUMAN
Potassium29	UniRef50_Q96KK3 Potassium voltage-gated channel subfamily S member 1 n=183 Tax=Gnathostomata RepID=KCNS1_HUMAN
Potassium30	UniRef50_Q96L42 Potassium voltage-gated channel subfamily H member 8 n=339 Tax=Gnathostomata RepID=KCNH8_HUMAN
Potassium31	UniRef50_Q9BQ31 Potassium voltage-gated channel subfamily S member 3 n=172 Tax=Vertebrata RepID=KCNS3_HUMAN
Potassium32	UniRef50_Q9H252 Potassium voltage-gated channel subfamily H member 6 n=107 Tax=Euteleostomi RepID=KCNH6_HUMAN
Potassium33	UniRef50_Q9H3M0 Potassium voltage-gated channel subfamily F member 1 n=109 Tax=Euteleostomi RepID=KCNF1_HUMAN
Potassium34	UniRef50_Q9NR82 Potassium voltage-gated channel subfamily KQT member 5 n=390 Tax=Euteleostomi RepID=KCNQ5_HUMAN
Potassium35	UniRef50_Q9UIX4 Potassium voltage-gated channel subfamily G member 1 n=404 RepID=KCNG1_HUMAN
Potassium36	UniRef50_Q9UJ90 Potassium voltage-gated channel subfamily E regulatory beta subunit 5 n=70 Tax=Eutheria RepID=KCNE5_HUMAN
Potassium37	UniRef50_Q9UK17 Potassium voltage-gated channel subfamily D member 3 n=524 Tax=Bilateria RepID=KCND3_HUMAN
Potassium38	UniRef50_Q9ULD8 Potassium voltage-gated channel subfamily H member 3 n=149 Tax=Eutheria RepID=KCNH3_HUMAN
Potassium39	UniRef50_Q9Y2W7 Calsenilin n=606 Tax=Gnathostomata RepID=CSEN_HUMAN
Potassium40	UniRef50_Q9Y6H6 Potassium voltage-gated channel subfamily E member 3 n=48 Tax=Amniota RepID=KCNE3_HUMAN
Potassium41	UniRef50_Q9Y6J6 Potassium voltage-gated channel subfamily E member 2 n=141 Tax=Euteleostomi RepID=KCNE2_HUMAN
Sodium1	UniRef50_P35498 Sodium channel protein type 1 subunit alpha n=2204 Tax=Chordata RepID=SCN1A_HUMAN
Sodium2	UniRef50_P35498-3 Isoform 3 of Sodium channel protein type 1 subunit alpha n=332 Tax=Euteleostomi RepID=P35498-3
Sodium3	UniRef50_Q01118 Sodium channel protein type 7 subunit alpha n=123 Tax=Eutheria RepID=SCN7A_HUMAN
Sodium4	UniRef50_Q14524 Sodium channel protein type 5 subunit alpha n=1337 Tax=Gnathostomata RepID=SCN5A_HUMAN
Sodium5	UniRef50_Q8IZF0 Sodium leak channel non-selective protein n=334 Tax=Gnathostomata RepID=NALCN_HUMAN
Sodium6	UniRef50_Q9UI33 Sodium channel protein type 11 subunit alpha n=33 Tax=Boreoeutheria RepID=SCNBA_HUMAN

Supplementary Table 5. Character state matrix of the total number of channels and toxin

