

A Polychaete's Powerful Punch: Venom Gland Transcriptomics of *Glycera* Reveals a Complex Cocktail of Toxin Homologs

Björn M. von Reumont^{1,*}, Lahcen I. Campbell^{1,†}, Sandy Richter², Lars Hering³, Dan Sykes⁴, Jörg Hetmank², Ronald A. Jenner^{1,*}, and Christoph Bleidorn^{2,5,*}

¹Department of Life Sciences, The Natural History Museum, London, United Kingdom

²Molecular Evolution and Systematics of Animals, Institute of Biology, University of Leipzig, Germany

³Animal Evolution & Development, Institute of Biology, University of Leipzig, Germany

⁴Imaging and Analysis Centre, The Natural History Museum, London, United Kingdom

⁵German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

*Corresponding author: E-mail: bmvr@arcor.de; r.jenner@nhm.ac.uk; bleidorn@uni-leipzig.de.

†These authors contributed equally to this work.

Accepted: August 26, 2014

Data deposition: Raw data and transcriptome contigs have been deposited at GenBank under the accessions: *Glycera tridactyla* (body tissue); GART00000000; PID PRJNA234022, *Glycera tridactyla*; GARU00000000; PID PRJNA234021, *Glycera fallax*; GARV00000000; PID PRJNA234017, *Glycera dibranchiata*; GASB00000000; PID PRJNA234012.

Abstract

Glycerids are marine annelids commonly known as bloodworms. Bloodworms have an eversible proboscis adorned with jaws connected to venom glands. Bloodworms prey on invertebrates, and it is known that the venom glands produce compounds that can induce toxic effects in animals. Yet, none of these putative toxins has been characterized on a molecular basis. Here we present the transcriptomic profiles of the venom glands of three species of bloodworm, *Glycera dibranchiata*, *Glycera fallax* and *Glycera tridactyla*, as well as the body tissue of *G. tridactyla*. The venom glands express a complex mixture of transcripts coding for putative toxin precursors. These transcripts represent 20 known toxin classes that have been convergently recruited into animal venoms, as well as transcripts potentially coding for *Glycera*-specific toxins. The toxins represent five functional categories: Pore-forming and membrane-disrupting toxins, neurotoxins, protease inhibitors, other enzymes, and CAP domain toxins. Many of the transcripts coding for putative *Glycera* toxins belong to classes that have been widely recruited into venoms, but some are homologs of toxins previously only known from the venoms of scorpaeniform fish and monotremes (stonustoxin-like toxin), turrid gastropods (turriptide-like peptides), and sea anemones (gigantoxin I-like neurotoxin). This complex mixture of toxin homologs suggests that bloodworms employ venom while predating on macroscopic prey, casting doubt on the previously widespread opinion that *G. dibranchiata* is a detritivore. Our results further show that researchers should be aware that different assembly methods, as well as different methods of homology prediction, can influence the transcriptomic profiling of venom glands.

Key words: venomics, venom evolution, Annelida, bloodworms, *Glycera*, ShKT toxin.

Introduction

Complex proteinaceous venoms have evolved convergently in many animal groups, with primary roles in defense, predation, and competition (Fry et al. 2009; Casewell et al. 2013). Individual venoms can have an extraordinary range of physiological effects depending on their specific composition. The diverse bioactivities of venom cocktails do not only play central roles in securing the ecological and evolutionary success of

venomous taxa as different as arthropods, jellyfish, and cone snails, they are also being exploited for important applied uses, such as the development of new drugs and insect-resistant crops (King 2011; King and Hardy 2013). Venomics, the scientific study of venoms, is also yielding insights into general biological questions, such as the evolutionary importance of orphan genes, the evolution of neofunctionalization of duplicate genes, and the evolutionary pressures regulating the

© 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.

optimal use of metabolically expensive resources such as venoms (Khalturin et al. 2009; Chang and Duda 2012; Morgenstern and King 2013).

Rapid technological advances brought by next generation sequencing (NGS) platforms and improvements in mass spectrometry techniques are currently unlocking transcriptomic, genomic, and proteomic resources for venomomics research at an unparalleled rate and in unprecedented depth (Favreau et al. 2006; Terrat et al. 2012; Aird et al. 2013; Francischetti et al. 2013; Vonk et al. 2013; von Reumont et al. 2014). These approaches are generating a host of new insights into the venom composition of poorly studied taxa such as bats, echidnas, and remipede crustaceans, while at the same time providing more detailed data on venom complexity in better studied taxa (Terrat et al. 2012; Francischetti et al. 2013; He et al. 2013; Wong et al. 2013; von Reumont et al. 2014). All these studies facilitate further insights into venomomics' central theme: The convergent recruitment of protein families into venoms (Fry et al. 2009; Casewell et al. 2013).

However, it is important that these technological advances are based on a broad and solid empirical foundation so that taxon- and method-specific peculiarities can be distinguished from general insights into the composition and biology of venoms. In this respect Annelida is a promising group for gaining new insights into the composition, biology, and evolution of venoms.

Annelida currently comprise around 17,000 described species that are classified into more than a hundred families, reflecting the huge morphological diversity of this animal phylum (Zhang 2011). Given the extreme disparity of annelid life styles it may be no surprise that several annelid taxa use noxious substances to defend themselves against predators and parasites, or to secure prey. For instance, fireworms (Amphinomidae) bear irritating calcareous chaetae that upon breaking can release a trimethylammonium compound (complanine) that can cause serious skin inflammation in humans (Nakamura et al. 2008; Borda et al. 2012). Another defensive annelid toxin is a protein called Lysenin that is found in the coelomic fluid of the earthworm *Eisenia fetida* (Sukumwang and Umezawa 2013). *Eisenia fetida* expels its coelomic fluid when attacked, and the fluid is known to be toxic to vertebrates, probably as a result of the presence of Lysenin (Kobayashi et al. 2001). Lysenin is hemolytic and can lyse cells by inserting into cell membranes, an ability which probably also allows it to play a role in innate immunity as it is able to attack the cell membranes of parasites (De Colibus et al. 2012).

In addition to these defensive uses of toxins, two annelid taxa are known to employ toxins for predatory and parasitic purposes. Parasitic leeches express a complex mixture of anti-coagulant polypeptides in their salivary glands to assist in blood feeding and to prevent coagulation of blood inside the animal's crop (Min et al. 2010; Kvist et al. 2014). The glycerids, also known as bloodworms, are the only

annelids known to use a complex venom for overwhelming prey. There are at least 42 described species of glycerids (Böggemann 2002) that are characterized by a uniform morphology. All Glyceridae possess a pharynx equipped with four strong jaws which are connected to venom glands (fig. 1). The jaws are largely composed of a melanin-like network, making them highly resistant to abrasion, and each jaw bears a channel and pores for venom release (Moses et al. 2006).

The biological activity of *Glycera* venom has been described on the protein level for two species, *Glycera tridactyla* (formerly *Glycera convoluta*) and *Glycera dibranchiata* (Michel and Robin 1972; Michel and Keil 1975; Manaranche et al. 1980; Bon et al. 1985; Ungerer 2002), revealing both neurotoxic and enzymatic activities. A severe local inflammatory response has been described for humans bitten by these worms (Klawe and Dickie 1957). Moreover, *Glycera* venom can cause cardiac arrest, progressive paralysis, convulsions, and death when injected into crustaceans (Michel and Keil 1975; Manaranche et al. 1980; Bon et al. 1985; Ungerer 2002).

Several interesting neurotoxins have been identified in the venom cocktail of bloodworms. Meunier et al. (2002) demonstrated that one neurotoxin of *G. tridactyla*, which is called glycerotoxin, acts selectively on $Ca_v2.2$ channels (N-type Ca^{2+} channels) and is able to stimulate an increase in miniature potentials that is fully reversible. Due to its properties, glycerotoxin has been used as a research tool in several recent studies (Schennig et al. 2006; Meunier et al. 2010). At least one additional component with high toxicity to crustaceans was found in the venom of *G. tridactyla*, but it has not yet been studied further (Bon et al. 1985). An initial study of the venom of *G. dibranchiata* revealed the existence of another toxin that differs in its mode of action from glycerotoxin. Intriguingly, this protein is able to form pores in plasma membranes in a manner similar to α -latrotoxin (Kagan et al. 1982), the potent vertebrate-specific neurotoxin from black widow spider venom (Garb and Hayashi 2013). However, all this work has been conducted on the protein level, often on isolated protein fractions. There is no toxin profile available for any polychaete venom on either the proteomic or transcriptomic level, which hinders identifying and characterizing the toxins responsible for envenomation symptoms.

In this article, we investigate glycerid venom composition by providing the first Illumina-based NGS transcriptomic analyses of glycerid venom glands and body tissue obtained from three species of *Glycera*. We perform phylogenetic analyses of individual toxin families including bloodworm homologs to illuminate their evolutionary relationships and possible evolutionary origins. Additionally, we present a methodological investigation of the difficulties of profiling the expression of putative toxin transcripts with transcriptomic techniques. We pay special attention to the effects of using different methods of transcriptome assembly and different strategies to identify putative venom toxins.

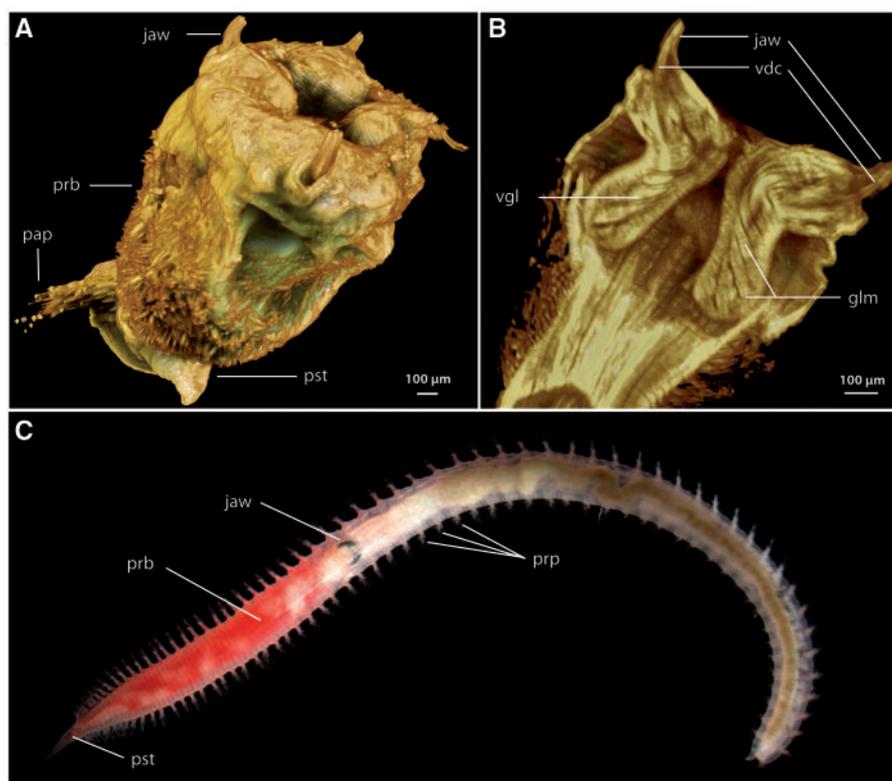


Fig. 1.—Morphology of *Glycera* and its proboscis and venom apparatus. (A) Rendered Micro-CT image of everted proboscis of *Glycera tessellata* showing the four jaws surrounding the terminal mouth. (B) Rendered Micro-CT sagittal section through the proboscis of *G. tessellata* showing the outlines of two venom glands associated with the jaws. (C) General bloodworm morphology illustrated by an unidentified *Glycera* specimen with an inverted proboscis. glm, muscles associated with the venom glands; pap, parapodium; prb, proboscis; pst, prostomium; vgl, venom gland.

