

RESEARCH

Open Access



High-throughput identification of novel conotoxins from the Chinese tubular cone snail (*Conus betulinus*) by multi-transcriptome sequencing

Chao Peng^{1†}, Ge Yao^{2†}, Bing-Miao Gao^{3†}, Chong-Xu Fan², Chao Bian^{1,4}, Jintu Wang¹, Ying Cao², Bo Wen¹, Yabing Zhu¹, Zhiqiang Ruan^{1,4}, Xiaofei Zhao⁵, Xinxin You^{1,4}, Jie Bai^{1,4}, Jia Li^{1,4}, Zhilong Lin¹, Shijie Zou¹, Xinhui Zhang^{1,4}, Ying Qiu^{1,4}, Jieming Chen^{1,4}, Steven L. Coon⁶, Jiaan Yang⁵, Ji-Sheng Chen^{2*} and Qiong Shi^{1,4,7*}

Abstract

Background: The venom of predatory marine cone snails mainly contains a diverse array of unique bioactive peptides commonly referred to as conopeptides or conotoxins. These peptides have proven to be valuable pharmacological probes and potential drugs because of their high specificity and affinity to important ion channels, receptors and transporters of the nervous system. Most previous studies have focused specifically on the conopeptides from piscivorous and molluscivorous cone snails, but little attention has been devoted to the dominant vermivorous species.

Results: The vermivorous Chinese tubular cone snail, *Conus betulinus*, is the dominant *Conus* species inhabiting the South China Sea. The transcriptomes of venom ducts and venom bulbs from a variety of specimens of this species were sequenced using both next-generation sequencing and traditional Sanger sequencing technologies, resulting in the identification of a total of 215 distinct conopeptides. Among these, 183 were novel conopeptides, including nine new superfamilies. It appeared that most of the identified conopeptides were synthesized in the venom duct, while a handful of conopeptides were identified only in the venom bulb and at very low levels.

Conclusions: We identified 215 unique putative conopeptide transcripts from the combination of five transcriptomes and one EST sequencing dataset. Variation in conopeptides from different specimens of *C. betulinus* was observed, which suggested the presence of intraspecific variability in toxin production at the genetic level. These novel conopeptides provide a potentially fertile resource for the development of new pharmaceuticals, and a pathway for the discovery of new conotoxins.

Keywords: Cone snail, Conopeptide, Conotoxin, *Conus betulinus*, Transcriptome, Venom duct, Venom bulb

* Correspondence: chenjsh@cae.cn; shiqiong@genomics.cn

† Equal contributors

²Research Institute of Pharmaceutical Chemistry, Beijing 102205, China

¹BGI-Shenzhen, Shenzhen 518083, China

Full list of author information is available at the end of the article



Background

The genus *Conus* is classified in the Coninae subfamily within the Conidae family that belongs to the Conoidea superfamily (a branch of the Neogastropoda clade) [1, 2]. With an estimation of 700 species [3, 4], all cone snails are classified in *Conus*, which is the largest genus among marine invertebrates. Venomous cone snails are carnivorous and predatory marine gastropod mollusks that use a complex cocktail of venom components for many reasons, including capture and digestion of prey, defense against foes [5], avoidance of competitors and other biological purposes [6]. According to variation in diet, cone snails are divided into three groups: piscivorous species that hunt small fish, molluscivorous species that feed on other marine snails including other cone snails, and vermivorous cone snails that prey on polychaetes and hemichordates [7].

As slow-moving predatory marine gastropods, cone snails have developed successful strategies to subdue quicker or stronger prey during more than 55 million years of evolution [8], including a hollow, harpoon-like radular tooth and potent toxins targeted to the nervous system and musculature of the prey [9]. The modified radular tooth (along with a venom gland) can be launched out from the snail's mouth deep into the prey's flesh in a harpoon-like action. The injected venom rapidly enters the victim's circulatory system, interacts with a range of molecular targets in the nervous system, and causes paralysis in a short time (sometimes within a few seconds) [10, 11].

In contrast to the well-known large protein toxins in snake venom, *Conus* venom mainly contains a diverse array of unique bioactive peptides commonly referred to as conopeptides or conotoxins. These small polypeptide conotoxins typically range from seven to 46 amino acids in length, with many of them consisting of 12–30 amino acids [12]. They have high specificity and affinity to voltage-gated ion channels, ligand-gated ion channels, G-protein-coupled receptors and neurotransmitter transporters in the central and peripheral nervous systems [3, 6, 12–18]. Because of their bioactive specificity, *Conus* venoms have become a potent resource for pharmacological neuroscience research [19–22] and a promising source for the discovery of new drugs to treat a wide variety of human neurological diseases [6, 23–30]. To date, several conotoxins have already demonstrated potential therapeutic effects in preclinical or clinical trials. The most well-known is ω -MVIIA (commercially known as ziconotide), derived from the venom of *C. magus*, which has been approved by the US Food and Drug Administration (FDA) to treat previously unmanageable chronic pain in cancer and AIDS patients [30–32]. Another conotoxin, conantokin G, a specific antagonist against the NR2B subunit of the NMDA receptor, is in

human clinical trials for intractable epilepsy [33]. In addition, more and more conopeptides are undergoing development for the treatment of pathologies including pain, Parkinson's disease, cardiac infarction, hypertension and various neurological diseases [12, 28, 34–37].

With a few exceptions, each conopeptide precursor generally consists of three distinct regions: a highly conserved N-terminal signal peptide region, a less conserved intervening propeptide region, and a hypervariable C-terminal mature toxin region [38, 39]. Based on the sequence similarities of signal peptides in the precursors [40], conopeptides are currently classified into 26 gene superfamilies (A, B1, B2, B3, C, D, E, F, G, H, I1, I2, I3, J, K, L, M, N, O1, O2, O3, P, S, T, V and Y) [41–47] and 13 temporary gene superfamilies for those identified in the early divergent clade species [40, 42, 48, 49]. Although amino acid conservation in the mature peptide sequences of conopeptides within a same gene superfamily is much lower than in the signal and propeptide regions, certain characteristic cysteine frameworks within the mature conotoxins are often (but not always) specific to a conotoxin superfamily. So far, 26 distinct cysteine frameworks have been described, and they may be associated with particular pharmacological families [39, 40, 50].

Early views of the conotoxin-producing structures concluded that the muscular bulbous organ (the venom bulb, located at the end of the venom duct) was likely to participate in venom biosynthesis [51]. With the development of molecular biology, researchers found that the epithelial cells lining the cone snail's venom duct were rich in mature mRNAs encoding precursor conopeptides, and the venom bulb may function to propel the venom toward the pharynx while preying or defending [51–53].

With early estimates of an average of 100 conotoxins per species, recent reports proved the presence of 1000 to 2000 different conopeptides in a single sample of venom using high-sensitivity mass spectrometry [44, 54, 55]. A single mutation in a mature sequence can add one more conopeptide at the DNA level and, subsequently, owing to the 14 different post-translational modifications known in addition to other undefined alterations, an average of 20 different toxin variants for each conopeptide precursor can be characterized at the protein level [44]. This raises the possibility that the total of 500 to 700 species of cone snails, providing upwards of 50,000 conopeptide genes and 1,000,000 mature conotoxins as potential pharmacological targets, constitutes the largest single library of natural drug candidates.

Despite the large number of potential conopeptide genes and mature conotoxins, only approximately 1400 nucleotide sequences of conotoxin genes have been reported from 100 *Conus* species by traditional approaches

over the past decades, with as few as 210 peptides being validated at the protein level [40, 41]. Traditional methods, which may isolate and sequence these potential bioactives, are generally time-consuming, of low sensitivity, and often limited by sample availability. In contrast, high-throughput sequencing can achieve greater sequencing depth and larger coverage of the transcriptome so that even rare transcripts with low expression levels can be identified [56]. Recent studies on the venom duct transcriptome of several *Conus* species, using next-generation sequencing technologies, have uncovered about 100 conopeptide genes per *Conus* species [5, 44, 53, 57–62].

Data description

To date, most studies have specifically focused on the piscivorous and molluscivorous cone snails, whereas there is still relatively little research on the abundant vermivorous species (which account for about 75 % of all cone snails) [40, 58, 63, 64]. As a worm-hunting species, the Chinese tubular cone snail (*C. betulinus* (Linnaeus)) is a dominant *Conus* species inhabiting the South China Sea. In previous works on this

species [42], only 53 mature conotoxins from nine gene superfamilies (Fig. 1b) were derived from precursors or via traditional approaches (see Additional file 1). Next-generation whole-transcriptome sequencing of the *C. betulinus* venom duct has never been attempted. Our current study therefore surveyed conotoxin cDNA precursors using a variety of strategies, including sampling individuals with different body sizes, sampling different tissues, preparing samples with different normalization strategies, and employing different sequencing methodologies. This resulted in six datasets (see Methods for details). Body sizes were categorized as Big, Middle and Small: the Big specimen was 10 cm in body length, the Middle one was 8.7 cm and the Small specimen was 6 cm.

