ARTICLE

Diversity and novelty of venom peptides from *Conus* (*Asprella*) *rolani* revealed by analysis of its venom duct transcriptome

Ryoichi S. Taguchi, Dan Jethro M. Masacupan, and Arturo O. Lluisma*

Marine Science Institute, College of Science, University of the Philippines, Diliman

ABSTRACT

Conus species in the sub-genus *Asprella* are poorly studied because they inhabit deep-water habitats. To date, only a few peptides have been characterized from this clade. In this study, the venom duct transcriptome of a member of this clade, *Conus rolani,* was mined for potential conopeptides. Using a high-throughput RNA sequencing platform (Illumina) and a multiple k-mer *de novo* assembly, we found 103 putative conopeptide precursor amino acid sequences, including the few peptides previously reported for this species. The sequences, predominantly novel based on amino acid sequence, were diverse, comprising 36 gene superfamilies (including the "unassigned" superfamilies). As observed in other *Conus* species, the O1 gene superfamily was the most diverse (12 distinct sequences) but interestingly none of the sequences were found to contain the conserved amino acids associated with certain bioactivities in peptides found in piscivorous *Conus* species. The O2 superfamily was also highly diverse but con-ikot-ikot and an unassigned superfamily (MMSRMG) were more diverse than the rest of the superfamilies. In terms of gene expression levels, the understudied MEFRR paralog of the ancestral divergent M---L-LTVA superfamily was found to be the most highly expressed in the transcriptome, suggesting a novel role. Additionally, a conopeptide with high sequence similarity to A_2 secretory group XII phospholipases is the first reported member of this phospholipase group in *Conus* and potentially represents a novel superfamily, expanding the catalog of known phospholipases present in cone snail venoms. The discovery of these putative conopeptides provides the first but early glimpse of the diversity and novelty of the peptides in the *Asprella* group and sets the stage for their functional characterization.

*Corresponding author
Email Address: aolluisma@up.edu.ph
Date received: December 13, 2023
Date revised: February 12, 2024
Date accepted: February 20, 2024
DOI: https://doi.org/10.54645/202417SupQCH-42

INTRODUCTION

Cone snails are the largest marine invertebrate genus, comprising over 800 extant species (Himaya and Lewis 2018). They use a wide array of venom peptides for defensive and prey-capture purposes (Dutertre *et al*. 2014). This dependence on envenomation strategies for survival spurs a chemical evolutionary arms race against their predators and/or prey and drives the diversification of venom components (Duda Jr. and Kohn 2005). This chemical diversity has been highlighted by multiple *Conus* venom transcriptomics studies that reveal typically 100 to 500 putatively expressed precursor peptide sequences, each precursor giving rise to around 10 post-translationally modified peptides by variable peptide processing (Peng *et al*. 2021). In this light, transcriptomics has been an indispensable tool in shedding insight into these snails' venom components, so much so that venom researchers describe the use of this technology as arguably the single best advance in venom peptide discovery of the last decade (Robinson *et al*. 2017).

Since these venom components target neurological and neuromuscular receptors with high selectivity (McIntosh and Jones 2001), they constitute a promising source of molecular probes and/or drugs. In 2004, the US Food and Drug

KEYWORDS

Conus, conopeptides, transcriptome, venom, toxinology

Administration approved ziconotide, a 25-amino acid natural product from the *Conus magus*. Along with its clinical application, the discovery of this peptide led to the identification of the voltage-gated calcium ion channel CaV 2.2 as a target for pain management (Miljanich 2014). Though ziconotide was discovered using an "activity-first" approach, recent technological advances in nucleotide sequencing have made "sequence-first" approaches facile and feasible (Safavi-Hemami *et al*. 2019; Robinson *et al*. 2017). Peptide drug leads from cone snails that reached clinical trials, such as Vc1.1 and RgIA, were discovered using this integrated approach (Sandall *et al*. 2003; Ellison *et al*. 2008).

The pharmacological diversity of the peptides found in the venom gland of the cone snails is suspected to be underestimated since most of the studies focused on species that belong to a few lineages (Puillandre *et al*. 2014). One poorly studied group is *Asprella* (**Figure 1**); members of this sub-genus are presumed as fish hunters that live offshore, typically at depths of 60 to 250m (Olivera *et al*. 2021). One study that investigated the conopeptides from this group described the potent NMDA-specific bioactivity of conantokins from *C. brettinghami*, *C. bocki*, and *C. rolani* (Twede *et al*. 2009; Platt *et al*. 2014; Gowd *et al*. 2012). Additionally, using a bioactivity-guided purification, a new class of conopeptide was discovered from the venom duct of *C. rolani*, the first evidence of somatostatin-like peptides in animal venoms (Ramiro *et al*. 2022).

To date, no transcriptome analysis is available for any member of this clade. To further investigate conopeptide diversity in *Asprella*, and to contribute to the cataloging of conopeptide diversity in general, a high-throughput transcriptome sequencing of the venom duct from *C. rolani* was carried out. Here we report 103 putative precursor venom peptide sequences from 36 gene superfamilies, of which ten are unassigned superfamilies, and two are proposed new superfamilies. We also report a putative Group XII Secretory Phospholipase A_2 enzyme annotated from the transcriptome assembly.

METHODOLOGY

Sample collection, RNA extraction, and sequencing

A *C. rolani* specimen collected from Sogod, Cebu, Philippines using tangle nets at depths of around 150 to 300m was selected for this study. Before dissection, a live specimen was acclimatized for 24 hours in an aerated (improvised) aquarium filled with seawater. The snail was dissected directly on ice, and its venom duct was carefully obtained and stored in 1 µL of RNAlater® (Invitrogen) at 4°C before long-term storage at -80°C. Total RNA was isolated from the venom duct using the Trizol extraction method. Briefly, the thawed venom duct was homogenized using 2.0 mm of ZR bashing beads (ZYMO research) in a bead beater (Precellys, Berlin Technologies) with 1 ml TRIzol reagent (Sigma-Aldrich TRI reagent®). The total RNA was extracted from the homogenate using chloroform and precipitated using isopropanol. The pellet was washed with 75% EtOH and air-dried. The pelleted RNA was resuspended in 50 µL of RNAase-free water, and clean-up was performed using the Qiagen purification kit following the manufacturer's protocol. The quality and quantity of total RNA were evaluated using the Agilent TapeStation and Qubit, respectively. One library was constructed using the TruSeq Stranded mRNA Library prep kit and the sequencing was performed in Illumina NextSeq550 using Mid Output v2.5 with 300 cycles (paired ends, 2 x 150 base pairs). All library preparations and sequencing were done at the Philippine Genome Center, University of the Philippines.

De novo Transcriptome Assembly

Random sequencing errors in the raw sequencing reads were corrected using Rcorrector (Song and Florea 2015). The reads were then filtered, and the adapters were trimmed using the default settings of Fastp (Chen *et al*. 2018), which resulted in 6,213,532 reads for forward and reverse reads. To rid of unwanted ribosomal RNA sequences, reads with identical sequences in the SILVA database (Quast *et al*. 2013) were removed from the working data using Bowtie2 (Langmead *et al*. 2012).

The filtered reads were then *de novo* assembled using TransAbyss v. 2.0.1 (Robertson *et al*. 2010) using four k-mer sizes (31, 49, 55, 77), based on the predicted optimum k-mer size using rnaSPAdes (Bushmanova *et al*. 2019). The assembled sequences were compiled using a merge package from TransAbyss 2.0.1. To assess the alignment rates of the individual and merged assemblies, filtered reads were mapped to the assembly using the Bowtie 2 read aligner (Langmead *et al*. 2012).

To calculate gene transcription values in the assembly, the obtained alignments were used as input to RNA-Seq by Expectation Maximization (RSEM) (Li and Dewey 2011) to estimate transcript abundance in terms of transcripts per million (TPM).

Species identification of *Conus* sample

To confirm the species identity of the sample, the assembly was used as a query for the *Conus* molecular markers database from the Barcode of Life Database (BOLD) (Ratnasingham and Hebert 2007). BlastN was used to match the transcripts to the marker database, with a 97% percent identity and a 50% High Scoring Pair (HSP) query coverage cut-off (Boratyn *et al*. 2012). The COI gene sequence match from the transcriptome assembly was used to construct a phylogenetic tree altogether with COI sequences from the *Conus (Asprella)* clade (Olivera *et al*. 2021), conducted in MEGA X (Kumar *et al*. 2018). ClustalW (Thompson *et al*. 1994) was used for multiple sequence alignment via Jalview (Waterhouse *et al*. 2009). The phylogenetic relationship of the sequences was inferred by using the Maximum Likelihood method and the Hasegawa-Kishino-Yano model (Hasegawa *et al*. 1985) with a discrete Gamma distribution (5 categories) (+*G*, parameter = 0.9598) and a rate variation model that allowed some sites to be evolutionarily invariable ([+*I*], 33.57% sites). The consensus tree was inferred from 1,000 bootstrap replicates (Felsenstein 1985).

