

## Nemertean Toxin Genes Revealed Through Transcriptome Sequencing

Nathan V. Whelan<sup>1\*</sup>, Kevin M. Kocot<sup>1,2</sup>, Scott R. Santos<sup>1</sup>, Kenneth M. Halanych<sup>1</sup>

<sup>1</sup>: Department of Biological Sciences & Molette Biology Laboratory for Environmental and Climate Change Studies, Auburn University, 101 Rouse Life Sciences, Auburn, Alabama 36849.

<sup>2</sup>: Current address: School of Biological Sciences, University of Queensland, Brisbane QLD 4072, Australia.

\*Author for Correspondence: Nathan Whelan, Department of Biological Sciences, Auburn University, Auburn, Alabama, USA, Phone: 334-844-3223, Fax: 334-844-2333, NVW: [nwhelan@auburn.edu](mailto:nwhelan@auburn.edu) KMH: [ken@auburn.edu](mailto:ken@auburn.edu)

Raw transcriptome data have been submitted at GenBank under accessions SRX731465, SRX731466, SRX732127

© The Author(s) 2014. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract**

Nemerteans are one of few animal groups that have evolved the ability to utilize toxins for both defense and subduing prey, but little is known about specific nemertean toxins. In particular, no study has identified specific toxin genes even though peptide toxins are known from some nemertean species. Information about toxin genes is needed to better understand evolution of toxins across animals and possibly provide novel targets for pharmaceutical and industrial applications. We sequenced and annotated transcriptomes of two free-living and one commensal nemertean and annotated an additional six publicly available nemertean transcriptomes to identify putative toxin genes. Approximately 63-74% of predicted open reading frames in each transcriptome were annotated with gene names, and all species had similar percentages of transcripts with each higher-level GO term. Every nemertean analyzed possessed genes with high sequence similarities to known animal toxins including those from stonefish, cephalopods, and sea anemones. One toxin-like gene found in all nemerteans analyzed had high sequence similarity to Plancitoxin-1, a DNase II hepatotoxin that may function well at low pH, which suggests the acidic body walls of some nemerteans could work to enhance the efficacy of protein toxins. The highest number of toxin-like genes found in any one species was seven and the lowest was three. The diversity of toxin-like nemertean genes found here is greater than previously documented, and these animals are likely an ideal system for exploring toxin evolution and industrial applications of toxins.

**Keywords:** Nemertea, ribbon worm, toxin, cytotoxin, hepatotoxin, transcriptome.

## Introduction

Many animal lineages have independently evolved the use of toxins to enhance their fitness (Fry et al. 2009a, b; Casewell et al. 2013; Undheim et al. 2014), and humans have capitalized on a multitude of these chemicals for pharmaceuticals and other industrial applications (Kalman et al. 1998; Lewis and Garcia 2003; King 2011; Upadhyay et al. 2013). Depending on the species, toxins utilized by animals may be obtained through ingestion of other organisms (Paul and Van Alstyne 1988; Paul and Pennings 1991), produced by bacterial symbionts (Carroll et al. 2003) or synthesized by the animal itself (Casewell et al. 2013). Generally, toxins are employed for either prey capture [e.g., cone snails (Olivera et al. 1985), venomous snakes (Lee 1979), cnidarians (Turk and Kem 2009)] or defense from predators [e.g., poison dart frogs (Daly et al. 1978; Saporito et al. 2012) and rough skinned newts (Brodie 1968)]. Toxin genes are particularly interesting from an evolutionary perspective as toxic proteins with similar form and function show convergent evolution in different animal groups (reviewed by Fry et al. 2009a; Casewell et al. 2013). Despite the potential utility of toxins for biomedical and other applications, precise chemical properties of many toxins and the breadth of toxins produced across life, remain unknown. This lack of information limits our understanding of toxin evolution as well as evolution of predation and defense mechanisms (Brodie and Brodie 1990; Jost et al. 2008).

One group that has received relatively little attention in regards to toxins is Nemertea, or ribbon worms, with multiple recent reviews on animal toxins having failed to mention this group (Fry et al. 2009a, b; Casewell et al. 2013). Nemertea contains approximately 1,500 species of mostly marine organisms (Gibson 1995; Kajihara et al. 2008), and they are perhaps best known for a long, eversible proboscis, used in prey capture (Stricker and Cloney 1981; McDermott and

Roe 1985; Thiel and Kruse 2001). Although the proboscis is typically used to capture and immobilize prey, nemerteans can also employ toxins to paralyze or kill prey (McDermott and Roe 1985; Asakawa et al. 2013). Nemerteans also utilize toxins for predator defense (Bacq 1937; Kem 1971, 1976; Kem et al. 1971; McClintock and Baker 1997; Carroll et al. 2003) by secreting toxins from cells in their epidermis (Norenburg 1985). For example, toxins in the body wall of *Parbolasia corrugatus* provide defense from potential predators in Antarctic waters (McClintock and Baker 1997). Nemerteans can be prominent members of marine benthic (Gibson 1983; McClintock and Baker 1997) and pelagic communities (Roe and Norenburg 1999), but many species appear to have few or no predators, apparently because of the effectiveness of their acidic and/or chemically defended body walls (Ling 1970; Dayton et al. 1974; Ferraris 1979; Ali et al. 1990; Heine et al. 1991; McClintock and Baker 1997). Despite toxins being an integral part of nemertean biology, few have been characterized.

Information about nemertean toxin genes and their products is needed to understand how nemerteans work to enhance their predatory and defensive effectiveness. To date, the limited studies on nemertean toxin genes and their products have mostly focused on non-peptides such as tetrodotoxin (Ali et al. 1990; Carroll et al. 2003; Tanu et al. 2004) and anabaseine (Kem 1971; Kem et al. 1971). Thus far, only three peptide toxins have been described from nemerteans: cytotoxin A-III in *Cerebratulus lacteus* and *Parbolasia corrugatus* and neurotoxin B-II and neurotoxin B-IV in *C. lacteus* (Kem 1976, 1994; Blumenthal et al. 1981; Barnham et al. 1997; Berne et al. 2003). Overall, genomic underpinnings of most nemertean toxins and whether the aforementioned peptide toxins are found in other nemertean species have not been assessed. As such, whether nemerteans primarily obtain toxins from the environment and/or symbionts (e.g.,

tetrodotoxin from bacteria; Carroll et al. 2003) or if they possess an unrecognized diversity of toxin genes is unclear.

Transcriptome sequencing can be a critical tool for understanding the ecology and evolution of understudied animals and provides resources for future scientific study. Furthermore, advances in sequencing technologies and bioinformatics allow for comprehensive characterization of animal transcriptomes and exploration of toxin genes with relative ease (Garber et al. 2011; Hass et al. 2013). In this study, we searched transcriptomes of nine nemerteans for toxin genes. We sequenced and characterized transcriptomes of two free-living nemerteans, *Paranemertes peregrina* and *Tubulanus polymorphus*, and one commensal nemertean, *Malacobdella grossa*. We also analyzed publicly available transcriptome data of *Cephalothrix hongkongiensis*, *Cephalothrix linearis*, *Cerebratulus marginatus*, two *Lineus* species, and *Ramphogordius lacteus* to further characterize nemertean transcriptomes and survey toxin gene diversity across the group. We consider any gene coding for a protein with high similarity to known toxins in other animal groups as a putative toxin gene. *Malacobdella grossa*, which lives in the mantle cavity of bivalves, was hypothesized to have fewer toxin genes than other species to potentially minimize harm to its host. . Similarities between nemertean genes and toxins from other animals add to the growing body of evidence supporting convergent evolution in animal-derived peptide toxins.

