



Article

# High Throughput Identification of Novel Conotoxins from the Vermivorous Oak Cone Snail (*Conus quercinus*) by Transcriptome Sequencing

Bingmiao Gao <sup>1,2,3,†</sup> , Chao Peng <sup>2,†</sup>, Yabing Zhu <sup>4,†</sup>, Yuhui Sun <sup>4,5</sup>, Tian Zhao <sup>6</sup>, Yu Huang <sup>2,7,\*</sup> and Qiong Shi <sup>2,7,\*</sup>

<sup>1</sup> Hainan Provincial Key Laboratory of Research and Development of Herbs, College of Pharmacy, Hainan Medical University, Haikou 571199, China; gaobingmiao@hainmc.edu.cn

<sup>2</sup> 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

<sup>3</sup> Institute for Molecular Bioscience, The University of Queensland, St. Lucia, Brisbane, QLD 4072, Australia

<sup>4</sup> BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China; zhuyabing@genomics.cn (Y.Z.); sunyuhui@genomics.cn (Y.S.)

<sup>5</sup> Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

<sup>6</sup> Chemistry Department, College of Art and Science, Boston University, Boston, MA 02215, USA; domzhao@bu.edu

<sup>7</sup> BGI Education Center, University of Chinese Academy of Sciences, Shenzhen 518083, China

\* Correspondence: huangyu@genomics.cn (Y.H.); shiqiong@genomics.cn (Q.S.); Tel.: +86-755-3630 7807 (Q.S.)

† These authors contributed equally to this work.

Received: 2 November 2018; Accepted: 3 December 2018; Published: 5 December 2018



**Abstract:** The primary objective of this study was to realize the large-scale discovery of conotoxin sequences from different organs (including the venom duct, venom bulb and salivary gland) of the vermivorous Oak cone snail, *Conus quercinus*. Using high-throughput transcriptome sequencing, we identified 133 putative conotoxins that belong to 34 known superfamilies, of which nine were previously reported while the remaining 124 were novel conotoxins, with 17 in new and unassigned conotoxin groups. A-, O<sub>1</sub>-, M-, and I<sub>2</sub>- superfamilies were the most abundant, and the cysteine frameworks XIII and VIII were observed for the first time in the A- and I<sub>2</sub>-superfamilies. The transcriptome data from the venom duct, venom bulb and salivary gland showed considerable inter-organizational variations. Each organ had many exclusive conotoxins, and only seven of all the inferred mature peptides were common in the three organs. As expected, most of the identified conotoxins were synthesized in the venom duct at relatively high levels; however, a number of conotoxins were also identified in the venom bulb and the salivary gland with very low transcription levels. Therefore, various organs have different conotoxins with high diversity, suggesting greater contributions from several organs to the high-throughput discovery of new conotoxins for future drug development.

**Keywords:** cone snail; conotoxin; *Conus quercinus*; transcriptome; venom duct; venom bulb; salivary gland

## 1. Introduction

Cone snails (*Conus* spp.) are venomous predators living in tropical marine waters all over the world. They rapidly immobilize their prey using a complex cocktail of small and disulfide-rich peptides, usually 7–46 amino acids (aa) in length and collectively known as conotoxins or conopeptides [1,

2]. Each cone snail is believed to synthesize 50–200 such peptides, and it is estimated that more than 80,000 different conotoxins may exist in approximately 800 species of cone snails around the world [2–4]. Most of these conotoxins can selectively target certain voltage-gated ion channels, ligand-gated ion channels, G-protein coupled receptors and neurotransmitter transporters in the central and peripheral nervous systems [5–7]. Thus, these conotoxins are widely used as pharmacological tools in neuroscience research, and some have already demonstrated therapeutic potential in preclinical and clinical trials [7]. The most famous,  $\omega$ -MVIIA (Ziconotide) is derived from the venom of a fish-hunting species *C. magus*, and has been approved by the American Food and Drug Administration (FDA) for treatment of chronic pain in patients with cancer or acquired immune deficiency syndrome (AIDS) [8,9].

Although over 80,000 different conotoxins may exist, to date less than 0.1 percent of them have been characterized from 100 *Conus* species by traditional approaches [3,10,11]. However, these methods, which are used to isolate and identify the potential bioactive substances directly from venoms are now widely considered to be time-consuming, as having low sensitivity and to be limited by sample availability. In contrast, high-throughput sequencing can achieve higher sequencing depth (amplification with more sequencing reads) and greater coverage of transcriptomes, so that even rare transcripts with low transcription levels can be identified [12]. Therefore, several recent studies using next-generation sequencing technologies on the venom duct transcriptomes of *Conus* species have uncovered about 3000 conotoxin genes in 14 cone snails [1,2,13–25]. Altogether, these transcriptome studies show that each *Conus* species produces at least 100–400 different conotoxins, which has been very beneficial for the discovery of novel superfamilies and new cysteine frameworks.

To date, only 41 conotoxins have been previously identified in the Oak cone snail (*C. quercinus*) using traditional methods, and they were classified into 6 superfamilies (A, M, T, Q, O<sub>1</sub> and K) [26–28]. To further contribute to the cataloguing of conotoxin diversity in the main lineages of cone snails, we employed the high-throughput transcriptome sequencing approach to explore additional conotoxins from the transcriptomes of different organs, including the venom duct, venom bulb and salivary gland in *C. quercinus*. For simplification, the three transcriptomes from the different organs were named as “VB” (venom bulb), “VD” (venom duct), and “SG” (salivary gland) datasets.

## 2. Results

### 2.1. Assembly of Transcriptome Sequences

Our high-throughput sequencing generated approximately 29.45, 30.18 and 30.78 million paired-end short reads with a length of 150 bp for VD, SG and VB, respectively. After removal of low-quality reads and adapter sequences, we obtained 4.04, 4.01 and 4.10 gigabases (Gb) of clean data for each sample, and the clean data ratio was estimated to be 91.48%, 88.63% and 88.83%, respectively (Table 1). De novo assembling of these clean reads produced 171 k, 225 k and 124 k contigs for VD, SG and VB, respectively, and N50 of each assembly was calculated to be 586, 669 and 651 bp in size. After clustering contigs into consensus unigenes based on sequence similarity, the final assembly of each transcriptome contained 91 k, 113 k and 66 k unigenes for VD, SG and VB, respectively, with similar GC contents (44.21%, 44.87% and 44.72%; Table 1, Supplementary Figure S1).

To investigate the transcription levels of genes including conotoxins in each sample, we mapped the clean reads back to the de novo assembled unigenes (Supplementary Table S1) and estimated the abundance of each gene. There were 72 k, 85 k and 61 k transcribed genes in VD, SG and VB, respectively, (Supplementary Table S2) in which 33.49% (24,327 genes), 21.97% (18,801 genes) and 50.06% (30,676 genes; Supplementary Figure S2) had high transcription levels (values of fragments per kilobase of exon per million fragments mapped (FPKM)  $\geq 10$ ; see more details in Section 2.4).

Subsequently, functional annotations of the assembled transcripts were realized by blasting against several public databases and a total of 30,830 (23.11%) (Supplementary Tables S3 and S4) were successfully annotated in at least one database, including 8136 (6.10%) genes with common hits in all databases (Supplementary Figure S3). There were 11,719 genes enriched into 44 KEGG (Kyoto

Encyclopedia of Genes and Genomes) pathways, with signal transduction being the most enriched one (Supplementary Figure S4). Another 11,323 genes were assigned into 57 GO (Gene Ontology) terms, which can be classified into three categories, including biological process, cellular component and molecular function (Supplementary Figure S5).

**Table 1.** Sequencing statistics and assembly summary for the venom duct (VD), salivary gland (SG) and venom bulb (VB).

Samples	VD	SG	VB
<b>Raw data</b>			
Total Reads	29,451,202	30,184,621	30,775,070
Total length (bp)	4,417,680,300	4,527,693,150	4,616,260,500
Read length (bp)	150	150	150
<b>Clean data</b>			
Total Reads	28,865,798	28,664,492	29,290,756
Total length (bp)	4,041,211,720	4,013,028,880	4,100,705,840
Read length (bp)	140	140	140
Clean data ratio	91.48%	88.63%	88.83%
<b>Contigs</b>			
Total Number	171,606	225,404	124,936
Total Length (bp)	80,150,026	111,578,962	61,180,077
Mean Length (bp)	467	495	489
N50 (bp)	586	669	651
N70 (bp)	331	353	349
N90 (bp)	212	214	215
GC Content	43.70%	44.29%	44.32%
<b>Unigenes</b>			
Total Number	91,392	113,472	66,549
Total Length (bp)	53,668,190	72,145,553	41,085,333
Mean Length (bp)	587	635	617
N50 (bp)	800	934	884
N70 (bp)	436	485	464
N90 (bp)	254	262	259
GC Content	44.21%	44.87%	44.72%