Annotation of conopeptides

Nucleotide sequences were translated into peptide sequences with TransDecoder, considering six possible reading frames (https://github.com/TransDecoder/TransDecoder/releases).
Using a compiled conopeptide database (combined NCBI and Conoserver (Kaas *et al*. 2012), retrieved October 31, 2021), sequences similar to the translated sequences were then searched using BlastP (Boratyn *et al*. 2012), with a 1e-5 e-value cut-off. Only matched sequences with a recognizable stop codon, and with >1 TPM were considered as putative conopeptides (i.e., precursor sequences) and used for further analysis.

The putative conopeptide precursor sequences were then submitted to ConoPrec (Kaas *et al*. 2012) for prediction of their structural features. Matches with no detected Signal or Pre region were N-terminally trimmed based on the closest database match and re-submitted to ConoPrec. If more than one translation product from a nucleotide sequence were obtained from the set cut-offs, additional parameters such as proper partitioning result from ConoPrec, NCBI-BLAST match score, cysteine framework, and signal sequence detection using SignalP v.6.0 or TransP v.2.0 were taken into account to choose

only one translation product or if the sequence would be included as an annotated conopeptide. For conopeptide precursor sequences that correspond to two or more distinct transcripts (i.e., at the nucleotide level), the expression values (TPM) of the transcripts were summed.

Top hits from the Uniprot database, with search restricted to *Conus* clade and with set e-value: < 0.001 of the mature region sequences of the top 15 most highly expressed precursor peptides were compiled (**Table 2**). Only hits with bioassay data were reported.

Classification of Gene Superfamilies

The predicted conopeptide sequences were classified with the aid of ConoPrec, an online precursor analysis tool implemented in the ConoServer (Kaas *et al*. 2012), and/or the closest matches of signal sequences in the GenBank database (highest scoring conopeptide signal sequence hit with >90% sequence identity) identified using BlastP (Barghi *et al*. 2015).

Signal sequences that were not assigned by ConoPrec to a superfamily but have been previously reported were named using the first six letters of the signal sequence. This set of sequences would be referred to as unassigned superfamilies. These sequences were verified using SignalP 6.0 (Teufel *et al*. 2022) and TargetP (Mishra *et al*. 2014). DeepLoc (Thumuluri *et al*. 2022) was also used to assess the protein localization of some sequences. Data on the closest matches of these signal sequences from GenBank (sequences, taxonomic identity of source species, etc.) reported were then collated.

The predicted signal sequences that were >15 amino acid long were aligned using MAFFT (Katoh *et al*. 2019; Kuraku *et al*. 2013) with default setting to construct a maximum likelihood tree. Using MEGA X (Kumar *et al*. 2018), the tree was inferred from 1,000 bootstrap samples using a maximum likelihood model, Jones-Taylor-Thornton (Jones *et al*. 1992) matrix-based model, and a discrete gamma-distribution evolutionary rate. Branches corresponding to partitions reproduced in less than 50% of bootstrap replicates were collapsed. **Figures 3 & 5** were prepared using a Zappo color scheme to highlight the differences in the physicochemical features on the aligned amino acid residues.

RESULTS AND DISCUSSION

I. Transcriptome assembly

The multiple k-mer assembly workflow using four k-mer sizes (33, 49, 55, 77) with TransAbyss and TransAbyss merge package resulted in 94.30% of the 6,213,532 paired reads aligning concordantly. This alignment rate is higher compared to assemblies that tested the k-mer sizes individually (83.24% - 88.27% alignment rate). The highest alignment rate was used as a basis for selecting the assembly to be used later for further analysis (as described in Barghi *et al*. 2014).

From this assembly, a total of 103 putative precursor conopeptide sequences were identified, totaling 151,092.95 TPM, relative to the total transcripts of the venom duct.

II. Species confirmation of the *C. rolani* sample

Early studies on the identification of members of the *Asprella* clade encountered difficulties in resolving what appeared to be similar forms from diverse localities (Olivera *et al*. 2021). To confirm the species identity of the sample used in this study, the COI sequence was extracted from the transcriptome assembly and, along with other *Asprella* COI sequences, was used to construct a phylogenetic tree. This study's COI sequence labeled "Olango *C. rolani*" clustered strongly with the previously identified *C. rolani* sequences (**Figure 1**). This confirms the sample used for transcriptome analysis was *C. rolani.*

III. Conopeptide toxin precursor diversity in *C. rolani*

A total of 103 putative conopeptide sequences were annotated from the transcriptome assembly, assigned to 25 known superfamilies, 10 unassigned superfamilies but previously reported in the literature, and putatively one novel superfamily (a single precursor sequence but multiple reads that map to the predicted transcript). Because the identification of these putative conopeptides was based on sequence similarity, it is possible if not likely that there were novel gene-encoded venom components that could not have been detected by this method.

To investigate relationships among identified conopeptide precursor sequences, a phylogenetic tree based on the amino acid sequences of the signal sequence region was constructed (Figure 2.). The tree confirms that most of the annotated conopeptides could generally be assigned to canonical superfamilies based on cluster membership.

III.A O1 superfamily is most diverse.

Of the identified conopeptides, the O1 superfamily has the highest number (12) of distinct precursors in *C. rolani*. This is a common observation for cone snails; on the Conoserver database, this superfamily has the most precursor sequences reported (as of May 28, 2023). However, only the VI/VII cysteine framework was observed among the O1 sequences from *C. rolani* (**Figure 3**). In contrast, per records of the Conoserver database, multiple Cys frameworks are found within the O1 superfamily.

It was also observed that none of the *C. rolani* O1 sequences contain the critical residues at specific positions associated with certain bioactivities in conopeptides from other piscivorous species. For example, O1 superfamily ω-conotoxins from piscivorous cone snails are reported to have a tyrosine residue at position 13 (aligned residue 24 in **Figure 3**) that is critical for voltage-gated Ca^{2+} channel bioactivity (Lewis *et al*. 2012). This was not observed in the *C. rolani* O1 superfamily sequences, with this position predominantly substituted with positively charged residues (Ro9, Ro31, Ro36, Ro43, Ro54, Ro58, and Ro72), similar to GVIIA from piscivorous *Conus geographus* (Olivera *et al*. 1985), and to MoVIA and MoVIB from vermivorous *Conus moncuri* (Sousa *et al*. 2018). Another known critical residue for bioactivity among fish-hunting cone snails is the lysine residue at position 2 (aligned residue 11 in **Figure 3**) (Lewis *et al*. 2012); in *C. rolani* O1 peptides, this position is mostly occupied by aliphatic/hydrophobic residues.

Also found to be relatively more diverse are the A, con-ikot-ikot, and the MMSRMG superfamilies, each of which has 6 distinct precursors observed in the transcriptome.

III.B Coninsulins

A previous study reported that cone snails produce venom peptides that are structurally and functionally similar to insulin; these peptides induce glycemic shock in prey (Safavi-Hemami et al. 2019). Con-Ins G1 is a minimized version of human insulin that could bind and activate the human insulin receptor (Menting et al. 2016). It lacks a region equivalent to the canonical eight-residue C-terminal segment of the human insulin B chain. In the current study, two putative homologs were found in the *C. rolani* transcriptome (Figure 4) and one of these (Ro23_insulin) has a B-chain region that is even shorter. An interesting feature is the longer A chain region found in the *C. rolani* sequences (Figure 4), which is also reported to be found in homologs in the Phasmaconus clade and which could be compensatory for amino acid deletions in the B chain to improve its activity on insulin receptors (Xiong et al. 2022).

III.C Unassigned superfamilies

Thirty-six conopeptide superfamilies were identified in the transcriptome assembly on the basis of distinct signal sequences. Majority of these are established and named superfamilies. However, putative conopeptides were found in the transcriptome that belong to 10 unassigned superfamilies, i.e., superfamilies previously observed in the transcriptomes of other *Conus* species but still lack peptide-level validation. The signal peptide region of the sequences inferred using SignalP (ver. 6.0) or TargetP was predicted with high probability values (**Table 1**). These observations provide confirmation of the existence of these superfamilies and of their wider distribution in *Conus*. Of note is the putative superfamily Ro94_MVREVT which potentially represents a novel conopeptide superfamily, as discussed in the next section.

III.D Secretory phospholipases

Animal venoms are known to contain a wide range of secreted phospholipase A2 (sPLA2) enzymes. In the current study, three putative sequences were found to have significant sequence similarity to sPLA2 phospholipases. Two of these, Ro102_conodipine and Ro103_conodipine, could be identified as homologs of previously reported *Conus* conodipines, which belong to the sPLA2 group IX of phospholipases (Möller et al. 2019), increasing the taxonomic distribution of these enzymes in the genus. However, the transcripts of these peptides were not highly expressed (with expression values 7.1 and 2.87 TPM, respectively).