Readers should keep in mind that our transcriptomic approach identifies putative toxin precursor transcripts, and that proteomic and functional data need to be collected in order to validate whether the transcripts indeed code for active venom toxins.

Materials and Methods

Specimen Collection and Dissection

Several specimens of *G. tridactyla* Schmarada, 1861 and *Glycera fallax* Quatrefages, 1850 were collected from sandy intertidal flats at Roscoff, France. *Glycera dibranchiata* Ehlers, 1868 specimens were obtained from the Marine Biological Laboratory in Woods Hole, USA. Venom glands were dissected from individuals of all three species. Samples of gland tissue of *G. tridactyla* and *G. fallax* were preserved in RNAlater (Applied Biosystems, Warrington, UK). Additionally, a sample of whole pharynx tissue without venom glands, but composed of epidermal, muscle, nervous system and digestive system tissues was preserved in RNAlater (Applied Biosystems) from *G. tridactyla* to cover also nonvenom gland-related transcripts. Venom glands of one specimen of *G. dibranchiata* were

dissected from fresh material and directly used for total RNA extraction. All RNAlater preserved samples were stored for a few days at 4 °C and transferred then to −80 °C.

RNA Extraction, Library Reconstruction, and Illumina Sequencing

Total RNA was extracted from all samples by homogenization and cell lysis using TRIzol (Invitrogen, Carlsbad, CA). Further sample processing, library preparation and TruSeq RNA sequencing (one lane) of the *G. dibranchiata* sample, as well as quality control and base calling were performed at the GenePool genomics facility of the University of Edinburgh by dedicated technical staff. For the other two species, total RNA was purified using the RNeasy MinElute Cleanup Kit (Qiagen, Hilden, Germany). Afterwards, mRNA was isolated from total RNA using oligo(dT) beads and then fragmented using divalent cationic ions. First and second strand cDNA syntheses were performed by applying SuperScript II, RNase H and DNA polymerase I reactions, respectively, according to the manufacturer's protocols (Invitrogen). The library preparations of *G. tridactyla* and *G. fallax* were performed according to the Illumina multiplex protocol starting with the blunt end repair

described by Meyer and Kircher (2010) with one modification. Instead of the unmodified P7 Illumina adapter mentioned therein, a biotinylated P7 adapter was used to immobilize the libraries by binding to streptavidin beads. The libraries were sequenced (101 bp, paired-end) on the Genome Analyzer Ix platform (Illumina, San Diego) at the Max Planck Institute for Evolutionary Anthropology (Leipzig, Germany).

Sequence Assembly and Processing

The raw sequencing data were processed with IBIS 1.1.2 (Kircher et al. 2009) to split up the reads according to their indices and to clip off adapter sequences. Reads below a certain threshold (quality filter 15: 95% of the nucleotides of the read with a Phred score above 15; quality filter 20: 95% of the nucleotides of the read with a Phred score above 20) were removed by using the program ConDeTri (Smeds and Kunstner 2011). To retain the highest transcript diversity, the most thorough sequence analyses were conducted using filter 15 settings. Subsequently, sequence reads were assembled de novo using IDBA-tran v1.1.1 (Peng et al. 2013) and CLC Genomics Workbench v5.5.x (CLC bio, Aarhus, Denmark). IDBA-tran assemblies were performed using an initial kmer size of 20, an iteration size of 5, and a maximum *k*-mer size of 60. The option to keep more than one putative isoform per transcript was switched off. CLC assemblies were conducted using the following parameters: mismatch cost = 3, insertion cost = 3, deletion cost = 3, length fraction = 0.5, similarity fraction = 0.8, minimum contig length = 200, automatic bubble size = yes, automatic word size = yes, perform scaffolding = yes. N50 weighted median statistics were calculated for all assemblies (Earl et al. 2011), see [supplementary figures S6 and S7, Supplementary Material](#) online. Comparing the assemblies, IDBA-tran assemblies were chosen as the basis for our analyses, based on both the generally longer average and maximum contig lengths.

Identification and Annotation of Putative Venom Protein Contigs

To identify transcripts for candidate venom proteins, we used an updated customized bioinformatic pipeline as described in von Reumont et al. (2014). This updated bioinformatic pipeline was implemented in Perl, Python, and shell scripts (scripts and pipeline are provided in the [supplementary material, Supplementary Material](#) online) and was used to integrate sequence translation, Basic Local Alignment Search Tool (BLAST) analyses, duplicate BLAST hit identification and sequence retrieval. Secreted proteins (UniProt, location_sl_0243) were used as baits and searched against every 6-frame translated transcriptome by using local BLAST (*e* value $\leq 1e^{-5}$). The resulting sequence contig-files were subsequently loaded into BLAST2GO (Conesa and Götz 2008) to obtain BLAST annotation, GO-term annotation, and InterPro ID's for all of the contigs (see [supplementary tables S3–S6, Supplementary Material](#)

online). Signal peptides for all sequences were predicted separately from InterPro's signal peptide prediction feature implemented in Geneious R6 (Biomatters, <http://www.geneious.com/>, last accessed September 10, 2014) as well as using SignalP 4.1 (applying sensitivity setting of 3.0; see Petersen et al. 2011). Disulfide bridge patterns were predicted with the DBCP web server (Lin and Tseng 2010).

Contigs of all libraries were analyzed for the presence of homologs of venom proteins in a 2-fold strategy. First, matches for venom protein classes identified by InterProScan in BLAST2GO were extracted utilizing a Perl script. Second, HMMER 3 (<http://hmmer.org/>, last accessed September 10, 2014) was used to train hidden Markov models (HMMs) using `-hmmbuild` on alignments of known venom protein classes, which included annotated UniProt sequences of all non-*Glycera* sequences (both venomous and nonvenomous species) that were included in our trees. The alignments used to train the HMMER searches can be found in the [supplementary material, Supplementary Material](#) online. HMMER searches were then performed using `-hmmsearch` (default parameters with *e* value threshold of $10e^{-4}$) to identify venom protein classes in our contig-files. Two different strategies were applied to build HMMER profiles for the final HMMER search: 1) Complete sequence alignments and 2) only specific domain regions of the particular proteins were used. Domains, chains, and signalpeptides were identified by the InterProScan function of Geneious R6 and additionally checked by the UniProt annotation. Upon completion of each HMMER analysis, hit sequences identified by the HMMER for both complete and domain profiles were extracted from the respective contig libraries using a Perl script. Transcripts with sequences coding for signal peptides, but for which neither InterProScan nor HMMER provided any annotation may represent novel venom toxins unique for the *Glycera* lineage (see [supplementary table S2, Supplementary Material](#) online).

It should be noted that we did not attempt to distinguish between allelic variation and paralogy as causes of the transcript diversity we report here. Our primary interest is to estimate the diversity of putative toxin precursor transcripts as this is likely to be informative about the diversity of toxin proteins in crude venom.

Phylogenetic Reconstruction of Putative Venom Protein Transcripts

Prior to generating multiple sequence alignments of venom proteins, redundant contig sequences identified by the three independent search strategies (InterProScan, HMMER full alignment, HMMER Domain) were removed. Redundant contigs were removed using an in-house coded Perl script, that according to contig fasta headers, collated identical sequence contigs, and summarized those identical contigs under a single fasta header comprised the search method(s) that identified a

particular contig. Sequences were aligned using MAFFT L-INS-i (Kato and Standley 2013), including UniProt reviewed manually curated venom protein structure constraints (Jungo et al. 2012), if available. Using Geneious, contigs were annotated and sequences in which stop-codons interrupted the open reading frames (potentially representing sequencing artifacts or pseudogenes), or sequence fractions with stop codons located external to the venom domain of interest were removed from downstream alignments. Note that based on available sequence annotations divergent N-terminal and C-terminal regions were clipped from the alignments. This has in some cases resulted in the inclusion of seemingly identical sequences in several of our trees as in those cases distinct and unique sequences were represented only by their conserved domain regions. After choosing the best fitting model for each protein with ProtTest 3 (Darriba et al. 2011), a maximum-likelihood analysis was conducted with raxmlHPC-PTHREADS-SSE3 (Stamatakis and Alachiotis 2010) (-f a, 1,000 bootstrap pseudoreplicates, see figure legends of reconstructed trees and [supplementary table S7, Supplementary Material](#) online, for the chosen evolutionary model for each protein). Being most interested in the relationships of the *Glycera* sequences and their nearest relatives, we rooted all but two trees with a sequence from a nonvenomous vertebrate, one tree with a cnidarian ([supplementary fig. S14, Supplementary Material](#) online) and one tree with a tunicate ([supplementary fig. S15, Supplementary Material](#) online). It should be noted that although our chosen rootings are not ad hoc, they are provisional, and may need to be revised as more taxa are sampled and a more complete sampling of paralogs becomes available. All alignments used for tree reconstructions of putative venom proteins are provided in the [supplementary material, Supplementary Material](#) online.

Additional sequences from venomous and nonvenomous taxa were predominantly obtained from UniProt in order to maximize the number of annotated sequences in our analyses. We strove to broadly represent the phylogenetic breadth of Metazoa, including nonbilaterians, deuterostomes, ecdysozoans, and especially lophotrochozoans, including both venomous and nonvenomous taxa to represent transcripts coding for putative toxins and nontoxin homologs.

