Conotoxin Diversity in *Chelyconus ermineus* (Born, 1778) and the Convergent Origin of Piscivory in the Atlantic and Indo-Pacific Cones

Samuel Abalde[1], Manuel J. Tenorio[2], Carlos M.L. Afonso[3], and Rafael Zardoya[1,*]

[1]Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain

[2]Departamento CMIM y Q. Inorgánica-INBIO, Facultad de Ciencias, Universidad de Cadiz, Puerto Real, Spain

[3]Fisheries, Biodiversity and Conervation Group, Centre of Marine Sciences (CCMAR), Universidade do Algarve, Campus de Gambelas, Faro, Portugal

*Corresponding author: E-mail: rafaz@mncn.csic.es.

Accepted: July 28, 2018

Data deposition: Raw RNA seq data:

SRA database: project number SRP139515 (SRR6983161-SRR6983169)

Abstract

The transcriptome of the venom duct of the Atlantic piscivorous cone species *Chelyconus ermineus* (Born, 1778) was determined. The venom repertoire of this species includes at least 378 conotoxin precursors, which could be ascribed to 33 known and 22 new (unassigned) protein superfamilies, respectively. Most abundant superfamilies were T, W, O1, M, O2, and Z, accounting for 57% of all detected diversity. A total of three individuals were sequenced showing considerable intraspecific variation: each individual had many exclusive conotoxin precursors, and only 20% of all inferred mature peptides were common to all individuals. Three different regions (distal, medium, and proximal with respect to the venom bulb) of the venom duct were analyzed independently. Diversity (in terms of number of distinct members) of conotoxin precursor superfamilies increased toward the distal region whereas transcripts detected toward the proximal region showed higher expression levels. Only the superfamilies A and I3 showed statistically significant differential expression across regions of the venom duct. Sequences belonging to the alpha (motor cabal) and kappa (lightning-strike cabal) subfamilies of the superfamily A were mainly detected in the proximal region of the venom duct. The mature peptides of the alpha subfamily had the α4/4 cysteine spacing pattern, which has been shown to selectively target muscle nicotinic-acetylcholine receptors, ultimately producing paralysis. This function is performed by mature peptides having a α3/5 cysteine spacing pattern in piscivorous cone species from the Indo-Pacific region, thereby supporting a convergent evolution of piscivory in cones.

Key words: conotoxin, conopeptide, convergence, transcriptome, Conidae, expression.

Introduction

The family of Conidae (Fleming, 1822 sensu lato) that includes cone snails is well known for their astonishing species diversity (> 800 species; Tucker and Tenorio 2013) as well as for their sophisticated feeding behavior, which includes the production and injection of venom in preys through a specialized harpoon-like radular tooth (Salisbury et al. 2010; Dutertre et al. 2014; Olivera et al. 2015). Although all cone snails were traditionally classified into the single genus *Conus*, recent phylogenetic studies based on morphological (Tucker and Tenorio 2009) and molecular (Puillandre, Bouchet, et al. 2014; Uribe et al. 2017) data supported the split of *Conus* into several lineages, which are ranked either at the family or genus levels, respectively. According to Puillandre, Duda, et al. (2014) and Uribe et al. (2017), the following six genera are recognized: *Profundiconus*, *Californiconus*, *Lilliconus*, *Pygmaeconus*, *Conasprella*, and *Conus*. The latter genus holds most of the species diversity with up to 60 monophyletic groups, either recognized as subgenera (Puillandre, Duda, et al. 2014) or genera (Tucker and Tenorio 2009) depending on the author (herein we will use the taxonomy of Tucker and Tenorio 2009).

The last common ancestor of Conidae likely fed on worms, as most of all living species have been suggested to do (Duda et al. 2001; Puillandre, Bouchet, et al. 2014). During the evolution and diversification of the group, there was suggested to be one diet shift to prey on other snails in the last common ancestor of genera *Calibanus*, *Cylinder*, *Conus*, *Darioconus, Eugeniconus*, and *Leptoconus* (Puillandre, Bouchet, et al. 2014). Instead, phylogenetic analyses suggested at least two diet shifts to prey on fishes as the Atlantic/Eastern Pacific genus *Chelyconus* did not share a most recent common ancestor with Indo-Pacific piscivorous genera: *Phasmoconus*, *Gastridium, Pionoconus*, *Textilia*, *Afonsoconus*, *Embrikena*, and *Asprella* (for the latter three there is no direct observation of prey capture; Duda et al. 2001; Duda and Palumbi 2004; Puillandre, Bouchet, et al. 2014; Olivera et al. 2015). Here, we reconstructed a simplified maximum likelihood (ML) phylogeny of cone snails based on complete mitochondrial (mt) genomes showing the same evolutionary trends in feeding behavior (fig. 1). A remarkable singularity within the group is *Californiconus californicus*, which has a diverse diet including fish, snails, worms, and shrimps (Biggs et al. 2010). The shape and number of barbs of the hollow radular tooth as well as the feeding behavior of cone snails appear to be, at least in some cases, adapted to capturing most efficiently the different types of prey. For instance, some molluscivorous cones make successive injections of radular teeth into the prey (Prator et al. 2014) whereas piscivorous cones show up to three different hunting modes including electrical stunning and tethering of single preys using the proboscis, engulfing of several prey fish at once by the rostrum, and flailing the proboscis around the fish without tethering (Olivera et al. 2015).

The success of a strike relies on the high efficacy of the injected venoms, which readily elicit sedation, paralysis or sensory overload in the prey (Robinson, Li, Bandyopadhyay, et al. 2017). Cone venoms are for the most part complex mixtures of short bioactive peptides termed conotoxins or conopeptides (Lavergne et al. 2013; Robinson and Norton 2014). It is possible to distinguish at least two different components based on their ultimate target: 1) specific peptides, which target voltage-gated and ligand-gated ion channels, neurotransmitter transporters, and receptors in the central and peripheral nervous system of the preys (Olivera 2002; Terlau and Olivera 2004; Lewis et al. 2012; Lavergne et al. 2013) and 2) hormone/neuropeptide-like components, which target neuroendocrine processes in the prey (Safavi-Hemami et al. 2015; Robinson, Li, Bandyopadhyay, et al. 2017). In addition, the venom duct also produces several proteins that are involved in the processing of conotoxins or in enhancing venom activity (Safavi-Hemami et al. 2010, 2014; Hu et al. 2011; Terrat et al. 2012; Barghi et al. 2015b). Each cone species biosynthesizes in the apical secretory cells of the venom duct (Endean and Duchemin 1967) its own venom profile, which shows remarkable intraspecific variability (Davis et al. 2009; Rivera-Ortiz et al. 2011; Rodriguez et al. 2015; Chang and Duda 2016; Peng et al. 2016) as well as striking changes in composition over time within single individuals (Dutertre et al. 2010; Prator et al. 2014). Moreover, it has been shown that conotoxin expression is regionalized along the venom duct (Garrett et al. 2005; Tayo et al. 2010; Hu et al. 2012; Jin et al. 2015; Prashanth et al. 2016) and that the proximal and distal regions (with respect to the bulb) of the venom duct of several cones produce distinct defense- and predation-evoked conotoxin cocktails, respectively (Dutertre et al. 2014).

Conotoxins are generally synthesized as precursors with a typical three domain structure including: 1) a highly conserved hydrophobic N-terminal signal region, which guides the conotoxin precursor to the endoplasmic reticulum and the cellular secretory pathway; 2) an intervening, moderately conserved propeptide region, which for some conotoxins participates in secretion, post-translational modification and folding (Bandyopadhyay et al. 1998; Conticello et al. 2003; Buczek et al. 2004); and 3) a C-terminal region, which constitutes the mature functional peptide (Woodward et al. 1990; Kaas et al. 2010). The conserved sequence profiles of the signal region have been used to classify conotoxin precursors into 40–50 protein superfamilies, originally named with alphabet letters (Terlau and Olivera 2004; Corpuz et al. 2005; Kaas et al. 2010; Puillandre et al. 2012; Lavergne et al. 2013; Robinson and Norton 2014; Li et al. 2017). Furthermore, sequence comparison of mature peptides revealed up to 26 conserved cysteine frameworks named with roman numbers, which generally correlate with conotoxin precursor superfamilies (Lavergne et al. 2015), although most of these protein superfamilies have been shown to have more than one cysteine framework (Terlau and Olivera 2004; Corpuz et al. 2005; Luo et al. 2006; Kaas et al. 2010; Puillandre et al. 2010, 2012; Robinson and Norton 2014; Lavergne et al. 2015). Notably, some mature peptides are cysteine-poor or completely lack cysteines (Puillandre et al. 2012).

The renowned conotoxin hyperdiversity, which is typical for gene families that mediate interactions between organisms (Conticello et al. 2001; Barghi et al. 2015a), can be explained by a combination of several (evolutionary) processes including: 1) extensive gene duplication (Duda and Palumbi 1999; Espiritu et al. 2001; Puillandre et al. 2010; Chang and Duda 2012); 2) high mutation rates and diversifying selection of the mature domain (Conticello et al. 2001); 3) recombination events (Espiritu et al. 2001; Wu et al. 2013); and 4) variable peptide processing (Dutertre et al. 2013) and post-translational modifications (Lu et al. 2014; Peng et al. 2016).

The advent of high-throughput sequencing techniques, and in particular of RNA sequencing (first using the 454 GS FLX Titanium and currently the Illumina HiSeq platforms) has produced a quantum leap in the characterization of whole conotoxin precursor repertoires (Prashanth et al. 2012; Barghi et al. 2015b) compared with the more traditional sequencing

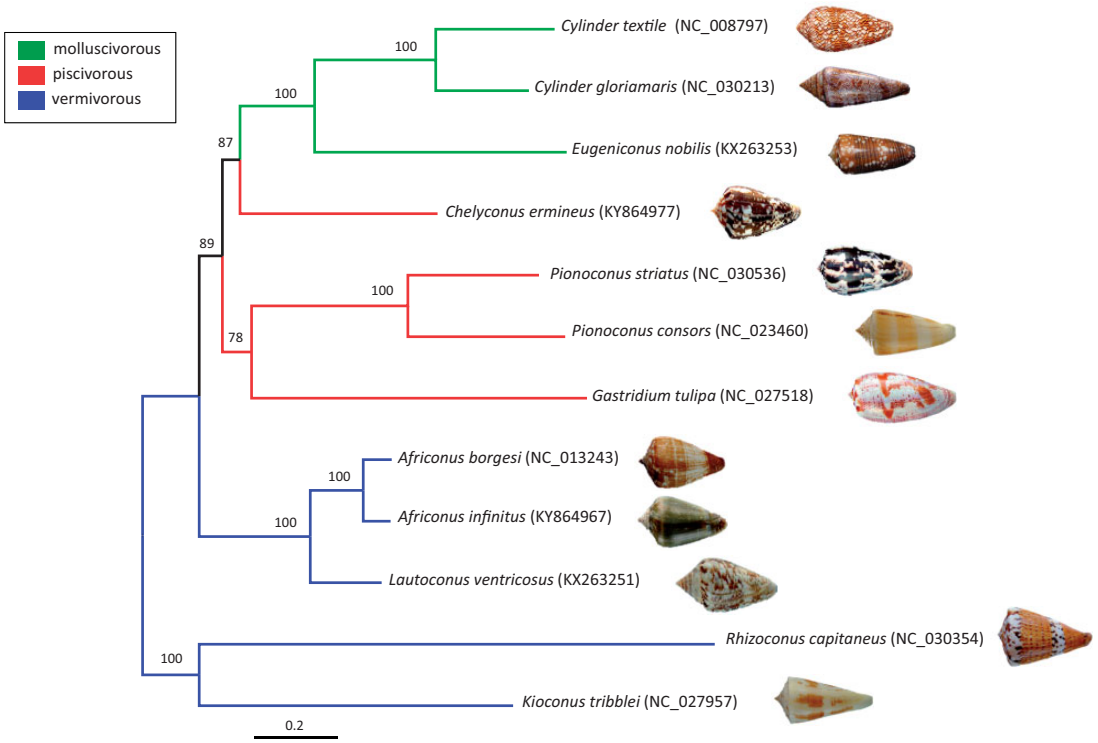


Fig. 1.—Simplified ML phylogeny of cone snails based on complete mitochondrial genomes (concatenated 13 protein-coding genes plus two rRNA genes analyzed at the nucleotide level). The evolution of diet is mapped onto the phylogeny. Bootstrap values are indicated above each node. Scale bar indicates substitutions/site. GenBank accession numbers are indicated for each species mt genome.

of individual cDNA clones (Pi, Liu, Peng, Jiang, et al. 2006; Pi, Liu, Peng, Liu, et al. 2006; Liu et al. 2012; Lu et al. 2014). RNA sequencing is highly sensitive and even rare transcripts with low expression levels can be identified (Barghi et al. 2015b). Therefore, several recent studies have determined the complete conotoxin diversity of 1) piscivorous species such as *Textilia bullatus* (Hu et al. 2011), *Pionoconus consors* (Terrat et al. 2012; Violette et al. 2012), *Gastridium geographus* (Hu et al. 2012; Dutertre et al. 2014; Safavi-Hemami et al. 2014), and *Pionoconus catus* (Himaya et al. 2015); 2) molluscivorous species such as *Conus marmoreus* (Dutertre et al. 2013; Lavergne et al. 2013), *Darioconus episcopatus* (Lavergne et al. 2015), and *Cylinder gloriamaris* (Robinson, Li, Lu, et al. 2017); and 3) vermivorous species such as *Puncticulis pulicarius* (Lluisma et al. 2012), *Rhizoconus miles* (Jin et al. 2013), *Kioconus tribblei* (Barghi et al. 2015b), *Dendroconus betulinus* (Peng et al. 2016), and the congeneric species *Turriconus andremenezi* and *Turriconus praecellens* (Li et al. 2017), among others. Altogether, these transcriptome studies show that each cone species produces at least 100–400 different conotoxin precursors, and have been very successful in discovering new superfamilies and cysteine frameworks (Barghi et al. 2015b; Lavergne et al. 2015; Peng et al. 2016; Li et al. 2017). The emerging general pattern is that several gene superfamilies are widespread among cone species, although with different degrees of expansion, whereas others are restricted to a few lineages (Duda and Remigio 2008; Puillandre et al. 2012).