Animals	Total Sodium	Total Calcium	Total Potassium	Actinoporins	Cellulase Toxin	Cystat-toxin	Dermato-toxin	Total Jellyfish	Peptidase Toxin	Total Phospholipase	Total Toxin	Total SCRIPs	Sodium Toxin	Total Toxin Eight	Total Toxin	Total Neurotoxin	Total HKG	Total Channels
Amphimedon Queenslandica	1	4	2	0	1	0	0	0	1	0	1	0	0	0	3	1	54	7
Drosophila Melanogaster	2	5	11	0	0	0	0	0	1	2	4	0	0	0	7	4	56	18
Enteromyxum Leei	2	4	3	1	0	0	0	0	1	0	0	0	0	0	2	0	37	9
Kudoa Iwatai	1	1	3	0	0	0	0	0	1	0	2	0	0	0	3	2	48	5
Sphaeromyxa Zaharoni	1	0	0	0	1	0	0	0	1	0	3	0	0	0	5	3	41	1
Myxobolus Cerebralis	0	0	4	0	0	0	0	0	1	0	1	0	0	0	2	1	47	4
Thelohanellus Kitauei	0	1	1	0	0	0	0	0	1	0	1	0	0	0	2	1	29	2
Polypodium Hydriforme	1	5	5	0	0	0	0	0	1	0	3	0	0	0	4	3	55	11
Aurelia Aurita	2	8	9	0	1	1	0	1	1	1	3	0	0	1	9	4	56	19
Alatina Alata	2	3	3	0	1	0	1	2	1	3	3	0	0	0	11	3	58	8
Podocoryna Carnea	2	7	6	1	1	0	1	1	1	2	5	0	0	0	12	5	58	15
Turritopsis	2	4	6	0	1	1	1	1	1	2	2	0	0	0	9	2	57	12
Hydractinia Symbiolongicarpus	2	7	7	1	1	1	1	2	1	1	2	0	0	0	10	2	56	16
Hydra Vulgaris	1	5	6	1	1	0	0	0	1	2	2	0	0	0	7	2	59	12
Hydra Oligactis	1	6	5	2	1	1	1	0	1	1	1	0	0	0	8	1	52	12
Millepora Alcicornis	2	6	5	0	1	1	1	0	1	2	3	0	0	0	9	3	58	13
Acanthogorgia Aspera	1	5	5	1	0	0	0	0	1	1	2	1	0	0	6	3	55	11
Madracis Auretenca	2	6	6	0	1	0	1	0	1	3	4	2	1	0	13	7	55	14
Stylophora Pistillata	2	5	8	0	1	0	1	0	1	2	3	1	0	0	9	4	53	15
Pocillopora Damicornis	2	7	8	0	1	0	1	0	1	1	4	1	0	0	9	5	53	17
Seriatopora Hystrix	2	7	6	0	1	0	1	0	1	2	4	5	0	0	14	9	52	15
Fungia Scutaria	2	5	7	1	1	1	0	0	1	2	6	2	1	0	15	9	57	14

Ctenactis Echinata	1	4	4	0	0	1	0	0	1	0	1	0	0	0	3	1	59	9
Monstastaea Cavernosa	2	5	8	1	1	1	1	0	1	2	3	2	0	0	12	5	59	15
Platygyra Daedalea	1	4	3	2	0	1	0	0	1	1	2	1	0	0	8	3	49	8
Favia Lizardensis	1	5	7	1	0	1	0	0	1	0	2	2	0	0	7	4	56	13
Cyphastrea Serailia	2	7	5	1	1	1	1	0	1	1	3	4	0	0	13	7	58	14
Diploria Strigosa	1	6	8	1	1	0	1	1	1	1	3	5	0	0	14	8	51	15
Acropora Hyacinthus	2	3	6	0	0	0	1	0	1	2	2	1	0	0	7	3	40	11
Acropora Cervicornis	2	5	8	1	1	0	1	0	1	2	2	2	0	0	10	4	59	15
Acropora Digitifera	1	2	4	1	1	0	1	1	1	0	3	3	0	0	11	6	28	7
Porites Astreoides	2	5	9	1	1	0	1	1	1	3	6	1	0	0	15	7	59	16
Actinia Tenebrosa	2	6	8	1	1	1	1	0	1	3	11	1	0	3	23	15	53	16
Anthopleura Elegantissima	2	7	9	1	1	0	1	0	1	3	11	1	0	3	22	15	58	18
Exaiptasia Pallida	1	0	4	1	1	0	1	1	1	2	1	0	0	0	8	1	28	5
Nematostella Vectensis	2	6	10	0	1	1	1	0	1	3	3	0	0	1	11	4	58	18
Protopalythoa Variabilis	2	5	4	1	1	0	1	1	1	2	4	0	0	0	11	4	54	11
Palythoa Caribaeorum	2	6	8	0	1	0	1	1	1	2	7	0	0	0	13	7	56	16



Supplementary Figure 1. Cnidarian species tree with support values using a likelihood approach and a Bayesian approach

Appendix II: Supplementary information for Chapter 2

Supplementary File 1. Annotation report for RNA-seq assembled transcriptome of *Condylactis gigantea* is available at *https://drive.google.com/file/d/1IzAmytll9W5qWWyxXtvln6t1QML43SlA/view?usp=sharing*

Supplementary File 2. Annotation report for RNA-seq assembled transcriptome of *Entacmaea quadricolor* is available at https://drive.google.com/file/d/1Lii0YlPNVujyXlSUVHI1wFy5-CzQf9jl/view?usp=sharing

Supplementary File 3. Annotation report for MS/MS spectral results of *Condylactis gigantea* and *Entacmaea quadricolor* is available at https://drive.google.com/file/d/19DVYGE9jIJqFbLLrofuoCcdtoymUQSm8/view?usp=sharing