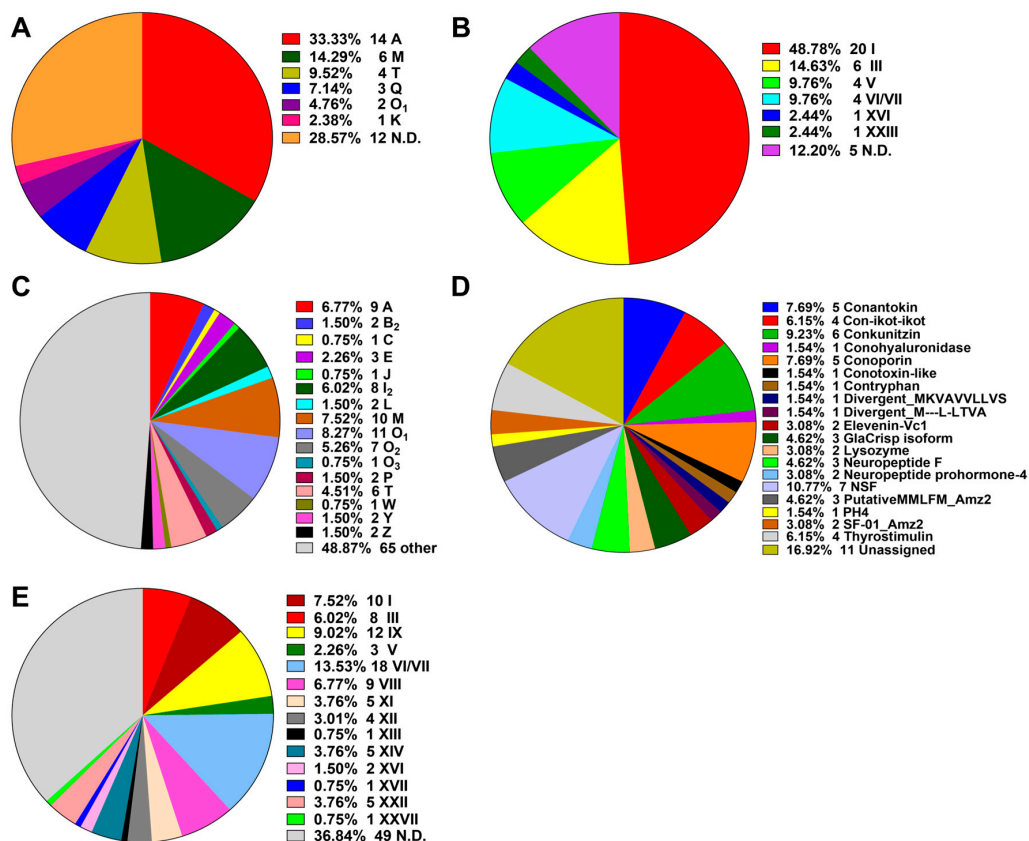
## 2.2. Summary of Conotoxins in the Three Transcriptomes

The putative conotoxin sequences were predicted by BLASTX search and HMMER analysis [24] against a local reference database of known conotoxins from the ConoServer database [29], and then checked manually using the ConoPrec tool [30]. After removal of the transcripts with duplication or truncated mature region sequences, we obtained a total of 65, 55 and 52 putative conotoxin transcripts for the three datasets of VD, SG and VB, respectively (Supplementary Tables S5–S7). We then combined the three conotoxin datasets into the total transcript dataset of *C. quercinus* (Qc-transcript) and named the predicted 133 conotoxins as Qc-001 to Qc-133 (Supplementary Table S8). They have at least one aa difference between each other in the mature regions. The 41 conotoxins previously identified by traditional methods in *C. quercinus* [26–28] were classified into six superfamilies, including A, M, T, Q, O<sub>1</sub> and K (Figure 1A; Supplementary Table S9); 14 of these conotoxins belong to the A-superfamily, six belong to the M-superfamily and four belong to the T-superfamily, which account for 33.33%, 14.29% and 9.52%, respectively.

In the present study, each putative conotoxin was assigned to a superfamily based on percentage of sequence identity to the highly conserved signal regions of the known superfamilies in the ConoServer database (see more details in Section 4). Here, 133 novel putative conotoxins in *C. quercinus* were assigned to 34 known conotoxin superfamilies (Figure 1, Supplementary Table S8), in which 17 were described in the ConoServer database and the rest were identified by serial BLAST homology searches. In addition, seven belonged to a new superfamily (NSF) and 11 were unassigned to any superfamily. However, 68 conotoxins belonged to the majority of superfamilies (A, B<sub>2</sub>, C, E, J, I<sub>2</sub>, O<sub>1</sub>, O<sub>2</sub>, O<sub>3</sub>, L, M, P,

T, W, Y and Z; Figure 1C), and 65 were classified into other families such as conkunitzin, conoporphin and con-ikot-ikot (Figure 1D). Interestingly, we discovered 12 new superfamilies (B<sub>2</sub>, C, E, I<sub>2</sub>, J, O<sub>2</sub>, O<sub>3</sub>, L, P, W, Y and Z) from the Qc-transcript, which have never been reported in this species before, although neither the Q- or K-superfamily was discovered. Similar to the ConoServer database, A-, M-, O-, and I<sub>2</sub>-superfamilies accounted for the major proportion of venom conotoxins in *C. quercinus* (Figure 1C).

The cysteine patterns of conotoxins in *C. quercinus* are summarized in Figure 1B,E. The 41 conotoxins previously identified by traditional methods have six types of cysteine frameworks, including I (CC-C-C), III (CC-C-C-CC), V (CC-CC), VI/VII (C-C-CC-C-C), XVI (C-C-CC) and XXIII (C-C-C-CC-C). In addition, other frameworks such as C-C, CC-C and C-C-C, were also identified. In contrast, except for the cysteine frameworks determined in the 41 conotoxins, nine other types of cysteine frameworks were also discovered from the other 64 conotoxins, including IX (C-C-C-C-C-C), VIII (C-C-C-CC-C-C-C-C), XI (C-C-CC-CC-C-C), XII (C-C-C-C-CC-C-C), XIII (C-C-C-CC-C-C-C), XIV (C-C-C-C), XVII (C-C-CC-C-CC-C), XXII (C-C-C-C-C-C-C-C) and XXVII (C-CC-C-C-C), respectively. Meanwhile, other limited frameworks, C-C, CC-C and C-C-C, were determined as well.

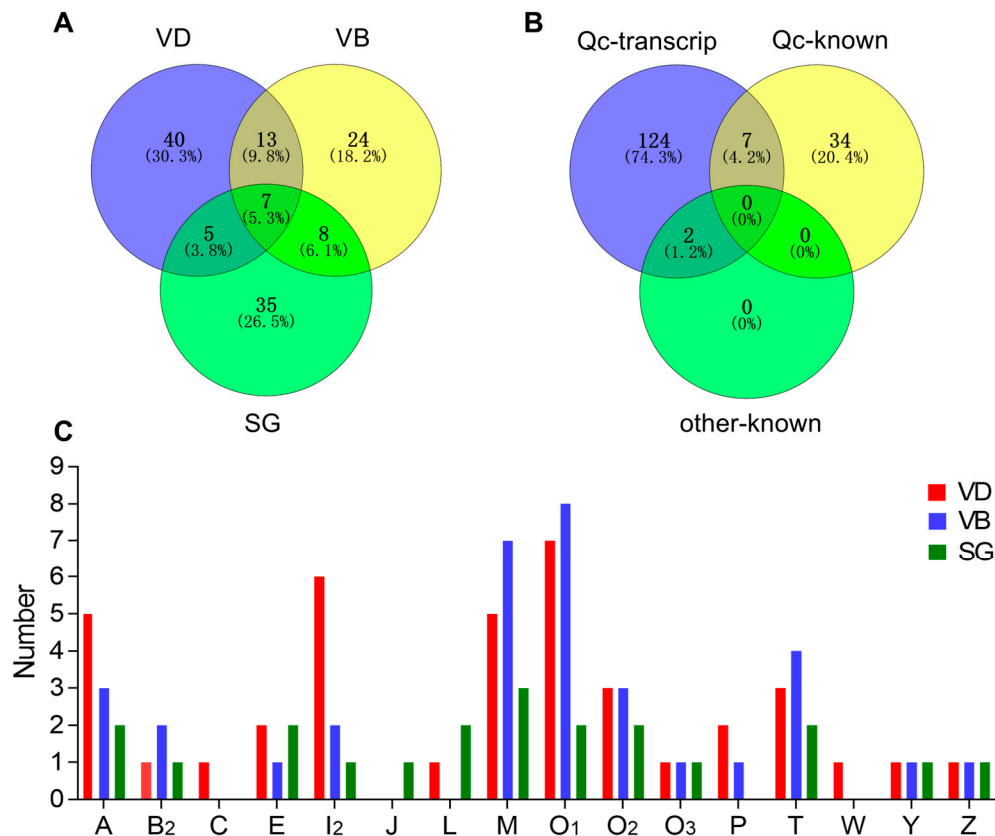


**Figure 1.** Summary of the superfamilies and cysteine patterns of conotoxins in *C. quercinus*. (A) The superfamilies reported previously. (B) The cysteine patterns reported previously. (C) Total superfamilies identified in this study. (D) Subdivision of the conotoxins summarized from the 'Other' group in (C) into further categories. (E) The cysteine patterns identified in this study.