The third sequence exhibiting sequence similarity to sPLA2 enzymes is Ro94_MVREV (Figure 5). A BLASTP search in the GenBank database using the whole precursor sequence as query showed that the top 100 closest hits to Ro94_MVREV are phospholipases that belong to another group, sPLA2 group XII (e-value: ≤ 2e-29), of which 29 are molluscan in origin (retrieved May 28, 2023). Interestingly, none of the hits are from cone snails. Upon limiting the database search to the genus *Conus*, no significant matches were found (e-value cutoff: ≤ 1e-4). The domain hit (sequences 41-177) has a very low domain hit e-value score (7.58e-40), indicating the significance of the protein annotation. Results of the analysis using the tool DeepLoc indicated that the protein is a secretory protein (88% probability), consistent with it being a possible venom component.

The signal peptide of the sequence Ro94_MVREV (TPM value: 2.82) is also different from the other two *C. rolani* phospholipases (conodipines). Hence, it potentially represents a novel superfamily (**Figure 2**) and hence is the first report for *Conus*. The signal peptide was predicted with high probability values by both SignalP and TargetP (99.9% probability on both tools) and thus appears to be a valid eukaryotic signal peptide.

However, the function of this protein remains to be established. Among the group XII A_2 phospholipases, only the human group XII $sPLA_2$ enzyme has been previously studied. Constitutive expression of this gene in HEK293 cells exhibited no detectable phospholipase activity *in vitro*, nor in arachidonic acid release and prostaglandin production. Interestingly, transfected cells exhibited abnormal morphology, suggesting an unknown functional aspect of the enzyme (Murakami *et al*. 2003). The $sPLA_2$ catalytic site with a known *DXCCXXHD* motif (Dennis *et al*. 2011) is well conserved for all presented sequences, except for the first aspartic acid residue (Residue no. 120-127, **Figure 5**).

Expression profile of conopeptides in *C. rolani*

The relative level of expression of the predicted peptides in the transcriptome is shown in **Figure 6**. The most highly expressed peptides (> 4 log10 TPM) are an A superfamily peptide and two peptides that belong to the so-called Divergent superfamily M---L-LTVA (a superfamily identified in early diverging clades in the *Conus* lineage). Other highly expressed peptides (expression level close to 4 log10 TPM) include a B1 superfamily (Conantokin) peptide and a C superfamily peptide. Collectively, these most highly expressed peptides constitute more than fifty percent of the TPM attributed to conopeptides in the transcriptome.

IV.A Superfamily paralog identified from the early divergent clade is the most highly expressed.

Interestingly, a putative homolog of the early divergent superfamily M---L-LTVA is the most highly expressed peptide in the transcriptome (**Figure 6 B**). The signal sequences in the *C. rolani* transcriptome start with MEFRR, which resembles a sequence previously reported from *C. ebraeus* and *C. judaeus* transcriptomes as SF-mi2 paralog 02 and/or Cerm11 paralog 1. SF-mi2-02 is reported to be highly expressed in *C. ebraeus* (>20,000 TPM), and moderately expressed in *C. judaeus* (>3,000 TPM) (Pardos-Blas *et al*. 2022). This MEFRR motif was also reported in *C. betulinus* (Peng *et al*. 2016) and *C. magus* (Pardos-Blas *et al*. 2019) transcriptomes. No function nor protein evidence of the MEFRR motif M---L-LTVA superfamily has yet to be described. However, using SignalP, this motif was verified as a signal sequence (99.9% probability). Additionally, subcellular localization by Ro1_M---L-LTVA was highly predicted by DeepLoc 2 to be extracellular (95.12% probability), which is consistent with the expectation that the peptide is secreted as a venom component.

IV.B Only ancestral loop patterns were observed in A-superfamily conopeptides.

The A superfamily, consisting of 6 identified conopeptide precursors in the transcriptome, is the second highly expressed superfamily as most of the peptides are moderately to highly expressed. Among these, only the cysteine framework I (4/7 loop size) was observed (Figure 7). This loop pattern is predicted to be the ancestral form of the gene superfamily as this motif is highly distributed across cone snail taxa irrespective of prey preferences (Santos et al. 2004).

IV.C Conantokin sequences previously reported from *C. rolani* were not observed in the transcriptome.

Four putative peptide sequences (Ro5_B1, Ro16_B1, Ro20_B1, and Ro78_B1) were identified as superfamily B1 (Conantokin) peptides in this study. Ro5_B1 is the fifth most highly expressed conopeptide (7150.97 TPM). The other B1 peptides have lower expression (Ro16_B1, 2505.1 TPM; Ro20_B1, 2065.06 TPM) or very low expression (Ro78_B1, 9.21 TPM). Ro16_B1 has the same mature sequence as Ro5_B1 but differ in the N-terminal pro-region.

Interestingly, none of these have identical sequences to three B1 peptides previously found in *C. rolani* (Gowd *et al*. 2012) which the authors isolated using a PCR-capture method. It was observed, however, that the mature sequence of ConRI-C (Gowd *et al*. 2012) only has two amino acid residue differences in the mature region compared to Ro5_B1: substitution of the fifth amino acid (V5L), and deletion of the C-terminal asparagine (**Figure 8**). The V5L substitution is important as such a substitution in the sequence of conantokin-R from *C. radiatus* significantly diminished the time of inhibition effect from >120 seconds (native sequence) to 45 seconds (conantokin-R V5L) in NR1A/NR2B NMDA receptors (Klein *et al*. 2001). However, the reduction of maximum current evoked was not significantly different.

IV.D Previously identified Consomatins from *C. rolani* are highly expressed

Four C superfamily (hence putative Consomatin) peptides were found in the *C. rolani* transcriptome (**Figure 9**). One of these,

Ro4_C (whole precursor sequence), completely matched in sequence a peptide isolated from the venom gland of *C. rolani*, reported as Consomatin Ro1 (Ramiro *et al*. 2022) and the first to be characterized somatostatin-like Conoidean peptide. Ro4_C is the fourth most highly expressed conopeptides (**Figure 9, Table 2**). Another peptide, Ro19_C, has a sequence that is identical to Consomatin Ro2, a previously reported consomatin mRNA sequence acquired through database mining (Koch *et al*. 2022).

IV.E The highly expressed conopeptides: pharmacological diversity and novelty

The 15 most highly expressed conopeptide transcripts in the *C. rolani* venom duct transcriptome are shown in **Table 2**. The peptides belong to diverse superfamilies. These highly expressed peptides thus likely have diverse pharmacological targets. To obtain insights into their potential pharmacological properties, their sequences were used as query to search for their closest hits (e-value cut-off: < 0.001) in the Uniprot database, with the search restricted to sequences from *Conus*. Only matches with bioassay data are shown (**Table 2**).

Eight out of the fifteen highly expressed peptides (Ro1_Divergent_M---L-LTVA, Ro2_Divergent_M---L-LTVA, Ro3_A, Ro6_MKMYL-, Ro10_T, Ro12, Ro14_MSRLFL, and Ro15_B2) have no close match in the Uniprot database with bioassay data, given the cut-off value. Hence, the molecular targets of these peptides cannot be determined based on sequence similarity. The two most highly expressed precursors belong to the divergent M---L-LTVA superfamily. No peptide belonging to this superfamily has yet been pharmacologically characterized. Ro3_A, classified as belonging to the well-characterized A-superfamily conopeptide, is expected to target nAChRs given its 4/7 motif (Lebbe *et al*. 2014), but no sequence is currently similar enough to support this claim. Similarly, Ro10_T, Ro12_MSRLFL, Ro14_I4, and Ro15_B2 could be assigned to superfamilies based on their conserved signal sequences, but they have no close hits with respect to their mature region.

In the case of Ro11_O2, the closest match is R7a. This is reported to have multiple brominated tryptophan residues, affecting mice with light sleep when injected intracranially. No molecular target has yet been identified for R7a (Jimenez *et al*. 2004).

For the other peptides with matches in the database (Ro4_C, Ro5_B1, Ro7_02, Ro8_M, Ro13_A), it might be possible to infer their molecular targets from those of their closest hits (at least at the receptor type level). Based on the search results, the targets of the *C. rolani* venom components could include somatostatin receptors, NMDA receptors, voltage-gated calcium ion channels (CaV 1.2, & CaV 2.2), tetrodotoxin-sensitive voltage-gated sodium ion channels (Ttx-sensitive NaVs), and nicotinic acetylcholine receptors (nAChRs) (**Table 2**).

Collectively, the transcripts of these 15 sequences represent 75% of the total TPM that could be attributed to putative conopeptides. Further investigating the venom repertoire of *C. rolani* may reveal new molecular targets of these peptides, as most of these sequences do not have homologs with reported bioactivity.

