# Transcriptome Analysis in Venom Gland of the Predatory Giant Ant *Dinoponera quadriceps*: Insights into the Polypeptide Toxin Arsenal of Hymenopterans

**Alba F. C. Torres[1], Chen Huang[2], Cheong-Meng Chong[2], Siu Wai Leung[2], Álvaro R. B. Prieto-da-Silva[3], Alexandre Havt[4], Yves P. Quinet[5], Alice M. C. Martins[1]*, Simon M. Y. Lee[2], Gandhi Rádis-Baptista[6]***

**1** Department of Clinical and Toxicological Analysis, Faculty of Pharmacy, Federal University of Ceara, Fortaleza, Ceara, Brazil, **2** State Key Laboratory of Quality Research in Chinese Medicine and Institute of Chinese Medical Sciences, University of Macau, Macao, China, **3** Laboratory of Genetics, Institute Butantan, Sao Paulo, Brazil, **4** Department of Physiology and Pharmacology, Faculty of Medicine, Federal University of Ceara, Fortaleza, Ceara, Brazil, **5** Laboratory of Entomology, State University of Ceara, Fortaleza, Ceara, Brazil, **6** Laboratory of Biochemistry and Biotechnology, Institute for Marine Sciences, Federal University of Ceara, Brazil

## Abstract

***Background:*** *Dinoponera quadriceps* is a predatory giant ant that inhabits the Neotropical region and subdues its prey (insects) with stings that deliver a toxic cocktail of molecules. Human accidents occasionally occur and cause local pain and systemic symptoms. A comprehensive study of the *D. quadriceps* venom gland transcriptome is required to advance our knowledge about the toxin repertoire of the giant ant venom and to understand the physiopathological basis of Hymenoptera envenomation.

***Results:*** We conducted a transcriptome analysis of a cDNA library from the *D. quadriceps* venom gland with Sanger sequencing in combination with whole-transcriptome shotgun deep sequencing. From the cDNA library, a total of 420 independent clones were analyzed. Although the proportion of dinoponeratoxin isoform precursors was high, the first giant ant venom inhibitor cysteine-knot (ICK) toxin was found. The deep next generation sequencing yielded a total of 2,514,767 raw reads that were assembled into 18,546 contigs. A BLAST search of the assembled contigs against non-redundant and Swiss-Prot databases showed that 6,463 contigs corresponded to BLASTx hits and indicated an interesting diversity of transcripts related to venom gene expression. The majority of these venom-related sequences code for a major polypeptide core, which comprises venom allergens, lethal-like proteins and esterases, and a minor peptide framework composed of inter-specific structurally conserved cysteine-rich toxins. Both the cDNA library and deep sequencing yielded large proportions of contigs that showed no similarities with known sequences.

***Conclusions:*** To our knowledge, this is the first report of the venom gland transcriptome of the New World giant ant *D. quadriceps*. The glandular venom system was dissected, and the toxin arsenal was revealed; this process brought to light novel sequences that included an ICK-folded toxins, allergen proteins, esterases (phospholipases and carboxylesterases), and lethal-like toxins. These findings contribute to the understanding of the ecology, behavior and venomics of hymenopterans.

**Citation:** Torres AFC, Huang C, Chong C-M, Leung SW, Prieto-da-Silva ÁRB, et al. (2014) Transcriptome Analysis in Venom Gland of the Predatory Giant Ant *Dinoponera quadriceps*: Insights into the Polypeptide Toxin Arsenal of Hymenopterans. PLoS ONE 9(1): e87556. doi:10.1371/journal.pone.0087556

**Editor:** Paulo Lee Ho, Instituto Butantan, Brazil

**Received** September 19, 2013; **Accepted** December 23, 2013; **Published** January 31, 2014

**Copyright:** © 2014 Torres et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The giant ant project was sponsored by the program for funding new research groups of The National Council for Scientific and Technological Development (CNPq), Ministry of Science and Technology, Federal Government of Brazil (program PRONEM/FUNCAP/CNPq, Ref. numbers PRN-0040-00067.01.00/10 and SPU 10582724-0). This study was partly supported by grants from the Science and Technology Development Fund of Macau, China (Ref. no. from 045/2007 to 078/2011/A3). GR-B and ARBPS are grateful to CNPq and the Coordination for the Improvement of Higher Education Personnel (CAPES) Program on Toxinology (Issued 2010), Ministry of Education of the Federal Government of Brazil for partial financial support for this work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: martinsalice@gmail.com (AMCM); gandhi.radis@ufc.br (GRB)

These authors contributed equally to this work.

## Introduction

In recent years, the order Hymenoptera, which comprises numerous species of bees, wasps and ants, has received increasing attention because of their direct and indirect influences on human health, ecological balance, agriculture and the forestry economy. The importance of investigating hymenopterans is well-known; thus, information from diverse studies has culminated with the establishment of a genome database (http://hymenopteragenome.org). Presently, this database maintains scientific information about one species of bee (*Apis mellifera*), two species of bumblebees (*Bombus terrestris* and *B. impatiens*), a parasitoid wasp (*Nasonia vitripennis*), and seven species of ants [1]. The seven ant species detailed in this database include the fungus-growing (leaf cutter) ants *Acromyrmex echinatior* and *Atta cephalotes*, the Florida carpenter ant *Camponotus floridanus*, Jerdon's jumping ant *Harpegnathos saltator*,

the Argentine ant *Linepithema humile*, the red harvester ant *Pogonomyrmex barbatus*, and the fire ant *Solenopsis invicta* (http://hymenopteragenome.org/ant_genomes/).

From the perspective of molecular toxinology and venomics, scientific interest and public concerns about animal venoms have sharply increased due the vast diversity of bioactive compounds, the safety hazards, and the potential therapeutic uses of isolated compounds [2]. Hymenopterans offer a myriad of biologically active molecules that incorporate the physiopathological bases of the causes of ant venom allergies and envenomation and the traditional and modern uses of ants for medicines and in programs of drug discovery. For example, in countries of East Asia, Africa and South America, ant venoms have been used to treat rheumatoid arthritis [3]. Badr and collaborators [4] demonstrated that the crude venom from *Pachycondyla senaarensis* induces death by apoptosis in breast carcinoma cells. The same venom has been proven to display anti-inflammatory and anti-oxidant properties [5]. Additionally, the venom of Chinese medicinal ant *Polyrhachis lammellidens* elicits analgesic and anti-inflammatory effects [6]. Peptides from *Pachycondyla goeldii* and *Myrmecia* sp. have been demonstrated to be effective against Gram-negative and Gram-positive bacteria, including the methicilin-resistant *Staphyloccocus aureus* and the pathogenic fungus *Candida albicans* [7,8]. Moreover, the antimicrobial effects of crude venom from *Crematogaster pygmaea*, which is composed of electrophilic aldehydes, have recently been demonstrated [9].

Our group has been investigating the biological effects of *D. quadriceps* crude venom in different experimental models. In a system of isolated perfused rat kidney, *D. quadriceps* venom triggers increases in urinary flow, decreases in vascular resistance, and reductions in sodium tubular transport; these findings demonstrate the natriuretic and diuretic-induction effects of this crude venom [10]. In pentylenetetrazol-induced seizure models, this giant ant venom exhibits neuroprotective activities that depend on the route of administration [11,12]. Additionally, models of antinociception, including those evoked by formalin, writhing, von Frey fibers and hot plates, the analgesic effects of *D. quadriceps* crude venom have been verified to be efficacious [13].

The genus *Dinoponera* (Hymenoptera: Vespoidea: Formicidae) belongs to the subfamily Ponerinae and comprises the following eight species with strict Neotropical distributions [14–16]: *D. australis*, *D. gigantea*, *D. hispida*, *D. longipes*, *D. lucida*, *D. mutica*, *D. quadriceps*, and *D. snellingi*. Collectively, these species form a group of primitive poneromorphs ants with body sizes ranging from 3 to 4 cm, which places them among the largest ants in the world [17,18]. Along with some other ponerine genera, *Dinoponera* ants have queenless colonies, and reproductive functions are performed by one mated dominant worker called a gamergate [19]. In Brazil, *Dinoponera* ants, popularly called "falsas tocandiras", exemplify the genera with strict New World distributions. The giant ant *D. quadriceps* preferentially inhabits the Brazilian northeast and can be found in a variety of vegetation types including *caatinga* (a xerophyllous vegetation), *cerrado* (a savannah-like vegetation), *brejos de altitude* (humid mountain forests) and Atlantic forest. *D. quadriceps* nests (Fig. 1-A) have an average population of 60 workers, one or two entrances, a vertical gallery that is between 30 to 120 cm deep, and up to 16 chambers [15,20]. As in all Aculeata hymenopterans for which anatomical observations exist, the venom gland of *Dinoponera* is formed by free tubules, a convoluted gland and a reservoir (venom sac) (Fig. 1-B) [21,22]. The free tubules and convoluted gland are involved in venom production, and the venom is a fluid that is rich in polypeptides and characteristic of the stinging ants that belong to the subfamilies Myrmicinae, Ponerinae and Pseudomyrmecinae [23].

*Dinoponera* ants preferentially prey on insects and small arthropods, but inevitable interactions with humans occur, and, upon venom injection, the location of the sting becomes extremely painful. Additionally, the stings may have systemic manifestations such as sweating, nausea, vomiting, fever, lymphadenopathy, cardiac arrhythmias and tissue damage [24–27]. These symptoms are mostly attributable to a cocktail of low molecular weight organic compounds and pharmacologically active polypeptides that is present in the venom. Indeed, regarding the polypeptide core of the venom, short peptides isolated from *Paraponera clavata* venom (i.e., poneratoxins (PoTx)) have been proven to prolong the action of sodium channels and interfere with neurotransmission [28,29]. To identify individual venom peptides and proteins in the venom of *Dinoponera australis*, Johnson and colleagues [30] conducted a proteomic survey via high performance liquid chromatography coupled with mass-spectrometry (HPLC/DAD/MS). However, only a limited number of peptides and related partial protein fragments were fully characterized. Consequently, a comprehensive description of the toxic polypeptide repertoire of the venom of the *Dinoponera* giant ant does not yet exist at either the genus or species level. Moreover, despite the large amounts of information about the compositions of the venoms of some species of hymenopterans of the family Formicidae that are available, information about the effects of the components of the venoms on biological systems and the potential uses of the individual polypeptides to manipulate cell biology are still scarce and deserving of deeper investigation. Given this context and the ecological and pharmacological interest in *D. quadriceps* venomics, in this study we report on the transcriptome of the giant ant venom gland and present new data that should further our comprehension of the evolution of the toxin arsenal of Hymenoptera.

## Results

### Sanger sequencing and cluster assemblies

A full-length cDNA library was constructed with approximately 20 venom glands from female (wingless) *D. quadriceps*. A cloning efficiency of $1\times10^5$ cfu/μg DNA was achieved, and the average insert size was 563 bp. A total of 420 individual randomly selected clones were sequenced by with the dideoxy chain-termination (Sanger) method. Nucleotide sequences with greater base calling values were processed via biocomputation (CLC Main Workbench v. 6.8.4, CLC Bio, U.S.A.). The high-quality trimmed ESTs generated an assemblage of 22 contigs and 44 singletons. As summarized in figure 2, the comparison of translated sequences with non-redundant protein databanks using BlastX analysis allowed for the classification and identification the following clusters (Fig. 2-A): (1) proteins involved in general metabolism (56% of contigs and singlets; e.g., transferases, ATP synthase, dehydrogenases, ribosomal proteins, and cytocrome c); (2) venom-linked polypeptide gene sequences (15%) that highlighted the predominant expression of dinoponeratoxins, allergen peptides, and an ICK motif-containing toxin; (3) cDNA precursor sequences that yielded no hits (20%), and (4) ESTs that represented hypothetical proteins with unknown function (5%). Comparisons of the amino acid and nucleotide sequences of the hypothetical proteins with a specific arthropod database revealed that the ESTs of *D. quadriceps* shared homology with polypeptides from scorpions and others ants, including *H. saltator*, *S. invicta* and *S. saevissima* (Fig. 2-B; Table S1).

### Ion Torrent deep RNA sequencing and read assembly

Next generation RNA-sequencing of the same *D. quadriceps* venom gland starting material that was used to construct the

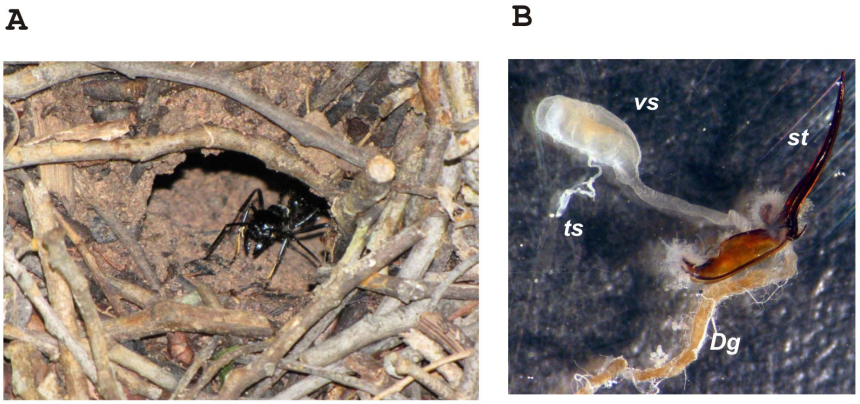



**Figure 1. Dinoponera quadriceps in the field and its dissected venom apparatus.** Part A - A single specimen of D. quadriceps protecting the nest's entrance. Part B - Dissected *D. quadriceps* venom apparatus (×40). Abbreviations: *Dg* - Doffur's gland; *cv* - convoluted gland (not observable); *st* - secretory tubule; *vs*- venom sac; sg – sting.
doi:10.1371/journal.pone.0087556.g001

cDNA library generated a total of 2,514,767 raw reads (average read length = 230 bps). The sequencing data were subjected in-depth bioinformatics analyses. Initially, the raw reads were filtered by an in-house Perl script to remove low quality reads, which resulted in a total of 18,546 assembled transcripts with contig sizes ranging from 101 to 8114 bp, and an average transcript length of 232.09 bp (Table S2). The size distribution of the assembled transcripts is shown in Figure S1. The processed contigs were annotated by querying the non-redundant (nr) protein database via transcript sequence comparison using the BLASTx algorithm, and 6429 contigs with high gene homology were retrieved (Table 1 and Fig. 3). The majority of *D. quadriceps* transcript sequences (3807 contigs) were matched to those of another ant species (*H. saltator*), which reflects their close phylogenetic relationship of these species. Interestingly, some contigs showed high sequence similarity with Hymenoptera species such as *A. florae* and *Megachile rotundata* (see Fig. 4 for details). Nearly all (~98%) EST sequences retrieved from cDNA library matched and strongly supported the assembled contigs that were obtained by next generation RNA-sequencing. Based on non-redundant annotation information, these RNA-derived contigs were then categorized into protein families involved in, for example, cellular processes, Hymenoptera venom-related polypeptides and ribosomal proteins (Fig. 3). This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GANS00000000. The version described in this paper is the first version, GANS01000000.

## Annotation of transcript clusters related to *D. quadriceps* venomics

From the RNA-sequencing data, Gene Ontology (GO) assignments were first used to classify the functions of these assembled transcripts. Based on sequence homology, 3024 transcript sequences were assigned at least one GO term, including 53 functional groups at the second level (Figure S2). Additionally, all assembled transcripts were compared against the Cluster of Orthologous Groups (COG) database to predict possible functions. The results revealed 1592 sequences with significant homology and assigned them into the appropriate clusters. These COG classifications were grouped into 23 functional categories (Figure S3, A to Z). Furthermore, according to the species and protein annotation information of the nr database, the transcripts related to toxins and venom components were grouped into the following four main categories: dinoponeratoxins, venom allergens, phospholipase-like toxin peptides and lethal-like proteins (Fig. 5). Again, as revealed by the cDNA library, there was a high preponderance of dinoponeratoxin transcripts. Venom phospholipase-related toxins and lethal-like proteins corresponded to the largest proportion of contigs. Specifically, the predominant encoded proteins found in the venom gland of *D. quadriceps* were the venom allergen 1 (Sol i 1) and venom allergen 3 (Sol i 3) antigens. Multiple alignments with sequences of other Hymenoptera species revealed several unique amino-acid sequence regions in Sol i 1 and Sol i 3 of *D. quadriceps* (Fig. 6, Fig. 7). The Sol i 3 protein sequences that was orthologous to those of other hymenoptera species, especially bees, were subjected to phylogenetic analyses (Fig. 8). As expected, the phylogenetic tree showed that *D. australis* was more closely related to *S. invicta* (the red imported fire ant) than other bee species. We found four contig sequences that coded for dinoponeratoxins and were homologous to the same toxin class of another ponerine ant species; i.e., *D. australis*. Multiple alignments of the peptide sequences from *D. quadriceps* and the *D. australis* dinoponeratoxins (Fig. 9) showed that these toxin peptides are highly conserved but also exhibited several amino acid substitutions that reflect the species-specific diversification of this toxin.