To further contribute to the cataloguing of conotoxin diversity in the main lineages of cone snails, we characterized the full transcriptome of *Chelyconus ermineus* (Born 1778). This is a cone species, which can be found on both shores of the Atlantic Ocean and feeds on fishes. Together with *Chelyconus purpurascens* from the Eastern Pacific region, it forms a clade, which according to the phylogeny of Conidae (Duda and Palumbi 2004; Puillandre, Bouchet, et al. 2014; fig. 1) may underwent a shift to piscivory independent to the one occurred in the ancestor of Indo-Pacific piscivorous genera. Thus far, the study of conotoxins in *C. ermineus* has been limited to the identification of few mature peptides (Martinez et al. 1995; Jacobsen et al. 1997; Barbier et al. 2004; Rivera-Ortiz et al. 2011; Echterbille et al. 2017) and some conotoxin precursors belonging to the A, B1, and O1 superfamilies (Duda and Palumbi 2004; Gowd et al. 2008; Puillandre et al. 2010; 18 entries in ConoServer, Kaas et al. 2012). Notably, however, some of the mature conotoxins of *C. ermineus* such as EVIA are among the few peptides whose functional activity (Barbier et al. 2004) and tertiary structure (Volpon et al. 2004) have been determined experimentally. There are a few more conotoxin precursors identified in the closely related *C. purpurascens* belonging to the A, B1, M, O1, and T superfamilies (e.g., Shon et al. 1995, 1998; Duda and

Palumbi 2004; 45 entries in ConoServer, Kaas et al. 2012). By sequencing the transcriptome of *C. ermineus*, the first one of an Atlantic cone species, we aimed: 1) to catalogue the diversity of conotoxin precursors in this species and classify them into superfamilies; 2) to identify other proteins that are transcribed in the venom duct and are potentially involved in the processing of conotoxins or in enhancing venom activity; 3) to estimate intraspecific variation of conotoxin precursors; 4) to determine differences in the spatial distribution of conotoxin precursors along the distal, medium, and proximal regions of the venom duct; 5) to quantify the expression levels of conotoxin genes in the different individuals and along the venom duct; and 6) to compare the venom composition of Atlantic and Indo-Pacific piscivorous cones, and to identify putative differences with the venoms of cones preying on snails and worms.

Materials and Methods

Sampling and RNA Extraction

Three adult specimens of *C. ermineus* were captured, respectively, in Boa Vista (CVERM3; hereafter ERM1), Sal (CVERM13; hereafter ERM2), and Santa Luzia (CV1446; hereafter ERM3) islands in Cabo Verde with corresponding permits (table 1). Each individual, in a resting stage, was extracted from the shell and dissected to remove the venom duct, which was excised into three equal parts: proximal, medium, and distal with respect to the venom bulb (following Tayo et al. 2010). These fragments were stored in 1 ml RNA*later* (Invitrogen, Life Technologies), first at 4°C and for the long term at −20°C.

For RNA extraction, each venom duct portion was incubated independently in a 2 ml eppendorf with 500 µl of TRIzol LS Reagent (Invitrogen, Life Technologies) and grinded with ceramic beads in a Precellys Evolution tissue homogenizer. The solution was mixed with 100 µl of chloroform. After centrifugation (12,000 x g for 15 minutes at 4°C), the aqueous phase was recovered and RNA precipitated with 250 µl of isopropanol and stored overnight at −80°C. The Direct-zol RNA miniprep kit (Zymo Research, Irvine) was used to purify total RNA (5–15 µg) following manufacturer's instructions.

Library Preparation and Sequencing

Dual-indexed cDNA libraries (307–345 bp insert average size) for each sample were constructed using the TruSeq RNA Library Prep Kit v2 (Illumina, San Diego) and following manufacturer's instructions at Sistemas Genómicos (Valencia, Spain). Briefly, the poly(A)+ mRNA fraction was isolated using oligo-$(dT)_{25}$ magnetic beads. Subsequently purified mRNA was chemically fragmented prior to reverse transcription and the construction of the cDNA library. The quality of the libraries was analyzed with the TapeStation 4200, High Sensitivity assay; the quantity of the libraries was determined by real-time PCR in LightCycler 480 (Roche). The pool of libraries (including other cone species for different projects) was split into different lanes and sequenced by paired-end sequencing (100×2) in an Illumina HiSeq2500 (two flowcells) following standard procedures at Sistemas Genómicos (Valencia, Spain).

Assembly

The reads corresponding to the different regions of the venom duct and individuals were sorted using the corresponding library indices. Adapter sequences were removed using SeqPrep (St John 2011). Assembly was performed using the TRUFA webserver (Kornobis et al. 2015). Briefly, the quality of the sequencing was checked using FastQC v.0.10.1 (Andrews 2010). Ends of reads were trimmed (PHRED < 30) and resulting trimmed reads were filtered out according to their mean quality scores (PHRED < 20) using PRINSEQ v.0.20.3 (Schmieder and Edwards 2011). This step also ensured minimizing cross-contamination resulting from potential index misassignment, as this tends to be associated to low quality scores (Wright and Vetsigian 2016). Filtered reads were used for de novo assembly of trancriptomes with Trinity r2012-06-08 (Grabherr et al. 2011) with default settings (minimum contig length: 200; sequence identity threshold: 0.95). The transcriptome raw reads produced in this project have been deposited at the NCBI SRA database under accession SRP139515 (see also table 1).

Prediction and Annotation of Conotoxin Precursors and Associated Proteins

The sequences of all conotoxin precursors and associated proteins of cone venoms available in GenBank release 217 (Benson et al. 2005), Uniprot release 2016_11 (Uniprot Consortium 2017), and ConoServer release 12-26-2016 (Kaas et al. 2012) were downloaded in December 26, 2016 to construct a local reference database. Redundant entries from the three databases were removed. Subsequently, BLASTX was used to identify those sequences encoding putative conotoxin precursors and associated proteins (with an E-value of 1e-5) among the assembled contigs by similarity searches against the reference database. These sequences were translated into amino acids using the universal genetic code and manually inspected, in order to discard false positives (hits not corresponding to canonical conotoxins) or assembly artifacts (due to indels that interrupt open reading frames). Duplicate and highly truncated (>55% of the estimated total length of a precursor) sequences were removed to produce the final working list of conotoxin precursors and associated proteins of *C. ermineus* (provided in supplementary table 1, Supplementary Material online). The three domains of the predicted conotoxin precursors and the cysteine frameworks of the mature peptides were identified using the Conoprec tool (Kaas et al. 2012). Assignment of amino acid

Table 1
Specimens of *Chelyconus ermineus* Analyzed in This Study and Main Statistics of Illumina Sequencing and Assembly

Specimen	Voucher ID MNCN	Island	Segment	SRA Accesion No.	# Raw Reads	# Clean Reads	# Contigs	% Mapping[a]	# Conotoxins	% Mapping[b]
ERM1	15.05/80980	Boa Vista	Proximal	SRR6983168	13,023,114	12,882,970	64,233	92	59	61
ERM1	15.05/80980	Boa Vista	Medium	SRR6983169	25,823,481	25,541,087	69,836	83	75	70
ERM1	15.05/80980	Boa Vista	Distal	SRR6983166	27,702,513	27,160,103	119,384	88	117	17
ERM2	15.05/80013	Sal	Proximal	SRR6983167	26,754,509	26,754,509	52,506	69	75	69
ERM2	15.05/80013	Sal	Medium	SRR6983164	26,986,678	26,986,220	57,887	76	89	63
ERM2	15.05/80013	Sal	Distal	SRR6983165	26,107,666	26,107,195	73,809	91	109	40
ERM3	15.05/78606	Santa Luzia	Proximal	SRR6983162	27,163,849	27,163,368	49,195	76	71	78
ERM3	15.05/78606	Santa Luzia	Medium	SRR6983163	31,223,312	31,222,733	68,103	92	90	58
ERM3	15.05/78606	Santa Luzia	Distal	SRR6983161	31,717,505	31,716,948	71,785	56	83	58

[a]Percentage of clean reads that map onto assembled contigs.
[b]Percentage of clean reads that map onto assembled conotoxin precursors.

sequences to different superfamilies was based on the two highest scoring full-length conotoxin precursor hits in the BLAST results and taking into account the percentage of sequence identity (>90%) to the highly conserved signal region (Robinson and Norton 2014; Barghi et al. 2015b). Further refinement of the superfamily assignment was achieved by aligning conotoxin precursor amino acid sequences of *C. ermineus* to selected canonical representatives of each superfamily using Mafft v7 (Katoh and Standley 2013) with default parameters (see supplementary file 1, Supplementary Material online). This step revealed important diversity (i.e., presence of potential paralogs) at the propeptide domain within several superfamilies, which was further analyzed. All *C. ermineus* conotoxin precursor amino acid sequences are deposited (as nucleotide sequences) in GenBank under accession numbers MH360289–MH360712.

Phylogenetic Analyses of the M and T Superfamilies

In order to infer the evolutionary origin of cysteine-poor conotoxins, we performed phylogenetic analyses of the M and T conotoxin precursor superfamilies, which have both cysteine-rich and cysteine poor members. Concatenated amino acid alignments of the signal and propeptide domains of the M and T superfamilies, respectively, were constructed using Mafft v7 (Katoh and Standley 2013) with default parameters. Phylogenetic relationships were inferred using ML (Felsenstein 1981) with PhyML v3.0 (Guindon et al. 2010) with default settings in the ATGC platform (http://www.atgc-montpellier.fr/phyml/; last accessed September 05, 2018) and using the smart model selection option. Statistical support was assessed with 1,000 bootstrap pseudoreplicates (BP).

Expression Analyses

Approximate relative expression levels were estimated by mapping only clean reads, back to all assembled contigs of *C. ermineus*. TPM (transcripts per kilobase million), which normalize for gene length and sequencing depth, were estimated with the RSEM package (which uses the mapper Bowtie 2; Langmead and Salzberg 2012) included in Trinity r2012-06-08 (Grabherr et al. 2011). In addition, we run the EBSeq software (Leng et al. 2013) as implemented in Trinity to estimate the posterior probability of being differentially expressed (PPDE), setting the False Discovery Rate (FDR) at 0.95, of conotoxins as a whole and of each of the different superfamilies along the different regions of the venom duct using the three individuals as biological replicates.

Reconstruction of Cone Snail Phylogeny

In order to determine diet shifts during the evolutionary history of cone snails, a simplified phylogeny was reconstructed using ML based on complete mt genomes (13 protein-coding and two rRNA genes) available in GenBank. Protein-coding genes were individually aligned using TranslatorX (Abascal et al. 2010), which generates a nucleotide alignment based on corresponding deduced amino acid alignments. The rRNA genes were aligned using Mafft v7 (Katoh and Standley 2013). All ambiguously aligned positions were removed using GBlocks v.0.9.1b (Castresana 2000) with the following settings: minimum sequence for flanking positions: 85%; maximum contiguous nonconserved positions: 8; minimum block length: 10; gaps in final blocks: no. Finally, the different single alignments were concatenated using Geneious 8.1.8.