CONCLUSION

Analysis of the venom duct transcriptome of *C. rolani* revealed a significant amount of not only diversity but also structural novelty of putative conopeptides in this species notwithstanding the significant amount of data already collected from various *Conus* species. The potential novelty of the pharmacological properties of these peptides is clearly of interest. In addition, potential functional differences in the role played by the different superfamilies in *Conus* venoms are apparent from the observation that the level of expression of the peptides at the superfamily level in *C. rolani* differs from those of other *Conus* species. And because this study represents the first transcriptomics-enabled venomics study of a species from the *Asprella* group, it also highlights the greater level of diversity and novelty of conopeptides that await further discovery in this group.

ACKNOWLEDGMENT

This work was funded by the Department of Science and Technology- Philippine Council for Health Research and Development (DOST-PCHRD). The collection of *C. rolani* samples was made possible under Gratuitous Permit No. 0252-23 granted by the Department of Agriculture — Bureau of Fisheries and Aquatic Resources, Philippines (DA-BFAR). The fieldwork was done with the help of Olango island fishermen, led by Antonio Mosqueda.
We extend our gratitude to Antonio Catalig, Zae-Zae Aguinaldo, Kreighton Cadorna, Jonathan Wong, and Niño Dan Posadas for troubleshooting and assisting in the generation of the figures.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

CONTRIBUTIONS OF INDIVIDUAL AUTHORS

D.J.M. conducted the experiments, R.S.T. analyzed the results, and R.S.T., D.J.M., and A.O.L. conceived the experiments. All authors wrote and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

REFERENCES

Barghi N, Concepcion GP, Olivera BM, Lluisma AO. HIGH conopeptide diversity in Conus tribblei revealed through analysis of venom duct transcriptome using two high-throughput sequencing platforms. Mar Biotechnol 2015; 17(1): 81–98. https://doi.org/10.1007/s10126-014-9595-7

Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. DOMAIN enhanced lookup time accelerated BLAST. Biol Direct 2012; 7(12). http://doi.org/10.1186/1745-6150-7-12. PMID: 22510480; PMCID: PMC3438057.

Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. rnaSPAdes: A *de novo* transcriptome assembler and its application to RNA-Seq data. GigaScience 2019; 8(9): 1–13. https://doi.org/10.1093/gigascience/giz100.

Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinform 2018; 34(17): i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Dennis EA, Cao J, Hsu YH, Magrioti V, Kokotos G. PHOSPHOLIPASE A2 enzymes: physical structure, biological function, disease implication, chemical inhibition, and therapeutic intervention. Chem Rev 2011; 111(10): 6130–6185. https://doi.org/10.1021/cr200085w.

Duda Jr. TF, Kohn AJ. SPECIES-level phylogeography and evolutionary history of the hyperdiverse marine gastropod genus Conus. Mol Phylogenet Evol 2005; 34(2): 257–272. https://doi.org/10.1016/j.ympev.2004.09.012.

Dutertre S, Jin AH, Vetter I, Hamilton B, Sunagar K, Lavergne V, Dutertre V, Fry BG, Antunes A, Venter DJ, Alewood PF. EVOLUTION of separate predation-and defence-evoked venoms in carnivorous cone snails. Nature communications. 2014 Mar 24;5(1):3521.

El Hamdaoui Y, Wu X, Clark RJ, Giribaldi J, Anangi R, Craik DJ, King GF, Dutertre S, Kaas Q, Herzig V, and Nicke A. PERIPLASMIC Expression of 4/7 α-Conotoxin TxIA Analogs in E. coli Favors Ribbon Isomer Formation - Suggestion of a Binding Mode at the α7 nAChR. Front. Pharmacol 2019; 10(577). https://doi.org/10.3389/fphar.2019.00577.

Ellison M, Feng ZP, Park AJ, Zhang X, Olivera BM, McIntosh JM, Norton RS. Alpha-RgIA, a novel conotoxin that blocks the alpha9alpha10 nAChR: structure and identification of key receptor-binding residues. J Mol Biol 2008; 377(4): 1216–1227. https://doi.org/10.1016/j.jmb.2008.01.082.

Felsenstein J. CONFIDENCE limits on phylogenies: An approach using the bootstrap. Evol 1985; 39: 783-791.

Gowd KH, Han TS, Twede V, Gajewiak J, Smith MD, Watkins M, Platt RJ, Toledo G, White HS, Olivera BM, Bulaj G. CONANTOKINS derived from the Asprella clade impart conRl-B, an N-methyl d-aspartate receptor antagonist with a unique selectivity profile for NR2B subunits. Biochem 2012; 51(23): 4685–4692. https://doi.org/10.1021/bi300055n.

Hansson K, Ma X, Eliasson L, Czerwiec E, Furie B, Furie BC, Rorsman P, Stenflo J. THE first gamma-carboxyglutamic acid-containing contryphan. A selective L-type calcium ion channel blocker isolated from the venom of Conus marmoreus. J Biol Chem 2004; 279(31): 32453–32463. https://doi.org/10.1074/jbc.M313825200.

Hasan MM, Starobova H, Mueller A, Vetter I, Lewis RJ. SUBCUTANEOUS ω-Conotoxins Alleviate Mechanical Pain in Rodent Models of Acute Peripheral Neuropathy. Mar Drugs 2021; 19(2): 106. https://doi.org/10.3390/md19020106.

Hasegawa M, Kishino H, Yano T. DATING the human-ape split by a molecular clock of mitochondrial DNA. J Mol Evol. 1985; 22: 160-174.

Himaya S, Lewis RJ. VENOMICS-Accelerated Cone Snail Venom Peptide Discovery. Int J Mol Sci 2018; 19(3): 788. https://doi.org/10.3390/ijms19030788.

Jimenez EC, Watkins M, Olivera BM. MULTIPLE 6-bromotryptophan residues in a sleep-inducing peptide. Biochem. 2004; 43(38): 12343–12348. https://doi.org/10.1021/bi0489412.

Jones DT, Taylor WR, Thornton JM. THE rapid generation of mutation data matrices from protein sequences. Computer applications in the biosciences: CABIOS 1992; 8(3): 275–282. https://doi.org/10.1093/bioinformatics/8.3.275.

Kaas Q, Yu R, Jin AH, Dutertre S, Craik DJ. ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. Nucleic Acids Res 2012; 40(Database issue): D325-30.

Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform 2019; 20(4): 1160-1166. https://doi.org/ 10.1093/bib/bbx108.

Klein RC, Prorok M, Galdzicki Z, Castellino FJ. THE amino acid residue at sequence position 5 in the conantokin peptides partially governs subunit-selective antagonism of recombinant N-methyl-D-aspartate receptors. J Biol Chem 2001; 276(29): 26860–26867. https://doi.org/10.1074/jbc.M102428200.

Koch TL, Ramiro IBL, Salcedo PF, Engholm E, Jensen KJ, Chase K, Olivera BM, Bjørn-Yoshimoto WE, Safavi-Hemami H. Reconstructing the origins of the somatostatin and allatostatin-C signaling systems using the accelerated evolution of biodiverse cone snail toxins. Mol Biol Evol 2022; 39(4): msac075. https://doi.org/10.1093/molbev/msac075.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 2018; 35(6): 1547-1549. https://doi.org/10.1093/molbev/msy096.

Kuraku S, Zmasek CM, Nishimura O, Katoh K. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res. 2013; 41(W1): W22-W28. https://doi.org/10.1093/nar/gkt389.

Langmead B, Salzberg S. FAST gapped-read alignment with Bowtie 2. Nat Methods 2012; 9: 357–359. https://doi.org/10.1038/nmeth.1923.

Lebbe EK, Peigneur S, Wijesekara I, Tytgat J. CONOTOXINS targeting nicotinic acetylcholine receptors: an overview. Mar Drugs 2014; 12(5): 2970–3004. https://doi.org/10.3390/md12052970.

Lewis RJ, Dutertre S, Vetter I, Christie MJ. CONUS venom peptide pharmacology. Pharmacol. Rev. 2012; 64(2): 259–298. https://doi.org/10.1124/pr.111.005322.

Lewis RJ, Schroeder CI, Ekberg J, Nielsen KJ, Loughnan M, Thomas L, Adams DA, Drinkwater R, Adams DJ, Alewood PF. ISOLATION and structure-activity of mu-conotoxin TIIIA, a potent inhibitor of tetrodotoxin-sensitive voltage-gated sodium channels. Mol Pharmacol 2007; 71(3): 676–685. https://doi.org/10.1124/mol.106.028225.

Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform. 2011; 12(323). https://doi.org/10.1186/1471-2105-12-323.

Livingstone CD, Barton GJ. PROTEIN Sequence Alignments: A Strategy for the Hierarchical Analysis of Residue Conservation. CABIOS 1993; 9(6): 745-756. https://doi.org/10.1093/bioinformatics/9.6.745

McIntosh JM, Jones RM. CONE venom–from accidental stings to deliberate injection. Toxicon 2001; 39(10): 1447–1451. https://doi.org/10.1016/s0041-0101(01)00145-3.