Results and Discussion

Diversity and Molecular Evolution of *Glycera* Venom Toxin Homologs

The transcriptomes of bloodworm venom glands reveal an unexpectedly complex cocktail of transcripts coding for putative venom protein precursors (fig. 2). The most deeply sequenced library (*G. dibranchiata*) expresses the greatest diversity of putative venom toxin transcripts, representing 20 toxin classes that have been convergently recruited into animal venoms, as well as 12 putative toxins that are possibly

unique for bloodworms (see fig. 2 and [supplementary tables S1 and S2, Supplementary Material](#) online). For convenience the identifiable putative *Glycera* toxins are classified into five functional categories: 1) Pore-forming and membrane-disrupting toxins: Actinoporin-like toxin, stonustoxin (SNTX)-like toxin, and sphingomyelinase; 2) neurotoxins: ShKT domain neurotoxin, gigantoxin-like neurotoxin, and turriptide-like neurotoxin; 3) protease inhibitors: Cystatin, Kazal domain protease inhibitor, Kunitz domain protease inhibitor, lipocalin, and serpin; 4) other enzymes: C-type lectin, chitinase, hyaluronidase, phospholipases, peptidase S1, peptidase S10, and metalloproteinase M12; and 5) CAP domain proteins.

The five most abundant toxin homologs transcribed in the venom glands of *G. dibranchiata* are peptidase S1, CAP, ShKT domain neurotoxin, metalloproteinase M12, and Kazal domain protease inhibitor, together representing 80% of putative toxin sequence reads. The venom gland libraries of *G. fallax* and *G. tridactyla* are sequenced at a shallower depth than that of *G. dibranchiata*, and consequently we detected a lower diversity of expressed toxin homologs in these species. The relative abundance of toxin homologs also differs between the species. Although peptidase S1 and Kazal domain protease inhibitors are among the most abundantly expressed toxins homologs in the venom glands of all species, the second most abundantly expressed type of toxin homolog in *G. fallax* is lipocalin, whereas in *G. tridactyla* the second most abundantly expressed toxin category represents a diversity of potentially lineage-specific toxin precursor transcripts without known homologs in other taxa. Potentially novel toxin precursor transcripts are also expressed in the venom glands of the other two species, albeit representing smaller percentages of the total number of toxin transcripts ([supplementary fig. S1, Supplementary Material](#) online). In the following, we discuss a selection of the transcripts coding for putative *Glycera* toxins from the major categories defined above. Further discussion is provided in the [supplementary material, Supplementary Material](#) online.

Pore-Forming and Membrane-Disrupting Toxins: Actinoporin-Like Toxin, SNTX-Like Toxin, and Sphingomyelinase

The intense pain and rapid and prolonged swelling and redness that can result from bloodworm envenomation (Klawe and Dickie 1957) may be mediated by a mix of putative pore-forming and membrane-disrupting toxins expressed in the venom glands of *G. dibranchiata*. These potential toxins may similarly account for the results of activity assays of *G. dibranchiata* venom (Kagan et al. 1982), which have shown that bloodworm venom can induce ion-permeable pores in lipid bilayers. The glands express a diversity of transcripts coding for precursors of toxins that are known to be able to cause cytolysis and hemolysis, and to induce severe pain. Three transcripts are homologous to actinoporins, which are highly

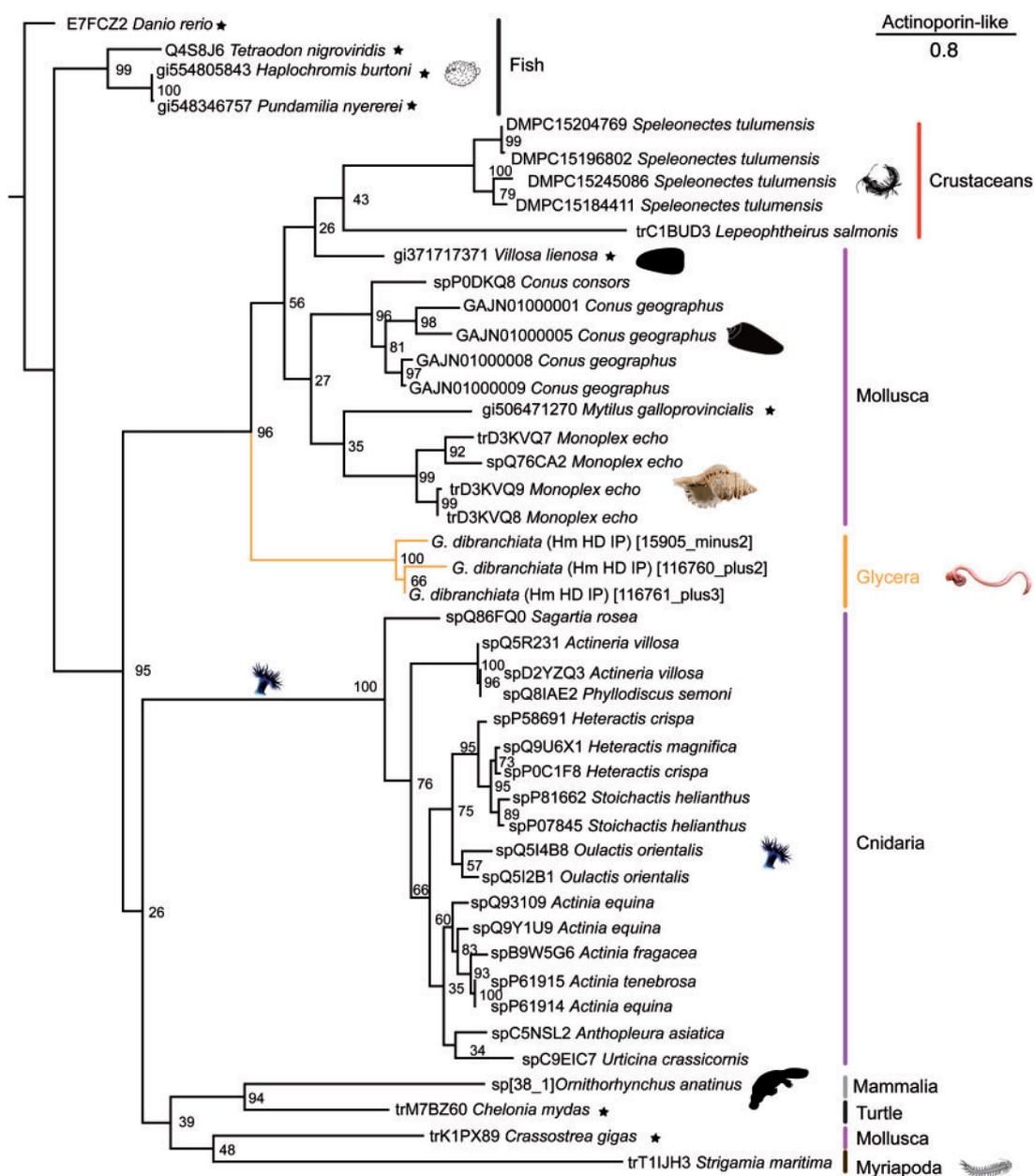


FIG. 3.—Phylogenetic tree of actinoporin-like sequences. Phylogenetic reconstruction was performed with RAxML (7.4.2 SSE3-PTHREADS) (Stamatakis and Alachiotis 2010), -f a, PROTGAMMAIWAG, 1,000 bootstrap pseudoreplicates. Bootstrap support values given for all nodes. Clade names are indicated by colored vertical bars, common names are given for clades with disputed monophyletic status. *Glycera* and polychaete transcripts are highlighted in orange. Sequences obtained from UniProt are denoted by “tr” or “sp” prefixes. Search strategies that identified a given *Glycera* sequence are labeled behind the species names as follows: Hm, HMMER; HD, HMMERDomain; IP, Interpro. The tree is rooted with a nonvenomous taxon (indicated by a star).

sphingomyelinase B has recently been detected in the venom of a hydrozoan jellyfish (Aларcon-Chaidez et al. 2009; Alves-Silva et al. 2010; Weston et al. 2013).

The third and most highly expressed pore-forming toxin homolog in *Glycera* venom glands is represented by SNTX-like transcripts. SNTXs and the related verrucotoxins, neoverrucotoxins, and trachynilysin are nonenzymatic cytolytic toxins

that have been identified in the venom of a diversity of scorpaeniform fish (Low et al. 1994; Chen et al. 1997; Ueda et al. 2006; Kiriake et al. 2013). Homologs have recently also been found to be expressed in the venoms of monotremes (Whittington et al. 2010; Wong et al. 2013). The fish SNTX can cause lethal hemolysis through the formation of pores in the cell membrane, as well as a variety of other effects that