The best-fit partition scheme and models of substitution for the data set were identified using PartitionFinder (Lanfear et al. 2012) with the Bayesian Information Criterion (Schwarz 1978). The following partitions were tested: all genes together, all genes arranged in subunits (*atp*, *cob*, *cox*, *nad*, and *rrn*), and all genes separated (except atp6-atp8 and nad4-nad4L). In addition, we also tested separately the three codon positions in the protein-coding genes. The best partition scheme was the one considering each codon position separately, all protein-coding genes concatenated, and

rRNA genes concatenated. For each partition, the selected best-fit model was GTR +I + G.

In order to reconstruct the ML tree, we used RAxML-HPC2 on XSEDE 8.2.10 (Stamatakis 2014) as implemented in the CIPRES Science Gateway v 3.3 (http://www.phylo.org/; last accessed September 05, 2018) with the rapid hill-climbing algorithm and 1,000 BP. The outgroups were *Rhizoconus capitaneus* and *K. tribblei* based on Puillandre et al. (2014).

Results

Sequencing and Assembly

A total of nine samples were sequenced corresponding to three regions (distal, medium, and proximal) of the venom duct of three individuals (ERM1-3) of *C. ermineus*. The main statistics associated to the sequencing and assembly procedures are summarized in table 1. Sequencing generated between 13 and 32 million raw reads per sample. Most (99–100%) of the reads were kept as clean after adapter and quality trimming. The number of assembled contigs varied between 49,195 and 119,384 with a mean of 69,637.6 per sample. Mapping of clean reads onto assembled contigs indicated that on an average 80% of the reads were used for further analyses (table 1). After BLASTX searches against a local reference database, the number of distinct (with at least one amino acid difference) putative conotoxin precursor sequences per sample (i.e., venom duct portion of an individual) varied between 59 and 117 (table 1). The majority of these sequences were full-length but a few were slightly truncated at the N- or C-terminus (see supplementary table 1, Supplementary Material online). Mapping of clean reads onto assembled transcripts indicated that conotoxin production made up on an average 57% of the transcriptome in the venom duct, although there was important variability among venom duct regions within the same individual (table 1).

Intraspecific Variation in Venom Composition

The sequences of the conotoxin precursor transcripts expressed in the venom duct of individuals ERM1-3 of *C. ermineus* inhabiting three different islands (Boa Vista, Sal, and Santa Luzia) of the archipelago of Cabo Verde were determined (fig. 2). The total numbers of inferred conotoxin precursors in each specimen were 161, 176, and 144, respectively. Of these, 25 were found in all specimens, nine were common to ERM1 and ERM2 (but not found in ERM3), 18 were shared by ERM1 and ERM3 (not present in ERM2), and 26 were common to ERM2 and ERM3 (and not found in ERM1). The number of peptides common to all three specimens rose up to 33 when only differences in the functional mature peptide were taken into consideration (fig. 2). In such analyses, the numbers of sequences exclusive to each specimen were 96, 104, and 56, respectively (fig. 2). We also estimated intraspecific diversity taking into account putative allele variation by clustering together sequences, which diverged in one or less, two or less, and three or less amino acids, respectively. The numbers of conotoxin precursor sequences common to all three specimens rose up to 44, 48, and 53, respectively (supplementary fig. 1, Supplementary Material online).

Diversity of Conotoxin Precursor Sequences in *C. ermineus*

Of the 422 distinct transcripts related to venom activity identified as produced in the venom duct of the three individuals of *C. ermineus*, a total of 296 could be assigned (based on the signal region sequence) to 33 known conotoxin precursor superfamilies already described in other cone venom ducts (fig. 3). In addition, 82 conotoxin precursor sequences were grouped, using reciprocal BLASTs and taking into account a 90% identity threshold per superfamily, into 22 unassigned conotoxin superfamilies, not formally described in other cone species but also present in some of them (see supplementary file 1, Supplementary Material online). Finally, 44 peptides corresponded to six associated protein families (see supplementary table 1 and file 1, Supplementary Material online). All but one (alpha conotoxin EI; P50982) previously reported conotoxin precursors and mature peptides from *C. ermineus* (Martinez et al. 1995; Jacobsen et al. 1997; Barbier et al. 2004; Duda and Palumbi 2004; Gowd et al. 2008; Puillandre et al. 2010; Rivera-Ortiz et al. 2011) were detected (see supplementary file 2, Supplementary Material online). Homologs to conotoxin precursors and mature peptides from *C. purpurascens* were identified as well (Shon et al. 1995, 1998; Duda and Palumbi 2004). The six most diverse conotoxin precursor superfamilies (in terms of the number of distinct members) were T, W, O1, M, O2, and Z, accounting for 57% of all observed diversity (fig. 3). Conversely, several of the conotoxin precursor superfamilies were restricted to only one or two representatives (e.g., A2, E, J, K, P, R, and several unassigned superfamilies).

Some of the inferred mature domains in the conotoxin precursors showed no cysteine framework. In some cases, these mature peptides belonged to superfamilies exclusively formed by members without cysteines, such as the W and Z superfamilies. In other cases, mature conotoxins with and without framework were grouped together within the same superfamily (fig. 3). Notably, the M and T superfamilies had both types of mature conotoxins. Phylogenetic trees of both superfamilies were reconstructed based on the amino acid sequences of the signal and propeptide regions, and allowed distinguishing several paralog groups within each conotoxin precursor superfamily (fig. 4). While mature conotoxins without cysteine framework form a distinct paralog group within the M superfamily, they seem to have originated independently and recurrently in the different paralogs within the T superfamily (fig. 4).

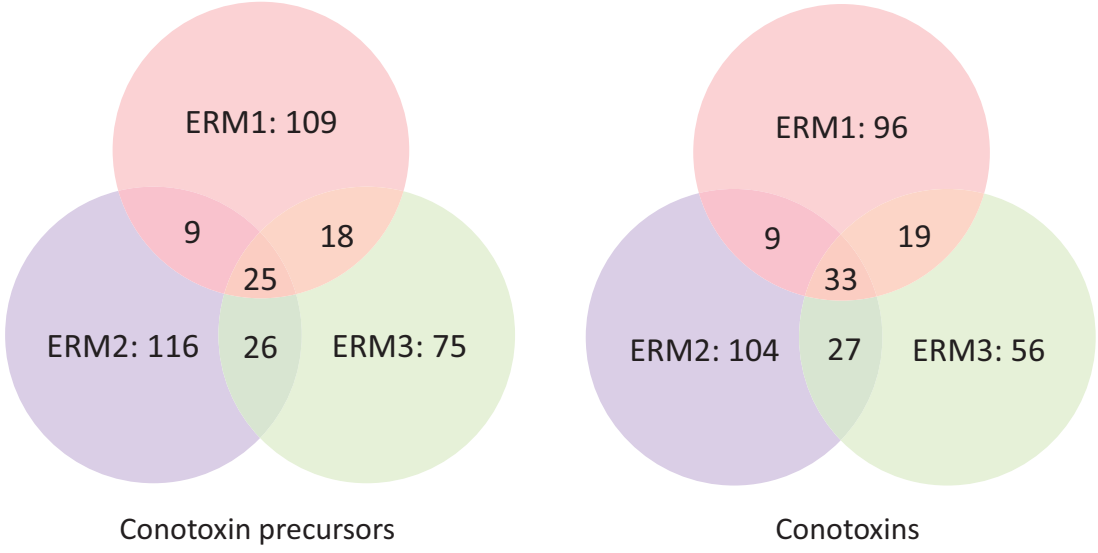


Fig. 2.—Distinct conotoxin precursors (left) and mature peptides (right) identified in the three analyzed individuals of *Chelyconus ermineus*.

Within unassigned superfamilies, most precursors have the canonical three-domain structure (see signal sequences in table 2). However, an interesting case was that of several precursors (unassigned superfamilies 16–22), which could not be assigned to any known superfamily because they lacked a signal region, but had a mature peptide, which could be confidently aligned to the mature peptides normally associated to the O3 and T superfamilies in other cone species (see supplementary file 1, Supplementary Material online). In trying to assign these sequences to known superfamilies, we used hidden Markov model searches as implemented in Conodictor (Koua et al. 2012) but retrieved no significant hit.

Differential Spatial Distribution of Conotoxin Precursor Transcripts along the Venom Duct

Changes in the diversity of conotoxin precursors were analyzed along the three regions (distal, medium, and proximal) of the venom duct (fig. 5). The numbers of distinct conotoxin precursor sequences found in the distal, medium, and proximal regions were 190, 162, and 130, respectively. Up to 30–33 (depending on the individual) precursor transcripts were expressed throughout the venom duct (fig. 5). Of these, seven were common to the three individuals (not shown). A total of 33–64, 24–34, and 15–25 conotoxin precursors were found exclusively in the distal, medium, and proximal portions, respectively (fig. 5). Most conotoxin precursor superfamilies were identified in the three regions of the venom duct with the exception of E, K, and Thyrostimulin β, which were only detected in the distal portion; L and Q, which were missing in the proximal portion; A2, which was absent in the medium portion; and J, which was missing in the distal portion (fig. 5). In general, the diversity of members of the different conotoxin precursor superfamilies was relatively uniform across venom duct segments. However, the H, I2, I3, L, M, O1, Q, T, and W superfamilies, and the hormone conophysin showed more diversity toward the distal portion; A2 and B2 superfamilies had more diversity toward the proximal portion; and most of the members of Z superfamily were detected in the medium portion (fig. 5).

Expression of Conotoxin Precursor Transcripts along the Venom Duct

The relative expression levels of the different conotoxin precursor superfamilies along the distal, medium and proximal regions of the venom duct in the three individuals were estimated as TPMs, thus normalizing for gene length and sequencing depth (fig. 6). Expression levels varied extensively among individuals hindering the inference of expression patterns. TPM values were declared reliable when they were of similar level in at least two out of the three individuals. Taking this into consideration, the most expressed conotoxin precursor superfamilies were: A, which accounted for much of the conotoxin precursor expression in the medium and proximal fractions, O2, which was expressed abundantly in the distal region; and O1, which showed expression throughout the venom duct but particularly in the distal and medium portions (fig. 6*A*). A second batch of midexpressed superfamilies included: T, which showed higher values in the distal region; M, which was mostly expressed in the proximal region; and S with expression levels higher in the medium and distal regions (fig. 6*A*). The remaining conotoxin precursor superfamilies had much lower expression levels, mostly concentrated in the medium and distal regions of the venom duct with the exception of O3 members, which were higher expressed in the medium and proximal regions (fig. 6*B*). We tested, within a Bayesian framework, whether any of the conotoxin

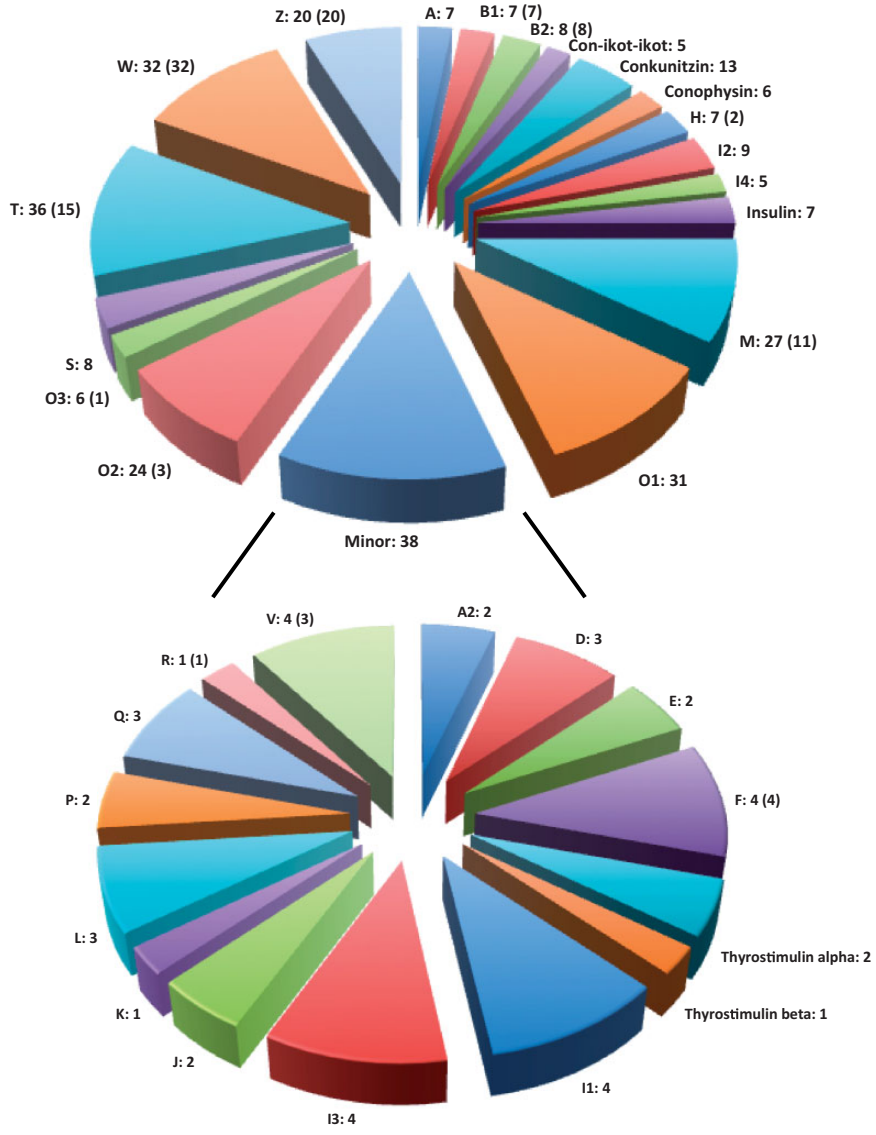


Fig. 3.—Distribution in superfamilies of the 296 identified conotoxin precursors. Numbers in parentheses indicate the number of mature peptides without cysteine framework.

precursor superfamilies had differential expression among regions of the venom duct using the individuals as biological replicates (table 3). Only the A and I3 superfamilies showed significant posterior probabilities. Within the A superfamily, paralogs A-1 (named alpha 4/4; see below) and A-2 (name kappa; see below) had PPDE values of 0.94 and 1, respectively (table 3). Within each paralog, only conotoxin precursors Cerm_405 (A-1) and Cerm_342 (A-2) both detected in ERM3 showed significant differential expression in the proximal and distal portions, respectively (table 3).

