High-Throughput Identification and Analysis of Novel Conotoxins from Three Vermivorous Cone Snails by Transcriptome Sequencing

Ge Yao [1,†], Chao Peng [2,†], Yabing Zhu [3], Chongxu Fan [1], Hui Jiang [1], Jisheng Chen [1], Ying Cao [1,*] and Qiong Shi [2,4,*]

[1]   State Key Laboratory of NBC Protection for Civilian, Beijing 102205, China; bzyaoge@163.com (G.Y.); chongxu_fan@hotmail.com (C.F.); jiangtide@sina.cn (H.J.); chenjsh@cae.cn (J.C.)
[2]   Shenzhen Key Lab of Marine Genomics, Guangdong Provincial Key Lab of Molecular Breeding in Marine Economic Animals, BGI Academy of Marine Sciences, BGI Marine, BGI, Shenzhen 518083, China; pengchao@genomics.cn
[3]   BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China; zhuyabing@genomics.cn
[4]   Laboratory of Aquatic Genomics, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518060, China
*   Correspondence: caoying01@sina.com (Y.C.); shiqiong@genomics.cn (Q.S.); Tel.: +86-755-3630-7807 (Q.S.)
†   These authors contributed equally to this work.

check for updates

Abstract: The venom of each *Conus* species consists of a diverse array of neurophysiologically active peptides, which are mostly unique to the examined species. In this study, we performed high-throughput transcriptome sequencing to extract and analyze putative conotoxin transcripts from the venom ducts of 3 vermivorous cone snails (*C. caracteristicus*, *C. generalis*, and *C. quercinus*), which are resident in offshore waters of the South China Sea. In total, 118, 61, and 48 putative conotoxins (across 22 superfamilies) were identified from the 3 *Conus* species, respectively; most of them are novel, and some possess new cysteine patterns. Interestingly, a series of 45 unassigned conotoxins presented with a new framework of C-C-C-C-C-C, and their mature regions were sufficiently distinct from any other known conotoxins, most likely representing a new superfamily. O- and M-superfamily conotoxins were the most abundant in transcript number and transcription level, suggesting their critical roles in the venom functions of these vermivorous cone snails. In addition, we identified numerous functional proteins with potential involvement in the biosynthesis, modification, and delivery process of conotoxins, which may shed light on the fundamental mechanisms for the generation of these important conotoxins within the venom duct of cone snails.

Keywords: *Conus*; conotoxin; transcriptome sequencing; phylogeny; venom duct

1. Introduction

Cone snail is the common name for predatory marine mollusks in the family Conidae, with over 700 extant species and a categorization of four genera and 71 subgenera [1–3]. Within the *Conus*, the largest genus in the Conidae, 57 subgenera have been recognized [3]. As venomous predators distributed throughout tropical and subtropical coastal waters all over the world, the living cone snails are typically divided into 3 groups based on their feeding habits, including fish hunters, mollusc hunters, and worm hunters [4–6]. Some phylogenetic data have suggested that the ancestral cone snails preyed on marine worms [7,8]. The fish-hunting and mollusc-hunting groups account for ~30% of *Conus* species, and they are assumed to be dangerous to humans; however, the largest

worm-hunting group seems to be nonthreatening [5,9,10]. An analysis of 141 human injuries reported from 34 responsible *Conus* species during the period of 1670–2017 [11] supports the fact that the venom of worm-hunting cone snails has only mild effects on humans, compared with those from fish-hunting and mollusc-hunting groups.

Although they are slow-moving creatures, cone snails can defeat fast-moving preys, competitors, and predators because of their specialized envenomation apparatus with potent venom components [12–14]. These venom components are commonly named conotoxins, a unique and remarkably diverse group of bioactive peptides with various pharmacological functions [6,14–17], which target a wide variety of ion channels, receptors, and even their subtypes in preys, predators, and humans with high affinity and specificity [6,18–20]. Consequently, conotoxins have become a research hotspot for the treatment of various neuropathic diseases, such as neuralgia, epilepsy, addiction, and Parkinson's disease [6,21–31].

With popular estimates of 50~200 classical conotoxins in a single *Conus* species, more than 80,000 natural conotoxins may exist in cone snails on a global scale [32–34]. Recent studies have shown that new methods, such as mass spectrometry, next-generation sequencing (NGS), and bioinformatics technologies, have predicted hundreds to thousands of venom peptides or transcripts from a single *Conus* species [34–37]. Therefore, cone snails, a tremendous store of natural conotoxins, are an underexploited resource for the development of potential drug candidates to treat a wide variety of human diseases [38].

Our present work reports a high-throughput transcriptome research on 3 worm hunting *Conus* species, *C. (Puncticulis) caracteristicus* (*C. caracteristicus*), *C. (Lividoconus) quercinus* (*C. quercinus*) and *C. (Strategoconus) generalis* (*C. generalis*), which are resident in offshore waters of the South China Sea. To date, there have been few studies on the screening of conotoxins from cone snails by transcriptome sequencing. One of our earlier studies on *C. quercinus* identified 65, 52, and 55 conotoxins from the venom duct, venom bulb and salivary gland, respectively [39]. Furthermore, only 38 and 4 conotoxins have been previously identified in *C. caracteristicus* and *C. generalis*, with classification into 10 (A, I3, M, O1, O2, O3, Q, S, T and Y) and 2 (D, O1) superfamilies, respectively [40–46]. In order to improve our understanding of the diversity of conotoxins, we performed transcriptome sequencing for the high-throughput identification and analysis of conotoxins from the venom duct of the 3 frequently collected cone snails.

2. Results

2.1. Summary of De Novo Assembled Transcriptome Data

After removal of low-quality reads, ambiguous reads and adapter sequences, we generated 4.57, 3.21, and 4.37 gigabases (Gb) of clean reads (with a mean length of 90 bp) for the venom duct transcriptomes of the 3 *Conus* species. Corresponding quality score 20 (Q20) of these sequencing data were 95.99%, 98.31%, and 96.00% respectively (Table 1). *De novo* assembling of all the high-quality clean reads using SOAPdenovo produced 213 k, 153 k, and 219 k contigs for the 3 species, respectively, which were subsequently assembled into scaffolds and unigenes. In total, the assembly of each transcriptome possessed 72 k, 61 k, and 95 k unigenes. More details of scaffold number, unigene number, mean length, and N50 value are summarized in Table 2.

Table 1. Statistics of venom duct transcriptome sequencing data for the 3 *Conus* species.

Species	Raw Data (Gb)	Clean Data (Gb)	Q20* (%)	Nonsequenced (%)	GC Content (%)
C. caracteristicus	5.51	4.57	95.99	0	47.84
C. quercinus	3.47	3.21	98.31	0.01	47.3
C. generalis	5.32	4.37	96.00	0	47.09

* A quality score for the percentage of incorrect bases at less than 1%.

Table 2. Summary of sequences produced by the assembling for the 3 *Conus* species.

Species	C. caracteristicus	C. generalis	C. quercinus
Clean reads			
Total reads (n)	50,788,576	48,557,734	35,694,024
Base pairs (Mb)	4,570.97	4,370.2	3,212.46
Mean length (bp)	90	90	90
Contigs ($\geq$100 bp)			
Total number	213,155	219,692	153,249
Base pairs (Mb)	47.84	60.75	40.22
Mean length (bp)	224	276	262
N50 (bp)	236	307	313
Scaffolds ($\geq$200bp)			
Total number	79,324	103,682	61,926
Base pairs (Mb)	47.57	65.38	34.96
Mean length (bp)	599	630	564
N50 (bp)	794	891	717
Unigenes ($\geq$200 bp)			
Total number	72,462	95,438	61,002
Base pairs (Mb)	39.61	54.87	33.67
Mean length (bp)	546	574	552
N50 (bp)	670	749	688

2.2. Screening of Conotoxins in the Venom Duct Transcriptomes

To annotate conotoxin coding sequences among the unigenes, we searched all six-frame translations of the unigenes against a local reference database of known conotoxins constructed from the public ConoServer database by running Genewise and Agustus with an E-value cut-off of 1.0×10^{-5} [18], and then manually checked them using the ConoPrec tool [19]. After the removal of the transcripts with duplication, frame-shifting, and truncated mature region sequences, we identified 118, 61, and 48 putative conotoxin sequences from the 3 transcriptome datasets of *C. caracteristicus*, *C. generalis,* and *C. quercinus*, respectively (Tables 3–5). Interestingly, most of these sequences are reported for the first time and some possess new cysteine patterns. We then summarized and named these predicted conotoxins from the 3 *Conus* species as Ca-1 to Ca-118, Ge-1 to Ge-61, and Qu-1 to Qu-48, respectively (see more details in Supplementary Tables S1–S3).