### 2.3. Comparison of Conotoxins in the Three Transcriptomes

The comparative distribution of putative conotoxins is summarized in Figure 2A. Of these putative conotoxins, seven were identified in all the three organs (VD, SG and VB), 13 were common to both VD and VB (but not in SG), 8 were shared by VB and SG (not in VD), and 5 were common to both VD and SG (not in VB). Among the 41 previously published conotoxin sequences of *C. quercinus* (Qc-known; Figure 2B and Supplementary Table S9), seven (Qc1.1b/Qc-072, Qc1.2/Qc-073,

Qc3.1/Qc-077, Qc3-IP01/-Qc-100, Qc1.16/Qc-104, Qc5.3/Qc-105, and Qc5.2/Qc-118) were recovered in our present transcript dataset of *C. quercinus* (Qc-transcript; Figure 2B and Supplementary Table S8). In addition, only two (Qc-056 and Qc-093) of the 133 identified conotoxins have the same mature region sequences as previously reported in other *Conus* species (*C. lividus* and *C. characteristicus*; other-known in Figure 2B), and they are both classified in the M-superfamily (Lv3-IP02 and Qc3-HGS02, respectively) [31].



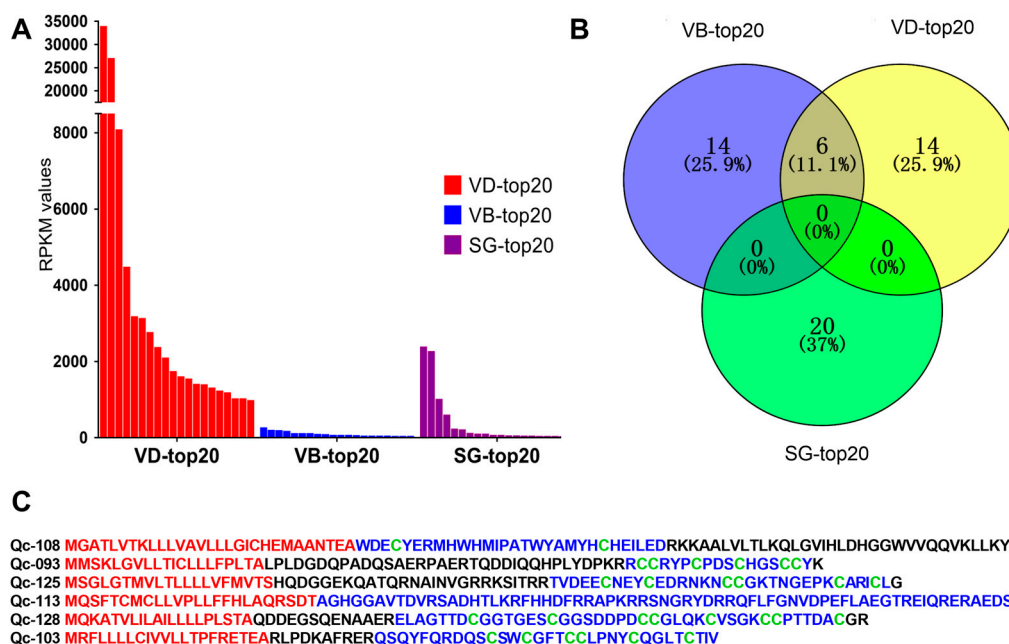
**Figure 2.** Comparison of conotoxin transcripts from various *C. quercinus* datasets. (A) Relationship of the identified conotoxins from the VD, VB and SG datasets. (B) Comparison among the total identified conotoxins from *C. quercinus* (Qc-transcript), the conotoxins reported previously in *C. quercinus* (Qc-known), and those from other cone snails (other-known). (C) Subdivision of 16 conotoxin superfamilies from our VD, VB and SG datasets.

Figure 2C shows the differences in spatial distribution of different conotoxin superfamilies among the VD, VB, and SG datasets of *C. quercinus*. Conotoxin sequences of A-, B<sub>2</sub>-, E-, I<sub>2</sub>-, M-, O<sub>1</sub>-, O<sub>2</sub>-, O<sub>3</sub>-, T-, Y- and Z-superfamilies were identified in VD, VB and SG, and A-, O<sub>1</sub>-, M- and I<sub>2</sub>-superfamilies were the most predominant conotoxin groups in *C. quercinus* (Figures 1 and 2C). In addition, the J-superfamily was identified only in SG, not in VD and VB; C-, P- and M-superfamilies were not found in SG, and C-, J-, L-, and W-superfamilies were not found in VB (Figure 2C). Therefore, the venom duct, venom bulb and salivary gland had different conotoxins with high diversity. However, our results confirm that the venom duct is the most high-throughput factory for organizational manufacture of conotoxins.

#### 2.4. Differential Transcription of Conotoxins in Different Organs

FPKM values were calculated to represent the transcription levels of conotoxins. We discovered that different organs have differential transcription profiles of toxins. The top 20 putative conotoxins (with the highest FPKM values) were selected from each of the three datasets for comparison.

We observed that transcription levels in the VD (RPKM values above 961.5) were generally higher than those in the SG and VB (Figure 3A, Supplementary Table S10). In contrast, the highest RPKM values in the SG-top20 and VB-top20 datasets were 2364.27 and 247.12, respectively.



**Figure 3.** The top 20 conotoxins with the highest values of fragments per kilobase of exon per million fragments mapped (FPKM) from the three transcriptome datasets. (A) Comparison of the FPKM values for individual conotoxins within each dataset. (B) A Venn diagram of the top 20 conotoxins from each of the VD, VB and SG datasets. (C) Comparison of six conotoxins sequences found in the VD-top20 and VB-top20. The signal regions predicted by the ConoPrec tool are marked in red, and the mature regions (in blue) and cysteine residues (in green) are presented for comparison.

Among the top 20 conotoxins of the three datasets, six putative conotoxins were found in the lists of the VD-top20 and VB-top20, but not in the SG-top20 list (Figure 3B). Interestingly, these six putative conotoxins with high transcription levels belonged to different families with various cysteine patterns (Figure 3C). For example, Qc-108 with a framework of C-C belongs to an unassigned superfamily, although it was similar to the new Superfamily-4 conotoxin G127-VD from *C. geographus* [32] through sequence alignment (the identity is about 85%). Qc-093 with the framework III (CC-C-C-CC) belongs to the M-superfamily, identical to the MMSK group conotoxin Qc3-HGS02 from *C. characteristicus* [31], and sharing 93% identity with the M-superfamily MMSK group conotoxin Qc3-YGS01 from *C. quercinus* [31]. Qc-126 with a framework of VI/VII (C-C-CC-C-C) belongs to the O<sub>3</sub>-superfamily, showing 81% identity with the O<sub>3</sub>-superfamily conotoxin Tr7.4 from *C. terebra*. This is the only O<sub>3</sub>-superfamily conotoxin identified in *C. quercinus*. Qc-114 with non-cysteine belongs to the conantokin superfamily, sharing 75% identity with the conantokin-F Fla-7 from *C. flavidus*. Qc-129 with the framework XVII (C-C-CC-C-CC-C) belongs to the Y-superfamily, sharing 70% identity with a conotoxin from *C. betulinus*. Qc-103 with the framework VI/VII(C-C-CC-C-C) belongs to Divergent\_ M-L-LTVA, which showed 83% identity with a conotoxin from *C. betulinus* [24].

As expected, it appears that most of the identified conotoxins were synthesized at high levels in the venom duct, although some conotoxins were also identified in the venom bulb and the salivary gland at low levels. Obviously, these putative conotoxins from various organs were variable in diversity.

### 2.5. Diversity of Conotoxin Superfamilies

The conotoxin mixture in *C. quercinus* consisted of 133 putative conotoxins, which were classified into 34 known superfamilies, seven NSF superfamilies, and eleven unassigned superfamilies.

Sequences of the A-, M-, O<sub>1</sub>- and I<sub>2</sub>-superfamilies were among the most abundant conotoxin groups in *C. quercinus* (Figures 1 and 3). Despite the observed high diversity, D-, F-, G-, H-, N-, S-, and V-superfamilies were not identified in *C. quercinus*.

A-superfamily conotoxins contain six cysteine patterns (I, II, IV, VI/VII, XIV, and XXII), and they act on nicotinic acetylcholine receptors (nAChRs), the GABAB receptor and  $\alpha$ 1-adrenoceptor [33]. Here, we determined that nine conotoxins belonged to this family, with eight having the cysteine pattern I and one having the cysteine pattern XIII (Figure 4). Qc-072, Qc-073 and Qc-104 are identical to the previously identified Qc1.1b, Qc1.2 and Qc1.16 from *C. quercinus*, respectively [26–28]. Qc-109 is the first A-superfamily conotoxin with the cysteine framework XIII (8 cysteine residues separated by 7 loops: C-C-C-CC-C-C-C). De13.1 from *C. delessertii* with framework XIII has been characterized as a member of the G-superfamily [34]. Qc-109 shared very low sequence similarity with De13.1, but does not belong to the G-superfamily; which therefore suggests it might have exciting prospects as a functionally novel conotoxin.

```

Qc1.18  MGMRMMFTMFLLVVLAITVVSFTSDHASDGRNTAAND--KASDLMAL--RDGCCSSPCS--VNNPDICGGGR----- : 67
Qc-039  MGMRMMFTMFLLVVLAITVVSFTSDHASDGRNTAAND--KASDLMAL--RDGCCSNPCS--VNNPDICGGGR----- : 67
Qc-074  MGMRMIFTVFLVVAITVVSFTSDHAFDGRNAAAND--KASDLMALAVRQCCSNPTCA--VSNPDICGGGR----- : 69
Qc-009  MGMRMIFTVFLVVAITVVSFTLDRASNGRDAADSD--KAADQIAQTVRDECCSNPCS--VSNPDICGGGR----- : 69
Qc-054  MGMRMIFTVFLVVAITVVS---DRVSNGRKAAAKF--KAPALMELSVRQCCSDPCS--VNNPDICSGGR----- : 66
Qc1.16*/Qc-104  MGMRVMFTFLVAVLSTTVVSFTLDRASNGRDAADSD--KAADQIAQTVRDECCSNPCS--QTHPEICRRTLMLQNLNHDMSPSA : 83
Qc1.1b*/Qc-072  MGMRMMFTMFLLVVLAITVVSFTSDHASDGRNTAAND--KASKLMAL--RNECCDNPPCK--SSNPDICDWRS----- : 67
Qc1.2*/Qc-073  MGMRMIFTVFLVVAITVVSFTLDRASKGRNAAADD--KPSDWIALA-IKQCCANPPCK--HVN---CR----- : 62
Qc-010  MGMRMMFTMFLLVVLAITVVSFTSDHASDGRNTAAND--KASNLMAL--RDGCCSNPCK--ASNPDICDWRS----- : 67
Qc-109  MGMRMMFTMFLLVVLAITVVSFTSDHGKSIKSSCFVNWRQSVFVALDRRVLYCTISLYCFAVKCCTGLCSVLLYCSELYCVIF--- : 84