Altogether, conotoxin precursor transcripts showed differential expression along the venom duct (table 3), and accounted for ~60% and 70% of the overall expression in the medium and proximal regions, respectively (fig. 6*C*). In contrast, conotoxin precursor expression in the distal region was restricted to 30% whereas other (house-keeping) transcripts dominated (fig. 6*C*). Ferritin showed important levels of expression, being preferentially expressed in the distal region (fig. 6*C*).

Member Diversity of the a Superfamily across Cone Species

Given the importance of the A superfamily in the overall expression of the venom duct of *C. ermineus*, we performed a more detailed analysis of the diversity of its members across species (table 4; see also classification of the 288 A superfamily conotoxin precursors available in ConoServer in supplementary file 3, Supplementary Material online). Within the A superfamily (Santos et al. 2004), there are two main groups of conotoxins with very distinct structure and function (Azam

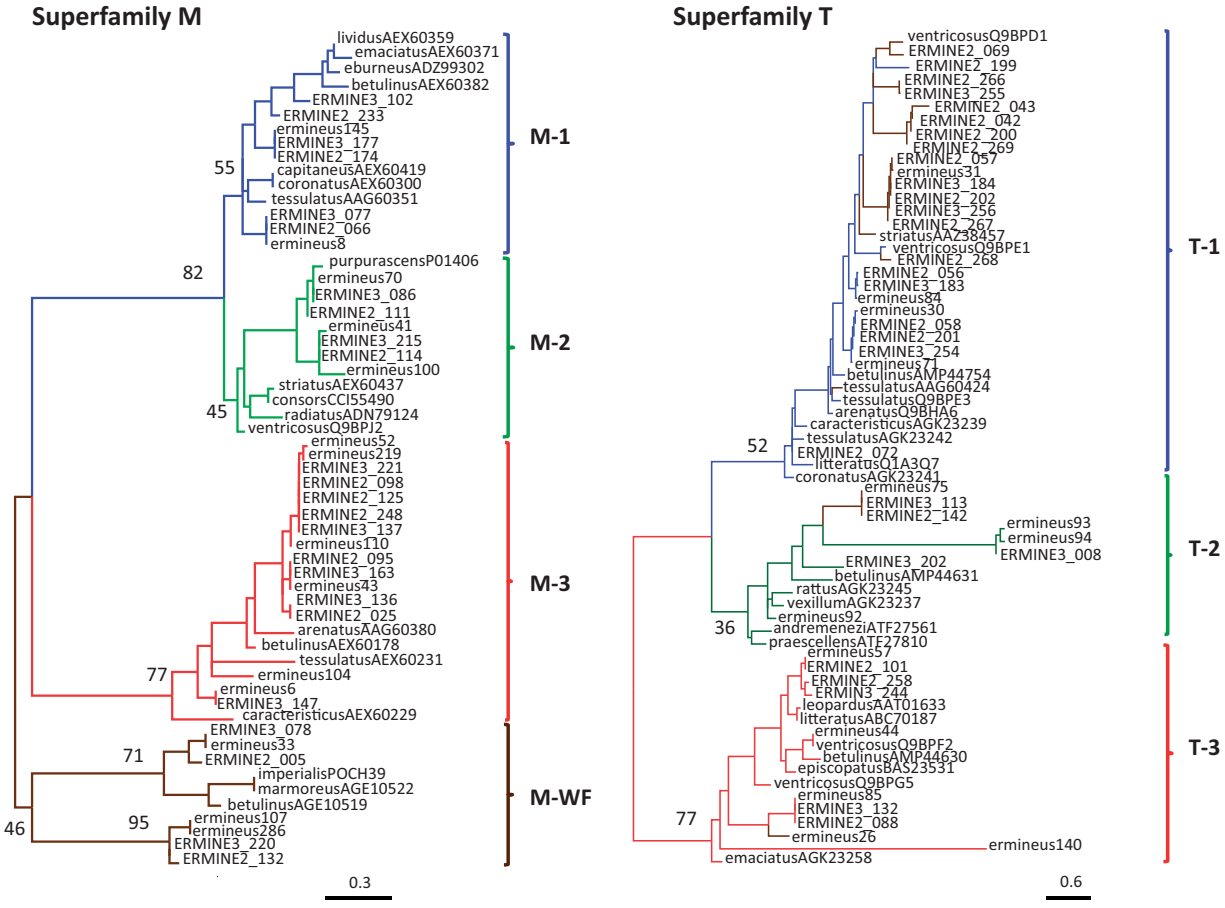


Fig. 4.—Reconstructed ML phylogenies of the M and T superfamilies, recovering several clades (paralogs; in different colors) and indicating the differential evolutionary origin of cysteine-poor (in brown; WF meaning without framework) mature peptides. Bootstrap values of main clades are indicated. Scale bar indicates substitutions/site. GenBank accession numbers are indicated after each species except for *C. ermineus*.

and McIntosh 2009; Puillandre et al. 2012; Robinson and Norton 2014). One group (alpha, α) has cysteine framework I and selectively target nicotinic-acetylcholine receptors (nAChRs), ultimately inhibiting neuromuscular transmission (Azam and McIntosh 2009). The other group (Kappa, κ) has cysteine framework IV and their members target preferentially K^+ channels, producing an excitatory effect (Robinson and Norton 2014). The Kappa subfamily is, thus far, only found in piscivorous cones (table 4; Santos et al. 2004). Within framework IV, the most frequent cysteine spacing pattern is cc7c2c1c3c (Puillandre et al. 2012), which is shared by *C. ermineus*, several *Pionoconus* species, and *Embrikena* (table 4). In contrast, specific cysteine spacing variations have been reported for *C. purpurascens*, *Gastridium*, and *Textilia* (table 4). Within the alpha subfamily, the most frequent cysteine spacing pattern is cc4c7c (named α4/7; Puillandre et al. 2012). Vermivorous (including *Rhombiconus imperialis*, which has a strict diet on amphinomids) and molluscivorous species show important diversity of α4/7 conotoxins whereas this subfamily is represented by only 1–2 members in most piscivorous species but *G. geographus* (table 4; a striking exception is *Asprella*, which has been proposed to be piscivorous, although not based on direct evidence, and has seven α4/7 conotoxin precursors, a pattern typical of vermivorous or molluscivorous cones). A second frequent cysteine spacing pattern is cc3c5c (named α3/5; Puillandre et al. 2012), which is almost exclusive of piscivorous species, and particularly diverse in *G. geographus*, *P. consors*, and *P. striatus* but not found in *Chelyconus* (table 4). Finally, a third cysteine spacing pattern, which is also relatively frequent, is cc4c4c (named α4/4; Puillandre et al. 2012). It is particularly diverse in the piscivorous genera *Chelyconus* and *Textilia*, and in the vermivorous genera *Virgiconus emaciatus* and *Calamiconus quercinus* (table 4). Interestingly, it has been also reported with lower diversity in the piscivorous genus *Pionoconus* and in the putative piscivorous genus *Afonsoconus* (Puillandre et al. 2012).

Discussion

The cocktail of bioactive peptides produced in the venom duct of a cone snail is a complex mixture aimed at paralyzing specific preys and deterring predators (Dutertre et al. 2014).

Table 2

New Signal Sequences in Conotoxin Precursors of *Chelyconus ermineus*

Unassigned Superfamily	Signal	Cysteine Framework	Also Found In:	Best-Hit Known Superfamily		
				% Coverage	% Identity	Superfamily
1	MRFYMLLAVALLLTSVMS	VI/VII	–	66	75	O2 (Q9NDA7)
2	MRFLLFLCIAVLLTSFRETEA	VI/VII	*betulinus*	85	35	T (BAS25421)
3	MKLSMMFILSLVLTLSMTDG	XIV	*praecellens*	90	67	L (ABC74975)
4	MKLSVMVIVLVLAMAFTPGLL	XIV	*betulinus*	80	67	L (ABC74975)
5	MNFSVMFILALVLTLSMTDA	XIV	*betulinus, praecellens*	90	61	L (ABC74975)
6	MKVVVVLLAVLVAASA	XIV	*betulinus*	100	56	Hyaluronidase (C0HKM3)
7	MCLSTMPSVILMMVLMFAFDNVDG	IX	*betulinus, imperialis*	58	57	P (ATF27727)
8	MKLFMFTAIIFTMASTTVT	VIII	*andremenezi, caracteristicus*	78	53	O1 (Q5K0B8)
9	MSKTGLVLVVLYLLSSPVNL	XIII	*miles, praecellens, andremenezi*	85	60	M (ACV87169)
10	MKFTTFVMVLMAAVLLTSILETEA	VI/VII	*betulinus, praecellens*	54	62	Con-ikot-ikot (BAO65537)
11	MEFRRLVTVGLLLTLVMSTDS	IX	*betulinus*	47	88	Insulin (AOF40168)
13	MLSMLAWTLMTAMVVMNAKS	$(C)_{12}$	*praecellens, gloriamaris*	55	73	O1 (BAS22670)
14	MNMRMTIIVFVVVATAATVVGST	CC-C-C-C-C-C-CC-C-C-C-C	*lenavati, tribblei*	100	61	Con-ikot-ikot (P0CB20)
15	MSVVYCKPSVPVDSVSSNFCVRGPDNGHQA	VI/VII	–	40	86	T (Q9BPD9)