Table 3. Classification and cysteine patterns of the conotoxins identified from *C. caracteristicus*.

Superfamily		Number	Cysteine Pattern (Number of Conotoxins)
A		11	CC-C-C (8), CC-C (3)
B1 (Conantokin)		2	Cysteine free
C (Contulakin)		1	Cysteine free
D		2	C-C-CC-C-C-C-C (1), C-CC-C-CC-C-C-C-C (1)
	I1	1	C-C-CC-CC-C-C
I	**I2**	6	C-C-CC-CC-C-C (1), C-C-C-C-CC-C-C (4), C-C-CC-C-C (1)
	I3	4	C-C-CC-CC-C-C (3), C-C-CC-C-C (1)
J		7	C-C-C-C
L		4	C-C-C-C
M		6	CC-C-C-CC (5), CC-C-C-C-C (1)
	O1	22	C-C-CC-C-C
O	**O2**	11	C-C-CC-C-C (3), C-C-CC-C-C-C-C (3), C-C (5)
	O3	6	C-C-CC-C-C
S		3	C-C-C-C-C-C-C-C-C-C
T		9	CC-CC (8), C-C-CC (1)
Y		1	C-C-CC-C-CC-C
Divergent M—L-LTVA		1	C-C-C-C-C-C
Unknown		21	C-C-C-C-C-C (19), C-C-C-C (1), CC-C-C-C-C (1)
Total		**118**	

Table 4. Classification and cysteine patterns of the conotoxins identified from *C. generalis*.

Superfamily			Number	Cysteine Pattern (Number of Conotoxins)
A			2	CC-C-C
B1 (Conantokin)			1	Cysteine free
C (Conotulakin)			1	Cysteine free
D			1	C-CC-C-CC-C-C-C-C
I	**I1**		1	C-C-CC-CC-C-C
	I2		4	C-C-CC-CC-C-C (2), C-C-C-C-CC-C-C (2)
	I3		1	C-C-CC-CC-C-C
L			3	C-C-C-C
M			4	CC-C-C-CC (3), C-C-CC (1)
O	**O1**		12	C-C-CC-C-C
	O2		4	C-C-CC-C-C (3), C-C-CC-C-C-C-C (1)
	O3		3	C-C-CC-C-C
P			2	C-C-C-C-C-C
S			1	C-C-C-C-C-C-C-C-C-C
T			5	CC-CC
Con-ikot-ikot			1	CC-C-C-C-CC-C-C-C
Conotoxin-like			1	CC-C-C
Divergent MSTLGMTLL-			1	C-C-C-CCC-C-C-C-C
Unknown			13	C-C-C-C-C-C
Total			**61**	

Table 5. Classification and cysteine patterns of the conotoxins identified from *C. quercinus*.

Superfamily		Number	Cysteine Pattern (Number of Conotoxins)
A		5	CC-C-C
B1 (Conantokin)		3	Cysteine free
I2		3	C-C-CC-CC-C-C (2), C-C-C-C-CC-C-C (1)
M		10	CC-C-C-CC (9), C-CC-C-C-C (1)
O	**O1**	8	C-C-CC-C-C
	O2	3	C-C-CC-C-C
	O3	1	C-C-CC-C-C
T		1	CC-CC
V		3	C-C-CC-C-C-C-C
Y		1	C-C-CC-C-CC-C
Con-ikot-ikot		1	CC-C-C-C-C-CC-C-C-C-C
Divergent M—L-LTVA		2	C-C-C-C-C-C
Unknown		7	C-C-C-C-C-C
Total		**48**	

In this study, each putative conotoxin was assigned to a superfamily based on its percentage of sequence identity to the highly conserved signal region of the known superfamily from the public ConoServer database (Figure 1). Here, among the 118 putative conotoxins in *C. caracteristicus*, 96 sequences were assigned to 16 previously reported superfamilies (A, B1, C, D, I1, I2, I3, J, L, M, O1, O2, O3, S, T, and Y), while only 1 sequence was classified into the "divergent M—L-LTVA" superfamily. In addition, 21 sequences were not assigned to any known superfamily (named "unknown"; see more details in Table 3).

Among the 61 putative conotoxins in *C. generalis*, 46 sequences were classified into 15 known superfamilies (A, B1, C, D, I1, I2, I3, L, M, O1, O2, O3, P, S, and T) and 1 cysteine-rich con-ikot-ikot family. In addition, 1 sequence was assigned to the "divergent MSTLGMTLL-" super-family, 1 was assigned to the conotoxin-like group, and the other 13 sequences were unknown (see more details in Table 4).
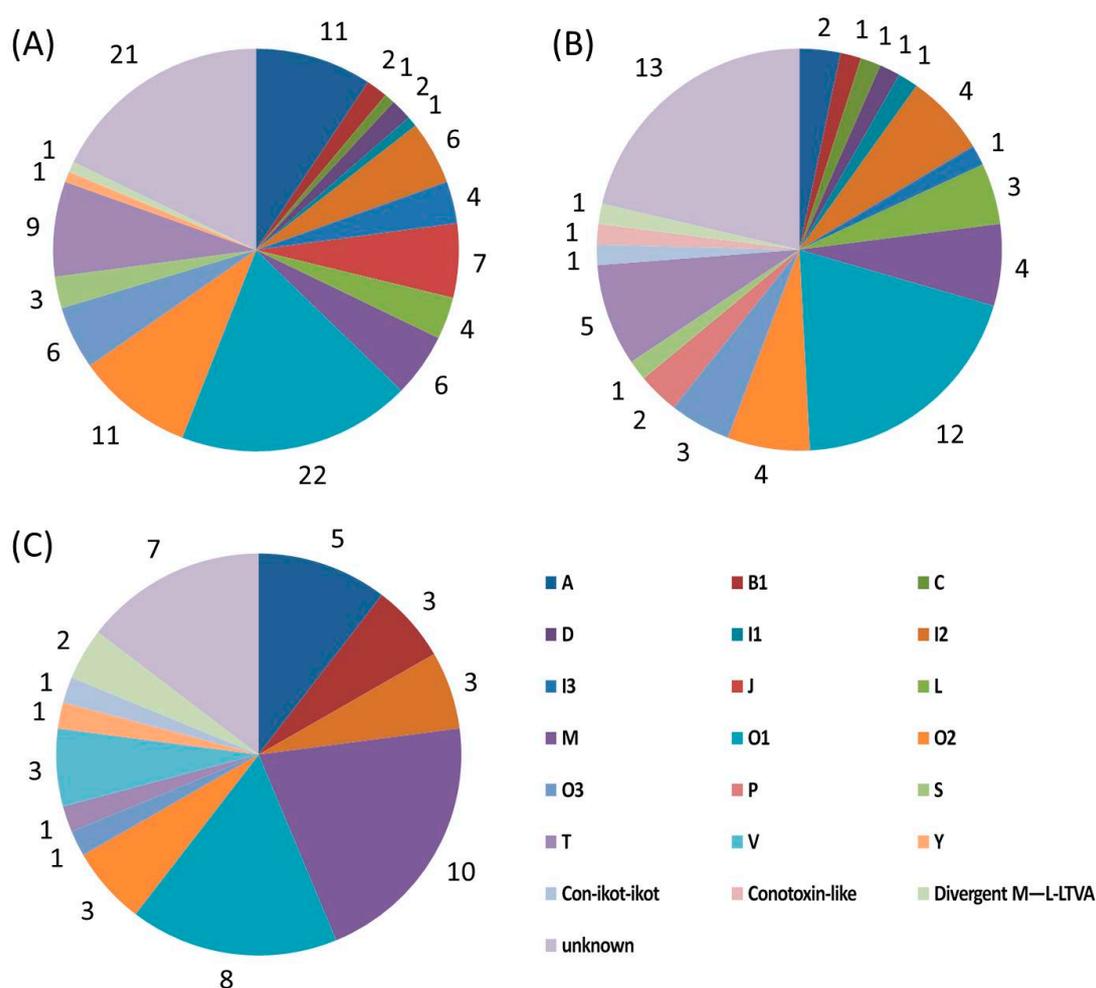
Figure 1. Summary of the conotoxins identified from the 3 *Conus* species. Many superfamilies or groups of conotoxins were classified in (**A**) *C. caracteristicus*, (**B**) *C. generalis*, and (**C**) *C. quercinus*.

