



HAL
open science

Transcriptomic Messiness in the Venom Duct of *Conus miles* Contributes to Conotoxin Diversity

Ai-Hua Jin, Sébastien Dutertre, Quentin Kaas, Vincent Lavergne, Petra Kubala, Richard J Lewis, Paul F Alewood

► **To cite this version:**

Ai-Hua Jin, Sébastien Dutertre, Quentin Kaas, Vincent Lavergne, Petra Kubala, et al.. Transcriptomic Messiness in the Venom Duct of *Conus miles* Contributes to Conotoxin Diversity. *Molecular and Cellular Proteomics*, 2013, 12 (12), pp.3824-3833. 10.1074/mcp.M113.030353 . hal-02306944

HAL Id: hal-02306944

<https://hal.science/hal-02306944>

Submitted on 7 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Transcriptomic Messiness in the Venom Duct of *Conus miles* Contributes to Conotoxin Diversity*

Ai-hua Jin[†], Sébastien Dutertre[†], Quentin Kaas[‡], Vincent Lavergne[‡], Petra Kubala[‡], Richard J. Lewis[‡], and Paul F. Alewood^{‡§}

Marine cone snails have developed sophisticated chemical strategies to capture prey and defend themselves against predators. Among the vast array of bioactive molecules in their venom, peptide components called conotoxins or conopeptides dominate, with many binding with high affinity and selectivity to a broad range of cellular targets, including receptors and transporters of the nervous system. Whereas the conopeptide gene precursor organization has a conserved topology, the peptides in the venom duct are highly processed. Indeed, deep sequencing transcriptomics has uncovered on average fewer than 100 toxin gene precursors per species, whereas advanced proteomics has revealed >10-fold greater diversity at the peptide level. In the present study, second-generation sequencing technologies coupled to highly sensitive mass spectrometry methods were applied to rapidly uncover the conopeptide diversity in the venom of a worm-hunting species, *Conus miles*. A total of 662 putative conopeptide encoded sequences were retrieved from transcriptomic data, comprising 48 validated conotoxin sequences that clustered into 10 gene superfamilies, including 3 novel superfamilies and a novel cysteine framework (C-C-C-CCC-C-C) identified at both transcript and peptide levels. A surprisingly large number of conopeptide gene sequences were expressed at low levels, including a series of single amino acid variants, as well as sequences containing deletions and frame and stop codon shifts. Some of the toxin variants generate alternative cleavage sites, interrupted or elongated cysteine frameworks, and highly variable isoforms within families that could be identified at the peptide level. Together with the variable peptide processing identified previously, background genetic and phenotypic levels of biological messiness in venoms contribute to the hypervariability of venom peptides and their ability to evolve rapidly. *Molecular & Cellular Proteomics* 12: 10.1074/mcp.M113.030353, 3824–3833, 2013.

Cone snails are predatory marine gastropods that feed on a variety of prey, including fish and invertebrates. With ~700

cone snail species described, *Conus* represents the largest genus of all invertebrate marine animals. The current classification and phylogeny of cone snails are still a matter of debate and are being refined using genomic DNA and radula morphology data (1–3). Cone snails are classified in the taxonomic class Neogastropoda, which comprises three superfamilies, Muricoidea, Cancellarioidea, and Conoidea (4). The Conidae family belongs to the Conoidea branch, and the only genus of this family is *Conus* (1).

Cone snails have evolved potent venoms that they use for defense and capturing prey (5). These venoms are highly complex mixtures of dominant cysteine-rich conotoxins as well as cysteine-poor conopeptides, enzymes, and proteins (6–10). Conopeptides are produced as mRNA precursors displaying a mostly conserved topological organization comprising an N-terminus signal sequence followed by an intervening propeptide region, the mature toxin region, and, for some, an additional C-terminal propeptide region (11). Based on signal sequence similarities, conopeptides have been classified into 18 gene superfamilies (12, 13), which reveal evolutionary relationships between different conopeptides. Indeed, the higher evolution rate of mature peptide regions prevents the establishment of reliable phylogeny using these regions only, and only the conservation of signal sequences offers the possibility of relating conopeptide precursors (14). During their journey in the endoplasmic reticulum and the export machinery, conopeptides are excised from the precursors with proteases (15) and at the same time are heavily post-translationally modified. Currently, 14 different post-translational modifications are identified in conopeptides (13).

The most common post-translational modification is the formation of disulfide bonds, and the conopeptides with more than one disulfide bond are commonly referred to as conotoxins (16). Conotoxins are currently divided into 24 cysteine frameworks, designated using Roman numerals, according to the arrangement of cysteines in the mature peptide region (17, 18). The disulfide bond connectivities are usually important for the folding and activity of conotoxins, although they are not part of the definition of the cysteine frameworks.

Second-generation transcriptomics has to date uncovered on average fewer than 100 toxin cDNA precursors per *Conus*

From the [†]Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland 4072, Australia

Received April 23, 2013, and in revised form, September 2, 2013
Published, MCP Papers in Press, September 16, 2013, DOI 10.1074/mcp.M113.030353

species (19–23). A more impressive molecular diversity has been described at the peptide level in cone snail venoms, with >1000 detected masses observed in a single specimen (19, 24), and even closely related species can display a completely different set of conopeptides in their venom (24). Phylogenetic studies of certain gene superfamilies of conopeptides revealed extensive gene turnover, rapid evolution, and diversification within relatively recent evolutionary time (25), with conopeptide genes being among the most rapidly evolving protein-coding genes in Metazoans, a phenomenon thought to be facilitated by extensive gene duplications (25). An understanding of how *Conus* venoms have evolved to generate this vast number of peptides from a limited set of genes is expected to shed light on the rapid molecular evolution of *Conus* venom peptides (19).

In the present study, a 454 pyrosequencing approach was applied to uncover the transcriptome of a worm-hunting species of cone snail, *Conus miles*. To date, only a cDNA cloning strategy using conserved signal peptides has been applied to the discovery of conotoxins from this species, with three superfamilies (O1, D, and I2) and 10 conopeptide sequences currently identified (26–28). To fully characterize the conopeptide isoforms, cysteine frameworks, and gene superfamilies within *C. miles* venom, we integrated transcriptomic and proteomic data using bioinformatics. This approach revealed unsuspected messiness at the mRNA level (29), where we identified a series of single amino acid variants (type I variants), pre-mature stop codons (type II variants), and frame shifts (type III variants). These variations produced conopeptides with alternative cleavage sites (types I and III), interrupted or elongated cysteine frameworks (types I, II, and III), and highly variable isoforms including deletions and elongations (types II and III). Interestingly, most of these unusual toxin variants were expressed at very low levels, and given the high rates of evolution of conotoxin genes within families and the presence of these single read (mRNA) peptides in the venom, we hypothesize that this “background” genetic noise or “transcriptomic messiness” contributes to venom peptide hypervariability and, more broadly, to the rapid evolution of bioactive peptides.

EXPERIMENTAL PROCEDURES

RNA Extraction, cDNA Library, 454 Sequencing, and Assembly—One single adult specimen of *C. miles* collected from the Great Barrier Reef (Queensland, Australia) and measuring 6 cm was dissected on ice. The venom duct was removed and directly placed in a 1.5-ml tube with 1 ml of TRIZOL reagent (Invitrogen). The extraction of total RNA was carried out following the manufacturer’s instructions. mRNA was purified from the tRNA using a Qiagen mRNA extraction kit (Qiagen, Valencia, CA). The Australian Genomic Research Facility conducted the next-generation sequencing using a Roche GS FLX Titanium sequencer. The assembly was carried out using Newbler 2.3.