Appendix III: Supplementary information for Chapter 3

Supplementary File 1. Abundance values in CPM for all transcripts in *Condylactis gigantea* for triplicates of tentacles and column is available at https://drive.google.com/open?id=1dxjlVUFmad91-hqyzhInldBqO5TvH7hA

Supplementary File 2. Abundance values in CPM for all transcripts in *Entacmaea quadricolor* for triplicates of tentacles and column is available at https://drive.google.com/open?id=1s6oPOkz_hvtHJaFZORMF24SiwwruMdhf

Supplementary File 3. Abundance values in TMM for all transcripts in *Condylactis gigantea* for triplicates of tentacles and column is available at https://drive.google.com/open?id=1B2rkOE6z8onxwzb2_7HVqeVoYuSWioK4

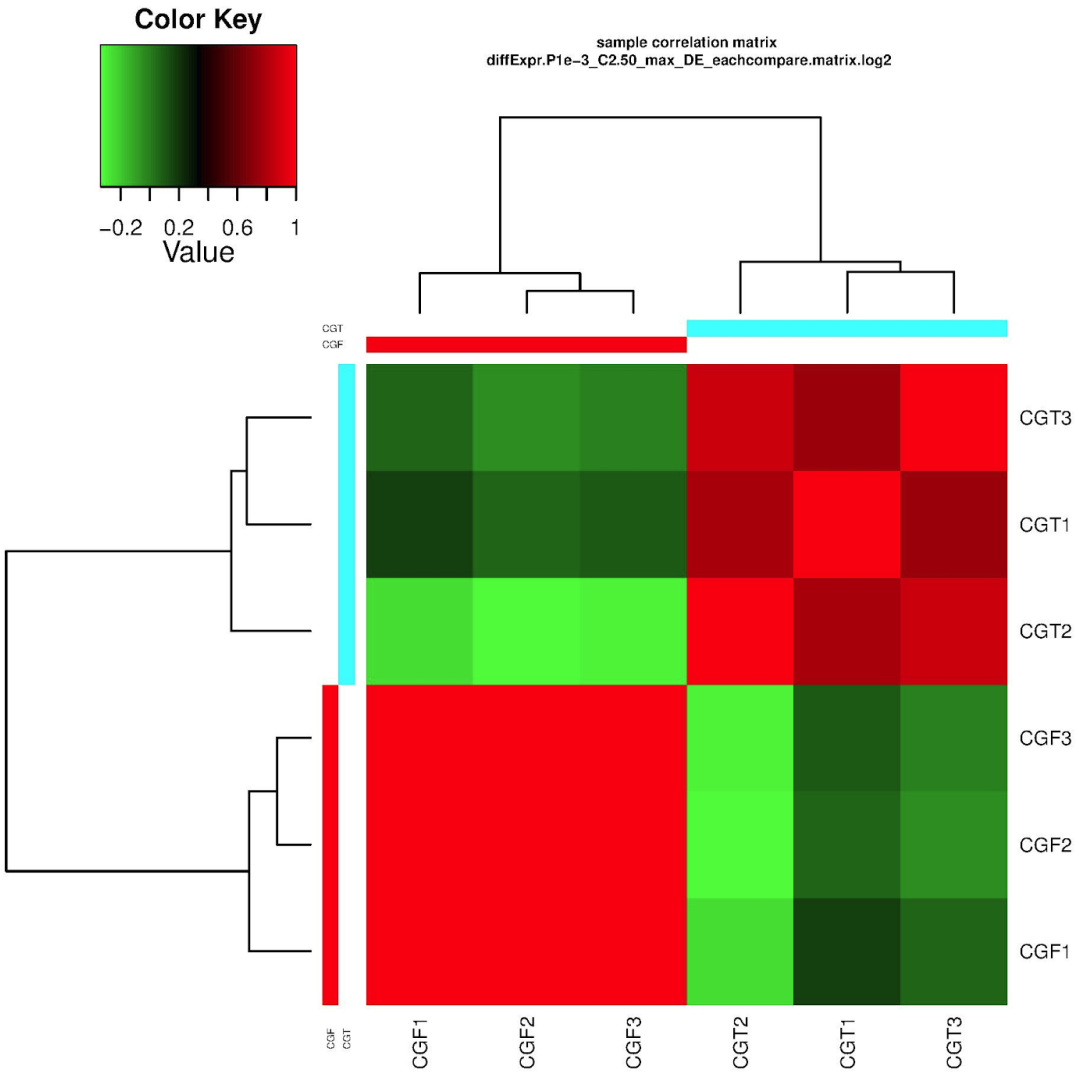
Supplementary File 4. Abundance values in TMM for all transcripts in *Entacmaea quadricolor* for triplicates of tentacles and column is available at https://drive.google.com/open?id=13NEh-oGeYnYM0TsAwMyDuLhSgnstBm9t