Compared with the conotoxin sequences reported in our previous transcriptome study of *C. quercinus* [39], the number of conotoxins from *C. quercinus* in this study was less, with the identification of only 48 putative conotoxins. Among them, 39 sequences were classified into 10 known superfamilies (A, B1, I2, M, O1, O2, O3, T, V, and Y) and the con-ikot-ikot family, 2 sequences were assigned to the "divergent M—L-LTVA" superfamily, and the remaining 7 sequences were unknown (see more details in Table 5).

2.3. Quantification of Conotoxin Abundance

To investigate the transcription levels of conotoxins in each species, we mapped clean reads back to the *de novo* assembled unigenes and calculated the fragments per kilobase of transcript per million mapped fragments (FPKM) values to quantify the abundance of each conotoxin transcript. We screened out those conotoxins with high transcription abundance, and the top 10 (with the highest FPKM values) were selected from each dataset for comparison. The number of mapped reads for the top 10 conotoxins accounted for 60.6%, 84.9%, and 80.4% of the total conotoxin reads from *C. caracteristicus*, *C. generalis*, and *C. quercinus* respectively. Interestingly, O- and M-superfamilies were always the most abundant within each transcriptome dataset (Figure 2, Supplementary Table S4), suggesting their critical roles in predation and defense for the 3 vermivorous *Conus* species.
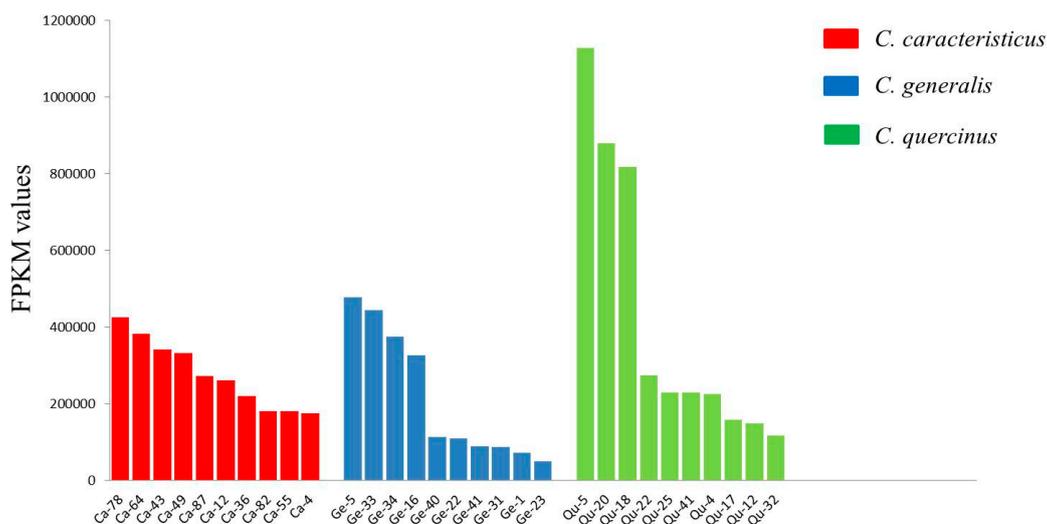
Figure 2. Comparison of the top 10 conotoxins (with the highest FPKM values) from the 3 transcriptome datasets.

The O-superfamily conotoxins specifically target a wide range of ion channels and receptors. In this study, a Bayesian phylogenetic tree was constructed with the Markov Chain Monte Carlo algorithm to analyze the relationships among these predicted O-superfamily conotoxins (Figure 3), in which 14 were presented with known bioactivities (previously reported from different *Conus* species) and 11 demonstrated high transcription levels from this study. Our phylogenetic assessment indicated that the O-superfamily conotoxin clades from the same *Conus* species arise as distinct lineages, suggesting that there was no correlation between the evolution of conotoxin sequences and interspecific genetic relationship (see more details in Figure 3). In turn, Ca-55 and ω-conotoxin PnVIA/PnVIB formed a monophyletic clade [47], and the Posterior probability of Ca-55 and PnVIB was 0.69. Ge-23 and κ-conotoxin PVIIA formed an individual clade [48], and the Posterior probability of both was 0.68. Qc-32 and γ-conotoxin PnVIIA/TxVIIA formed a monophyletic clade [49,50], but the homology between them was not supported by the low Posterior probability; they all exhibited distant evolutionary relationships to the other O-super-family conotoxins. Similarly, Ge-34 and δ-conotoxin TxVIA/TxVIB also formed a separate clade but with low Posterior probability [51].
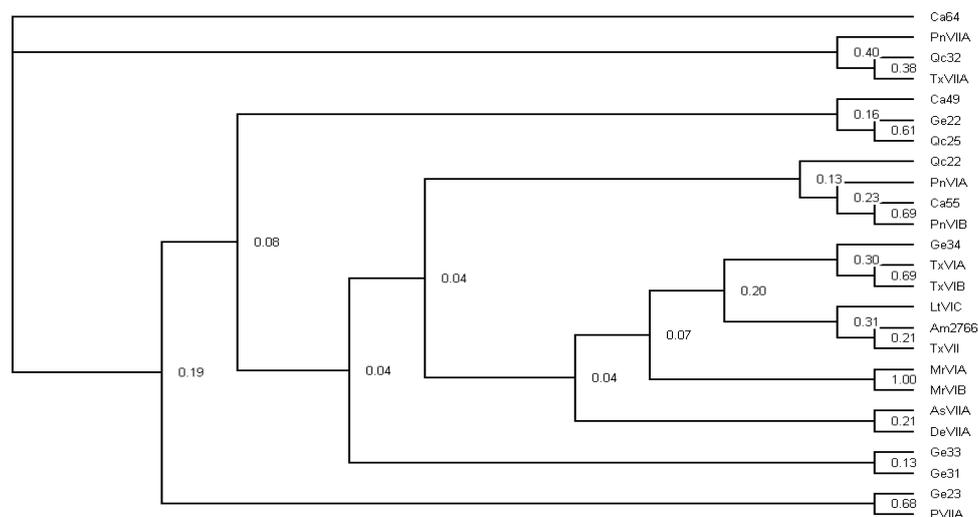
Figure 3. A Bayesian phylogenetic tree of the O-superfamily conotoxins. Posterior probabilities are labeled at each node.

2.4. Diversity of Conotoxin Structures

Among the putative conotoxin sequences in the 3 venom duct transcriptomes, most of them were discovered for the first time, and some possessed new cysteine frameworks or belonged to unknown superfamilies. The O-superfamily, including O1, O2, and O3, was the most abundant group in terms of conotoxin number (Tables 3–5). All of the O1- and O3-superfamily members exhibited the conventional VI/VII (C-C-CC-C-C) cysteine framework, which provides a stable three-disulfide inhibitor cysteine knot (ICK) motif [46]. Four O2-superfamily sequences from *C. caracteristicus* and *C. generalis* exhibited a C-terminal elongated XV (C-C-CC-C-C-C-C) cysteine framework (see Tables 3 and 4), and 5 short single disulfide-containing contryphan peptides with high identity were identified from only *C. caracteristicus* (Table 3). Interestingly, 5 O-superfamily members had the same mature regions as reported sequences from other *Conus* species. For example, Ge-24 from *C. generalis* had exactly the same mature peptide and prepro-region as the reported MgJr94 from piscivorous *C. magus* [52], while Ge-22 and Ge-23 showed mutations in the pro-peptide regions but with identical mature peptides to MiK41 and MiK42 respectively from *C. miles* [53].