Conopeptide Sequence Analysis—Raw cDNA reads (expressed sequence tags) and isotigs were up-loaded in a Web-based searchable database set up by the Australian Genomic Research Facility. Sorting of raw cDNA reads was performed with ConoSorter, a stand-

alone program developed in-house to classify conopeptides into gene superfamilies and classes. Briefly, after translating nucleic acid sequences in the six reading frames, the algorithm isolates the corresponding coding regions and classifies them into superfamilies and classes by employing an approach based on the complementarity of regular expressions and profile hidden Markov models. Finally, the program searches the ConoServer database for sequences already characterized and generates additional statistical information about the matching hits (frequency of identical sequences in the raw data set, percent hydrophobicity of the signal region, sequence length, and number of cysteine residues present).

Manual identification of conopeptide sequences was carried out from the retrieved data. Gene superfamilies, signal peptides, and cleavage sites were predicted using the ConoPrec tool implemented in ConoServer. The cut-off value for assigning a signal peptide to a gene superfamily was set at >75% sequence identity, as extrapolated from a recent analysis of all precursors deposited in ConoServer (17, 19).

Venom Sample Preparation—The pooled venom obtained from three adult (≥ 6 cm) specimens of *C. miles* collected from the Great Barrier Reef (Queensland, Australia) was used for proteomic studies. Dissection was performed on ice, the venom ducts were squeezed, and the contents were collected in 1 ml of 0.1% formic acid and stored at -20° C until further use.

Reduction-Alkylation and Enzyme Digestion—Reduction and alkylation of the cysteine bonds was carried out as previously described (19). Sigma proteomics sequencing-grade trypsin and endoproteinase Glu-C were used to digest the reduced and alkylated venom samples, and the enzymes were activated in 40 mM NH_4HCO_3 buffer. A ratio of 1:100 (w/w) of enzyme to venom peptides was used. The digestion was carried out in a microwave apparatus for 4 min on the lowest power setting.

Mass Spectrometry and Proteomic Analysis—Liquid chromatography-electrospray mass spectrometry (LC-ESI-MS)¹ was performed on an AB Sciex 5600 TF as previously described (19). Briefly, the dissected venom extracted as described above (~ 8 μl supernatant) was directly subjected to LC-ESI-MS in order to obtain a complete mass list of underivatized peptides. Information Dependent Acquisition was performed on the reduced, reduced/alkylated, and enzymatically digested venom samples (*i.e.* four sets of samples for MS/MS). We used ProteinPilot 4.0 software for peaklist generation and sequence identification by searching the LC-ESI-MS/MS spectra against the raw cDNA database (1,534,974 entries) generated by the Roche 454 GS FLX Titanium sequencer. For comparison, the spectra were also searched against a publicly accessible database extracted from UniProtKB using “venom protein” as the keyword (3906 entries). With the alkylated samples, the fixed modification was set as maleimide for cysteine alkylation. Nine different types of variable modifications that have been identified on conopeptides were considered: amidation, deamidation, hydroxylation of proline and valine, oxidation of methionine, carboxylation of glutamic acid, cyclization of N-terminal glutamine (pyroglutamate), bromination of tryptophan, and sulfation of tyrosine. The mass tolerance was set as 0.05 Da for precursor ions and 0.1 Da for the fragment ions. Tandem mass spectra were only acquired to the 2 to 5 charged ions, and the switch criteria were set to exclude former target ions for 8 s and to exclude isotopes within 4 Da. The threshold score for accepting individual peptide spectra was 99. The detected peptide pieces were manually inspected and validated.

¹ The abbreviations used are: LC-ESI-MS, liquid chromatography-electrospray mass spectrometry.

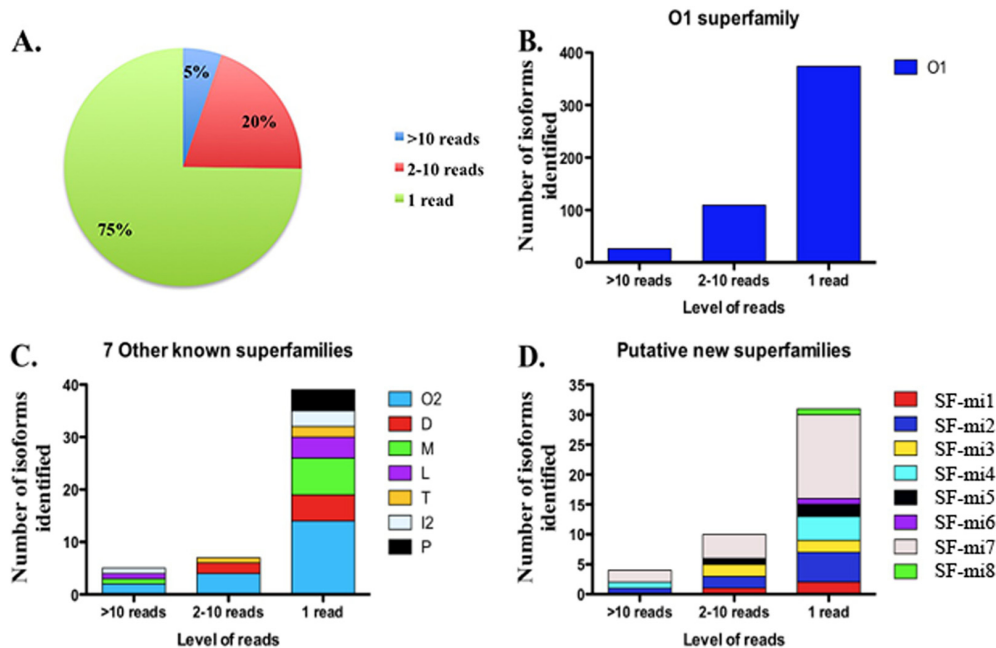


Fig. 1. **Isoforms and transcription levels of putative conopeptides.** A, variations in the level of transcription. Only 5% of the putative sequences had more than 10 cDNA reads, 20% of the putative sequences had moderate cDNA reads of 2 to 10, and 75% of the putative sequences were present as rare transcripts with only a single cDNA read discovered for the full-length precursor. B, isoforms of superfamily O1. C, isoforms of other known superfamilies. D, isoforms of eight putative new superfamilies.

RESULTS

Transcriptomic and Bioinformatic Data Analysis—A single run on the Roche GS FLX Titanium sequencer (one-quarter of a plate equivalent for *C. miles*; see “Experimental Procedures”) generated 255,829 cDNA reads averaging 325 bp (minimum of 19 bp) after trimming and removal of low-quality sequences. The raw cDNA reads were assembled using Newbler 2.3 software. Both the raw cDNA reads and assembled isotigs (including contigs) were sorted by our in-house program ConoSorter. After translation and motif searching using parameters generated from the ConoServer database, 17,215 and 50 peptide precursors were retrieved from the raw data and the isotigs, respectively. These peptide precursors were manually examined according to homology analysis generated by the ConoSorter program. Interestingly, only a small fraction of the total number of conopeptide precursors found in the raw data were retrieved from the assembled isotigs, indicating that genetic diversity is underestimated if only isotigs are analyzed. Overall, 662 precursors were characterized as putative conopeptide sequences based on the conserved precursor structures. Because the raw cDNA reads library was not normalized, the level of mRNA transcription could be inferred from the number of cDNA reads that coded for each conopeptide (19).

Dramatic differences at the mRNA level were observed between the different conopeptide precursors. Isoform MiEr95, belonging to the O1 gene superfamily, largely dominated the transcriptome with 4128 cDNA reads, whereas 495

putative conopeptide precursors were identified with only 1 cDNA read. These rare transcripts constituted ~75% of the total putative conopeptide precursors retrieved (Fig. 1). In addition to these sequences, 35 high-level precursors (>10 cDNA reads) and 132 low-level precursors (2 to 10 cDNA reads) were also identified.