Menting JG, Gajewiak J, MacRaild CA, Chou DH, Disotuar MM, Smith NA, Miller C, Erchegyi J, Rivier JE, Olivera BM, Forbes BE, Smith BJ, Norton RS, Safavi-Hemami H, Lawrence MC. A minimized human insulin-receptor-binding motif revealed in a Conus geographus venom insulin. Nat Struct Mol Biol 2016; 23(10): 916–920. https://doi.org/10.1038/nsmb.3292.

Miljanich GP. ZICONOTIDE: neuronal calcium channel blocker for treating severe chronic pain. Curr. Med. Chem. 2004; 11(23): 3029–3040. https://doi.org/10.2174/0929867043363884.

Mishra NK, Chang J, Zhao PX. PREDICTION of Membrane Transport Proteins and Their Substrate Specificities Using Primary Sequence Information. PLoS One 2014; 9(6): e100278. https://doi.org/10.1371/journal.pone.0100278.

Möller C, Davis WC, Clark E, DeCaprio A, Marí F. Conodipine-P1-3, the First Phospholipases A_2 Characterized from Injected Cone Snail Venom. Mol Cell Proteomics: MCP 2019; 18(5): 876–891. https://doi.org/10.1074/mcp.RA118.000972.

Murakami M, Masuda S, Shimbara S, Bezzine S, Lazdunski M, Lambeau G, Gelb MH, Matsukura S, Kokubu F, Adachi M, Kudo I. CELLULAR arachidonate-releasing function of novel classes of secretory phospholipase A2s (groups III and XII). J Biol Chem 2003; 278(12): 10657–10667. https://doi.org/10.1074/jbc.M211325200.

Olivera BM, Watkins M, Puillandre N, Tenorio MJ. HIDDEN diversity in the Asprella clade: description of Conus (Asprella) neocostatus sp. nov. (Gastropoda, Conidae). Xenophora Taxonomy 2021; 33: 22–29.

Olivera BM, Gray WR, Zeikus R, McIntosh JM, Varga J, Rivier J, de Santos V, Cruz LJ. PEPTIDE neurotoxins from fish-hunting cone snails. Science 1985; 230(4732): 1338–1343. https://doi.org/10.1126/science.4071055.

Pardos-Blas JR, Irisarri I, Abalde S, Tenorio MJ, Zardoya R. CONOTOXIN Diversity in the Venom Gland Transcriptome of the Magician's Cone, Pionoconus magus. Mar Drugs 2019; 17(10) 553. https://doi.org/10.3390/md17100553.

Pardos-Blas JR, Tenorio MJ, Galindo J, Zardoya R. COMPARATIVE Venomics of the Cryptic Cone Snail Species Virroconus ebraeus and Virroconus judaeus. Mar Drugs 2022; 20(2): 149. https://doi.org/10.3390/md20020149.

Peng C, Huang Y, Bian C, Li J, Liu J, Zhang K, You X, Lin Z, He Y, Chen J, Lv Y, Ruan Z, Zhang X, Yi Y, Li Y, Lin X, Gu R, Xu J, Yang J, … Shi Q. THE first Conus genome assembly reveals a primary genetic central dogma of conopeptides in C. betulinus. Cell Discov 2021; 7(1). https://doi.org/10.1038/s41421-021-00244-7.

Peng C, Yao G, Gao BM, Fan CX, Bian C, Wang J, Cao Y, Wen B, Zhu Y, Ruan Z, Zhao X, You X, Bai J, Li J, Lin Z, Zou S, Zhang X, Qiu Y, Chen J, Coon SL, … Shi Q. HIGH-throughput identification of novel conotoxins from the Chinese tubular cone snail (Conus betulinus) by multi-transcriptome sequencing. GigaScience 2016; 5(17). https://doi.org/10.1186/s13742-016-0122-9.

Platt RJ, Curtice KJ, Twede VD, Watkins M, Gruszczyński P, Bulaj G, Horvath MP, Olivera BM. FROM molecular phylogeny towards differentiating pharmacology for NMDA receptor subtypes. Toxicon 2014; 81: 67–79. https://doi.org/10.1016/j.toxicon.2014.01.016.

Puillandre N, Bouchet P, Duda Jr. TF, Kauferstein S, Kohn AJ, Olivera BM, Watkins M, Meyer C. MOLECULAR phylogeny and evolution of the cone snails (Gastropoda, Conoidea). Mol Phylogenet Evol 2014; 78: 290–303. https://doi.org/10.1016/j.ympev.2014.05.023.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. THE SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 2013; 41(Database issue): D590–D596. https://doi.org/10.1093/nar/gks1219.

Ramiro I, Bjørn-Yoshimoto WE, Imperial JS, Gajewiak J, Salcedo PF, Watkins M, Taylor D, Resager W, Ueberheide B, Bräuner-Osborne H, Whitby FG, Hill CP, Martin LF, Patwardhan A, Concepcion GP, Olivera BM, Safavi-Hemami H. SOMATOSTATIN venom analogs evolved by fish-hunting cone snails: From prey capture behavior to identifying drug leads. Sci Adv 2022; 8(12): eabk1410. https://doi.org/10.1126/sciadv.abk1410.

Ratnasingham S, Hebert PD. BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Notes 2007; 7(3), 355-364. https://doi.org/10.1111/j.1471-8286.2007.01678.x. PMID: 18784790.

Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman S, Mungall K, Lee S, Okada HM, Qian JQ, Griffith M, Raymond A, Thiessen N, Cezard T, Butterfield YS, Newsome R, Chan SK, She R, Varhol R, … Birol I. *De novo* assembly and analysis of RNA-seq data. Nat Methods 2010; 7: 909–912. https://doi.org/10.1038/nmeth.1517.

Robinson SD, Undheim EAB, Ueberheide B, King GF. VENOM peptides as therapeutics: advances, challenges and the future of venom-peptide discovery. Expert Rev Proteom 2017; 14(10): 931-939. https://doi.org/10.1080/14789450.2017.1377613.

Safavi-Hemami H, Brogan SE, Olivera BM. PAIN therapeutics from cone snail venoms: From Ziconotide to novel non-opioid pathways. J Proteomics 2019; 190: 12–20. https://doi.org/10.1016/j.jprot.2018.05.009.

Sandall DW, Satkunanathan N, Keays DA, Polidano MA, Liping X, Pham V, Down JG, Khalil Z, Livett BG, Gayler KR. A novel alpha-conotoxin identified by gene sequencing is active in suppressing the vascular response to selective stimulation of sensory nerves in vivo. Biochem 2003; 42(22): 6904–6911. https://doi.org/10.1021/bi034043e.

Santos AD, McIntosh JM, Hillyard DR, Cruz LJ, Olivera BM. THE A-superfamily of conotoxins: structural and functional divergence. J Biol Chem 2004; 279(17): 17596–17606. https://doi.org/10.1074/jbc.M309654200.

Song L, Florea L. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. GigaScience 2015; 4(48). https://doi.org/10.1186/s13742-015-0089-y.

Sousa SR, McArthur JR, Brust A, Bhola RF, Rosengren KJ, Ragnarsson L, Dutertre S, Alewood PF, Christie MJ, Adams DJ, Vetter I, Lewis RJ. NOVEL analgesic ω-conotoxins from the vermivorous cone snail Conus moncuri provide new insights into the evolution of conopeptides. Sci Rep 2018; 8(1): 13397. https://doi.org/10.1038/s41598-018-31245-4.

Teufel F, Armenteros JJA, Johansen AR, Gislason MH, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol 2022; 40: 1023–1025. https://doi.org/10.1038/s41587-021-01156-3.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence

alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994; 22(22): 4673–4680. https://doi.org/10.1093/nar/22.22.4673.

Thumuluri V, Armenteros JJA, Johansen AR, Nielsen H, Winther O. DeepLoc 2.0: multi-label subcellular localization prediction using protein language models, Nucleic Acids Res 2022; 50(W1): W228–W234. https://doi.org/10.1093/nar/gkac278.

Twede VD, Teichert RW, Walker CS, Gruszczyński P, Kaźmierkiewicz R, Bulaj G, Olivera BM. Conantokin-Br from Conus brettinghami and selectivity determinants for the NR2D subunit of the NMDA receptor. Biochem 2009; 48(19): 4063–4073. https://doi.org/10.1021/bi802259a.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2 - A multiple sequence alignment editor and analysis workbench. Bioinform 2009; 25: 1189-1191. https://doi.org/:10.1093/bioinformatics/btp033.