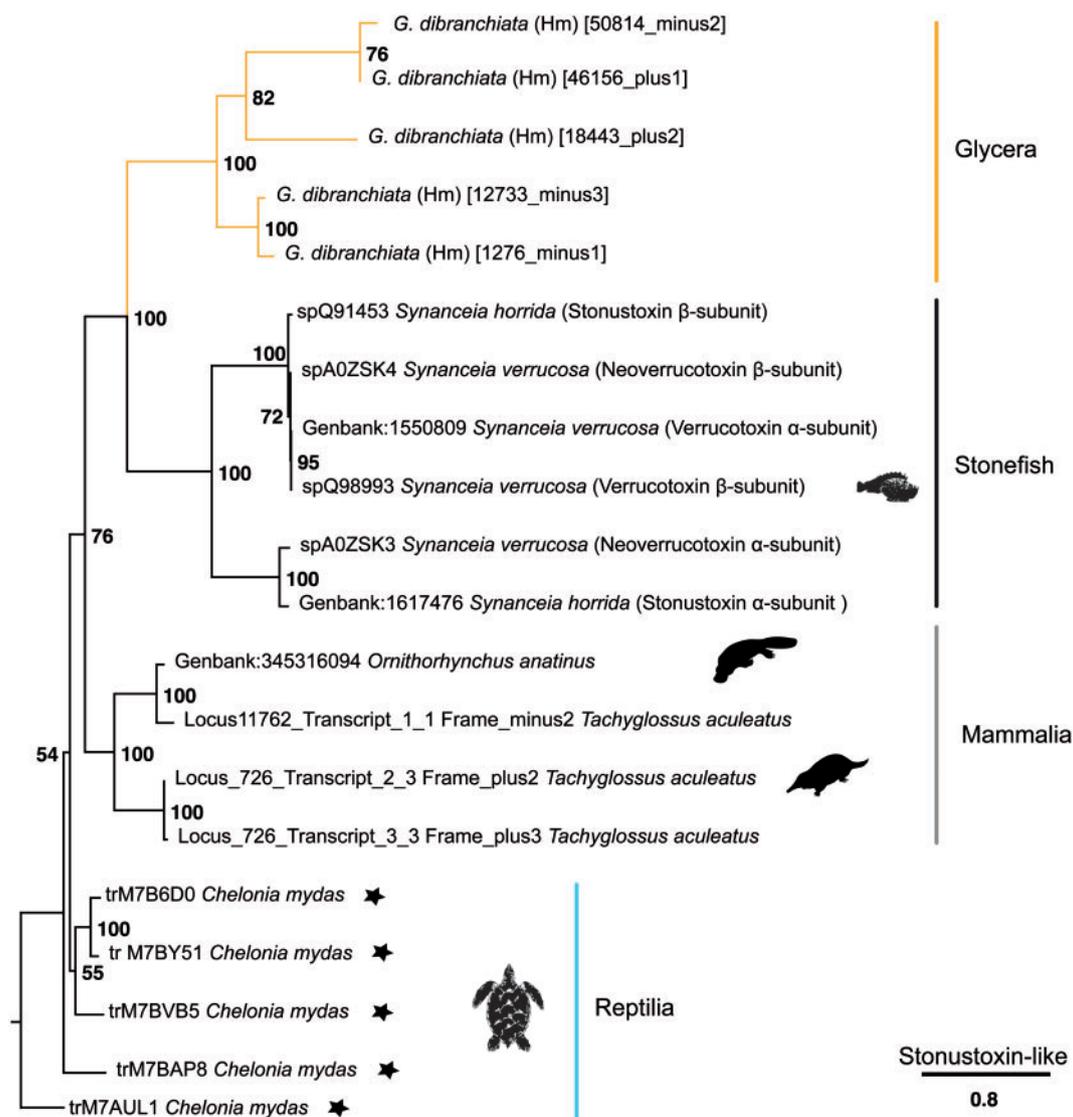


Fig. 5.—Phylogenetic tree of SNTX-like sequences. Phylogenetic reconstruction was performed with RAxML (7.4.2 SSE3-PHREADS) (Stamatakis and Alachiotis 2010), -f a, PROTGAMMAIJTTF, 1,000 bootstrap pseudoreplicates. Bootstrap support values given for all nodes. See the legend of figure 3 for further information.

1980; Bon et al. 1985; Meunier et al. 2002, 2010). These effects include cardiac arrest, progressive paralysis, and sudden convulsions followed by death in crustaceans, which are known bloodworm prey (Fauchald and Jumars 1979). Our results suggest that *Glycera* venom glands express precursors of different putative neurotoxins that could explain these observations.

The venom glands of both *G. dibranchiata* and *G. tridactyla* (as well as the body tissue of *G. tridactyla*) express 13 different transcripts containing ShKT domains. ShKT domains code for two peptide toxins, ShK and BgK, originally described from sea anemones, which can block voltage-gated potassium channels (Castaneda et al. 1995; Castaneda and Harvey

2009). ShK and BgK peptides are, respectively, 35 and 37 amino acids long, and they contain six conserved cysteine residues that form three disulfide bridges that play important roles in facilitating their ion channel blocking activities (Pennington et al. 1999). ShKT domains have been incorporated into a wide diversity of animal proteins, including venom toxins, which also contain a variety of other domains (Rangaraju et al. 2010).

The *Glycera* ShKT domain transcripts contain the six conserved cysteine residues that are characteristic for this domain (fig. 6). The transcripts show three basic domain arrangements: Metalloproteinase M12 + ShKT, SUEL-like Lectin + ShKT, and CAP + ShKT (see [supplementary fig. S2](#),

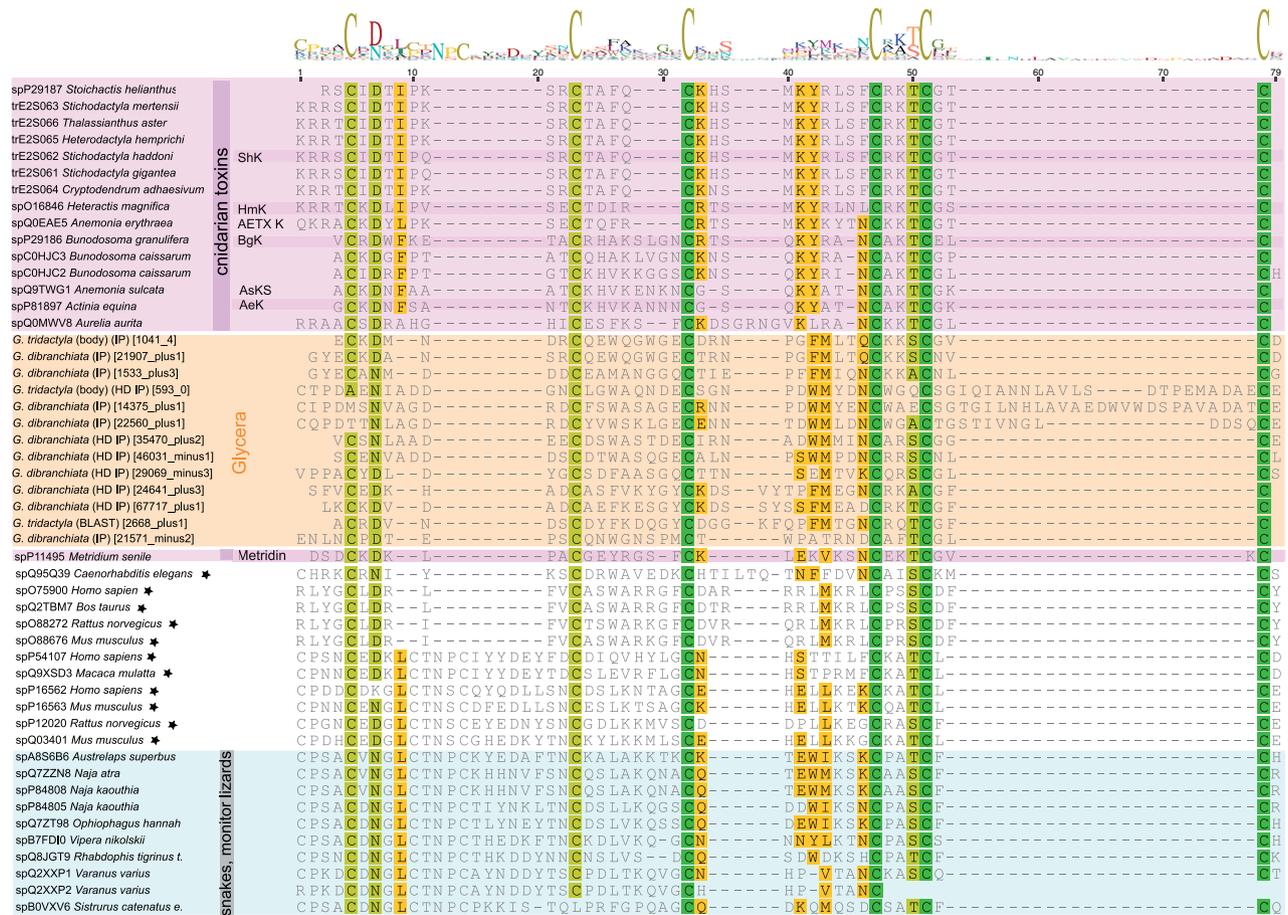


Fig. 6.—Multiple sequence alignment of ShKT toxin domains generated by MAFFT-L-INS-i (Kato and Standley 2013). Conserved residues are highlighted. Cnidarian sequences are highlighted in purple, *Glycera* sequences in orange, squamates in blue, and nonvenomous taxa are indicated by stars. See [supplementary figure S2, Supplementary Material](#) online, for an overview of the different domain arrangement patterns found in ShKT transcripts.

[Supplementary Material](#) online). However, within and between these three basic types the transcripts vary widely in the presence of putative cleavage sites, signal peptides, transmembrane regions, tandem repeats of ShKT domain variants, and the presence of additional domain types.

We performed a phylogenetic analysis on an alignment of the ShKT domains present in our sequences and those in a range of outgroups. This produced a tree with two distinct, but poorly supported, clades (fig. 7). Nine *Glycera* sequences group together with the sequences from other bilaterians, whereas four sequences group together exclusively with cnidarian sequences. Each of these two clades contains a transcript that together are the two most highly expressed putative toxin transcripts in the venom gland library of *G. dibranchiata*, but apart from their shared ShKT domains they have a very different overall organization. The transcript that groups with cnidarian sequences (21571_minus2) has a CAP domain upstream of the ShKT domain, whereas the other (14375_plus1) is preceded by two lectin domains, and

also has a string of ShKT-like domains ([supplementary fig. S2, Supplementary Material](#) online).

Given these very different structures it is nearly impossible to predict the effects of these putative toxins. Most of the cnidarian peptides are known to be potassium channel blockers, with the exception of metridin. Metridin does not have ion channel modulating activities, but instead has a hemolytic effect (Krebs and Habermehl 1987). This may be correlated with the absence of the lysine residue (located in position 41 in the alignment of fig. 6) that is present in the other cnidarian sequences, and which is needed for blocking the pore of the potassium ion channel (Rangaraju et al. 2010). Substitution of this lysine in ShKT peptides into other amino acids diminishes or abolishes potassium channel blocking activity. None of the bloodworm sequences have this lysine residue, but this does not necessarily imply that they cannot function as neurotoxins. Most of the snake ShKT domains also lack this lysine residue, yet many of them have the ability to block potassium or calcium channels. This is probably related to the fact that as in the

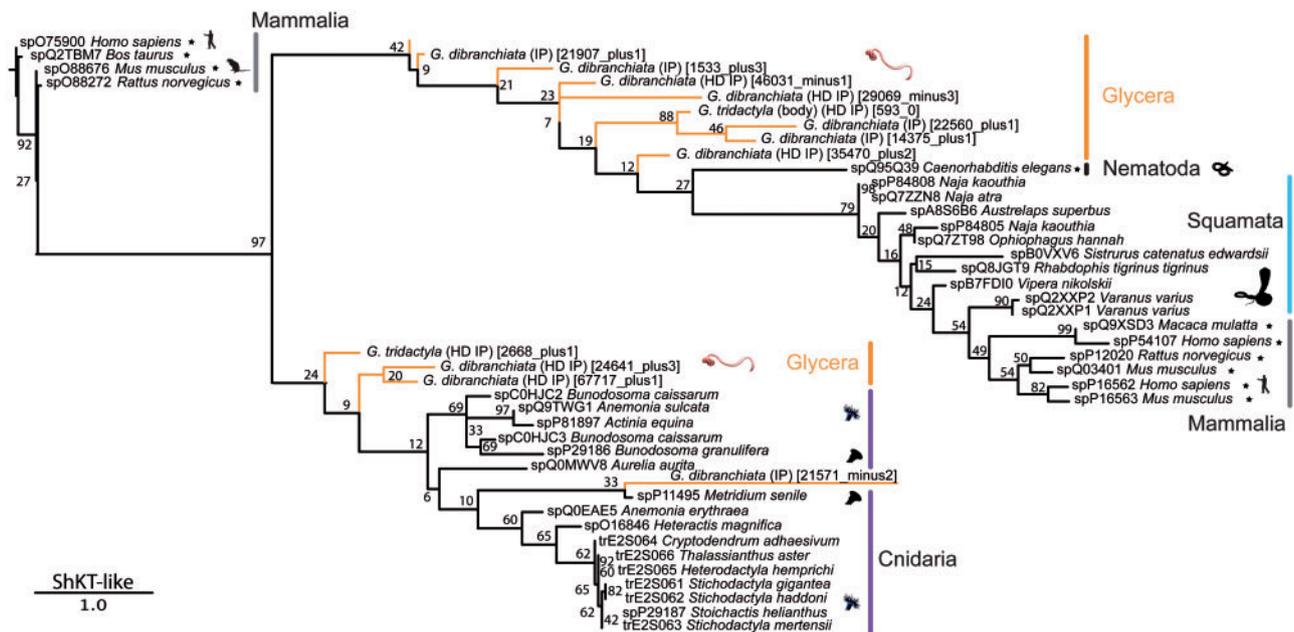


Fig. 7.—Phylogenetic tree of ShKT domain sequences. Phylogenetic reconstruction was performed with RAxML (7.4.2 SSE3-PTHREADS) (Stamatakis and Alachiotis 2010), -f a, PROTGAMMAIVTf, 1,000 bootstrap pseudoreplicates. Bootstrap support values given for all nodes. See the legend of figure 3 for further information.

bloodworms, snake ShKT domains are part of larger proteins that often contain an N-terminal CAP domain (supplementary fig. S2, Supplementary Material online), just as one of the two most highly expressed *Glycera* ShKT domain proteins.