Supplementary File 5. Abundance values in CPM for toxins in *Condylactis gigantea* for triplicates of tentacles and column is available at https://drive.google.com/open?id=1tTBnohxHJc_8V6TpG_06D2KE6kKUrJeD6CwAueNEpoU

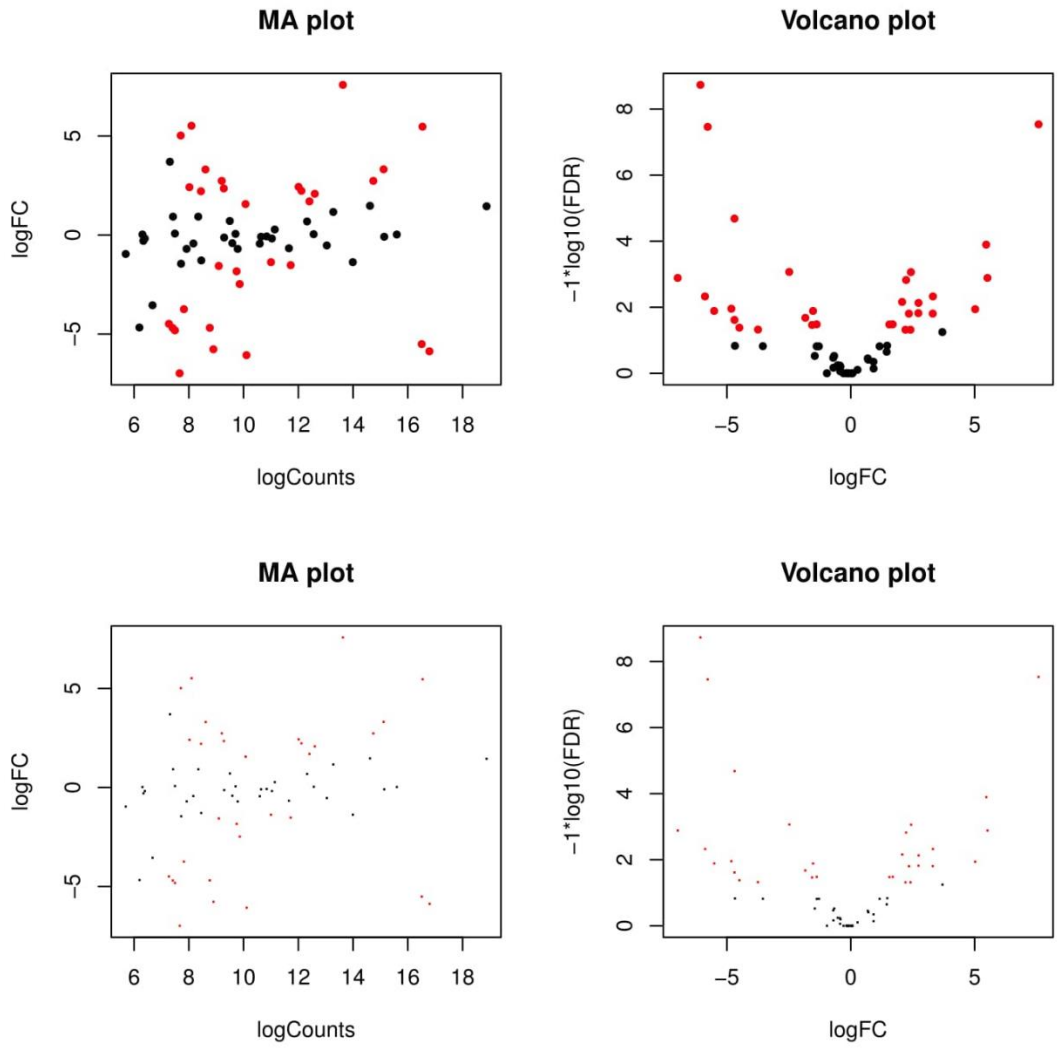
Supplementary File 6. Abundance values in CPM for toxins in *Entacmaea quadricolor* for triplicates of tentacles and column is available at https://drive.google.com/open?id=1Vj0LkteS9t3uxtsvQ_0Y23k807v4hKqVqA1kTvQQbJE