A summary of five transcriptome assemblies (excluding the expressed sequence tag (EST) dataset) is presented in Table 1. The Illumina sequencing generated 5.46, 4.58, 8.39, 4.67 and 9.77 Gb of raw sequences in the datasets Normalized, Small, Middle, Big and Bulb, respectively (see more details in Methods). After trimming low-quality reads, 5.23, 4.37, 7.95, 4.42 and 9.16 Gb of corresponding clean reads were obtained and used

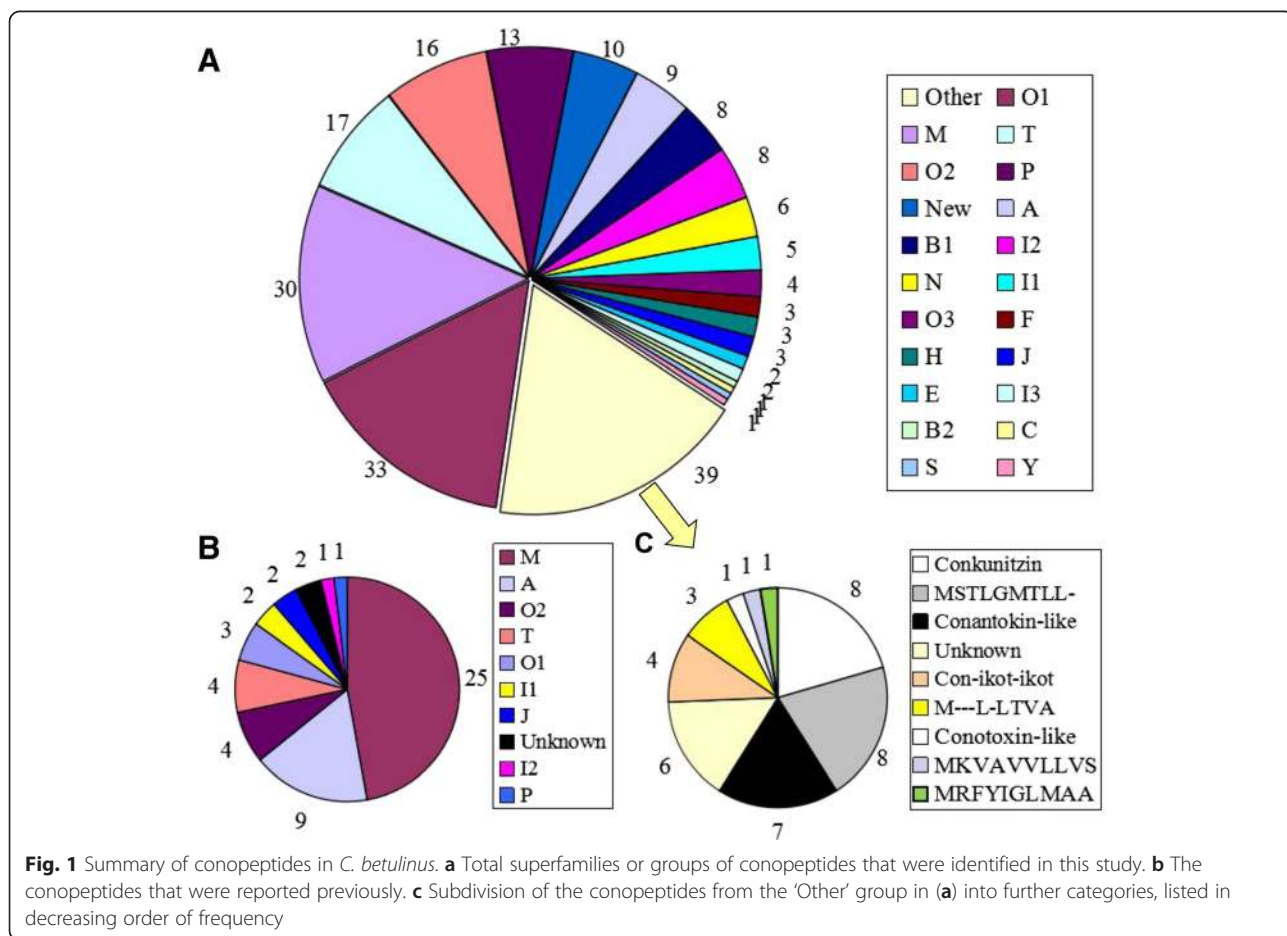


Table 1 Summary of the transcriptome assemblies

		Venom Duct				Venom Bulb
		Normalized	Small	Middle	Big	
Total raw reads		60,707,418	50,926,032	93,247,084	51,896,884	108,589,106
Total raw nucleotides (nt)		5,463,667,620	4,583,342,880	8,392,237,560	4,670,719,560	9,773,019,540
Total clean reads		58,091,534	48,557,542	88,318,918	49,161,726	101,724,490
Total clean nucleotides (nt)		5,228,238,060	4,370,178,780	7,948,702,620	4,424,555,340	9,155,204,100
Unique gene	Number	87,714	114,057	52,387	94,026	124,004
	Total length	45,438,256	44,918,779	23,128,493	37,880,261	67,451,577
	Max length	16,974	14,747	12,264	9,564	18,070
	Mean length	518	394	441	403	544
	N50	612	398	464	413	681

*The Normalized transcriptome was assembled by SOAPdenovo-Trans 1.02 and the transcriptomes of Big, Middle, Small and Bulb were assembled using Trinity software. Evaluation of two assemblers for the Normalized sample showed essentially equal performance.

for subsequent assembly. Using the *de novo* transcriptome assemblers Trinity and SOAPdenovo-Trans [65, 66], the clean reads from the five datasets were separately assembled into contigs. For improved assembly quality, a clustering step was performed by eliminating redundant contigs [67]. Contigs were then further assembled into between 136,569 and 180,492 unique transcripts with a mean length of 394 to 544 bp and an N50 length of 398 to 681 bp for the five transcriptomes.

In parallel, a cDNA library, generated from a pool of total RNA from the venom ducts of six specimens, was sequenced by using an ABI 3730 (Sanger-type). After removing vector sequences, primer sequences and poly(A) tails, 11,026 clean ESTs were obtained, with an average length of 663 bp. Redundancy among these ESTs was further eliminated and a total of 5798 unique transcripts, averaging 692 bp in length, were finally acquired.

Results

Total conopeptides identified in the current study

The putative conopeptide sequences were predicted by BLASTX search and HMMER analysis [68] (Additional files 2 and 3) against a local reference database of known conopeptides from the ConoServer databases [41], and then examined manually using the ConoPrec tool [42]. After removal of the transcripts with duplication or truncated mature region sequences, we obtained totals of 46, 123, 98, 94, 95 and 39 putative conopeptide transcripts for the six datasets of EST, Normalized, Small, Middle, Big (non-normalized venom ducts) and (venom) Bulb, respectively. The majority of these identified conopeptides were full-length or nearly full-length, although a few conopeptides contained mature region sequences only. For these partial sequences, we tried to derive the missing regions with RT-PCR and Sanger sequencing.

We combined the six conopeptide datasets into a 'Total conopeptide dataset' and named the 215 putative conopeptides identified as Bt001 to Bt215 (Additional

file 4). They each have at least one amino acid (aa) difference from one another in the mature regions. Among these 215 conopeptides, 178 were classified into 20 previously reported superfamilies (A, B1, B2, C, E, F, H, I1, I2, I3, J, M, N, O1, O2, O3, P, S, T and Y; Fig. 1a) and two cysteine-rich families (like Conkunitzin and Con-ikot-ikot; Fig. 1c). Seven of the conopeptides were highly similar to the 'conantokin-like' group (Fig. 1c), which belongs to the peculiar cysteine-poor Conantokin family but is not classified with the B1 superfamily in the ConoServer database because of the obvious difference in the signal region sequences.

In addition, the sequence identities in the signal regions of ten putative conopeptides (Fig. 2) in the 'Total conopeptide dataset' were below the threshold values for any empirical superfamilies (see more details in Methods). Therefore, the ConoPrec tool [42] was used to analyze these ten conopeptide precursors and identify their signal sequences. It was confirmed that the majority of these conopeptides (Bt101, Bt103, Bt110 and Bt113) contain three common regions, i.e. signal region, pro-region and mature region. However, several conopeptides, including Bt104, Bt106, Bt112, Bt116 and Bt119, have a short peptide sequence after the mature region, and Bt102 contains only partial pro- and mature regions. Finally, all these novel putative conopeptides were classified into nine new conopeptide superfamilies (Fig. 2), designated as NSF-bt01 to NSF-bt09.

Among the 53 previously published conopeptide sequences of *C. betulinus* (Additional file 1), 26 were recovered in our 'Total conopeptide dataset' (Fig. 1b and Additional file 1). Half of these 26 conopeptides belong to the M-superfamily, which accounts for 52 % of the M-conotoxins published previously for this species. Four conotoxins of the T-superfamily have been reported in *C. betulinus* before, and all of these were re-identified in

NSF-bt01	Bt101	MEFRRLLVTVALLLTLVMSIDSAPADQETGRVSLREDDRFPPNSDQACALPREGSSTSYQCQSTASASTADVNNNGITISEW
	Bt102DQETETGRVSLREDNRFPPNSGQACALPKAGSTLSYQCQSTANASRCVNDNCITTEEN
NSF-bt02	Bt103	MEKLMTLMVVAALVLTQVQVQDQNSPKDVRKRLQGSVFKRTEVSLPTEDDSELCAPYLCKCTMGSTDSRST
NSF-bt03	Bt104	MGKLTQVLLFAAVLMLTLDQGGEGDTSRNRKSRVTHRALRAQVKEGLMKNEAECCWPNTCRATIGETPPGRPPPAHKRAAWL
NSF-bt04	Bt106	MKLSVMVIVLVLATAFTPLGLLHDDNEGKALRRHVRHVRDTLADERAKRSSNPPPIPKICRLDVEDVNSDMGR
NSF-bt05	Bt110	MKYPTFVMVMAAVLLTGILETEARASLRARLAKKTYVPEQALEEFPPEVEACEGTSWCVGKPDGTACCRSGHFNAAICYI
NSF-bt06	Bt112	MMWKAGAVLFALLLVFSLAAVQPDLDERTFGRMKRDMFVRLKSKRDGQSTTYCPGPGQMTCCGHLNEINSAFCLFLRHR
NSF-bt07	Bt113	MMWKMVRMLFVVLLFLPLASLQEGDITQAKKDLKSNFYGALKRPARDRGTGHTDQGTGTECEGHDSCEGREHYPGTISE
NSF-bt08	Bt116	MNLVSMSTILLLLAQQLFAATSIAERKTAQDNTDLNLIIDLNAREDKPKKGGEGGFHTAECEDTYVCRQSFKSVCEIG
NSF-bt09	Bt119	MRLPTMHSVILMLLMLFAFDNVDGQGTARDVDNGKFMSSLSQSEGMPARFFMLGEKRLCESRSPNDGPDRCDDANNHEDYWK