## Material and Methods

### *Specimen Sampling and Sequencing*

*Malacobdella grossa* was collected off Rhode Island by a commercial vessel harvesting the bivalve *Arctica islandica* as part of the study by Dahlgren et al. (2000). *Paranemertes*

*peregrina* and *T. polymorphus* were collected in False Bay, San Juan Island, Washington. RNA extraction, cDNA library preparation, and Illumina sequencing generally followed the methods of Weigert et al. (2014). Briefly, we extracted RNA from *M. grossa* and *P. peregrina* using whole animals and from *T. polymorphus* using the anterior three quarters of the specimen with TRIzol (Invitrogen). RNA was purified with the Qiagen RNeasy kit (Valencia, California) using on-column DNase digestion. Complimentary DNA (cDNA) libraries were constructed with the SMART cDNA library construction kit (Clontech Laboratories, Mountain View, California) following the manufacturer's protocol except that the provided 3' oligo was replaced with the Cap-Trsa-CV oligo as per Meyer et al. (2009). Full-length cDNA was then amplified using the Advantage 2 PCR system (Clontech) with a minimum number of PCR cycles (i.e., 17-21) and sent to HudsonAlpha Institute for Biotechnology (Huntsville, AL, USA) for library preparation and sequencing on an Illumina HiSeq 2000 using 2 x 100 bp paired-end (PE) chemistry. Data for five other nemerteans were retrieved from NCBI (Table 1; Riesgo et al. 2012).

### *Transcriptome Assembly*

Transcriptome read quality was assessed with the FASTX toolkit (Gordon 2011). Given overall high read quality, sequences were not filtered prior to assembly. All raw data underwent digital normalization using the python script normalize-by-median.py (Brown et al. 2012) with a k-mer size of 20, a desired coverage (i.e., cutoff) of 30, and four hash tables with a lower bound of  $2.5 \times 10^9$ . Trinity version November 2013 (Grabherr et al. 2011; Haas et al. 2013) was utilized for transcriptome assembly of normalized reads for each species with a k-mer size of 25. Raw reads were assembled as paired-end data except for the two *Lineus* species and *R. lacteus*, for which only single-end reads were available.

### *Assessment of Assembly Quality*

A rarefaction curve (Sanders 1968) of assembly statistics was used to evaluate quality and completeness of assemblies from the newly sequenced species because a nemertean reference genome was not available for comparison. Specifically, we removed 10-70% of paired-end sequences from the end of raw read files to produce datasets with reduced numbers of reads. These subsampled datasets were assembled as above, and N50 and total number of contigs for each assembly was plotted. A plateau on the rarefaction curve would indicate that adding more sequence data would not considerably change the characteristics of the transcriptome assembly. We further measured completeness of each transcriptome assembly with CEGMA 2.4 (Parra et al. 2007), which determines how many of 248 core eukaryotic genes were present in each transcriptome. Core genes annotated by CEGMA are ones that are highly conserved and chosen from the eukaryotic orthologous groups database (Tatusov et al. 2003). An advantage to an approach like CEGMA is that potential differences in which genes are present among assemblies, as a result of expression differences in various tissue types used for cDNA library preparation, is minimized because housekeeping genes should be expressed in virtually all cells.

### *Transcriptome Annotation*

Annotation of each assembled transcriptome was done with the Trinotate annotation suite (<http://trinityrnaseq.sourceforge.net/annotation/Trinotate.html>). Briefly, TransDecoder (Haas et al. 2013) was first used to predict open reading frames (ORFs) of at least 300 bp. If multiple, overlapping ORFs were present in the same contig, only the longest ORF was retained. In contrast, if multiple but non-overlapping 300 bp ORFs were identified, all were retained. Thus,



two or more ORFs could originate from the same transcript (i.e. ORFs on both forward and reverse strands and/or multiple ORFs on the same strand for long contigs). Untranslated transcripts and translated ORFs were then queried against the Swiss-Prot database (UniProt Consortium 2014) using BLASTx and BLASTp, respectively (Altschul et al. 1997), with annotation coming from the best blast hit and associated Gene Ontology terms (Ashburner et al. 2000). Trinotate then used HMMER 3.1 tool hmmscan (Eddy 2001; Finn et al. 2011) and the Pfam-A database (Punta et al. 2014) to annotate protein domains for each predicted protein sequence. Trinotate results were populated into a SQLite database and placed into a tab delimited file with scripts provided in the Trinotate package and a custom wrapper (available from <http://github.com/halocaridina/bioinformatic-scripts>). To roughly characterize the protein composition of each nemertean transcriptome, a custom python script (available from <http://github.com/NathanWhelan>) was used to place GO terms for each UniProt annotated transcript from Trinotate into WEGO format, and annotated GO terms were visualized using the WEGO web service (Ye et al. 2006).

Toxin genes were identified based on sequence similarity to previously characterized animal toxins genes under the assumption that sequence similarity is generally indicative of function (Gabaldón and Huynen 2004). Putative toxin genes were initially distinguished if top BLASTx and/or BLASTp hits in the Trinotate output were a previously characterized eukaryotic toxin gene (as defined by Swiss-Prot or presence of a Pfam domain with “toxin” in the description). Genes of putative viral or bacterial origin were discarded, which eliminated putative toxin genes that may have been from bacterial endosymbionts or from horizontal gene transfer events. Amino acid sequences of remaining transcripts were then manually searched against the NCBI non-redundant GenBank database (nr) and the Pfam protein domain database using the



HMMER 3.1 tool phmmer (Eddy 2001; Finn et al. 2011). Annotated sequences initially identified as a toxin gene by Swiss-Prot were further considered a toxin gene if either 1) cross validation produced a significant hit for a toxin domain family in Pfam or 2) if the best annotated hit from the nr database was labeled a toxin gene. In some instances, a toxin as identified by Trinotate and Swiss-Prot did not have a toxin gene as highest hit against the nr database using phmmer, nor did these transcripts possess a toxin protein domain according to Pfam-A; such transcripts were not further considered as they were potential false positives. Toxin genes passing the above filters were then reciprocally queried against the other transcriptomes to identify putative orthologs via BLAST searches. We also queried previously identified nemertean peptide toxins (i.e., Cytotoxin A-III, Neurotoxin B-II, and Neurotoxin B-IV; Kem 1976; Blumenthal et al. 1981) against all transcriptomes with a tBLASTn search.

### *Gene Tree Reconstruction*

Gene trees were inferred for SNTX-like and Plancitoxin-1-like genes because they were found in all nine nemerteans. Putative stonefish toxin (SNTX) and Plancitoxin genes were translated with TransDecoder (Haas et al. 2013) using default settings. Redundant nemertean protein sequences were then removed from the dataset. The SNTX dataset of von Reumont (2014) was added to the nemertean SNTX-like genes. Putative Plancitoxin-1 genes and non-toxic DNase II genes were retrieved from UniProt and GenBank (Table s1; figs. s1, s2). Alignments were done in MAFFT 3 with the E-INS-i algorithm (Katoh and Standley 2013). The appropriate model of protein evolution was selected for each gene using ProtTest 3.4 (LG+F+ $\Gamma$  for both genes; Darriba et al. 2011). RAxML 8 (Stamatakis 2014) was used to infer maximum likelihood gene trees, and 1000

non-parametric bootstrap replicates were performed to assess nodal support. Trees were rooted with non-toxic homologs.