Contigs that were not related to any known sequences in the nr database accounted for 65% of the total transcripts. This high percentage of "no match" contigs was not surprising because the *E*-value used in the BLAST search was stringent to ensure the reliability of the alignment results. Additionally, the numbers of annotated sequences related to venom from other ant species that are available in the nr database are limited. Moreover, *de novo* RNA-Seq assembly is a complicated process and probably resulted in some sequencing errors, repetitive sequences and aberrant amplicons that brought several "acceptable" incorrect assembled sequence fragments together. Nevertheless, these "no-match" contigs might represent mRNA transcripts that encode novel genes. To further identify and characterize the contigs that had no match in the nr database, these contigs were subjected to a blastx search against the UniProtKB database (2012). This analysis resulted in 34 best hit contigs. For the other "no match" contigs, we performed a BLASTn search against 411 ESTs obtained from the Sanger sequencing of the cDNA library of the venomous gland of *D. quadriceps*, and six contigs were well mapped by numerous ESTs from the venom gland of *D. quadriceps*'s (Table S4). Additionally, ORF prediction was implemented with the OrfPre-

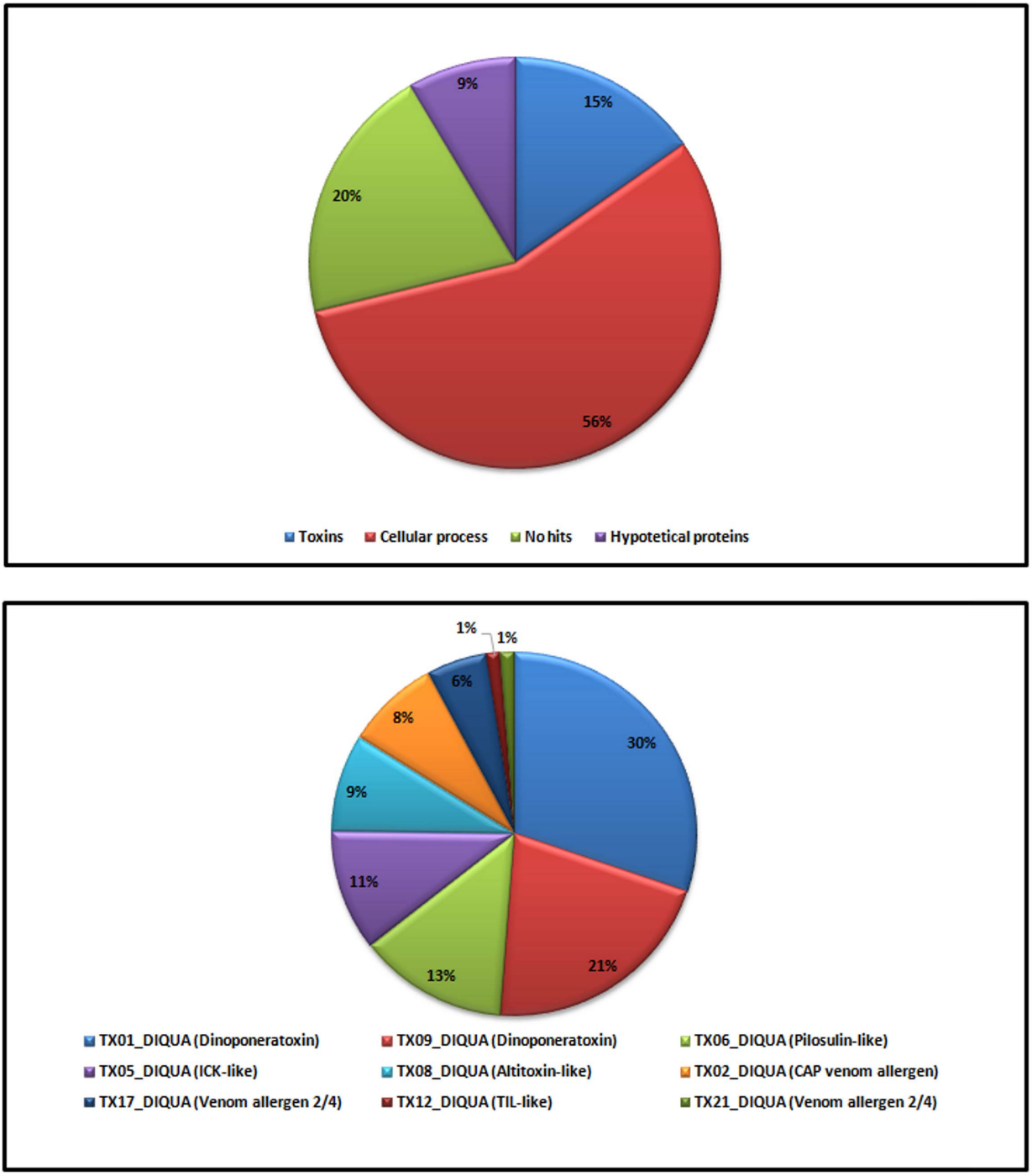



**Figure 2. Overall classification of ESTs from the *D. quadriceps* venom gland cDNA (ESTs) library based on function.** EST sequences were annotated by comparing contigs and singlets (E-values≤1.0E-5) with the non-redundant Genbank translated protein database using BLASTX. Search stringency was eventually adjusted (E-values lower than 1.0E-5) and database restricted to hymenopterans with the aim of finding relevant familial- and structural-related sequences of hymenopterans toxins. Part A – relative number of polypeptides which were counted as individual random sequenced clones. The group of <cellular process> refers to enzymes and polypeptides involved in the venom gland metabolism. Part B – relative numbers of venom-related sequences found in the overrepresented small toxin dataset. Under the term <toxin> is included dinoponeratoxins, venom allergens, and venom cysteine rich peptides (part B).
doi:10.1371/journal.pone.0087556.g002

dictor server, which was designed for ORF prediction and translation of batches of EST or cDNA sequences [31]; then, signal peptides were analyzed with PrediSi [32]. These seven contigs were novel candidate venom encoding transcripts in *D. quadriceps* because they were dissimilar proteins that did not match with any other sequences deposited the nr or UniProtKB databases.

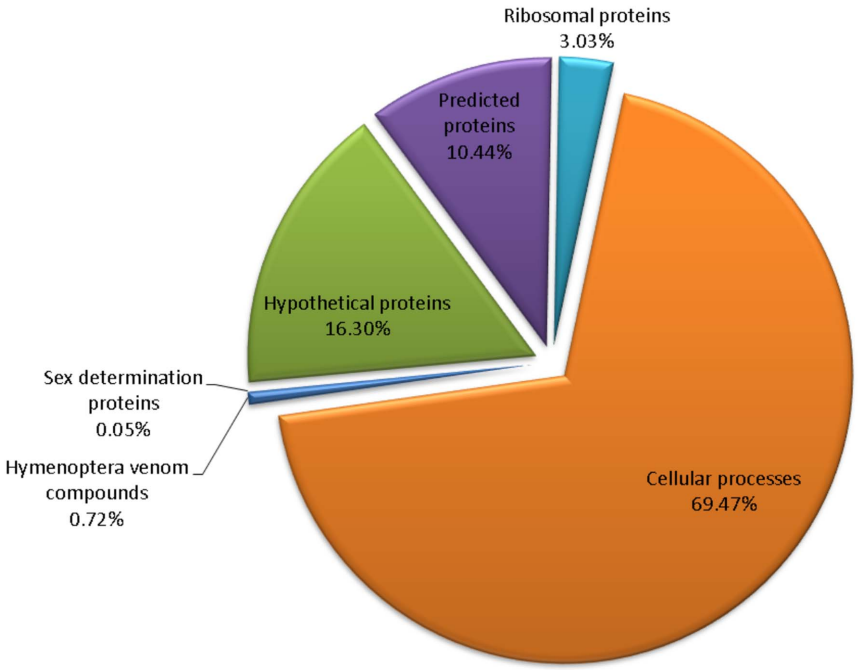



**Figure 3. General overview of the transcripts in the *D. quadriceps* venom gland identified by deep RNA sequencing.** Annotation of contigs from the RNA-Seq assembly of the giant ant venom following the conventions of figure 2; i.e., the E-value cut-off was at least 1.0E-5 for BLASTX functional comparison.
doi:10.1371/journal.pone.0087556.g003

## Clusters related to proteins involved in venom gland physiology

The results of deep RNA-sequencing showed that the venom gland transcriptome was composed of a large proportion of housekeeping genes, such as ribosomal proteins, cellular process-related proteins and hypothetical proteins, but the proportion of venom proteins was small (Fig. 3). Additionally, we identified two sex determination proteins in the venom transcriptome that could be used to predict the phylogenetic relationships of selected ant and bee species. Our phylogenetic analyses indicated that *H. saltator* likely shares the closest parental species with the giant ant *D. quadriceps* and suggests these species might have evolved from a common ancestor (Fig. S4).

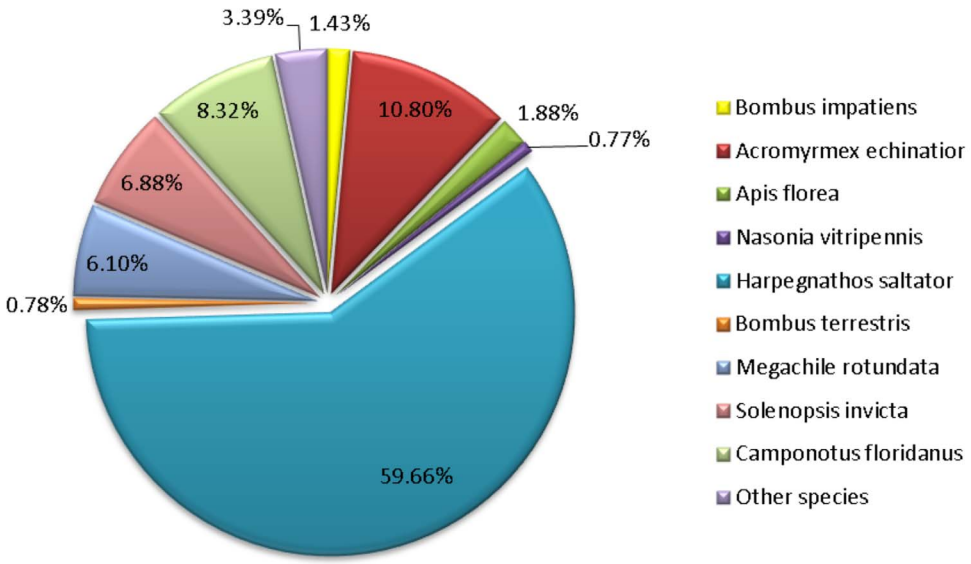



**Figure 4. Distribution of the Hymenoptera species as determined by best protein hits.** The percentages of homologs from distinct hymenopterans species with which the 6,429 contigs of the *D. quadriceps* venom gland shared the high sequence similarities.
doi:10.1371/journal.pone.0087556.g004

**Table 1.** Summary of assembled contigs derived from RNA sequencing and comparison with the ESTs of the *D. quadriceps* venom gland transcriptome.

| | |
|---|---|
| Total number of contigs | 18546 |
| (Average length, bp) | (232) |
| BLASTx-hits | 6429 |
| BLASTx-hits supported ESTs | 32 |
| No-BLASTx-hits | 12117 |
| No-BLASTx-hits supported ESTs | 16 |

doi:10.1371/journal.pone.0087556.t001

### Inhibitor cystine knot- (ICK-) toxin transcripts in the venom gland of *D. quadriceps*

A deduced amino acid sequence encoded by contig 05 (Table S1) has a putative signal peptide of 23 amino acids followed by a mature core toxin of 30 amino acids that includes six cysteine residues. A pro-piece (pro-peptide) was absent. As was inferred by similarity comparisons, the cysteine residues impair the structural configuration of the so-called "disulfide through disulfide knot" and allow for the grouping of this peptide into the class of knottins (Fig. 10-A and -B). The *Dinoponera* ICK-like toxin precursor contains a cysteine framework that is typified by VI/VII; i.e., -C-C-CC-C-C-, which is characteristic of the ω-toxin super-family (PFAM Clan CL0083) that encompasses several type-1 conotoxins, funnel web spider toxins and assassin bug toxins, among others. In figure 10-C, ICK-like toxins from *D. quadriceps*, *Conus* species, and spiders are aligned for primary structure comparison. Another type of ICK-folded toxin transcript found in the venom of the giant ant was similar to the so-called cysteine-rich venom proteins that share structural similarities with the toxins in the venoms of distinct species of animals (e.g., contig 1144, table S3).

## Discussion

Ants play significant ecological roles not only in diverse terrestrial environments, particularly in the tropical regions of the planet, but also in extreme biomes. Ants are engaged in several beneficial interrelationships with other organisms that include pollination, microbial dispersal, insect predation, and symbiotic associations with microorganisms [33]. In agriculture, ants modulate and interfere with soil conditions and may cause economic losses due to several factors including damage to the foliage of commercial plants, direct and indirect reductions in crop yields and pest management-related increases in the costs production [34]. With over 12,000 known species, the ants are among the most successful insects in the animal kingdom, and many of these species depend on venom for defense against other organisms, including predators and other ant species [35].

The genus *Dinoponera* inhabits the tropical regions of the New World, and its nests are located in diverse Brazilian biomes including semi-arid and savannah-like vegetation and humid and rain forests. *D. quadriceps* preys exclusively on insects and is normally harmless to humans except when they are disturbed or inadvertently encountered; the resultant interactions may result in accidents of varying levels of severity. Therefore, knowledge of venom composition is desirable not only to comprehend the pathophysiological basis of the effects and symptoms elicited by the sting of *D. quadriceps* but also to characterize proteins and peptides that will potentially be useful in the diagnosis and prospective treatment of chronic human diseases [5,36]. The biotechnological applications of ant venom include not only the development of pharmaceuticals but also the development of bio-insecticides; the latter benefit is exemplified by poneratoxin, a small 25-residue neuropeptide that is active at the Na+- and $K^+$-channels of invertebrates and was originally isolated in the venom of *Paroponera clavata* [29].

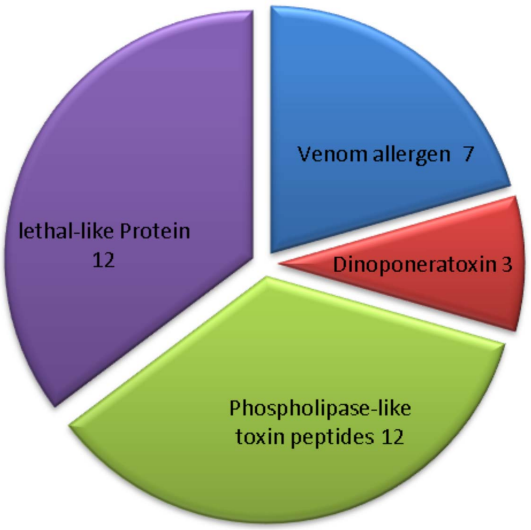



**Figure 5. Proportions of transcript toxin and toxin-related components expressed in the venom gland of *D. quadriceps*.** Toxins and venom-related protein precursors represented less than 1% of all transcripts (18,524) in the giant ant venome. As shown, the major toxin core was composed of four groups of polypeptides.
doi:10.1371/journal.pone.0087556.g005

A robust strategy for the study of the venom profiles of small poisonous creatures with tiny venom glands relies on molecular cloning and DNA/RNA sequencing [37]. Thus, we prepared a *D. quadriceps* venom gland cDNA library and investigated the toxin repertoire of cloned transcribed genes by Sanger and RNA deep sequencing. Hundreds randomly picked clones from the Sanger sequencing are presented in figure 2 and table S1. Regarding the contigs and singlets datasets, proteins classified as "no hits" accounted for ~20% of all ESTs, and dinoponeratoxins accounted for about 5% of all sequences categorized as toxins (Fig. 2-A). In terms of the absolute numbers of individual clones, over 37% of all retrieved random picked sequences were dinoponeratoxins (130 clones out 420) and over 50% of all venom-related sequences (Fig. 2-B). Two novel full-length dinoponeratoxin isoform precursors were identified in EST library (discussed below). Additionally, representative sequences of allergens and other minor components were also annotated and catalogued. However, due to the steep rise in the redundancy in the number of transcript sequences rescued from the cDNA (ESTs) library, we decided to conduct a deep investigation using next generation RNA-sequencing (NG RNA-Seq) to characterize the entire transcriptome of the *D. quadriceps* venom gland. To describe the overall transcript profile, we obtained our sequencing dataset with the Ion Torrent Personal Genome Machine (PGM™) and the Trinity program, which is specially designed for RNA-seq *de novo* assembly. This strategy yielded over 4,000,000 bp, and over 2,000,000 reads with an average length of ~230 nucleotides were deep sequenced (Table S2). These reads generated over 18,500 contigs, of which 6,463 contigs corresponded to BLASTx hits. Most of the expressed sequences in the venom gland encoded proteins related to biological processes, cellular architecture, and molecular functions (Fig. 3 and Fig. S2). Importantly, based on functional classification,

```
Polistes_dominulus        --------MNFKYSILFICFVKVLDNCYAADDLTTLRNGTLDRGITPDCTFNEKDIELHV
Polistes_annularis        ----------------------------------------------MSPDCTFNEKDIVFYV
Vespula_maculifrons       -------------------------------------------GPKCPFNSDTVSIII
Vespula_squamosa          -------------------------------------------GSKCPFSDDTVAMVI
Vespa_crabro              -------------------------------------------FNPCPYSDDTVKMIV
Dolichovespula_maculata   ----------------------------------------GILPECKLVPEEISFVL
Solenopsis_invicta        -------------------------MRKFAAIFVVFFVQCTHLYSLAQARAEPDPGVVEYLK
Dinoponera_quadriceps     YSLLCTMKFITGILVIFCVYLLSTGGDSKILPLKKLPSKIFHLKSHVDNTVKKPLKVFHL

Polistes_dominulus        YSRDKRNGIILKKEILKNYDLFQKS----------------------------QISHQIAI
Polistes_annularis        YSRDKRDGIILKKETLTNYDLFTKS----------------------------TISKQVVF
Vespula_maculifrons       ETRENRNRDLYTLQTLQNHPEFKKK----------------------------TITRPVVF
Vespula_squamosa          VTRENRNRDFYTLQTLRNHDEFKKK----------------------------AITRPVVF
Vespa_crabro              LTRENKKYDFYTLDTIKNHNEFKDT----------------------------ITLKPHVF
Dolichovespula_maculata   STRENRDGVYLTLQKLKNGKMFKNSD---------------------------LSSKKVPF
Solenopsis_invicta        QSCVYGNSSYINVYLYNSRFQGKNLGNQQSCQDI------------------NASLPVVF
Dinoponera_quadriceps     KSHVENSVGPLRMNKLTPNCIFGVKSMSMVLFTKNIPDGKYISLDSDLGRDLDLTKTIYF

Polistes_dominulus        LIHGFLSTGNNENFDAMAKALIEIDNFLVISVDWKKGACNAFASTNDVLGYSQAVGNTRH
Polistes_annularis        LIHGFLSTGNNENFVAMSKALIEKDDFLVISVDWKKGACNAFASTKDALGYSKAVGNTRH
Vespula_maculifrons       ITHGFTSSASEKNFINLAKALVDKDNYMVISIDWQTAACTNEYPGLKYAYYPTAASNTRL
Vespula_squamosa          ITHGFTSSATVESFVDLQTAILEXXXXKVTVSDWRVAACNRTTG---LLYYVTAVSNTRL
Vespa_crabro              ITHGFTSSATAENFVVMAKALLDKGNYLVILTDWRMAACTNEIAGLKLAYYPYAASNTRL
Dolichovespula_maculata   LIHGFISSATNKNYADMTRALLDKDDIMVISIDWRDGACSNEFALLKFIGYPKAVENTRA
Solenopsis_invicta        ITHGFTSSAQVSTFKDLANAFVQKG-HTAFIVDWSEAACTDGLPGVQFAEYNAAASNTYD
Dinoponera_quadriceps     TGHGFISNVNHSLSNRLSRALVEKD-YTVFSLDWSDAACTTGGLPLVKLLYPSGVQNTRE

Polistes_dominulus        VGKYVADFTKLLVEQYKVPMSNIRLIGHSLGAHTSGFAGKEVQRLKLGKYKEIIGLDPAG
Polistes_annularis        VGKFVADFTKLLVEKYKVLISNIRLIGHSLGAHTSGFAGKEVQKLKLGKYKEIIGLDPAG
Vespula_maculifrons       VGQYIATITQKLVKDYKISMANIRLIGHSLGAHVSGFAGKRVQELKLGKYSEIIGLDPAR
Vespula_squamosa          VGRYIATVTKKLVTDYNVSMADIRLIGHSLGAHVSGFAGKEVQKLKLEKYSEIIGLDPAG
Vespa_crabro              VGNYIATVTKMLVQKYNVPMANIRLIGHSLGAHISGFAGKKVQELGLGKYPEIIGLDPAG
Dolichovespula_maculata   VGKYIADFSKILIQKYKVLLENIRLIGHSLGAQIAGFAGKEFQRFKLGKYPEIIGLDPAG
Solenopsis_invicta        IGQLMAKYTVDLMNKCKIPLNNIQYVGHSLGSHVCGFAAKHVKKLINKTMPYILALDPAD
Dinoponera_quadriceps     IGNLMADYVMSLIDHG-ASLRNMGFIGHSLGSHVC-FASKKIYESGYGKVPLLFAADPAQ

Polistes_dominulus        PSFLTNKCPNRLCETDAEYVQAIHTSAILGVYYNVGSVDFYVNYGKSQPGCSEP----SC
Polistes_annularis        PYFHRSDCPDRLCVTDAEYVQVIHTSIILGVYYNVGSVDFYVNYGKNQPGCNEP----SC
Vespula_maculifrons       PSFDSNHCSERLCETDAEYVQIIHTSNYLGTEKILGTVDFYMNNGKNNPGCGRF-FSEVC
Vespula_squamosa          PSFESNDCAERLCKTDAHYVQIIHTSKKFGIEKSIGHVDFYVNQGNNQPGCGIIPLKDVC
Vespa_crabro              PSFKSNDCSQRICETDANYVQIIHTSNRLGTERTLGTVDFYMNNGYNQPGCGLPIIGETC
Dolichovespula_maculata   PSFKKKDCPERICETDAHYVQILHTSSNLGTERTLGTVDFYINDGSNQPGCTYI-IGETC
Solenopsis_invicta        PSFGSNKCGERICKSDAKRIVVFKTSILGIGENIIGHLLIVFDGGKSQPACSWY--DVPC
Dinoponera_quadriceps     PLFQLKQCPDRLCDTDGKLVITLHTSQIG-LGYPIGGLDLYFN-GFVQPKCHLD---ITC

Polistes_dominulus        SHTKAVKYLTECIKRECCLIGTPWKSYF------------STPKPISQCKRDTCVCVGLN
Polistes_annularis        SHTKAVKYLTECIKHECCLIGTPWKKYF------------STPKPISQCRGDTCVCVGLN
Vespula_maculifrons       SHTRAVIYMAECIKHECCLIGIPRSK---------------SSQPISRCTKQECVCVGLN
Vespula_squamosa          SHSRAITYMTECIKRECCLIGIPQSK---------------SSKSISSCTRQECVCVGLK
Vespa_crabro              SHTRAVKYFTECIKHECCLIGVPKSK---------------NPQPVSKCTRNECVCVGLN
Dolichovespula_maculata   SHTRAVKYLTECIRRECCLIGVPQSK---------------NPQPVSKCTRNECVCVGLN
Solenopsis_invicta        SHSESIVYATGMVSGRCQHLAVPWTAQQRINPIQWKFWRVFTSNIPAYPTSDTTNCVVLN
Dinoponera_quadriceps     GHIRSVLYLINMVEKKCSFPGIPGTYKQILN----------PFSKFPYPNSKTTDCFVMD

Polistes_dominulus        AQSYPAK--------GSFYVPVDKDAPYCHNEGIKL--
Polistes_annularis        AKSYPAR--------GAFYAPVEANAPYCHNEGIKL--
Vespula_maculifrons       AKKYPSR--------GSFYVPVESTAPFCNNKGKII--
Vespula_squamosa          AKSYPNT--------GSFYVPVESTAPFCNNKGKII--
Vespa_crabro              AKTYPKT--------GSFYVPVESKAPYCNNKGKII--
Dolichovespula_maculata   AKEYPKK--------GSFYVPVEAKAPFCNNNGKII--
Solenopsis_invicta        TNVFKND--------NTFEGEYHAFPDCARNLFKCRQQ
Dinoponera_quadriceps     DSIFNPRRKSLQNLAGIYYMFVDPDTFCTRKNFNCQRV
```