Fig. 2 New superfamilies of conopeptides identified in *C. betulinus*. The ten conopeptides in the 'New' group of Fig. 1a have been clustered into nine new superfamilies (designated NSF-btXX), according to their signal peptide sequences. The signal regions predicted by the ConoPrec tool are underlined, and the mature regions (shown in red) and cysteine residues (highlighted yellow) are marked for comparison

our current study. Some published conopeptides from this species, belonging to the A, I1, I2, O1 and O2 superfamilies, were also confirmed, whereas none belonged to the J and P superfamilies (Additional file 1).

Interestingly, only seven of the 215 total conopeptides have the same mature region sequences as previously reported in other cone snail species (Fig. 3 and Table 2). Among these, three conopeptides (Bt072, Bt079 and Bt091) belong to the M-superfamily, and the remaining four (Bt006, Bt148, Bt177 and Bt185) are classified into the A, O1, O2 and O3 superfamilies, respectively. There is no difference in their precursor sequences between Bt072 and the reported conotoxin S3-E02 of piscivorous *C. striatus*. The same situation also occurs between Bt185 and S6.18 (*C. striatus*), and between Bt177 and Vr15b (*C. varius*, vermivorous). Two precursor sequences of Ts3.2 from the vermivorous *C. tessulatus* have only one aa difference in the signal region; our

newly identified Bt079 has exactly the same mature region, but has four and three aa differences in the signal sequences and pro-regions, respectively. Bt091 has apparently lost part of its signal region, but its remaining sequences are consistent with Vx3-F01 from the vermivorous *C. vexillum*. Despite missing its signal sequence and partial pro-region, Bt148 showed an identical mature region and almost the same pro-region (only one aa mutation) as MaIr94 of the molluscivorous *C. marmoreus*. The mature peptide sequence of Bt006 had been reported in both *C. betulinus* (named as Bt1.4) and *C. pergrandis* (PeIA) [69, 70]; however, there is a one-aa difference in the pro-region between Bt006 and Bt1.4, and PeIA has more aa mutations (nine) in the pro-region when compared with Bt006 and Bt1.4.

These seven conotoxins demonstrated significant differential expression during the growth of *C. betulinus* (Table 2). For example, Bt079 and Bt148 were only

M superfamily	Conus Species	Food Habits
Bt072 <u>MLKMGVVFTEFLVFLPLATLQLDADQPVVERYVENKQDLNPDERSNFRLLPVRRCCSVSICQPPVCECCA</u>	<i>C. betulinus</i>	V
S3-E02 <u>MLKMGVVFTEFLVFLPLATLQLDADQPVVERYVENKQDLNPDERSNFRLLPVRRCCSVSICQPPVCECCA</u>	<i>C. striatus</i>	P
Bt079 <u>MMSKLGVLITICLLFPLAVSLDGGQPADLPELRQDFAPERSPWFDPVRRCCSQDCRVCIPCCPY</u>	<i>C. betulinus</i>	V
Ts3.2 <u>MMSKLGVLITICLLFPLAVSLDGGQPADLPELRQDFAPERSPWFDPVRRCCSQDCRVCIPCCPY</u>	<i>C. tessulatus</i>	V
Ts3.2 <u>MMSKLGVLITICLLFPLAVSLDGGQPADLPELRQDFAPERSPWFDPVRRCCSQDCRVCIPCCPY</u>	<i>C. tessulatus</i>	V
Bt091 <u>VLFPPLATLQLDADQHVVERNAENKQDLIPDKRRVILLRALRQRCCPMEWCDGDCFCVY</u>	<i>C. betulinus</i>	V
Vx3-F01 <u>MLKMGVVFTEFLVFLPLATLQLDADQHVVERNAENKQDLIPDKRRVILLRALRQRCCPMEWCDGDCFCVY</u>	<i>C. vexillum</i>	V
A superfamily		
Bt006 <u>MMFTVFLVLLVATTVVVSDRAFGRNPAANDKRSDLAALSVRRCGCCSHIPACSVNHPELGGRRR</u>	<i>C. betulinus</i>	V
Bt1.4 <u>MMFTVFLVLLVATTVVVSDRAFGRNPAANDKRSDLAALSVRRCGCCSHIPACSVNHPELGGRRR</u>	<i>C. betulinus</i>	V
PeIA <u>FDGRNPAANDKRSDLAALTVRGCSSHIPACSVNHPELGG</u>	<i>C. pergrandis</i>	Un
O1 superfamily		
Bt148 <u>EVQRSEDPKLAKRRCLESGSLCFAGYGHSSCCSGACLDYGGLVGACR</u>	<i>C. betulinus</i>	V
MaIr94 <u>MKLTQVLLITVFLPLATLQLDADQVVERYVENKQDLNPDERSNFRLLPVRRCCSVSICQPPVCECCA</u>	<i>C. marmoreus</i>	M
O2 superfamily		
Bt177 <u>MEKLTTLVAVTLLAIAQVLVQSDREKPLKGRVKQYAAKRLSALMRGPRQCTPKNQICEEDGECPLNCKCFTRPDCQSGYKCRP</u>	<i>C. betulinus</i>	V
Vr15b <u>MEKLTTLVAVTLLAIAQVLVQSDREKPLKGRVKQYAAKRLSALMRGPRQCTPKNQICEEDGECPLNCKCFTRPDCQSGYKCRP</u>	<i>C. varius</i>	V
O3 superfamily		
Bt185 <u>MSGLGIMVLTLLLLVIMATSHQDAGEKQMTQRDAINVRRRRSTITGRVVTAEACEEAEQEKNCCTGTTNGQPCNFPCIG</u>	<i>C. betulinus</i>	V
S6.18 <u>MSGLGIMVLTLLLLVIMATSHQDAGEKQMTQRDAINVRRRRSTITGRVVTAEACEEAEQEKNCCTGTTNGQPCNFPCIG</u>	<i>C. striatus</i>	P

Fig. 3 Comparison of seven *C. betulinus* conopeptides with their homologous sequences reported from other *Conus* species. The conopeptides identified in our current study are shown in black, and the reference sequences are marked in blue. Note that the mature regions (shown in red) are highly conserved. The names of the reference sequences are derived from the ConoServer database. The signal regions are underlined, and the dissimilar residues are highlighted in yellow. (Food habits: M, molluscivorous; P, piscivorous; Un, undescribed; V, vermivorous)

Table 2 Expression levels of the seven conotoxins covered in Fig. 3

Conopeptide	Superfamily	Small		Middle		Big		Reference conotoxin ^a
		RPKM	Ranking	RPKM	Ranking	RPKM	Ranking	
Bt072	M	3,110	19/98	25,964	8/94	—	—	S3-E02
Bt079		—	—	—	—	1331	38/96	Ts3.2
Bt091		19	85/98	—	—	—	—	Vx3-F01
Bt006	A	1,380	33/98	—	—	—	—	PelA/Bt1.4
Bt148	O1	—	—	—	—	3034	24/96	Malr94
Bt177	O2	437	51/98	649	55/94	628	52/96	Vr15b
Bt185	O3	—	—	17,584	13/94	—	—	S6.18