Using a 75% signal peptide homology cut-off (17, 19), we clustered the 662 putative conopeptide precursors into eight known (*i.e.* O1, O2, D, M, T, I2, L, and P) and eight putative new (1 to 8) gene superfamilies. The signal peptides and cysteine frameworks are listed in Table I. The identification of known gene superfamilies was confirmed using the ConoPrec tool in ConoServer. Conopeptide isoforms from all three previously discovered superfamilies (O1, D, and I2) from *C. miles* (26–28) were observed. However, only 3 (MiEr95, MiEr93, and Mi20.1) of the 10 known sequences described in the literature were identified in this transcriptome, probably because of the well-known phenomena of intraspecific variation in cone snails (24, 30, 31); previous discoveries were made using venom pooled from 9 to 15 specimens. In addition to these known sequences, new isoforms from five other gene superfamilies (M, O2, L, T, and P) were also found. Finally, eight putative new gene superfamilies (coded SF-mi1–8) were identified. SF-mi1 and -2 are closely related to superfamily M (64% and 57%), whereas SF-mi3 is closely related to superfamily O2 (69%). SF-mi4, -5, and -7 contained only two cysteine residues in their mature conopeptides; they all showed less than 50% homology to the signal peptides of other

TABLE I
Representative signal sequence(s) and cysteine framework for *C. miles* superfamilies

Gene superfamily	Signal sequence(s)	Cysteine pattern	Framework
O1	MKLLCVLIVAMLPLMACHLIIA MKLTCALIITLLFLSITAG	C-C-CC-C-C	VI/VII
O2	MEKLTVLILVATVLLTIQVLG	C-C-CC-C-C-C-C	XV
D	MPKLEMMLLVLLILPLSSFSAA	C-CC-C-CC-C-C-C-C	XX
M	MSKLGVLVFLVFLFTMATLQLDA	CC-C-C-CC	III
L	MKLSVMFIVFLMLTMTDMPMTDG	C-CC-C	XXIV
T	MCCLPVFIILLIPAS	CC-CC	V
I2	MMCRLTSLCCLLVIVLLNSAVDG	C-C-CC-CC-C-C	XI
P	MHLSLAGSAVLVLLLLFALGNFAGVQP	C-C-C-C-C-C	IX
SF-mi1	MSKTGLVLVLYLLSSPVNL	C-C-C-CC-C-C-C	XIII
SF-mi2	MRFFLLLTVALFLTSITG	C-C-C-CCC-C-C	Novel
SF-mi3	MGILTVLLPLVAVLVT	C-C-CC-C-C	VI/VII
SF-mi4	MTPRMNLMLTFVVMVTVPLLLA	C-C	-
SF-mi5	MGLLPLQTSVLLAPVHQ	C-C	-
SF-mi6	MSTLGKVVLLLLLPLGNP	C-C-CC-C-C-C-C	XV
SF-mi7	MSTLNPLTRIWYWRASLVPAAAVPIPIAYT	C-C	-
SF-mi8	MMYRLTLFCCLLVIVPLNMA	C-C-C-CC-C-C-C	XIII

known superfamilies. In contrast, SF-mi6 and -8 had eight cysteine residues in their mature peptides, and their signal peptides were ~ 60% homologous to M and I2, respectively.

The total numbers of isoforms and cDNA reads for each superfamily are plotted in Fig. 1. Overall, the greatest number of isoforms was discovered for superfamily O1, accounting for approximately four times more sequences than the isoforms from the remaining superfamilies combined, irrespective of read number (Fig. 1B). Only four other known superfamilies (O2, M, L, and I2) contain isoforms with high-level cDNA reads (>10 reads). D and T superfamily isoforms display only low-level cDNA reads (2 to 10 reads and 1 read, respectively). Four identified isoforms belonged to the P superfamily, all with only one cDNA read (Fig. 1C). Some isoforms in the three putative new superfamilies were identified with high cDNA reads (SF-mi2, -4, and -7). The remaining five putative new superfamilies had low- or very low-level cDNA read numbers (Fig. 1D).

Proteomic Data Analysis—To interrogate our transcriptomic sequences at the peptide level, we employed a proteomic strategy involving LC-ESI-MS and LC-ESI-MS/MS to uncover the complexity of *C. miles* dissected venom. To improve MS/MS fragmentation, the whole venom was reduced and alkylated prior to digestion (32) by either endo-GluC or trypsin. The total ion chromatogram (Fig. 2) illustrates the mass profile of the native venom sample. The pooled dissected venom of *C. miles* from three specimens was complex; nevertheless, 9 out of the 10 previously published conopeptide sequences from *C. miles* could be identified as major components (see Fig. 2). MS/MS coverage was obtained across all gene precursors regardless of their level of transcription; that is, some mRNA precursors with high-level cDNA reads could be identified at the peptide level, and surprisingly some mRNA transcripts with low-level and very low-level cDNA reads were confirmed at the peptide level as well. Specific examples are illustrated in the following sections.

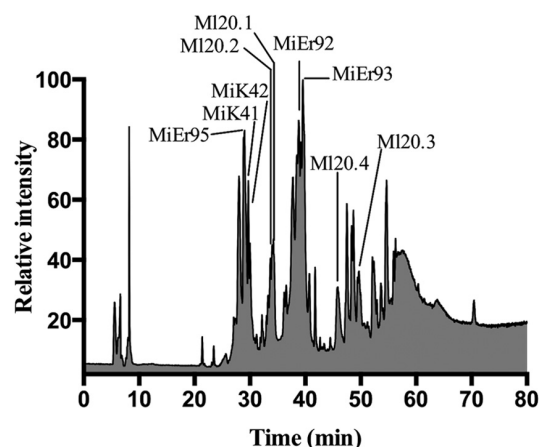


Fig. 2. Total ion chromatogram of *C. miles*. LC-MS run on the TripleTOF 5600 revealed the complexity of the dissected venom of *C. miles*. Detected masses corresponding to the previously identified conotoxins are indicated.

Conotoxin Precursors, Mature Peptides, and Cysteine Frameworks—From the 662 putative conopeptide sequences identified in the venom gland transcriptome of *C. miles*, only those that complied with the following four criteria were selected for further analysis: (i) the mature conopeptide region should contain more than four cysteine residues; (ii) the full length of the mRNA precursor must be validated by two or more cDNA reads; (iii) mutations should occur in the mature region excluding propeptides (identical mature peptides from different propeptides can complicate the integration process of the transcriptomics and proteomics); and (iv) the predicted mature peptides should contain an uninterrupted canonical cysteine framework (or frameworks) within the gene superfamily (*i.e.* no odd-number cysteines due to single cysteine residue mutation). The sequences of 48 conotoxin precursors selected with these criteria are shown in Fig. 3. ConoPrec was