Supplementary File 7. Abundance values in CPM for Swissprot IDs in *Condylactis gigantea* for triplicates of tentacles and column is available at https://drive.google.com/open?id=1k4jnrBOkiZ9MS0FdychvTvcvx9DBkUDZ

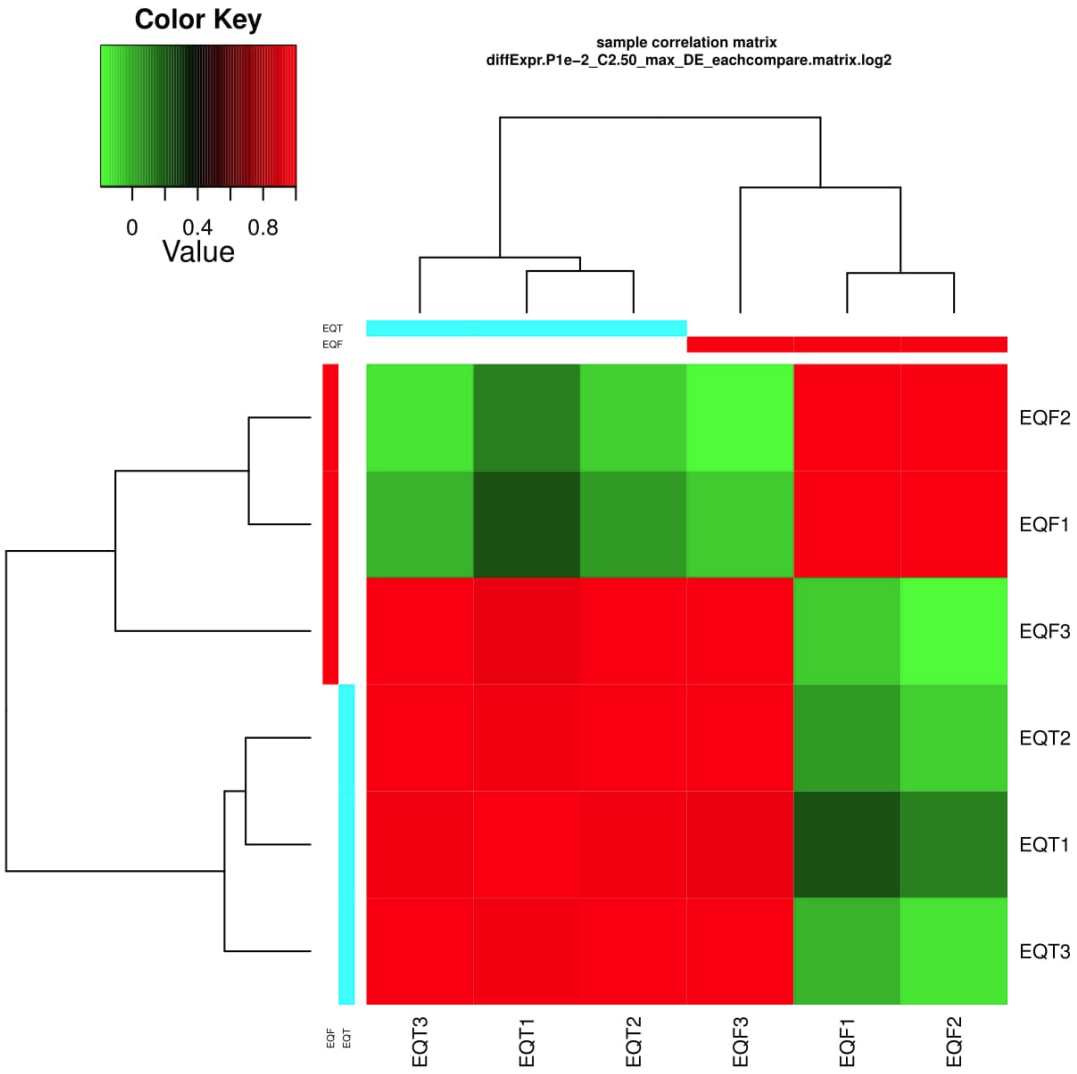
Supplementary File 8. Abundance values in CPM for Swissprot IDs in *Entacmaea quadricolor* for triplicates of tentacles and column is available at https://drive.google.com/open?id=1SG16xo2CD-13YW_uGGZVyxPWdutvdpte