Another putative *Glycera* neurotoxin is represented by transcripts that show strong sequence similarity to gigantoxin I, a neurotoxic peptide first found to be expressed in the nematocysts of the giant sea anemone *Stichodactyla gigantea* (Honma et al. 2003; Shiomi et al. 2003). Gigantoxin I is potently paralytic to crustaceans, but its epidermal growth factor (EGF)-like domain also gives it EGF activity. The presence of EGF domains in many transcripts expressed in the venom glands of all three investigated species of *Glycera* is probably the reason why significant BLAST matches (e value ≤ 0.00001) for gigantoxin I are found in all of our libraries. Most of these transcripts, however, code for large proteins. In contrast, two transcripts expressed in the venom glands of *G. dibranchiata* code for peptides of about the same length as gigantoxin I (117671_minus1 and 101222_plus1 are 88 and 89 amino acids long, respectively) (fig. 8). These peptides are also predicted to have signal peptides and propeptides with remarkable similarity to the gigantoxin I-like neurotoxins found in sea anemones, and the annelid and cnidarian peptide precursors also have an identical cysteine scaffold. Interestingly, transcript 117671_minus1 is predicted to form the same three disulfide bridges as the cnidarian toxins, whereas transcript 101222_plus1 is predicted to form two of these. The homology of the gigantoxin I-like peptide sequences in cnidarians

and *Glycera* is limited to the EGF domains of longer proteins. Because of the very short and conserved alignment we did not perform a phylogenetic analysis of these sequences. However, we did explore whether the EGF-like domains of the gigantoxin I-like sequences of *Glycera* are more similar to the EGF domains of proteins from other bilaterians (reflecting the phylogenetic position of *Glycera*) or whether they are more similar to those of cnidarians (possibly reflecting changes associated with being expressed as peptides rather than larger proteins). To do this, we constructed a neighbor-joining network with Splitstree. Neighbor-joining networks graphically summarize the presence of conflicting signals in aligned sequence data. The results (supplementary fig. S3, Supplementary Material online) show that the sea anemone peptides and bloodworm homologs are indeed similar, and that the two *Glycera* peptide transcripts group more closely with the cnidarian peptide sequences than with the *Glycera* transcripts that code for larger proteins.

Cnidarian gigantoxin I-like peptides are able to elicit acute pain by indirect activation of the TRPV1 channel, a cation channel also known as the capsaicin receptor because it is targeted by the active component of chilli peppers. The toxin peptides do this by binding to EGF receptors with their EGF domain, which results in the activation of phospholipase A₂ and the production of metabolites that activate the TRPV1 channel (Cuypers et al. 2011). We expect that their similar structure, including the presence of an EGF domain, allows the two putative *Glycera* peptides to exert a similar effect,

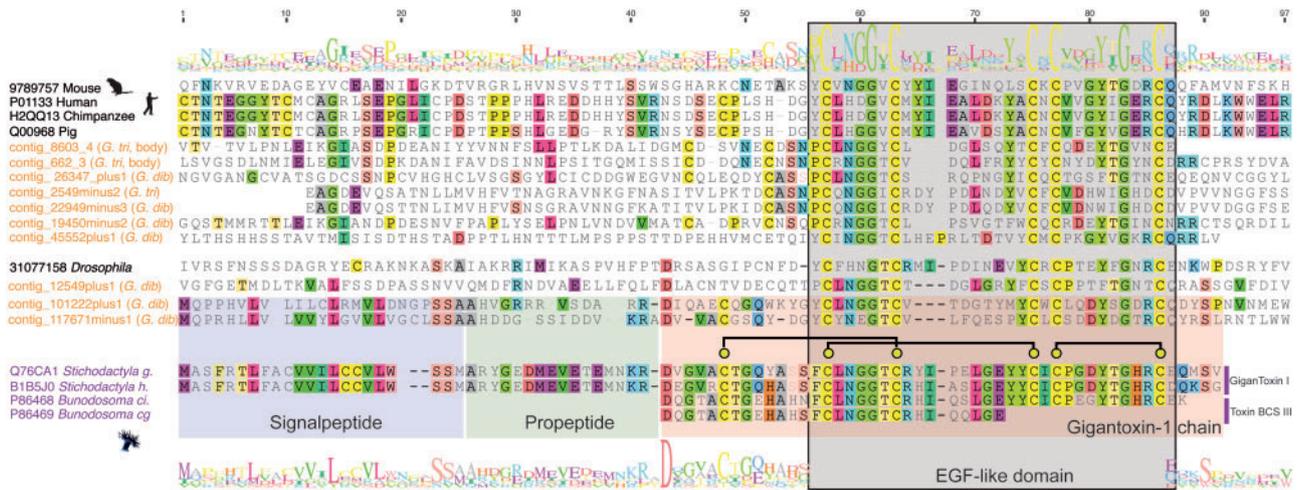


Fig. 8.—Multiple sequence alignment of gigantoxin domain sequences generated by MAFFT-L-INS-i (Katoh and Standley 2013). The signal peptide, propeptide, and domain regions of the cnidarian sequences and the two *Glycera* peptides that most strongly resemble them are highlighted in colored boxes. *Glycera* contig 117671 minus1 has a leucine residue in position 91 of the alignment, which is thought to be involved in mediating the high-affinity binding of the EGF domain to the TRPV1 receptor, triggering the associated pain pathway.

which may cause the sharp pain induced by bloodworm bites (Klawe and Dickie 1957). We expect that one of the *Glycera* peptides will be an especially potent activator of the TRPV1 pain pathway as it possesses a leucine residue in alignment position 91 (fig. 8) that is essential for high-affinity binding to the EGF receptor, and which the cnidarian toxins lack (Shiomi et al. 2003).

Glycera venom glands also express transcripts coding for the precursors of a third putative type of neurotoxin that shows high similarity to the turrpeptides of turrid gastropods. These peptides are dominated by a Kazal domain and we therefore discuss them in the section on Kazal protease inhibitors in the [supplementary material, Supplementary Material online](#) (see also [supplementary fig. S12, Supplementary Material online](#)).

There is no indication that any of the candidate neurotoxins identified here represents glycerotoxin, the calcium channel activating neurotoxin known from the venom of *G. tridactyla* (Bon et al. 1985; Meunier et al. 2002, 2010). Glycerotoxin is expected to be much larger than the putative neurotoxins identified here. Identifying and characterizing glycerotoxin in future studies will require our approach to be complemented by protein sequencing of purified fractions of *Glycera* venom.

Enzymes: Phospholipase A₂ and Phospholipase B

The venom glands of *G. dibranchiata* express PLA₂ and phospholipase B (PLB) transcripts, which is consistent with earlier biochemical studies that have shown phospholipase activity of *G. tridactyla* venom (Bon et al. 1985). The expression of phospholipases might, among other things,

enhance the neurotoxicity of the *Glycera* venom. PLA₂ is an enzyme that has been widely recruited into animal venoms, from cnidarians to cone snails, snakes, and arthropods (Fry et al. 2009; Terrat et al. 2012). PLA₂ catalyzes the hydrolysis of phospholipids, and as a result its venom effects are varied, including cytotoxicity, myotoxicity, neurotoxicity, antiplatelet activity, and inflammation. *Glycera dibranchiata* venom glands express three different PLA₂ transcripts at a low abundance, whereas the body tissues of *G. tridactyla* express two transcripts (of unequal length and only differing by a single amino acid). All these sequences have the two typical active sites (histidine and aspartic acid) for PLA₂. The clade formed of two of the *G. dibranchiata* sequences and the two *G. tridactyla* sequences represents group XII secreted PLA₂ ([supplementary fig. S4, Supplementary Material online](#)).

We also found two transcripts of PLB expressed in the *G. dibranchiata* gland library. Although PLA₂ has been frequently recruited into venom cocktails, this is more rarely the case for PLB. As far as we know PLB activity has been reported from some hymenopteran venoms (Rosenberg et al. 1977; Watala and Kowalczyk 1990), and PLB transcripts have only been found to be expressed in the venom glands of several species of snakes (Rokyta et al. 2011; Fry et al. 2012; Margres et al. 2013).

PLA₂ is known to be responsible for neurotoxic activity at the presynaptic membrane, where it hydrolyzes phospholipids into fatty acids and lysophospholipids (Rigoni et al. 2005; Zimmerberg and Chernomordik 2005). These products cannot form a lipid bilayer. Fatty acids are able to change the conformation of the membrane into a hemifusion state, thereby reducing the threshold for synaptic vesicle release. PLB

does not generate lysophospholipids, but instead it hydrolyzes twice the amount of fatty acids compared with PLA₂, which might help to increase the stimulatory effect of glycerotoxin at the presynaptic membrane.

Bloodworm Feeding Biology and the Unexpected Diversity of Expressed Toxin Homologs

In addition to the components discussed above, bloodworm venom glands express transcripts coding for another dozen different putative toxins, as discussed in the [supplementary material, Supplementary Material](#) online. Bloodworm venom is clearly not dominated by toxin classes that are more or less variations on a single theme, such as the neurotoxic peptides of cone snails and their relatives (Olivera et al. 2012), which represent the nearest venomous outgroups of bloodworms outside Annelida. The different components of *Glycera* venom are most likely involved in different aspects of prey capture and processing, although conclusions can only be tentative until more comprehensive functional assays have been performed.