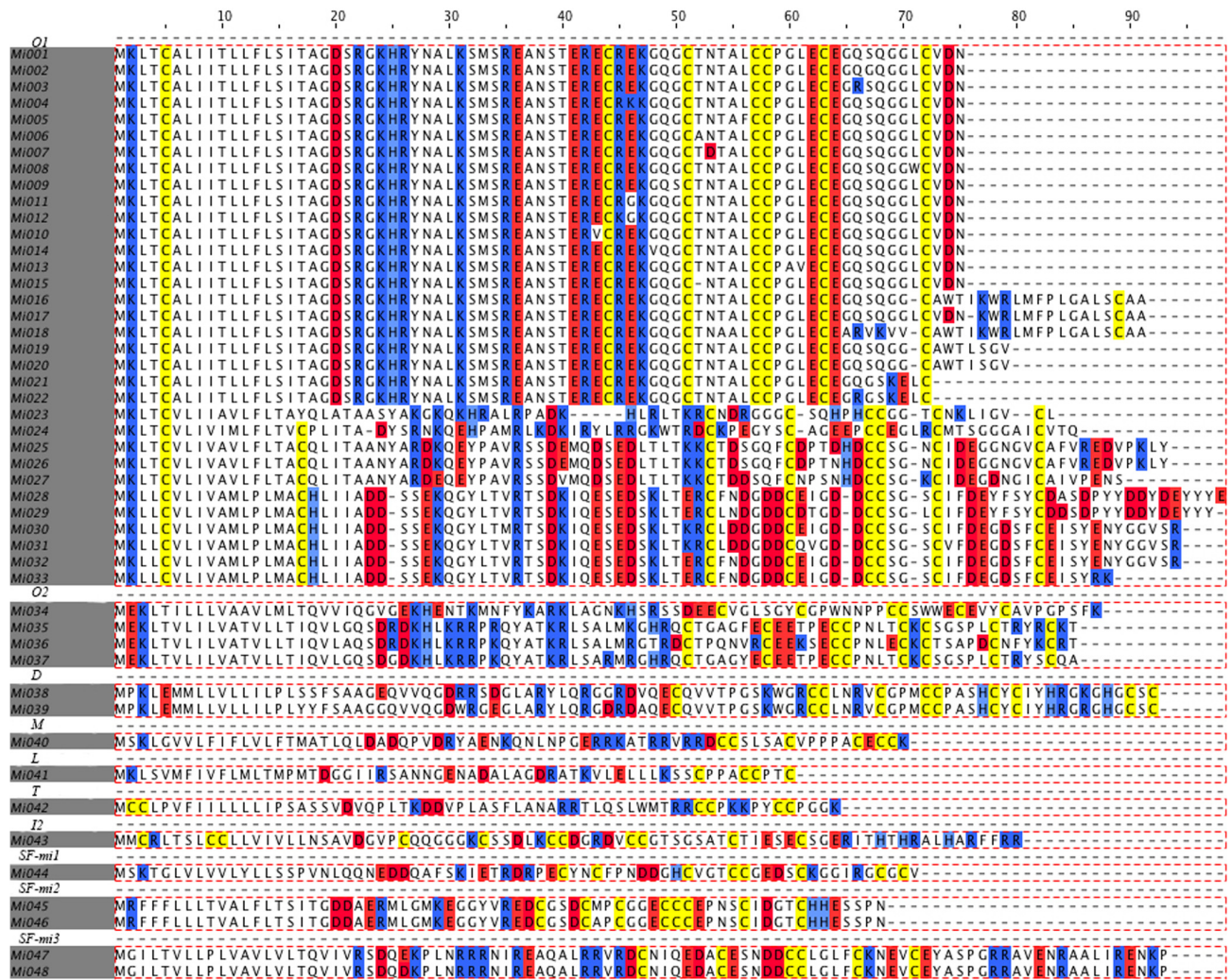


Fig. 3. Alignment of 48 conotoxin precursors. Sequences are clustered by gene superfamily according to their signal peptide, with gaps introduced to optimize alignment. Color-coding has been applied using the following scheme: cysteine residues are in yellow, negatively charged residues are in red, and positively charged residues are in blue.

used to predict the mature peptides, and the mature sequences that could be identified with an MS/MS confidence value of more than 99% are listed in Table II.

Superfamily O1—The numbers of cDNA reads and conotoxin variants were greatest in this superfamily, with a total of 5912 cDNA reads and 33 isoforms that clustered into two subfamilies (Fig. 3). The first subfamily contained 22 isoforms (Mi001–Mi022), including the known MiEr95 (Mi001). The full precursor of this subfamily contained a pre-sequence cleavage site (ER or KR), resulting in predicted mature peptides containing between 31 and 52 amino acids and a Type VI/VII framework. MiEr95 was the most highly expressed isoform in the whole transcriptome, with 4128 cDNA reads for the full precursor and 7667 for the mature peptide alone. In addition to MiEr95, 14 variants were also unambiguously identified via MS/MS within this subfamily (Table II).

The second subfamily (Mi023–Mi033) had a different signal sequence than the first subfamily. The full precursor of this subfamily had a presequence cleavage site (KR, ER, and RR), resulting in predicted mature peptides comprising 30 to 40 amino acids, three disulfide bonds, and a classic Type VI/VII (C-C-CC-C-C) framework. Remarkably, isoforms in this subfamily were enriched in acidic residues (~24%) and contained more than twice as many aspartic acid residues (~17%) as glutamic acid residues (~7%). Such high levels of negatively charged residues appear remarkable, as the average frequencies of aspartic acid and glutamic acid residues in mature conopeptides are only ~4.4% and 3.7%, respectively (13). Although the known MiEr93 (Mi023) has smaller loop sizes and contains more basic residues than other precursors, it was included in this subfamily based upon its conserved signal peptide. MiEr93, along with three other sequences

TABLE II
MS/MS sequences identified at >99% confidence value

Sequence number	MS/MS (>99% confidence)	Number of reads	Cysteine modification	Post-translational modifications	Precursor <i>m/z</i>	<i>z</i>
Mi001	ECREKGGGCTNTALCCP <u>GLECE</u> EGQSQQGGLCVDN	4128			1134.80	3
Mi003	ECREKGGGCTNTALCCP <u>GLECE</u> GRSQGGLCVDN	5	Y	1Hyp	1007.87	4
Mi004	ECRKKGGGCTNTALCCP <u>GLECE</u> EGQSQQGGLCVDN	3	Y		996.63	4
Mi005	GQGCTNTAFCCP <u>GLECE</u> EGQSQQGGLCVDN	2	Y		1092.73	3
Mi007	ECREKGGGCTDTALCCP <u>GLECE</u> EGQSQQGGLCVDN	2	Y	1Hyp	1001.13	4
Mi008	ECREKGGGCTNTALCCP <u>GLECE</u> EGQSQQGWCVDN	2	Y		1015.11	4
Mi009	EKGQSCNTALCCP <u>GLECE</u> EGQSQQGGLCVDN	2	Y		1177.11	3
Mi010	VCREKGGGCTNTALCCP <u>GLECE</u> EGQSQQGGLCVDN	21	Y		989.37	4
Mi012	ECKGKGQGCNTA	2	Y		745.83	2
Mi013	ECREKGGGCTNTALCC <u>PAVE</u> CEGQSQQGGLCVDN	2		1Hyp	1140.13	3
Mi014	ECREKVQGCNTALCCP <u>GLECE</u> EGQSQQGGLCVDN	2			1148.80	3
Mi015	ECREKGGGCTNTALCC <u>PG</u>	2	Y		1078.91	2
Mi016	ECREKGGGCTNTALCCP <u>GLECE</u> EGQSQQGGA	29		1Hyp,NH ₂	1016.44	3
Mi018	ECREKGGGCTNAA	2	Y		780.81	2
Mi020	ECREKGGGCTNTALCCP <u>GLECE</u> EGQSQQGCAWTI	2	Y	1Hyp,NH ₂	1008.13	4
Mi023	GGGCSQHPHCCGGTGNK	9	Y	1Hyp	683.91	3
Mi025	PTDHDCCSGNCIDEGGNGVCAVREDVPKLY	4			829.10	4
Mi026	PTNHDCCSGNCIDEGGNGVCAVREDVPKLY	2			828.86	4
Mi027	CTDDSQFCNPSNHDCSSGKCIDEGDNGICAIPENS	31			948.37	4
Mi035	CPNLTCKCSGSPLCTRYRCKT	12		2Hyp	592.27	4
Mi036	CKCTSAPDCNFYKCR	8		1Hyp	619.26	3
Mi037	CKCSGSPLCTRYSCQA	5		1Hyp	861.86	2
Mi040	DCCSLSACVPPACECK	32		1Hyp	922.85	2
Mi041	SSCPPACCP	18			534.68	2
Mi042	CCPKPYCCPG	5		NH ₂	599.25	2
Mi043	VPQQGGGK	10	Y	1Hyp	493.73	2
Mi044	DRPECYNCFPNDGHCVGTCCGEDSCKGGIRGCGCV	3			754.69	5
Mi045	CMPCGGECCEPNSCIDGTCHHE	8	Y		1036.28	3
Mi046	EDCGSDCAPCGECCPEPNSCIDGTCHHESSPN	12		1Hyp	1125.70	3