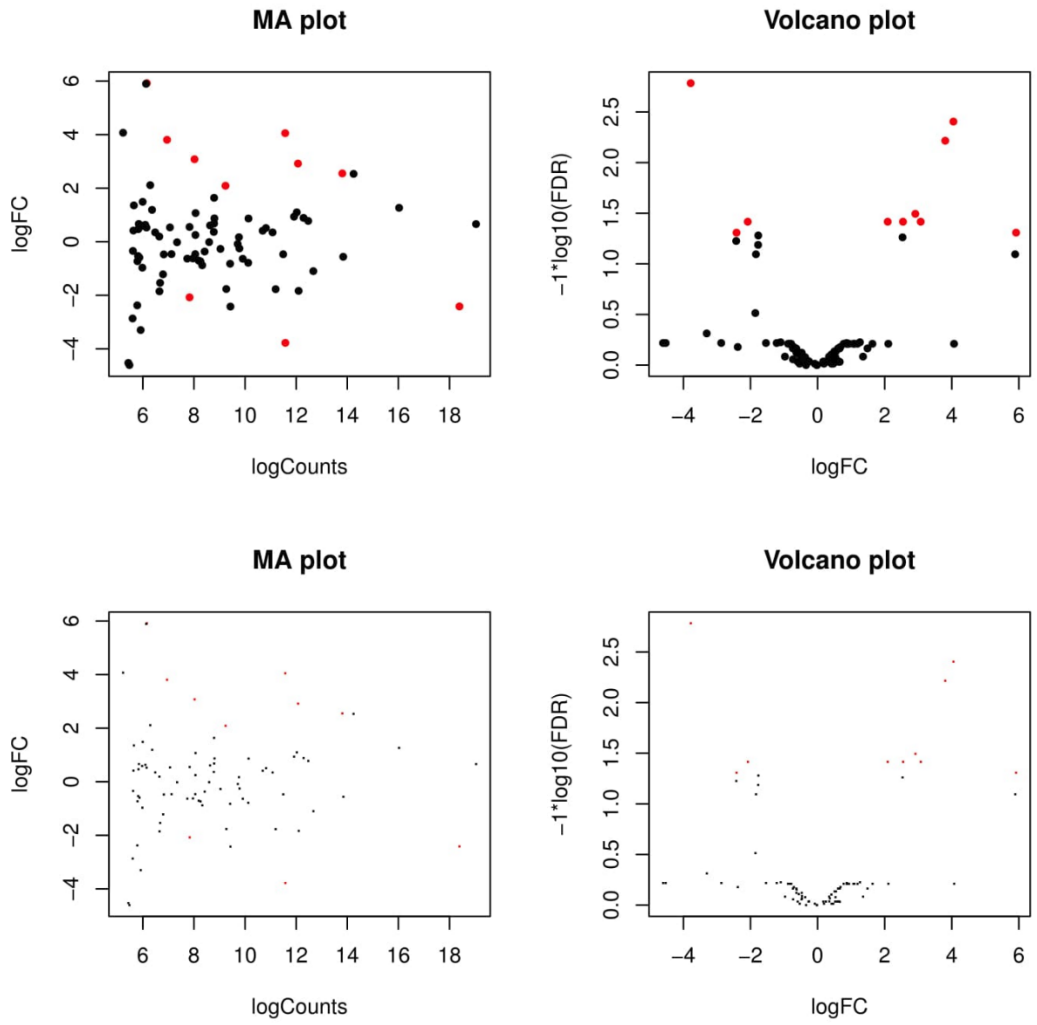
Supplementary Figure 1. Correlation map for toxins of *Condylactis gigantea* between biological replicates. CGF1, CGF2, and CGF3 represent the three tissue samples from the column of *Condylactis gigantea*. CGT1, CGT2, and CGT3 represent the three tissue samples from the tentacles of *Condylactis gigantea*. Red blocks represent high correlation between the samples on the x and y axis. The dendogram on the left of the heatmap represent the similarity between the samples based on the expression levels of the tissue samples. The figure was generated using Trinity.