Based on the results of available experiments that assessed the effects of bloodworm venom on crustacean prey, and supplemented by the known effects of homologs of the putative bloodworm toxins in other taxa, we speculate that the various pore-forming toxins and neurotoxin transcripts identified in *Glycera* venom glands code for proteins that can rapidly paralyze or kill invertebrate prey, especially crustaceans. The various highly expressed CAP domain proteins might contribute to overwhelming the prey. Enzymes such as the phospholipases probably enhance the neurotoxicity of the venom, whereas others, such as the various peptidases, and especially hyaluronidase, may act as spreading factors. Enzymes, including the various proteases and chitinase, may act predominantly as aides in the digestion of prey. Although bloodworms can inflict painful bites in defense, we do not know how effective these are in deterring predation by their natural predators, such as fish and birds.

The toxin category that remains most mysterious is that of the protease inhibitors. The homologs of five different types of protease inhibitors expressed in bloodworm venom glands are particularly prominent in the oral secretions of blood-feeders, such as ticks, mosquitos, leeches, and vampire bats (Anatriello et al. 2010; Chagas et al. 2013; Francischetti et al. 2013; Low et al. 2013; Kvist et al. 2014). But despite the fact that these protease inhibitors may have antihemostatic effects, no evidence exists suggesting that bloodworms are blood feeders.

It is possible that the observed differences in the complement of putative toxin transcripts expressed by the three species of bloodworm investigated here reflect different dietary specializations, but not enough is known about the prey preferences of the different species of *Glycera* to address this issue at this time. Our results do contradict, however, the long held belief that *G. dibranchiata* is chiefly a detritivore (Klawe and

Dickie 1957; Fauchald and Jumars 1979; Rouse and Pleijel 2001). The anatomy of the venom apparatus and the diverse mix of toxin homologs transcribed in the venom glands of *G. dibranchiata* leave little doubt that it is a capable predator of macroscopic prey that likely uses potent venom to catch its prey. This conclusion is supported by the presence of the remains of other polychaetes, mollusks, and crustaceans in the guts and feces of this species (Ambrose 1984; Ungerer 2002), as well as by predation experiments in which it consumed enteropneusts (Giray and King 1997).

Two other factors need to be kept in mind as well when considering the issue of differences in the toxin mix of different species. First, the library of *G. dibranchiata* was sequenced more deeply than those of the other two species, which may have led to an underestimate of toxin homolog diversity in these species. Second, the venom gland libraries for all three species were based on tissue from single individuals, and they therefore only provide a snapshot of possible toxin variation. It is known that substantial intraspecific variation in the toxin composition of venom can result from differences in the physiological state of the venom glands (active or resting), as well as sex differences, ontogenetic differences, and geographical differences (Abdel-Rahman et al. 2009; Morgenstern et al. 2011; Safavi-Hemami et al. 2011; Sunagar et al. 2014). Consequently, to achieve a more detailed understanding of species differences in toxin composition future studies need to take these different factors into account.

Interestingly, the venom glands and body tissue of *G. tridactyla* express transcripts coding for the same convergently recruited types of putative toxins. Indeed, as indicated in [supplementary table S8, Supplementary Material](#) online, several *G. tridactyla* transcripts expressed in both the venom gland and body tissue appear to be identical. However, *G. tridactyla* venom glands express several unique putative toxin transcripts ([supplementary table S2, Supplementary Material](#) online) not expressed in the body tissue, and which do not correspond to any known convergently recruited toxin types. In fact, this class of potentially novel toxins is the most highly expressed class of putative toxins in the venom glands of *G. tridactyla*. Interestingly, the widespread expression of putative toxins in nonvenom gland tissues has recently also been found in snakes (Hargreaves et al. 2014), and has been taken to suggest that the expression of venom toxins may become restricted to the venom glands after gene duplication and subfunctionalization events.

Identifying Toxin Homologs: Comparison of Transcriptome Assembly Methods

All transcriptome-based studies rely on the assembly of sequencing reads, and different kinds of assembly software have been published for both 454 and Illumina data. Different transcriptome de novo assemblies are generally

difficult to compare (Grabherr et al. 2011; Martin and Wang 2011; O'Neil and Emrich 2013). Assembly software can differ in the number, length, and quality of the contigs they produce, as well as their sensitivity for finding low expressed transcripts (Bradnam et al. 2013; Peng et al. 2013), especially during de novo assembly. In our analyses, we compared assemblies produced by the commonly used licensed software suite CLC Genomics Workbench (CLC bio) and IDBA-tran (Peng et al. 2013). IDBA-tran is a de novo transcriptome assembler based on de Bruijn graphs for Illumina platform based data that uses an iterative *k*-mer optimization approach. IDBA-tran outperforms widely used software such as Trinity (Grabherr et al. 2011) with respect to sensitivity, specificity, and the number of correctly assembled contigs (Peng et al. 2013). Moreover, it is specifically designed to assemble low abundance transcripts, which is important as highly potent toxins may be expressed at lower levels.

Our results confirm that the choice of transcriptome assembly method can substantially influence the results (see [supplementary table S1, Supplementary Material](#) online). We found that overall IDBA was more sensitive than CLC. For several putative toxins, only IDBA was able to generate contigs. For instance, putative actinoporin-like toxins were never identified in the CLC assemblies, and only IDBA produced for the *G. tridactyla* library contigs of the ShKT domain toxin, and for the *G. fallax* library contigs of CAP domain toxins. However, IDBA did not generate a larger number of contigs than CLC for every toxin. The assemblers also differed in the average length of contigs produced. This is especially important for identifying N-terminal signal peptides. With some exceptions, such as the absence of signal peptides from many secreted cnidarian peptides and stonefish SNTXs (Ghadessy et al. 1996; Balasubramanian et al. 2012; Brinkman et al. 2012), most venom proteins are secreted and should therefore have a signal peptide (Fry et al. 2009). For our data IDBA generally produced assemblies comprising longer transcripts including signal peptides, with the exception of hyaluronidase, serpin and sphingomyelinase, for which only CLC produced contigs with predicted signal peptides. We therefore based our analyses on the IDBA-tran assembled data.

Multipronged Approach to Toxin Homolog Identification

Many transcriptomic studies of venom composition rely either on BLAST (Altschul et al. 1990) or on InterProScan (Zdobnov and Apweiler 2001) based identification of putative venom proteins. Instead, our approach for annotating and identifying all putative toxin families of *Glycera* consists of three independent search strategies: 1) BLAST2GO based InterPro annotation (Conesa et al. 2005), 2) profile HMMs (Eddy 1998) based on profiles of full-length toxin sequences, and 3) profile HMM based on single protein domains (for details see Materials and Methods and [supplementary fig. S5, Supplementary Material](#) online).

BLAST identification is especially useful for analyses of large numbers of sequences. However, BLAST is typically less sensitive than model-based approaches when it comes to identifying protein homologs showing a wide evolutionary diversity (Koua et al. 2013). Moreover, a lot of manual editing of BLAST-outputs is needed if toxins are composed of domain regions that are found also in proteins with normal physiological functions unrelated to venom proteins. In contrast, BLAST2GO utilizes an automated model-based approach and scans whole transcript sequences for domains through HMM profiles based on InterProScan. Thereby, it searches against protein databases such as PFAM, PROSITE, and SUPERFAMILY (Conesa and Götz 2008). The protein family IDs and related protein or domain names are output into a ".xml" file that can be searched by text or script-based strategies. However, with this approach obviously only toxins which are curated in these databases can be discovered. This is true, for instance, for glycerotoxin, the activity of which has been characterized, but for which no sequence data are available (Meunier et al. 2002, 2010). Additionally, searching toxins in.xml files is hindered by the lack of a consistent vocabulary and high numbers of synonyms for names of protein families in these databases (Koua et al. 2013).

The identification of gigantoxin I-like transcripts in our data represents a good example of the difficulty of identifying new toxins that were neither revealed by InterProScan (domain IDs) nor by our HMMER approaches. The use of probabilistic models of sequence composition in HMMER is beneficial as these are based on large multiple sequence alignments of specific protein families. HMMs use position-specific insertion/deletion probabilities instead of the arbitrary, position-invariant gap costs of the more traditional BLAST Smith–Waterman algorithm, and this allows profile HMMs to model a varying frequency of indels (Eddy 1998; Koua et al. 2013). One downside is that HMMER requires alignments of a minimum number of homologous sequences to train HMM profiles. Because only a small number of gigantoxin I-like sequences is present in public databases we could not train a HMM profile to identify this toxin in our assemblies. Moreover, the domain of gigantoxin I strongly resembles an EGF domain, which is present in many proteins that are abundant in nonvenom gland tissue. Hence, a search for gigantoxin through InterProScan would lead to many false positives. In the case of gigantoxin, it was the BLAST output from BLAST2GO (against NCBI [National Center for Biotechnology Information] nr) that gave the correct match for one of our transcripts. The same was true for the identification of the two turripeptide-like toxins. InterProScan and HMMER annotated their Kazal domains. However, only careful study of our BLAST results provided the seed for further investigation that resulted in our identification of the first nonmollusk turripeptide-like toxin transcripts.

Generally, the different search strategies produced similar results for our data. There was a high degree of overlap in

identification between the methods, yet a number of transcripts could have been overlooked if only one of the search strategies would have been applied. Our methodological conclusions are therefore 2-fold. First, only a combination of the above mentioned strategies is able to exhaustively identify putative toxins. Second, unusual or novel toxins require off-pipeline attention due to the lack of information present in public databases.

Conclusion

In summary, we present the first transcriptomic profiles of polychaete venom glands. *Glycera* venom glands express a complex cocktail of putative toxin precursor transcripts, including neurotoxic peptides, pore-forming proteins, and enzymes. Our study further expands the number and taxonomic range of protein types known to have been convergently recruited into animal venoms (Casewell et al. 2013). Notably, bloodworm venom glands express homologs of toxins previously only known from venoms of scorpaeniform fish and monotremes (SNTX-like toxin), turrid gastropods (turriptide-like peptides), and sea anemones (gigantoxin I-like neurotoxin), as well as many toxins with a broader phylogenetic distribution. Given this unexpected diversity of new putative bloodworm toxins, it will be interesting to profile more species of bloodworms.

The deeply sequenced venom gland transcriptome of *G. dibranchiata* shows that with the single exception of the protease inhibitor cystatin, all identified venom toxin transcripts have undergone varying degrees of lineage-specific diversification. Highly expressed putative toxins, such as the enzymes peptidase S1 and metalloprotease M12, as well as CAP and Kazal domain proteins, have diversified into large numbers of species- or genus-specific paralogs. Although proteomic and functional studies are needed to confirm our transcriptomic predictions of venom composition and activities, our results clearly suggest that bloodworms are effective predators that likely use complex toxin cocktails to subdue and process macroscopic prey. These predatory habits are confirmed by the anatomy of their venom apparatus, as well as field observations of their feeding habits, even for *G. dibranchiata*, which had long been thought to be a detritivore.