Notes: Cys-mod (Y), cysteine alkylation by maleimide; Hyp, hydroxyproline; NH₂, C-terminal amidation. Underlined residues indicate post-translational modification sites.

within this subfamily, was identified via MS/MS analysis (Table II).

Superfamily O2—Three precursors (Mi035–037) with 100% signal sequence homology and one precursor (Mi034) with 82% homology to the O2 superfamily were identified. All the full-length precursors contained an obvious presequence cleavage site (K/R-R), resulting in predicted mature peptides containing 29 to 39 amino acids. Two cysteine frameworks exist in the O2 superfamily, including three precursors (Mi035–037) with a Type XV (C-C-CC-C-C-C) framework that were identified via MS/MS after enzymatic digestion. In contrast, the predicted mature peptide for Mi034 contained only six cysteine residues within the Type VI/VII (C-C-CC-C-C) framework. However, despite the abundance of this sequence at the transcript level (42 cDNA reads), MS/MS support had a confidence value of <50%.

Superfamily D—Of the two superfamily D precursors (Mi038 and Mi039), Mi038 had been previously discovered (MI20.1) (26). Both contain a pre-sequence cleavage site (R-R). MI20.1 and the novel Mi039 were expressed with four and two cDNA reads, respectively. MS/MS analysis confirmed both sequences at the peptide level with >69% of confidence value.

Superfamily M—No conopeptides from superfamily M had been previously discovered from *C. miles*. One precursor (Mi040), with a signal peptide with 96% homology to classic M superfamily members, was discovered with 32 cDNA reads. It contained an obvious presequence cleavage site (RR), resulting in a short mature peptide of 18 amino acids and cysteine framework III (CC-C-C-CC). The mature sequence was confirmed via MS/MS analysis and contained three sequential proline residues.

Superfamily L—One precursor (Mi041) from the L superfamily was identified with an L-K cleavage site predicted by ConoPrec (13, 33). Precursor Mi041 had 18 cDNA reads, and the predicted mature peptide was identified with good MS/MS coverage, containing only 11 amino acid residues. This peptide was a member of the recently reported cysteine framework XXIV (C-CC-C) (18) and included three proline residues in the two loops despite its short sequence.

Superfamily T—One precursor (Mi042) with 100% sequence homology to classic T superfamily members was identified with five cDNA reads and an RR cleavage site. The sequence of the resulting short mature peptide (13 residues) had a classic T superfamily cysteine framework V (CC-CC) that was confirmed by MS/MS.

TABLE III
Interrupted cysteine frameworks from the *C. miles* O1 superfamily

Variations		Sequences	Copies
	Reference precursor	<i>KRECREKGQGCTNTALCCPGLCEGGQSQQGGLCVDN</i>	4128
Type I single mutant	New pre-cleavage site	<i>KRERRREKGQGCTNTALCCPGLCEGGQSQQGGLCVDN</i>	3
	5-cysteine framework	<i>KRECREKGQGCTNTALCYPGLECEGGQSQQGGLCVDN</i>	2
Type II pre-mature stop codon	Truncated framework	<i>KRECREKGQGCTNTALCCPG</i>	223
Type III frame shift	Truncated framework	<i>KRECREKGQGCTNTALAAPARM</i>	7
	New post-cleavage site	<i>KRECREKGQGCTNTALFPCIPSQWNSGKKVRVLTQLFAAPA</i>	5
	Different loop size	<i>KRECREKGQGCTNTALCCPGLDVKARVKVCAWTIKWRLMFPLGALSCAA</i>	3

Notes: Underlined residues indicate sequence changes caused by mutations. N-terminal and post-cleavage sites are shown in italics. Cysteine residues are in bold to highlight frameworks.

Superfamily I2—One sequence (Mi043) was found in the I2 superfamily with 13 cDNA reads. The I2 superfamily has a different precursor structure, with the mature sequence placed between the signal peptide and the propeptide regions. There was a post-cleavage site (ER) for Mi043, leaving the predicted mature peptide with 39 amino acid residues, including 8 cysteine residues and an amidated C terminus. The cysteine framework was Type XI (C-C-CC-CC-C-C), and MS/MS identified a partial sequence of a tryptic digest (Table II).

Superfamily SF-mi1—One precursor (Mi044) was identified whose signal peptide was 64% homologous to the M superfamily, though with a different cysteine framework. The full precursor contained a presequence cleavage site (E-R), resulting in a predicted mature peptide with 35 amino acids and four disulfide bonds in a Type XIII framework (C-C-C-CC-C-C). The predicted mature sequences contained six acidic residues and three basic residues. Despite only moderate transcription at the mRNA level (three cDNA reads), the mature peptide of Mi044 produced good MS/MS coverage.

Superfamily SF-mi2—This novel superfamily contained two precursors (Mi045 and Mi046), with signal peptides having only 57% homology to the M superfamily. Both gave a moderate number of cDNA reads (8 and 12) and contained a novel cysteine framework with eight cysteine residues (C-C-C-CCC-C-C). The pre-cleavage sites were predicted after the first K (position 29) by ConoPrec to generate mature peptides of 39 amino acids with an eight-residue-long N-terminal tail. A conserved arginine residue at position 35 produced a second cleavage site, resulting in mature peptides 33 amino acids in length with a two-residue-long N-terminal tail. Both the K29 and R35 cleavage sites were confirmed by MS/MS sequencing, with the shorter peptide dominating the mature products for Mi045 and Mi046.

Superfamily SF-mi3—Two precursors (Mi047 and Mi048) contained a signal peptide with 69% homology to the O2 superfamily that was distinct from SF-mi1 and SF-mi2 signal peptides and their corresponding cysteine frameworks. We therefore classified these two sequences as belonging to a new superfamily. The full precursor contained both a presequence cleavage site (RR) and a post-cleavage site (RR) that

resulted in a predicted mature peptide of 33 amino acids and three disulfide bonds with a classic VI/VII (C-C-CC-C-C) framework. However, MS/MS coverage was weak for both peptides, even after Glu-C enzymatic treatment.

Messiness at the mRNA Level—We identified a large number (~600) of putative conopeptides of three distinct types comprising a series of single amino acid variants (type I), pre-mature stop codons (type II), and frame shifts (type III) (see Table III). Some of these variations produced conopeptides with alternative cleavage sites (types I and III), interrupted cysteine frameworks (types I, II, and III), and highly variable isoforms with deletions and elongations (types II and III).