Xiong X, Blakely A, Kim JH, Menting JG, Schäfer IB, Schubert HL, Agrawal R, Gutmann T, Delaine C, Zhang YW, Artik GO, Merriman A, Eckert D, Lawrence MC, Coskun Ü, Fisher SJ, Forbes BE, Safavi-Hemami H, Hill CP, Chou DH. SYMMETRIC and asymmetric receptor conformation continuum induced by a new insulin. Nat Chem Biol 2022; 18(5): 511–519. https://doi.org/10.1038/s41589-022-00981-0.

Yu C, Yu H, Li P. HIGHLIGHTS of animal venom research on the geographical variations of toxin components, toxicities and envenomation therapy. International journal of biological macromolecules. 2020 Dec 15;165:2994-3006.

LIST OF FIGURES

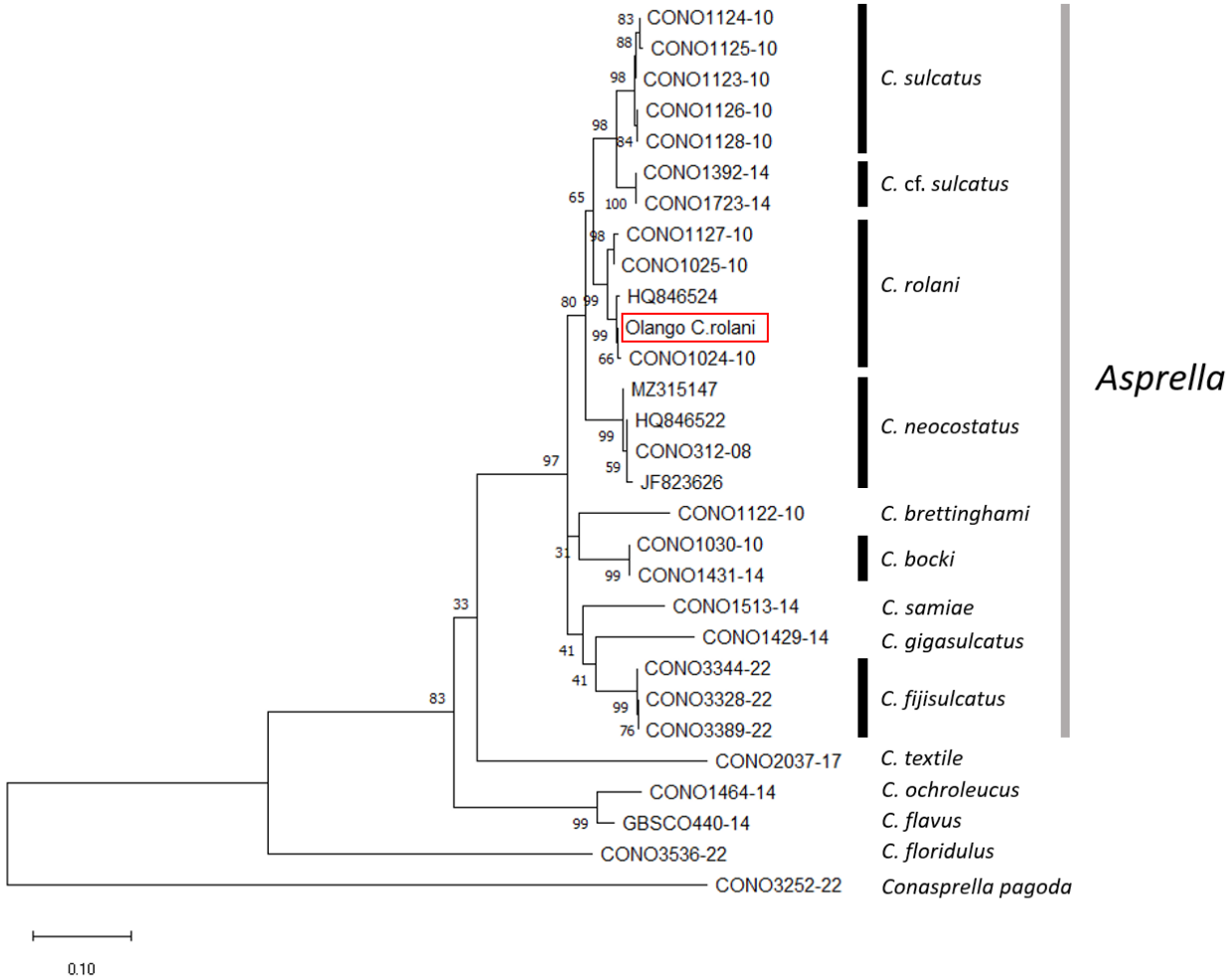

Figure 1: Phylogeny of selected *Conus* species in the subgenus *Asprella* and phylogenetic placement of the specimen used in this study. The maximum likelihood tree was constructed using a fragment of the *cox1* gene. The numbers shown next to the nodes are the bootstrap support values in % (from 1,000 bootstraps). "Olango *C. rolani*", boxed in red, was the sample used in this study.

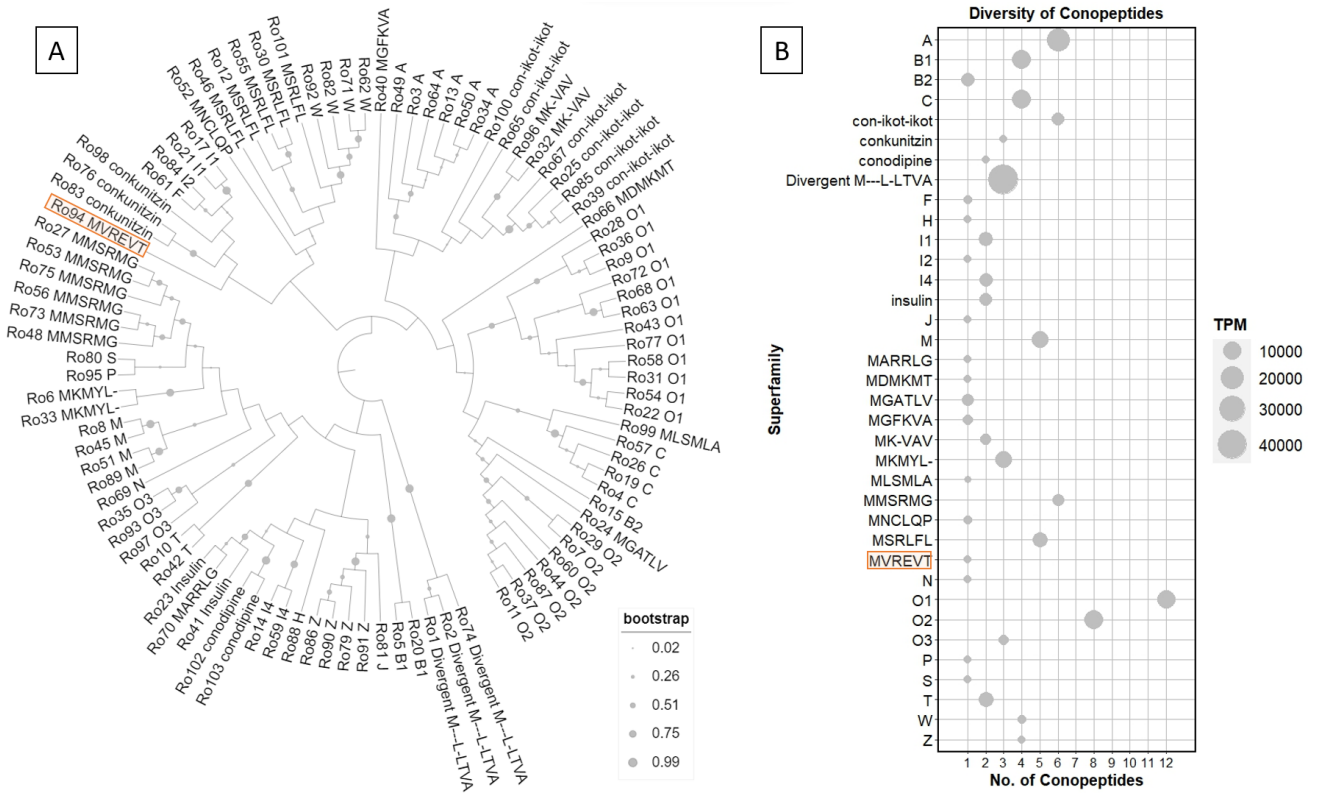

Figure 2: Putative conopeptides superfamilies found in *C. rolani*. A) A bootstrap consensus tree of the 98 out of the 103 signal sequences from identified conopeptides was derived using a Maximum Likelihood method, with statistical support from 1,000 pseudo-replicates. Only signal sequences with annotation and >15 amino acids long were used for the analysis. B) A bubble plot of known, unassigned, and a possibly novel superfamily (highlighted in orange) from *C. rolani*. The bubble size is proportional to the cumulative expression values (TPM) of each superfamily. In total, conopeptides were assigned into 36 superfamilies.