The M-superfamily was also the predominant one in terms of transcription abundance and diversity of cysteine frameworks. Besides 17 sequences with the typical III (CC–C–C–CC) cysteine pattern, 3 conotoxins with IV (CC-C-C-C-C), XVI (C-C-CC) and XXVII (C-CC-C-C-C) cysteine frameworks were also observed (Tables 3–5).

For the I-superfamily, a variety of conotoxin (including I1, I2 and I3) transcripts were retrieved from the 3 transcriptomes. A total of 20 members were identified, in which most (13 sequences) belonged to the I2-superfamily and possessed the typical post-peptide and pro-region-free structure. These I-superfamily conotoxins generally had various signal regions and cysteine-rich frameworks, of which 11 exhibited the representative XI (C-C-CC-CC-C-C) pattern and 7 possessed the XII (C-C-C-C-CC-C-C) pattern, whereas only 2 had the framework VI/VII (C-C-CC-C-C) with distinct signal sequences and loop length (Tables 3–5).

Concerning the A-superfamily, 18 conotoxins were determined with 15 sequences of the common I (CC-C-C) pattern, and 3 transcripts with only 3 cysteine residues (the CC-C framework). In contrast to notable abundance and variety of A-superfamily conotoxins in previously reported piscivorous species [54–56], the number and diversity of the A-superfamily identified from these 3 vermivorous species were scarce. Meanwhile, 15 T-superfamily conotoxins were also identified; however, most of them exhibited the simple V (CC-CC) framework with high identity to several known τ-conotoxins, and only one contained the XVI (C-C-CC) framework.

In addition to the abovementioned major superfamilies, many less representative B1 (conontokin)-, C (contulakin)-, D-, J-, L-, P-, S-, V-, Y-superfamilies and the con-ikot-ikot family were also discovered in the 3 transcriptomes. For example, 6 conontokin sequences and 2 contulakin sequences with the cysteine free pattern were identified in this study. Three D-superfamily sequences with the XII (C-CC-C-CC-C-C-C-C) and XV (C-C-CC-C-C-C-C) patterns, 3 V-superfamily sequences with the XV (C-C-CC-C-C-C-C) pattern, and 2 Y-superfamily sequences with the XVII (C-C-CC-C-CC-C) pattern were also observed; interestingly, these 3 superfamilies have been isolated only from vermivorous cone snails to date [57]. Meanwhile, 2 con-ikot-ikot peptides with the novel CC-C-C-C-CC-C-C-C and CC-C-C-C-C-CC-C-C-C-C frameworks were identified. Con-ikot-ikot toxins were reported to specifically target post-synaptic AMPA receptors [58], but in consideration of the complex cysteine patterns and variable loop length, the functions of both conotoxins with new cysteine frameworks are worth investigating further.

In this study, combined with our published report of *C. betulinus* [32], we also identified 45 putative unassigned conotoxins, which possessed the IX (C-C-C-C-C-C) cysteine framework and loop lengths same as Cal9.1a~d from *Californiconus californicus* [3,59]. In fact, Cal9.1a~d contained unassigned signal peptide sequences, and their loop lengths and mature regions appeared to be sufficiently distinct from other known conotoxins. This group of conotoxins was previously found only in *Californiconus californicus*, an unusual species with special prey-capture behavior and prey preferences,

and phylogenetic analysis also indicated that *Californiconus californicus* has large evolutionary distance from the *Conus* species [60]. The experts at WoRMS placed this genus *Californiconus* in the family Conidae, but as indicated it is highly divergent from the Conidae; hence, some researchers have placed this genus in a proposed separate (sub)family [1]. Our present work confirmed that these types of conotoxins are possibly synthesized in various species with different feeding habits; therefore, they may represent a new superfamily with potential specificity in pharmacological activity (Figure 4).
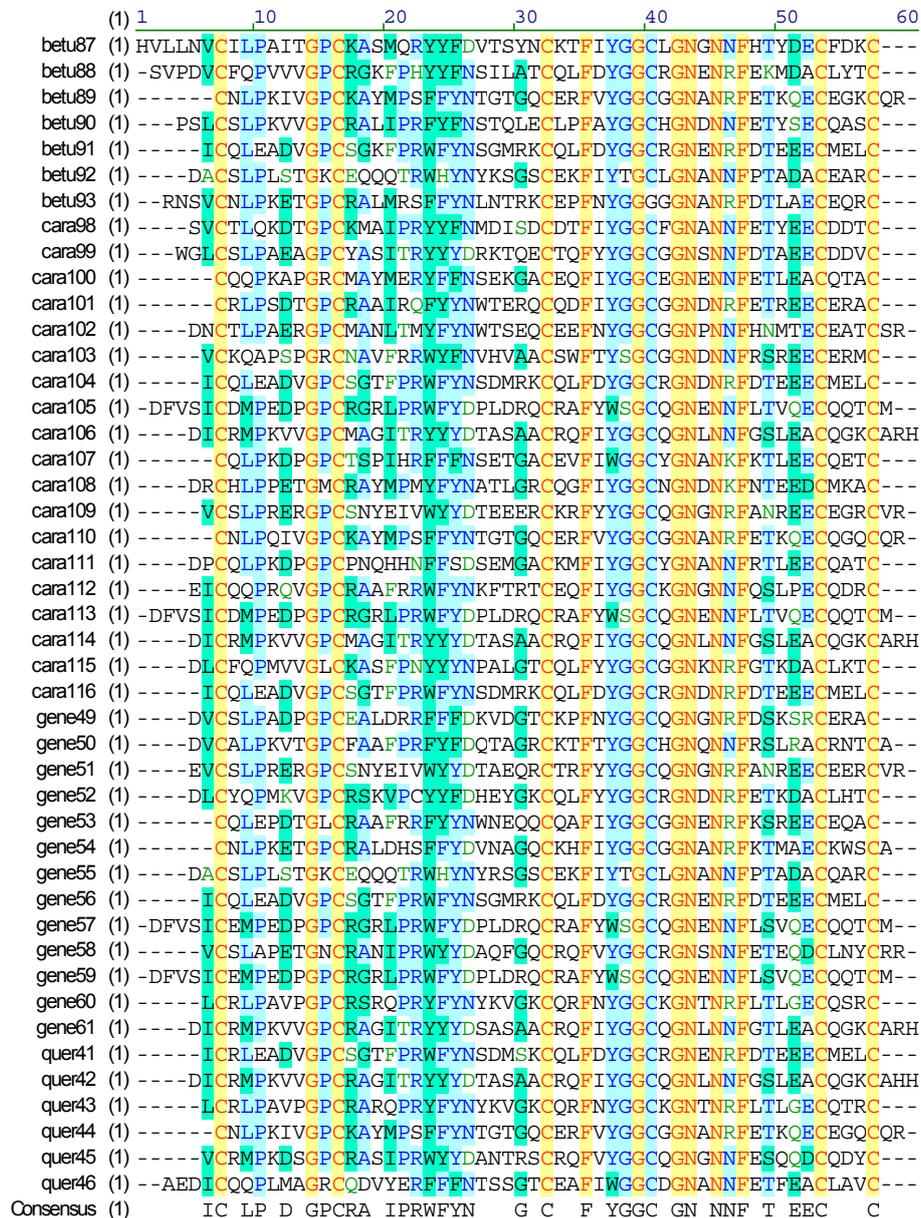
Figure 4. Alignment of the achieved new superfamily conotoxins from venom duct transcriptomes of *C. betulinus* [32] (betu), *C. caracteristicus* (cara), *C. generalis* (gene), and *C. quercinus* (quer).

2.5. Identification of Conotoxin Biosynthesis Related Proteins

Transcripts for genes encoding functional proteins that are potentially involved in conotoxin biosynthesis were also annotated in the 3 transcriptome datasets. By homology comparison to known post-translational modification enzymes, we presumed that multiple proteins possess enzymatic activities for involvement in conotoxin maturation and modification in the venom duct lumen. From the transcriptomes, we also identified some isoforms of endoprotease, including sequences with high

similarity to Tex31 [61], which has the hydrolytic activity to separate mature conotoxins from the precursor constituents.