Interrupted Cysteine Framework—Each of the type I–III variations can interrupt the normally conserved cysteine framework patterns, as illustrated by the examples from superfamily O1 listed in Table III. The single mutation of a cysteine residue (type I) creates an odd number of cysteine residues, allowing dimers to be formed. Pre-mature stop codons (type II) produce truncated isoforms as well as truncated frameworks. Frame shifts (type III) produce highly variable isoforms including cysteine deletions/additions or loop sizes that are significantly different from those in the canonical cysteine frameworks. Alternative cleavage sites were also observed and are shown in Table III. These interrupted cysteine frameworks gave high cDNA reads, though no MS/MS data were found to support their expression in the venom. Why such precursors were not identified at the peptide level is unclear, but it may have been the result of unidentified heterodimer formation. A similar phenomenon was observed at the cDNA level by Terrat *et al.*, and sequences containing five and seven cysteine residues have been listed in their new scaffolds table (23).

Rare Transcripts—An unusual aspect of the *C. miles* transcriptome is that ~75% of the 662 putative conopeptides were identified as rare transcripts with only single cDNA reads retrieved for the full-length precursors (Fig. 2A). Some of these may result from sequencing artifacts, as the sequencing error rate for the 454 is reported to reach ~1% (34). Therefore, out of the total of 6271 cDNA reads encoding full-length conopeptide precursors, about 63 are estimated to result from sequencing errors, and as 495 isoforms are detected with a

TABLE IV
Rare transcripts with MS/MS coverage

Superfamily	Sequence	Cysteine modification	Post-translational modifications	Precursor <i>m/z</i>	<i>z</i>
D	MPKLEMMLLVLLILPLSSFSAAGDKVVQGD RRSDGLAR YLQ RGGRDVQDCQVVT <u>PGSKWGR</u> CCLN RVC GPMCCPAS HCYCIYHRGKGHGCS		Hyp	597.62	3
M	MLKVG VFLVFLVLLSLADSWNGD NPGRQRGEKQSPQR NVFRSNLRKYNSYQKRRCANSTPCGECTDEGK <u>ICQVQ</u> <u>GGK</u> GTGECVPNTR	Y	Hyp	521.75	2
SF-mi1	MSKTGLV LVLVLYLLSSPVNLQQNEDDQAF SKIET DRP <u>ECY</u> <u>NCFP</u> <u>NDDGH</u> <u>CVGTCCGEDS</u> <u>CKGGIRGCGCL</u>			946.61	4
SF-mi4	MTPRMNLLLMTFV MTVPLLLTQ <u>PNVRCET</u> <u>NDNGDIIC</u> <u>KKANGQT</u> LHN PDRSDNGTTDLRQEGPLMTALRGP WHR	Y	Pyro-glu	899.07	3
SF-mi8	MMYRLTLFC LLLVIVPLNMARKSGMLAR <u>WNEVCSTSS</u> <u>CGCNQDER</u> CCL ECPLRIA EKQLSVMC RTFNAD TKELTA CRMLQTDSDS	Y		1104.87	2

Notes: >99% confidence regions highlighted in bold. Cys-mod (Y), cysteine alkylation by maleimide; Hyp, hydroxyproline; Pyro-glu, glutamine to pyroglutamate. Underlined residues indicate post-translational modification sites.

single cDNA read, the sequencing error accounts for only one-eighth of these. Importantly, many of these rare reads are unrelated to common transcripts and thus unlikely to be read errors. Moreover, a number of these rare transcripts produced confident MS/MS coverage (see Table IV), with MS/MS validating the new single cDNA read isoforms from D and SF-mi1 superfamilies and the propeptides for SF-mi4 and SF-mi8 superfamilies. In addition, enzymatic digestion allowed partial sequence identification for a new isoform from the M superfamily via MS/MS. These proteomic results confirm that the single cDNA read transcriptomic sequences previously overlooked contribute to diversity at the mRNA level and that the majority of this “transcriptomic messiness” does not simply arise from 454 systematic sequencing errors.

DISCUSSION

Deep sequencing combined with mass spectrometry was used to discover new peptides and gene families from the worm-hunting cone snail *C. miles*. From the 662 putative conopeptide precursors retrieved from transcriptomic data, 48 conotoxin sequences within seven known superfamilies and three new superfamilies were analyzed in detail. In addition, one new cysteine framework and eight other known cysteine frameworks were revealed (Table I). High-confidence MS/MS matches (>99% of confidence value) within the proteomic data were achieved for 29 peptides. Prior to this work, only three superfamilies had been identified from the venom of *C. miles* (26–28). Our present work confirmed the presence of these three superfamilies, along with four other known superfamilies and three novel superfamilies.

We obtained ~60% MS/MS coverage for our selected set of 48 transcriptomic sequences, similar to results obtained in our previous study of *C. marmoratus* (~66% MS/MS coverage of 105 transcriptomic sequences) (19). Because the correlation between mRNA and protein levels often deviates from a simple 1:1 relationship, levels of expression likely differ de-

pending on the nature of the peptide (35, 36), with nuclear retention potentially suppressing expression even for transcripts with higher copy numbers. Modern MS can provide unambiguous identification of peptides at the MS/MS level; however, the efficiency of MS fragmentation depends upon the abundance, sequence composition, and local chemical environment (37). Thus, limitations of current MS/MS likely contribute to incomplete reconciliation between proteomic and transcriptomic data.

Remarkably, the transcriptome of the *C. miles* venom duct was largely dominated by one toxin superfamily, in terms of both the level of mRNA transcription and the number of conopeptide isoforms present. Indeed, superfamily O (including O1 and O2) accounted for more than 90% of all the conopeptide cDNA reads identified. This high level of transcription for the superfamily also generated a high number of isoforms (77%). Interestingly, the O superfamily is also abundant (44%) in the transcriptome of another worm-hunting cone snail, *C. pulicarius* (22). The O superfamily has diverse pharmacology, with ω -, μ O, δ , γ -, and κ -conotoxins targeting voltage-gated calcium, sodium, and potassium channels (38). Although most biologically active O-superfamily members characterized so far are identified from fish-hunting species (17, 39), the prevalence of O-superfamily toxins in worm-hunting cone snails suggests an important role in prey capture across diverse phyla.

Surprisingly, no precursors from the A superfamily were discovered in the *C. miles* transcriptome, despite its being the most widely distributed gene family (38). Interestingly, although A superfamily isoforms Pu1.1–Pu1.3 were previously discovered in the venom of *C. pulicarius* (40), they were not detected in the venom gland transcriptome (22). At the protein level, we were also unable to identify A-superfamily-related peptides when using NCBI cone-snail-related proteins as the searching database. These results suggest that A-superfamily conopeptides are not critical for vermivorous cone snails. In

contrast, sequences belonging to superfamily D, which targets nicotinic acetylcholine receptors, were found, albeit at low transcription levels, but they were reasonably abundant at the peptide level, with several identified via MS/MS analysis. Interestingly, all the D-superfamily peptides (<30 isoforms) reported so far are from three clades (*i.e.* Rhizoconus, Rhombiconus, and Strategoconus) of worm-hunting cone snails that include *C. miles* (41).

Surprisingly, a high density of aspartic and glutamic amino acid residues was found in the mature sequences of three groups of *C. miles* conotoxins: SF-mi2, SF-mi3, and one subgroup of O1 superfamilies. In SF-mi2, four glutamic and three aspartic residues are evenly distributed along the sequences of these five-loop peptides, indicating that these residues might be exposed in its novel cysteine framework that includes three consecutive cysteine residues and very small loop sizes (3–2–3–4–4). Three sequential cysteine residues (CCC) have also been found in two other conotoxin frameworks (II (42) and IXI (43)) and in funnel-web spider atracotoxins (44). In contrast, SF-mi3 superfamily mature peptides containing acidic residues are mostly distributed in loop 1 and loop 2 of the peptides, reminiscent of O1 superfamily subgroup peptides that contain aspartic acids in loop 1, loop 2, and loop 4. In comparison to these acidic peptide families, the conantokins have significantly more glutamic acid than aspartic acid residues (E:D > 2.5:1) and are cysteine poor. In this NMDA antagonist peptide family, the heavily γ -carboxylated glutamic acid residues are clustered on one face of the helical structured N terminus in an arrangement that might bind calcium (45–49). γ -Carboxylation has also been observed in *C. miles* acidic peptides, but infrequently.