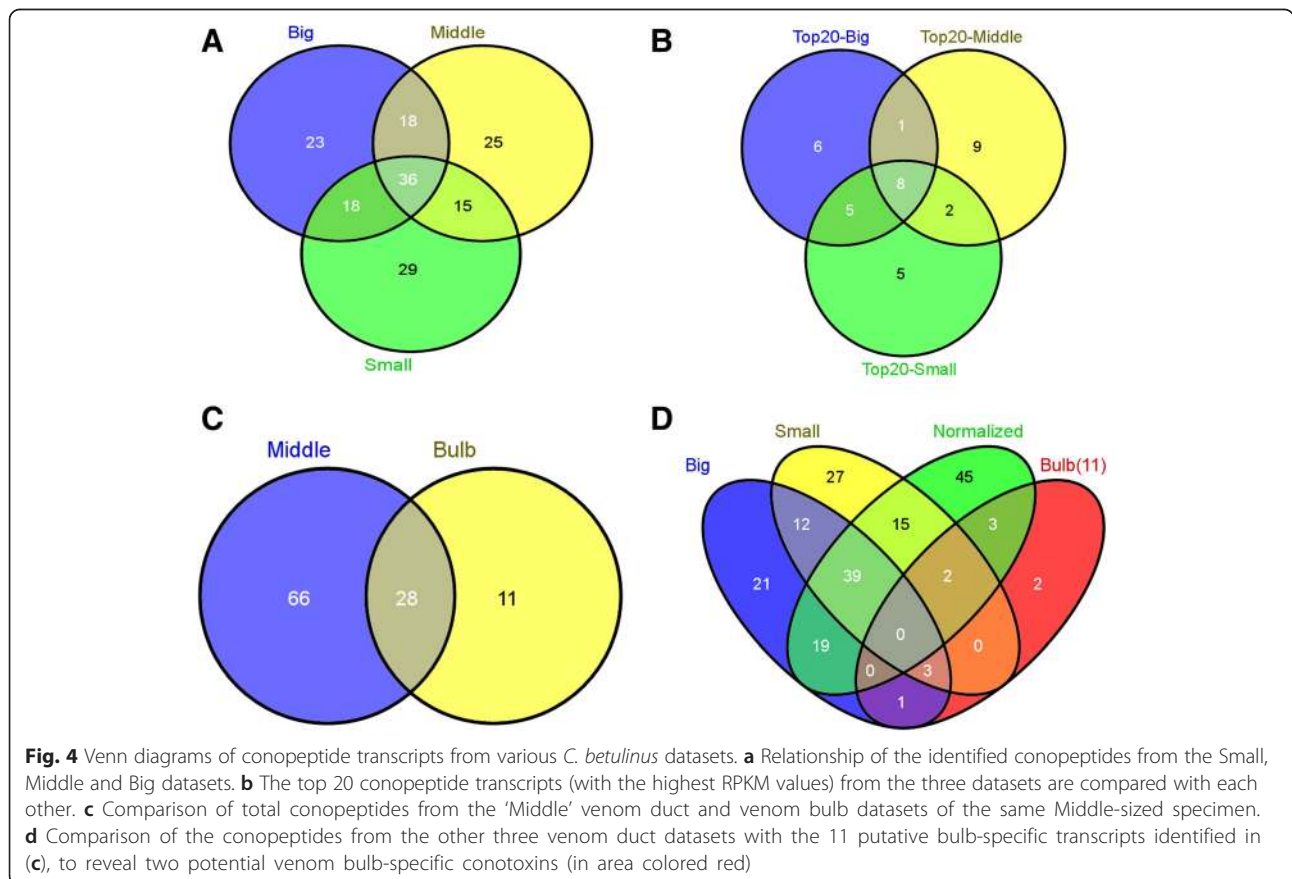
^aNames of the reference conotoxins are derived from the ConoServer database. — indicates undetectable

present in the Big dataset, Bt006 and Bt091 were only expressed in the Small dataset, and Bt185 was expressed only in the Middle dataset; Bt072 was expressed in both the Small and Middle datasets, but was absent from the Big dataset; Bt177 was the only conotoxin identified in all three of the datasets from differently sized snails.

Comparison of conopeptides in the three venom duct transcriptomes

A total of 98, 94 and 95 putative conopeptide precursors, respectively, were identified from the Small, Middle and Big datasets from the venom ducts of three body-

sized *C. betulinus*. The comparative distribution of conopeptides is summarized in the Venn diagrams of Fig. 4a,b. The 36 common conopeptides were classified into 21 superfamilies, of which 17 were described in the ConoServer database and three were new superfamilies. Most of these identified peptides belong to the O1 and M superfamilies and the Conkunitzin group. Around the same number of unique conopeptides were present together within each pairing of the three transcriptomes (51, 54 and 54 common precursors in the Small & Middle, Middle & Big, and Big & Small, respectively). In the same way, the number of conopeptides identified as



specific to any one of the three body-sized specimens was similar (29, 25 and 23 from the Small, Middle and Big datasets, respectively).

RPKM (reads per kilobase of transcript per million mapped reads) values were calculated to represent the expression levels of each conopeptide. The top 20 conopeptides (with the highest RPKM values) were selected from each of the datasets, and it was found that expression levels in the Middle snail were generally higher than those in the Small or Big specimens. The top 18 conopeptides in the Middle specimen, as well as the top eight conopeptides in both the Small and Big specimens, had RPKM values above 10,000 (Fig. 5). Eight transcripts were common to the top 20 conopeptides of the three datasets (Fig. 4b), but their RPKM ranking was variable (Table 3). For example, Bt035, the highest-ranked in both the Small and the Big specimens, ranked only fourth in the Middle snails. Using the National Center for Biotechnology Information’s (NCBI) BLASTP tool, we also found that Bt035 is similar to the Eb-conantokin-like protein from *C. eburneus* and the conantokin-F peptide from *C. flavidus* in its precursor sequences (73 and 65 % identity, respectively), suggesting Bt035’s conantokin-like status and potential neuronal

NMDAR-inhibiting activity. Interestingly, Bt018, with high expression levels in all three specimens (Table 3), may be the first B2-superfamily conopeptide identified in *C. betulinus*, because its signal sequence possesses high similarity (only one aa substitution) to this superfamily.

Bt075 and Bt082 (a conomorphin) belong to the M-superfamily. Bt075 is identical to the Bt3-3-VP02 conotoxin previously reported in *C. betulinus* [71]. However, Bt082 contains the cleavage site KR, producing a mature 17-aa linear peptide with no cysteine framework, which is different from the more common 15-aa mature peptide in previously examined conomorphins (Conomorphin-Bt1, 2 and 3).

Bt055, a representative of the I2-superfamily, was the highest-ranked conopeptide in the Middle specimen; its mature region sequence matches the cysteine framework of the previously identified kappa-Btx (Additional file 1). The A-superfamily Bt005 is highly similar to both the α -conotoxin-like Lp1.7 from *C. leopardus* [72] and Lt1c from *C. litteratus* [73].

There are only ten conopeptides classified into the H-superfamily in the ConoServer database, including seven from *C. marmoreus* [44] and three from *C. victoriae* [74]. Our Bt043 is the first H-superfamily conopeptide

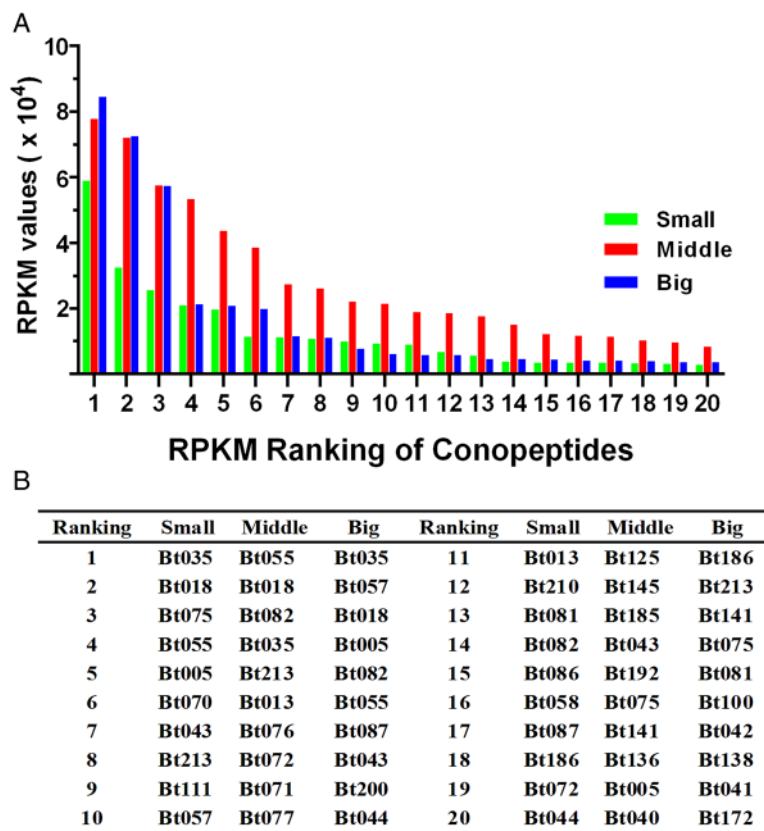


Fig. 5 The top 20 conopeptides (with the highest RPKM values) from three transcriptomes. **a** Comparison of the top 20 conopeptides from each of the Small, Middle and Big datasets. **b** The RPKM ranking of individual conopeptides within each of the datasets

Table 3 The conopeptides that are common among the three top 20s

Conopeptide	Ranking of RPKM			Superfamily	Possible activity ^a
	Small	Middle	Big		
Bt035	1	4	1	Conantokin-like	NMDA receptor inhibitors
Bt018	2	2	3	B2	unknown
Bt075	3	16	14	M	$\alpha, \iota, \kappa, \mu$
Bt055	4	1	6	I2	κ
Bt005	5	19	4	A	α
Bt043	7	14	8	H	$\delta, \gamma, \kappa, \mu, \omega$
Bt213	8	5	12	T	ϵ, μ, τ
Bt082	14	3	5	M	unknown

^aThe Greek letters denote pharmacological families defined in ConoServer as conopeptides sharing the same receptor specificities

identified in *C. betulinus*, with high expression levels and a classical VI/VII cysteine framework. On the other hand, all the known T-superfamily conotoxins in *C. betulinus* belong to cysteine framework pattern V and have four cysteines; however, our Bt213 has the longest mature peptide sequence, with 23–33 aa more than any previously reported in the same species (Additional files 1 and 4).