Formation of disulfides is the most ubiquitous modification within conotoxins. This process and related proper peptide-folding are mediated by protein disulfide isomerases (PDIs), peptidyl-prolyl cis-trans isomerases, immunoglobulin-binding proteins, and chaperones (e.g., hsp70, hsp60, and calreticulin) [33]. All of these proteins, especially multi-isoforms of PDI were identified (Supplementary Tables S1–S3). Complete sequences of peptidylglycine alpha-amidating monooxygenase with two domains were identified as well, which may mediate the C-terminal amidation process for the full activity of various neuroactive peptides [62]. In addition, other candidate enzymes participating in post- translational modification were also predicted, such as prolyl/lysyl-hydroxylase, vitamin K-dependent γ-carboxylase, and Glutaminyl-peptide cyclotransferase.

Our data also revealed numerous sequences with potential roles in transportation, synergy, and degradation of conotoxins. Translocator-like sequences including the Sec family and diverse transmembrane proteins were discovered. In particular, we predicted several transcripts with high similarity to the Sec61 and Sec14 translocon that were identified in the spider venom duct with the ability to bind specific polypeptide toxins and to induce subsequent localization and transportation [63,64]. Large arrays of conotoxin-related proteins (widely existing in other animal venoms) with high transcription levels were also predicted. These sequences include multiple enzymes, such as the phospholipase A2 (PLA2) family, nucleotidase and hyaluronidase, as most of them may have neurotoxic and cytotoxic activities themselves or participate in anti-hemostatic effects [65], and may enhance the diffusion of conotoxins and cooperate with them for prey capture or synergistic predator defense. Meanwhile, many proteins homologous to the ubiquitin and ubiquitylation system were retrieved from the 3 transcriptome datasets in this study. These proteins may contribute to the degradation of poor-quality conotoxin molecules synthesized in the venom duct, which will ensure the effectiveness of venom to a higher degree in the prey capture process [66] (Supplementary Table S5).

3. Discussion

A total of 118, 61, and 48 putative conotoxin transcripts were identified from the 3 transcriptome datasets of *C. caracteristicus*, *C. generalis*, and *C. quercinus*, respectively. Given that these *Conus* species have similar feeding habits and distribution (in the offshore waters of the South China Sea), the interspecific divergence in toxin numbers and transcription levels were beyond our expectation. The interspecies variability of venom may contribute to the dietary preferences for different worms. Furthermore, the conotoxin composition in the venom duct of *C. quercinus* from this study was only 23% (11 of 48), identical to our previous report [39]. This remarkable variance of conotoxin types in the same species, especially in the same organ and using the same sequencing method, may be due to differences between *Conus* individuals that were from different geographical populations or at different developmental stages (for a more detailed discussion, please refer to our previous report on *C. betulinus* [32]). The intraspecies variability of venom composition has also been recently observed by other researchers, who recommended sequencing transcriptomes of more than one individual for the solid analysis of the conotoxin inventory in any examined species [67]. In addition, we also identified some identical conotoxin sequences from different *Conus* species; this convergent evolution trend is widespread presented among various *Conus* species with diverse phylogenetic clades. Related biological significance needs more investigation.

Conotoxins have a variety of mechanisms for actions, which however have been limited by the lack of high-throughput functional screening methods; therefore, most of them have not been determined by far. In this study, we attempted to apply the phylogenetic analysis strategy to explore the evolutionary relationships of these highly transcribed O-superfamily members, and then predicted their potential bioactivities. Among the top 10 conotoxins with the highest FPKM values in the 3 *Conus* species, 4 O-superfamily conotoxins formed monophyletic clades with several known pharmacological conotoxins, but the poor Posterior probabilities (0.30~0.69) at the nodes do not

support related bioactivity prediction. The detailed functions of these high-abundance conotoxins in predation or defense deserve further investigation (such as by patch-clamp detection). Construction of electrophysiological platforms for the analysis of multiple neural ion channels and receptors is underway in our laboratory, which in turn proposes a potential hope for the development of novel conotoxin-based marine drugs for the treatment of neuroreceptor associated human diseases.

This study also identified many complete and partial sequences of 11 enzymes, which are potentially involved in the post-translational modification of conotoxins, as well as numerous functional proteins that may be related to the conotoxin biosynthesis processes including translation, protein folding, translocation, delivery and degradation. Gene Ontology (GO) functional classification (Supplementary Table S5) showed that conotoxin synthesis may be closely related to binding, catalytic activity, metabolic process and cellular process. The abundance of functional proteins underscores the fact that the venom duct is a metabolically active organ. Although there have been many reports on various novel conotoxins, our present work improves the understanding of conotoxin biosynthesis processes in vivo, which may provide insights into the fundamental mechanisms underlying the generation of complexly modified peptides in general.

4. Materials and Methods

4.1. Sample Collection, RNA Extraction and Sequencing

C. caracteristicus, C. quercinus, and *C. generalis* were collected in the offshore areas of Sanya City, Hainan Province, China. Specimen identification (using *COI* gene sequences [39]) was performed after they were collected and dissected on ice. Three intact venom ducts were separated and the total RNAs were extracted using TRIzol® LS Reagent (Invitrogen, Life Technologies, Carlsbad, CA, USA) following the manufacturer's instructions. Total RNAs were further treated with oligo-(dT)- attached magnetic beads (Invitrogen, Life Technologies, Carlsbad, CA, USA) to extract the mRNAs. Three non-normalized Illumina cDNA libraries were constructed separately and sequenced on the Illumina HiSeq2000 platform (Illumina, San Diego, CA, USA) by BGI-Tech (BGI, Shenzhen, China).

4.2. Sequence Analysis and Assembling

Raw sequencing reads from the 3 sets of transcriptome sequencing were cleaned up using SOAPnuke software (BGI, Shenzhen, Guangdong, China) [68] to ensure high quality for downstream analyses. Adapters and reads with over 10% of non-sequenced (N) bases or more than 50% of low-quality bases (base quality $\leq$ 10) were removed. Then the filtered reads were assembled into unigenes with SOAPdenovo-Trans v1.02 (BGI, Shenzhen, Guangdong, China) for *de novo* transcriptome assembling [69]. The FPKM value, a general parameter for quantification of gene transcription, was calculated for comparison [70].

4.3. Prediction and Identification of Conotoxins

All previously known conotoxins in the ConoServer database [19] were downloaded to construct a local reference dataset for conotoxin prediction from our 3 transcriptome datasets using the traditional homology search method. Subsequently, unigenes from each transcriptome were run against the local conotoxin dataset using Genewise v2.4.1 [71] and Agustus v2.7 [72] with an E-value of 1.0×10^{-5}. Those unigenes with the best hits were translated into peptide sequences. A conotoxin generally consists of a highly conserved N-terminal signal peptide region, a less conserved intervening pro-peptide region, and a hypervariable C-terminal mature peptide region with conserved cysteine patterns [73,74]. A few conotoxins also have a post-peptide region at the C-terminal after the mature peptide region [75]. The predicted conotoxin transcripts were manually inspected using the ConoServer's web-based ConoPrec and NCBI's blastp. Those transcripts with duplication or truncated mature region sequences were removed.

4.4. Classification of Conotoxin Superfamilies

The distinct regions and cysteine frameworks of these predicted conotoxins were analyzed using the ConoServer's web-based ConoPrec. Based on 75% identity in the conserved signal peptide sequences [57], these identified conotoxins could be assigned to most of the 27 known superfamilies in the ConoServer. The particular threshold values for I1, I2, L, M, P, S, con-ikot-ikot and divergent superfamilies were 71.85%, 57.6%, 67.5%, 69.3%, 69.1%, 72.9%, $64.5 \pm 20.2\%$, and $64.22 \pm 20.53\%$, respectively [33]. Those conotoxins without signal regions but still showing high similarity either in the proregion or mature region were considered as the "Unknown" group.

4.5. Annotation of Predicted Functional Proteins

Unigenes were firstly translated into amino acids in six frames and aligned with BLASTX to public protein databases (E-value $\leq 1.0 \times 10^{-5}$) including NCBI non-redundant (Nr), Swiss-Prot [76], and Clusters of Orthologous Groups (COG) [77]. The protein with the highest sequence similarity was retrieved and annotated to each unigene. For the Nr annotation, Blast2GO v4.1 (Instituto Valenciano de Investigaciones Agrarias, Moncada, Valencia, Spain) [78] was used to determine GO annotation, which was defined by molecular function, cellular component, and biological process ontologies.