## Results

### *Transcriptome Assembly and Characterization*

Average read quality score was above 30 for all taxa except *P. peregrina*, which had an average quality score of 28.7. Assemblies from Trinity (figshare, doi:10.6084/m9.figshare.1203580) possessed an average of 111,441.7 ( $\pm 40,859$  sd) contigs per taxon—full assembly and sequencing statistics are presented in Table 1. Rarefaction curves of the three taxa sequenced here showed a steady increase in N50 and number of contigs as the amount of total data increased (fig. s1), with a flat plateau not being reached for either measurement. However, the overall trend strongly suggested that increasing the number of reads would result in diminishing returns regarding the total number of transcripts recovered. Furthermore, the high percentage of CEGMA core genes present ( $> 85\%$  of core genes had complete sequences;  $> 93\%$  of core genes were partially sequenced) in each full assembly of newly sequenced transcriptome (Table 1) suggests that each assembly possesses a representative snapshot of the mRNA content expected from a typical metazoan cell. In contrast, assemblies from *Cerebratulus marginatus* and *Cephalothrix hongkongiensis* possessed much lower percentages of CEGMA core genes ( $< 30\%$  core genes had complete sequences; Table 1).

[[Table 1]]

Annotation reports from Trinotate for each transcriptome can be found on figshare (doi:10.6084/m9.figshare.1203580). In total, 18.2-46.2% of putative transcripts had predicted ORFs and 63.1-91.7% of inferred ORFs were annotated with a significant BLASTp hit (Table 2).

Of these annotated transcripts, most had multiple GO terms. For all species, the highest percentage of genes were annotated as biological process (44-48%). Lower percentages of genes were annotated as either cellular component (28-30%) or molecular function (24-26%) (fig. 1A-G; data for *Cerebratulus marginatus* and *Cephalothrix hongkongiensis* not shown, but can be seen in Riesgo et al. 2012). Overall, lower-level GO term annotations were similar across all nine species, but *M. grossa*, *P. peregrina*, and *T. polymorphus* typically had more genes annotated for any given term than other species (fig. 1H). Auxiliary transport protein, metabolic chaperone, and toxin binding are notable exceptions to this pattern (fig. 1H).

[[Table 2]]

#### *Nemertean Toxin Genes*

Putative toxin genes were recovered from all nemertean transcriptomes examined here (Table 3), and the number of such genes ranged from three in the commensal *M. grossa* to seven in *Cerebratulus marginatus*, *L. ruber*, and both *Cephalothrix* species. In some cases, multiple transcripts of the same toxin-like gene were found in each assembly. The two most common putative toxin genes were Plancitoxin-1-like (Shiomi et al. 2004) and stonefish-like toxin genes (SNTX/VTX family; Ueda et al. 2006). All Plancitoxin-1-like genes possessed a DNase II protein domain. Six other types of toxin-like genes were identified, but less frequently, among species analyzed. Putative toxin genes with at least two ShK protein domains downstream of cysteine rich secretory domains (CAP; Gibbs et al. 2008) were found in *M. grossa*, *P. peregrina*, and *Cephalothrix linearis*. Natterin-4-, Echotoxin-2-, and SE-cephalotoxin-like genes were recovered in four, five, and six of the nemerteans analyzed here, respectively (Table 3). Toxin-like genes in the membrane-attack complex/perforin family (MACPF) (Nagai et al. 2002; Oshiro et al. 2004; Satoh et al. 2007) were found only in *Cerebratulus* species. Previously identified

nemertean peptide toxins, Neurotoxin B-II and Neurotoxin B-IV were not found, but we did not sequence the only species from which they are known (i.e., *Cerebratulus lacteus*). Four species expressed the other known nemertean peptide toxin, Cytotoxin A-III.

[[Table 3]]

### *Toxin Gene Trees*

The SNTX dataset had 33 sequences and 577 aligned amino acid positions; the best tree had a likelihood value of -17869.9071 (fig. s2). The Plancitoxin/DNase II dataset had 23 sequences and 442 aligned amino acid positions; the best tree had a likelihood value of -10952.2682 (fig. s3). Both datasets have been deposited on figshare (doi:10.6084/m9.figshare.1203580). Nemertean SNTX-like genes were all more closely related to the recently described SNTX-like genes from *Glycera* polychaetes than stonefish genes. A large clade of all nemertean SNTX-like transcripts and *Glycera* SNTX genes were sister to toxins from stonefish. Plancitoxin-1 like genes were more closely related to Plancitoxin-1 from *Acanthaster planci*—the species from which it was described—, but nodal support values were so low for other relationships that they should be approached with caution.

## **Discussion**

### *Nemertean Toxins*

Nemerteans have long been known to use toxins in predation and defense (Kem 1971; Ali et al. 1990; Heine et al. 1991; McClintock and Baker 1997; Asakawa et al. 2013), but our findings indicate a previously unidentified diversity of putative toxin genes among members of the phylum (Table 3). Prior studies of nemertean toxins have focused primarily on *Parbolasia corrugatus* or *Cerebratulus* spp. and only identified three protein toxins (Kem 1976, 1994;

McClintock and Baker 1997; Berne et al. 2003). In contrast, annotation of nemertean transcriptomes found expression of multiple putative toxin genes, of which only one was previously known from nemerteans. Differences in the presence of specific toxin genes hint that their expression may depend on the life habits and/or evolutionary history of each species, particularly for the commensal *M. grossa*, which had the fewest observed toxin genes. However, other factors such as active or resting state of venom glands when each animal was sacrificed, gender, or physiological state could also be responsible for such differences (Morgenstern et al. 2011; Menezes et al. 2006).

Plancitoxin-1, a protein with homologs found in all nine nemerteans analyzed here and one of the few known toxic DNase II proteins (Shiomi et al. 1990; Shiomi et al. 2004; Ota et al. 2006), is a hepatotoxin found in the crown-of-thorn starfish (*Acanthaster planci*). *Acanthaster* Plancitoxin-I is inactive below pH 5.5 and most active at pH 7.2 (Shiomi et al. 2004), but mammalian DNase II proteins function best at pH 5.0 (Counis and Torriglia 2000; Evans and Arguiera 2003). Some nemerteans are known to have acidic body walls (e.g., *Pa. corrugatus*, *Cephalothrix spiralis*, *Lineus socialis* Ferraris 1979; Heine et al. 1991), and putatively toxic DNase II proteins in nemerteans may function well at low pH levels like mammalian DNase II proteins. As such, nemerteans that secrete acidic mucus may do so to improve the effect of toxin peptides, rather than as a standalone defensive mechanism.

Genes with high similarity to toxins in the SNTX/VTX protein family were also found in all nine nemerteans analyzed. SNTX/VTX toxins have cytolytic properties and were found previously in fish (Ueda et al. 2006, 2007), monotremes (Wong et al. 2013), and *Glycera* annelids (von Reumont et al. 2014). These toxins are utilized for defense, sexual competition, and predation in fish, monotremes and *Glycera*, respectively. Because whole animals, rather than

only epidermis or proboscis associated glands, were used for most RNA extractions here, tissue specific isolation of SNTX/VTX toxins will be necessary to determine if such toxins are used by nemerteans for defense, predation, or both.

Of the three previously identified nemertean peptide toxins, we only found transcripts for Cytotoxin A-III-like proteins (Kem 1994; Berne et al. 2003). This may indicate that Neurotoxin B-II and Neurotoxin B-IV are not commonly expressed in many nemerteans. Cytotoxin A-III has been found in the epidermis of some nemerteans, where it serves as a defensive neurotoxin (Kem and Blumenthal 1978; Berne et al. 2003). This toxin was initially described from the heteronemertean *Cerebratulus lacteus*. As expected, it was also present in the closely related *Cerebratulus marginatus*. Furthermore, both *Lineus* species and *R. lacteus*, all three of which are also heteronemerteans (Sundberg et al. 2001; Thollesson and Norenburg 2003; Kvist et al. 2014), also expressed Cytotoxin A-III-like genes.