Differential expression of conopeptides in the venom duct and the venom bulb

To determine if conopeptides are transcribed in the venom bulb, we dissected this tissue away from the venom duct of the Middle specimen of *C. betulinus* (Fig. 6) for further transcriptome sequencing. The total number of unique conopeptide sequences identified in the venom bulb (39) was less than half of the number identified in the venom duct (94). In the Bulb dataset, 16 known superfamilies, four new superfamilies and two groups of conopeptides were identified. Most of the conopeptides (17) belong to the M and O1 superfamilies and the Conkunitzin group. Sequences of the A, B, C, F,



Fig. 6 Dissection of the venom duct and the venom bulb in *C. betulinus*. Although the two regions are morphologically connected, our transcriptomic data demonstrated differential expression of conopeptides between them

H, I, P and T superfamilies, as well as the conantokin-like group, were also observed. As expected, the expression levels of conopeptides identified in the venom bulb were far lower than those in the venom duct. The RPKM values of all conopeptides in the Bulb dataset were below 180; in contrast, in the Middle dataset, the highest and median RPKM values were 77,776 and 1021 respectively. We randomly picked several conopeptides (Bt018, Bt054, Bt055 and Bt082) and confirmed the differential expression between the venom bulb and the venom duct by RT-PCR (Fig. 7 and Additional file 5).

Between the two transcriptomes of venom duct and venom bulb from the Middle specimen, 28 conopeptides were revealed to be common, whereas 66 and 11 were unique to the venom duct and the venom bulb, respectively (Fig. 4c). Interestingly, among these 11 potential venom bulb-specific conopeptides, two M-superfamily members, Bt070 (Bt3-D05) and Bt089 (Bt3-TP04), had been identified in a previous study [71]; the other nine, belonging to nine different superfamilies or groups (B1, F, I1, O1, O2, O3, T, NSF-bt01 and Conkunitzin), were reported for the first time. When we compared the venom bulb dataset with all the datasets of the venom duct, nine of the sequences, including Bt070 (Bt3-D05) and Bt089 (Bt3-TP04), among these 11 potential unique conopeptides were revealed to be common to the two parts, whereas Bt048 and Bt168 remained unique to the venom bulb (Fig. 4d). However, there may have been a certain amount of contamination, since a tiny portion of the venom duct inside the venom bulb could not be removed and had to be considered as part of the venom bulb.

Discussion

Sanger sequencing technology previously generated 19 conotoxin cDNA sequences of *C. striatus* and 42 of *C. litteratus* [73, 75]. In contrast, next-generation sequencing, a relatively inexpensive and efficient technology, has been applied recently to the venom duct transcriptomes of several *Conus* species [44, 53, 58, 59, 74, 76] and between 61 and 136 conopeptide sequences, belonging to 11–30 superfamilies, were discovered. To investigate the diversity of conopeptides in a single species using high-throughput methodologies, we applied both traditional Sanger sequencing and the next-generation Illumina Hiseq2000 sequencing platforms to study the venom duct and the venom bulb transcriptomes of *C. betulinus*.

We report here over 200 conopeptide transcripts in this one *Conus* species. Based on traditional large-scale cloning of cDNA libraries and Sanger sequencing, only 46 conotoxin transcripts were detected from a mixed library of venom duct mRNAs. However, the number of conopeptide sequences identified

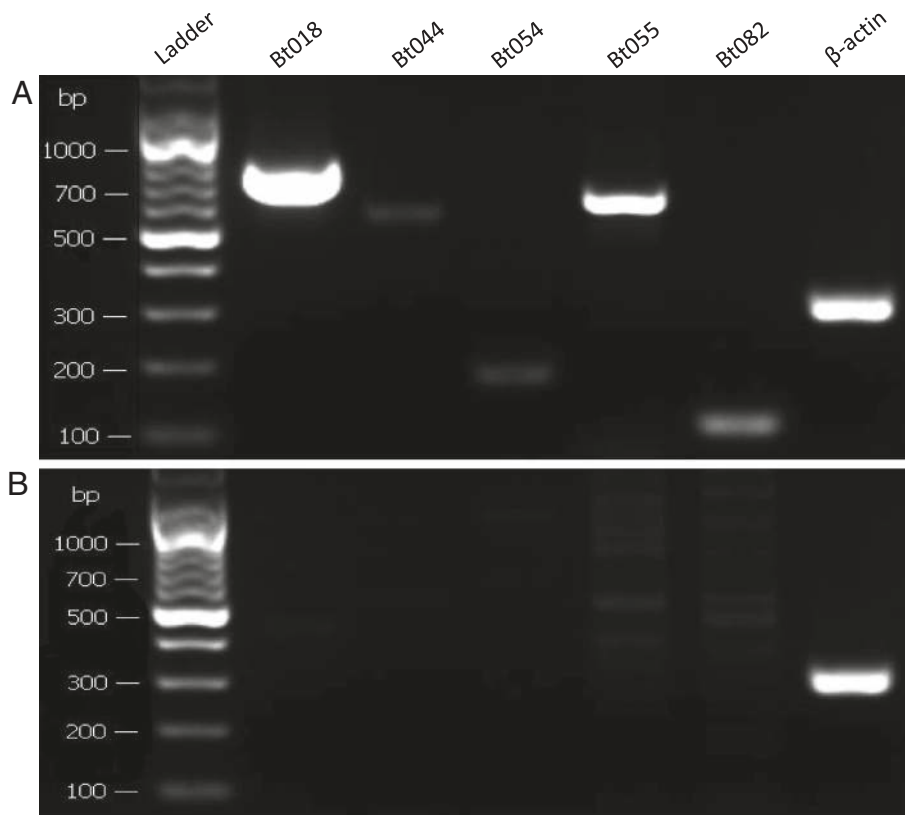


Fig. 7 Confirmation by RT-PCR of expression differences of five randomly selected conopeptides. The PCR templates were from the venom duct (a) and the venom bulb (b). Beta-actin was used as the internal control

from three non-normalized venom duct transcriptomes (from Small, Middle and Big specimens) and normalized venom duct/bulb transcriptomes (from the Middle specimen) was as high as 123 in each dataset with next-generation sequencing (Fig. 8). The dramatic difference between the two approaches once again demonstrates the value of this new transcriptome sequencing technology as a high-throughput method for discovering novel conotoxin genes. In addition, normalization during sample preparation can significantly increase the total conopeptide numbers compared to non-normalized libraries (from 94–98 up to 123; Fig. 8) but, as found here, additional non-overlap of transcripts was detected (Fig. 4d).

Our transcriptome of the venom bulb from *C. betulinus* demonstrates the presence of conotoxins in this tissue. Although only 39 conotoxin transcripts with very low expression levels were confirmed by transcriptome sequencing and RT-PCR, at least two of them (Bt048 and Bt168) were not identified from any venom duct dataset, which indicates they may be unique to the venom bulb. This provides confirmatory evidence for previous work comparing the venom duct and the venom bulb in which one-dimensional gel electrophoresis and RT-PCR showed

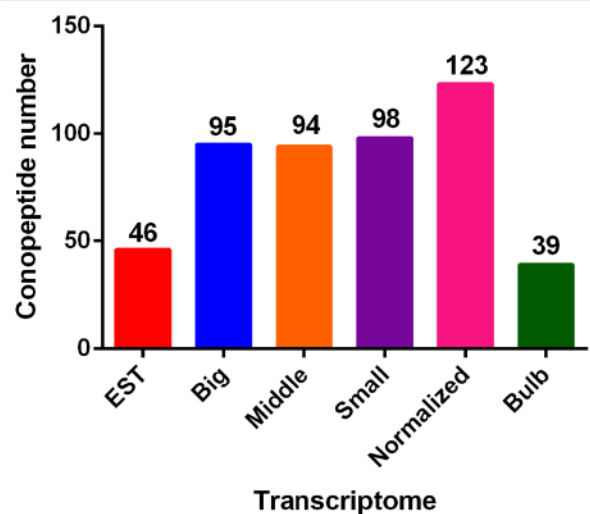
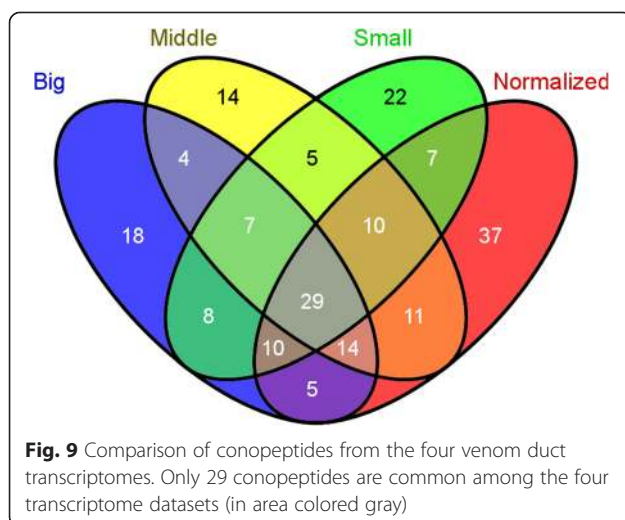


Fig. 8 Number of conopeptides identified from different transcriptomes. The differences could be the result of differential sequencing method, sample preparation method, and/or specimen and tissue differences. The Normalized transcriptome was assembled by SOAPdenovo-Trans 1.02 and the transcriptomes of Big, Middle, Small and Bulb were assembled using Trinity software. Evaluation of the two assemblers for the Normalized sample showed essentially equal performance.

expression levels (albeit low) of several proteins important for toxin biosynthesis, suggesting that the venom bulb engages in conotoxin biosynthesis [51].

After removal of duplications from the five transcriptomes and one EST sequencing dataset (Fig. 8), we obtained a total of 215 unique putative conopeptide transcript sequences, which were classified into 37 superfamilies and groups with at least one aa difference in the mature regions between each other. This represents a very high number of novel conopeptide transcripts discovered from a single species of *Conus*.