Moreover, our results suggest that in order to produce exhaustive transcriptomic profiles of venom glands it is important to be aware that different transcriptome assembly methods, as well as different methods of homology prediction, can yield different results. Although our analyses benefit from the use of a customized bioinformatic pipeline, several putative toxins would not have been identified without manually scrutinizing the results. But when a bioinformatic pipeline is combined with a thorough inspection of the results, deep transcriptomic profiling of venom glands of poorly studied taxa is one of the most effective ways to broaden the empirical foundation of venomomics.

Supplementary Material

Supplementary files S1–S3 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

Phylogenetic analyses were performed with the assistance of SBCS Informatics (<http://informatics.sbcs.qmul.ac.uk>), the EPSRC funded MidPlus cluster at Queen Mary University in London, United Kingdom. The authors are grateful to Birgit Nickel (MPI EVA Leipzig) for assisting in the generation of GAll transcriptome libraries and to Anne Weigert (University of Leipzig) for providing a photograph of a glycerid. Giampietro Schiavo (Cancer Research UK London Research Institute) gave valuable comments on the possible function of some toxin candidates. This work was supported by German Science Foundation grants RE 3454/1-1 and RE 3454/1-2 to B.M.v.R., BL7877-1 to C.B., and a Biotechnology and Biological Sciences Research Council grant BB/K003488/1 and a Departmental Investment Fund grant from the Natural History Museum to R.A.J. R.A.J. and C.B. designed the study. B.M.v.R., J.H., L.C., L.H., and S.R. designed and performed the bioinformatic analyses. B.M.v.R., L.C., R.A.J., and C.B. led data interpretation. D.S. produced the micro-CT scan of *Glycera tessellata*. B.M.v.R., L.C., C.B., and R.A.J. wrote the article. All authors read and approved the manuscript.

Literature Cited

- Abdel-Rahman MA, Omran MAA, Abdel-Nabi IM, Ueda H, McVean A. 2009. Intraspecific variation in the Egyptian scorpion *Scorpio maurus palmatus* venom collected from different biotopes. *Toxicon* 53: 349–359.
- Aird SD, et al. 2013. Quantitative high-throughput profiling of snake venom gland transcriptomes and proteomes (*Ovophis okinavensis* and *Protobothrops flavoviridis*). *BMC Genomics* 14:790.
- Alarcon-Chaidez FJ, Boppana VD, Hagymasi AT, Adler AJ, Wikel SK. 2009. A novel sphingomyelinase-like enzyme in *Ixodes scapularis* tick saliva drives host CD4 +T cells to express IL-4. *Parasite Immunol.* 31: 210–219.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Alves-Silva J, et al. 2010. An insight into the sialome of *Glossina morsitans morsitans*. *BMC Genomics* 11:213.
- Ambrose W. 1984. Influences of predatory polychaetes and epibenthic predators on the structure of a soft-bottom community in a Maine estuary. *J Exp Mar Biol Ecol.* 81:115–145.
- Anatriello E, et al. 2010. An insight into the sialotranscriptome of the brown dog tick, *Rhipicephalus sanguineus*. *BMC Genomics* 11:450.
- Anderluh G, Macek P. 2003. Dissecting the actinoporin pore-forming mechanism. *Structure* 11:1312–1313.
- Bakrac B, Anderluh G. 2010. Molecular mechanism of sphingomyelin-specific membrane binding and pore formation by actinoporins. In: Anderluh G, Lakey J, editors. *Proteins: membrane binding and pore formation*. Austin (TX): Springer Science+Business Media, LLC. p. 106–115.

- Bakrac B, et al. 2008. Molecular determinants of sphingomyelin specificity of a eukaryotic pore-forming toxin. *J Biol Chem.* 283: 18665–18677.
- Balasubramanian PG, et al. 2012. Proteome of *Hydra* nematocyst. *J Biol Chem.* 287:9672–9681.
- Barlic A, et al. 2004. Lipid phase coexistence favors membrane insertion of equinatoxin-II, a pore-forming toxin from *Actinia equina*. *J Biol Chem.* 279:34209–34216.
- Binford GJ, et al. 2009. Molecular evolution, functional variation, and proposed nomenclature of the gene family that includes sphingomyelinase D in sicariid spider venoms. *Mol Biol Evol.* 26:547–566.
- Binford GJ, Wells MA. 2003. The phylogenetic distribution of sphingomyelinase D activity in venoms of Haplogyne spiders. *Comp Biochem Phys B.* 135:25–33.
- Böggemann M. 2002. Revision of the Glyceridae Grube 1850 (Annelida: Polychaeta). *Abh Senckenb Naturforsch Ges.* 555:1–249.
- Bon C, Saliou B, Thieffry M, Manaranche R. 1985. Purification of alpha-glycerotoxin, a neurotoxin from *Glycera convoluta* that triggers neurotransmitter release. *Toxicon* 23:24–25.
- Borda E, Kudenov JD, Bienhold C, Rouse GW. 2012. Towards a revised Amphinomidae (Annelida, Amphinomida): description and affinities of a new genus and species from the Nile Deep-sea Fan, Mediterranean Sea. *Zool Scr.* 41:307–325.
- Bradnam K, et al. 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience* 2:10.
- Brinkman DL, et al. 2012. Venom proteome of the box jellyfish *Chironex fleckeri*. *PLoS One* 7:e47866.
- Casewell NR, Wuester W, Vonk FJ, Harrison RA, Fry BG. 2013. Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol Evol.* 28: 219–229.
- Castaneda O, et al. 1995. Characterization of a potassium channel toxin from the Caribbean Sea anemone *Stichodactyla helianthus*. *Toxicon* 33:603–613.
- Castaneda O, Harvey AL. 2009. Discovery and characterization of cnidarian peptide toxins that affect neuronal potassium ion channels. *Toxicon* 54:1119–1124.
- Chagas AC, et al. 2013. A deep insight into the sialotranscriptome of the mosquito, *Psorophora albipes*. *BMC Genomics* 14:875.
- Chang D, Duda TF Jr. 2012. Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. *Mol Biol Evol.* 29:2019–2029.
- Chen D, Kini RM, Yuen R, Khoo HE. 1997. Haemolytic activity of stonustoxin from stonefish (*Synanceja horrida*) venom: pore formation and the role of cationic amino acid residues. *Biochem J.* 325:685–691.
- Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.
- Conesa A, Götz S. 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics.* 2008:619832.
- Cuyppers E, Peigneur S, Debaveye S, Shiomi K, Tytgat J. 2011. TRPV1 channel as new target for marine toxins: example of giantoxin I, a sea anemone toxin acting via modulation of the PLA2 pathway. *Acta Chim Slov.* 58:735–741.
- Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27: 1164–1165.
- De Colibus L, et al. 2012. Structures of Lysenin reveal a shared evolutionary origin for pore-forming proteins and its mode of sphingomyelin recognition. *Structure* 20:1498–1507.
- Earl D, et al. 2011. Assemblathon 1: a competitive assessment of de novo short read assembly methods. *Genome Res.* 21:2224–2241.
- Eddy SR. 1998. Profile hidden Markov models. *Bioinformatics* 14:755–763.
- Fauchald K, Jumars PA. 1979. The diet of worms: a study of polychaete feeding guilds. *Oceanogr Mar Biol.* 17:193–284.
- Favreau P, et al. 2006. Mass spectrometry strategies for venom mapping and peptide sequencing from crude venoms: case applications with single arthropod specimen. *Toxicon* 47:676–687.
- Fernandes-Pedrosa M, et al. 2008. Transcriptome analysis of *Loxosceles laeta* (Araneae, Sicariidae) spider venomous gland using expressed sequence tags. *BMC Genomics* 9:279.
- Francischetti IMB, et al. 2013. The “Vampirome”: transcriptome and proteome analysis of the principal and accessory submaxillary glands of the vampire bat *Desmodus rotundus*, a vector of human rabies. *J Proteomics.* 82:288–319.
- Fry BG, et al. 2009. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu Rev Genom Hum G.* 10: 483–511.
- Fry BG, et al. 2012. The structural and functional diversification of the Toxicofera reptile venom system. *Toxicon* 60:434–448.
- Garb JE, Hayashi CY. 2013. Molecular evolution of α -latrotoxin, the exceptionally potent vertebrate neurotoxin in black widow spider venom. *Mol Biol Evol.* 30:999–1014.
- Garcia T, et al. 2009. Pharmacological effects of two cytolytins isolated from the sea anemone *Stichodactyla helianthus*. *J Biosci.* 34:891–898.
- Garcia-Ortega L, et al. 2011. The behavior of sea anemone actinoporins at the water-membrane interface. *Biochim Biophys Acta.* 1808: 2275–2288.
- Ghadessy F, et al. 1996. Stonustoxin is a novel lethal factor from stonefish (*Synanceja horrida*) venom cDNA cloning and characterization. *J Biol Chem.* 271:25575–25581.
- Giese C, Mebs D, Werding B. 1996. Resistance and vulnerability of crustaceans to cytolytic sea anemone toxins. *Toxicon* 34:955–958.
- Giray C, King GM. 1997. Predator deterrence and 2,4-dibromophenol conservation by the enteropneusts *Saccoglossus bromophenolosus* and *Protoglossus graveolensis*. *Mar Ecol Prog Ser.* 159:229–238.
- Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29:644–652.
- Hargreaves AD, Swain MT, Hegarty MJ, Logan DW, Mulley JF. 2014. Restriction and recruitment—gene duplication and the origin and evolution of snake venom toxins. *Genome Biol Evol.* 6:2088–2095.
- He Q, et al. 2013. The venom gland transcriptome of *Latrodectus tede-cinguttatus* revealed by deep sequencing and cDNA library analysis. *PLoS One* 8:e81357.
- Honma T, Nagai H, Nagashima Y, Shiomi K. 2003. Molecular cloning of an epidermal growth factor-like toxin and two sodium channel toxins from the sea anemone *Stichodactyla gigantea*. *Biochim Biophys Acta.* 1652:103–106.
- Jungo F, Bougueleret L, Xenarios I, Poux S. 2012. The UniProtKB/Swiss-Prot Tox-Prot program: a central hub of integrated venom protein data. *Toxicon* 60:551–557.
- Kagan BL, Pollard HB, Hanna RB. 1982. Induction of ion-permeable channels by the venom of the fanged bloodworm *Glycera dibranchiata*. *Toxicon* 20:887–893.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.
- Kawashima Y, Nagai H, Ishida M, Nagashima Y, Shiomi K. 2003. Primary structure of echotoxin 2, an actinoporin-like hemolytic toxin from the salivary gland of the marine gastropod *Monoplex echo*. *Toxicon* 42: 491–497.
- Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG. 2009. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.* 25:404–413.
- King GF. 2011. Venoms as a platform for human drugs: translating toxins into therapeutics. *Exp Opin Biol Ther.* 11:1469–1484.
- King GF, Hardy MC. 2013. Spider-venom peptides: structure, pharmacology, and potential for control of insect pests. *Annu Rev Entomol.* 58: 475–496.