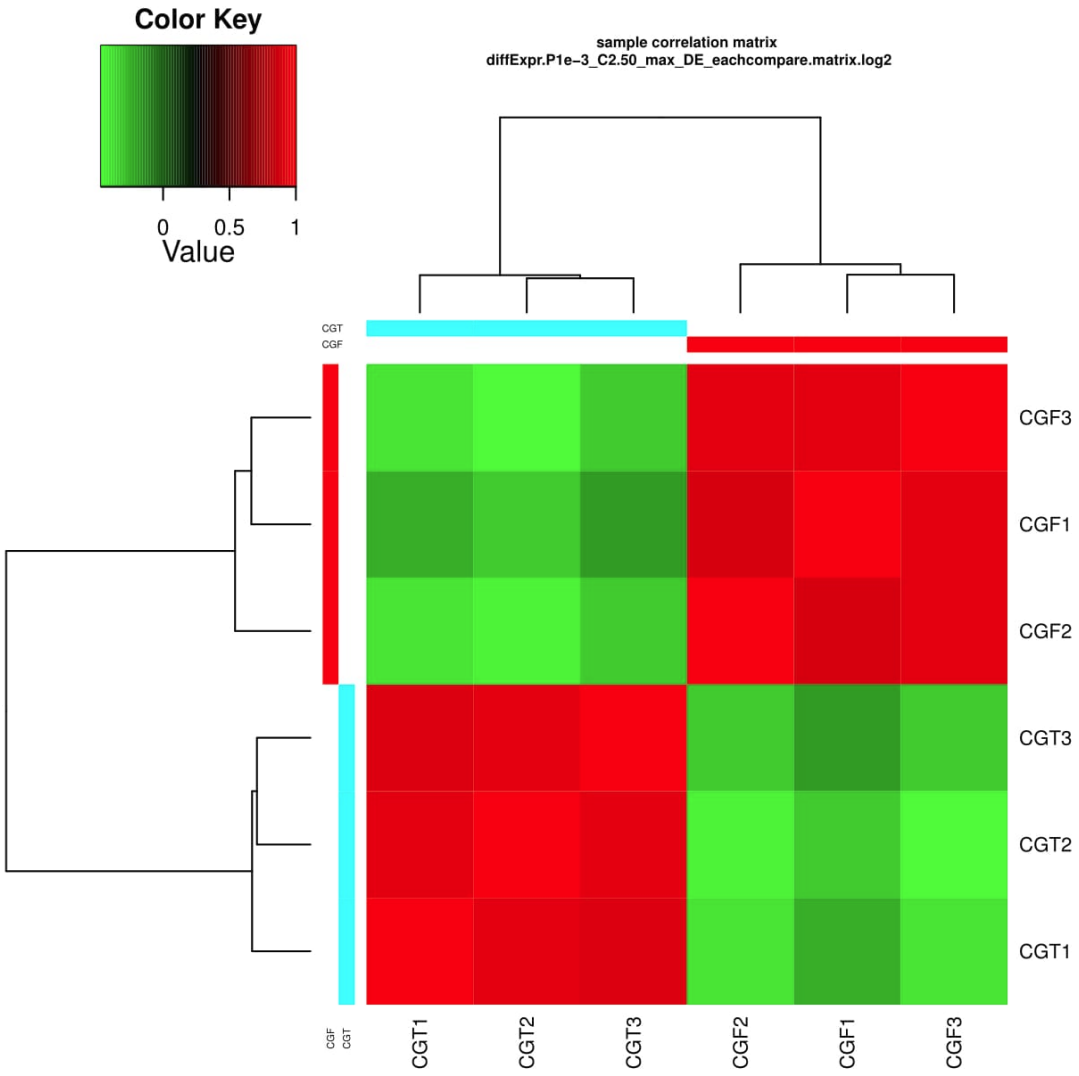
Supplementary Figure 2. MA plot and Volcano plot for toxins of *Condylactis gigantea* Tentacle Vs *Condylactis gigantea* column. The red dots represent the differentially expressed toxins, while the black dots represent toxins that are not differentially expressed. The MA plot is constructed between the log of the CPM and log of the Fold Counts (FC). The volcano plot is constructed between the log of the Fold Counts and log of the False Discovery Correction Rate (FDR).