**Figure 6. Multiple alignments of the deduced amino acid sequences of venom allergen I (Sol i 1) from *D. quadriceps* with known sequences from different ant and wasp species.** The identical and conserved amino acid residues of diverse ant and wasp hymenopterans are highlighted in black and gray, respectively. Dots represent gaps.
doi:10.1371/journal.pone.0087556.g006

the most abundant proteins (over 200) were related to ribosomal structure and biogenesis, translational and post-translational modification, protein turnover and chaperones (Fig. S3), as it would be expected for a specialized and dynamic metabolic active tissue that needs to synthesize venom components for ant defense and feeding. Another interesting finding from the giant ant transcriptome is related to the proportion of Hymenoptera species with which *D. quadriceps* sequences share high similarity based on BLASTx queries. Nearly 60% of all *D. quadriceps* transcript sequences had the greatest homology with sequences from *H. saltator*, followed by *A. echinator* (~11%), *S. invicta* (~7%) and *M. rotundata* (~6%) (Fig. 4). As *D. quadriceps* and *H. saltator* belong to

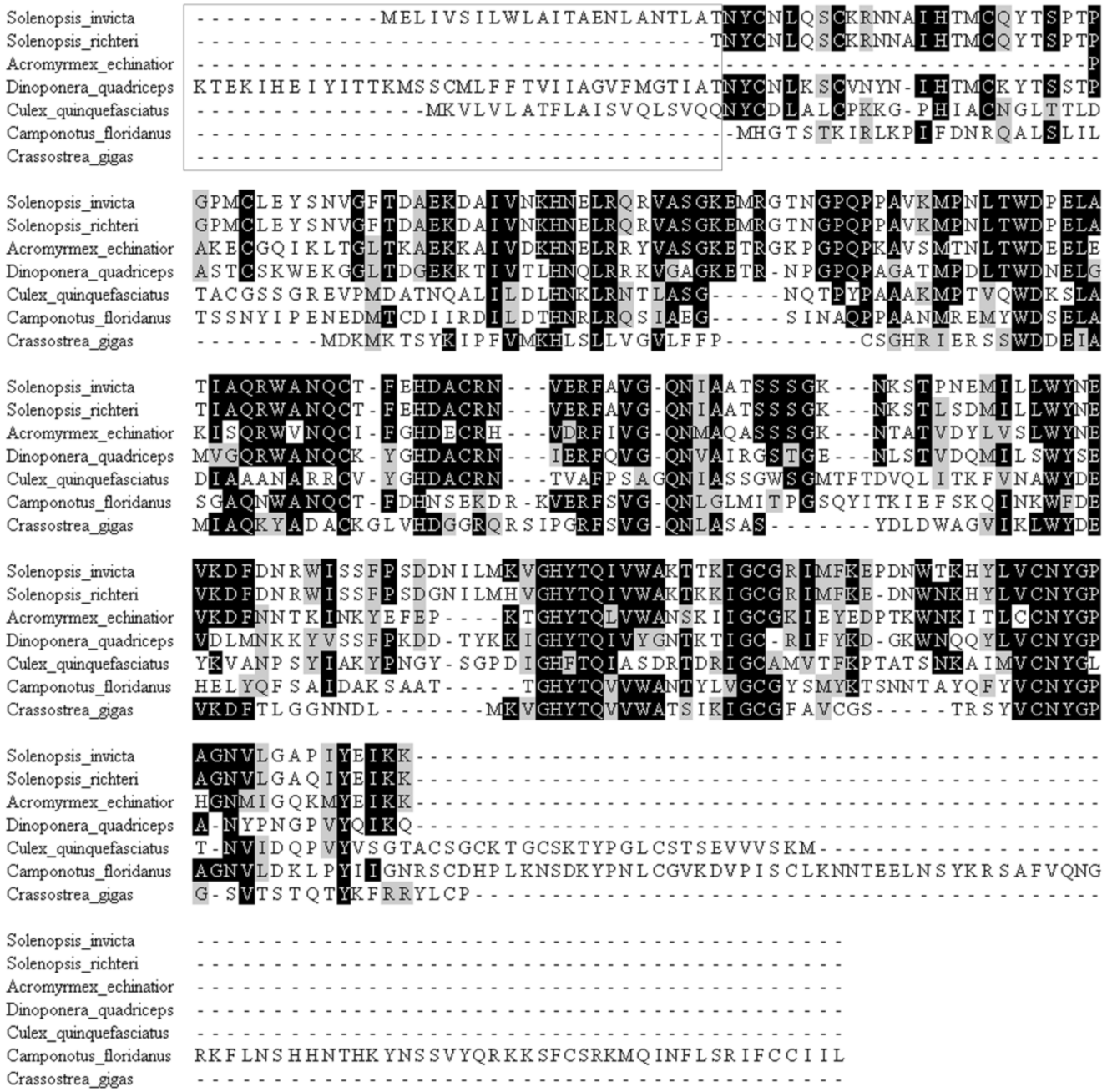



**Figure 7. Multiple alignments of the deduced amino acid sequences of the venom allergen antigen 5 (Sol i 3) of *D. quadriceps* with known sequences from different Hymenoptera species.** The conserved regions of the Sol i 3/allergen antigen 5 (Ag 5) polypeptide sequences of *D. quadriceps* and other hymenopterans are presented in black boxes. Conserved amino acid residue mutations are highlight in light gray boxes. Dots represent gaps. The unique amino acid sequences of the *D. quadriceps* allergen homologous to Sol i 3 protein are displayed in open boxes.
doi:10.1371/journal.pone.0087556.g007

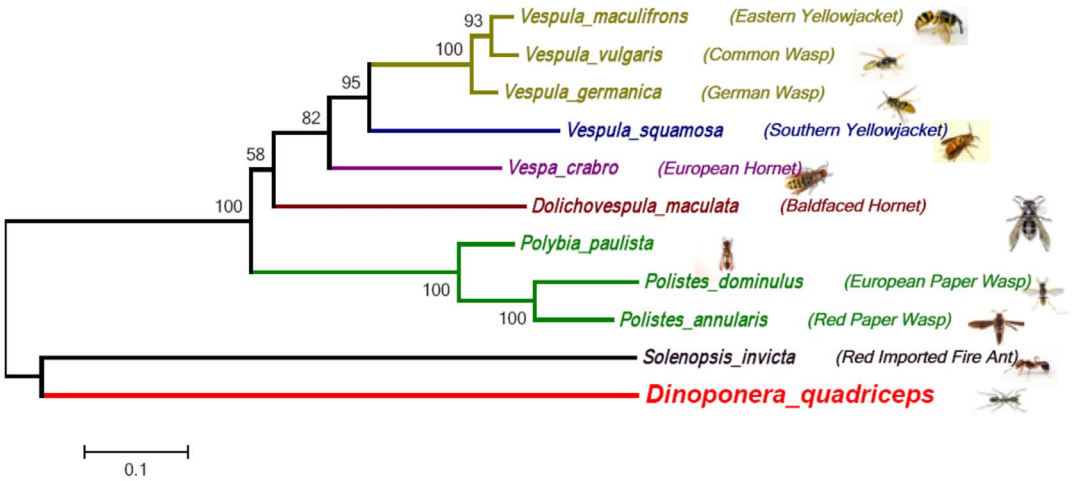



**Figure 8. Comparison of Sol i 3/Allergen Antigen 5 polypeptides from *D. quadriceps* with their hymenopteran orthologs.** Phylogenetic tree based on neighbor-joining analyses of a concatenated alignment of allergen 5/Sol i 3 and the orthology relationships of multiple insects. The scale bar indicates 0.05 substitutions per site. *Vespula maculifrons*, *Vespula vulgaris*, *Vespula germanica*, *Vespula squamosal*, *Vespa crabro*, *Dolichovespula maculate*, *Polybia paulista*, *Polistes dominulus*, *Polistes annularis*, and *Solenopsis invicta*.
doi:10.1371/journal.pone.0087556.g008

the same subfamily (Ponerini) and evolved from a common ancestor, it is reasonable to argue that they have a common framework of transcript sequences in their venom glands. Two groups of highly homologous proteins expressed by hymenopterans in the venom gland transcriptome of *D. quadriceps* that were disclosed here are related to "sex determination" and mixed-function "lethal-like" proteins.

The so-called "sex determination proteins" were first discovered in *Drosophila* [38] and thereafter shown to be present in ant species, such as *A. echinatior*, *H. saltator* and *C. floridanus* (Figure S4). This group of proteins belongs to the BTB-ZF family of transcriptional regulators, which are part of the somatic sex determination hierarchy [39,40]. These proteins can switch fruitless gene (fru) splicing from the male-specific pattern to the female-specific

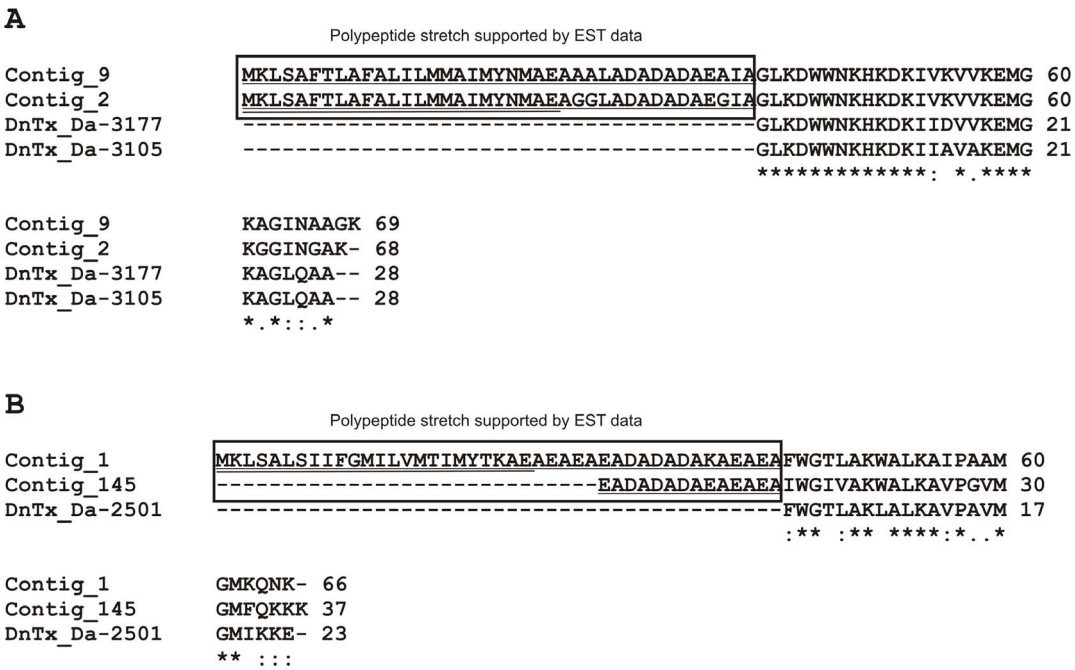



**Figure 9. Multiple alignments of deduced amino acid sequences of different identified dinoponeratoxins from the *D. quadriceps* venom gland transcriptome and *D. australis*.** Deduced *D. quadriceps* dinoponeratoxin cDNA precursor sequences (contig_1 and contig_9). RNA deep sequencing contigs (contig_2 and contig_145) were compared to mature peptide sequences from *D. australis* (Da-3177 and Da-3105) (part A) and from another species of *Dinoponera* (Da-2501) (part B). ClustalW was used to multi-align the sequences. Identical amino acid residues are marked with asterisks. Stretches of deduced amino acid sequences supported by EST sequences are boxed. Signal peptides (pre-peptides) are doubled underlined, and pro-regions of pro-peptides are shown with a single line under the sequence. Contigs 1 and 9 were first identified in the EST library, and contigs 2 and 145 came primarily from RNA-Seq.
doi:10.1371/journal.pone.0087556.g009

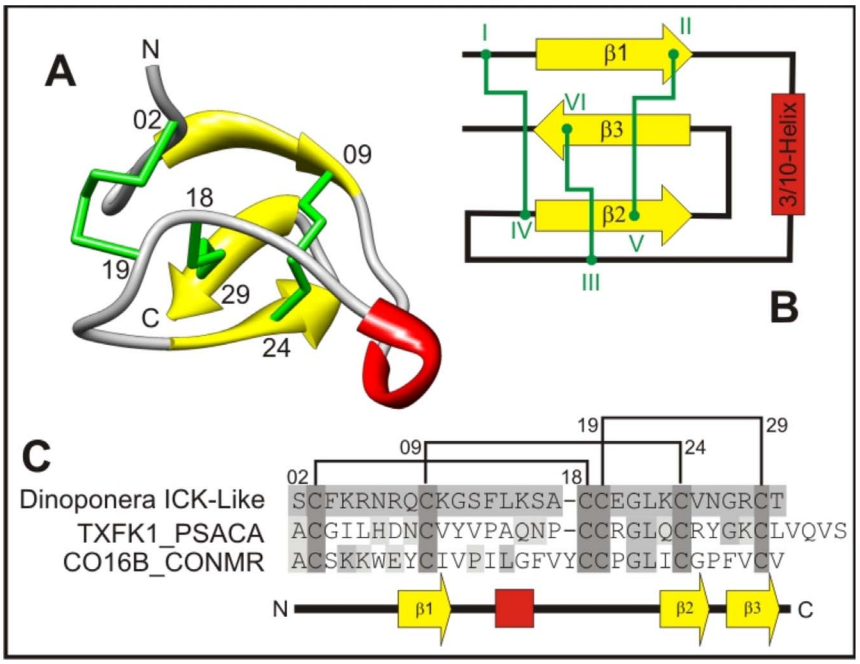



**Figure 10. The Inhibitor cystine knot- (ICK-) toxin from *D. quadriceps* venom.** A) *Dinoponera ICK-like* tridimensional homology model from contig 05 (Table S1) mature peptide (yellow – β-sheet; green – cysteine; ted – 3/10-helix. N-terminus (**N**); C-terminus (**C**). The numbers indicate the cysteine positions in the *Dinoponera* sequence. The presence of a 'disulfide through disulfide knot' structurally defines this peptide as a knottin. B) Schematic representation of a knottin obtained when one disulfide bridge crosses the macrocycle formed by two other disulfides, and the interconnecting backbone (disulfide III–VI) goes through disulfides I–IV and II–V. C) Alignments of the *Dinoponera ICK-like* mature peptide with the tarantula and Conus snail sequences of the templates used in the homology model showing the highly conserved cysteine framework [C-C-CC-C-C] that is characteristic of the omega-toxin-like family. **TXFK1_PSACA**: psalmopeotoxin I (PcFK1) from the venom of the tarantula *Psalmopoeus cambridgei* (NMR structure 1X5V; UNIPROT Accession P0C201). **CO16B_CONMR**: Mu-Conotoxin MrVIB from conus snail *Conus marmoreus* (NMR structure 1RMK; UNIPROT Accession Q26443).
doi:10.1371/journal.pone.0087556.g010

pattern through the activation of the female-specific fru 5′-splice site [41–42]. Our analysis of the molecular phylogeny of sex determination proteins indicated that the amino acid sequences of the members of this class of proteins are phylogenetically related to the insect clade and, conversely, that the orthologous relationship is closely related to insect species (Figure S4), which suggests a common mechanism of sex determination in hymenopterans. Moreover, the sex determination proteins are usually expressed in specific regions of the adult male brain that are associated with the control of courtship songs and the steps of male courtship and are also expressed in the larval and pupal male mushroom bodies and optic lobes and thereby possess tissue-specific characteristics [38,39]. Importantly, our results have revealed for the first time that this group of proteins is expressed in giant ant venom glands. Further investigation is required to understand the role of the orthologous sex determination proteins in the venom gland, although the presence of this group of proteins might suggest that they influence courtship behavior and social organization in the queenless nests of the giant ant species.