In contrast to Cytotoxin A-III being present only in heteronemerteans, evolutionary relatedness alone does not appear to explain which species were found to express SE-cephalotoxin-, Echotoxin-2- and Natterin-4-like genes. For example, SE-cephalotoxin-like genes were expressed in four heteronemerteans and the paleonemertean *Cephalothrix hongkongiensis* but were not found in the remaining species. However, this could be a result of different expression patterns rather than SE-cephalotoxin-like genes being absent from the genomes of the other four species. Presence or absence of Echotoxin-2- and Natterin-4-like genes was also similar to that of SE-cephalotoxin-like genes. Interestingly, SE-cephalotoxin and natterins have previously been found only in cephalopods and fish, respectively (Tamura et al. 2011; Barathkuma et al. 2013). SE-cephalotoxin is found in salivary glands of the cuttlefish *Sepia esculenta* and is utilized in predation, but its precise biochemical function is unknown (Ueda et

al. 2008). On the other hand, natterins from fish all have aerolysin domains and induce edema and nociception (Xue et al. 2012). Nemertean Natterin-4-like genes also possess an aerolysin domain, which is suggestive of a similar function.

While ShK toxins and their characteristic domains were initially described from the sea anemone *Stichodactyla helianthus* (Castañeda et al. 1995), they have also been found in snake venom (Shikamoto et al. 2005; Wang et al. 2006) and *Glycera* annelids (von Reumont et al. 2014). Interestingly, ShK toxin-like genes identified here were more similar at the amino acid level to ShK toxin genes in snakes (Table 3) with whom nemerteans share a more recent common ancestor than sea anemones. ShK domains in nemertean ShK toxin-like proteins were always downstream of cysteine-rich secretory domains, which have toxic properties in other invertebrates (e.g., in Hymenoptera; Lu et al. 1993) further indicating toxic properties of these peptides. Because ShK toxins function as potassium-channel blockers, they have been proposed as human pharmaceutical targets to treat obesity (Upadhyay et al. 2013), graft rejection, and autoimmune diseases (Kalman et al. 1998). Thus, nemertean ShK-like toxins may have applicability in future drugs.

Membrane-attack complex/perforin family proteins (MACPF) are found in bacteria and many eukaryotes where they function in innate immunity (Young et al. 1986; Morito-Yamamuro et al. 2005; Tosi, 2005), pathogenesis (Kaiser et al. 2004), or less commonly, in predatory/prey interactions (Oshiro et al. 2003; Satoh et al. 2007). Given such a wide utility, it is not surprising that we found genes encoding MACPF protein domains. Genes coding for MACPF proteins in both *Cephalothrix* species had high sequence similarity at the amino acid level to toxin genes found in sea anemones *Actinaria villosa* (Oshiro et al. 2003) and *Phyllodiscus semoni* (Satoh et



al. 2007). This suggests that some nemerteans may use MACPF toxins in defense and/or prey capture in an analogous fashion as sea anemones.

### *Nemertean Transcriptome Characterization*

Fewer than half of all contigs from assemblies of each species were annotated with ORFs or other annotations, but of those contigs with predicted ORFs over 60% in each species were annotated with a Swiss-Prot hit (Table 2). Unannotated contigs likely result from a combination of incompletely assembled transcripts, non-coding regions, and proteins of unknown function. Notably, the percentage of contigs with predicted open reading frames ranged from 18.2% in *Cerebratulus marginatus* to 46.2% in *T. polymorphus*. In contrast, the percentage of ORFs with Swiss-Prot annotation hits for each species was higher and more uniform (63.1%-74.4%; Table 2). This indicates the presence of a group of nemertean proteins with unknown function—some of which could be toxins. Additionally, uniformities in the percentage of genes classified for higher level GO terms among all taxa provide evidence of conserved higher-level function of many nemertean genes and conserved cellular physiology in nemerteans (fig. 1A-G). Such characterizations fail to capture similarities or differences in unannotated genes, but they highlight the general homogeneity of transcriptome function among nemerteans. For example, nearly half (44-48%) of all annotated ORFs in each species were categorized as biological processes, and differences among species for the other two higher-level GO terms varied by no more than 2%. Percentages of genes annotated for lower-level GO terms were more varied (fig. 1), which provides evidence of some transcriptome specialization and/or differences in gene expression among nemertean lineages. Despite advances in sequencing technologies, functional

work is still needed to characterize the large fraction of transcripts and putative proteins of unknown function.

### *Conclusions*

Transcriptome analysis of nine nemertean species uncovered nine different types of toxin-like genes—eight of which were previously unknown from nemerteans. Nemerteans are one of three animal lineages known to utilize toxins for both predation and defense (Casewell et al. 2013), and peptide toxins clearly contribute to the overall toxin suite in ribbon worms. Two recent reviews emphasized convergent evolution of predatory and defensive toxins in animals (Fry et al. 2009a; Casewell et al. 2103). Furthermore, Fry et al. (2009b) demonstrated that toxin genes in cephalopods resemble toxins in other lineages and concluded that peptide toxins have evolved from non-toxic proteins that are found across animals. These studies did not include nemerteans, but similarities among putative nemertean toxins and toxin proteins in disparate lineages (e.g., cnidarians, molluscs, fish) indicate that convergent evolution of toxins from non-toxic proteins may have also occurred in nemerteans. Identifying putative nemertean toxins here relied on sequence similarities to well-known, annotated animal toxins, and many more toxins that are unique to nemerteans may await discovery. Sequence and structural similarity to peptide toxins in other organisms is evidence of shared function, but experimental studies are needed to understand the precise function of putative toxin genes identified here and to explore how different toxins work together to provide defense from predators or enhance predatory behaviors.

### **Acknowledgments**

We thank David Branson, Damien Waits, and Pamela Brannock for help with bioinformatics and data collection. This work was supported by National Science Foundation [DEB-1036537, DEB-1210518, OPP-0338218, ANT-1043745]; and National Aeronautics and Space Administration [NASA- NNX13AJ31G]. Use of the CASIC and Skynet computational resources at Auburn University is acknowledged. Two anonymous reviewers improved a previous version of this manuscript. This is Molette Biology Laboratory contribution #### and Auburn University Marine Biology Program contribution #####.

## References

- Ali AE et al. 1990. Tetrodotoxin and related substances in a ribbon worm *Cephalothrix linearis* (Nemertean). *Toxicon*. 28:1083-1093.
- Altschul SF et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25:3389-3402.
- Asakawa M, Ito K, Kajihara H. 2013. Highly toxic ribbon worm *Cephalothrix simula* containing tetrodotoxin in Hiroshima Bay, Hiroshima Prefecture, Japan. *Toxins*. 5:376-385.
- Ashburner M et al. 2000. Gene ontology: tool for the unification of biology. *Nat Gen*. 25:25-29.
- Bacq ZM. 1937. L' Amphiporine et La Némertine poisons de vers Némertiens. *Archives of Phys and Biochem*. 44:190-204.
- Barathkuma TR, Manivannan J, Thangaraj M. 2013. *In silico* sequence analysis, homology modeling and functional annotation of toxin I hypothetical protein of catfish, *Plotosus lineatus*. *International Journal of Pharm Biol Arch*. 4:753-757.