```

**Figure 4.** Alignment of the achieved A-superfamily sequences from our *C. quercinus* transcriptomes. \* represents the conotoxins that have been reported previously. The signal regions are highlighted in yellow, the mature regions are underlined, and cysteine residues (in red) are marked for comparison.

M-superfamily conotoxins have eight cysteine patterns (I, II, III, IV, VI/VII, IX, XIV and XVI), and include three distinct pharmacological families, including  $\mu$ -conotoxins,  $\kappa$ M-conotoxins, and  $\psi$ -conotoxins for blocking voltage-gated sodium channels, voltage-gated potassium channels and nicotinic acetylcholine receptors, respectively [35]. However, in this study, all the related eight conotoxins presented with the cysteine pattern III (Figure 5). Qc-077 and Qc-100 were identical to Qc3.1 and Qc3-IP01 from *C. quercinus*, respectively. Qc-056 and Qc-093 were also the same as Lv3-IP02 from *C. ermineus* and Qc3-HGS02 from *C. characteristicus*, respectively. In addition, the cysteine patterns of Qc-033 and Qc-132 were C-C.

```

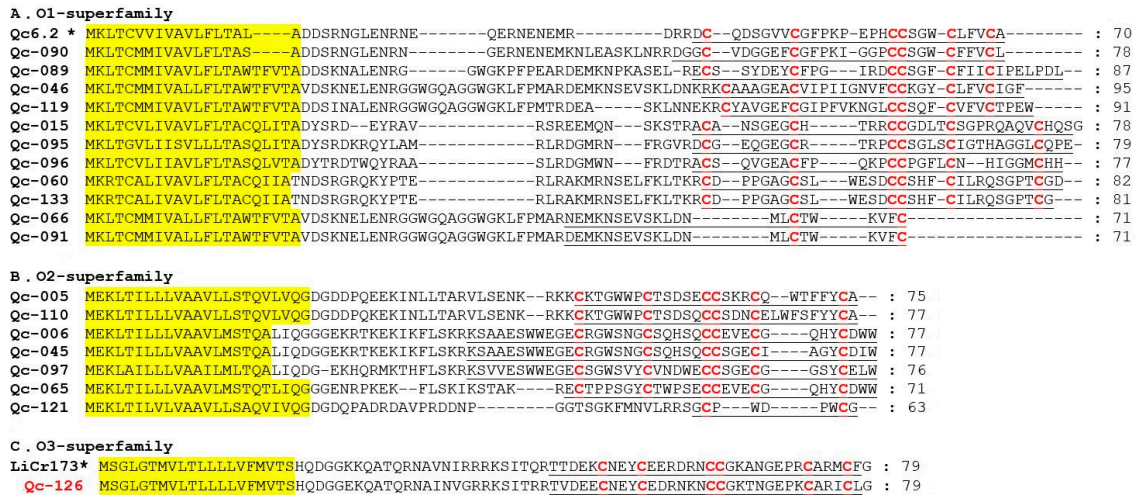
Qc3-TD01*  -MLKMGVLFVFLVFLPLATLQLDADQP----VARYAENKQDFNPNERMKIMLSALRQRECCEPSWCDAGCTDGCC----- : 71
Qc-062    -MLKMGVLFVFLVFLPLATLQLDADQP----VARYAENKQDFNPNERMKTMLSALRQRECCEPSWCDAGCYDGCC----- : 71
Qc3.1*/Qc-077  -MLKMGVLFVFLVFLPLATLQLDADQP----AARYAENKQDFNPNERKEMMLSALRQRACCDPDWCDAGCYDGCC----- : 71
Qc-023    MMSKLGVLITICLLVFLPLATLQLDGDQP----ADRPAERKQDISSEQYR---LFDQRQSRCCQ-WFCGYRCL--CCRVG-- : 69
Qc3-IP01*/Qc-100  MMSKLGVLITICLLVFLPLAVQLDGDQP----VDLPALRTQDFAPEHSP---WFNPVK-RCCS-RHC-WVCIP-CCPNGSP : 70
Lv3-IP02*/Qc-056  MMSKLGVLITICLLVFLPLAVQLDGDQP----VDLPALRTQDFAPEHSP---WFNPVK-RCCS-RHC-WVCIP-CCPNGSA : 70
Qc3-HGS02*/Qc-093  MMSKLGVLITICLLVFLPLALPLDGDQPADQSAERPAERTQD-DIQQHP---LYDPKR-RCCR-YFCPDSCHGSCCYK--- : 72
Qc-042    MMSKLGVLITICLLVFLPLALPLDGDQPADQSAERPAERTQDISSEQYR---KFDQRQ-RCCR-WFCPDSCHGSCCYK--- : 73
Qc-092    MMSKLGVLITICLLVFLPLALRLVGDQP----AERPAERTQDISSEQYR---KFDQRQ-RCCQ-WFCPGSCR--CCRYR-- : 68
Qc-033    -MSKLGAVLFTFLVFLPLAALQPVRDQP----ADRNAEPRGKIRNGMHG--FWPYMFLGHPGHHRHASWACP-WCG----- : 68
Qc-132    -MSKLGAVLFTFLVFLPLAALQPVRDQP----ADRNAEPRGKIRNGMHG--FWPYM-FGHPGHHRHASWACP-WCG----- : 67

```

**Figure 5.** Alignment of the achieved M-superfamily sequences from our *C. quercinus* transcriptomes. \* represents the conotoxins that have been reported previously. The signal regions are highlighted in yellow, the mature regions are underlined, and cysteine residues (in red) are marked for comparison.

O-superfamily members are unusually hydrophobic peptides with the same cysteine framework of C-C-CC-C-C, which provides these peptides with a similar inhibitory cystine knot (ICK) motif [36]. The O-superfamily is the most predominant superfamily in terms of conotoxin numbers (18 peptides; including O<sub>1</sub>, O<sub>2</sub> and O<sub>3</sub>), but only with the VI/VII and C-C patterns of cysteine frameworks (Figure 6). Eleven O<sub>1</sub>-superfamily sequences were identified in our present study (Figure 6A), in which nine presented the framework VI/VII and two showed the C-C pattern. Seven O<sub>2</sub>-superfamily conotoxins were identified (Figure 6B), in which six presented with framework VI/VII and one possessed the C-C

pattern. Qc-121 is 92% identical to the contryphan-R from *C. radiatus* [37]. Signal peptide sequence of Qc-126 indicated that this conotoxin was related to the O<sub>3</sub>-superfamily (Figure 6C). Interestingly, only one O<sub>3</sub>-superfamily was identified as sharing the cysteine framework VI/VII from *C. quercinus* [38]. Qc-126 showed 72% identity and 100% similarity to the LiCr173 from *C. lividus* [39].



**Figure 6.** Alignment of the achieved O-superfamily sequences from our *C. quercinus* transcriptomes. (A) The sequences of O<sub>1</sub>-superfamily. (B) The sequences of O<sub>2</sub>-superfamily. (C) The sequences of O<sub>3</sub>-superfamily. \* represents the conotoxins that have been reported previously. The signal regions are highlighted in yellow, the mature regions are underlined, and cysteine residues (in red) are marked for comparison.

I<sub>2</sub>-superfamily peptides have three cysteine frameworks in the primary sequences (XI, XII, and XIV), with the characteristic pattern of C-C-CC-CC-C-C. They are longer than those belonging to other *Conus* peptide families [40]. Eight I<sub>2</sub>-superfamily conotoxins were identified in the present study (Figure 7A), in which five presented with the framework XI, one owned the framework XII, and two possessed the framework VIII. Interestingly, framework VIII is often present in the B- and S-superfamilies, but Qc-082 and Qc-083 with the framework VIII were not similar to those conotoxins of the B- and S-superfamilies [41,42]. In fact, Qc-082 and Qc-083 are the first I<sub>2</sub>-superfamily conotoxins with the type VIII cysteine framework C-C-C-C-C-C-C-C-C-C. Meanwhile, six T-superfamily conotoxins were also identified (Figure 7B), in which three presented with the framework V (CC-CC), one owned the framework I (CC-C-C), and two possessed the framework XVI. For example, Qc-105 and Qc-118 showed 100% identity with Qc5.3 and Qc5.2 from *C. quercinus*, respectively.