Several classes of polypeptides with multiple functions are grouped together under the term "lethal-like proteins", despite the fact that these proteins have distinct structures and protein motifs (Fig. 5 and Table S2). For example, we found transcript that coded for *protein lethal essential for life*, *multidrug resistance-associated protein lethal*, *lethal neighbor of tid protein*, and *lethal malignant brain tumor-like protein*, among other lethal-like sequences. In the category of *protein lethal essential for life*, we found sequences that possessed the signatures of small heat shock proteins (sHSPs) and **α**-crystallins (**α**C); i.e., domain sHSP/**α**C. Polypeptides with these domains, apart from being responsive to a variety of stresses, are involved in a wide range of functions in multiple cell types and multiple organisms [43]. In the venom glands of ants, the *protein lethal* with domain sHSP/**α**C proteins might work as co-chaperones that assist in toxin folding and post-translational processing of key proteins. The other lethal-like or lethal-related proteins, namely, the *lethal neighbor of tid proteins*, contain the typical domain of ALG3 proteins; i.e., a structural motif dedicated to the enzymatic linkage of N-glycosidic groups to glycoproteins (pfam05208). The catalytic process of glycosilation post-translationally modifies numerous toxins including ant venom phospholipases [44]. Lethal-like proteins are also associated with multidrug resistance. A number of transcripts code for *multidrug resistance-associated protein lethal* proteins, which include the ABC transporter cassette motif in their structures. In ants, these polypeptides might be associated with the active transport of glycoproteins and xenobiotics to sub-cellular compartments like those that have been observed in several organisms and described in more detail in *Drosophila* Malpighian tubule physiology [45,46]. Another lethal-related protein class in the giant ant transcriptome includes the *lethal malignant brain tumor-like protein*. The members in this class of proteins contain the domain of the MBT superfamily (smart00561) as exemplified by proteins of *Drosophila* and vertebrates that act in the process of transcriptional regulation. Additionally, proteins in this class possess the "sterile alpha motif (SAM)", which interacts with diverse proteins, RNAs and membrane lipids, and plays a role in signal transduction and transcription. Thus, the categorization of lethal-like protein in the venomics of the giant ant relies on the fact that members of this group of polypeptides participate in critical

biochemical events of venom gland physiology and might additionally induce disruption of the cellular circuitries of prey or victims after ant sting and envenomation.

Notably, less than 1% (exactly 0.72%) of all sequenced transcripts corresponded to toxins or venom-related peptides in hymenopterans (Fig. 3), and this percentage corresponded with hundreds of polypeptides. Accordingly, Johnson and colleagues [30] performed a proteomic-based investigation on the venom of *D. australis* and found over 75 polypeptides, but only described the six most abundant dinoponeratoxins and their congeners. In another recent study on the venom transcript profile of ants, Bouzid and collaborators [2] identified five major classes of toxins and toxin-related peptides (i.e., pilosulin 5-like, insect cytokine precursor, Sol I 3 antigen, secapins and allergen peptide) that corresponded to a total of 42 of the 85 ESTs that were Hymenoptera toxin-like candidates.

Despite of relative simplicity of ant venoms, which are comprised of a basic framework of allergen peptides, hyaluronidases, phospholipases and phosphatases, the small repertoire of venom-related components is highly effective in causing tissue permeability, edema, warmth of the local area, life-threatening systemic anaphylactic reactions, and, occasionally, delayed hypersensitivity; these reactions are observed after most Hymenoptera stings [47]. Moreover, due to the high levels of structural conservation and homology among hymenopterans allergens, venom cross-reactivity is one of the major risk factors related to insect stings [48].

In our transcriptome analysis, we determined that the giant ant venom gland possesses a relative simple toxin and venom-related peptide repertoire, the sequences of which predominantly fit into the following three major venomic classes: esterases (phospholipase-related polypeptides and carboxylesterases), venom allergens, and dinoponeratoxins (Fig. 2-B and Fig. 5; tables S1 and S3). Minor venom components, which are not less toxic or harmless to human cells and tissue, included cysteine-rich peptides stabilized by disulfide bonds that are distributed across the venoms from animals of multiple phyla.

In table 2, the general venom polypeptide cores from *D. quadriceps* and several species of ants are shown and compared. There are venom components common to all species (e.g., phospholipases) and others, such as venom allergens and dinoponeratoxin, which are genus-specific. However, given that most of the common venom components share high levels of amino acid and epitope similarity, cross reactivity and anaphylaxis after Hymenoptera stings is a recurrent risk, as previously mentioned. This is a particular concern for phospholipases and other major venom allergens, which exhibit the highest levels of similarities in their primary structures and tri-dimensional topologies.

Notably, two of the most conserved venom allergens from fire ants (*S. invicta*) are also expressed in the venom gland of *D. quadriceps*: the homologous of Sol i 1 and Sol i 3 (Fig. 6 and Fig. 7, respectively). *D. quadriceps* venom allergen 1 belongs to the phospholipase $AB_1$ group and, consequently, displays a propensity for cross-reactivity with the Sol i 1 homologs of several hymenopteran-related sequences [49]. Allergen proteins are potent components of the Hymenoptera venom arsenal that induce hypersensitive allergic responses. As can be observed in figures 7 and S5, the structural conservation of the *D. quadriceps* Sol i 3/allergen 5-like transcript is not only related to sequences from other species of ants but also to sequences from wasp species. Indeed, these sequences are collectively called CAP proteins (**c**ysteine-rich secretory proteins - CRISPs, insect venom **a**llergen antigen 5, and **p**athogenesis-related 1 proteins), and they compose a superfamily of polypeptide sequences that are found in a wide range of organisms, including prokaryotes. In figure S5, the *D. quadriceps* CAP allergen Sol i 3/allergen 5-like is compared with homologous sequences from *S. invicta* and *Vespula vulgaris*, and its tri-dimensional model depicted. The spatial structure of *D. quadriceps* Sol i 3/allergen 5-like proteins extremely highly conserved and fits well with fire ant Sol i 3/Ag 5 experimental crystal structure data [50]. Interestingly, polypeptides in the mammalian male reproductive tract and in the venom secretory ducts (e.g., stecrisp, PSTx, triflin, helothermine) of several snake and lizard species are CAP proteins in the CRISP group. CRISPs have two domains: A CAP N-terminal domain followed by a C-terminal cysteine rich domain (CRD) containing 10 absolutely conserved cysteine residues. Importantly, from the perspective of venomics, the CRD possesses a fold in common with sea anemone ion channel inhibitors (Bgk- and Shk-type toxins), which are involved in the blocking of voltage gated $Ca^{2+}$- and $K^{+}$-channels and modulate ryanodine receptors. Interestingly, these biological activities are only observed when these polypeptides are associated with the CAP domain [51,52]. Despite belonging to CAP superfamily, the Sol i 3/allergen antigen 5 (Sol i 3/Ag 5) homologs from *Hymenoptera* venom form a major and distinct clade. Sol i 3/Ag 5 homologs, besides having been initially characterized in the venoms of fire ants and several species of wasps, have also been identified in the midgut of *Drosophila* and in the saliva of ticks, sand flies, and mosquitoes. In the secretions of blood-feeding insects, the Ag5 proteins are part of a cocktail of salivary proteins that are believed to be active either in the suppression of the host immune system or in the prevention of clotting to prolong feeding [52]. Thus, the structural conservation of these allergens in the venom of *D. quadriceps* (Ponerinae), other ant species (e.g., *Pachycondyla* and *S. invicta*) and wasps is an interesting aspect of Hymenoptera venomics because these insect groups have a long natural history of divergence that dates to more than 150 million years ago [35].

In addition to the venom allergen homologous of Sol i 1, which are related to the group of phospholipase $AB_1$,esterases (phospholipases and carboxylesterases) comprises another class of predominantly expressed venom components in the transcriptome of the giant ant *D. quadriceps*. We found members and isoforms of $PLA_1$, $PLA_2$, $PLB_1$, PLD, which indicates that phospholipases compose an important part of the toxin core of giant ant venoms (Fig. 5, table S3). Broadly, the class of esterase-lipases comprises a superfamily of enzymes that includes diverse groups of phospholipases (A1, A2, B, C and D). The most well-known phospholipases, $A_1$, $A_2$ and C ($PLA_1$, $PLA_2$ and PLC), hydrolyze the ester bonds of phospholipids at the specific positions of *sn-1*, *sn-2* and *sn-3*, respectively [53–55], and the resulting products are used as precursors in the synthetic pathways for pharmacological mediators such as leukotrienes and prostaglandins or play a role in signal transduction as secondary messengers [56]. $PLA_1$s can be classified into two groups based on their structures and cellular localizations. The first group encompasses intracellular enzymes that include phosphatidic acid-preferring phospholipase $A_1$, and the second group is composed of other extracellular enzymes that include phosphatidylserine-specific $PLA_1$ [57]. Members of the latter group of enzymes have been detected in hymenopterans, mainly in wasp species. The structures of these enzymes are conserved across a wide range of organisms, and, in humans, they induce platelet activation, allergic reactions and hemolysis [54,58,59]. Phospholipases $A_2$ is subdivided into more than a dozen subtypes (from I to XV, with subgroups) and is the most abundantly expressed type of phospholipase in the venom of hymenopterans. The venoms of bees and wasps have been shown

**Table 2.** Overview of major venom polypeptide components of *D. quadriceps* and other ant species.

| Subfamily | Genus and species | Major (core) venom polypeptides | Reference |
|---|---|---|---|
| **Ponerinae** | *Dinoponera quadriceps* | Dinoponeratoxins, allergens (homologous to Sol i 1/$PLA_1B$, Sol i 2/4 and Sol i 3/Ag 5), esterases (phospholipases A and B, acid phosphatases, carboxylesterase) | This work |
| | *Pachycondyla sp.* | 16 and 24 kDa (Pac c 3/Sol i 3 allergen) IgE-binding proteins | [44] |
| | *Harpegnathos saltator* | Esterases | [75] |
| **Myrmicinae** | *Solenopsis invicta* | Allergens (Sol i 1, 2, 3 and 4), phospholipases (including Sol i 1/$PLA_1B$) | [44] |
| | *Tetramorium bicarinatum* | Pilosulin-like peptides (1,3, 5), allergen (Sol i 3/Ag 5), insect cytokine precursor uENF2, metalloproteases | [2] |
| | *Acromyrmex echinatior* | Carboxylesterase-3, Esterase FE4, Phospholipase A1, | GenBank GI68854, GI68747 |
| **Formicinae** | *Camponotus floridanus* | Carboxylesterase-3 | [75] |
| **Myrmeciinae** | *Myrmecia sp.* | Phospholipase $A_2$, Phospholipase B, Hyaluronidase, Acid and Alkaline phosphatases | [92] |

doi:10.1371/journal.pone.0087556.t002

to possess high levels of phospholipase activity; thus, this class of enzymes has been implicated in tissue damage and hypersensitivity reactions [60–62]. The pioneering report of Schmidt and Blum [63] detailed the high phospholipase $A_2$ and B contents of the venom of the Harvester ant (*Pogonomyrmex badius*). Since that early work, phospholipase activity has been described in the venoms of several ant species, including *Neoponera apicalis*, *N. olivaceae*, *P. clavata*, *Pachycondyla cressiondes*, *Pogonomyrmex occidentalis* and *Dinoponera grandis* [53]. These enzymes have been related to be the major allergen in hymenoptera venoms [49,64,65] and, notably, are responsible for the antigenic cross-reactivity between hymenoptera venoms [49,60,66,67].

PLCs are ubiquitous enzymes that hydrolyze membrane lipid phosphoinositides to yield two important second messengers, inositol phosphates and diacylglycerol (DAG) (cl14615); however, despite finding numerous types of phospholipases, we did not find any PLC venom in our study of giant ant venomics, Two additional subfamilies of phospholipases should be mentioned: (1) phospholipase B, which has both esterase and phospholipase-A/lysophospholipase activity (cd01824) and is involved in the conversion of phosphatidylcholine to fatty acids and glycerophosphocholine (perhaps in the context of dietary lipid uptake); and (2) phospholipase D (PLD), which is a signal-activated enzyme that catalyzes the hydrolysis of the phosphodiester bond of glycerophospholipids to generate phosphatidic acid (PA); PA is an intracellular signaling lipid that has been implicated in vesicular trafficking [68,69]. In the *D. quadriceps* venome, we found transcripts for the first giant ant PLD (sphingomyelinase-D, E.C. 3.1.4.41), a dermonecrotic toxin that, until now, had only been described venoms of spiders, particularly the brown spider *Loxoceles gaucho*. In spider venom, PLD is responsible for the necrotic arachnidism and massive inflammatory response observed in loxoscelism, which is the main effect of brown spider envenomation [70,71]. Thus, the presence of PLD in the *D. quadriceps* venome is another remarkable finding that advances the comprehension of the pathophysiological effects of the venom of the giant ant. Moreover, the expression of several isoforms and subgroups of phospholipases in the giant ant venom gland indicate that lipid homeostasis and the lipid-mediated signaling circuitry are targets that mediate the tissue disruption of, and damage to, prey and/or victims.

We also identified transcripts for carboxylesterases in the venom of *D. quadriceps*; carboxylesterases are enzymes that hydrolyze carboxylic acid esters into their corresponding acids and alcohols (GO0004091). Enzymes in this class are considered to be protective molecules because promote cellular detoxification by inactivating toxicants and carcinogens. As described by Hatfield and Potter [72], diverse molecules such as drugs, pesticides and veterinary products contain ester moieties that are susceptible to catalytic conversion by these enzymes. Interestingly, carboxylesterases have been identified in diverse hymenoptera venoms, including the venoms of *Apis mellifera* [73], *N. vitripennis* [74], *H. saltator*, *C. floridanus* [75], *A. echinatior* (GI332028825). Thus, it is likely that such this class of venom components might be involved in the detoxification xenobiotics and confer resistance to insecticides [74,75]. Despite this physiological function, the *A. mellifera* carboxylesterase has been shown to cause allergic reactions in humans [76], and this type of venom component might make a similar contribution to the symptoms of by giant ant stings.

Dinoponeratoxin transcripts were identified in high proportions in the transcriptome of *D. quadriceps*, and four novel dinoponeratoxin sequences were discovered (Fig. 9). As noted by Johnson and co-workers [30], dinoponeratoxins share variable primary sequence similarities with antimicrobial peptides from ponerine ants (i.e., ponericin G and W3) and other organisms; e.g., frogs (gaegurin-5 and brevinin-1 PTa). Additionally, it is notable that bioactive peptides seem to be encrypted into the dinoponeratoxin precursors, as numerous peptide fragments were detected in the venom proteomic of *D. australis*. The alignment of dinoponeratoxins with antimicrobial peptide sequences provides a clue regarding the functions of these peptides in the venom of *D. quadriceps* and other ponerine ants, although the actual biological activity of dinoponeratoxin - other than microbicide - deserves further investigation. It is known that many antimicrobial peptides cause cellular membrane damage via pore formation or micellization as a part of their modes of action [77,78]; these processes might intensify toxin dispersal in the tissues of prey or victims and increase the efficacy of ant envenomation. Another interesting aspect of dinoponeratoxin transcripts is related to their cDNA precursors at the nucleotide level. We observed that the dinoponeratoxin (contig 9, table S1) precursor shares a certain degree of identity with the 5′-end of the pilosulin 5 cDNA precursor (not shown). First discovered in the venom of Australian jumper jack ant (*Myrmecia pilosula*), pilosulins comprise a family of short monomeric and dimeric peptides with antimicrobial activity that is mediated through histamine-releasing and hemolytic effects [79,80]. In light of toxin evolution and diversification, the sharing of stretches of nucleotide sequences between dinoponeratoxin and pilosulin-5 is indicative of the conservation of parental ancestor

genes and their recruitment as a template for the evolution of bioactive peptide variants in the venom that, in this case, cause the degranulation of mast cells and histamine release. Another pilosulin-related sequence of giant ant venom gland was observed to be encoded by the contig 6 (table S1).

Minor toxin-related components identified in the venom gland transcriptome of *D. quadriceps* include sequences with high homologies with the spider U8 agatoxin, the venom protein VN50, the cysteine-rich venom proteins of the endoparasitic wasp, and the ICK-like toxins from *Conus* and spiders. The U8-agatoxin-Ao1a-like isoform peptide precursor was found in an initiative to identify toxin-like structures based on new algorithms for the interrogation of the cDNA library of the spider *Agelena orientalis* and the data bank for neurotoxin sequences [81]. Based on this strategy, a contig (1144_B2-2, table S3) from the *D quadriceps* transcriptome was found to meet the criteria for the "extra structural motif" (ESM), in which a pair of CXC fragments in the C-region (or C1CZC1C, where Z is variable number of amino acid residues), code for a novel neurotoxin precursor that is homologous to spider toxin and diverse hymenopteran-related sequences (Fig. S6).