4.6. Phylogenetic Inference of Abundant Conotoxins

Bayesian analyses of the combined data were performed with MrBayes v.2.01 (University of Rochester, Rochester, NY, USA) [79] using the best-fit model indicated by Modeltest 3.06 (Brigham Young University, Provo, UT, USA) [80]. A Metropolis-coupled Markov Chain Monte Carlo algorithm running four Markov chains simultaneously was employed to estimate the posterior probability of phylogenetic trees. Each Markov chain was initiated with a random tree and run for 1,000,000 generations, sampling every 100 generations for a total of 10,000 samples per run. The first 2,500 samples of each run were discarded as burn-in, and the remaining samples were applied to construct a consensus tree using PAUP*4.0 (Florida State University, Tallahassee, FL, USA) [81].

4.7. Availability of Supporting Data

The datasets supporting the results of this article are included within the article and its supplementary files. The transcriptome reads generated in this study have been deposited in China National GeneBank Nucleotide Sequence Archive with accession numbers of CNS0048931 for *C. caracteristicus*, CNS0048933 for *C. generalis*, and CNS0048932 for *C. quercinus* under the project CNP0000360.

5. Conclusions

In this report, we have examined the diverse transcription repertoire in the venom ducts of 3 vermivorous cone snails, which are resident in the offshore waters of the South China Sea. We not only succeeded in characterizing the abundant conotoxin-encoding transcripts across 22 known superfamilies and 1 new superfamily, but also identified a variety of functional proteins that may be responsible for biosynthesis and delivery of these conotoxins. As expected, the majority of the identified conotoxins were novel, based on their transcript sequences, and some possessed new cysteine frameworks and divergent signal regions. Comparison analysis indicated surprising interspecific and intraspecific divergences in the conotoxin numbers and transcription levels, thus we made a primary conclusion that the abundant O-superfamily conotoxins in all 3 venom duct transcriptomes probably play major roles in the prey capture strategy of these vermivorous species. Our study with various cone snail species provides new insights into the complex biosynthesis mechanisms that lead to the remarkable variability of the venom composition. Our present work adds more conotoxins, which will definitely improve our genetic resource to develop new drugs.

References

1. World Register of Marine Species. Available online: http://www.marinespecies.org (accessed on 21 January 2019).

2. Puillandre, N.; Bouchet, P.; Duda, T.F., Jr.; Kauferstein, S.; Kohn, A.J.; Olivera, B.M.; Watkins, M.; Meyer, C. Molecular phylogeny and evolution of the cone snails (Gastropoda, Conoidea). *Mol. Phylogenet. Evol.* **2014**, *78*, 290–303. [CrossRef]

3. Puillandre, N.; Duda, T.F.; Meyer, C.; Olivera, B.M.; Bouchet, P. One, four or 100 genera? A new classification of the cone snails. *J. Molluscan Stud.* **2015**, *81*, 1–23. [CrossRef]

4. Prashanth, J.R.; Dutertre, S.; Jin, A.H.; Lavergne, V.; Hamilton, B.; Cardoso, F.C.; Griffin, J.; Venter, D.J.; Alewood, P.F.; Lewis, R.J. The role of defensive ecological interactions in the evolution of conotoxins. *Mol. Ecol.* **2016**, *25*, 598–615. [CrossRef] [PubMed]

5. Himaya, S.W.; Jin, A.H.; Dutertre, S.; Giacomotto, J.; Mohialdeen, H.; Vetter, I.; Alewood, P.F.; Lewis, R.J. Comparative venomics reveals the complex prey capture strategy of the piscivorous cone snail *Conus catus*. *J. Proteom. Res.* **2015**, *14*, 4372–4381. [CrossRef] [PubMed]

6. Lewis, R.J.; Dutertre, S.; Vetter, I.; Christie, M.J. *Conus* venom peptide pharmacology. *Pharmacol. Rev.* **2012**, *64*, 259–298. [CrossRef]

7. Gao, B.; Peng, C.; Chen, Q.; Zhang, J.; Shi, Q. Mitochondrial genome sequencing of a vermivorous cone snail *Conus quercinus* supports the correlative analysis between phylogenetic relationships and dietary types of *Conus* species. *PLoS ONE* **2018**, *13*, e0193053. [CrossRef]

8. Aman, J.W.; Imperial, J.S.; Ueberheide, B.; Zhang, M.M.; Aguilar, M.; Taylor, D.; Watkins, M.; Yoshikami, D.; Showers-Corneli, P.; Safavi-Hemami, H.; et al. Insights into the origins of fish hunting in venomous cone snails from studies of *Conus tessulatus*. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5087–5092. [CrossRef]

9. Robinson, S.D.; Safavi-Hemami, H.; McIntosh, L.D.; Purcell, A.W.; Norton, R.S.; Papenfuss, A.T. Diversity of conotoxin gene superfamilies in the venomous snail, *Conus victoriae*. *PLoS ONE* **2014**, *9*, e87648. [CrossRef] [PubMed]

10. Kumar, P.S.; Kumar, D.S.; Umamaheswari, S. A perspective on toxicology of *Conus* venom peptides. *Asian Pac. J. Trop. Med.* **2015**, *8*, 337–351. [CrossRef]

11. Kohn, A.J. *Conus* envenomation of humans: In fact and fiction. *Toxins* **2019**, *11*, 10. [CrossRef]

12. Olivera, B.M. *Conus* venom peptides, receptor and ion channel targets and drug design: 50 million years of neuropharmacology. *Mol. Biol. Cell.* **1997**, *8*, 2101–2109. [CrossRef] [PubMed]

13. Dutertre, S.; Jin, A.H.; Vetter, I.; Hamilton, B.; Sunagar, K.; Lavergne, V.; Dutertre, V.; Fry, B.G.; Antunes, A.; Venter, D.J.; et al. Evolution of separate predation- and defence-evoked venoms in carnivorous cone snails. *Nat. Commun.* **2014**, *5*, 3521. [CrossRef] [PubMed]

14. Endean, R.; Duchemin, C. The venom apparatus of *Conus magus*. *Toxicon* **1967**, *4*, 275–284. [CrossRef]

15. Safavi-Hemami, H.; Li, Q.; Jackson, R.L.; Song, A.S.; Boomsma, W.; Bandyopadhyay, P.K.; Gruber, C.W.; Purcell, A.W.; Yandell, M.; Olivera, B.M.; et al. Rapid expansion of the protein disulfide isomerase gene family facilitates the folding of venom peptides. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 3227–3232. [CrossRef] [PubMed]

16. Neves, J.; Campos, A.; Osório, H.; Antunes, A.; Vasconcelos, V. Conopeptides from Cape Verde *Conus crotchii*. *Mar. Drugs* **2013**, *11*, 2203–2215. [CrossRef] [PubMed]

17. Phuong, M.A.; Mahardika, G.N.; Alfaro, M.E. Dietary breadth is positively correlated with venom complexity in cone snails. *BMC Genomics* **2016**, *17*, 401. [CrossRef] [PubMed]

18. Kaas, Q.; Westermann, J.C.; Halai, R.; Wang, C.K.; Craik, D.J. ConoServer, a database for conopeptide sequences and structures. *Bioinformatics* **2007**, *24*, 445–446. [CrossRef] [PubMed]

19. Kaas, Q.; Yu, R.; Jin, A.H.; Dutertre, S.; Craik, D.J. ConoServer: Updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Res.* **2012**, *40*, D325–D330. [CrossRef]

20. Terlau, H.; Olivera, B.M. *Conus* venoms: A rich source of novel ion channel-targeted peptides. *Physiol. Rev.* **2004**, *84*, 41–68. [CrossRef]

21. Tosti, E.; Boni, R.; Gallo, A. μ-Conotoxins modulating sodium currents in pain perception and transmission: A therapeutic potential. *Mar. Drugs* **2017**, *15*, 295. [CrossRef]

22. Barton, M.E.; White, H.S.; Wilcox, K.S. The effect of CGX-1007 and CI-1041, novel NMDA receptor antagonists, on NMDA receptor-mediated EPSCs. *Epilepsy Res.* **2004**, *59*, 13–24. [CrossRef]

23. Romero, H.K.; Christensen, S.B.; Di Cesare Mannelli, L.; Gajewiak, J.; Ramachandra, R.; Elmslie, K.S.; Vetter, D.E.; Ghelardini, C.; Iadonato, S.P.; Mercado, J.L.; et al. Inhibition of alpha9alpha10 nicotinic acetylcholine receptors prevents chemotherapy-induced neuropathic pain. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 1825–1832. [CrossRef]

24. Hannon, H.E.; Atchison, W.D. Omega-conotoxins as experimental tools and therapeutics in pain management. *Mar. Drugs* **2013**, *11*, 680. [CrossRef]

25. Crooks, P.A.; Bardo, M.T.; Dwoskin, L.P. Nicotinic receptor antagonists as treatments for nicotine abuse. *Adv. Pharmacol.* **2014**, *69*, 513–551.