**Figure 7.** Alignment of the achieved I<sub>2</sub>- and T-superfamily sequences from our *C. quercinus* transcriptomes. (A) The sequences of I<sub>2</sub>-superfamily. (B) The sequences of T-superfamily. \* represents the conotoxins that have been reported previously. The signal regions are highlighted in yellow, the mature regions are underlined, and cysteine residues (in red) are marked for comparison.



In this study, besides these major superfamilies of conotoxins, many others of B<sub>2</sub>-, C-, E-, P-, J-, L-, W-, Y-, Z-superfamilies and conantokin family were also discovered in our *C. quercinus* transcriptomes. For instance, one B<sub>2</sub>-superfamily sequence (Qc-088) with non-cysteine, a C-superfamily sequence (Qc-101) with the framework C-C, a W-superfamily sequence (Qc-112) with non-cysteine, and one J-superfamily sequence (Qc-002) with a framework of CC-C were identified in the present work. Two P-superfamily sequences with the framework IX (Qc-124 and Qc-128) were identified in the venom gland. However, to date only nine P-superfamily conotoxins have been reported previously, and they all owned the framework IX (C-C-C-C-C) [43]. Meanwhile, three E-superfamily sequences with the framework XXII (Qc-024, Qc-075 and Qc-076), two Y-superfamily sequences with the framework CC-C and XVII respectively (Qc-032 and Qc-129), two Z-superfamily sequences with the framework C-C (Qc-037 and Qc-111), and two L-superfamily sequences with the framework XIV (Qc-014 and Qc-123) were also identified in our current study.

Within the new superfamilies (NSF), seven conotoxins showing divergent signal regions were assigned to seven new conotoxin groups (Figure 8A). However, considering the minimal similarity of these sequences in the signal regions, we are not clear whether these groups should be recognized as distinct superfamilies or subgroups within a superfamily. In addition, within the unassigned superfamilies, there are eleven sequences without any known superfamily (Figure 8B).

**A. NSF-superfamily**

```

Qc-004 --MEALTIFRCLLAALISSVVVAPLKEKVSEQEDECVPVSGGSNPLRD-----CMRA-----CLTTSTY : 58
Qc-113 MEMEALTIFRCLLAALISSVVVAPLKEKVSEQEDECVPVSGGSNPLRD-----CMRA-----CLTTSTV : 60
Qc-050 --MGSL-YTPCLASILIMSDSNARTFPSAPR-----FPGLP-----CVRA-----SLNEAQE : 46
Qc-007 --MESTSSSSSVVFAVWLRNVWAVVVVCVFSILTG-C-VR--ILPTSS-----SRS-----FWWIHSTRPSV : 57
Qc-057 --MGFRRLATVALLLTLVMGIDSAPAGQTEAGRVSLREGLENTFPCNPGR-----CA-----CLPASGT : 57
Qc-059 --MEFRRLVTVTLLLTLVMSIDSVPADETETGRVSLR-E--DDRFPCNS-----NCCA-----CLPAEGS : 55
Qc-052 --MTRCLTSLLICIAMAVLATGIYYPSDETIGPCNTNDISKTETYSGRQKYVVDRTYCTCASGEVHFAADDTSSSTTVPHKIVYCITPTYS : 91

Qc-004 YQCRYGYCKNCRN-G----- : 72+
Qc-113 YQCRYGYCKNCRNGRA-CLTTST---VYQCRYGYCKNCRNG : 97+
Qc-050 GDTHFLPVEEVGWASSDCLITH----- : 68+
Qc-007 VTPRPTGSATLSTSTARCVA----- : 77+
Qc-057 TTSY--VQSTQSTADCMDNEC---VTEAEWGRRR--- : 89+
Qc-059 -STSY-QQSLDASTDDDCFDNEC---VTQSEWGRR--- : 86+
Qc-052 CTGGTSPVTDPDDEGPRRMQCTCEEYKYFVSHGWHVRCK-- : 131+

```

**B. Unassigned superfamily**

```

Qc-063 ----MRCYVLLVTLLLMWFS---GGDAG-PWHAIGHFVR---R-DCQTG-CVGCNNR---AG----CC-----CGG-----Q : 53
Qc-078 ----MKISLIICLLLIAFMNG---DGSPGN-TMYSRRS---AGVAAA-IKRFQKN-FLKKS---CN-----NCPE----- : 54
Qc-047 ----MIRMGFFLMLTVAVLLTS---LTCTDG-APAENVKLKSAR-DDCSWD-CETCDPD--VPCG---C-----CGN----- : 57
Qc-084 ----MIRMGFFLMLTVAVLLTS---LTCTDG-APAENVKLKSAR-DDCSWD-CETCDPD--VPCG---C-----CGN-----V : 58
Qc-085 ----MIRMGFFLMLTVAVLLTS---LTCTDG-APAENVKLKSAR-DDCSWD-CETCDPD--VPCG---C-----CGN----- : 54
Qc-117 ----MNFSEMFILALVSTQSMTDGFILPAENG-GRTFWRH----RPDSMDLQTRMRQT---RN---LCP-----QCPN----- : 58
Qc-122 -NWKDGRGPPSLVAPRSLPVAT---DICGGDPDEGCHEPFEF-----MGCYMN-CLMEERG--RHFG---CYER-----CKY----- : 61
Qc-051 MSNIFVLVAVFSLVVHSERAAG-FMEPMEGGLQRVMRAEMTSSLDDYPVEDVDLLQPLFRTPLKR--QWCR-----RG----- : 72
Qc-108 --MGATLVTKLLLVAVLLLGICH---EMAANTEAWDECYERMHWHMIPATWYAMYHCEILEDRKK---AALVLTLKQLG----- : 72
Qc-049 ---MMLFVLAAILTMASTTVTA---DPCSEN-GVCAWGGQEGVN-CDCPGTPCQKDDHSISAAGSILYTCKLISDFDECGDDEVAMEGDFSE : 86

Qc-063 -VCVD--NNRCE--G----GS----- : 65
Qc-078 EPCCA--GDQCMPDPGYEPFCGN----- : 75
Qc-047 VICHS--DGTCQLVSDPYNYGSY--- : 80
Qc-084 -ICHS--NGVCQLVSDPYNYGSY--- : 80
Qc-085 VICHS--SGVCQHVSDTYNYGSY--- : 78
Qc-117 -GCHV--DRTCID----- : 68
Qc-122 ---CAYANLAGCWASVG----- : 75
Qc-051 MAYNPA-LGTCTFSVAAMRGRGRKYRRH : 99
Qc-108 VIHLDHGGVVQVKLLKYLGVEN----- : 96
Qc-049 LNCRCS-SGVYEINADEVVVCG----- : 107

```

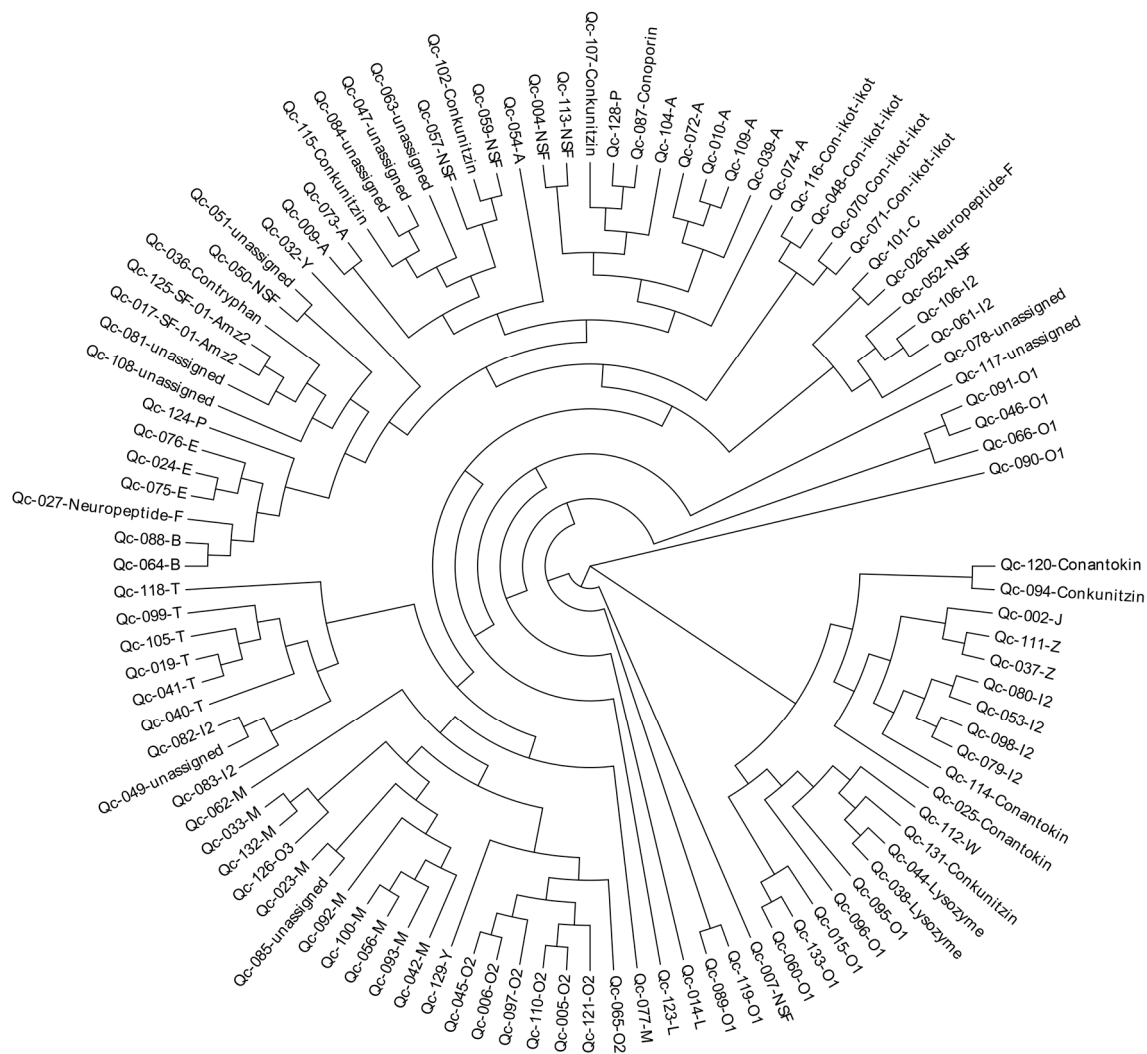
**Figure 8.** Alignment of the achieved new superfamily (NSF) and unassigned superfamily sequences from our *C. quercinus* transcriptomes. (A) The sequences of NSF-superfamily. (B) The sequences of unassigned superfamily. The signal regions are highlighted in yellow, the mature regions are underlined, and cysteine residues (in red) are marked for comparison.