- Barnham KJ, Dyke TR, Kem WR, Norton RS. 1997. Structure of Neurotoxin B-IV from the marine worm *Cerebratulus lacteus*: a helical hairpin cross-linked by disulphide bonding. *J Mol Biol.* 268:886-902.
- Berne S et al. 2003. Isolation and characterisation of a cytolytic protein from mucus secretions of the Antarctic heteronemertine *Parbolasia corrugatus*. *Toxicon.* 41:483-491.
- Blumenthal KM, Keim PS, Heinrikson RL, Kem WR. 1981. Structure and action of heteronemertine polypeptide toxins. Amino acid sequence of *Cerebratulus lacteus* Toxin B-II and revised structure of Toxin B-IV. *J Biol Chem.* 256:9063-9067.
- Brodie ED. 1968. Investigations on the skin toxin of the adult Rough-Skinned Newt, *Taricha granulosa*. *Copeia.* 2:307-313.
- Brodie ED, Brodie ED. 1990. Predatory-prey arms races. *Bioscience.* 49:557-568.
- Brown T, Howe C, Zhang A, Pyrkosz Q, Brom AB. 2012. A reference-free algorithm for computational normalization of shotgun sequencing data. *ArXiv e-prints*:1203.4802.
- Carroll S, McEvoy EG, Gibson R. 2003. The production of tetrodotoxin-like substances by nemertean worms in conjunction with bacteria. *J Exp Mar Biol Ecol.* 288:51-63.
- Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG. 2013. Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol Ecol.* 28:219-229.
- Castañeda O et al. 1995. Characterization of a potassium channel toxin from the Caribbean sea anemone *Stichodactyla helianthus*. *Toxicon.* 33:603-613.
- Counis MF, Torriglia A. 2000. DNases and apoptosis. *Biochem and Cell Biol.* 78:405-414.
- Dahlgren TG, Weinberg JR, Halanych KM. 2000. Phylogeography of the ocean quahog (*Arctica islandica*): influences of paleoclimate on genetic diversity and species range. *Mar Biol.* 137:487-495.

- Daly JW, Brown GB, Mensah-Dwumah M. 1978. Classification of skin alkaloids from neotropical poison-dart frogs (*dendrobatidae*). *Toxicon*. 16:163-188.
- Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 52: 696-704.
- Dayton PK, Robilliard GA, Paine RT, Dayton LB. 1974. Biological accommodation in the benthic community at McMurdo Sound, Antarctica. *Ecol Monogr*. 44:105-128.
- Eddy SR. 2001. Accelerated profile HMM searches. *PLOS Computational Biology* 7: e1002195.
- Evans CJ, Arguiera RJ. 2003. DNase II: genes, enzymes and function. *Gene*. 322:1-15.
- Ferraris JD. 1979. Histological study of secretory structure of nemerteans subjected to stresses: III cephalic glands. *Gen Comp Endocr*. 39:451-466.
- Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 39:W29-W37.
- Fry BG et al. 2009a. The toxicogenomic multiverse: convergent recruitment of proteins in animal venoms. *Annu Rev Genom Hum G*. 10:483-511.
- Fry BG, Roelants K, Norman JA. 2009b. Tentacles of venom: toxic protein convergence in the kingdom Animalia. *J Mol Evol*. 68:311-321.
- Gabaldón T, Huynen MA. 2004. Prediction of protein function and pathways in the genome era. *Cell Mol Life Sci*. 61:930-944.
- Garber M, Grabherr MG, Guttman M, Trapnell C. 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods*. 6:469-477.
- Gibbs GM, Roelants K, O'Bryan MK. 2008. The CAP superfamily; cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins—roles in reproduction, cancer, and immune defense. *Endocr Rev*. 29:865-897.

- Gibson R. 1983. Antarctic nemerteans: The Anatomy, Distribution and Biology of *Parborlasia corrugatus* (McIntosh, 1876) (Heteronemertea, Lineidae). In: Biology of the Antarctic Seas XIV: American Geophysical Union. p. 289-316.
- Gibson R. 1995. Nemertean genera and species of the world: an annotated checklist of original names and description citations, synonyms, current taxonomic status, habitats and recorded zoogeographic distribution. J Nat Hist. 29:271-562.
- Gordon A. 2011. FASTX-toolkit, computer program distributed by the author.
- Grabherr MG et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol. 29:644-652.
- Haas BJ et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 8:1494-1512.
- Heine JN, McClintock JB, Slattery M, Weston J. 1991. Energetic composition, biomass, and chemical defense in the common Antarctic nemertean *Parborlasia corrugatus* McIntosh. J Exp Mar Biol Ecol. 153:15-26.
- Jost MC et al. 2008. Toxin-resistant sodium channels: parallel adaptive evolution across a complete gene family. Mol Biol and Evol. 31:1016-1024.
- Kaiser et al. 2004. A member of a conserved *Plasmodium* protein family with membrane attack complex/perforin (MACPF)-like domains localizes to the micronemes of sporozoites. Mol Biochem Parasit. 133:15-26.
- Kajihara H, Chernyshev AV, Shi-Chin S, Sundberg P, Crandall FB. 2008. Checklist of nemertean genera and species published between 1995 and 2007. Species Diversity. 13:245-274.

- Kalman K et al. 1998. ShK-Dap 22, a potent Kv1.3-specific immunosuppressive polypeptide. *J Biol Chem.* 273:32697-32707.
- Katoh K, Standley DM 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol and Evol.* 30: 772-780.
- Kem WR. 1971. A study of the occurrence of anabaseine in *Paranemertes* and other nemertines. *Toxicon.* 9:23-32.
- Kem WR. 1976. Purification and characterization of a new family of polypeptide neurotoxins from the heteronemertine *Cerebratulus lacteus* (Leidy). *J Biol Chem.* 251:4184-4192.
- Kem WR. 1994. Structure and membrane actions of a marine worm protein cytolyisin, *Cerebratulus* Toxin A-III. *Toxicology.* 87:189-203.
- Kem WR, Abbott BC, Coates RM. 1971. Isolation and structure of a hoplonemertine toxin. *Toxicon.* 9:15-22.
- Kem WR, Blumethal, KM. 1978. Purification and characterization of the cytolytic *Cerebratulus* A toxins. *J Biol Chem.* 253:5752-5757.
- King GF. 2011. Venoms as a platform for human drugs: translating toxins into therapeutics. *Expert Opin Biol Th.* 11:1469-1484.
- Kvist S, Laumer CE, Junoy J, Giribet G. 2014. New insights into the phylogeny, systematics and DNA barcoding of Nemertea. *Invertebr Syst.* 28:287-308.
- Lee C-Y. 1979. Snake Venoms. In: Born GVR, Farah A, Herken H, Welch AD, editors. *Handbook of Experimental Pharmacology.* Berlin: Springer-Verlag.
- Lewis R, Garcia ML. 2003. Therapeutic potential of venom peptides. *Nat Rev Drug Discov.* 2: 790-802.