26. Gandini, M.A.; Sandoval, A.; Felix, R. Toxins targeting voltage-activated Ca2+ channels and their potential biomedical applications. *Curr. Top. Med. Chem.* **2015**, *15*, 604–616. [CrossRef] [PubMed]

27. Vetter, I.; Lewis, R.J. Therapeutic potential of cone snail venom peptides (conopeptides). *Curr. Top. Med. Chem.* **2012**, *12*, 1546–1552. [CrossRef] [PubMed]

28. Layer, R.T.; Mcintosh, J.M. Conotoxins: Therapeutic potential and application. *Mar. Drugs* **2006**, *4*, 119–142. [CrossRef]

29. Han, T.S.; Teichert, R.W.; Olivera, B.M.; Bulaj, G. *Conus* venoms-a rich source of peptide-based therapeutics. *Curr. Pharm. Des.* **2008**, *14*, 2462–2479. [CrossRef] [PubMed]

30. Olivera, B.M.; Teichert, R.W. Diversity of the neurotoxic *Conus* peptides: A model for concerted pharmacological discovery. *Mol. Interv.* **2007**, *7*, 251–260. [CrossRef]

31. Fedosov, A.E.; Moshkovskii, S.A.; Kuznetsova, K.G.; Olivera, B.M. Conotoxins: From the biodiversity of gastropods to new drugs. *Biomed. Khim.* **2013**, *59*, 267–294. [CrossRef]

32. Peng, C.; Yao, G.; Gao, B.M.; Fan, C.X.; Bian, C.; Wang, J.; Cao, Y.; Wen, B.; Zhu, Y.; Ruan, Z.; et al. High-throughput identification of novel conotoxins from the Chinese tubular cone snail (*Conus betulinus*) by multi-transcriptome sequencing. *Gigascience* **2016**, *5*, 17. [CrossRef] [PubMed]

33. Barghi, N.; Concepcion, G.P.; Olivera, B.M.; Lluisma, A.O. High conopeptide diversity in *Conus tribblei* revealed through analysis of venom duct transcriptome using two high-throughput sequencing platforms. *Mar. Biotechnol.* **2015**, *17*, 81–98. [CrossRef] [PubMed]

34. Dutertre, S.; Jin, A.H.; Kaas, Q.; Jones, A.; Alewood, P.F.; Lewis, R.J. Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol. Cell. Proteom.* **2013**, *12*, 312–329. [CrossRef]

35. Lavergne, V.; Harliwong, I.; Jones, A.; Miller, D.; Taft, R.J.; Alewood, P.F. Optimized deep-targeted proteotranscriptomic profiling reveals unexplored *Conus* toxin diversity and novel cysteine frameworks. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 3782–3791. [CrossRef] [PubMed]

36. Biass, D.; Dutertre, S.; Gerbault, A.; Menou, J.L.; Offord, R.; Favreau, P.; Stöcklin, R. Comparative proteomic study of the venom of the piscivorous cone snail *Conus consors*. *J. Proteome* **2009**, *72*, 210–218. [CrossRef]

37. Davis, J.; Jones, A.; Lewis, R.J. Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. *Peptides* **2009**, *30*, 1222–1227. [CrossRef]

38. Gao, B.; Peng, C.; Yang, J.; Yi, Y.; Zhang, J.; Shi, Q. Cone snails: A big store of conotoxins for novel drug discovery. *Toxins* **2017**, *9*, 397. [CrossRef] [PubMed]

39. Gao, B.; Peng, C.; Zhu, Y.; Sun, Y.; Zhao, T.; Huang, Y.; Shi, Q. High throughput identification of novel conotoxins from the Vermivorous Oak cone snail (*Conus quercinus*) by transcriptome sequencing. *Int. J. Mol. Sci.* **2018**, *19*, 3901. [CrossRef] [PubMed]

40. Zhangsun, D.; Luo, S.; Wu, Y.; Zhu, X.; Hu, Y.; Xie, L. Novel O-superfamily conotoxins identified by cDNA cloning from three vermivorous *Conus* species. *Chem. Biol. Drug Des.* **2006**, *68*, 256–265. [CrossRef] [PubMed]

41. Liu, L.; Wu, X.; Yuan, D.; Chi, C.; Wang, C. Identification of a novel S-superfamily conotoxin from vermivorous *Conus caracteristicus*. *Toxicon* **2008**, *51*, 1331–1337. [CrossRef]

42. Yuan, D.D.; Liu, L.; Shao, X.X.; Peng, C.; Chi, C.W.; Guo, Z.Y. Isolation and cloning of a conotoxin with a novel cysteine pattern from *Conus caracteristicus*. *Peptides* **2008**, *29*, 1521–1525. [CrossRef]

43. Yuan, D.D.; Liu, L.; Shao, X.X.; Peng, C.; Chi, C.W.; Guo, Z.Y. New conotoxins define the novel I3-superfamily. *Peptides* **2009**, *30*, 861–865. [CrossRef]

44. Luo, S.; Zhangsun, D.; Harvey, P.J.; Kaas, Q.; Wu, Y.; Zhu, X.; Hu, Y.; Li, X.; Tsetlin, V.I.; Christensen, S.; et al. Cloning, synthesis, and characterization of αO-conotoxin GeXIVA, a potent α9α10 nicotinic acetylcholine receptor antagonist. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 4026–4035. [CrossRef]

45. Xu, S.; Zhang, T.; Kompella, S.N.; Yan, M.; Lu, A.; Wang, Y.; Shao, X.; Chi, C.; Adams, D.J.; Ding, J.; et al. Conotoxin αD-GeXXA utilizes a novel strategy to antagonize nicotinic acetylcholine receptors. *Sci. Rep.* **2015**, *5*, 14261. [CrossRef] [PubMed]

46. Jiang, S.; Tae, H.S.; Xu, S.; Shao, X.; Adams, D.J.; Wang, C. Identification of a novel O-conotoxin reveals an unusual and potent inhibitor of the human α9α10 nicotinic acetylcholine receptor. *Mar. Drugs* **2017**, *15*, 170. [CrossRef] [PubMed]

47. Kits, K.S.; Lodder, J.C.; van der Schors, R.C.; Li, K.W.; Geraerts, W.P.; Fainzilber, M. Novel omega-conotoxins block dihydropyridine-insensitive high voltage-activated calcium channels in molluscan neurons. *J. Neurochem.* **1996**, *67*, 2155–2163. [CrossRef]

48. Terlau, H.; Shon, K.J.; Grilley, M.; Stocker, M.; Stuhmer, W.; Olivera, B.M. Strategy for rapid immobilization of prey by a fish-hunting marine snail. *Nature* **1996**, *381*, 148–151. [CrossRef] [PubMed]

49. Fainzilber, M.; Nakamura, T.; Lodder, J.C.; Zlotkin, E.; Kits, K.S.; Burlingame, A.L. gamma-Conotoxin-PnVIIA, a gamma-carboxyglutamate-containing peptide agonist of neuronal pacemaker cation currents. *Biochemistry* **1998**, *37*, 1470–1477. [CrossRef] [PubMed]

50. Fainzilber, M.; Gordon, D.; Hasson, A.; Spira, M.E.; Zlotkin, E. Mollusc-specific toxins from the venom of *Conus textile* neovicarius. *Eur. J. Biochem.* **1991**, *202*, 589–595. [CrossRef]

51. Pi, C.; Liu, J.; Peng, C.; Liu, Y.; Jiang, X.; Zhao, Y.; Tang, S.; Wang, L.; Dong, M.; Chen, S.; et al. Diversity and evolution of conotoxins based on gene expression profiling of *Conus litteratus*. *Genomics* **2006**, *88*, 809–819. [CrossRef]

52. Luo, S.; Zhangsun, D.; Zhang, B.; et al. Novel O-Superfamily conotoxins, and their coding polynucleotides and use. Patent CN (200410103561.0)-A, 30 December 2004.