Previous studies have shown that venom composition varied dramatically even among individuals from the same species, and suggested the presence of intraspecific variability in the *Conus* venom peptides [51, 55, 77–81]. However, a recent report by Dutertre *et al.* [44] revealed that thousands of conotoxins may be derived from only hundreds of conopeptide genes. Our comparison of the venom duct transcriptomes of three body-sized *C. betulinus* (the Small, Middle and Big datasets) has authenticated the presence of 28 to 32 gene superfamilies in each individual. By comparing each pair of the three transcriptomes, we found that the number of common conopeptides ranged from 51 to 54 and the individual-specific conopeptides averaged 42 (40–47; Fig. 4a). A wider comparison, in Fig. 9, reduced the individual-specific numbers to 22, 14 and 18, respectively. Hence, it is likely to be possible to identify more novel conotoxins via the sequencing and comparison of additional specimens, and the total number of conopeptide genes existing in a species may be comfortably above early estimates that ranged from 50 to 200 conopeptides per *Conus* species. The exact numbers can be confirmed by whole genome sequencing, which is underway for *C. betulinus* in our laboratories.



Conclusions

This study is the first report to examine the diverse conopeptide expression repertoire in the vermivorous Chinese tubular cone snail, *Conus betulinus*. We used multi-transcriptome sequencing and complemented by traditional Sanger sequencing and a total of 215 unique putative conopeptide sequences were identified. As anticipated, most (183) of the identified conopeptides were novel, and nine new superfamilies were classified. We also demonstrated the differential expression of conopeptide genes among three individuals with different body sizes (at potentially different developmental stages), and our data suggest the existence of remarkable intraspecific variability in the venom duct. It is therefore probable that the discovery of more novel conotoxins will be accomplished via sequencing and analyzing additional specimens. Meanwhile, comparison of conopeptide expression in the venom duct and the venom bulb sheds light on the presence of conotoxins outside the venom duct. We demonstrated for the first time the existence of a handful of conotoxins in the venom bulb, although their expression levels were relatively low.

Methods

Sample collection and RNA extraction

Eight specimens of *Conus betulinus* (Additional file 6), 6–10 cm in length, were collected in the offshore areas of Sanya City, Hainan Province, China. Immediately after collection, the snails were placed on ice and dissected. Taxonomic identification was confirmed by COI sequences (DNA barcoding), which are identical to the reported data (GenBank accession numbers HQ834088.1, KJ549869.1, JN053043.1 and JF823627.1). Total RNA was extracted from the venom ducts of all specimens and the venom bulb of one middle-sized snail using TRIzol[®] LS Reagent (Invitrogen, Life Technologies, USA) according to the manufacturer's instructions. Isolation of mRNA molecules containing poly(A) tails was carried out via the use of oligo-(dT)-attached magnetic beads (Invitrogen). All experiments were performed in accordance with the guidelines of the Animal Ethics Committee and were approved by the Institutional Review Board on Bioethics and Biosafety of BGI (No. FT15103).

Construction and sequencing of cDNA libraries

To maximize the numbers of conopeptides identified from the specimens, three methods were applied to construct different cDNA libraries. The first was to construct a full-length cDNA library of mixed mRNAs from the venom ducts of six snails of various sizes, and around 11,000 clones were sequenced from 5' to 3' using an automated ABI 3730 sequencer. The second approach was to choose a medium-sized specimen

for construction of a normalized Illumina cDNA library. cDNAs were normalized using a duplex-specific nuclease (DSN) approach according to the DSN Normalization Sample Preparation Guide (Early Access Protocol, Part number 15014673 Rev. C, Illumina, 2010). Third, and most importantly, we constructed four non-normalized Illumina cDNA libraries using mRNAs from, respectively, the venom ducts of three snails with different body length and body weight, and a venom bulb from the middle-sized specimen. For nomenclature, the four non-normalized transcriptome datasets were named 'Big' (venom duct of a snail 10 cm in body length), 'Middle' (venom duct of a snail 8.7 cm in length), 'Small' (venom duct of a snail 6 cm in length) and 'Bulb' (venom bulb from the middle-sized snail). In addition, a normalized transcriptome of another medium-sized snail was referred to as the 'Normalized' dataset. The traditional transcriptome of cDNA libraries (reverse transcribed from total mRNAs, cloned, and sequenced by ABI 3730 Sanger methodology) was called the 'EST' dataset.

Sequence data processing (analysis and assembly)

In order to obtain high-quality clean reads for *de novo* assembly, the raw reads generated from transcriptome sequencing were filtered with the following steps: (1) adaptor sequences were removed; (2) reads with more than 10 % of unknown nucleotides were removed; (3) reads with more than 50 % of low-quality bases (base quality ≤ 10) were discarded. The clean reads that remained were assembled into unique genes using Trinity software [65] with an optimized k-mer length of 25 for *de novo* assembly, except for the Normalized transcriptome of venom duct that was assembled by SOAPdenovo-Trans 1.02 [66]. The expression of unique genes was calculated using RPKM, which is a general method of quantifying gene expression from RNA sequencing data by normalizing for total read length and the number of sequencing reads [82]. We represent the expression of each unique transcript using RPKM values, instead of sequencing depth/coverage, because the values are normalized and facilitate comparisons.

Meanwhile, raw sequences of the EST dataset from the ABI 3730 sequencing were trimmed by removal of vector sequences, primer sequences, and poly(A) tails with ABI Prism DNA Sequencing Analysis v5.4 software to obtain high-quality clean EST sequences. Redundant sequences were removed from the dataset.

Prediction and identification of conopeptides

We applied homology searches and an *ab initio* prediction method [83] to predict conotoxins from the five transcriptomes and one EST sequencing dataset.

For homology searches, all previously known conopeptides were downloaded from the ConoServer database [42] to construct a local reference dataset. We subsequently used BLASTX (with an E-value of $1e-5$) to run our assembled sequences against the local dataset. Those unique genes/ESTs with the best hits in the BLASTX data were translated into peptide sequences.

In addition, an *ab initio* prediction method using a Hidden Markov Model (HMM) was adopted to discover new conopeptides. First, the reference dataset of known conopeptides from the ConoServer database [42] was grouped into different classes according to their published superfamily and family classifications. Second, sequences of each class were aligned with the ClustalW tool [84] and the ambiguous results were checked using the ConoPrec tool [42] and manual inspection. Finally, a profile HMM was built for the conserved-domain of each class using *hmmbuild* from the HMMER 3.0 package [68] to find the best HMM parameter, and the *hmmsearch* tool was then applied, using this trained HMM parameter, to scan every unique assembled gene/EST for identification of conopeptides.

Classification of gene superfamilies

The predicted conotoxin transcripts were manually inspected using the ConoPrec tool implemented in the ConoServer database [42]. Those transcripts with duplication or truncated mature region sequences were removed. The signal peptides, gene superfamilies and cysteine frameworks of these predicted conopeptides were also checked for confirmation. Based on 75 % identity in the highly conserved signal peptide sequences [40], the conopeptide precursors could be assigned to most of the gene superfamilies present in the ConoServer database. Particular cut-off values were then used for some gene superfamilies with lower conservation of the signal region. The threshold values for assigning the conopeptides to I1, I2, L, M, P, S, con-ikot-ikot and divergent superfamilies were adjusted to 71.85, 57.6, 67.5, 69.3, 69.1, 72.9, 64.5 ± 20.2 and 64.22 ± 20.53 %, respectively [76]. If the conservation of a signal region was below the threshold value for any reported conotoxin superfamily, the conopeptide was regarded as a member of a new gene superfamily named in the form 'NSF-bt' plus an Arabic number suffix. Those conopeptides without signal regions but still showing similarity either in the pro- or mature region were considered as an 'Unknown' group.

Reverse transcription PCR

After extraction of total RNA from the venom duct and venom bulb of the Middle specimen, cDNAs were

reverse transcribed using the M-MuLV First Strand cDNA Synthesis Kit (Sangon, China). We randomly selected five conopeptides and employed Primer Premier 5.0 to design primer pairs (see detailed nucleotide sequences in Additional file 5). Reverse transcription PCR (RT-PCR) was performed in 50- μ l reactions, containing 0.5 μ l of cDNA, 0.5 μ l of rTaq DNA Polymerase (Takara, Japan), 1 \times PCR reaction buffer (Takara), 200 μ M of each dNTP, and 0.2 μ M of forward and reverse primers. The targeted DNAs were amplified in an ABI 9700 thermal cycler (Life Technologies, USA) as follows: initial denaturation at 95 °C for 5 min; 35 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 45 s; final extension at 72 °C for 10 min. Beta-actin was used as an internal control. All the PCR amplicons were checked by 1.5 % agarose gel electrophoresis for comparison of relative expression levels.

Availability of supporting data

The datasets supporting the results of this article are included within the article and its additional files. The transcriptome reads produced in this study have been deposited in the NCBI SRA database with accession numbers SRS1009725 for the Big dataset, SRS1009729 for the Middle dataset, SRS1009726 for the Small dataset, SRS1009727 for the Normalized dataset, and SRS1009728 for the Bulb dataset. The clean reads for 11,026 clones were sequenced by ABI 3730 and submitted to the NCBI as EST data (PRJNA290540). Additional file 7 provides the translated sequences of conopeptides identified from the EST dataset; all of the 215 transcripts identified have been submitted to GenBank (accession numbers are included in Additional file 8). Additional supporting data is also hosted in the *GigaScience* GigaDB repository [85], as well as linked to NCBI bioproject number PRJNA290540.