## 2.6. Phylogeny of the Superfamily Signal Sequences

To investigate the relationships among these identified conotoxins from *C. quercinus*, we constructed a phylogenetic tree of all conotoxin superfamilies based on the aa sequences of the signal regions (Figure 9). This analysis showed that most of the conotoxins from *C. quercinus* transcriptomes were successfully allocated to the known superfamilies, such as B-, E, M-, O<sub>2</sub>- and T-superfamilies. For

example, the I<sub>2</sub>-superfamily conotoxins are divided into three groups, including group I (Qc-082 and Qc-083), group II (Qc-098, Qc-079, Qc-053, Qc-080), and group III (Qc-106, Qc-061).

In addition, seven NSF superfamily conotoxins were also divided into five groups, including NSF-group I (Qc-004 and Qc-113) in the A-superfamily, NSF-group II (Qc-052) in the I<sub>2</sub>-superfamily, NSF-group III (Qc-007) in the O<sub>1</sub>-superfamily, NSF-group IV (Qc-050) and NSF-group V (Qc-059 and Qc-057). However, there are still a few superfamilies that cannot be assigned to any group.



**Figure 9.** A phylogenetic tree of the signal sequences from identified conotoxins of *C. quercinus*. The phylogenetic analysis was performed using the maximum likelihood method, and the statistical supports were assessed with 1000 bootstrap pseudo-replicates.

### 3. Discussion

Next-generation sequencing technologies have recently been applied to study the venom duct transcriptomes of several *Conus* species, and 61~215 conotoxin sequences, belonging to 11~44 superfamilies have been discovered [1,2,13–25]. In previous work on *C. quercinus*, only 41 conotoxins within six superfamilies (Figure 1A) were derived from traditional approaches [26–28]. Thenext-generation transcriptome sequencing on the whole mRNA reservoir of *C. quercinus* has never been reported before. To investigate the diversity of conotoxins in a high throughput way, we employed the next-generation Illumina HiSeq2000 platform to sequence the venom duct, venom bulb and salivary gland transcriptomes of *C. quercinus*.

We obtained a total of 65, 55 and 52 putative conotoxin transcripts from the three datasets for VD, SG and VB, respectively. By comparing each two of the three transcriptomes, we determined that seven identical putative conotoxins were commonly identified in VD, SG and VB; 13 were common to both VD and VB, eight were shared by VB and SG, and five were common to both VD and SG. After removal of duplications from the three transcriptomes, we obtained a total of 133 unique putative conotoxin transcript sequences, which were classified into 34 known superfamilies with at least 1-aa difference in the mature regions between each other. In particular, conotoxin sequences of the A-, B<sub>2</sub>-, E-, I<sub>2</sub>-, -M-, O<sub>1</sub>-, O<sub>2</sub>-, O<sub>3</sub>-, T-, Y- and Z-superfamilies were found in the VD, VB, and SG, and A-, I<sub>2</sub>-, M- and O-superfamilies were the most predominant conotoxin groups in *C. quercinus*. In summary, differential transcription of conotoxins happens in different organs.

Although a total of 133 putative conotoxin sequences were identified from *C. quercinus*, only seven of which were identical to the previously discovered 41 conotoxins. This remarkable difference in the same *Conus* species may be due to differences between *Conus* individuals from different geographical populations or at different developmental stages. In fact, the previously reported 41 conotoxins were pooled from different *C. quercinus* individuals. We previously also obtained various numbers of putative conotoxin sequences from different datasets (variable in organs or sequencing methods) of *C. betulinus* [24].

Early studies of the conotoxin-producing organization demonstrate that conotoxins are synthesized in a muscular, bulbous organ located at the end of the venom duct [44,45]. With the development of molecular biology, researchers found that the epithelial cells lining the cone snails' venom duct are rich in mature mRNAs encoding conotoxins, and the venom bulb may provide the impetus to propel the venom toward the pharynx while hunting or defending [46,47]. Currently, many studies show that not all the *Conus* venom components are synthesized in the local venom duct and that other glands in the anterior gut may also make a minor contribution [17,24,48]. For example, Biggs et al. [48] discovered that PuSG1.1 and PuSG1.2 belonged to the A-superfamily with a cysteine pattern of CC-C-C in the SG of a worm-hunting species *C. pulicarius*. In addition, Lavergne et al. [17] recently reported the discovery of 3303 (99.19%) novel conotoxin sequences in the VD, as well as 22 (0.66%) in the radular sac (RS) and 5 (0.15%) in the SG in another worm-hunting species, *C. episcopatus* by transcriptome sequencing. However, similar types of transcripts were not found in the salivary glands of several fish-hunting *Conus* species [48]. For instance, the results of Dutertre et al. [32] showed that 127 conotoxin sequences were recovered from the VD, while no conotoxin sequences were identified in the SG, and only three rare conotoxin transcripts were identified in the RS in the fish-hunting species, *C. geographus* through transcriptome sequencing and proteomics. We previously demonstrated the existence of a handful of conotoxins in the venom bulb of *C. betulinus* at very low transcription levels [24]. In this study, 65, 55 and 52 putative conotoxin sequences in the VD, SG and VB, respectively, were also identified in the worm-hunting species *C. quercinus*.

It appears that most of the identified conotoxins were synthesized in the venom duct at relatively high levels, while many conotoxins were also identified in the venom bulb and salivary gland with very low transcription levels. Although we identified a number of conotoxins in the venom bulb and salivary gland, their expression levels are relatively low. It is interesting to note that such low yields are not enough for predation, hence, we propose that these conotoxins may function endogenously within the *Conus* species [48]. In addition, these conotoxins may enhance the potency of the venom by producing bioactive components [48,49]. Thus, the precise physiological functions of these salivary gland transcripts still need to be explored further. Furthermore, identification of diverse and divergent conotoxins and new cysteine frameworks XIII and VIII in *C. quercinus* will provide a potentially fertile resource for the discovery of peptides that may target different subtypes of receptors, which provides valuable leads for drug development and biomedical applications.

## 4. Materials and Methods

### 4.1. Sample Collection and RNA Extraction

Three specimens of *C. quercinus* were collected in the offshore areas of Lingshui City, Hainan province, China. Species identification was performed using mitochondrial genes as we have reported previously [50]. The venom bulb, venom duct and salivary gland were dissected immediately before RNA extraction in TRIzol Reagent (Thermo Fisher Scientific, Waltham, MA, USA). The quality of the isolated RNA was assessed by an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). Three Illumina cDNA libraries were constructed separately using the pooled RNA from the same organ of different individuals. Afterwards, the libraries were sequenced on an Illumina HiSeq4000 platform at BGI-Tech (BGI, Shenzhen, Guangdong, China) according to the manufacturer's protocols (Illumina, San Diego, CA, USA).

### 4.2. Sequence Analysis and Assembling

The paired-end short reads produced from the sequencer were firstly filtered with SOAPnuke software [51] to discard contaminants with adapters and those reads with over 10% of non-sequenced bases or more than 50% of low-quality bases (base quality score  $\leq 10$ ). The remaining clean reads were assembled into contigs using Trinity v2.5.1 [52]. These obtained contigs were finally clustered based on sequence similarity and assembled to consensus unigenes using TGICL v2.1 [53]. To calculate gene transcription values in the assembled transcriptomes, clean reads were aligned to the de novo assemblies with the Bowtie 2 read aligner [54]. Then, the obtained alignments were presented to RNA-Seq by Expectation Maximization (RSEM) v1.2.31 to estimate transcript abundance in terms of FPKM [55].

### 4.3. Functional Annotation of Transcripts

Functional annotations of the identified transcripts were conducted by searching several public databases (with the threshold of E-value  $\leq 10^{-5}$ ) including NCBI Nr and Nt, UniProtKB/Swiss-Prot [56], InterPro [57], Kyoto Encyclopedia of Genes and Genomes (KEGG) [58], Clusters of Orthologous Groups (COG) [59] and Gene Ontology (GO). Blast2GO v4.1 [60] was employed to perform GO annotation of the Nr blast results.