- Kircher M, Stenzel U, Kelso J. 2009. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol.* 10: R83.
- Kiriake A, Suzuki Y, Nagashima Y, Shiomi K. 2013. Proteinaceous toxins from three species of scorpaeniform fish (lionfish *Pterois lunulata*, devil stinger *Inimicus japonicus* and waspfish *Hypodytes rubripinnis*): close similarity in properties and primary structures to stonefish toxins. *Toxicon* 70:184–193.
- Klawe WL, Dickie LM. 1957. Biology of the bloodworm, *Glycera dibranchiata* Ehlers, and its relation to the bloodworm fishery of the Maritime Provinces. *Bull Fish Res Board Can.* 115:1–33.
- Kobayashi H, Ohtomi M, Sekizawa Y, Ohta N. 2001. Toxicity of coelomic fluid of the earthworm *Eisenia foetida* to vertebrates but not invertebrates: probable role of sphingomyelin. *Comp Biochem Phys C.* 128:401–411.
- Koua D, et al. 2013. Position-specific scoring matrix and hidden Markov model complement each other for the prediction of conopeptide superfamilies. *Biochim Biophys Acta.* 1834:717–724.
- Krebs HC, Habermehl GG. 1987. Isolierung und Strukturklärung eines hämolytisch aktiven Peptids aus der Seeanemone *Metridium senile*. *Naturwissenschaften* 74:395–396.
- Kristan KC, Viero G, Dalla Serra M, Macek P, Anderluh G. 2009. Molecular mechanism of pore formation by actinoporins. *Toxicon* 54: 1125–1134.
- Kvist S, Brugler MR, Goh TG, Giribet G, Siddall ME. 2014. Pyrosequencing the salivary transcriptome of *Haemadipsa interrupta* (Annelida: Clitellata: Haemadipsidae): anticoagulant diversity and insight into the evolution of anticoagulation capabilities in leeches. *Invertebr Biol.* 133:74–98.
- Lin H-H, Tseng L-Y. 2010. DBCP: a web server for disulfide bonding connectivity pattern prediction with the prior knowledge of the bonding state of cysteines. *Nucleic Acids Res.* 38:W503–W507.
- Low DHW, et al. 2013. Dracula's children: molecular evolution of vampire bat venom. *J Proteomics.* 89:95–111.
- Low K, Gwee M, Yuen R, Gopalakrishnakone P, Khoo H. 1994. Stonustoxin: effects on neuromuscular function *in vitro* and *in vivo*. *Toxicon* 32:573–581.
- Manaranche R, Thieffry M, Israel M. 1980. Effect of the venom of *Glycera convoluta* on the spontaneous quantal release of transmitter. *J Cell Biol.* 85:446–458.
- Mancheno JM, Martin-Benito J, Martinez-Ripoll M, Gavilanes JG, Hermoso JA. 2003. Crystal and electron microscopy structures of sticholysin II actinoporin reveal insights into the mechanism of membrane pore formation. *Structure* 11:1319–1328.
- Margres MJ, Aronow K, Loyacano J, Rokyta DR. 2013. The venom-gland transcriptome of the eastern coral snake (*Micrurus fulvius*) reveals high venom complexity in the intragenomic evolution of venoms. *BMC Genomics* 14:1–10.
- Martin JA, Wang Z. 2011. Next-generation transcriptome assembly. *Nat Rev Genet.* 12:671–682.
- Meunier FA, et al. 2010. Sustained synaptic-vesicle recycling by bulk endocytosis contributes to the maintenance of high-rate neurotransmitter release stimulated by glycerotoxin. *J Cell Sci.* 123: 1131–1140.
- Meunier FA, Feng ZP, Molgo J, Zamponi GW, Schiavo G. 2002. Glycerotoxin from *Glycera convoluta* stimulates neurosecretion by up-regulating N-type Ca₂₊ channel activity. *EMBO J.* 21:6733–6743.
- Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* 2010. pdb.prot5448.
- Michel C, Keil B. 1975. Biologically active proteins in venomous glands of polychaetous annelid, *Glycera convoluta* Keferstein. *Comp Biochem Physiol.* 50:29–32.
- Michel C, Robin Y. 1972. Preliminary biochemical data on venomous glands of *Glycera convoluta* Keferstein (Annelida, Polychaeta). *C R Seances Soc Biol Fil.* 166:853–857.
- Min G-S, Sarkar IN, Siddall ME. 2010. Salivary transcriptome of the North American medicinal leech, *Macrobodella decora*. *J Parasitol.* 96: 1211–1221.
- 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.
- Morgenstern D, King GF. 2013. The venom optimization hypothesis revisited. *Toxicon* 63:120–128.
- Moses DN, Mattoni MA, Slack NL, Waite JH, Zok FW. 2006. Role of melanin in mechanical properties of *Glycera* jaws. *Acta Biomater.* 2: 521–530.
- Nakamura K, et al. 2008. Complanine, an inflammation-inducing substance isolated from the marine fireworm *Eurythoe complanata*. *Org Biomol Chem.* 6:2058–2060.
- Olivera BM, et al. 2012. Adaptive radiation of venomous marine snail lineages and the accelerated evolution of venom peptide genes. *Ann N Y Acad Sci.* 1267:61–70.
- O'Neil ST, Emrich SJ. 2013. Assessing De Novo transcriptome assembly metrics for consistency and utility. *BMC Genomics* 14:465.
- Peng Y, et al. 2013. IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. *Bioinformatics* 29:326–334.
- Pennington MW, et al. 1999. Role of disulfide bonds in the structure and potassium channel blocking activity of ShK toxin. *Biochemistry* 38: 14549–14558.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 8:785–786.
- Rangaraju S, et al. 2010. Potassium channel modulation by a toxin domain in matrix metalloprotease 23. *J Biol Chem.* 285: 9124–9136.
- Rigoni M, et al. 2005. Equivalent effects of snake PLA2 neurotoxins and lysophospholipid-fatty acid mixtures. *Science* 310:1678–1680.
- Rokyta DR, Wray KP, Lemmon AR, Lemmon EM, Caudle SB. 2011. A high-throughput venom-gland transcriptome for the Eastern Diamondback Rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes. *Toxicon* 57:657–671.
- Rosenberg P, Ishay J, Gitter S. 1977. Phospholipases A and B activities of oriental hornet (*Vespa orientalis*) venom and venom apparatus. *Toxicon* 15:141–155.
- Rouse GW, Pleijel F. 2001. *Polychaetes*. Oxford: Oxford University Press.
- Safavi-Hemami H, et al. 2011. Embryonic toxin expression in the cone snail *Conus victoriae*: primed to kill or divergent function? *J Biol Chem.* 286: 22546–22557.
- Schenning M, et al. 2006. Glycerotoxin stimulates neurotransmitter release from N-type Ca₂₊ channel expressing neurons. *J Neurochem.* 98: 894–904.
- Shiomi K, et al. 2003. An epidermal growth factor-like toxin and two sodium channel toxins from the sea anemone *Stichodactyla gigantea*. *Toxicon* 41:229–236.
- Shiomi K, Kawashima Y, Mizukami M, Nagashima Y. 2002. Properties of proteinaceous toxins in the salivary gland of the marine gastropod (*Monoplex echo*). *Toxicon* 40:563–571.
- Smeds L, Kunstner A. 2011. ConDeTri—a content dependent read trimmer for Illumina data. *PLoS One* 6:e26314.
- Stamatakis A, Alachiotis N. 2010. Time and memory efficient likelihood-based tree searches on phylogenomic alignments with missing data. *Bioinformatics* 26:i132–i139.
- Sukumwang N, Umezawa K. 2013. Earthworm-derived pore-forming toxin Lysenin and screening of its inhibitors. *Toxins* 5:1392–1401.
- Sunagar K, et al. 2014. Intraspecific venom variation in the medically significant Southern Pacific Rattlesnake (*Crotalus oreganus helleri*): bio-discovery, clinical and evolutionary implications. *J Proteomics.* 99: 68–83.

- Terrat Y, et al. 2012. High-resolution picture of a venom gland transcriptome: case study with the marine snail *Conus consors*. *Toxicon* 59: 34–46.
- Ueda A, et al. 2006. Purification, properties and cDNA cloning of neoverrucotoxin (neoVTX), a hemolytic lethal factor from the stonefish *Synanceia verrucosa* venom. *Biochim Biophys Acta*. 1760:1713–1722.
- Ungerer NK. 2002. Characterization of the venom of *Glycera dibranchiata* [Senior Honors thesis][Sweet Briar (VA)]: Department of Biology, Sweet Briar College.
- Violette A, et al. 2012. Large-scale discovery of conopeptides and conoproteins in the injectable venom of a fish-hunting cone snail using a combined proteomic and transcriptomic approach. *J Proteomics*. 75: 5215–5225.
- von Reumont BM, et al. 2014. The first venomous crustacean revealed by transcriptomics and functional morphology: remipede venom glands express a unique toxin cocktail dominated by enzymes and a neurotoxin. *Mol Biol Evol*. 31:48–58.
- Vonk FJ, et al. 2013. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc Natl Acad Sci U S A*. 110:20651–20656.
- Warren WC, et al. 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453:175–183.
- Watala C, Kowalczyk JK. 1990. Hemolytic potency and phospholipase activity of some bee and wasp venoms. *Comp Biochem Phys C*. 97: 187–194.
- Weston AJ, et al. 2013. Proteomic characterisation of toxins isolated from nematocysts of the South Atlantic jellyfish *Olindias sambaquiensis*. *Toxicon* 71:11–17.
- Whittington CM, et al. 2010. Novel venom gene discovery in the platypus. *Genome Biol*. 11:R95.
- 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.
- Zdobnov EM, Apweiler R. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847–848.
- Zhang Z-Q. 2011. Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness. *Zootaxa* 3148:1–237.
- Zimmerberg J, Chernomordik LV. 2005. Synaptic membranes bend to the will of a neurotoxin. *Science* 310:1626–1627.
- Zobel-Thropp PA, Kerins AE, Binford GJ. 2012. Sphingomyelinase D in sicariid spider venom is a potent insecticidal toxin. *Toxicon* 60: 265–271.

Associate editor: Dorothee Huchon