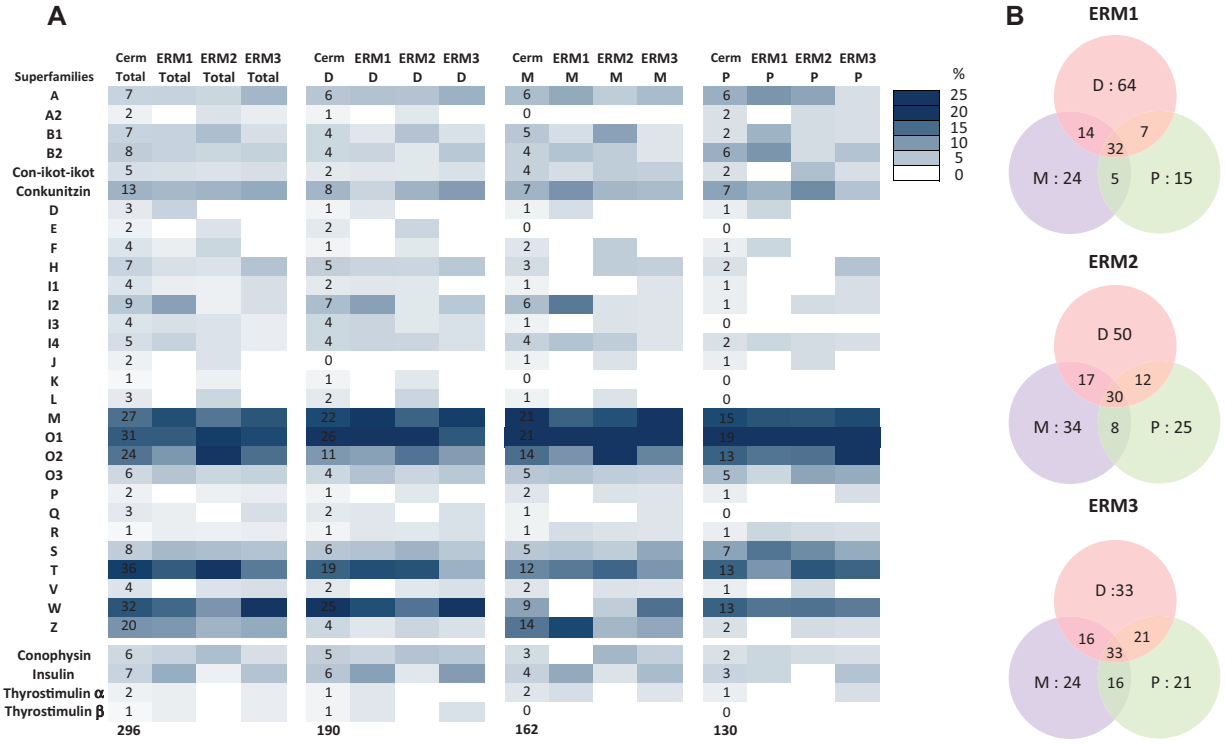
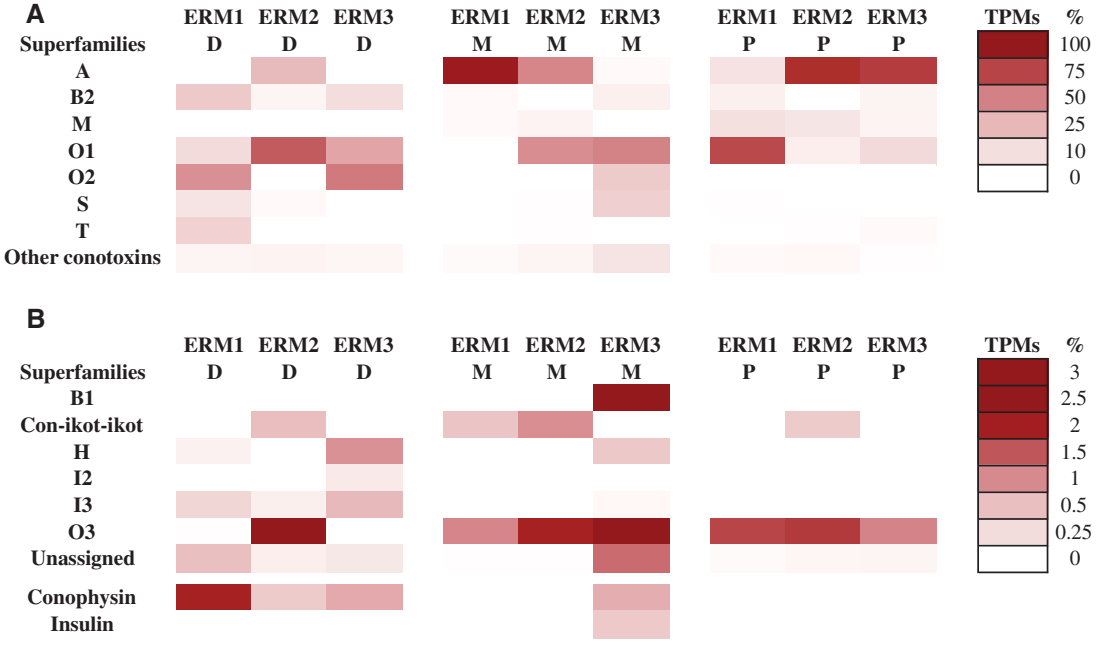


FIG. 5.—Distribution of conotoxin precursor superfamily diversity (number of members) along the distal (D), medium (M) and proximal (P) portion of the venom duct (with respect to the venom bulb) is shown in panel (*A*). Number of common members in the three analyzed individuals per venom duct region is depicted in panel (*B*). Cerm code indicates unique peptide sequences after considering the three analyzed individuals (see supplementary table 1, Supplementary Material online).



C

	ERM1	ERM2	ERM3	ERM1	ERM2	ERM3	ERM1	ERM2	ERM3
	D	D	D	M	M	M	P	P	P
Others	78.356	55.851	69.227	28.144	37.243	51.307	26.676	30.860	29.893
Conodipine	0.028	0.003	0.074	0.344	0.023	0.001	0.000	0.048	0.012
Conohyaluronidase	0.001	0.291	0.000	0.207	0.430	0.523	0.688	0.265	0.000
Ferritin	3.112	0.848	1.774	0.325	0.114	1.115	0.318	0.133	0.497
PDI	0.200	0.082	0.167	0.081	0.125	0.119	0.149	0.083	0.120
Conotoxins	18.3023	42.923	28.758	70.898	62.064	46.934	72.168	68.611	69.479

Fig. 6.—Distribution of conotoxin precursor superfamily transcript relative expression (TPMs) along the distal (D), medium (M) and proximal (P) portion of the venom duct (with respect to the venom bulb). Highly-medium expressed transcripts are shown in panel (*A*). Low expressed components are depicted in panel (*B*). Overall conotoxin expression with respect to other protein expression in the venom duct is shown in panel (*C*).

Table 3

Differential Expression of Conotoxins in the Different Regions of the Venom Gland of *Chelyconus ermineus*

	Region	PPDE
Conotoxins	PM	1
Superfamilies		
A	PM	0.96
I3	D	0.99
Superfamily A		
A-1 (alpha 4/4)	PM	0.94
A-2 (Kappa)	P	1
Superfamily A-1		
Cerm_138	M	0.71
Cerm_255	P	0.70
Cerm_405	P	1
Superfamily A-2		
Cerm_008	PD	0.89
Cerm_145	PM	0.82
Cerm_268	P	0.76
Cerm_342	D	1

The composition of this cocktail greatly varies among species and currently, we are just starting to catalogue the repertoire of conotoxins and associated proteins produced by the more than 800 species of cones (e.g., Peng et al. 2016; Phuong et al. 2016; Li et al. 2017; Robinson, Li, Lu, et al. 2017). The recent advent of next generation RNA sequencing allows a robust approach to cataloguing all the transcripts expressed in the venom duct of a cone snail: not only it is possible to determine the bulk of mRNAs that are synthesized but also the variability of expression at distinct portions of the venom duct (Dutertre et al. 2014), among different individuals (Li et al. 2017), and under variable external conditions (Dutertre et al. 2010). Beyond identifying and describing all the key components of the venom, the ultimate, more ambitious goal of cataloguing studies is to tackle long-standing evolutionary and ecological questions, for example, how the diversity of conotoxins evolved (Duda and Palumbi 1999), how the venom mixtures were adapted to highly distinct diets (Duda et al. 2001) and predation strategies (Olivera et al. 2015), how a cone individual is able to modulate the composition of the

Table 4

Superfamily A Diversity in Several Cone Snails with Different Diets Based on Conoserver Entries

		Framework IV				Framework I									Framework IX	
Genus	**species**	**cc7c2c1c3c**	**cc7c2c1c6c**	**cc6c2c1c3c**	**cc6c2c1c4c**	**cc4c7c**	**cc4c3c**	**cc4c4c**	**cc4c5c**	**cc4c6c**	**cc4c8c**	**cc3c5c**	**cc3c6c**	**cc5c10c**	**c2c4c6c**	**c10c1c3c**
Chelyconus	*ermineus*	4 (3)				2 (2)		5 (3)								
Chelyconus	*purpurascens*		1 (1)			3 (2)		(4)[a]								
Gastridium	*geographus*			3 (2)		5 (5)					1 (1)	6 (3)		1 (1)		
Pionoconus	*achatinus*	4 (4)		1 (1)								5 (2)	1 (1)			
Pionoconus	*catus*	3 (3)										1 (1)				
Pionoconus	*consors*	5 (5)				1 (1)						7 (4)	2 (1)			
Pionoconus	*magus*	3 (3)				1 (1)						1 (1)				
Pionoconus	*monachus*											1 (1)	1 (1)			
Pionoconus	*striatus*	4 (4)						1 (1)				8 (6)				
Textilia	*bullatus*				3 (3)	1 (1)		5 (4)								
Embrikena	*pergrandis*	(2)				1 (1)										
Afonsoconus	*kinoshitai*							2 (2)								
Asprella	*sulcata*					7 (5)										
Conus	*marmoreus*					7 (5)										
Cylinder	*textile*					7 (5)										
Rhombiconus	*imperialis*					1 (1)	4 (4)									
Calamiconus	*quercinus*					12 (11)		3 (2)								
Dendroconus	*betulinus*					8 (8)						1 (1)				
Fraterconus	*distans*					2 (2)				1 (1)						
Lithoconus	*leopardus*					11 (9)				3 (2)						
Lividoconus	*lividus*					6 (6)	3 (2)									
Puncticulis	*pulicarius*					5 (4)			1 (1)	1 (1)					2 (2)	1 (1)
Virgiconus	*emaciatus*					1 (1)		4 (1)	1 (1)	2 (2)						
Virgiconus	*flavidus*					5 (4)										
Vituliconus	*vitulinus*					1 (1)					2 (2)					

NOTE.—Number of precursors and mature proteins (in parentheses) are shown.

Red, piscivorous; green, molluscivorous; brown, vermivorous (fire worms); blue, vermivorous.

[a]Puillandre et al. (2012), Hoggard et al. (2017).

venom both to prey and to deter predators (Dutertre et al. 2014), or what is the role of conotoxin diversification in speciation (Li et al. 2017).

According to reconstructed phylogenies of the family Conidae, the piscivorous diet evolved independently in some Atlantic and Indo-Pacific cone genera (Duda et al. 2001; Duda and Palumbi 2004; Puillandre, Bouchet, et al. 2014; this work). The transcriptomes of several piscivorous genera (*Gastridium, Pionoconus*, and *Textilia*) from the Indo-Pacific have been reported but none from Atlantic piscivorous genera was available yet. Hence, the importance of sequencing the transcriptome of *C. ermineus*: its comparison with those of piscivorous cone snails from the Indo-Pacific, and against those of cones eating snails and worms could provide important clues on which conotoxins are needed specifically for fish hunting and how they evolved (Duda and Palumbi 2004).

The obtained number of assembled contigs per *C. ermineus* venom duct transcriptome is comparable to those typically reported in equivalent studies also based on the Illumina platform (Peng et al. 2016; Li et al. 2017; Robinson, Li, Lu, et al. 2017). Moreover, the total number of clean reads, which mapped onto conotoxin transcripts constituted 57%. This number is somewhat higher than those reported by studies based on direct sequencing of individual cDNA clones such as, for example, the 39–50% of transcripts being conotoxins reported for *Virgiconus virgo*, *Tesselliconus eburneus*, *R. imperialis*, and *C. marmoreus* (Liu et al. 2012) or based on 454 sequencing, for example, the 42.7% of transcripts being conotoxins reported for *P. consors* (Terrat et al. 2012) but lower to the 88% reported for *G. geographus* (Hu et al. 2012). A total of 378 transcripts encoding conotoxin precursors were identified in the venom duct of *C. ermineus*. This number is similar to those reported for the venom duct transcriptomes of *Virroconus coronatus* (331; Phuong et al. 2016), *Puncticulis arenatus* (326; Phuong et al. 2016), or *Harmoniconus sponsalis* (401; Phuong et al. 2016) and larger than others such as those reported in *D. betulinus* (215; Peng et al. 2016), *C. marmoreus* (158; Lavergne et al. 2013), *T. praecellens* (149–155; Li et al. 2017), *T. andremenezi* (107–128; Li et al. 2017), or *C. gloriamaris* (108; Robinson, Li, Lu, et al. 2017). Until recently, most studies were based on single individuals whereas the current trend is to sequence several specimens as here (Barghi et al. 2015a; Peng et al. 2016; Li et al. 2017), which affects the comparison of numbers. In fact, the *C. ermineus* individuals produced each 145–176 conotoxin precursors, which is consistent with most reported single individual transcriptomes (see above). In *C. ermineus*, ~20% of the inferred mature conotoxins were common to the three analyzed individuals and strikingly, each of them showed an important number of sequences not found in the other two. Therefore, it is very likely that even the number here reported underestimates the whole diversity of conotoxins produced by this species. Such interindividual differences are congruent with results reported in the closely related *C. purpurascens* (Rodriguez et al. 2015) and for three specimens of *D. betulinus* with different body sizes (Peng et al. 2016). In this regard, it has been suggested that differences in age/size could be a factor fostering intraspecific variation (Barghi et al. 2015a; Peng et al. 2016). The three specimens of *C. ermineus* here analyzed were presumably adults and they differed in their size: ERM1-3 had shells of 73.1, 55 and 46 mm in length, respectively. In contrast, other studies found little intraspecific variation within species of the genus *Turriconus* (Li et al. 2017) and of the genus *Kioconus* (Barghi et al. 2015a). Here, it is important to note that despite our specimens were from different islands (Boa Vista, Sal, and Santa Luzia), they have not accumulated larger *cox1* sequence divergences than individuals within species of *Turriconus* or *Kioconus* (see additional supplementary fig. 2, Supplementary Material online). Altogether, our results suggest that, in the near future, many more individuals than those currently analyzed would be required to describe the whole richness of the venom of any cone species (Dutertre et al. 2010).