- Ling EA. 1970. Further investigations on the structure and function of cephalic organs of a nemertine *Lineus ruber*. *Tissue Cell*. 2:569-588.
- Lu G, Villalba M, Coscia MR, Hoffman DR, King TP. 1993. Sequence analysis and antigenic cross-reactivity of a venom allergen, Antigen 5, from hornets, wasps, and yellow jackets. *J Immunol*. 150:2823-2830.
- McClintock JB, Baker BJ. 1997. A review of chemical ecology of Antarctic marine invertebrates. *Am Zool*. 37:329-342.
- McDermott JJ, Roe P. 1985. Food, feeding behavior and feeding ecology of nemerteans. *Am Zool*. 25:113-125.
- Menezes MC, Furtado MF, Travaglia-Cardoso SR, Camargo ACM, Serrano SMT. 2006. Sex-based individual variation of snake venom proteome among eighteen *Bothrops jararaca* siblings. *Toxicon*. 47: 304-312.
- Meyer E et al. 2009. Sequencing and *de novo* analysis of a coral larval transcriptome using 454 GSFlx. *BMC Genomics*. 10:219.
- Morita-Yamamuro C et al. 2005. The *Arabidopsis* gene *CADI* controls programmed cell death in the plant immune system and encodes a protein containing a MACPF domain. *Plant Cell Physiol*. 46: 902-912
- Morgenstern D et al. 2011. The tale of a resting gland: transcriptome of a replete venom gland from the scorpion *Hottentotta judaicus*. *Toxicon*. 57: 695-703.
- Nagai H et al. 2002. Novel proteinaceous toxins from the nematocyst venom of the Okinawan sea anemone *Phyllodiscus semoni* Kwietniewski. *Biochem Bioph Res Co*. 292:760-763.
- Norenburg JL. 1985. Structure of the nemertine integument with consideration of its ecological and phylogenetic significance. *Amer Zool*. 25: 37-51.

- Olivera BM et al. 1985. Peptide neurotoxins from fish-hunting cone snails. *Science*. 230:1338-1343.
- Oshiro N et al. 2004. A new membrane-attack complex/perforin (MACPF) domain lethal toxin from the nematocyst venom of the Okinawan sea anemone *ActinERIA villosa*. *Toxicon*. 43:225-228.
- Ota E et al. 2006. Caspase-independent apoptosis induced in rat liver cells by Plancitoxin I, the major lethal factory from crown-of-thorns starfish *Acanthaster planci* venom. *Toxicon*. 48:1002-1010.
- Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*. 23:1061-1067.
- Paul VJ, Pennings SC. 1991. Diet-derived chemical defenses in the sea hare *Stylocheilus longicauda* (Quot et Gaimard 1824). *J Exp Mar Biol and Ecol*. 151:227-243.
- Paul VJ, Van Alstyne KL. 1988. Use of ingested algal diterpenoids by *Elysia halimeda* Macnae (Opisthobranchia: Ascoglossa). *J Exp Mar Biol and Ecol*. 119:15-29.
- Punta M et al. 2014. The Pfam protein families database. *Nucleic Acids Res*. 42:D222-D230.
- Riesgo A et al. 2012. Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model organisms. *Front in Zool*. 9:33.
- Roe P, Norenburg JL. 1999. Observations on the depth distribution, diversity and abundance of pelagic nemerteans from the Pacific Ocean off California and Hawaii. *Deep-Sea Res Pt I*. 46:1201-1220.
- Saporito RA, Donnelly MA, Spande TF, Garraffo HM. 2012. A review of chemical ecology in poison frogs. *Chemoecology*. 22:159-168.

- Satoh H, Oshiro N, Namikoshi M, Nagai H. 2007. Characterization of PsTX-60B, a new membrane-attack complex/perforin (MACPF) family toxin, from the venomous sea anemone *Phyllodiscus semoni*. *Toxicon*. 49:1208-1210.
- Shikamoto Y, Suto K, Yamazaki Y, Morita T, Mizuno H. 2005. Crystal structure of a CRISP Ca<sup>2+</sup>-channel blocker derived from snake venom. *J Mol Biol*. 350:735-743.
- Shiomi K, Midorikawa S, Ishida M, Nagashuma Y, Nagai H. 2004. Plancitoxins, lethal factors from the crown-of-thorns starfish *Acanthaster planci*, are deoxyribonucleases II. *Toxicon*. 44:499-506.
- Shiomi K, Yamamoto S, Yamanaka H, Kikuchi T, Konno K. 1990. Liver damage by the crown-of-thorns starfish (*Acanthaster planci*) lethal factor. *Toxicon*. 28:469-475.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30: 1312-1313.
- Stricker SA, Cloney RA. 1981. The stylet apparatus of the nemertean *Paranemertes peregrina*: its ultrastructure and role in prey capture. *Zoomorphology*. 97:205-223.
- Sundberg P, Turbeville JM, Lindh S. 2001. Phylogenetic relationships among higher nemertean (Nemertea) taxa inferred from 18S rDNA sequences. *Mol Phylogenet Evol*. 20:327-334.
- Tamura S, Yamakawa M, Shiomi K. 2011. Purification, characterization and cDNA cloning of two natterin-like toxins from the skin secretion of oriental catfish *Plotosus lineatus*. *Toxicon*. 58:430-438.
- Tanu MB et al. 2004. Immunoenzymatic visualization of tetrodotoxin (TTX) in *Cephalothrix* species (Nemertea: Anopla: Palaeonemertea: Cephalotrichidae) and *Planocera reticulata* (Platyhelminthes: Turbellaria: Polycladida: Planoceridae). *Toxicon*. 44:515-520.

- Tatusov RL et al. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*. 4:41.
- Thiel M, Kruse I. 2001. Status of the Nemertea as predators in marine ecosystems. *Hydrobiologia*. 456:21-32.
- Thollessen M, Norenburg JL. 2003. Ribbon worm relationships: a phylogeny of the phylum Nemertea. *Proc R Soc Lond B*. 270:407-415.
- Tosi, MF. 2005. Innate immune response to infection. *Current reviews of allergy and clinical immunology*. 116:241-249.
- Turk T, Kem WR. 2009. The phylum Cnidaria and investigations of its toxins and venoms until 1990. *Toxicon*. 54:1031-1037.
- Ueda A et al. 2006. Purification, properties and cDNA cloning of neoverrucotoxin (neoVTX), a hemolytic factor from stonefish *Synanceia verrucosa* venom. *Biochim et Biophys Acta*. 1760:1713-1722.
- Ueda A, Nagai H, Ishida M, Nagashima Y, Shiomi K. 2008. Purification and molecular cloning of SE-cephalotoxin, a novel proteinaceous toxin from the posterior salivary gland of cuttlefish *Sepia esculenta*. *Toxicon*. 52:574-581.
- Undheim EAB et al. 2014. Clawing through evolution: toxin diversification and convergence in the ancient lineage Chilopoda (Centipedes). *Mol Biol and Evol*. 31:2124-2148.
- UniProt Consortium. 2014. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res*. 42:D191-D198.
- Upadhyay SK et al. 2013. Selective Kv1.3 channel blocker as therapeutic for obesity and insulin resistance. *P Natl Acad Sci USA*. 110:E2239-E2248.

- von Reumont BM et al. 2014. A polychaete's powerful punch: venom gland transcriptomics of *Glycera* reveals a complex cocktail of toxin homologs. *Genome Biol and Evol.* 6:2406-2423.
- Wang F et al. 2006. Structural and functional analysis of natrin, a venom protein that targets various ion channels. *Biochem Bioph Res Co.* 351:443-448.
- Weigert A et al. 2014. Illuminating the base of the Annelid tree using transcriptomics. *Mol Biol and Evol.* 31:1391-1401.
- Wong ESW, Nicol S, Warren WC, Belov K. 2013. Echidna venom gland transcriptome provides insights into the evolution of monotreme venom. *PLoS One* 8:e79092.
- Xue Z et al. 2012. Characterization, phylogenetic analysis and cDNA cloning of natterin-like gene from the blood of lamprey, *Lampetra japonica*. *Immunol Lett.* 148:1-10.
- Ye J et al. 2006. Wego: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34:W293-W297.
- Young JD, Liu CC, Leong LG, Cohn ZA. 1986. The pore-forming protein (perforin) of cytolytic T lymphocytes is immunologically related to the components of membrane attack complex of complement through cysteine-rich domains. *J Exp Med.* 164:2077-2082.