In addition to new superfamilies and cysteine frameworks, the *C. miles* transcriptome contained the largest number of toxin variants yet identified at the mRNA level from a single *Conus* species. From the >600 putative conopeptide sequences discovered, we found (i) mutations in the propeptide region only but identical mature peptide, (ii) interrupted cysteine frameworks, and (iii) isoforms that appeared at very low transcription levels. Strikingly, the majority of this mRNA messiness was also transcribed at low levels, explaining why these rare sequences have eluded previous studies using traditional transcriptomic approaches. Nevertheless, the diversity discovered here at the genetic level is impressive, suggesting that it might play an evolutionary role. Previously, we reported a high level of sequence diversity at the peptide level in the molluscivorous *C. marmoreus*, also at background levels (19). Many of these peptides were N- and C-terminal truncations of the major mature peptide toxins, likely resulting from promiscuous enzymatic activity that we describe as “variable peptide processing” (19). Although not analyzed to the same extent, variable peptide processing has also been observed in the venom of *C. miles* (data not shown). Based on the results from *C. miles*, it now appears that both genetic and post-translational messiness contribute to the overall cono-

toxin diversity. A theory recently developed in the field of enzymology stipulates that the origins of evolutionary innovations are the results of infidelity and heterogeneity inherent to most biological processes, leading to genetic and phenotypic variations (29). We propose that background genetic diversity, in addition to the peptide diversity arising from variable peptide processing, contributes to venom peptide diversity. Venom peptide messiness thus provides a nascent pool of accumulated chemical diversity that could contribute to the rapid evolution of venom peptides with new functions and provides a novel mechanism of adaptation in cone snails.

Acknowledgments—We thank Sandy Pineda-Gonzales for her help with RNA extraction and purification, and members of the Brisbane Shell Club for collecting the specimens of *C. miles* used in this study.

* This work was supported by an NHMRC Program Grant (R.J.L. and P.F.A.), an Early Career Research grant from The University of Queensland (to S.D. and A.H.J.), an NHMRC Principal research Fellowship (R.J.L.), and a grant from the Australian Research Council (Q.K.). S.D. was the recipient of a UQ postdoctoral fellowship. The AB SCIEX 5600 mass spectrometer was supported by an ARC LIEF grant.

§ To whom correspondence should be addressed: Professor Paul F. Alewood, Institute for Molecular Bioscience, the University of Queensland, Queensland 4072, Australia, Tel.: 61-7-3346-2982, Fax: 61-7-3346-2101, E-mail: p.alewood@imb.uq.edu.au.

¶ These authors contributed to this work equally.

REFERENCES

1. Bouchet, P., Kantor, Y. I., Sysoev, A., and Puillandre, N. (2011) A new operational classification of the conoidea. *J. Molluscan Stud.* **77**, 273–308
2. Biggs, J. S., Watkins, M., Corneli, P. S., and Olivera, B. M. (2010) Defining a clade by morphological, molecular and toxinological criteria: distinctive forms related to *Conus praececellens* A. Adams, 1854. *Nautilus* **124**, 1–19
3. Duda, T. F., Jr., Bolin, M. B., Meyer, C. P., and Kohn, A. J. (2008) Hidden diversity in a hyperdiverse gastropod genus: discovery of previously unidentified members of a *Conus* species complex. *Mol. Phylogenet. Evol.* **49**, 867–876
4. Modica, M. V., and Holford, M. (2010) The Neogastropoda: evolutionary innovations of predatory marine snails with remarkable pharmacological potential. In *Evolutionary Biology—Concepts, Molecular and Morphological Evolution*, pp. 249–270, Pontarotti, P. (ed.), Springer-Verlag Berlin Heidelberg
5. Olivera, B. M. (1997) E. E. Just Lecture, 1996. *Conus* venom peptides, receptor and ion channel targets, and drug design: 50 million years of neuropharmacology. *Mol. Biol. Cell* **8**, 2101–2109
6. Violette, A., Biass, D., Dutertre, S., Koua, D., Piquemal, D., Pierrat, F., Stocklin, R., and Favreau, P. (2012) Large-scale discovery of conopeptides and conoproteins in the injectable venom of a fish-hunting cone snail using a combined proteomic and transcriptomic approach. *J. Proteomics* **75**, 5215–5225
7. Czerwiec, E., Begley, G. S., Bronstein, M., Stenflo, J., Taylor, K., Furie, B. C., and Furie, B. (2002) Expression and characterization of recombinant vitamin K-dependent gamma-glutamyl carboxylase from an invertebrate, *Conus textile*. *Eur. J. Biochem.* **269**, 6162–6172
8. Marsh, H. (1971) The caseinase activity of some vermivorous cone shell venoms. *Toxicon* **9**, 63–67
9. Moller, C., Vanderweit, N., Bubis, J., and Mari, F. (2013) Comparative analysis of proteases in the injected and dissected venom of cone snail species. *Toxicon* **65**, 59–67
10. Violette, A., Leonardi, A., Piquemal, D., Terrat, Y., Biass, D., Dutertre, S., Noguier, F., Ducancel, F., Stocklin, R., Krizaj, I., and Favreau, P. (2012) Recruitment of glycosyl hydrolase proteins in a cone snail venomous arsenal: further insights into biomolecular features of *Conus* venoms.