### 4.4. Prediction and Identification of Conotoxins

We applied homolog searches and an ab initio prediction method [61] to predict conotoxins from the three transcriptome datasets. For homologous prediction, all those previously known conotoxins were downloaded from the ConoServer database to construct a local reference dataset. We subsequently used BLASTX (with an E-value of  $1 \times 10^{-5}$ ) to run our assembled sequences against the local dataset. Those unigenes with the best hits in the BLASTX data were translated into aa sequences.

In addition, the ab initio prediction method using the HMM model [24] was adopted to discover new conotoxins. First, the three conotoxin datasets were grouped into different classes according to the superfamily and family classification in the ConoServer database. Sequences of each class were aligned with ClustalW [32], and the ambiguous results were checked by manual correction. A pHMM was built for conserved domain of each class using hmmbuild from the HMMER 3.0 package (<http://hmmmer.janelia.org>), and the hmmsearch tool was then applied to scan every assembled unigene/EST for identification of conotoxins.

### 4.5. Classification of Gene Superfamilies

The predicted conotoxin transcripts were manually inspected using the ConoPrec: precursor analysis implemented in the ConoServer [61]. Those transcripts with duplication or truncated mature region sequences were removed. Signal peptides, gene superfamilies and cysteine frameworks of these

predicted conotoxins were also checked for confirmation. Based on the 75% identity of the highly conserved signal peptide sequences [62], conotoxins could be assigned to the known superfamilies presented in the ConoServer [63]. If the conservation of a signal region was below the threshold value for any reported conotoxin superfamily, the conotoxin was regarded as a member of a new superfamily named after “NSF-Qc” plus Arabic numbers suffix. Those conotoxins without signal regions but still showing similarity either in the pro- or mature region, were considered as the Unknown group.

#### 4.6. Alignment and Phylogenetic Analysis

Concatenated amino-acid alignments of the signal domains of all superfamilies were performed using MUSCLE v.3.8.31 [64]. Prior to the phylogenetic analysis, divergent and ambiguously aligned blocks were removed using Gblocks software. To choose the best-fitted model, JModelTest and Prottest3.2 [65] were used. The phylogenetic analysis was carried out using the maximum likelihood method with the RAxML8.1, and the trees was visualized and annotated using the tree viewer of MEGA4 [66]. Statistical supports were assessed with 1000 bootstrap pseudo-replicates.

#### 4.7. Availability of Supporting Data

Datasets supporting the results of the present work are included within this article and related supplementary files. The transcriptome reads generated in this study have been deposited in China National GeneBank (CNGB) Nucleotide Sequence Archive (CNSA) with the accession number of CNP0000233 and the NCBI under PRJNA500655.

## 5. Conclusions

This study is the first to examine the diverse conotoxin transcription repertoire in different organs (including the venom duct, venom bulb and salivary gland) of the vermivorous Oak cone snail, *C. quercinus*. A total of 133 unique conotoxin sequences were identified by using a high-throughput transcriptome sequencing approach. These conotoxins could be classified into 34 known gene superfamilies, while 17 conotoxins were new or unassigned to any group. A-, O<sub>1</sub>-, M-, and I<sub>2</sub>-superfamilies were the most abundant, and the cysteine frameworks XIII and VIII were observed for the first time in the A- and I<sub>2</sub>-superfamilies, respectively. In addition, comparison of the conotoxins from the venom duct, the venom bulb and the salivary gland demonstrated that most of the identified conotoxins were synthesized in the venom duct, while a number of conotoxins were also identified in the venom bulb and salivary gland at low levels. Therefore, different organs have various conotoxins with high diversity. Interestingly, it is the first time that the existence of a number of conotoxins in the salivary gland has been revealed, although their exact functions are not clear. These novel conotoxins provide a potentially valuable resource for the development of new pharmaceuticals, and a pathway for the discovery of new conotoxins.

**Supplementary Materials:** Supplementary materials can be found at <http://www.mdpi.com/1422-0067/19/12/3901/s1>.

**Author Contributions:** B.G. and Q.S. conceived and designed the experiments; C.P., Y.Z., Y.H., Y.S. and T.Z. analyzed the data; B.G., C.P. and Y.Z. wrote the paper; Q.S. and Y.H. revised the manuscript. All authors approved the submission of the manuscript.

**Funding:** This research was funded by Hainan Provincial Keyjoint Research and Invention Program, grant number ZDYF2018138; China Scholarship Council, grant number 201708460079; Shenzhen Special Project for High-Level Talents, grant number SZYSGZZ-2018001; China 863 Project, grant number 2014AA093501 and Shenzhen Dapeng Special Project for Industrial Development, grant number PT20170302.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. 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](#)]
2. Himaya, S.W.; Jin, A.H.; Dutertre, S.; Giacomotto, J.; Mohialdeen, H.; Vetter, I.; Alewood, P.F.; Lewis, R.J. Comparative venomomics reveals the complex prey capture strategy of the piscivorous cone snail *Conus catus*. *J. Proteome Res.* **2015**, *14*, 4372–4381. [[CrossRef](#)]
3. Macrander, J.; Panda, J.; Janies, D.; Daly, M.; Reitzel, A.M. Venomix: A simple bioinformatic pipeline for identifying and characterizing toxin gene candidates from transcriptomic data. *Peer J.* **2018**, *6*, e5361. [[CrossRef](#)]
4. 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](#)]
5. 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](#)]
6. Prashanth, J.R.; Brust, A.; Jin, A.H.; Alewood, P.F.; Dutertre, S.; Lewis, R.J. Cone snail venomomics: From novel biology to novel therapeutics. *Future Med. Chem.* **2014**, *6*, 1659–1675. [[CrossRef](#)]
7. Vetter, I.; Lewis, R.J. Therapeutic potential of cone snail venom peptides (conopeptides). *Curr. Top. Med. Chem.* **2012**, *12*, 1546–1552. [[CrossRef](#)]
8. Rigo, F.K.; Dalmolin, G.D.; Trevisan, G.; Tonello, R.; Silva, M.A.; Rossato, M.F.; Klafke, J.Z.; Cordeiro Mdo, N.; Castro Junior, C.J.; Montijo, D.; et al. Effect of omega-conotoxin MVIIA and Phalpa1beta on paclitaxel-induced acute and chronic pain. *Pharmacol. Biochem. Behav.* **2013**, *114*, 16–22. [[CrossRef](#)]
9. Eisapoor, S.S.; Jamili, S.; Shahbazzadeh, D.; Ghavam Mostafavi, P.; Pooshang Bagheri, K. A New, High yield, rapid, and cost-effective protocol to deprotection of cysteine-rich conopeptide, omega-conotoxin MVIIA. *Chem. Biol. Drug Des.* **2016**, *87*, 687–693. [[CrossRef](#)]
10. Olivera, B.M.; Seger, J.; Horvath, M.P.; Fedosov, A.E. Prey-capture strategies of fish-hunting cone snails: Behavior, neurobiology and evolution. *Brain Behav. Evol.* **2015**, *86*, 58–74. [[CrossRef](#)]
11. Mir, R.; Karim, S.; Kamal, M.A.; Wilson, C.M.; Mirza, Z. Conotoxins: Structure, therapeutic potential and pharmacological applications. *Curr. Pharm. Des.* **2016**, *22*, 582–589. [[CrossRef](#)] [[PubMed](#)]
12. Ansorge, W.J. Next-generation DNA sequencing techniques. *Nat. Biotechnol.* **2009**, *25*, 195–203. [[CrossRef](#)]
13. Hu, H.; Bandyopadhyay, P.K.; Olivera, B.M.; Yandell, M. Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. *BMC Genom.* **2011**, *12*, 60. [[CrossRef](#)]
14. 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 Genom.* **2012**, *13*, 284. [[CrossRef](#)] [[PubMed](#)]
15. Terrat, Y.; Biass, D.; Dutertre, S.; Favreau, P.; Remm, M.; Stocklin, R.; Piquemal, D.; Ducancel, F. High-resolution picture of a venom gland transcriptome: Case study with the marine snail *Conus consors*. *Toxicon* **2012**, *59*, 34–46. [[CrossRef](#)] [[PubMed](#)]
16. Lavergne, V.; Dutertre, S.; Jin, A.H.; Lewis, R.J.; Taft, R.J.; Alewood, P.F. Systematic interrogation of the *Conus marmoreus* venom duct transcriptome with ConoSorter reveals 158 novel conotoxins and 13 new gene superfamilies. *BMC Genom.* **2013**, *14*, 708. [[CrossRef](#)] [[PubMed](#)]
17. Lavergne, V.; Harliwong, I.; Jones, A.; Miller, D.; Taft, R.J.; Alewood, P.F. Optimized deep-targeted proteo-transcriptomic profiling reveals unexplored *Conus* toxin diversity and novel cysteine frameworks. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E3782–E3791. [[CrossRef](#)]
18. 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](#)]
19. Lluisma, A.O.; Milash, B.A.; Moore, B.; Olivera, B.M.; Bandyopadhyay, P.K. Novel venom peptides from the cone snail *Conus pulicarius* discovered through next-generation sequencing of its venom duct transcriptome. *Mar. Genom.* **2012**, *5*, 43–51. [[CrossRef](#)]
20. Jin, A.H.; Dutertre, S.; Kaas, Q.; Lavergne, V.; Kubala, P.; Lewis, R.J.; Alewood, P.F. Transcriptomic messiness in the venom duct of *Conus miles* contributes to conotoxin diversity. *Mol. Cell Proteom.* **2013**, *12*, 3824–3833. [[CrossRef](#)]