### Figure Legend

fig. 1: A-G) Pie charts with percentages of the higher-level GO terms for each species except *Cephalothrix hongkongensis* and *Cerebratulus marginatus* (see Riesgo et al. 2012). A) *M. grossa*, B) *P. peregrina*, C) *T. polymorphus*, D) *Cephalothrix linearis*, E) *L. longissimus*, F) *L. ruber*, G) *R. lacteus*. H) Bar chart of genes annotated with select "Biological Process" lower-level GO terms. I) Bar chart of genes annotated with select "Cellular Component" and "Molecular Function" lower-level GO terms. Bar plot colors are referenced in H.

Table 1: Assembly statistics

Species	GenBank Number	Sequencing Platform	Raw reads	N50	GC%	Mean Contig Length	# of Contigs	Longest Contig	% Partial CEGMA Genes	% Complete CEGMA Genes
<i>Iacobdella</i> <i>ssa</i>	SRX731465	HiSeq 2x100	83,045,507	1720	43.8	1047.06	109120	15014	99.19	93.55
<i>ranemertes</i> <i>egrina</i>	SRX731466	HiSeq 2x100	86,271,698	1493	45.6	869.65	99203	25470	93.95	86.69
<i>qualnus</i> <i>ymorphus</i>	SRX732127	HiSeq 2x100	39,262,732	1492	39.3	916.39	79313	13710	98.79	88.31
<i>phalothrix</i> <i>gkongiensis</i>	SRX205320	Genome Analyzer II 2x100	26,112,259	471	36.9	431.00	122233	4383	65.32	16.53
<i>phalothrix</i> <i>earis</i>	SRX524866- SRX534868	Genome Analyzer IIx 2x75	12,062,451	1241	37.4	738.21	67317	14570	97.58	86.69
<i>rebratulus</i> <i>rginatus</i>	SRX205323	Genome Analyzer II 2x100	26,688,280	515	41.6	458.84	194257	8534	83.47	28.23
<i>eus</i> <i>gissimus</i>	SRX565176- SRX565181	HiSeq 1x100	95,576,010	2292	43.0	1186.57	79452	25184	98.79	93.55
<i>eus ruber</i>	SRXX565182- SRX565183	HiSeq 1x100	59,763,351	2105	43.6	1082.25	156014	26007	98.39	92.74
<i>nphogordius</i> <i>teus</i>	SRX565174, SRX565175	HiSeq 1x100	25,452,210	3246	43.1	1539.37	96066	23116	98.79	91.94

Downloaded from <http://gbe.oxfordjournals.org/> by Nathan Whelan on December 9, 2014

Table 2: Annotation statistics

Species	Number (%) of transcripts annotated with BLASTn	Number (%) of transcripts with predicted ORF	Number (%) of ORFs annotated with BLASTp
<i>Malacobdella grossa</i>	33,964 (41.8)	40,421 (37.0)	30,072 (74.4)
<i>Paranemertes peregrina</i>	36,452 (35.7)	40,530 (40.9)	28,460 (70.2)
<i>Tubulanus polymorphus</i>	38,364 (34.3)	51,660 (46.2)	33,904 (65.6)
<i>Cephalothrix hongkongiensis</i>	39,525 (32.2)	37,688 (30.7)	24,329 (64.5)
<i>Cephalothrix linearis</i>	22,328 (32.9)	25,652 (37.8)	18,391 (71.7)
<i>Cerebratulus Marginatus</i>	32,011 (16.4)	35,451 (18.2)	22,357 (63.1)
<i>Lineus longissimus</i>	29,048 (35.4)	35,714 (47.8)	24,852 (70.0)
<i>Lineus ruber</i>	39,770 (24.6)	47,288 (47.2)	33,468 (70.8)
<i>Ramphogordius lacteus</i>	38,771 (38.7)	52,422 (32.4)	33,809 (64.5)

Table 3: Toxin genes identified in each species

Species	Identifier	Trinotate Swiss-Prot Annotation	Functional Annotation and Pfam domains	Putative Orthologs
<i>Malacobdella grossa</i>	M1	A0ZSK4: Neoverrucotoxin subunit beta	SNTX/VTX toxin; Hemolytic activity	P1, Ch2, Cl1, Cm1, Llo1, Lr8, P2, Cl2
	M2	Q3SB03: Cysteine-rich venom protein pseudochetoxin-like	CAP domain; ShK domains; Potassium channel blocker	
	M3	Q75WF2: Plancitoxin-1	Plancitoxin; DNase II domain; Hepatotoxin	P4, T2, Ch3, Cl3, Cm2, Llo2, Lr2, Rl2
<i>Paranemertes peregrina</i>	P1	A0ZSK3: Neoverrucotoxin subunit alpha	SNTX/VTX toxin; Hemolytic activity	M1, Ch2, Cm1, Llo1, Lr8
	P2	Q3SB03: Cysteine-rich venom protein pseudochetoxin-like	CAP domain; ShK domains; Potassium channel blocker	M2, Cl2



<i>Tubulanus polymorphus</i>	P3	Q98993: Verrucotoxin subunit beta	SNTX/VTX toxin; Hemolytic, cytolytic, hypotensive activities; Calcium channel inhibitor	T1, Ch1, Cl1, Cm1, Llo1, Lr8, Rl1
	P4	Q75WF2: Plancitoxin-1	Plancitoxin; DNase II domain; Hepatotoxin	M3, T2, Ch3, Cl3, Cm2, Llo2, Lr2, Rl2
	T1	Q91453: Stonustoxin subunit beta	SNTX/VTX toxin; Pore-forming; Hemolytic, edema inducing activities	P3, Ch1, Cl1, Cm1, Rl1, Llo1
	T2	Q75WF2: Plancitoxin-1	Plancitoxin; DNase II domain; Hepatotoxin	M3, P4, Ch3, Cl3, Cm2, Llo2, Lr2, Rl2
	T3	Q76CA2: Echotoxin-2	Echotoxin; Anemone cytotox domain; Pore-forming; Hemolytic activity	Ch4, Cl4, Cm3, Lr3
<i>Cephalothrix hongkongiensis</i>	T4	Q66S13: Natterin-4	Natterin; Aerolysin domain; Edema and nociception induction	Cm6, Lla6, Lr7
	Ch1	Q91453: Stonustoxin subunit beta	SNTX/VTX toxin; Pore-forming; Hemolytic, edema inducing activities	P3, T1, Cl1, Lr1, Rl1
	Ch2	A0ZSK3: Neoverrucotoxin subunit alpha	SNTX/VTX toxin; Hemolytic activity	M1, P1, Cm1, Llo1, Lr8,
	Ch3	Q75WF2: Plancitoxin-1	Plancitoxin; DNase II domain; Hepatotoxin	M3, P4, T2, Cl3, Cm2, Llo2, Lr2, Rl2
	Ch4	Q76CA2: Echotoxin-2	Echotoxin; Anemone cytotox domain; Pore-forming; Hemolytic activity	T3, Cm3, Lr3
	Ch5	B2DCR8: SE-cephalotoxin	Cephalotoxin; Toxic function unknown	Cl5, Cm4, Lr4, Llo3, Rl3
	Ch6	P58912: Toxin PsTX-60B	MACPF toxin domain; Hemolytic activity	Cl6