The Illumina-based venom repertoires of *C. ermineus* and of other recently investigated cone species with various diets were compared in table 5. Most conotoxin precursor superfamilies reported in other cone species (Peng et al. 2016; Phuong et al. 2016; Li et al. 2017; Robinson, Li, Lu, et al. 2017) were also identified in *C. ermineus*. The most diverse (in number of members) superfamilies in *C. ermineus* were O1, O2, M, and T, a pattern that is conserved in other cone species (table 5). Interestingly, the W and Z superfamilies, which have a cysteine-poor mature peptide and were originally reported in *C. marmoreus* (Lavergne et al. 2013), were particularly abundant in *C. ermineus*. Different cone species seem to have undergone diversification bursts of particular superfamilies, which are otherwise poorly represented in other species (Duda and Remigio 2008; Puillandre et al. 2012; Barghi et al. 2015a). A paradigmatic case is the A superfamily, which is highly diverse in the Indo-Pacific piscivorous *G. geographus* (Safavi-Hemami et al. 2014), *P. consors* (Terrat et al. 2012), *Pionoconus catus* (Himaya et al. 2015), and *T. bullatus* (Hu et al. 2011) but underrepresented in *C. ermineus*. Other examples are: the superfamilies P and O1d, which are particularly rich in *Turriconus* (Li et al. 2017), the B1 superfamily in *G. geographus*, the I2 superfamily in *V. virgo*, the A, I3, and N superfamilies in *Rolaniconus varius*, the D superfamily in *Rhizoconus vexillum* (Prashanth et al. 2016), and the con-ikot-ikot and B2 superfamilies in *K. tribblei* and *K. lenavati* (Barghi et al. 2015a).

About 20% of the newly identified conotoxin precursors could not be assigned to known superfamilies based on their signal domain, and new unassigned conotoxin precursor superfamilies had to be proposed temporarily as in other studies (e.g., Barghi et al. 2015b; Peng et al. 2016). However, most of these unassigned superfamilies had homologs in other cone species, and thus, it is foreseen that as more cone transcriptomes become available, formal (phylogeny-

Table 5

Comparison of the Diversity of Main Conotoxin Precursor Superfamilies in Cone Snails with Different Diets

Species	*Chelyconus ermineus*	*Gastridium geographus*	*Pionoconus consors*	*Cylinder gloriamaris*	*Conus marmoreus*	*Conus marmoreus*	*Dendroconus betulinus*	*Turriconus andremenezi*	*Turriconus praecellens*	*Rolaniconus varius*	*Virgiconus virgo*	*Rhombiconus imperialis*
												
Reference	**This work**	Hu et al. 2012	Terrat et al. 2012	Robinson et al. 2017	Phuong et al. 2016	Lavergne et al. 2013	Peng et al. 2016	Li et al. 2017	Li et al. 2017	Phuong et al. 2016	Phuong et al. 2016	Phuong et al. 2016
NGS platform	**HiSeq2000**	**FLX Titanium**	**454**	**HiSeq2000**	**HiSeq2000**	**454**	**HiSeq2000**	**HiSeq2000**	**HiSeq2000**	**HiSeq2000**	**HiSeq2000**	**HiSeq2000**
Diet	**fish**	**fish**	**fish**	**snail**	**snail**	**snail**	**worm**	**worm**	**worm**	**worm**	**worm**	**fire worms**
A	8	12	14	3	1	0	9	0	0	16	5	3
A2	2	–[a]	–	–	–	–	–	4	2	–	–	–
B1 (conantokin)	7	2	1	1	0	0	8	0	0	2	1	0
B2	8	0	0	1	1	0	1	2	3	2	0	1
B4	–	–	–	–	0	–	–	2	0	8	0	0
C	0	0	0	0	0	0	1	0	0	0	0	0
Con-ikot-ikot	5	6	0	2	0	0	4	2	5	8	1	0
Conkunitzin	13	1	7	0	1	0	8	0	0	2	1	0
D	3	0	0	0	0	0	0	0	2	0	0	1
E	2	0	0	1	4	0	2	0	0	2	1	1
F	4	0	0	1	4	0	3	0	0	2	1	0
H	7	0	0	3	3	0	3	1	3	0	0	0
I1	4	1	0	2	2	0	5	3	2	2	0	2
I2	9	0	0	3	0	0	8	4	5	3	14	2
I3	4	0	0	0	0	0	2	6	5	16	0	0
I4	5	–	–	4	4	2	0	4	6	0	0	0
J	2	4	0	4	0	0	3	0	9	1	0	0
K	1	0	0	0	0	0	0	0	0	2	0	5
L	3	0	0	0	0	0	0	0	3	3	0	0
M	27	2	8	7	20	55	30	28	33	23	7	8
N	0	0	0	1	2	0	6	0	1	11	6	1
O1	31	18	15	12	13	61	33	22	36	22	26	9
O1d	–	–	–	–	–	–	–	5	12	–	–	–
O2 (contryphan)	24	1	0	18	3	4	16	17	17	5	13	3
O3	6	0	0	1	0	0	4	0	0	2	1	0
P	2	0	1	3	0	0	13	26	32	14	0	12
Q	3	–	–	0	0	1	0	0	0	0	2	0
R	1	–	–	0	–	1	0	0	0	–	–	–
S	8	5	3	1	1	0	1	2	3	5	0	1
T	36	6	5	21	20	7	17	41	39	19	13	8
U	0	–	–	1	0	1	0	0	2	0	1	0

V	4	0	0	0	0	0	3	3	0	3	0	0	0	0	0	0	0	0	0	0
W	32	–	–	–	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
X					2															
Y	0	0	0	0	0	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0
Y3					1															
Z	20	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
conoCAP	0	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Conomap	0	–	–	0	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Conophysin (Conopressin)	6	1	1	1	0	0	1	2	7	0	1	0	0	0	0	0	0	0	0	1
Conorfamide	0	–	–	1	–	0	2	3	–	–	–	–	–	–	–	–	–	–	–	–
Contulakin	0	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Insulin	7	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Prohormone-4	0	–	–	3	–	0	2	1	–	–	–	–	–	–	–	–	–	–	–	–
Thyrostimulin α	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Thyrostimulin β	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

[a]Hyphen indicates that the study apparently did not search for the corresponding peptide.

based) classification of these unassigned superfamilies will be possible (Puillandre et al. 2012; Lavergne et al. 2013). Remarkably, we found a combination of mature peptides normally associated to the signal and propeptide domains of the O3 and T superfamilies, which in *C. ermineus* were associated to amino terminal sequences lacking signal and/or propeptide domains (unassigned families 16–22). This observation may evoke the possibility of domain shuffling as one of the underlying mechanisms for generating precursor diversity (Pi, Liu, Peng, Liu, et al. 2006). Despite the coverage is high and homogeneous throughout these assembled transcripts (including the boundary of the mature peptide and the rest of the putative precursor), the possibility of an assembly artifact cannot be fully excluded given that most of these sequences do not show a canonical three-domain structure (and should be validated experimentally).

One hot debate in the past has been the relevance of classifying conotoxins into cysteine-rich and cysteine-poor categories (Puillandre et al. 2012; Robinson, Li, Bandyopadhyay, et al. 2017). Our results support that this dichotomy is irrelevant from an evolutionary perspective, although it may have some functional meaning. According to our phylogenetic and sequence comparison analyses, there are at least three different evolutionary origins for cysteine-poor mature peptides: 1) whole conotoxin precursor superfamilies carrying mature peptides with no cysteine framework, such as, for example, R, W, and Z; 2) conotoxin precursor superfamilies in which cysteine-poor mature peptides are associated to a distinct signal and propeptide combination, and thus have a single evolutionary origin, such as, for example, M; and 3) conotoxin precursor superfamilies in which cysteine-poor mature peptides are associated to signal and propeptide combinations that also can be linked to mature peptides with a known cysteine framework, indicating multiple evolutionary origins, such as, for example, T. This third evolutionary pattern supports the importance of modularity as evolutionary mechanism for generating conotoxin diversity (Pi, Liu, Peng, Liu, et al. 2006).

Importantly, the use of the amino acid sequences of the propeptide domain in addition to those of the signal domain for phylogenetic analyses and sequence similarity comparisons proved to be very informative (Lavergne et al. 2013) and led to discrimination of potential distinct paralogs within the different conotoxin precursor superfamilies. For instance, at least four paralogs were identified within the M superfamily (named M1-3, M-WF), three within the T superfamily (T1-3; see Liu et al. 2012), which detected up to four clades in their phylogenetic analysis), six within the O1 superfamily (O1-1-6; see Li et al. 2017), which already distinguish O1d from the remaining members of O1) and six within the O2 superfamily (O2-1-6). This, thus far, mostly overlooked diversity within each superfamily will need to be taken into account in future studies when updating classifications of conotoxin precursor superfamilies (Lavergne et al. 2013) and when summarizing the composition of venoms in the different species as the

different paralogs may well have different functions (Altenhoff et al. 2012).

The current classification system of conotoxin superfamilies originally based on the alphabet can barely integrate the many novel signal sequences (and corresponding superfamilies) found in every new study and would have even more serious problems in dealing with paralog diversity within each superfamily, if this information is also incorporated. Therefore, a radically new evolutionary-based classification (beyond the goals of the present study) is urgently needed, and should come out from the consensus among experts in the field.

The distribution of conotoxin precursor diversity along the venom duct of *C. ermineus* showed some degree of regionalization. The A2 and B2 superfamilies had more diversity toward the proximal portion; Z was more diverse in the medium portion; and O1, M, and T superfamilies had more diversity toward the distal portion. The corresponding pattern in *G. geographus* showed the A and M superfamilies to be more abundant in the proximal portion, and O1 and T more diverse in the distal portion (Hu et al. 2012). The distribution pattern found in the molluscivorous *Cylinder textile*, showed the M and T superfamilies as more abundant in the proximal region whereas the O superfamily was more diverse in the distal region (Garrett et al. 2005). Overall, these comparisons show two discriminant patterns: 1) the T superfamily is more diverse in the distal portion of piscivorous cones and in the proximal region of the molluscivorous cone; and 2) the M superfamily is more diverse in the distal region of *C. ermineus* and in the proximal portion of *G. geographus*.

The venom duct of a cone snail is a specialized convoluted duct mostly devoted to the biosynthesis of conotoxins (Safavi-Hemami et al. 2014), as further demonstrated here by the elevated proportion of conotoxin transcripts detected in the transcriptome of the *C. ermineus* venom duct. Moreover, our results support that conotoxin expression is localized preferentially in the medium and proximal regions of the venom duct of *C. ermineus* whereas the distal region is mostly devoted to the expression of house-keeping genes. The inferred expression patterns of conotoxin precursor superfamilies in *C. ermineus* showed drastic variations among individuals. This may reflect natural conditions or potential methodological biases (despite the use of common sample handling, laboratory, sequencing, and analytical procedures), although we cannot discern between both possibilities. In any case, these results highlight the need of sequencing a fair amount of specimens to generate statistically robust quantitative comparisons and conclusions, as it is becoming the rule for model system species (Schurch et al. 2016). Being cautious in the interpretation of the results of our expression analyses (i.e., considering reliable only those TPM values, which are similar in at least two individuals), we observed that those conotoxin precursor superfamilies showing higher levels of expression are also those having more member diversity (O1, O2, M, and T). A striking exception to this pattern is the A superfamily, which has few distinct members in *C. ermineus*, but the highest levels of expression (see below). Most superfamilies showed low expression levels, suggesting a subtle contribution of them to the final venom composition. The expression of most superfamilies appears to reflect some degree of compartmentalization. In particular, the A superfamily is preferentially expressed in the medium and proximal regions of the venom duct whereas most other superfamilies tend to be expressed toward the distal region (as occurs in Indo-Pacific cones such as *G. geographus*; Hu et al. 2012). Moreover, we found that this compartmentalization of the expression of the A superfamily is n statistically significant, further indicating its functional importance in the venom of *C. ermineus*.