The venom protein VN50 is a serine protease homolog without enzymatic activity from the venom *Cotesia rubecula* that, once injected into the host, interferes with the immune defense mechanisms mediated by the prophenoloxidase (proPO) activation cascade [82]. The presence of a VN50-related sequence (contig 1219_B2-2) in the venom gland of *D. quadriceps* indicates that this venom component has the ability to disturb the homeostasis of invertebrates and augment the processes of their intoxication. Another less abundant, but not less important, transcript sequence in the *D. quadriceps* transcriptome (contig 23_B2-2) shares high similarity with the cysteine-rich venom proteins that make up a significant fraction of the polypeptides from the venom of the parasitoid wasp *Pimpla hypochondriaca* [83]. Venom peptides in this class include trypsin inhibitors (including the Kunitz type) and voltage-gated $Ca^{2+}$-channel antagonists (ω-conotoxins) with analgesic properties from the gastropod mollusk sea snail *Conus*. A structural fold that is highly conserved across diverse bioactive venom cysteine-rich peptides (or disulfide-rich domains) is known as the inhibitor cysteine-knot (ICK) motif. The ICK motif is an evolutionarily conserved structure that is found in the venoms of a wide range of organisms of diverse phyla, including arthropods (spiders and scorpions), marine cnidarians (sea anemones), sea snails, fungi, plants, and vertebrates. The ICK motif is a chemically and thermally very stable protein fold that is characterized by a knot formed by segments (α-helix, β-sheet or random coil) connected by two disulfide bonds and interlaced with a third disulfide bond that forms up to six loops with variable sequences [84]. Venom from the marine mollusk *Conus*, spiders and scorpions include numerous toxins containing the ICK fold. In the transcriptome of *D. quadriceps*, we identified the first ponerine ant ICK-containing toxin (Fig. 10). The BLAST hits with the highest scores for this novel sequence were for the superfamily-A conotoxins and the scorpion like-toxins. The high scores between this novel sequence and members of these protein families were mainly due to similarities with the signal peptides of the *Dinoponera* ICK-like toxin precursors. However, the *Dinoponera* ICK-like toxin cysteine framework was proven to be VI/VII (C-C-CC-C-C), and this pattern of disulfide bonds would place this ant ICK-like toxin into the ω-toxin superfamily (PFAM Clan cl0083). This superfamily includes several type-1 conotoxins, funnel web spider toxins, and assassin bug toxins, all of which are characterized as neurotoxic antagonists of several subtypes of ion channel [85]. Curiously, the translated sequence of the *Dinoponera* ICK-like toxin also has structural homology with an uncharacterized protein family expressed that is in *Schistosoma* (result not shown). Moreover, ICK-type peptides with antimicrobial have recently been identified in the genomes of seven ant species; i.e., *A. cephalotes*, *A. echinatior*, *L. humile*, *S. invicta*, *P. barbatus*, *C. floridanus*, and *H. saltator*. Based on molecular phylogenetic analyses, these ant antimicrobial peptides with ICK motifs have been were categorized into three subtypes [86]. Thus far, transcripts encoding cysteine-rich peptides with variable numbers of disulfide bonds and cysteine stabilized ICK-folded peptides form a minor proportion of all expressed components of *D. quadriceps* venom.

## Conclusions

In this study, we sequenced and characterized the transcriptome of the *D. quadriceps* venom gland and focused principally on the polypeptide core of the venom that is responsible for the symptoms caused by giant ant stings. With a combination of Sanger sequencing of the cDNA library and next generation deep RNA sequencing, a complete description of transcriptome of the venom gland was achieved. These experiments revealed the expression of coding sequences dedicated to specialized tissue physiology and toxin production. The polypeptide component framework of the venom is comprised of a major and a minor core of toxin classes, which, in combination, elicit local and systemic reactions to stings from this species of ponerine ant. The major toxin core is composed of venom allergens, phospholipases, carboxylesterases, lethal-like toxins, and dinoponeratoxins. With exception of dinoponeratoxins, all components of the major toxin core of the giant ant display high structural similarities with toxins of the same groups of proteins from different hymenopterans species (ants and wasps). This finding indicates that cross-reactivity is a recurrent phenomenon of hymenopteran envenomation and a property of the structures of venom components that has been maintained over the course of more than 100 million years of insect evolution and diversification. Moreover, the toxins that compose the major core may functionally interfere with lipid homeostasis and signaling, the trafficking of vesicles and protein-protein and protein-nucleic acid interactions. These toxins may also disrupt the biochemical circuitry of the tissues of the prey or victims. Dinoponeratoxins are another major toxic component of *D. quadriceps* venom and are exclusively expressed in the giant ant venom. This group of small disulfide-less toxins shares structural similarity with sequences from *D. australis* and, at the nucleotide level, shares an evolutionary relationship with pilosulins, which are antimicrobial and mast cell degranulating peptides from *Myrmecia* venom. Cysteine rich polypeptides that are shaped by disulfide bonds and folded into the typical ICK-motif comprise a fraction of the minor *D. quadriceps* venom protein core. The ICK-type peptides from the giant ant possess structural identity with peptides from scorpions, spiders and the marine venomous *Conus* gastropods. A component of the minor venom protein core of *D. quadriceps* is the homolog of a serine protease that is devoid of catalytic activity (i.e., the peptidase S1-derived scaffold venom protein of the solitary wasp and blood-feeding bat), which should interfere with innate immune responses and inhibit endogenous peptidase S1 homologs of the prey or victims. Taken together, the results of the transcriptome and venomic analysis of *D. quadriceps* provide an inestimable resource for the understanding of the molecular and physiopathological bases of giant ant envenomation and basic knowledge about the major and minor polypeptide venom components that may be useful in the future for medical biotechnology.

## Methods

### *Dinoponera quadriceps* sampling and processing

Specimens of the giant ant *Dinoponera quadriceps* were collected at twilight in the Serra de Maranguape (3°52′51″S and 38°42′39″W), which is a mountain range with peaks below 1000 m that is located about 40 km from the coast of the north-eastern state of Ceará, Brazil. A trap made with ground tuna meat was used to attract ants from inside their nests. The individuals were either kept in the laboratory in artificial nests or dissected with thin scissors and scalpels under a stereomicroscope (Nikon SMZ 800, Tokyo, Japan). The resting (not milked) venom gland was carefully separated from the other surrounding tissues, such as the Malpighi tubules and muscles, and quickly transferred to microtubes containing RNAlater solution (Life Technologies, U.S.A.). Authorization (no. 28794-1) to access Brazilian biodiversity was issued by the Brazilian Institute of the Environment and Renewable Natural Resources (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis, IBAMA) of the Ministry of the Environment of the Federal Government of Brazil.

### RNA isolation and venom gland cDNA library construction

Total RNA was prepared using the RNAeasy Mini Kit (Qiagen, U.S.A), which is based on a combination of guanidine-isothiocyanate lysis and silica-membrane purification in a mini-spin column. The procedures were performed according to the manufacturer's instructions. The quality of the RNA was analyzed with denaturating RNA electrophoresis in TAE agarose gel as described in [87] and quantified by UV absorption using NanoDrop (Thermo Scientific, Wilmington, DE-USA).

The cDNA library was constructed using a switching mechanism at the 5′ end of the RNA template to preferentially amplify the poly-A tail (IN-FUSION™ SMARTer™ cDNA Library Construction kit, Clontech, Mountain View, CA-USA). Briefly, the synthesis of single strand cDNA began with 1 µg of total RNA and 3′ In-fusion SMARTer CDS primers. The resulting first strands of cDNA were sequentially amplified with SMARTer V Primer. To synthesize double strand cDNA, the Advantage 2 Polymerase, the 5′ PCR primer IIA and 3′ In-fusion SMARTer PCR Primer with the lowest possible number of cycles (in this case 15) were used. The double stranded (ds-) cDNA was fractionated by size exclusion chromatography (CHROMA SPIN 400, Clontech, USA), and the fractions with specific molecular sizes (over 400 bp) were combined and precipitated with NaCl/ethanol. The cDNA was quantified by U.V. spectrophotometry and ligated into the plasmid vector (pSMART 2IFD). A commercial strain of *E. Coli* was used to as a host for the maintenance and propagation of the *D. quadriceps* venom gland cDNA library, which was then stored at −80°C.

### Sanger sequencing of independent clones

The cDNA library was seeded in LB agar plates supplemented with appropriate antibiotics, and individual clones were selected and analyzed for insert size. To screen for positive clones, specific primers flanking the multi-cloning site of the vector were used in a PCR reaction. The amplified screening products were analyzed in 1% agarose gel and photodocumented. The PCR products of 420 independent clones were cleaned up by EXO/SAP treatment and sodium acetate/ethanol precipitation. The precipitated DNA were resuspended in deionized formamide and sequenced in a 3100 Avant Sequencer (Applied Biosystems, USA) using Prism Big Dye Terminator v3.0 Cycle Sequencing Ready Reactions (Applied Biosystems, USA).

### Data analysis of Sanger sequencing

Sequences were analyzed with the CLC Bio Main Workbench v 6.8.3 Software. The sequences were trimmed to exclude vector contamination and the poly-A tail. The following parameters were selected for clustering and assembling: minimum aligned read length of 50 and medium (ambiguity level) alignment stringency. The resulting contigs (clusters and singlets) were compared against the non-redundant GenBank (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov) using BLASTx for all organisms and/or an arthropod databank (see below for deep sequencing analysis).

### Sample preparation and deep Sequencing using Ion Torrent Technology

A fragment library was generated with an input of 100 ng DNA using the Ion Xpress Plus Fragment Library Kit (Life Technologies, Carlsbad, USA). Next, DNA fragmentation, adaptor ligation, fragment size selection and library amplification were carried out for a total of 8 cycles according to manufacturer's instructions. The targeting of fragments of approximately 330 bp was performed with 2% agarose gel cassettes for the Pippin Prep instrument (Sage Science, Beverly, USA). Prior to emulsion PCR, the size distribution and concentrations of the library were assessed with a Bioanalyzer 2100 using the DNA High Sensitivity Kit (Agilent, Santa Clara, CA). The fragment library was adjusted to approximately 26 pM and amplified with Ion Sphere particles™ (ISPs) by emulsion PCR using the Ion OneTouch™ Instrument with the Ion OneTouch™ 200 Template Kit (Life Technologies, Grand Island, NY) and template-positive ISP enrichment according to the manufacturer's protocol (Part Number 4472430 Rev. E, 06/2012). Over 50% of the ISPs were and then sequenced on an Ion 316™ chip using the Ion Torrent Personal Genome Machine (PGM™) (Life Technologies, San Francisco, CA, USA) for 130 cycles (520 flows) with the Ion PGM™ 200 Sequencing Kit (Part Number 4474246 Rev. D, 06/2012).

### Sequence filtering and (*de novo*) assembly of deep sequencing data

Ion Torrent Personal Genome Machine (PGMTM) sequencing can generate 200–500 base pairs (bps) from each single end (SE) of a cDNA fragment. The raw reads were initially processed to obtain clean reads by removing the adapter sequences and removing low quality sequences (i.e., reads in which more than 50% of the bases had quality value $s \leq 5$). Next, de novo transcriptome assembly was carried out using the short-read assembling program Trinity [88]. Specifically the k-mer counting method was set to jellyfish, and the contigs that could not be extended on either end were defined as transcripts.

### Gene ontology and functional annotation

All of the assembled transcripts were searched against the NCBI non-redundant protein sequence (nr) and COG databases using BLASTX to identify the proteins with the highest sequence similarities with the given transcripts to retrieve the function annotations of the transcripts, and standard cut-off E-value of $<10^{-5}$ was used. Next, the species and function information from the nr database was used to perform taxonomic analyses with in-house Perl scripts. For the nr annotation, the in-house BLAST2GO embedded annotation pipeline was used to obtain the GO annotations of the assembled transcripts to describe the biological processes, molecular functions and cellular components of the transcripts [89], and the WEGO [90] software was used to conduct GO functional classification to understand the distribu-

tion of gene functions at the macroscopic level. Furthermore, the transcripts that had no matches in the nr database were subjected to another BLASTX search against the UniProtKB database.

### Bioinformatic analyses of similarities and molecular phylogenies

For particular transcript sequences, systematically similarity analyses were performed using the BLASTP tool against the non-redundant (nr) and UniProt databases, and these analyses were subsequently restricted to the *Arthropoda*, *Hymenoptera* or *Formicidae* taxa. Further, the analyses were expanded to include more specific databases such as ArachnoServer, Knottins [91] and SCOP KNOTTIN Superfamily. Predicted amino acid sequences were aligned with ClustalW software using the default parameters (http://www.ebi.ac.uk), and the putative cleavage sites related to signal peptides (and pro-peptides) were deduced and compared with the available experimental crystal structure or proteomic data

## Supporting Information

**Table S1 Annotations of identified ESTs retrieved from randomly selected clones in the *D. quadriceps* venom gland cDNA library.** Contigs assembled from Sanger sequencing data according to described in the 'materials and methods' section, The Uniprot and sequence identification (GI) numbers of homologous polypeptides are indicated. Number in parenthesis represents percentage of sequence similarity.
(DOC)

**Table S2 Overview of experimental data from the next generation RNA sequencing.**
(DOC)

**Table S3 Annotations of the identified venomous contigs in the *D. quadriceps* venom gland.** Contigs representing candidate toxin-like peptides based on the alignment of EST data set, prediction of signal peptides and ORFs, and BLASTx comparison. GI denotes sequence identification number of "top hit" toxin homologous. Number in parenthesis represents percentage of sequence similarity.
(DOC)

**Table S4 Summary of toxin-like peptide candidates from the *D. quadriceps* venom gland transcriptome.** Validation of some RNA-Seq assembled contigs according to the mapped EST fragments. Full-length open reading frames (ORFs) are indicated with a "+" sign. The presence or absence of a signal peptide in the ORFs is indicated by a "Y" (yes) or "N" (no).
(DOC)

**Figure S1 The length distribution of the assembled transcripts of *D. quadriceps* venom gland.** Sequence length is expressed as base pairs (bp), and the frequency is expressed as the absolute number of sequences. The y-axis indicates the number of sequences of different lengths.
(TIF)

**Figure S2 Gene ontology (GO) classification of contigs identified in the venom gland of *D. quadriceps*.** Distribution of unigenes associated with biological processes, cellular components and molecular functions of the giant ant venome transcripts.
(TIF)

**Figure S3 Clusters of orthologous groups (COG) function classification of the assembled giant ant venomic transcripts.** Query of the COG database allowed for the classification of venom-related transcripts into 24 group based on functions (from A to Z) and attributes. The distribution of these groups of the *D. quadriceps* transcriptome (y-axis) is shown.
(TIF)

**Figure S4 Comparison of sex determination proteins from *D. quadriceps* with their hymenopterans orthologs.** Phylogenetic tree based on neighbor-joining analyses of a concatenated alignment of a sex determination protein and orthology relationships in multiple insects. The scale bar indicates 0.05 substitution per site. *Apis mellifera, Megachile rotundata, Bombus impatiens, Nasonia vitripennis, Acromyrmex echinatior, Harpegnathos saltator, Solenopsis invicta, camponotus floridanus*.
(TIF)

**Figure S5 Structural comparison of *D. quadriceps* CAP venom allergen antigen 5 with two hymenopteran venom allergen and the tridimensional topology of CAP venom allergen antigen 5.** The *Dinoponera* venom allergen precursor was aligned with the following Hymenoptera allergen Ag5 precursors using the BLOSUM62 scores: VA3_SOLIN (*Solenopsis invicta* venom allergen 3, Sol I 3) and VA5_VESVU (*Vespula vulgaris* Venom allergen 5, Ves5). Part A depicts the primary sequence analysis. The signal peptide is indicated by an orange box; the predicted pattern of disulfide bridges is represented by connected lines and numbered based on *Dinoponera* CAP Venom allergen 5 (dark gray box). The major conserved secondary structural elements are indicated by marks above the relevant amino acid sequences (red bar, α-helix; yellow arrow, β-strand). CAP signature motifs are marked by cyan boxes, and the histidine and glutamine/glutamic acid residues involved in the putative active site are also indicated (blue*). The histidines, which have the ability to form complexes with divalent cations, form a structure with some similarity to the protease active sites, which is consistent with the calcium-activated serine protease-like activity that has been reported for the *Conus textile* cysteine-rich venom protein endoproteinase Tex31 (Q7YT83). Part B: tri-dimensional structure based the on structural characteristics described in part A.
(TIF)

**Figure S6 *D. quadriceps* U8-agatoxin-Ao1a-like venom homolog Contig1144_B2-2 was assembled with data from the RNA-Seq reads.** TXAG8 (Q5Y4U4) is a U8-agatoxin-Ao1a-like venom homolog that was discovered by a program for the identification of toxin-related candidates in arthropod databases [81]. The other sequences came from query of the database for homologous entries.
(TIF)

## Author Contributions

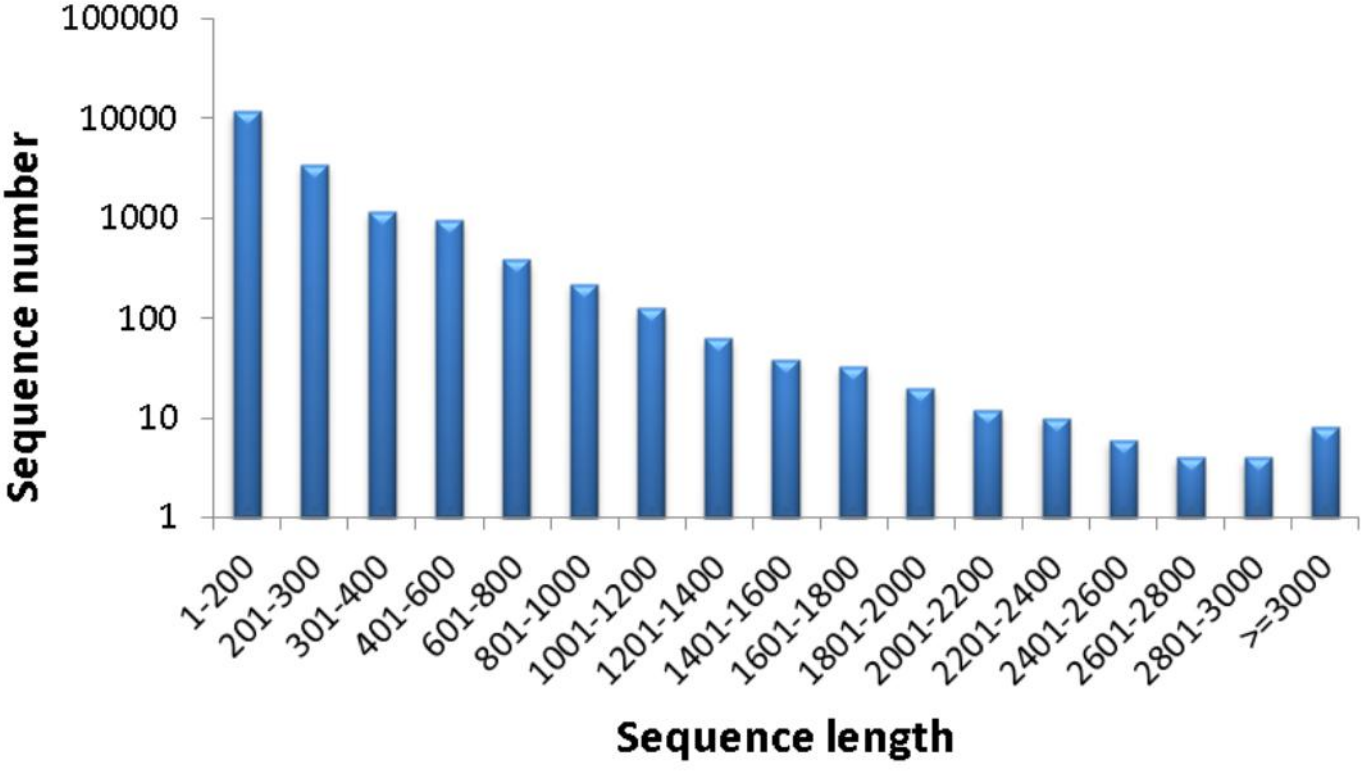
Conceived and designed the experiments: AMCM ARBPS SMYL GR-B. Performed the experiments: AFCT CH AH C-MC YPQ. Analyzed the data: AFCT CH C-MC ARBPS SWL SMYL GR-B. Contributed reagents/materials/analysis tools: AMCM ARBPS SMYL GR-B. Wrote the paper: AFCT CH ARBPS SMYL GR-B. Field trip missions and rearing ants: AFCT YPQ.