21. Jin, A.H.; Vetter, I.; Himaya, S.W.; Alewood, P.F.; Lewis, R.J.; Dutertre, S. Transcriptome and proteome of *Conus planorbis* identify the nicotinic receptors as primary target for the defensive venom. *Proteomics* **2015**, *15*, 4030–4040. [CrossRef]
22. 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. (NY)* **2015**, *17*, 81–98. [CrossRef]
23. Zhang, H.; Fu, Y.; Wang, L.; Liang, A.; Chen, S.; Xu, A.J. Identifying novel conopeptides from the venom ducts of *Conus litteratus* through integrating transcriptomics and proteomics. *J. Proteom.* **2018**. Available online: <https://europepmc.org/abstract/med/30267875> (accessed on 26 September 2018). [CrossRef]
24. 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]
25. Huang, Y.; Peng, C.; Yi, Y.; Gao, B.; Shi, Q. A Transcriptomic survey of ion channel-based conotoxins in the Chinese tubular cone snail (*Conus betulinus*). *Mar. Drugs* **2017**, *15*, 228. [CrossRef]
26. Ye, M.; Hong, J.; Zhou, M.; Huang, L.; Shao, X.; Yang, Y.; Sigworth, F.J.; Chi, C.; Lin, D.; Wang, C. A novel conotoxin, qc16a, with a unique cysteine framework and folding. *Peptides* **2011**, *32*, 1159–1165. [CrossRef]
27. Peng, C.; Chen, W.; Han, Y.; Sanders, T.; Chew, G.; Liu, J.; Hawrot, E.; Chi, C.; Wang, C. Characterization of a novel  $\alpha 4/4$ -conotoxin, Qc1.2, from vermivorous *Conus quercinus*. *Acta Biochim. Biophys. Sin. (Shanghai)* **2009**, *41*, 858–864. [CrossRef]
28. Peng, C.; Tang, S.; Pi, C.; Liu, J.; Wang, F.; Wang, L.; Zhou, W.; Xu, A. Discovery of a novel class of conotoxin from *Conus litteratus*, It14a, with a unique cysteine pattern. *Peptides* **2006**, *27*, 2174–2181. [CrossRef]
29. Kaas, Q.; Westermann, J.C.; Halai, R.; Wang, C.K.; Craik, D.J. ConoServer, a database for conopeptide sequences and structures. *Bioinformatics* **2008**, *24*, 445–446. [CrossRef]
30. 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]
31. Zhou, M.; Wang, L.; Wu, Y.; Zhu, X.; Feng, Y.; Chen, Z.; Li, Y.; Sun, D.; Ren, Z.; Xu, A. Characterizing the evolution and functions of the M-superfamily conotoxins. *Toxicon* **2013**, *76*, 150–159. [CrossRef]
32. 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]
33. Puillandre, N.; Watkins, M.; Olivera, B.M. Evolution of *Conus* peptide genes: Duplication and positive selection in the A-superfamily. *J. Mol. Evol.* **2010**, *70*, 190–202. [CrossRef] [PubMed]
34. Aguilar, M.B.; Ortiz, E.; Kaas, Q.; López-Vera, E.; Becerril, B.; Possani, L.D.; de la Cotera, E.P. Precursor De13.1 from *Conus delessertii* defines the novel G gene superfamily. *Peptides* **2013**, *41*, 17–20. [CrossRef] [PubMed]
35. Rajesh, R.P. Novel M-Superfamily and T-Superfamily conotoxins and contryphans from the vermivorous snail *Conus figulinus*. *J. Pept. Sci.* **2015**, *21*, 29–39. [CrossRef] [PubMed]
36. 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  $\alpha 9\alpha 10$  Nicotinic Acetylcholine Receptor. *Mar. Drugs* **2017**, *15*, 170. [CrossRef] [PubMed]
37. Jimenez, E.C.; Craig, A.G.; Watkins, M.; Hillyard, D.R.; Gray, W.R.; Gulyas, J.; Rivier, J.E.; Cruz, L.J.; Olivera, B.M. Bromocontryphan: Post-translational bromination of tryptophan. *Biochemistry* **1997**, *36*, 989–994. [CrossRef] [PubMed]
38. Vijayarathy, M.; Balaram, P. Mass spectrometric identification of bromotryptophan containing conotoxin sequences from the venom of *C. amadis*. *Toxicon* **2018**, *144*, 68–74. [CrossRef]
39. 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]
40. Zamora-Bustillos, R.; Aguilar, M.B.; Falcón, A. Identification, by molecular cloning, of a novel type of I<sub>2</sub>-superfamily conotoxin precursor and two novel I<sub>2</sub>-conotoxins from the worm-hunter snail *Conus spurius* from the Gulf of México. *Peptides* **2010**, *31*, 384–393. [CrossRef]
41. Phuong, M.A.; Mahardika, G.N. Targeted sequencing of venom genes from cone snail genomes improves understanding of conotoxin molecular evolution. *Mol. Biol. Evol.* **2018**, *35*, 1210–1224. [CrossRef]

42. Robinson, S.D.; Norton, R.S. Conotoxin gene superfamilies. *Mar. Drugs* **2014**, *12*, 6058–6101. [[CrossRef](#)] [[PubMed](#)]
43. Wang, L.; Liu, J.; Ren, Z.; Chen, Y.; Xu, A. Discovery of two P-superfamily conotoxins, It9a and It9b, with different modifications on voltage-sensitive sodium channels. *Toxicon* **2017**, *134*, 6–13. [[CrossRef](#)] [[PubMed](#)]
44. Himaya, S.W.A.; Lewis, R.J. Venomics-Accelerated Cone Snail Venom Peptide Discovery. *Int. J. Mol. Sci.* **2018**, *19*, 788. [[CrossRef](#)] [[PubMed](#)]
45. Marshall, J.; Kelley, W.P.; Rubakhin, S.S.; Bingham, J.P.; Sweedler, J.V.; Gilly, W.F. Anatomical correlates of venom production in *Conus californicus*. *Biol. Bull.* **2002**, *203*, 27–41. [[CrossRef](#)]
46. Safavi-Hemami, H.; Young, N.D.; Williamson, N.A.; Purcell, A.W. Proteomic interrogation of venom delivery in marine cone snails: Novel insights into the role of the venom bulb. *J. Proteome Res.* **2010**, *9*, 5610–5619. [[CrossRef](#)]
47. Prashanth, J.R.; Dutertre, S.; Lewis, R.J. Pharmacology of predatory and defensive venom peptides in cone snails. *Mol. Biosyst.* **2017**, *13*, 2453–2465. [[CrossRef](#)]
48. Biggs, J.S.; Olivera, B.M.; Kantor, Y.I.  $\alpha$ -Conopeptides specifically expressed in the salivary gland of *Conus pulicarius*. *Toxicon* **2008**, *52*, 101–105. [[CrossRef](#)]
49. Shimek, R.L. The morphology of the buccal apparatus of *Oenopota levidensis* (Gastropoda, Turridae). *Zoomorphology* **1975**, *80*, 59–96. [[CrossRef](#)]
50. 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](#)]
51. Li, R.; Li, Y.; Kristiansen, K.; Wang, J. SOAP: Short oligonucleotide alignment program. *Bioinformatics* **2008**, *24*, 713–714. [[CrossRef](#)]
52. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494. [[CrossRef](#)] [[PubMed](#)]
53. Pertea, G.; Huang, X.; Liang, F.; Antonescu, V.; Sultana, R.; Karamycheva, S.; Lee, Y.; White, J.; Cheung, F.; Parvizi, B.; et al. TIGR Gene Indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. *Bioinformatics* **2003**, *19*, 651–652. [[CrossRef](#)] [[PubMed](#)]
54. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357. [[CrossRef](#)] [[PubMed](#)]
55. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **2011**, *12*, 323. [[CrossRef](#)] [[PubMed](#)]
56. 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](#)]
57. Finn, R.D.; Attwood, T.K.; Babbitt, P.C.; Bateman, A.; Bork, P.; Bridge, A.J.; Chang, H.Y.; Dosztányi, Z.; El-Gebali, S.; Fraser, M.; et al. InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res.* **2016**, *45*, D190–D199. [[CrossRef](#)]
58. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)]
59. 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](#)]
60. 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](#)]
61. Kaas, Q.; Westermann, J.C.; Craik, D.J. Conopeptide characterization and classifications: An analysis using ConoServer. *Toxicon* **2010**, *55*, 1491–1509. [[CrossRef](#)]
62. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [[CrossRef](#)] [[PubMed](#)]
63. Laht, S.; Koua, D.; Kaplinski, L.; Lisacek, F.; Stöcklin, R.; Remm, M. Identification and classification of conopeptides using profile Hidden Markov Models. *Biochim. Biophys. Acta* **2012**, *1824*, 488–492. [[CrossRef](#)] [[PubMed](#)]



64. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)] [[PubMed](#)]
65. Diego, D.; Guillermo, L.; Taboada, R.D.; David, P. ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **2011**, *27*, 1164–1165.
66. Alexandros, S. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**, *22*, 2688–2690.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).