The different strategies of prey capture among piscivorous cone species determine the exact mixture (termed "cabal") of venom components, which will act coordinately to produce a specific physiological response (Olivera et al. 2016). While the Indo-Pacific species *G. geographus* has a "net engulfment" strategy, the Indo-Pacific species of the genus *Pionoconus* and *Textilia* as well as the Atlantic and Eastern Pacific species of the genus *Chelyconus* have a "taser and tether" or "hook and line" strategy (Olivera et al. 2015, 2016). The "net engulfment" strategy requires disorienting the fish by the release of the "nirvana cabal" into the water. This is a mixture, among others, of B1 superfamily (Hu et al. 2012) and insulin-like peptides (Safavi-Hemami et al. 2015; Robinson, Li, Bandyopadhyay, et al. 2017). Once fishes are disoriented and engulfed, the cone injects into each captured fish, a group of paralytic conotoxins, the "motor cabal," which includes the αA conotoxins, and the M superfamily μ- and ψ-conotoxins. In the case of the "taser and tether" strategy, the capture of the prey occurs through direct injection of two different mixtures, the "lightning-strike" and the motor cabals (Olivera 2002). The lightning-strike cabal induces an excitatory response, which ultimately causes tetanic paralysis. This cabal includes δ-conotoxins and κ-conotoxins from the O1 superfamily, conkunitzins, and κA conotoxins (Himaya et al. 2018).

Therefore, while the nirvana cabal is exclusive of some of species of the genus *Gastridium*, the lightning-strike cabal is found in genera such as *Pionoconus*, *Textilia*, and *Chelyconus*, and the motor cabal is found in all four above-mentioned genera (Olivera et al. 2016). The κA conotoxins of the lightning-strike cabal have generally the same cysteine spacing pattern in *Pionoconus*, *Textilia*, and *Chelyconus* (except *C. purpurascens*). However, the main blockers of K^+ channels belong to the O superfamily in *Chelyconus*, whereas this physiological role is accomplished by conkunitzins in *Pionoconus* (Olivera et al. 2016). With regards to the motor cabal, there are striking instances of differential recruitment of the αA conotoxins. All piscivorous genera have generally at least one member of the α4/7 subfamily, which blocks neuronal nAChRs (Azam and McIntosh 2009). In addition, *Gastridium*

and *Pionoconus* have members of the α3/5 subfamily, which inhibit muscle nAChRs. However, *Chelyconus* and *Textilia* have instead members of the α4/4 subfamily. The α4/4 conotoxins of *C. ermineus* and *C. purpurascens* have been shown to selectively bind muscle nAChRs (López-Vera et al. 2007; Quinton et al. 2013) whereas those of *Textilia* block neuronal nAChRs (Chi et al. 2006). The presence of α4/4 conotoxins has been detected also in *Afonsoconus*, and three species of *Pionoconus*, but their functions have not been determined (Puillandre et al. 2012). The α4/4 sequences of *Pionoconus* and *Textilia* (and probably *Afonsoconus*) are more closely related phylogenetically than those of *Chelyconus* (Puillandre et al. 2012). Altogether, our results suggest independent genetic and biochemical pathways to evolve the same diet adaptation, and thus, favor the hypothesis of a convergent origin of piscivory in cones from the Indo-Pacific and Atlantic oceans (Duda et al. 2001; Puillandre, Bouchet, et al. 2014).

Finally, it is interesting to note that recent analysis of conotoxin envenomation in *C. purpurascens* showed that different individuals could include alternatively either the lightning-strike, the motor or both cabals in the composition of their venoms when preying (Himaya et al. 2018). In our case (see supplementary table 1, Supplementary Material online), the three individuals of *C. ermineus* produced conotoxins belonging to the motor cabal including α4/4 conotoxins (the α4/7 conotoxin was not detected), and M superfamily μ- and ψ-conotoxins as well as had conotoxins belonging to the lightning-strike cabal including the κA conotoxins, and the O1 superfamily δ- (Aman et al. 2015) and κ-conotoxins. Moreover, the differential high levels of A superfamily transcripts in the proximal region of the venom duct in *G. geographus* as part of the motor cabal have been associated to defense-evoked responses (Dutertre et al. 2014). In *C. ermineus* (see supplementary table 1, Supplementary Material online), for the A superfamily, the members involved in the motor cabal (α4/4 conotoxins) and the lightning-strike cabal (κA conotoxins) showing differential expression are located in the proximal and distal regions of the venom duct, respectively, supporting the regionalization of the cabals as reported in *Protostrioconus obscurus* (Dutertre et al. 2014). Nevertheless, all the above-mentioned inferences need to be interpreted with caution and as tentative until further comparative analyses based on more individuals are carried out.

Conclusions

The venom duct of *C. ermineus* produces a great diversity of conotoxin precursors, most corresponding to known superfamilies and several showing novel signal domains. Comparison of these data to the venom repertoires reported from different cone species with various diets supports that some superfamilies (O1, O2, T, M) are widespread among cone species, making the basic venom toolkit, whereas others are restricted to fewer lineages. The different superfamilies show various degrees of expansion depending on the species. In the case of *C. ermineus*, the cysteine-poor superfamilies W and Z are particularly diverse. In this regard, the wide distribution of cysteine-poor mature peptides among superfamilies indicate multiple and diverse origins. Both, diversity and expression of conotoxins are regionalized along the venom duct. Diversity in the number of members of a superfamily increases toward the distal region whereas the less diverse superfamilies in the proximal region show higher expression levels. In particular, the A superfamily, which is highly diverse in piscivorous cones from the Indo-Pacific Ocean, consists of rather few and distinct (α4/4) members in the *C. ermineus* venom, but these show differentially and significantly high expression levels toward the proximal region. These contrasting patterns support convergent strategies to produce the motor cabal, which targets nicotinic acetylcholine receptors, and seems essential for deterring/preying fishes.

Our results show that each newly analyzed cone species uncovers additional conotoxin diversity and thus, that we are still far from covering the whole repertoire of conotoxins, as the venom duct transcriptome of the majority of cone species awaits sequencing and analysis. Moreover, the numerous unassigned superfamilies, which are discovered in every new cone transcriptome together with the emerging evidence of the existence of distinct paralogs within each superfamily prompt for a revision and an update of the nomenclature of conotoxins as the use of the alphabet-based classification seems to be too constrained and obsolete in evolutionary terms.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank four anonymous reviewers, Sébastien Dutertre, and Christoph Bleidorn for insightful comments on a previous version of the manuscript. We are indebted to Cabo Verde biology students Paulo Vasconcelos and Stiven Delgado for their valuable help during sampling in Santa Luzia (Cabo Verde). We thank Dr Rui Freitas from the Universidade de Cabo Verde (UniCV) for his continuous support of our research in Cabo Verde. We also thank Dr Iderlindo Silva dos Santos and Dra. Sonia Monteiro de Pina Araujo from the Direccçao Nacional do Medio Ambiente of the Ministério do Ambiente, Habitação e Ordenamento do Território (MAHOT) of the Republic of Cabo Verde for their help with collecting permits (Autorizações 07/2013, 26/2013, 01/2104, 04/2015, and 03/2016). We are grateful to Jesús Marco and Aida Palacios, who provided access to the supercomputer Altamira at the Institute of Physics of Cantabria (IFCA-CSIC), member of the Spanish Supercomputing Network, for performing assembling and

phylogenetic analyses. We are grateful to Etienne Kornobis for help with early analyses. Thanks to Alexander Medvedev, who shares free the pictures of the different species of cones from his collection (www.coneshells-am.ru; last accessed September 05, 2018). This work was funded by the Spanish Ministry of Economy, Industry and Competitiveness (CGL2013-45211-C2-2-P and CGL2016-75255-C2-1-P [AEI/FEDER, UE]) to R.Z.; BES-2014-069575 to S.A.

Literature Cited

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 38(Web Server issue):W7–13.

Altenhoff AM, Studer RA, Robinson-Rechavi M, Dessimoz C. 2012. Resolving the ortholog conjecture: orthologs tend to be weakly, but significantly, more similar in function than paralogs. PLoS Comput Biol. 8(5):e1002514.

Aman JW, et al. 2015. Insights into the origins of fish hunting in venomous cone snails from studies of *Conus tessulatus*. Proc Natl Acad Sci U S A. 112(16):5087.

Andrews S. 2010. FastQC. available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; last accessed September 05, 2018.

Azam L, McIntosh JM. 2009. Alpha-conotoxins as pharmacological probes of nicotinic acetylcholine receptors. Acta Pharmacol Sin. 30(6):771.

Bandyopadhyay PK, et al. 1998. Conantokin-G precursor and its role in γ-carboxylation by a vitamin K-dependent carboxylase from a conus-snail. J Biol Chem. 273(10):5447–5450.

Barbier J, et al. 2004. A δ-conotoxin from *Conus ermineus* venom inhibits inactivation in vertebrate neuronal Na+ channels but not in skeletal and cardiac muscles. J Biol Chem. 279(6):4680–4685.

Barghi N, Concepcion GP, Olivera BM, Lluisma AO. 2015a. Comparison of the venom peptides and their expression in closely related *Conus* species: insights into adaptive post-speciation evolution of *Conus* exogenomes. Genome Biol Evol. 7(6):1797–1814.

Barghi N, Concepcion GP, Olivera BM, Lluisma AO. 2015b. High conopeptide diversity in *Conus tribblei* revealed through analysis of venom duct transcriptome using two high-throughput sequencing platforms. Mar Biotechnol 17(1):81–98.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2005. GenBank. Nucleic Acids Res. 33:D34–D38.

Biggs JS, et al. 2010. Evolution of *Conus* peptide toxins: analysis of *Conus californicus* Reeve, 1844. Mol Phylogenet Evol. 56(1):1–12.

Born IEv. 1778. Index rerum naturalium Musei Cæsarei Vindobonensis. Pars I. Testacea. Verzeichniss der natürlichen Seltenheiten des k. k. Naturalien Cabinets zu Wien. Erster Theil. Schalthiere. Vindobonae: Vindobonæ ex officina Krausiana.

Buczek O, Olivera BM, Bulaj G. 2004. Propeptide does not act as an intramolecular chaperone but facilitates protein disulfide isomerase-assisted folding of a conotoxin precursor. Biochemistry 43(4):1093–1101.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17(4):540–552.

Chang D, Duda TF. 2012. Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. Mol Biol Evol. 29(8):2019–2029.

Chang D, Duda TF. 2016. Age-related association of venom gene expression and diet of predatory gastropods. BMC Evol Biol. 16(1):1–12.

Chi S-W, Kim D-H, Olivera BM, McIntosh JM, Han K-H. 2006. NMR structure determination of α-conotoxin BulA, a novel neuronal nicotinic acetylcholine receptor antagonist with an unusual 4/4 disulfide scaffold. Biochem Biophys Res Commun. 349(4):1228–1234.

Conticello SG, et al. 2001. Mechanisms for evolving hypervariability: the case of conopeptides. Mol Biol Evol. 18(2):120–131.

Conticello SG, et al. 2003. The prodomain of a secreted hydrophobic miniprotein facilitates its export from the endoplasmic reticulum by hitchhiking on sorting receptors. J Biol Chem. 278(29):26311–26314.

Corpuz GP, et al. 2005. Definition of the M-Conotoxin superfamily: characterization of novel peptides from molluscivorous *Conus* venoms. Biochemistry 44(22):8176–8186.

Davis J, Jones A, Lewis RJ. 2009. Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. Peptides 30(7):1222–1227.

Duda TF, Kohn AJ, Palumbi SR. 2001. Origins of diverse feeding ecologies within *Conus*, a genus of venomous marine gastropods. Biol J Linn Soc. 73(4):391–409.

Duda TF, Palumbi SR. 1999. Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. Proc Natl Acad Sci U S A. 96(12):6820–6823.

Duda TF, Palumbi SR. 2004. Gene expression and feeding ecology: evolution of piscivory in the venomous gastropod genus *Conus*. Proc R Soc B Biol Sci. 271(1544):1165–1174.

Duda TF, Remigio EA. 2008. Variation and evolution of toxin gene expression patterns of six closely related venomous marine snails. Mol Ecol. 17(12):3018–3032.

Dutertre S, Biass D, Stöcklin R, Favreau P. 2010. Dramatic intraspecimen variations within the injected venom of *Conus consors:* an unsuspected contribution to venom diversity. Toxicon 55(8):1453–1462.

Dutertre S, et al. 2013. Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom. Mol Cell Proteomics 12(2):312–329.

Dutertre S, et al. 2014. Evolution of separate predation- and defence-evoked venoms in carnivorous cone snails. Nat Commun. 5:3521.

Echterbille J, et al. 2017. Discovery and characterization of EIIB, a new α-conotoxin from *Conus ermineus* venom by nAChRs affinity capture monitored by MALDI-TOF/TOF mass spectrometry. Toxicon 130:1–10.