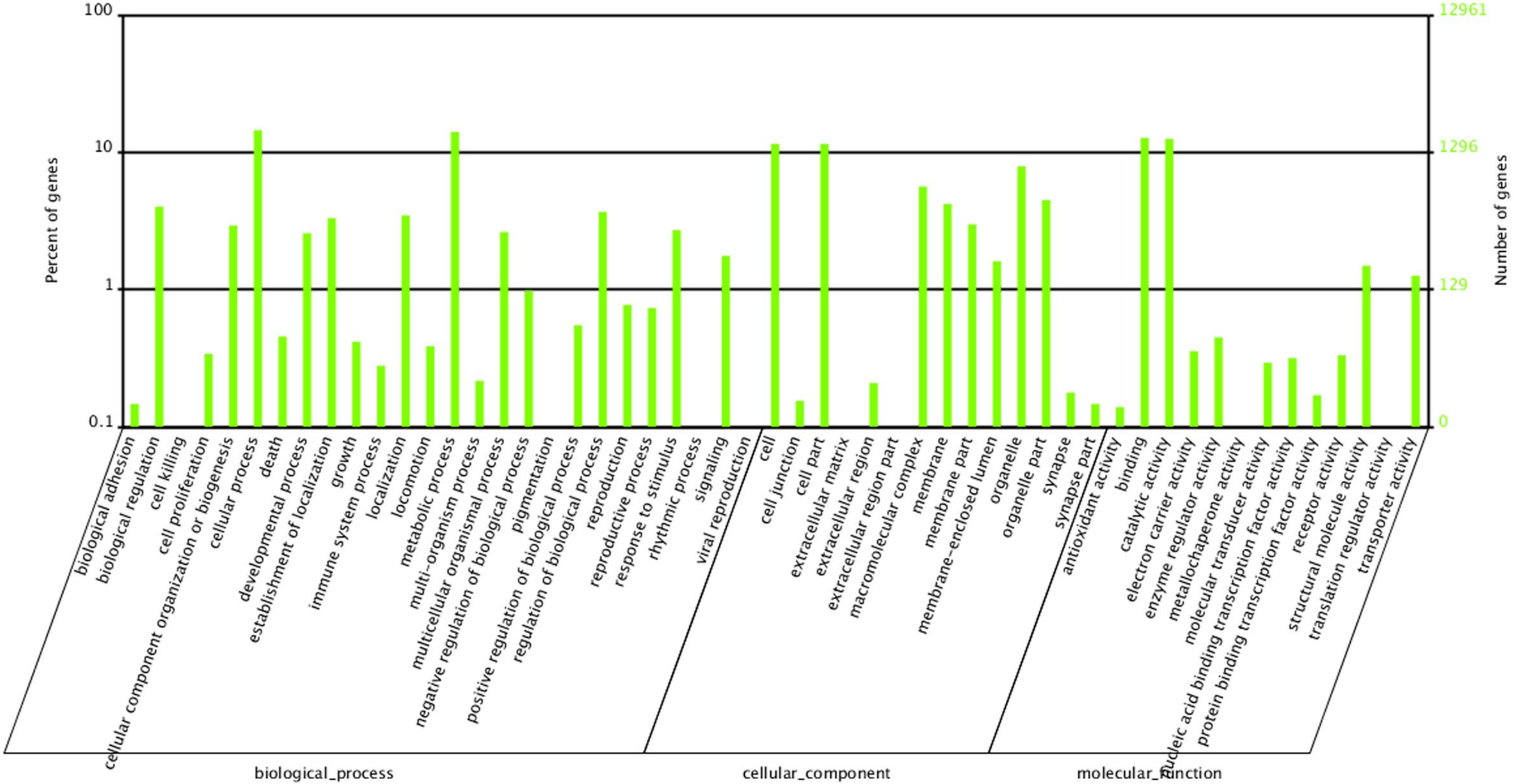
## References

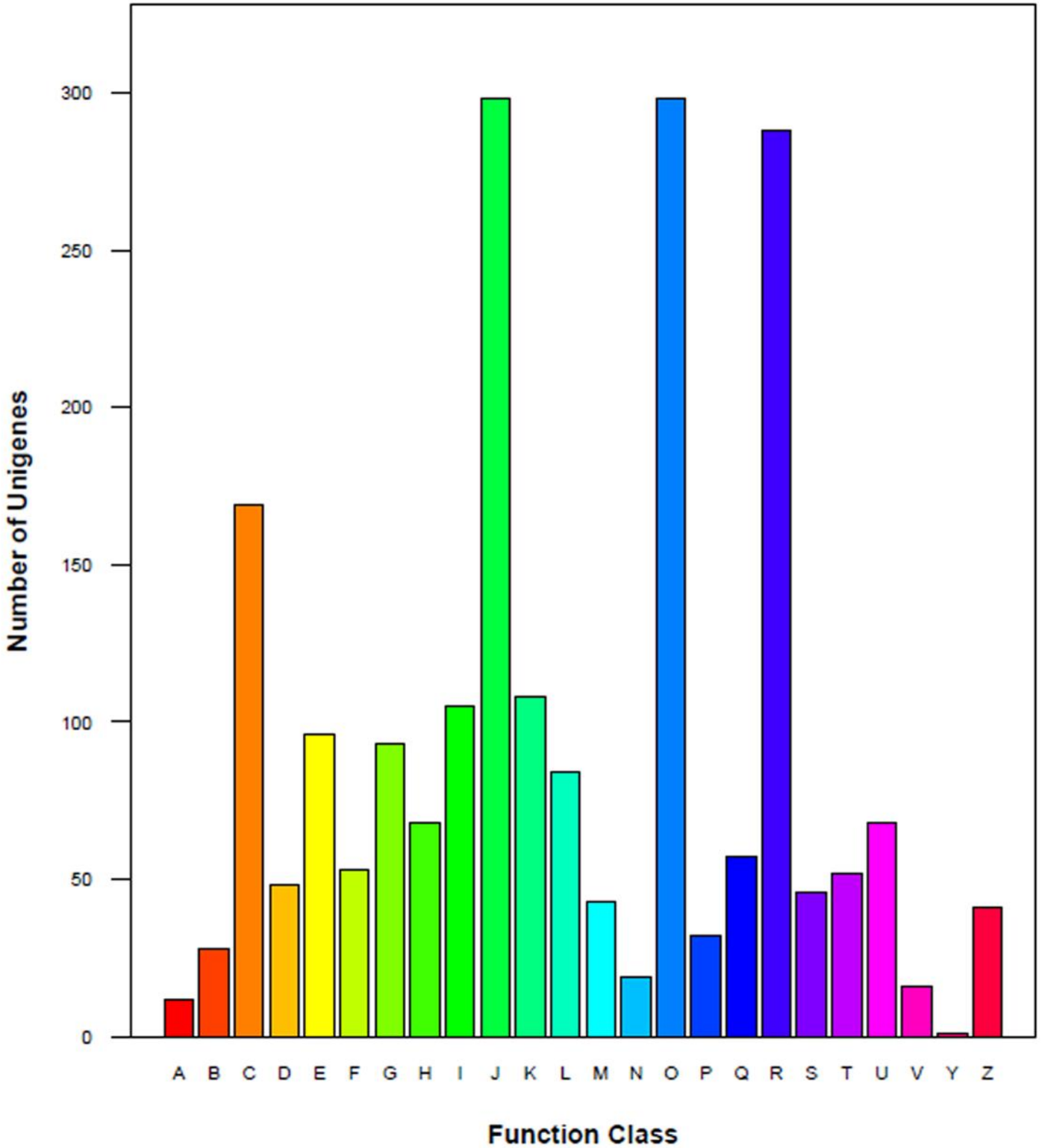
1. Munoz-Torres MC, Reese JT, Childers CP, Bennett AK, Sundaram JP, et al. (2012) Hymenoptera Genome Database: integrated community resources for insect species of the order Hymenoptera. Nucleic Acids Research 39:658–662.
2. Bouzid W, Klopp C, Verdenaud M, Ducancel F, Vétillard A (2013) Profiling the venom gland transcriptome of *Tetramorium bicarinatum* (Hymenoptera: Formicidae): The first transcriptome analysis of an ant species. Toxicon 70:70–81.

3. Cherniak EP (2010) Bugs as drugs, Part 1: Insects: the new alternative medicine for the 21st century? Altern Med Rev 15:124–135.
4. Badr G, Garraud O, Daghestani M, Al-Khalifa MS, Richard Y (2012) Human breast carcinoma cells are induced to apoptosis by samsum ant venom through an IGF-1-dependant pathway, PI3K/AKT and ERK signaling. Cel Immunol 273:10–16.
5. Ebaid H, Al-Khalifa M, Isa AM, Gadoa S (2012) Bioactivity of Samsum ant (*Pachycondyla sennaarensis*) venom against lipopolysaccharides through antioxidant and upregulation of Akt1 signaling in rats. Lipids in Health Disease 11:93.
6. Dkhil M, Abdel BA, Al Quraishi, Al Khalifa M (2010) Anti-inflammatory activities of the venom from samsumants *Pachycondyla sennaarensis*. Afr J Pharm Pharmacol 4:115–118.
7. Orivel J, Dejean A (2001) Comparative effect of the venoms of ants of the genus *Pachycondyla* (Hymenoptera: Ponerinae). Toxicon 39:195–201.
8. Zelezetsky I, Pag U, Antcheva N, Sahl HG, Tossi A (2005) Identification and optimization of an antimicrobial peptide from the ant venom toxin pilosulin. Biochem Biophys 434:358–364.
9. Quinet Y, Vieira RHSF, Sousa MR, Evangelista-Barreto NS, Carvalho FCT, et al. (2012) Antibacterial properties of contact defensive secretions in neotropical *Crematogaster* ants. JVAT 18:441–445.
10. Torres AFC, Quinet YP, Havt A, Rádis-Baptista G, Martins AMC (2013) Molecular pharmacology and toxynology of venom from ants.; Rádis-Baptista G, editor. An Integrated view of the molecular recognition and toxinology – From analytical procedures to biomedical applications. Croatia: InTech. pp. 207–222.
11. Lopes KS, Rios E, Dantas RT, Lima C, Linhares M, et al. (2011) Effect of *Dinoponera quadriceps* venom on chemical-induced seizures models in mice. Rencontres en Toxinologie – Meeting on Toxinology SFET Editions. Available: http://www.sfet.asso.fr. Accessed 20 August 2012.
12. Lopes KS, Rios ERV, Lima CNC, Linhares MI, Torres AFC, et al. (2013) The effects of the brazilian ant *Dinoponera quadriceps* venom on chemically induced seizure models. Neurochem Intern 63:141–145.
13. Sousa PL, Quinet Y, Ponte EL, do Vale JF, Torres AFC, et al. (2012) Venom's antinociceptive property in the primitive ant *Dinoponera quadriceps*. J Ethnopharmacol 144:213–216.
14. Kempf WW (1971) A preliminary review of the Ponerinae ant genus *Dinoponera* Roger (Hymenoptera: Formicidae). Studia Entomologica, 14: 369–394.
15. Paiva RVS, Brandão CRF (1995) Nests, worker population, and reproductive status of workers, in the giant queenless ponerinae ant *Dinoponera* Roger Hymenoptera Formicidae. Ethology Ecology and Evolution 7:297–312.
16. Lenhart PA, Dash ST, Mackay WP (2013) A revision of the giant Amazonian ants of the genus Dinoponera (Hymenoptera, Formicidae). J Hymenoptera 31:119–164.
17. Lattke, J. E. 2003. Subfamilia Ponerinae, p. 261–276 In: F. FERNÁNDEZ (ed.) Introducción a las hormigas de la región Neotropical. Bogotá, Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, 398 p.
18. Wilson EO, Hölldobler B (2005) Eusociality: origin and consequences. Proc Natl Acad Sci 102:13367–13371
19. Monnin T, Peeters C (1998). Monogyny and regulation of worker mating in the queenless ant *Dinoponera quadriceps*. Animal Behaviour 55:299–306.
20. Vasconcellos A, Santana GG, Souza AK (2004) Nest spacing and architecture, and swarming of males of *Dinoponera quadriceps* (Hymenoptera, Formicidae) in a remnant of the Atlantic forest in Northeast Brazil. Braz J Biol 64:357–362.
21. Schoeters E, Billen J (1995) Morphology and ultra structure of the convoluted gland in the ant *Dinoponera australis* (Hymenoptera:Formicidae). Int J Insect Morphol Embryol 24:323–332.
22. Mathias MIC (2007) Glândulas de veneno em formigas. Biológico 69:229–235.
23. Holldobler B, Wilson EO (1990). The ants. Cambridge: Harvard University, Belknap Press. 746 p.
24. Haddad Jr V, Cardoso JLC, França FOS, Wen FH (1996) Acidentes por formigas: um problema dermatológico. An Bras Derm 71:527–530.
25. Haddad Jr V(2001) Acidentes por formigas. In: Manual de diagnóstico e tratamento de acidentes por animais peçonhentos. Brasília, FNS. pp.65–66
26. Haddad Jr V, Cardoso JLC, Moraes RHP (2005) Description of an injury in a human caused by a false tocandira (*Dinoponera gigantea*, Perty, 1833) with a revision on folkloric, pharmacological and clinical aspects of the giant ants of the genera Paraponera and Dinoponera (sub-family ponerinae). Rev Inst Med Trop 47:235–238.
27. Cruz Lopez L, Morgan ED (1997) Explanation of bitter taste of venom of ponerine ant *Pachycondyla apicalis*. J Chem Ecol 23:705–712.
28. Duval A, Malécot CO, Pelhate M, Piek T (1992) Poneratoxin, a new toxin from ant venom, reveals an interconversion between two gating modes of the Na channels in frog skeletal muscle fibres. Pflugers Arch 420:239–247.
29. Szolajska E, Poznanski J, Ferber ML, Michalik J, Gout E, et al. (2004) Poneratoxin, a neurotoxin from ant venom. Structure and expression in insect cells and construction of a bio-insecticide. Europ J Biochem 271:2127–2136.
30. Johnson SR, Copello JA, Evans MS, Suarez AV (2010) A biochemical characterization of the major peptides from the Venom of the giant Neotropical hunting ant *Dinoponera australis*. Toxicon 55:702–710.
31. Min XJ, Butler G, Storms R, Tsang A (2005) OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res 33:677–680.
32. Menne KM, Hermjakob H and Apweiler R (2000) A comparison of signal sequence prediction methods using a test set of signal peptides. Bioinformatics 16:741–742.
33. Zientz E, Feldhaar H, Stoll S, Gross R (2005) Insights into the microbial world associated with ants. Arch Microbiol 184:199–206.
34. Boulogne I, Germosen-Robineau L, Ozier-Lafontaine H, Jacoby-Koaly C, Aurela L, Loranger-Merciris G (2012) *Acromyrmex octospinosus* (Hymenoptera: Formicidae) management. Part 1: Effects of TRAMIL's insecticidal plant extracts. Pest Manag Sci 68: 313–320.
35. Ward PS (2006) Ants. Curr Biol. 7:16:R152–155.
36. Gruber CW, Muttenthaler M (2012) Discovery of Defense- and Neuropeptides in Social Ants by Genome-Mining. PLoS ONE 7(3): e32559.
37. Raidis-Baptista G (2011). Molecular Toxinology – Cloning Toxin Genes for Addressing Functional Analysis and Disclosure Drug Leads.; Brown, G, editor. Molecular Cloning - Selected Applications in Medicine and Biology. Croatia: InTech. pp. 161–196.
38. Ryner LC, Goodwin SF, Castrillon DH, Anand A, Villella A, et al (1996) Control of male sexual behavior and sexual orientation in Drosophila by the fruitless gene. Cell 87:1079–1089.
39. Ito H, Fujitani K, Usui K, Shimizu-Nishikawa K, Tanaka S, et al (1996) Sexual orientation in Drosophila is altered by the satori mutation in the sex-determination gene fruitless that encodes a zinc finger protein with a BTB domain. Proc Natl Acad Sci U S A 93:9687–9692.
40. Zollman S, Godt D, Privé GG, Couderc JL, Laski FA (1994) The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in Drosophila, Proc Natl Acad Sci U S A 91:10717–10721.
41. Dauwalder B, Tsujimoto S, Moss J, Mattox W (2002) The Drosophila takeout gene is regulated by the somatic sex-determination pathway and affects male courtship behavior. Genes Dev 16:2879–2892.
42. Heinrichs V, Ryner LC, Baker BS (1998) Regulation of sex-specific selection of fruitless 5′ splice sites by transformer and transformer-2. Mol Cell Biol 18:450–458.
43. Basha E, O'Neill H, Vierling E (2012) Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions. Trends Biochem Sci 37:106–117.
44. Hoffman DR (2010) Ant venoms. Curr Opin Allergy Clin Immunol 10:342–346.
45. Karnaky KJ Jr, Hazen-Martin D, Miller DS (2003) The xenobiotic transporter, MRP2, in epithelia from insects, sharks, and the human breast: implications for health and disease. J Exp Zool A Comp Exp Biol. 300:91–97.
46. O'Donnell MJ (2009) Too much of a good thing: how insects cope with excess ions or toxins in the diet. J Exp Biol. 212:363–72.
47. Fitzgerald KT, Flood AA (2006) Hymenoptera stings. Clin Tech Small Anim Pract 21:194–204.
48. Biló BM, Rueff F, Mosbech H, Bonifazi F, Oude-Elberink JN (2005) Diagnosis of Hymenoptera venom allergy. Allergy 60:1339–1349.
49. Hoffman DR, Sakell RH and Schmidt M (2005) Sol i 1, the phospholipase allergen of imported fire ant venom. J Allergy Clin Immunol 115:611–616.
50. Padavattan S, Schmidt M, Hoffman DR, Marković-Housley Z (2008) Crystal structure of the major allergen from fire ant venom, Sol i 3. J Mol Biol 383:178–185.
51. Gibbs GM, Scanlon MJ, Swarbrick J, Curtis S, Gallant E, Dulhunty AF, O'Bryan MK (2006) The cysteine-rich secretory protein domain of Tpx-1 is related to ion channel toxins and regulates ryanodine receptor Ca2+ signaling. J Biol Chem. 281(7): 4156–4163.
52. Gibbs GM, Roelants K, and 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. Endocrine Reviews 29(7): 865–897.
53. Zalat S, Schmidt J, Moawad TI (2003) Lipase and phospholipase activities of Hymenoptera venoms (wasps and ants). Egyptian J Biol 5:138–147.
54. Seismann H, Blank S, Cifuentes L, Braren I, Bredehorst R, Grunwald T, Ollert M, Spillner E (2010) Recombinant phospholipase A1 (Ves v 1) from yellow jacket venom for improved diagnosis of hymenoptera venom hypersensitivity. Clin Mol Allergy 8:7.
55. Vines CM (2012) Phospholipase C. Adv Exp Med Biol 740:235–254.
56. Murakami M, Kudo I (2002) Phospholipase A2. JB Minireview-Lipid Signaling Biochem 131:285–292.
57. Richmond GS, Smith TK (2011) Phospholipases A1. Inter J Mol Sci 12:588–612.
58. Sonoda H, Aoki J, Hiramatsu T, Ishida M, Bandoh K, Nagai Y, Taguchi R, Inoue K, Arai H (2002) A Novel Phosphatidic acid-selective phospholipase A1 that produces lysophosphatidic acid. J Biol Chem 277: 34254–34263.
59. Aoki J, Nagai Y, Hosono H, Inoue K, Hiroyuki A (2002) Structure and function of phosphatidylserine-specific phospholipase A1. Bioch Biophys Acta 1582:26–32.
60. King TP, Spangfort MD (2000) Structure and biology of stinging insect venom allergens. Int Arch Allergy Immunol 123:99–106.
61. dos Santos Pinto JR, Fox EG, Saidemberg DM, Santos LD, da Silva Menegasso AR, et al. (2012) Proteomic view of the venom from the fire ant *Solenopsis invicta* Buren. J Proteome Res 11:4643–4653.
62. Huang WN, Chen YH, Chen CL, Wu W (2011) Surface pressure-dependent interactions of secretory phospholipase A2 with zwitterionic phospholipid membranes. Langmuir 27:7034–7041.