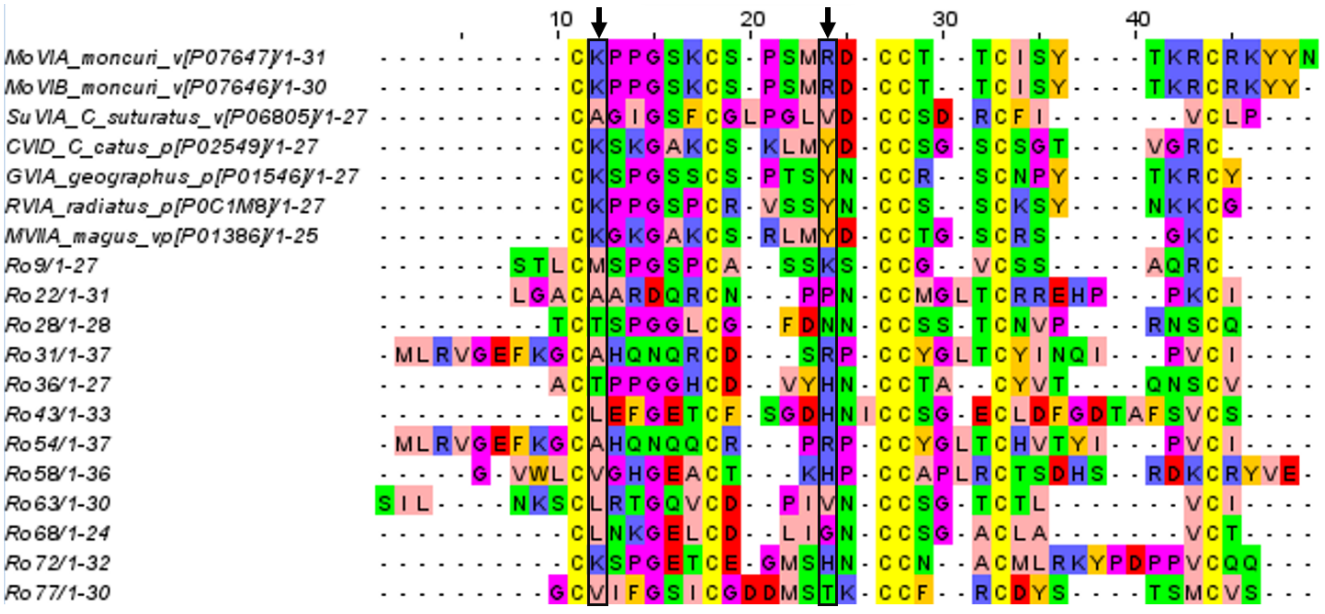

Figure 3: Sequence alignment of the mature peptide region of identified O1 superfamily from *C. rolani* and other O1 superfamily ω-conotoxins. The residues are colored using the Zappo color scheme to highlight the differences in their physicochemical properties.

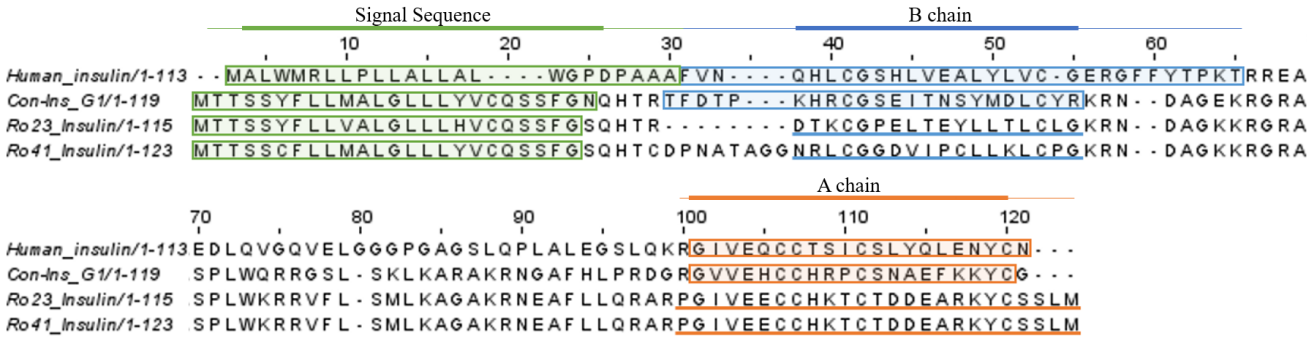

Figure 4: Sequences of human insulin, Con-Ins G1, and coninsulins found in the *C. rolani* transcriptome were aligned using MAFFT. In boxes are sequences annotated from protein evidence (orange or blue box) or conserved canonical signal peptides (green boxes). (Genbank accession numbers - Human_insulin: QMS45324.1, Con-Ins_G: A0A0B5AC95.1).

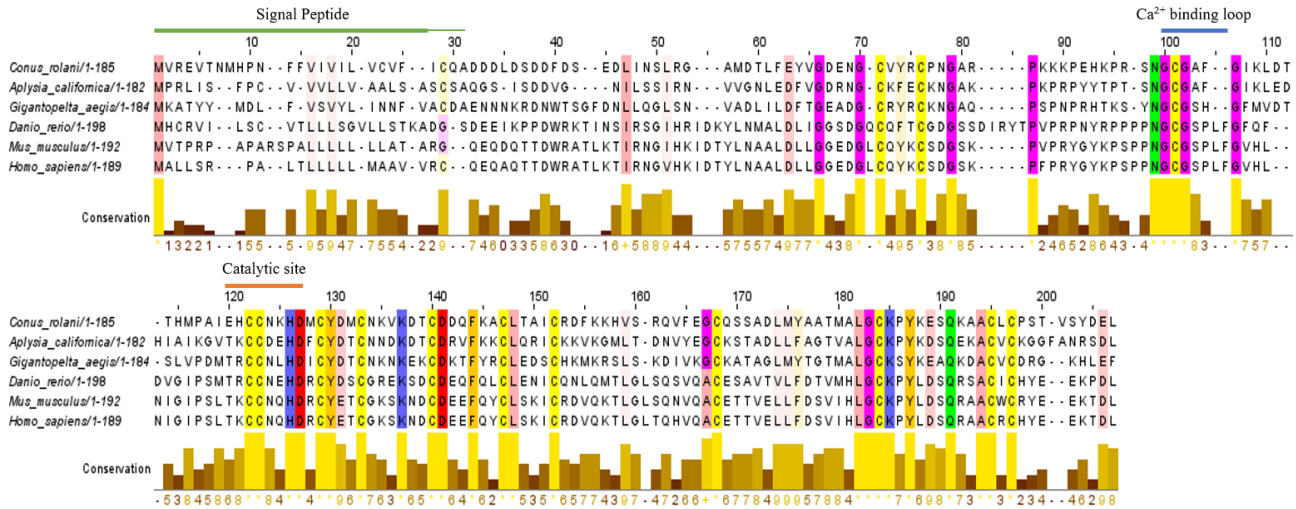

Figure 5: Alignment of Ro94_MVREV (labeled *Conus rolani*) with known secreted phospholipase A_2 ($sPLA_2$) group XII enzymes from different organisms with a reported genome assembly. Of note are the conserved HD residues in the catalytic site, which is a motif of all $sPLA_2$ enzymes. Conserved residues with a ≥9 score (Livingstone & Barton, 1993) are highlighted with a Zappo color scheme. (GenBank accession numbers – *Aplysia_californica*: XP_005096638.1, *Gigantopelta_aegis*: XP_041363593.1, *Danio_rerio*: NP_001076268.1, *Mus_musculus*: NP_075685.2, *Homo_sapiens*: NP_110448.2).

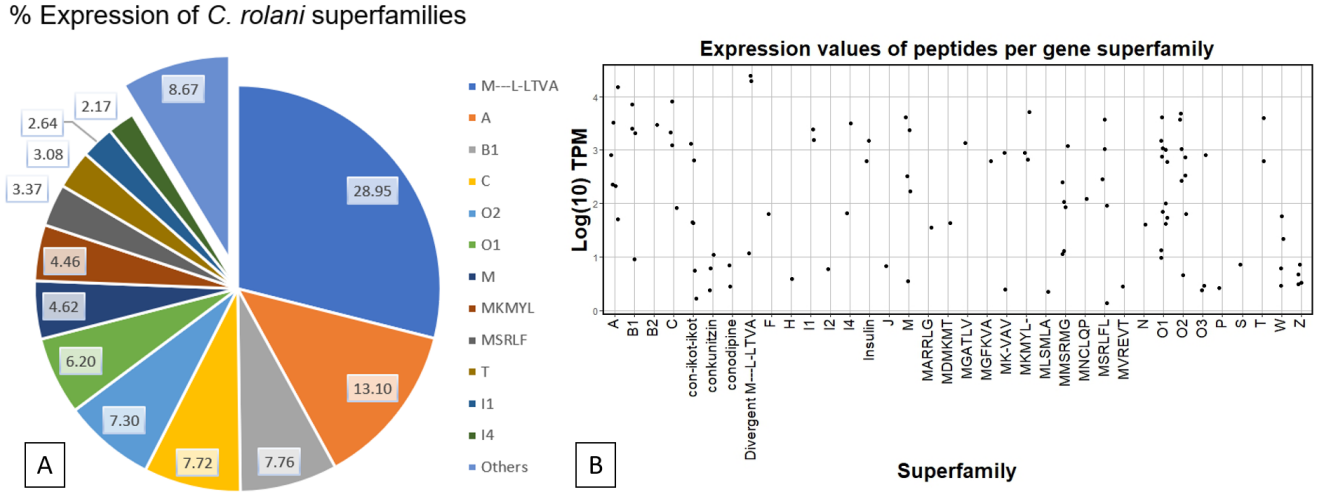

Figure 6: Relative expression of conopeptide superfamilies in *C. rolani*. A. Values reported are the percentage of respective superfamilies relative to all reads that aligned to conopeptides. B. Jitter plot of all conopeptide expression values (TPM) categorized by conopeptide superfamily.