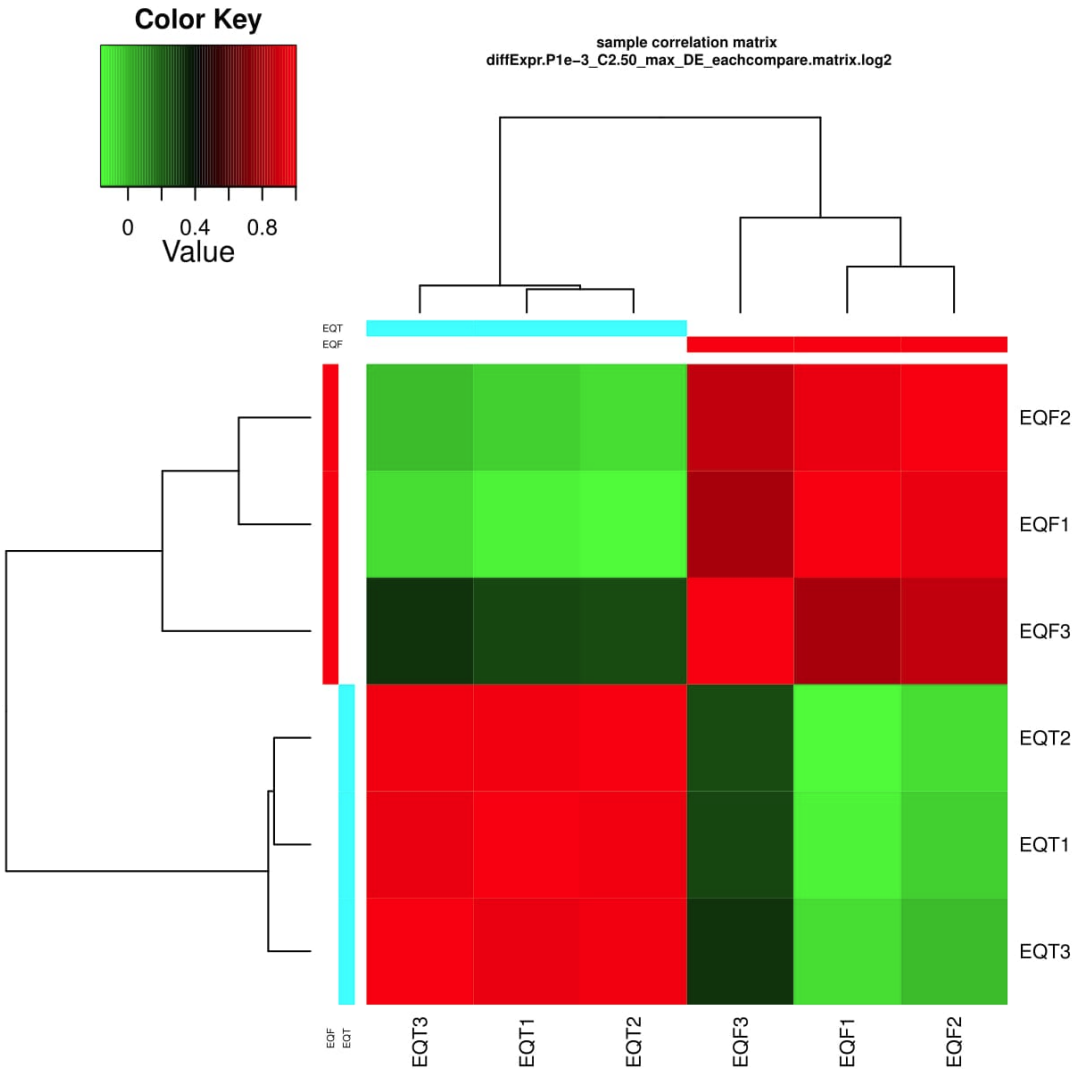
Supplementary Figure 3. Correlation map for toxins of *Entacmaea quadricolor* between biological replicates. EQF1, EQF2, and EQF3 represent the three tissue samples from the column of *Entacmaea quadricolor*. EQT1, EQT2, and EQT3 represent the three tissue samples from the tentacles of *Entacmaea quadricolor*. Red blocks represent high correlation between the samples on the x and y axis. The dendogram on the left of the heatmap represent the similarity between the samples based on the expression levels of the tissue samples. The figure was generated using Trinity.



Supplementary Figure 4. MA plot and Volcano plot for toxins of *Entacmaea quadricolor* Tentacles Vs *Entacmaea quadricolor* Column. The red dots represent the differentially expressed toxins, while the black dots represent toxins that are not differentially expressed. The MA plot is constructed between the log of the CPM and log of the Fold Counts (FC). The volcano plot is constructed between the log of the Fold Counts and log of the False Discovery Correction Rate (FDR).



Supplementary Figure 5. Correlation map for uniprot IDs of *Condylactis gigantea* between biological replicates. CGF1, CGF2, and CGF3 represent the three tissue samples from the column of *Condylactis gigantea*. CGT1, CGT2, and CGT3 represent the three tissue samples from the tentacles of *Condylactis gigantea*. Red blocks represent high correlation between the samples on the x and y axis. The dendogram on the left of the heatmap represent the similarity between the samples based on the expression levels of the tissue samples. The figure was generated using Trinity.



Supplementary Figure 6. Correlation map for Uniprot IDs of *Entacmaea quadricolor* between biological replicates. EQF1, EQF2, and EQF3 represent the three tissue samples from the column of *Entacmaea quadricolor*. EQT1, EQT2, and EQT3 represent the three tissue samples from the tentacles of *Entacmaea quadricolor*. Red blocks represent high correlation between the samples on the x and y axis. The dendogram on the left of the heatmap represent the similarity between the samples based on the expression levels of the tissue samples. The figure was generated using Trinity.