63. Schmidt JO, Blum MS (1978) A harvester ant venom: chemistry and pharmacology. Science 200:1064–1066
64. Hoffman DR, Peroutka CM, Schmidt M (2008) Imported fire ant venom phospholipase, Sol I 1, is a protein allergen. J Allergy Clin Immunol 121:S30.
65. Muller UR (2002) Recombinant Hymenoptera venom allergens. Allergy 57: 570–576.
66. Hoffman DR, Dove DE, Moffitt JE, Stafford CT (1988) Allergens in hymenoptera venom. XXI. Cross-reactivity and multiple reactivity between fire ant venom and bee and wasps venoms. J Allergy Clin Immunol 82:828–834.
67. Sin BA, Akdis M, Zumkehr J, Bezzine S, Bekpen C, et al. (2011) T-cell and antibody responses to phospholipase A2 from different species show distinct cross-reactivity patterns. Allergy 66:1513–1521.
68. Liscovitch M, Czarny M, Fiucci G, Tang X (2000) Phospholipase D: molecular and cell biology of a novel gene family. Biochem J 3:401–415
69. McDemott M, Wakelam MJ, Morris AJ (2004) Phospholipase D. Biochem Cell Biol 82: 225–253.
70. Chaim OM, Trevisan-Silva D, Chaves-Moreira D, Wille AC, Ferrer VP, et al (2011) Brown spider (*Loxosceles genus*) venom toxins: tools for biological purposes. Toxins 3:309–344.
71. Magalhães GS, Caporrino MC, Della-Casa MS, Kimura LF, Prezotto-Neto JP, et al (2013) Cloning, expression and characterization of a phospholipase D from *Loxosceles gaucho* venom gland. Biochimie 9:1773–1783.
72. Hatfield MJ, Potter PM (2011) Carboxylesterase inhibitors. Expert Opin Ther 8:1159–1171.
73. Cui F, Lin Z, Wang H, Liu S, Chang H, et al (2011) Two single mutations commonly cause qualitative change of nonspecific carboxylesterases in insects. Insect Biochem Mol Biol 41:1–8.
74. Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, et al. (2010) Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. Science 327:343–348.
75. Bonasio R, Zhang G, Ye C, Mutti NS, Fang X, et al (2010) Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Science 329:1068–1071.
76. Blank S, Seismann H, Bockisch B, Braren I, Bredehorst R, et al (2008) Identification and recombinant expression of a novel IgE-reactive 70 kDa carboxylesterase from *Apis mellifera* venom. Available: http://www.uniprot.org/uniprot/B2D0J5. Accessed 19 September 2013.
77. Shai Y (2002) Mode of action of membrane active antimicrobial peptides. Biopolymers 66: 236–248.
78. Huang Y, Huang J, Chen Y (2010) Alpha-helical cationic antimicrobial peptides: relationships of structure and function. Protein Cell 2:143–152.
79. Inagaki H, Akagi M, Imai HT, Taylor RW, Kubo T (2004) Molecular cloning and biological characterization of novel antimicrobial peptides, pilosulin 3 and pilosulin 4, from a species of the Australian ant genus *Myrmecia*. Arch Biochem Biophys 428:170–178.
80. Inagaki H, Akagi M, Imai HT, Taylor RW, Wiese MD, et al. (2008) Pilosulin 5, a novel histamine-releasing peptide of the Australian ant, *Myrmecia pilosula* (Jack Jumper Ant). Arch Biochem Biophys 477:411–416.
81. Kozlov S, Malyavka A, McCutchen B, Lu A, Schepers E, et al (2005) A novel strategy for the identification of toxin like structures in spider venom. Proteins 59:131–140.
82. Asgari S, Zhang G, Zareie R, Schmidt O (2003) A serine proteinase homolog venom protein from an endoparasitoid wasp inhibits melanization of the host hemolymph. Insect Biochem Mol Biol 33:1017–1024.
83. Parkinson NM, Conyers C, Keen J, MacNicoll A, Smith I, et al (2004) Towards a comprehensive view of the primary structure of venom proteins from the parasitoid wasp *Pimpla hypochondriaca*. Insect Biochem Mol Biol 34:565–571.
84. Daly NL, Craik DJ (2011) Bioactive cystine knot proteins. Curr Opin Chem Biol. 15:362–368.
85. Gracy J, Chiche L (2011) Structure and modeling of knottins, a promising molecular scaffold for drug discovery. Curr Pharm Des 17:4337–4350.
86. Zhang Z, Zhu S(2012) Comparative genomics analysis of five families of antimicrobial peptide-like genes in seven ant species. Dev Comp Immunol. 38:262–274.
87. Masek T, Vopalensky V, Suchomelova P, Pospisek M (2005) Denaturing RNA electrophoresis in TAE agarose gels. Anal Biochem. 336:46–50.
88. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652.
89. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676.
90. Ye J, Fang L, Zheng H, Zhang Y, Chen J, et al. (2006) WEGO: a web tool for plotting GO annotations. Nucleic Acids Res 34:293–297.
91. Gelly JC, Gracy J, Kaas Q, Le-Nguyen D, Heitz A, Chiche L (2004) The KNOTTIN website and database: a new information system dedicated to the knottin scaffold. Nucleic Acids Res 32:156–159.
92. Matuszek MA, Hodgson WC, Sutherland SK, King RG (1992) Pharmacological studies of jumper ant (*Myrmecia pilosula*) venom: evidence for the presence of histamine, and haemolytic and eicosanoid-releasing factors. Toxicon 30:1081–1091


100000
10000
1000
100
10
1
Sequence number
1-200
201-300
301-400
401-600
601-800
801-1000
1001-1200
1201-1400
1401-1600
1601-1800
1801-2000
2001-2200
2201-2400
2401-2600
2601-2800
2801-3000
>=3000
Sequence length


100
10
1
0.1
Percent of genes
12961
1296
129
0
Number of genes
biological adhesion
biological regulation
cell killing
cell proliferation
cellular component organization or biogenesis
cellular process
death
developmental process
establishment of localization
growth
immune system process
localization
locomotion
metabolic process
multi-organism process
multicellular organismal process
negative regulation of biological process
pigmentation
positive regulation of biological process
regulation of biological process
reproduction
reproductive process
response to stimulus
rhythmic process
signaling
viral reproduction
cell
cell junction
cell part
extracellular matrix
extracellular region
extracellular region part
macromolecular complex
membrane
membrane part
membrane-enclosed lumen
organelle
organelle part
synapse
synapse part
antioxidant activity
binding
catalytic activity
electron carrier activity
enzyme regulator activity
metallochaperone activity
molecular transducer activity
nucleic acid binding transcription factor activity
protein binding transcription factor activity
receptor activity
structural molecule activity
translation regulator activity
transporter activity
biological_process
cellular_component
molecular_function




A: RNA processing and modification

B: Chromatin structure and dynamics

C: Energy production and conversion

D: Cell cycle control, cell division, chromosome partitioning

E: Amino acid transport and metabolism

F: Nucleotide transport and metabolism

G: Carbohydrate transport and metabolism

H: Coenzyme transport and metabolism

I: Lipid transport and metabolism

J: Translation, ribosomal structure and biogenesis

K: Transcription

L: Replication, recombination and repair

M: Cell wall/membrane/envelope biogenesis

N: Cell motility

O: Posttranslational modification, protein turnover, chaperones

P: Inorganic ion transport and metabolism

Q: Secondary metabolites biosynthesis, transport and catabolism

R: General function prediction only

S: Function unknown

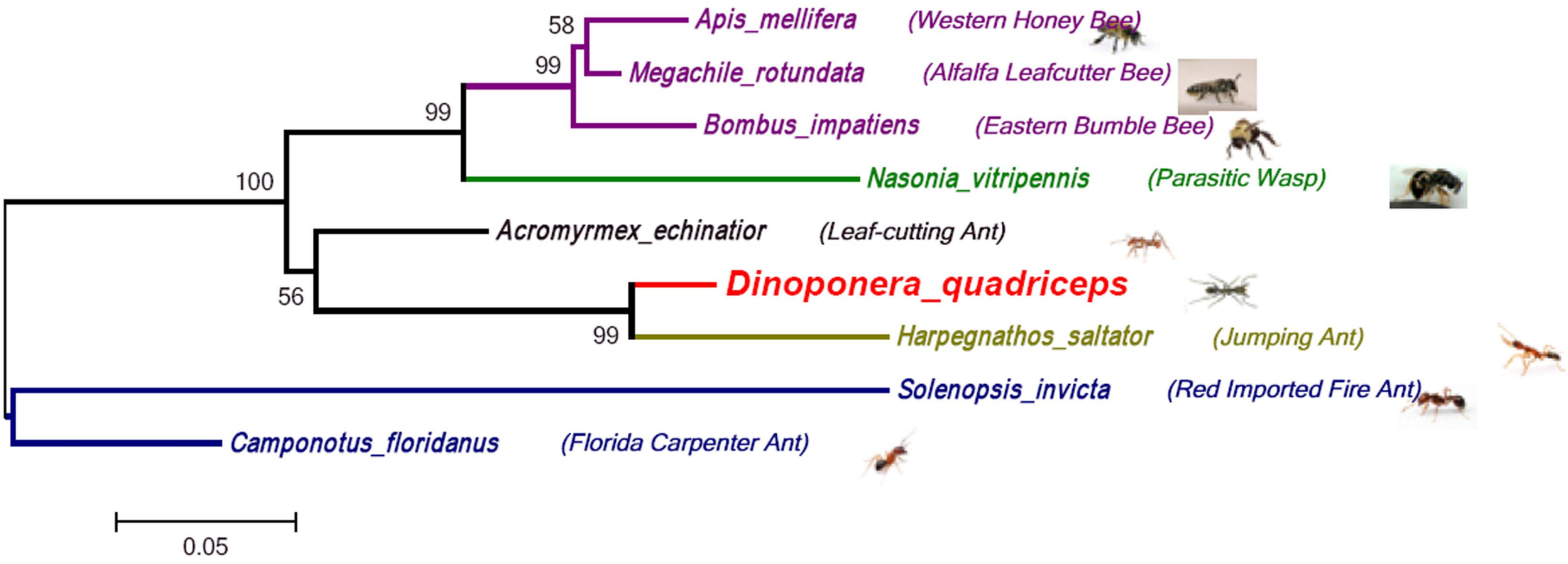
T: Signal transduction mechanisms

U: Intracellular trafficking, secretion, and vesicular transport

V: Defense mechanisms

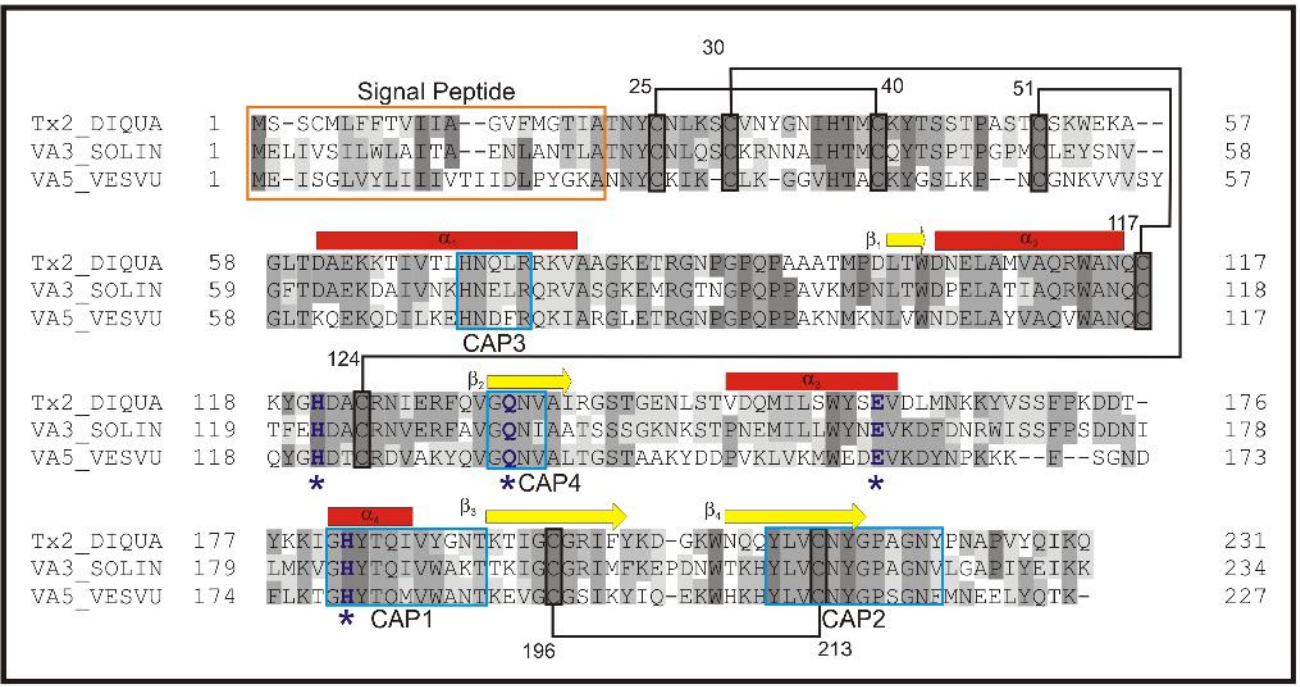
Y: Nuclear structure

Z: Cytoskeleton

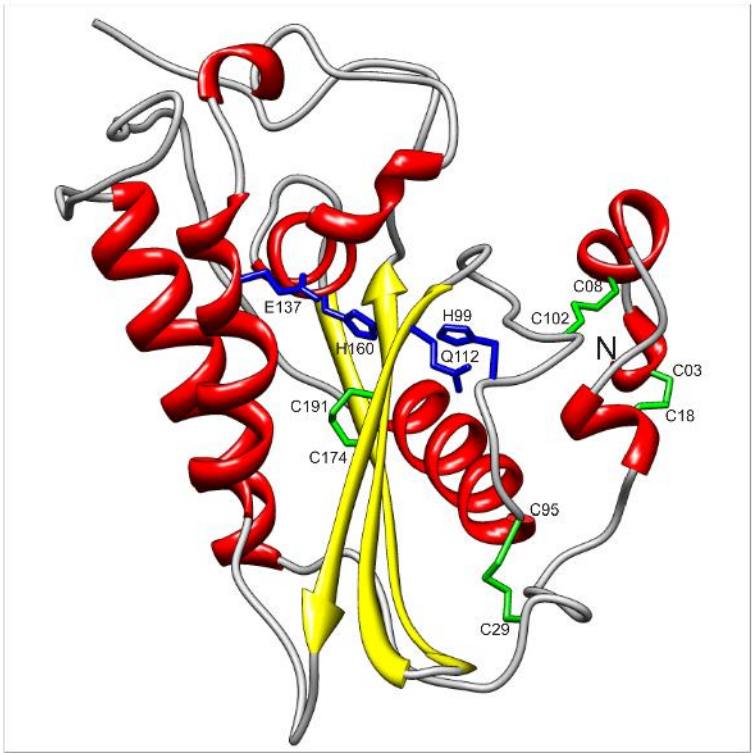

58
99
99
100
56
99
Apis_mellifera
(Western Honey Bee)
Megachile_rotundata
(Alfalfa Leafcutter Bee)
Bombus_impatiens
(Eastern Bumble Bee)
Nasonia_vitripennis
(Parasitic Wasp)
Acromyrmex_echinatior
(Leaf-cutting Ant)
Dinoponera_quadriceps
Harpegnathos_saltator
(Jumping Ant)
Solenopsis_invicta
(Red Imported Fire Ant)
Camponotus_floridanus
(Florida Carpenter Ant)
0.05