Additional files

- Additional file 1:** Summary of previously reported conopeptides in *C. betulinus*. (DOC 102 kb)
- Additional file 2:** The alignment file used to generate the profile HMM (pHMM). (RAR 10 kb)
- Additional file 3:** The actual pHMM file created for identifying conopeptides. (RAR 225 kb)
- Additional file 4:** The FASTA file for the 215 conopeptides identified by us from *C. betulinus*. (TXT 17 kb)
- Additional file 5:** Primer sequences used for RT-PCR (5' to 3'). (DOCX 13 kb)
- Additional file 6:** A picture of *Conus betulinus* (Linnaeus) native to the South China Sea. (JPG 4261 kb)
- Additional file 7:** List of conopeptide coding sequences identified in the EST dataset. (XLSX 15 kb)
- Additional file 8:** Accession numbers of the 215 identified conopeptides in the NCBI GenBank. (XLSX 15 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CP and QS designed the work and wrote the manuscript. GY, BG, CF and J-SC helped in planning and sketching the manuscript. CP, GY, JW, YZ, XfZ, CB and XhZ carried out bioinformatics and statistical analyses of the transcriptional data. CP, BG, BW, ZR, XY, JB, JL, ZL, SZ, YQ and JC performed experiments. YC, SLC and JY participated in revising the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by China 863 Program (No. 2014AA093501), Shenzhen Dapeng Special Program for Industrial Development (No. KY20150207), Special Project on the Integration of Industry, Education and Research of Guangdong Province (No. 2013B090800017), Shenzhen Scientific R&D Program (No. CXB201108250095A), State Key Laboratory of Agricultural Genomics (No. ZDSY20120618171817275), and the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development at the National Institutes of Health, USA.

Author details

¹BGI-Shenzhen, Shenzhen 518083, China. ²Research Institute of Pharmaceutical Chemistry, Beijing 102205, China. ³School of Pharmaceutical Sciences, Hainan Medical University, Haikou 571199, China. ⁴Shenzhen Key Laboratory of Marine Genomics, Guangdong Provincial Key Laboratory of Molecular Breeding in Marine Economic Animals, State Key Laboratory of Agricultural Genomics, Shenzhen 518083, China. ⁵Micro Pharmatech Ltd, Wuhan 430075, China. ⁶Molecular Genomics Laboratory, National Institutes of Health, Bethesda, MD 20892, USA. ⁷BGI-Zhenjiang Institute of Hydrobiology, Zhenjiang 212000, China.

Received: 14 August 2015 Accepted: 7 April 2016

Published online: 14 April 2016

References

- Bouchet P, Kantor YI, Sysoev A, Puillandre N. A new operational classification of the conoidea (Gastropoda). *J Molluscan Stud.* 2011;77: 273–308.
- Modica MV, Holford M. The Neogastropoda: evolutionary innovations of predatory marine snails with remarkable pharmacological potential. In: Pontarotti P, editor. *Evolutionary Biology - Concepts, Molecular and Morphological Evolution.* Heidelberg: Springer; 2010. p. 249–70.
- Olivera BM. Conus venom peptides, receptor and ion channel targets and drug design: 50 million years of neuropharmacology. *Mol Biol Cell.* 1997;8: 2101–9.
- Röckel D, Korn W, Kohn AJ. *Manual of the living Conidae.* Wiesbaden: Verlag Christa Hemmen; 1995.
- Dutertre S, Jin AH, Vetter I, Hamilton B, Sunagar K, Lavergne V, et al. Evolution of separate predation- and defence-evoked venoms in carnivorous cone snails. *Nat Commun.* 2014;5:3521.
- Terlau H, Olivera BM. Conus venoms: a rich source of novel ion channel-targeted peptides. *Physiol Rev.* 2004;84:41–68.
- Kohn AJ. The ecology of *Conus* in Hawaii. *Ecol Monogr.* 1959;29:47–90.
- Duda Jr TF, Kohn AJ. Species-level phylogeography and evolutionary history of the hyperdiverse marine gastropod genus *Conus*. *Mol Phylogenet Evol.* 2005;34:257–72.
- Endean R, Duchemin C. The venom apparatus of *Conus magus*. *Toxicon.* 1967;4:275–84.
- Salisbury SM, Martin GG, Kier WM, Schulz JR. Venom kinematics during prey capture in *Conus*: the biomechanics of a rapid injection system. *J Exp Biol.* 2010;213:673–82.
- Schulz JR, Norton AG, Gilly WF. The projectile tooth of a fish-hunting cone snail: *Conus catus* injects venom into fish prey using a high-speed ballistic mechanism. *Biol Bull.* 2004;207:77–9.
- Lewis RJ, Dutertre S, Vetter I, Christie MJ. Conus venom peptide pharmacology. *Pharmacol Rev.* 2012;64:259–98.
- Olivera BM, Cruz LJ. Conotoxins, in retrospect. *Toxicon.* 2001;39:7–14.
- Norton RS, Olivera BM. Conotoxins down under. *Toxicon.* 2006;48:780–98.