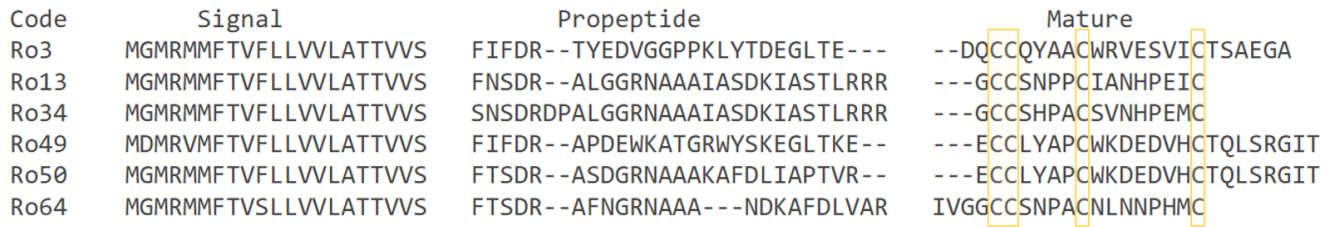

Figure 7: Alignment of A-superfamily precursor conopeptide sequences. Highlighted in yellow are cysteines that determine the 4/7 loop pattern.

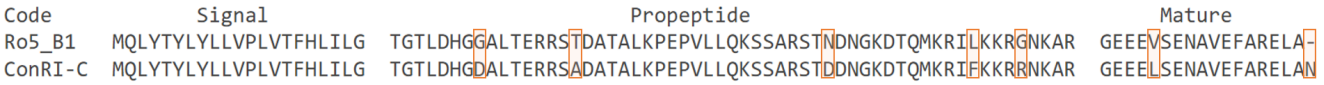

Figure 8: Comparison of the previously reported ConRI-C, obtained using a PCR-capture method, and Ro5_B1, from *C. rolani* (this study). Highlighted in orange are the amino acid differences between the two precursor sequences. (Genbank accession number – ConRI-C: P0DKZ1.1)

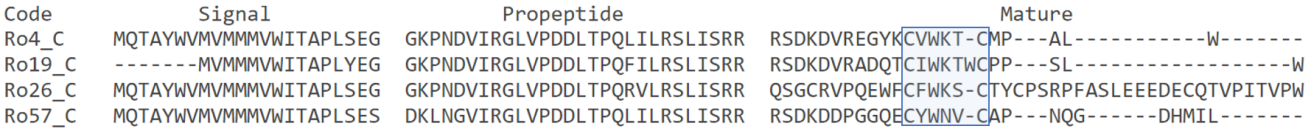

Figure 9: Sequence alignment of annotated consomatin precursor sequences. Boxed in blue are the regions that correspond to the core somatostatin receptor binding motif.

LIST OF TABLES

Table 1: Unassigned and a putative novel superfamily signal sequences from *C. rolani* with indicated SignalP or TargetP probability values.

C. rolani peptideCode	Signal Sequence	SignalP	TargetP	Cysteine Framework	Best-Hit of whole precursor conopeptide sequence			
					Species	%Identity	%Coverage	GenBank ID
Ro70	MARRLGFLIVALGLLLHWSQA	0.999	1	XIV	C. ermineus	68%	58%	AXL95338.1
Ro66	MDMKMTFSGLVLVVLVTTVFG	0.999	0.998	VIII	C. judaeus	77%	100%	UMA83597.1
Ro24	MGATLVTKLLLVAALLLGLCHEMAA	0.999	1	none	C. geographus	100%	67%	BAO65660.1
Ro40	MGFKVALIVLVLVVMATTSA	0.999	1	XIV	C. villepinii	85%	60%	P84704.2
Ro33	MKMYLSLAILLLLASTIVDS	0.999	1	none	C. betulinus	57%	99%	ALM87512.1
Ro32	MKVVAVFLVVALAAAYG	0.999	1	XXII	C. betulinus	86%	100%	ALM87501.1
Ro99	MLSMLAWTLMTAMVVMNAES	0.999	1	XXXIII	C. judaeus	73%	100%	UMA83620.1
Ro27	MMSRMGVEFFLLLLFALAST	0.999	1	VIII/none	C. ermineus	59%	96%	AXL95737.1
Ro52	MNCLQPLLVLLLISTITA	0.999	0.999	XIII	C. ebraeus	71%	98%	UMA82624.1
Ro12	MSRLFLVLLVISVVILHTDS	0.999	0.999	none	C. judaeus	71%	100%	UMA83294.1
Ro94	MVREVTNMHPNFFVIVILVCVFICQA	0.999	0.999	C-CC-C-C-C-C-C	Haliotis rufescens	55%	68%	XP_046382180.2

Table 2: Sequences of the predicted mature regions of the 15 most highly expressed conopeptide transcripts from *C. rolani* are shown with their TPM values. If the transcript has a close toxin match in UniProt (e-value: < 0.001) with bioassay data, the information is shown.

Code/Superfamily	Predicted mature peptide sequence	TPM	Closest hit with bioactivity data	Molecular target/s of hit	e-value	References
Ro1_Divergent M---L-LTVA	FPCDKDQCYCDHEEKDGQCMKISSKGTLCR GDKCKTKTDWKEPDNVLGTMYND	24299.82	-	-	-	-
Ro2_Divergent M---L-LTVA	FPCSTDQCYCEHKEEDKCMKTSSKDTLCKG DVCKTETDW	19429.5	-	-	-	-
Ro3_A	DQCCQYAACWRVESVICTSAEGA	15193.75	-	-	-	-
Ro4_C	RSDKDVREGYKCVWKTCMPALW	8182.22	Consomatin RoI*	Somatostatin receptors	exact match	Ramiro *et al*. 2022
Ro5_B1	GEEEVSENAVEFARELA	7150.97	conRI-C	NMDA receptors	1.9e-9	Gowd *et al*. 2012
Ro6_MKMYL-	CEQCSCYELNTARCTDKISCITPEACTPRGSC PTPGHCVCTRFASGCCTRAVECAPGKCH	5183.25	-	-	-	-
Ro7_O2	ESECPWRPWC	4826.24	contryphan-M	human L-type Ca^{2+} channels (Cav 1.2)	2.1e-7	Hansson *et al*. 2004
Ro8_M	GRCCIGPKGCSTRYCRNQQCC	4143.63	TIIIA	Ttx-Sensitive NaVs (rat NaV 1.2, NaV 1.4)	3.2e-11	Lewis *et al*. 2007
Ro9_O1	STLCMSPGSPCASSKSCCGVCSSAQRC	4105.79	GVIA	human CaV 2.2	8.1e-19	Hasan *et al*. 2021
Ro10_T	RICCFDDPDCCEA	4036.05	-	-	-	-
Ro11_O2	TLAEGGMWGECAEWLETCHWPVQCCSGN CEGHCRPWS	3730.99	R7A	-	0.00012	Jimenez *et al*. 2004
Ro12_MSRLFL	SRPLSRAATGDGFQGKLRARIFRRGALSDKT DSEEEETGEKELGKRQQAVQKRGIRRMKQ WNIN	3666.75	-	-	-	-
Ro13_A	GCCSNPPCIANHPEIC	3284.7	TxIA	nAChRs	1.1e-9	El Hamdaoui *et al*. 2019
Ro14_I4	QLWACSFDKTTESCSYETNELCSCAYHSCCTI PNKPPPVCIKDLMCPYFMSMGTSRRFTQM QERFLPML	3211.4	-	-	-	-
Ro15_B2	RQHSQFNADKSSFDSDSLGNIMDFMNENG NSLPFANMDSAATDLGNFEPSAENEDGKFR FFDKQQ	2990.11	-	-	-	-

*Exact match