**A**

```
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
```


Signal Peptide
30
25
40
51
117
124
Tx2_DIQUA 1 MS-SCMLFFTVIIA--GVFMGTIATNYCNLKSCVNYGNIHTMCKYTSSTPASICSKWEKA-- 57
VA3_SOLIN 1 MELIVSILWLAITA--ENLANTLATNYCNLQSCKRNNAIHTMCQYTSPTPGPMCLEYSNV-- 58
VA5_VESVU 1 ME-ISGLVYLIIIVTIIDLPYGKANNYCKIK-CLK-GGVHTACKYGSLKP--NCGNKVVVSY 57
Tx2_DIQUA 58 GLTDAEKKTIVTIHNQLRRKVAAGKETRGNPGPQPAAATMPDLTWDNELAMVAQRWANQC 117
VA3_SOLIN 59 GFTDAEKDAIVNKHNELRQRVASGKEMRGTNGPQPPAVKMPNLTWDPELATIAQRWANQC 118
VA5_VESVU 58 GLTKQEKQDILKEHNDFRQKIARGLETRGNPGPQPPAKNMKNLVWNDELAYVAQVWANQC 117
CAP3
Tx2_DIQUA 118 KYGHDACRNIERFQVGQNVAIRGSTGENLSTVDQMILSWYSEVDLMNKKYVSSFPKDDT- 176
VA3_SOLIN 119 TFEHDACRNVERFAVGQNIAATSSSGKNKSTPNEMILLWYNEVKDFDNRWISSFPSDDNI 178
VA5_VESVU 118 QYGHDTCRDVAKYQVGQNVALTGSTAAKYDDPVKLVKMWEDEVKDYNPKKK--F--SGND 173
CAP4
Tx2_DIQUA 177 YKKIGHYTQIVYGNTKTIGCGRIFYKD-GKWNQCYLVCNYGPAGNYPNAPVYQIKQ 231
VA3_SOLIN 179 LMKVGHYTQIVWAKTTKIGCGRIMFKEPDNWTKHYLVCNYGPAGNVLGAPIYEIKK 234
VA5_VESVU 174 FLKTGHYTQMVWANTKEVGCGSIKYIQ-EKWHKHYLVCNYGPSGNFMNEELYQTK- 227
CAP1
CAP2
196
213
100%
50%
0%
BLOSUM62

**B**


E137
H160
H99
Q112
C08
C102
N
C03
C18
C191
C174
C95
C29

```
gi|345480830|ref|XP_003424224.   1 ------MKIGLIGLVCIG-VFVLAESVSAGPYFDDEEEALASD--ADYSD 41NALENLMHAVQQK---------RACVRRNG
gi|345480832|ref|XP_003424225.   1 ------MKIGLIGLVCIG-VFVLAESVSAGPYFDDEEEALASD--ADYSD 41NALENLMHAVQQKRASWISMYRRACVRRNG
gi|350400650|ref|XP_003485909.   1 ------MKFSLFLLFCLF--AVFADKTLGSPYIDDEEDSLQVA-DSDYTD 41NALENLMRAAQQK---------RACIRRGG
gi|350400653|ref|XP_003485910.   1 ------MKFSLFLLFCLF--AVFADKTLGSPYIDDEEDSLQVA-DSDYTD 41NALENLMRAAQQKRTSLIYLFRRACIRRGG
gi|340711014|ref|XP_003394077.   1 ------MKFSLFLSICLF--VVFADRTLGSPYIDDEEDSLQVA-DFDYMD 41NALINLMRAPQQK---------RTCIRRGG
gi|340711016|ref|XP_003394078.   1 ------MKFSLFLSICLF--VVFADRTLGSPYIDDEEDSLQVA-DFDYMD 41NALINLMRAPQQKRTSLIYLFRRTCIRRGG
gi|328780473|ref|XP_003249808.   1 ------MKFGLFWLFCLLTVAIFTDRTFGSPYIDDEEDSLQIA-DSDYTD 43NALENLMRAAQQK---------RACIPRSG
gi|328780475|ref|XP_003249809.   1 ------MKFGLFWLFCLLTVAIFTDRTFGSPYIDDEEDSLQIA-DSDYTD 43NALENLMRAAQQKRTSLIYLFRRACIPRSG
gi|383850502|ref|XP_003700834.   1 ------MKFCLLWLFFLL--VVLADGTLGSPYIDDEEDSLQVA-DSDYTD 41NALENLMRAAQQK---------RACIKRGG
gi|383850504|ref|XP_003700835.   1 ------MKFCLLWLFFLL--VVLADGTLGSPYIDDEEDSLQVA-DSDYTD 41NALENLMRAAQQKRTSLIYLFRRACIKRGG
Contig1144_B2-2                  1 VSRVVGMRSCLLWLSILL--VVLADRSFASPYFDEEDDGLQAGGDSDYTD 48NALENLMRAAQQK---------RACIRRGG
gi|74845713|sp|Q5Y4U4.1|TXAG8_   1 ------MKSLLFVTIAVY----FVAQAVTANLLSN---FLGSS-LIDDDK 36GNMHKLYKRSEDQ-----------CIGRSC
consensus                        1       *.. *. ....    ..................*... ..*...*  .....*.......          ..*..*..
```

```
gi|345480830|ref|XP_003424224.   70 DCCFSSSCRCNLW 82GSNCRCQRVGLFQKWG---------- 98
gi|345480832|ref|XP_003424225.   79 DCCFSSSCRCNLW 91GSNCRCQRVGLFQKWG---------- 107
gi|350400650|ref|XP_003485909.   70 DCCYSSSCRCNLW 82GSNCQCQRMGLFQKWG---------- 98
gi|350400653|ref|XP_003485910.   79 DCCYSSSCRCNLW 91GSNCQCQRMGLFQKWG---------- 107
gi|340711014|ref|XP_003394077.   70 DCCYSSSCRCNLW 82GTNCQCQRMGLFQKWG---------- 98
gi|340711016|ref|XP_003394078.   79 DCCYSSSCRCNLW 91GTNCQCQRMGLFQKWG---------- 107
gi|328780473|ref|XP_003249808.   72 DCCYSSSCRCNLW 84GSNCRCQRMGLFQKWG---------- 100
gi|328780475|ref|XP_003249809.   81 DCCYSSSCRCNLW 93GSNCRCQRMGLFQKWG---------- 109
gi|383850502|ref|XP_003700834.   70 DCCNSSSCRCNLW 82GSNCRCQRVGLFQKWG---------- 98
gi|383850504|ref|XP_003700835.   79 DCCNSSSCRCNLW 91GSNCRCQRVGLFQKWG---------- 107
Contig1144_B2-2                  77 DCCFSSSCRCNLW 89GSNCQCQRMGLFQKWGASPL------ 109
gi|74845713|sp|Q5Y4U4.1|TXAG8_   63 SCCPYAACRCNLW 75KTSCKCQRTG--RKWATPCKEIYSPN 99
consensus                        91 .**....*******  ...*.***.*...**.          *
```

## Additional files

**Table S1 – Annotation of identified ESTs retrieved from randomly picked clones in the *D. quadriceps* venom gland cDNA library.**

Contigs assembled from Sanger sequencing data according to described in the 'materials and methods' section. The Uniprot and sequence identification (GI) numbers of homologous polypeptides are indicated. Number in parenthesis represents percentage of sequence similarity.

| Contig ID | Transcript Name | Length (bp) | Protein Definition | Matched species and maximum identify (%) | E-value | Uniprot / Gi number |
|---|---|---|---|---|---|---|
| Contig1 | TX01_VGDQ | 426 | Dinoponeratoxin Da-2501 | *Dinoponera australis* [ant]( 84,2) | $7.4x10^{-02}$ | 294863158 |
| Contig2 | TX02_VGDQ | 919 | CAP Venom allergen | *Solenopsis Invicta* [ant] (59,0) | $2x10^{-80}$ | 6136163 |
| Contig3 | HP03_VGDQ | 351 | Venom secreted hypothetical peptide | - | - | - |
| Contig4 | HP04_VGDQ | 767 | Hypothetical protein | *Harpegnathos saltator* [ant] (54,3) | $5x10^{-58}$ | 307212163 |
| Contig5 | TX05_VGDQ | 485 | ICK-like | *Androctonus amoreuxi* [scorpion] | $1.4x10^{-03}$ | PS01138 |
| Contig6 | TX06_VGDQ | 430 | Pilosulin | *Myrmecia banki* [ant] | $6x10^{-3}$ | Q68Y23 |
| Contig7 | HP07_VGDQ | 395 | Venom secreted hypothetical peptide | - | - | - |
| Contig8 | TX08_VGDQ | 184 | Altitoxin [(?)] | *Parabuthus transvaalicus* [scorpion] | 1.8 | PC1B5 |
| Contig9 | TX09_VGDQ | 578 | Dinoponeratoxin Da-3177 | *Dinoponera australis* [ant] (85,7) | $4x10^{-07}$ | 294863162 |
| Contig10 | - | 576 | No hit | - | - | - |
| Contig11 | CP11_VGDQ | 807 | Cytochrome c oxidase | *Camponotus vafer* [ant](69,0) | $2x10^{-53}$ | 359828649 |
| Contig12 | TX12_VGDQ | 585 | Chymotrypsin inhibitor | *Acromyrmex echinatior* [ant] | $1x10^{-19}$ | EGI67859 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Contig13 | HP13_VGDQ | 954 | Histone-lysine N-methyltransferase | *Hydra magnipapillata* [cnidarian] (76,9) | $5x10^{-3}$ | 449685068 |
| Contig14 | - | 357 | No hit | - | - | - |
| Contig15 | - | 388 | No hit | - | - | - |
| Contig16 | CP16_VGDQ | 804 | ATP synthase | *Vellenhovia benzai* [ant] (59,2) | $2x10^{-19}$ | 373432543 |
| Contig17 | TX17_VGDQ | 640 | PBP/GOBP Venom allergen 2/4-like | *Solenopsis saevissima* [ant] | $1x10^{-4}$ | A5X2H7 |
| Contig18 | CP18_VGDQ | 436 | 28S partial ribosomal RNA gene | *Dinoponera* sp. Pr01 voucher [ant] | 0.0 | DQ400983.1 |
| Contig19 | TX19_ VGDQ | 424 | Dinoponeratoxin Da-2501/Da-1585 | *Dinoponera australis* [ant] | 45 | P0CF01 |
| Contig20 | CP20_VGDQ | 717 | Mitochondrial DNA | mitochondrion *Pristomyrmex punctatus* [ant] | $2x10^{-155}$ | AB556947.1 |
| Contig21 | TX21_ VGDQ | 708 | PBP/GOBP Venom allergen 2/4-like | *Solenopsis invicta* [ant] | $7x10^{-5}$ | P35777 |

(?): Should be confirmed

**Table S2 – Overview of experimental data from the Next Generation RNA sequencing**

| Items | Number |
|---|---|
| Total transcripts | 18,546 |
| Max contig length(bp) | 8114 |
| Min contig length(bp) | 101 |
| Whole dataset length(bp) | 4,304,312 |
| Average contig lengths(bp) | 232.09 |
| N50 | 249 |
| N90 | 119 |

**Table S3 - Annotation of identified venomous contigs in *D. quadriceps* venom gland**

Contigs representing candidate toxin-like peptides based on the alignment of EST data set, prediction of signal peptides and ORFs, and BLASTx comparison. GI denotes sequence identification number of "top hit" toxin homologous. Number in parenthesis represents percentage of sequence similarity.

| Contig ID | Length (bp) | Protein Definition | Matched species and maximum identify (%) | E-value | Gi number |
|---|---|---|---|---|---|
| *Contig2 | 483 | Dinoponeratoxin Da-3177 | Dinoponera australis [ant]( 85.71) | 8e-07 | 294863162 |
| *Contig1 | 464 | Dinoponeratoxin Da-2501 | Dinoponera australis [ant]( 84.21) | 3e-06 | 294863158 |
| *Contig12 | 1051 | Venom allergen III; Flags: Precursor | Solenopsis invicta [ant](56.58) | 2e-72 | 14424466 |
| *Contig27 | 1656 | Venom allergen I; Flags: Precursor | Solenopsis invicta [ant](35.51) | 1e-44 | 75006920 |
| Contig141 | 3365 | Protein lethal(2)essential for life | Acromyrmex echinatior [ant](93.33) | 2e-33 | 332022897 |
| Contig351 | 3845 | Protein lethal(2)essential for life | Harpegnathos saltator [ant](86.90) | 1e-37 | 307197364 |
| Contig455 | 1497 | Probable multidrug resistance-associated protein lethal(2)03659 | Harpegnathos saltator [ant](86.25) | 4e-169 | 307196484 |
| Contig857 | 966 | Probable multidrug resistance-associated protein lethal(2)03659 | Harpegnathos saltator [ant](76.36) | 3e-131 | 307196484 |
| Contig1700 | 826 | Probable multidrug resistance-associated protein lethal(2)03659 | Harpegnathos saltator [ant](81.82) | 3e-36 | 307196484 |

| | | | | | |
|---|---|---|---|---|---|
| Contig5586 | 139 | Lethal(2)neighbour of tid protein | Harpegnathos saltator [ant](86.96) | 5e-19 | 307207923 |
| Contig8914 | 165 | Lethal(2)neighbour of tid protein | Acromyrmex echinatior [ant](85.19) | 9e-19 | 332018132 |
| Contig11936 | 176 | Lethal(2)neighbour of tid protein | Harpegnathos saltator [ant](91.38) | 6e-23 | 307207923 |
| Contig12935 | 189 | Lethal(3)malignant brain tumor-like 3 protein | Harpegnathos saltator [ant](85.48) | 2e-09 | 307202837 |
| Contig17436 | 120 | Pre-mRNA-splicing regulator female-lethal(2)D | Harpegnathos saltator [ant](87.18) | 2e-11 | 307211429 |
| Contig17643 | 159 | Probable multidrug resistance-associated protein lethal(2)03659 | Harpegnathos saltator [ant](77.50) | 6e-10 | 307207958 |
| Contig18270 | 128 | Probable multidrug resistance-associated protein lethal(2)03659 | Harpegnathos saltator [ant](90.48) | 4e-14 | 307196485 |
| Contig187 | 1410 | PREDICTED: phospholipase A2 isozymes PA3A/PA3B/PA5 | Apis mellifera[honey bee](85.51) | 9.00E-109 | 328778177 |
| Contig662 | 1603 | Putative phospholipase B-like lamina ancestor | Harpegnathos saltator [ant](60.42) | 8e-26 | 307214880 |
| Contig4464 | 699 | Phospholipase A2 isozyme PA4 | Harpegnathos saltator [ant](88.89) | 2.00E-35 | 307198397 |
| Contig5423 | 186 | Phospholipase A2 isozyme PA3A/PA3B/PA5 | Acromyrmex echinatior [ant](83.93) | 2e-18 | 332029748 |
| Contig6851 | 234 | Lysophospholipase-like protein 1 | Harpegnathos saltator [ant](89.80) | 5e-17 | 307200032 |
| Contig9856 | 272 | Phospholipase A-2-activating protein | Harpegnathos saltator [ant](84.85) | 2e-22 | 307196404 |

| | | | | | |
|---|---|---|---|---|---|
| Contig10011 | 233 | Group XIIA secretory phospholipase A2 | Camponotus floridanus [ant](87.01) | 1e-34 | 307187781 |
| Contig12747 | 163 | Phospholipase B1, membrane-associated | Acromyrmex echinatior [ant](53.70) | 9e-06 | 332028550 |
| Contig13100 | 134 | Phospholipase B1, membrane-associated | Harpegnathos saltator [ant](75.61) | 3e-11 | 307193256 |
| Contig17494 | 145 | Phospholipase D3 | Camponotus floridanus [ant](97.92) | 6e-11 | 307176418 |
| Contig18196 | 185 | Phospholipase DDHD2 | Harpegnathos saltator [ant](91.80) | 8e-28 | 307211672 |
| Contig1144 | 495 | PREDICTED: U8-agatoxin-Ao1a-like isoform 1 | Bombus impatiens[bee](85.71) | 6e-39 | 350400650 |
| Contig326 | 451 | PREDICTED: venom carboxylesterase-6-like | Bombus terrestris[bee](56.14) | 2e-31 | 340722695 |
| Contig391 | 921 | PREDICTED: venom carboxylesterase-6-like | Apis florea[honeybee](63.25) | 3e-102 | 380027409 |
| Contig584 | 933 | PREDICTED: venom carboxylesterase-6 isoform 1 | Nasonia vitripennis [bee](69.77) | 1e-103 | 345497204 |
| Contig1381 | 150 | PREDICTED: venom carboxylesterase-6-like | Apis mellifera [honeybee](75.51) | 8e-13 | 48097744 |
| Contig1632 | 149 | PREDICTED: venom carboxylesterase-6-like | Megachile rotundata [bee](83.67) | 1e-16 | 383858940 |
| Contig3130 | 165 | PREDICTED: venom carboxylesterase-6-like | Megachile rotundata [bee](64.81) | 2e-12 | 383858940 |
| Contig3290 | 474 | PREDICTED: venom carboxylesterase-6 | Nasonia vitripennis [bee](64.84) | 2e-44 | 345497204 |

| | |
|---|---|
| | isoform 1 |

* represent contig was supported by the ESTs data.

**Table S4 - Summary of toxin-like peptide candidates from *D. quadriceps* venom gland transcriptome**

Validation of some RNA-Seq assembled contigs according to the mapped EST fragments. Full-length open reading frames (ORFs) are indicated with a "+" sign. The presence or absence of a signal peptide in the ORFs is indicated by a "Y" (yes) or "N" (no).

| Contig ID | Length | Mapped EST fragments | ORF | Signal peptide |
|---|---|---|---|---|
| Contig25 | 539 | 26 | + | Y |
| Contig3 | 2257 | 22 | + | N |
| Contig4 | 476 | 49 | + | Y |
| Contig10 | 254 | 20 | + | N |
| Contig5 | 482 | 34 | + | N |
| Contig8 | 675 | 16 | + | Y |