Endean R, Duchemin C. 1967. The venom apparatus of *Conus magus*. Toxicon 4(4):275–284.

Espiritu DJD, et al. 2001. Venomous cone snails: molecular phylogeny and the generation of toxin diversity. Toxicon 39(12):1899–1916.

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol. 17(6):368–376.

Garrett JE, Buczek O, Watkins M, Olivera BM, Bulaj G. 2005. Biochemical and gene expression analyses of conotoxins in *Conus textile* venom ducts. Biochem Biophys Res. Communications 328(1):362–367.

Gowd KH, et al. 2008. Conantokin-P, an unusual conantokin with a long disulfide loop. Toxicon 52(2):203–213.

Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29(7):644–652.

Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59(3):307–321.

Himaya SWA, et al. 2015. Comparative venomics reveals the complex prey capture strategy of the piscivorous cone snail *Conus catus*. J Proteome Res. 14(10):4372–4381.

Himaya SWA, Marí F, Lewis RJ. 2018. Accelerated proteomic visualization of individual predatory venoms of *Conus purpurascens* reveals separately evolved predation-evoked venom cabals. Sci Rep. 8(1):330.

Hoggard MF, et al. 2017. *In vivo* and *in vitro* testing of native α-conotoxins from the injected venom of *Conus purpurascens*. Neuropharmacology 127:253–259.

Hu H, Bandyopadhyay PK, Olivera BM, Yandell M. 2011. Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. BMC Genomics 12:60.

Hu H, Bandyopadhyay PK, Olivera BM, Yandell M. 2012. Elucidation of the molecular envenomation strategy of the cone snail *Conus geographus* through transcriptome sequencing of its venom duct. BMC Genomics 13:284.
Jacobsen R, et al. 1997. Differential targeting of nicotinic acetylcholine receptors by novel αA-conotoxins. J Biol Chem. 272(36):22531–22537.
Jin A-h, et al. 2013. Transcriptomic messiness in the venom duct of *Conus miles* contributes to conotoxin diversity. Mol Cell Proteomics 12(12):3824–3833.
Jin A-H, et al. 2015. δ-Conotoxin SuVIA suggests an evolutionary link between ancestral predator defence and the origin of fish-hunting behaviour in carnivorous cone snails. Proc R Soc B Biol Sci. 282(1811):20150817.
Kaas Q, Westermann J-C, Craik DJ. 2010. Conopeptide characterization and classifications: an analysis using ConoServer. Toxicon 55(8):1491–1509.
Kaas Q, Yu R, Jin A-H, Dutertre S, Craik DJ. 2012. ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. Nucleic Acids Res. 40(D1):D325–D330.
Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.
Kornobis E, et al. 2015. TRUFA: a user-friendly web server for de novo RNA-seq analysis using cluster computing. Evol Bioinformatics 11:EBO.S23873–104.
Koua D, et al. 2012. ConoDictor: a tool for prediction of conopeptide superfamilies. Nucleic Acids Res. 40(Web Server issue):W238–W241.
Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol. 29(6):1695–1701.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357.
Lavergne V, et al. 2013. Systematic interrogation of the *Conus marmoreus* venom duct transcriptome with ConoSorter reveals 158 novel conotoxins and 13 new gene superfamilies. BMC Genomics 14:708.
Lavergne V, et al. 2015. Optimized deep-targeted proteotranscriptomic profiling reveals unexplored Conus toxin diversity and novel cysteine frameworks. Proc Natl Acad Sci U S A. 112(29):E3782–E3791.
Leng N, et al. 2013. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics 29(8):1035–1043.
Lewis RJ, Dutertre S, Vetter I, Christie MJ. 2012. *Conus* venom peptide pharmacology. Pharmacol Rev. 64(2):259–298.
Li Q, et al. 2017. Divergence of the venom exogene repertoire in two sister species of *Turriconus*. Genome Biol Evol. 9(9):2211–2225.
Liu Z, et al. 2012. Diversity and evolution of conotoxins in *Conus virgo*, *Conus eburneus*, *Conus imperialis* and *Conus marmoreus* from the South China Sea. Toxicon 60(6):982–989.
Lluisma AO, Milash BA, Moore B, Olivera BM, Bandyopadhyay PK. 2012. Novel venom peptides from the cone snail *Conus pulicarius* discovered through next-generation sequencing of its venom duct transcriptome. Marine Genomics 5:43–51.
López-Vera E, Jacobsen RB, Ellison M, Olivera BM, Teichert RW. 2007. A novel alpha conotoxin (α-PIB) isolated from C. purpurascens is selective for skeletal muscle nicotinic acetylcholine receptors. Toxicon 49(8):1193–1199.
Lu A, Yang L, Xu S, Wang C. 2014. Various conotoxin diversifications revealed by a venomic study of *Conus flavidus*. Mol Cell Proteomics 13(1):105–118.
Luo S, et al. 2006. Identification and molecular diversity of T-superfamily conotoxins from *Conus lividus* and *Conus litteratus*. Chem Biol Drug Des. 68(2):97–106.
Martinez JS, et al. 1995. alpha-conotoxin EI, a new nicotinic acetylcholine receptor antagonist with novel selectivity. Biochemistry 34(44):14519–14526.
Olivera BM. 2002. *Conus* venom peptides: reflections from the biology of clades and species. Annu Rev Ecol Syst. 33(1):25–47.
Olivera BM, Fedosov A, Imperial JS, Kantor K. 2016. Physiology of envenomation by conoidean gastropods. In: Saleuddin S, Mukai S, editors. Physiology of molluscs. New York: Apple Acdemic Press. p. 154–188.
Olivera BM, Seger J, Horvath MP, Fedosov AE. 2015. Prey-capture strategies of fish-hunting cone snails: behavior, neurobiology and evolution. Brain Behav Evol. 86(1):58–74.
Peng C, et al. 2016. High-throughput identification of novel conotoxins from the Chinese tubular cone snail (*Conus betulinus*) by multi-transcriptome sequencing. GigaScience 5:17.
Phuong MA, Mahardika GN, Alfaro ME. 2016. Dietary breadth is positively correlated with venom complexity in cone snails. BMC Genomics 17:401.
Pi C, Liu Y, Peng C, Jiang X, et al. 2006. Analysis of expressed sequence tags from the venom ducts of *Conus striatus*: focusing on the expression profile of conotoxins. Biochimie 88(2):131–140.
Pi C, Liu J, Peng C, Liu Y, et al. 2006. Diversity and evolution of conotoxins based on gene expression profiling of *Conus litteratus*. Genomics 88(6):809–819.
Prashanth JR, et al. 2016. The role of defensive ecological interactions in the evolution of conotoxins. Mol Ecol. 25(2):598–615.
Prashanth JR, Lewis RJ, Dutertre S. 2012. Towards an integrated venomics approach for accelerated conopeptide discovery. Toxicon 60(4):470–477.
Prator CA, Murayama KM, Schulz JR. 2014. Venom variation during prey capture by the cone snail, *Conus textile*. PLoS One 9(6):e98991.
Puillandre N, Duda TF, Meyer C, Olivera BM, Bouchet P. 2014. One, four or 100 genera? A new classification of the cone snails. J Mollus Stud. 81(1):1.
Puillandre N, Bouchet P, et al. 2014. Molecular phylogeny and evolution of the cone snails (Gastropoda, Conoidea). Mol Phylogenet Evol. 78:290–303.
Puillandre N, Koua D, Favreau P, Olivera BM, Stöcklin R. 2012. Molecular phylogeny, classification and evolution of conopeptides. J Mol Evol. 74(5–6):297–309.
Puillandre N, Watkins M, Olivera BM. 2010. Evolution of *Conus* peptide genes: duplication and positive selection in the A-superfamily. J Mol Evol. 70(2):190–202.
Quinton L, et al. 2013. Identification and functional characterization of a novel α-conotoxin (EIIA) from *Conus ermineus*. Anal Bioanal Chem. 405(15):5341–5351.
Rivera-Ortiz JA, Cano H, Marí F. 2011. Intraspecies variability and conopeptide profiling of the injected venom of *Conus ermineus*. Peptides 32(2):306–316.
Robinson SD, Li Q, Bandyopadhyay PK, et al. 2017. Hormone-like peptides in the venoms of marine cone snails. Gen Comp Endocrinol. 244:11–18.
Robinson SD, Li Q, Lu A, et al. 2017. The venom repertoire of *Conus gloriamaris* (Chemnitz, 1777), the glory of the sea. Mar Drugs 15:145.
Robinson SD, Norton RS. 2014. Conotoxin gene superfamilies. Mar Drugs 12(12):6058–6101.
Rodriguez A, Dutertre S, Lewis R, Marí F. 2015. Intraspecific variations in *Conus purpurascens* injected venom using LC/MALDI-TOF-MS and LC-ESI-TripleTOF-MS. Anal Bioanal Chem. 407(20):6105–6116.
Safavi-Hemami H, Bulaj G, Olivera BM, Williamson NA, Purcell AW. 2010. Identification of *Conus* Peptidylprolyl Cis-Trans Isomerases (PPIases) and assessment of their role in the oxidative folding of conotoxins. J Biol Chem. 285(17):12735–12746.

Safavi-Hemami H, et al. 2014. Combined proteomic and transcriptomic interrogation of the venom gland of *Conus geographus* uncovers novel components and functional compartmentalization. Mol Cell Proteomics 13(4):938–953.
Safavi-Hemami H, et al. 2015. Specialized insulin is used for chemical warfare by fish-hunting cone snails. Proc Natl Acad Sci U S A. 112(6):1743–1748.
Salisbury SM, Martin GG, Kier WM, Schulz JR. 2010. Venom kinematics during prey capture in *Conus*: the biomechanics of a rapid injection system. J Exp Biol. 213(5):673.
Santos AD, McIntosh JM, Hillyard DR, Cruz LJ, Olivera BM. 2004. The A-superfamily of Conotoxins: structural and functional divergence. J Biol Chem. 279(17):17596–17606.
Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27(6):863–864.
Schurch NJ, et al. 2016. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA 22(10):1641–1851.
Schwarz G. 1978. Estimating the dimension of a model. Ann Stat. 6(2):461–464.
Shon K-J, et al. 1995. Purification, characterization, synthesis, and cloning of the Lockjaw peptide from *Conus purpurascens* venom. Biochemistry 34(15):4913–4918.
Shon K-J, et al. 1998. μ-Conotoxin PIIIA, a new peptide for discriminating among tetrodotoxin-sensitive Na channel subtypes. J Neurosci. 18(12):4473.
St John J. 2011. SeqPrep. Available from: https://github.com/jstjohn/SeqPrep; last accessed September 05, 2018.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313.
Tayo LL, Lu B, Cruz LJ, Yates JR. 2010. Proteomic analysis provides insights on venom processing in *Conus textile*. J Proteome Res. 9(5):2292–2301.
Terlau H, Olivera BM. 2004. Conus venoms: a rich source of novel ion channel-targeted peptides. Physiol Rev. 84(1):41–68.
Terrat Y, et al. 2012. High-resolution picture of a venom gland transcriptome: case study with the marine snail *Conus consors*. Toxicon 59(1):34–46.
Tucker JK, Tenorio MJ. 2009. Systematic classification of recent and fossil conoidean gastropods: with keys to the genera of cone shells. Harxheim: Conchbooks.
Tucker JK, Tenorio MJ. 2013. Illustrated catalog of the living cone shells. Wellington (FL): MDM Publishing.
Uniprot Consortium. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45(D1):D158–D169.
Uribe JE, Puillandre N, Zardoya R. 2017. Beyond *Conus*: phylogenetic relationships of Conidae based on complete mitochondrial genomes. Mol Phylogenet Evol. 107:142–151.
Violette A, et al. 2012. Large-scale discovery of conopeptides and conoproteins in the injectable venom of a fish-hunting cone snail using a combined proteomic and transcriptomic approach. J Proteomics 75(17):5215–5225.
Volpon L, et al. 2004. NMR solution structures of δ-conotoxin EVIA from *Conus ermineus* that selectively acts on vertebrate neuronal Na+ channels. J Biol Chem. 279(20):21356–21366.
Woodward SR, Cruz LJ, Olivera BM, Hillyard DR. 1990. Constant and hypervariable regions in conotoxin propeptides. EMBO J. 9(4):1015–1020.
Wright ES, Vetsigian KH. 2016. Quality filtering of Illumina index reads mitigates sample cross-talk. BMC Genomics 17(1):876.
Wu Y, et al. 2013. Molecular evolution and diversity of *Conus* peptide toxins, as revealed by gene structure and intron sequence analyses. PLoS One 8(12):e82495.

Associate editor: Mandë Holford