53. Luo, S.; Zhangsun, D.; Feng, J.; Wu, Y.; Zhu, X.; Hu, Y. Diversity of the O-superfamily conotoxins from *Conus miles*. *J. Pept. Sci.* **2007**, *13*, 44–53. [CrossRef] [PubMed]

54. Hu, H.; Bandyopadhyay, P.K.; Olivera, B.M.; Yandell, M. Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. *BMC Genomics* **2011**, *12*, 60. [CrossRef] [PubMed]

55. Terrat, Y.; Biass, D.; Dutertre, S.; Favreau, P.; Remm, M.; Stöcklin, R.; Piquemal, D.; Ducancel, F. High-resolution picture of a venom duct transcriptome: case study with the marine snail *Conus consors*. *Toxicon* **2012**, *59*, 34–46. [CrossRef]

56. Hu, H.; Bandyopadhyay, P.K.; Olivera, B.M.; Yandell, M. Elucidation of the molecular envenomation strategy of the cone snail *Conus geographus* through transcriptome sequencing of its venom duct. *BMC Genomics* **2012**, *13*, 284. [CrossRef] [PubMed]

57. Kaas, Q.; Westermann, J.C.; Craik, D.J. Conopeptide characterization and classifications: An analysis using ConoServer. *Toxicon* **2010**, *55*, 1491–1509. [CrossRef] [PubMed]

58. Walker, C.S.; Jensen, S.; Ellison, M.; Matta, J.A.; Lee, W.Y.; Imperial, J.S.; Duclos, N.; Brockie, P.J.; Madsen, D.M.; Isaac, J.T.; et al. A novel *Conus* snail polypeptide causes excitotoxicity by blocking desensitization of AMPA receptors. *Curr. Biol.* **2009**, *19*, 900–908. [CrossRef] [PubMed]

59. Elliger, C.A.; Richmond, T.A.; Lebaric, Z.N. Diversity of conotoxin types from *Conus californicus* reflects a diversity of prey types and a novel evolutionary history. *Toxicon* **2011**, *57*, 311–322. [CrossRef] [PubMed]

60. Biggs, J.S.; Watkins, M.; Puillandre, N.; Ownby, J.P.; Lopez-Vera, E.; Christensen, S.; Moreno, K.J.; Bernaldez, J.; Licea-Navarro, A.; Corneli, P.S.; et al. Evolution of *Conus* peptide toxins: analysis of *Conus californicus* Reeve, 1844. *Mol. Phylogenet. Evol.* **2010**, *56*, 1–12. [CrossRef] [PubMed]

61. Milne, T.J.; Abbenante, G.; Tyndall, J.D.; Halliday, J.; Lewis, R.J. Isolation and characterization of a cone snail protease with homology to CRISP proteins of the pathogenesis-related protein superfamily. *J. Biol. Chem.* **2003**, *278*, 31105–31110. [CrossRef] [PubMed]

62. Kang, T.S.; Vivekanandan, S.; Jois, S.D.; Kini, R.M. Effect of C-terminal amidation on folding and disulfide-pairing of alpha-conotoxin ImI. *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 6333–6337. [CrossRef]

63. Romisch, K. Surfing the Sec61 channel: Bidirectional protein translocation across the ER membrane. *J. Cell Sci.* **1999**, *112*, 4185–4191.

64. Stock, S.D.; Hama, H.; DeWald, D.B. SECl4-dependent secretion in Saccharomyces cerevisiae Nondependence on sphingolipid synthesis-coupled diacylglycerol production. *J. Biol. Chem.* **1999**, *274*, 12979–12983. [CrossRef]

65. Jiang, Y.; Li, Y.; Lee, W. Venom gland transcriptomes of two elapid snakes (*Bungarus multicinctus* and *Naja atra*) and evolution of toxin genes. *BMC Genomics* **2011**, *12*, 1. [CrossRef]

66. Ciechanover, A.; Iwai, K. The ubiquitin system: From basic mechanisms to the patient bed. *IUBMB Life* **2004**, *56*, 193–201. [CrossRef] [PubMed]

67. Abalde, S.; Tenorio, M.J.; Afonso, C.M.L.; Zardoya, R. Conotoxin diversity in *Chelyconus ermineus* (Born, 1778) and the convergent origin of piscivory in the Atlantic and Indo-Pacific cones. *Genome Biol. Evol.* **2018**, *10*, 2643–2662. [CrossRef]

68. Li, R.; Li, Y.; Kristiansen, K.; Wang, J. SOAP: Short oligonucleotide alignment program. *Bioinformatics* **2008**, *24*, 713–714. [CrossRef] [PubMed]

69. Xie, Y.; Wu, G.; Tang, J.; Luo, R.; Patterson, J.; Liu, S.; Zhou, X.; Lam, T.W.; Li, Y.; Xu, X.; et al. SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. *Bioinformatics* **2014**, *30*, 1660–1666. [CrossRef]

70. Mortazavi, A.; Williams, B.A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* **2008**, *5*, 621–628. [CrossRef]

71. Birney, E.; Clamp, M.; Durbin, R. GeneWise and Genomewise. *Genome Res.* **2004**, *14*, 988–995. [CrossRef] [PubMed]

72. Keller, O.; Kollmar, M.; Stanke, M.; Waack, S. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **2011**, *27*, 757–763. [CrossRef]

73. Woodward, S.R.; Cruz, L.J.; Olivera, B.M.; Hillyard, D.R. Constant and hypervariable regions in conotoxin propeptides. *EMBO J.* **1990**, *9*, 1015–1020. [CrossRef] [PubMed]

74. Olivera, B.M.; Walker, C.; Cartier, G.E.; Hooper, D.; Santos, A.D.; Schoenfeld, R.; Shetty, R.; Watkins, M.; Bandyopadhyay, P.; Hillyard, D.R. Speciation of cone snails and interspecific hyperdivergence of their venom peptides. Potential evolutionary significance of introns. *Ann. N.Y. Acad. Sci.* **1999**, *870*, 223–237. [CrossRef] [PubMed]

75. Zamora-Bustillos, R.; Aguilar, M.B.; Falcón, A. Identification, by molecular cloning, of a novel type of I2-superfamily conotoxin precursor and two novel I2-conotoxins from the worm-hunter snail *Conus spurius* from the Gulf of México. *Peptides* **2010**, *31*, 384–393. [CrossRef]

76. Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M.C.; Estreicher, A.; Gasteiger, E.; Martin, M.J.; Michoud, K.; O'Donovan, C.; Phan, I.; et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **2003**, *31*, 365–370. [CrossRef] [PubMed]

77. Tatusov, R.L.; Galperin, M.Y.; Natale, D.A.; Koonin, E.V. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **2000**, *28*, 33–36. [CrossRef] [PubMed]

78. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676. [CrossRef]

79. Huelsenbeck, J.P.; Ronquist, F.R. MrBayes: Bayesian inference of phylogenic trees. *Bioinformatics* **2001**, *17*, 754–755. [CrossRef] [PubMed]

80. Posada, D.; Crandall, K.A. Selecting the best-fit model of nucleotide substitution. *Syst. Biol.* **2001**, *50*, 580–601. [CrossRef]

81. Swofford, D.L. *PAUP: Phylogenetic Analysis Using Parsimony (* And Other Methods)*; Version 4.0; Sinauer: Sunderland, MA, USA, 2001.