- Marine Drugs* **10**, 258–280
11. Woodward, S. R., Cruz, L. J., Olivera, B. M., and Hillyard, D. R. (1990) Constant and hypervariable regions in conotoxin propeptides. *EMBO J.* **9**, 1015–1020
 12. Kaas, Q., Westermann, J. C., Halai, R., Wang, C. K., and Craik, D. J. (2008) ConoServer, a database for conopeptide sequences and structures. *Bioinformatics* **24**, 445–446
 13. Kaas, Q., Yu, R., Jin, A. H., Dutertre, S., and Craik, D. J. (2012) ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Res.* **40**, D325–D330
 14. Puillandre, N., Koua, D., Favreau, P., Olivera, B. M., and Stocklin, R. (2012) Molecular phylogeny, classification and evolution of conopeptides. *J. Mol. Evol.* **74**, 297–309
 15. Milne, T. J., Abbenante, G., Tyndall, J. D., Halliday, J., and Lewis, R. J. (2003) Isolation and characterization of a cone snail protease with homology to CRISP proteins of the pathogenesis-related protein superfamily. *J. Biol. Chem.* **278**, 31105–31110
 16. Olivera, B. M., Rivier, J., Scott, J. K., Hillyard, D. R., and Cruz, L. J. (1991) Conotoxins. *J. Biol. Chem.* **266**, 22067–22070
 17. Kaas, Q., Westermann, J. C., and Craik, D. J. (2010) Conopeptide characterization and classifications: an analysis using ConoServer. *Toxicon* **55**, 1491–1509
 18. Luo, S., Christensen, S., Zhangsun, D., Wu, Y., Hu, Y., Zhu, X., Chhabra, S., Norton, R. S., and McIntosh, J. M. (2013) A novel inhibitor of alpha9alpha10 nicotinic acetylcholine receptors from *Conus vexillum* delineates a new conotoxin superfamily. *PLoS One* **8**, e54648
 19. Dutertre, S., Jin, A. H., Kaas, Q., Jones, A., Alewood, P. F., and Lewis, R. J. (2013) Deep venomomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol. Cell. Proteomics* **12**, 312–329
 20. Hu, H., Bandyopadhyay, P. K., Olivera, B. M., and Yandell, M. (2011) Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. *BMC Genomics* **12**, 60
 21. Hu, H., Bandyopadhyay, P. K., Olivera, B. M., and 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
 22. Lluisma, A. O., Milash, B. A., Moore, B., Olivera, B. M., and Bandyopadhyay, P. K. (2012) Novel venom peptides from the cone snail *Conus pulicarius* discovered through next-generation sequencing of its venom duct transcriptome. *Mar. Genomics* **5**, 43–51
 23. Terrat, Y., Biass, D., Dutertre, S., Favreau, P., Remm, M., Stocklin, R., Piquemal, D., and Ducancel, F. (2012) High-resolution picture of a venom gland transcriptome: case study with the marine snail *Conus consors*. *Toxicon* **59**, 34–46
 24. Davis, J., Jones, A., and Lewis, R. J. (2009) Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. *Peptides* **30**, 1222–1227
 25. Chang, D., and Duda, T. F., Jr. (2012) Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. *Mol. Biol. Evol.* **29**, 2019–2029
 26. Loughnan, M. L., Nicke, A., Lawrence, N., and Lewis, R. J. (2009) Novel alpha D-conopeptides and their precursors identified by cDNA cloning define the D-conotoxin superfamily. *Biochemistry* **48**, 3717–3729
 27. Luo, S., Zhangsun, D., Feng, J., Wu, Y., Zhu, X., and Hu, Y. (2007) Diversity of the O-superfamily conotoxins from *Conus miles*. *J. Pept. Sci.* **13**, 44–53
 28. Mondal, S., Babu, R. M., Bhavna, R., and Ramakumar, S. (2006) I-conotoxin superfamily revisited. *J. Pept. Sci.* **12**, 679–685
 29. Tawfik, D. S. (2010) Messy biology and the origins of evolutionary innovations. *Nat. Chem. Biol.* **6**, 692–696
 30. Dutertre, S., Biass, D., Stocklin, R., and Favreau, P. (2010) Dramatic intra-specimen variations within the injected venom of *Conus consors*: an unsuspected contribution to venom diversity. *Toxicon* **55**, 1453–1462
 31. Jakubowski, J. A., Kelley, W. P., Sweedler, J. V., Gilly, W. F., and Schulz, J. R. (2005) Intraspecific variation of venom injected by fish-hunting *Conus* snails. *J. Exp. Biol.* **208**, 2873–2883
 32. Hale, J. E., Butler, J. P., Gelfanova, V., You, J.-S., and Knierman, M. D. (2004) A simplified procedure for the reduction and alkylation of cysteine residues in proteins prior to proteolytic digestion and mass spectral analysis. *Anal. Biochem.* **333**, 174–181
 33. Duckert, P., Brunak, S., and Blom, N. (2004) Prediction of proprotein convertase cleavage sites. *Protein Eng. Des. Sel.* **17**, 107–112
 34. Gilles, A., Meglec, E., Pech, N., Ferreira, S., Malausa, T., and Martin, J. F. (2011) Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* **12**, 245
 35. Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* **25**, 117–124
 36. Gry, M., Rimini, R., Stromberg, S., Asplund, A., Ponten, F., Uhlen, M., and Nilsson, P. (2009) Correlations between RNA and protein expression profiles in 23 human cell lines. *BMC Genomics* **10**, 365
 37. Steen, H., and Pandey, A. (2002) Proteomics goes quantitative: measuring protein abundance. *Trends Biotechnol.* **20**, 361–364
 38. Lewis, R. J., Dutertre, S., Vetter, I., and MacDonald, J. C. (2012) *Conus* venom peptide pharmacology. *Pharmacol. Rev.* **64**, 259–98
 39. Olivera, B. M., Walker, C., Cartier, G. E., Hooper, D., Santos, A. D., Schoenfeld, R., Shetty, R., Watkins, M., Bandyopadhyay, P., and Hillyard, D. R. (1999) Speciation of cone snails and interspecific hyperdivergence of their venom peptides. Potential evolutionary significance of introns. *Ann. N. Y. Acad. Sci.* **870**, 223–237
 40. Yuan, D. D., Han, Y. H., Wang, C. G., and Chi, C. W. (2007) From the identification of gene organization of alpha conotoxins to the cloning of novel toxins. *Toxicon* **49**, 1135–1149
 41. Tucker, J. K. and Tenorio, M. J. (2009) *Systematic Classification of Recent and Fossil Conoidean Gastropods*, ConchBooks, Hackenheim, Germany
 42. Ramilo, C. A., Zafaralla, G. C., Nadasdi, L., Hammerland, L. G., Yoshikami, D., Gray, W. R., Kristipati, R., Ramachandran, J., Miljanich, G., Olivera, B. M., and Cruz, L. J. (1992) Novel alpha- and omega-conotoxins from *Conus striatus* venom. *Biochemistry* **31**, 9919–9926
 43. Chen, P., Garrett, J. E., Watkins, M., and Olivera, B. M. (2008) Purification and characterization of a novel excitatory peptide from *Conus distans* venom that defines a novel gene superfamily of conotoxins. *Toxicon* **52**, 139–145
 44. Alewood, D., Birinyi-Strachan, L. C., Pallaghy, P. K., Norton, R. S., Nicholson, G. M., and Alewood, P. F. (2003) Synthesis and characterization of delta-atracotoxin-Ar1a, the lethal neurotoxin from venom of the Sydney funnel-web spider (*Atrax robustus*). *Biochemistry* **42**, 12933–12940
 45. Ye, M., Khoo, K. K., Xu, S., Zhou, M., Boonyalai, N., Perugini, M. A., Shao, X., Chi, C., Galea, C. A., Wang, C., and Norton, R. S. (2012) A helical conotoxin from *Conus imperialis* has a novel cysteine framework and defines a new superfamily. *J. Biol. Chem.* **287**, 14973–14983
 46. Rigby, A. C., Baleja, J. D., Li, L., Pedersen, L. G., Furie, B. C., and Furie, B. (1997) Role of gamma-carboxyglutamic acid in the calcium-induced structural transition of conantokin G, a conotoxin from the marine snail *Conus geographus*. *Biochemistry* **36**, 15677–15684
 47. White, H. S., McCabe, R. T., Armstrong, H., Donevan, S. D., Cruz, L. J., Abogadie, F. C., Torres, J., Rivier, J. E., Paarmann, I., Hollmann, M., and Olivera, B. M. (2000) In vitro and in vivo characterization of conantokin-R, a selective NMDA receptor antagonist isolated from the venom of the fish-hunting snail *Conus radiatus*. *J. Pharmacol. Exp. Ther.* **292**, 425–432
 48. Gowd, K. H., Twede, V., Watkins, M., Krishnan, K. S., Teichert, R. W., Bulaj, G., and Olivera, B. M. (2008) Conantokin-P, an unusual conantokin with a long disulfide loop. *Toxicon* **52**, 203–213
 49. Twede, V. D., Teichert, R. W., Walker, C. S., Gruszczynski, P., Kazmierkiewicz, R., Bulaj, G., and Olivera, B. M. (2009) Conantokin-Br from *Conus bretinghami* and selectivity determinants for the NR2D subunit of the NMDA receptor. *Biochemistry* **48**, 4063–4073