15. England LJ, Imperial J, Jacobsen R, Craig AG, Gulyas J, Akhtar M, et al. Inactivation of a serotonin-gated ion channel by a polypeptide toxin from marine snails. *Science*. 1998;281:575–8.
16. Sharpe IA, Gehrman J, Loughnan ML, Thomas L, Adams DA, Atkins A, et al. Two new classes of conopeptides inhibit the alpha1-adrenoceptor and noradrenaline transporter. *Nat Neurosci*. 2001;4:902–7.
17. McIntosh JM, Santos AD, Olivera BM. Conus peptides targeted to specific nicotinic acetylcholine receptor subtypes. *Annu Rev Biochem*. 1999;68:59–88.
18. Olivera BM. Conus peptides: biodiversity-based discovery and exogenomics. *J Biol Chem*. 2006;281:31173–7.
19. Jones RM, Bulaj G. Conotoxins - new vistas for peptide therapeutics. *Curr Pharm Des*. 2000;6:1249–85.
20. Lewis RJ, Garcia ML. Therapeutic potential of venom peptides. *Nat Rev Drug Discov*. 2003;2:790–802.
21. Layer RT, McIntosh JM. Conotoxins: therapeutic potential and application. *Mar Drugs*. 2006;4:119–42.
22. Carstens BB, Clark RJ, Daly NL, Harvey PJ, Kaas Q, Craik DJ. Engineering of conotoxins for the treatment of pain. *Curr Pharm Des*. 2011;17:4242–53.
23. Lewis RJ. Conotoxin venom peptide therapeutics. *Adv Exp Med Biol*. 2009;655:44–8.
24. Gayler K, Sandall D, Greening D, Keays D, Polidano M, Livett B, et al. Molecular prospecting for drugs from the sea. Isolating therapeutic peptides and proteins from cone snail venom. *IEEE Eng Med Biol Mag*. 2005;24:79–84.
25. Halai R, Craik DJ. Conotoxins: natural product drug leads. *Nat Prod Rep*. 2009;26:526–36.
26. Leary D, Vierros M, Hamon G, Arico S, Monagle C. Marine genetic resources: a review of scientific and commercial interest. *Mar Policy*. 2009;33:183–94.
27. Molinski TF, Dalisay DS, Lievens SL, Saludes JP. Drug development from marine natural products. *Nat Rev Drug Discov*. 2009;8:69–85.
28. Han TS, Teichert RW, Olivera BM, Bulaj G. Conus venoms - a rich source of peptide-based therapeutics. *Curr Pharm Des*. 2008;14:2462–79.
29. Lewis RJ. Discovery and development of the chi-conopeptide class of analgesic peptides. *Toxicon*. 2012;59:524–8.
30. Miljanich GP. Ziconotide: neuronal calcium channel blocker for treating severe chronic pain. *Curr Med Chem*. 2004;11:3029–40.
31. Lynch SS, Cheng CM, Yee JL. Intrathecal ziconotide for refractory chronic pain. *Ann Pharmacother*. 2006;40:1293–300.
32. McGivern JG. Ziconotide. A review of its pharmacology and use in the treatment of pain. *Neuropsychiatr Dis Treat*. 2007;3:69–85.
33. McIntosh JM, Jones RM. Cone venom—from accidental stings to deliberate injection. *Toxicon*. 2001;39:1447–51.
34. King GF. Venoms as a platform for human drugs: translating toxins into therapeutics. *Expert Opin Biol Ther*. 2011;11:1469–84.
35. Olivera BM, Teichert RW. Diversity of the neurotoxic Conus peptides: a model for concerted pharmacological discovery. *Mol Interv*. 2007;7:251–60.
36. Wang CZ, Chi CW. Conus peptides—a rich pharmaceutical treasure. *Acta Biochim Biophys Sin Shanghai*. 2004;36:713–23.
37. Azam L, Dowell C, Watkins M, Stitzel JA, Olivera BM, McIntosh JM. Alpha-conotoxin BulA, a novel peptide from *Conus bullatus*, distinguishes among neuronal nicotinic acetylcholine receptors. *J Biol Chem*. 2005;280:80–7.
38. Woodward SR, Cruz LJ, Olivera BM, Hillyard DR. Constant and hypervariable regions in conotoxin prepeptides. *EMBO J*. 1990;9:1015–20.
39. Olivera BM, Walker C, Cartier GE, Hooper D, Santos AD, Schoenfeld R, et al. Speciation of cone snails and interspecific hyperdivergence of their venom peptides. Potential evolutionary significance of introns. *Ann N Y Acad Sci*. 1999;870:223–37.
40. Kaas Q, Westermann JC, Craik DJ. Conopeptide characterization and classifications: An analysis using ConoServer. *Toxicon*. 2010;55:1491–509.
41. Kaas Q, Westermann JC, Halai R, Wang CK, Craik DJ. ConoServer, a database for conopeptide sequences and structures. *Bioinformatics*. 2008;24:445–6.
42. Kaas Q, Yu R, Jin AH, Dutertre S, Craik DJ. ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Res*. 2012;40:D325–30. doi:10.1093/nar/gkr886.
43. Puillandre N, Koua D, Favreau P, Olivera BM, Stocklin R. Molecular phylogeny, classification and evolution of conopeptides. *J Mol Biol*. 2012;74:297–309.
44. Dutertre S, Jin AH, Kaas Q, Jones A, Alewood PF, Lewis RJ. Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol Cell Proteomics*. 2013;12:312–29.
45. Luo S, Christensen S, Zhangsun D, Wu Y, Hu Y, Zhu X, et al. A novel inhibitor of alpha9alpha10 nicotinic acetylcholine receptors from *Conus vexillum* delineates a new conotoxin superfamily. *PLoS ONE*. 2013;8:e54648. doi:10.1371/journal.pone.0054648.
46. Aguilar MB, Ortiz E, Kaas Q, Lopez-Vera E, Becerril B, Possani LD, et al. Precursor De13.1 from *Conus delessertii* defines the novel G gene superfamily. *Peptides*. 2013;41:17–20.
47. Ye M, Khoo KK, Xu S, Zhou M, Boonyalai N, Perugini MA, et al. A helical conotoxin from *Conus imperialis* has a novel cysteine framework and defines a new superfamily. *J Biol Chem*. 2012;287:14973–83.
48. Espiritu DJ, Watkins M, Dia-Monje V, Cartier GE, Cruz LJ, Olivera BM. Venomous cone snails: molecular phylogeny and the generation of toxin diversity. *Toxicon*. 2001;39:1899–916.
49. Biggs JS, Watkins M, Puillandre N, Ownby JP, Lopez-Vera E, Christensen S, et al. Evolution of Conus peptide toxins: analysis of *Conus californicus* Reeve, 1844. *Mol Phylogenet Evol*. 2010;56:1–12.
50. Bernáldez J, Román-González SA, Martínez O, Jiménez S, Vivas O, Arenas I, et al. A *Conus regularis* conotoxin with a novel eight-cysteine framework inhibits CaV2.2 channels and displays an anti-nociceptive activity. *Mar Drugs*. 2013;11:1188–202.
51. Safavi-Hemami H, Young ND, Williamson NA, Purcell AW. Proteomic interrogation of venom delivery in marine cone snails: novel insights into the role of the venom bulb. *J Proteome Res*. 2010;9:5610–9.
52. Marshall J, Kelley WP, Rubakhin SS, Bingham JP, Sweedler JV, Gilly WF. Anatomical correlates of venom production in *Conus californicus*. *Biol Bull*. 2002;203:27–41.
53. Hu H, Bandyopadhyay P, Olivera B, 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. doi:10.1186/1471-2164-13-284.
54. Biass D, Dutertre S, Gerbault A, Menou JL, Offord R, Favreau P, et al. Comparative proteomic study of the venom of the piscivorous cone snail *Conus consors*. *J Proteomics*. 2009;72:210–8.
55. Davis J, Jones A, Lewis RJ. Remarkable inter- and intraspecies complexity of conotoxins revealed by LC/MS. *Peptides*. 2009;30:1222–7.
56. Prashanth JR, Lewis RJ, Dutertre S. Towards an integrated venomics approach for accelerated conopeptide discovery. *Toxicon*. 2012;60:470–7.
57. Hu H, Bandyopadhyay PK, Olivera BM, Yandell M. Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. *BMC Genomics*. 2011;12:60. doi:10.1186/1471-2164-12-60.
58. Lluisma AO, Milash BA, Moore B, Olivera BM, Bandyopadhyay PK. Novel venom peptides from the cone snail *Conus pulicarius* discovered through next-generation sequencing of its venom duct transcriptome. *Mar Genomics*. 2012;5:43–51.
59. Terrat Y, Biass D, Dutertre S, Favreau P, Remm M, Stocklin R, et al. High-resolution picture of a venom gland transcriptome: case study with the marine snail *Conus consors*. *Toxicon*. 2012;59:34–46.
60. Barghi N, Concepcion GP, Olivera BM, Lluisma AO. Comparison of the Venom Peptides and Their Expression in Closely Related Conus Species: Insights into Adaptive Post-speciation Evolution of Conus Exogenomes. *Genome Biol Evol*. 2015;7:1797–814.
61. Himaya SWA, Jin AH, Dutertre S, Giacomotto J, Mohialdeen H, Vetter I, et al. Comparative Venomics Reveals the Complex Prey Capture Strategy of the Piscivorous Cone Snail *Conus catus*. *J Proteome Res*. 2015;14:4372–81.
62. Lavergne V, Harliwong I, Jones A, Miller D, Taft RJ, Alewood PF. Optimized deep-targeted proteotranscriptomic profiling reveals unexplored Conus toxin diversity and novel cysteine frameworks. *Proc Natl Acad Sci U S A*. 2015;112:E3782–91.
63. Kohn AJ, Perron FE. *Life History and Biogeography: Patterns in Conus*. Oxford: Clarendon Press; 1994.
64. Jin AH, Dutertre S, Kaas Q, Lavergne V, Kubala P, Lewis RJ, et al. Transcriptomic messiness in the venom duct of *Conus miles* contributes to conotoxin diversity. *Mol Cell Proteomics*. 2013;12:3824–33.
65. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29:644–52.
66. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, et al. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*. 2014;30:1660–6.

67. Perteau G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, et al. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*. 2003;19:651–2.
68. Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs. *Bioinformatics*. 2013;29:2487–9.
69. Watkins M, Olivera BM, Hillyard DR, McIntosh JM, Jones RM. Alpha-conotoxin peptides. 2002. Patent: JP 2002534996-A 11 22-OCT-2002.
70. McIntosh JM, Plazas PV, Watkins M, Gomez-Casati ME, Olivera BM, Elgoyhen AB. A novel alpha-conotoxin, PeIA, cloned from *Conus pergrandis*, discriminates between rat alpha9alpha10 and alpha7 nicotinic cholinergic receptors. *J Biol Chem*. 2005;280:30107–12.
71. Zhou M, Wang L, Wu Y, Zhu X, Feng Y, Chen Z, et al. Characterizing the evolution and functions of the M-superfamily conotoxins. *Toxicon*. 2013;76:150–9.
72. Yuan DD, Han YH, Wang CG, Chi CW. From the identification of gene organization of alpha conotoxins to the cloning of novel toxins. *Toxicon*. 2007;49:1135–49.
73. Pi C, Liu J, Peng C, Liu Y, Jiang X, Zhao Y, et al. Diversity and evolution of conotoxins based on gene expression profiling of *Conus litteratus*. *Genomics*. 2006;88:809–19.
74. Robinson SD, Safavi-Hemami H, McIntosh LD, Purcell AW, Norton RS, Papenfuss AT. Diversity of conotoxin gene superfamilies in the venomous snail, *Conus victoriae*. *PLoS ONE*. 2014;9:e87648. doi:10.1371/journal.pone.0087648.
75. Pi C, Liu Y, Peng C, Jiang X, Liu J, Xu B, et al. Analysis of expressed sequence tags from the venom ducts of *Conus striatus*: focusing on the expression profile of conotoxins. *Biochimie*. 2006;88:131–40.
76. Barghi N, Concepcion GP, Olivera BM, Lluisma AO. High Conopeptide Diversity in *Conus tribblei* Revealed Through Analysis of Venom Duct Transcriptome Using Two High-Throughput Sequencing Platforms. *Mar Biotechnol (NY)*. 2015;17:81–98.
77. Lavergne V, Dutertre S, Jin AH, Lewis RJ, Taft RJ, Alewood PF. Systematic interrogation of the *Conus marmoreus* venom duct transcriptome with ConoSorter reveals 158 novel conotoxins and 13 new gene superfamilies. *BMC Genomics*. 2013;14:708. doi:10.1186/1471-2164-14-708.
78. Jakubowski JA, Kelley WP, Sweedler JV, Gilly WF, Schulz JR. Intraspecific variation of venom injected by fish-hunting *Conus* snails. *J Exp Biol*. 2005;208:2873–83.
79. Rivera-Ortiz JA, Cano H, Marí F. Intraspecific variability and conopeptide profiling of the injected venom of *Conus ermineus*. *Peptides*. 2011;32:306–16.
80. Dutertre S, Biass D, Stocklin R, Favreau P. Dramatic intraspecific variations within the injected venom of *Conus consors*: an unsuspected contribution to venom diversity. *Toxicon*. 2010;55:1453–62.
81. Romeo C, Di Francesco L, Oliverio M, Palazzo P, Massilia GR, Ascenzi P, et al. *Conus ventricosus* venom peptides profiling by HPLC-MS: a new insight in the intraspecific variation. *J Sep Sci*. 2008;31:488–98.
82. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat Methods*. 2008;5:621–8.
83. 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;1824:488–92.
84. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23:2947–8.
85. Peng C, Yao G, Gao B, Fan C, Bian C, Wang J, Cao Y, Wen B, Zhu Y, Ruan Z, Zhao X, You X, Bai J, Li J, Lin Z, Zou S, Zhang X, Qiu Y, Chen J, Coon SL, Yang J, Chen J, Shi Q. Supporting material for “High throughput identification of novel conotoxins from the Chinese tubular cone snail (*Conus betulinus*) by multi-transcriptome sequencing”. *GigaScience Database*. 2016; <http://dx.doi.org/10.5524/100169>

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