<i>Cephalothrix linearis</i>	Ch7	Q76DT2: Toxin AvTX-60A	MACPF toxin domain; Hemolytic activity	Cl7
	Cl1	Q91453: Stonustoxin subunit beta	SNTX/VTX toxin; Pore-forming; Hemolytic, endema inducing activities	P3, T1, Ch1, Lr1, Rl1
	Cl2	Q2XXQ3: Cystein-rich venom protein ENH1	CAP domain; ShK domains; Potassium channel blocker	M2, P2
	Cl3	Q75WF2: Plancitoxin-1	Plancitoxin; DNase II domain; Hepatotoxin	M3, P4, T2, Ch3, Cm2, Llo2, Lr2, Rl2
	Cl4	Q76CA2: Echotoxin-2	Echotoxin; Anemone cytotox domain; Pore-forming; Hemolytic activity	T3, Ch4, Cm3, Lr3
	Cl5	B2DCR8: SE-cephalotoxin	Cephalotoxin; Toxic function unknown	Ch5, Cm4, Llo3, Lr4, Rl3
	Cl6	P58912: Toxin PsTX-60B	MACPF toxin domain; Hemolytic activity	Ch6
<i>Cerebratulus marginatus</i>	Cl7	Q76DT2: Toxin AvTX-60A	MACPF toxin domain; Hemolytic activity	Ch7
	Cm1	A0ZSK3: Neoverrucotoxin subunit alpha	Neoverrucotoxin; Hemolytic activity	P1, M1, Ch2, Llo1, Lr8
	Cm2	Q75WF2: Plancitoxin-1	Plancitoxin; DNase II domain; Hepatotoxin	P4, T2, M3, Ch3, Cl3, Llo2, Lr2, Rl2
	Cm3	Q76CA2: Echotoxin-2	Echotoxin; Pore-forming; Hemolytic activity	T3, Ch4, Cl4, Lr3
	Cm4	B2DCR8: SE-cephalotoxin	Cephalotoxin; Toxic function unknown	Ch5, Cl5, Llo3, Lr4, Rl3
	Cm5	P01527: Cytotoxin A-III	Cytotoxin A-III: Pore-forming; Kinase C, sodium, and calcium channel inhibitor	Ch5, Llo3, Lr4, Rl3

<i>Lineus longissimus</i>	Cm6	N/A: Inferred as Natterin-4 from reciprocal BLAST	Natterin; Aerolysin domain; Edema and nociception induction	T4, Lr7, R15
	Llo1	A0ZSK4: Neoverrucotoxin subunit alpha	SNTX/VTX toxin; Hemolytic activity	M1, P1, Ch2, Cm1, Lr8
	Llo2	Q75WF2: Plancitoxin-1	Plancitoxin; DNase II domain; Hepatotoxin	M3, P4, T2, Ch3, Cl3, Cm2, Lr2, R12
	Llo3	B2DCR8: SE-cephalotoxin	Cephalotoxin: Toxic function unknown	Ch5, Cl5, Cm4, Lr4, R13,
	Llo4	P01527: Cytotoxin A-III	Cytotoxin A-III: Pore-forming; Kinase C, sodium, and calcium channel inhibitor	Cm5, Llo4, Lr5, R14
<i>Lineus ruber</i>	Lr1	Q98989: Stonustoxin	SNTX/VTX toxin; Pore-forming; Hemolytic, endema-inducing activities	P3, T1, Ch1, Cl1, R11
	Lr2	Q75WF2: Plancitoxin-1	Plancitoxin; DNase II domain; Hepatotoxin	M3, P4, T2, Ch3, Cl3, Cm2, Llo2, R12
	Lr3	Q76CA2: Echotoxin-2	Echotoxin; Pore-forming; Hemolytic activity	T3, Ch4, Cl4, Cm3
	Lr4	B2DCR8: SE-cephalotoxin	Cephalotoxin; toxic function unknown	Ch5, Cl5, Cm4, Llo3, R13
	Lr5	P01527: Cytotoxin A-III	Cytotoxin A-III; Pore-forming; Kinase C, sodium, and calcium channel inhibitor	Cl6, Llo4, R14
<i>Ramphogordius lacteus</i>	R11	Q98989: Stonustoxin	SNTX/VTX toxin; Pore-forming; Hemolytic, endema-inducing activities	T1, P3, Ch1, Cl1, Llo1

R12	Q75WF2: Plancitoxin-1	Plancitoxin; DNase II domain; Hepatotoxin	M3, P4, T2, Ch3, Cl3, Cm2, Llo2, Lr2
R13	B2DCR8: SE-cephalotoxin	Cephalotoxin: Toxic function unknown	Ch5, Cl5, Cm4, Llo3, Lr4,
R14	P01527: Cytotoxin A-III	Cytoxoxin A-III: Pore-forming; Kinase C, sodium, and calcium channel inhibitor	Lr4, Cm5, Llo3
R15	Q66S13: Natterin-4	Natterin; Aerolysin domain; edema and nociception induction	T4, Cm6, Lr7

**Figure 1: Gene Ontology (GO) enrichment analysis of differentially expressed genes.**

**Panel A: Distribution of GO terms.** The pie chart shows the distribution of GO terms across three categories: Cellular component (29%), Molecular function (25%), and Biological process (46%).

**Panel B: Distribution of GO terms.** The pie chart shows the distribution of GO terms across three categories: Cellular component (29%), Molecular function (26%), and Biological process (45%).

**Panel C: Distribution of GO terms.** The pie chart shows the distribution of GO terms across three categories: Cellular component (30%), Molecular function (26%), and Biological process (44%).

**Panel D: Distribution of GO terms.** The pie chart shows the distribution of GO terms across three categories: Cellular component (28%), Molecular function (24%), and Biological process (48%).

**Panel E: Distribution of GO terms.** The pie chart shows the distribution of GO terms across three categories: Cellular component (28%), Molecular function (25%), and Biological process (47%).

**Panel F: Distribution of GO terms.** The pie chart shows the distribution of GO terms across three categories: Cellular component (28%), Molecular function (26%), and Biological process (46%).

**Panel G: Distribution of GO terms.** The pie chart shows the distribution of GO terms across three categories: Cellular component (28%), Molecular function (24%), and Biological process (48%).

**Panel H: GO term enrichment across species.** The bar chart shows the percent of genes (log scale) for various GO terms across five species: *C. linearis* (brown), *L. ruber* (teal), *M. grossa* (red), *L. longissimus* (blue), and *R. lacteus* (orange). The GO terms are: biological adhesion, biological regulation, cell killing, cell component biogenesis, developmental process, growth, immune system process, localization, metabolic process, multi-organism process, multicell organism process, response to stimulus, and rhythmic process. The number of genes for each term is listed on the right.

**Panel I: GO term enrichment across species.** The bar chart shows the percent of genes (log scale) for various GO terms across five species: *C. linearis* (brown), *L. ruber* (teal), *M. grossa* (red), *L. longissimus* (blue), and *R. lacteus* (orange). The GO terms are: cytoplasm, endoplasmic reticulum, mitochondrion, nucleus, ribosome, antioxidant, auxiliary transport protein, binding, catalytic, cytoskeletal protein binding, electron carrier, enzyme regulator, hydrolase, kinase, metallochaperone, molecular transducer, nucleic acid binding, structural molecule, toxin binding, and transcription regulator. The number of genes for each term is listed on the right.