

Molecular Phylogeny, Classification and Evolution of Conopeptides

N. Puillandre · D. Koua · P. Favreau ·
B. M. Olivera · R. Stöcklin

Received: 7 February 2012 / Accepted: 12 June 2012 / Published online: 4 July 2012
© Springer Science+Business Media, LLC 2012

Abstract Conopeptides are toxins expressed in the venom duct of cone snails (Conoidea, *Conus*). These are mostly well-structured peptides and mini-proteins with high potency and selectivity for a broad range of cellular targets. In view of these properties, they are widely used as pharmacological tools and many are candidates for innovative drugs. The conopeptides are primarily classified into superfamilies according to their peptide signal sequence, a classification that is thought to reflect the evolution of the multigenic system. However, this hypothesis has never been thoroughly tested. Here we present a phylogenetic analysis of 1,364 conopeptide signal sequences extracted from GenBank. The results validate the current conopeptide superfamily classification, but also reveal several important new features. The so-called “cysteine-poor” conopeptides are revealed to be closely related to “cysteine-rich” conopeptides; with some of them sharing very similar signal sequences, suggesting that a distinction based on cysteine content and configuration is not phylogenetically relevant and does not reflect the evolutionary history of conopeptides. A given cysteine pattern or

pharmacological activity can be found across different superfamilies. Furthermore, a few conopeptides from GenBank do not cluster in any of the known superfamilies, and could represent yet-undefined superfamilies. A clear phylogenetically based classification should help to disentangle the diversity of conopeptides, and could also serve as a rationale to understand the evolution of the toxins in the numerous other species of conoideans and venomous animals at large.

Keywords Cone snails · *Conus* · Conoidea · Cys-pattern · Venom · Molecular evolution

Introduction

Cone snails of the genus *Conus* are predatory venomous marine mollusks feeding on fish, worm or snails. After decades of biological prospecting, conopeptides expressed in their venom duct have emerged as one of the richest and most promising marine sources of natural products (Blunt et al. 2012). The analysis of cone snail venoms has revealed a complex exogenome that is characterized by an extremely high level of diversity. With more than 600 described *Conus* species, each producing an estimated 100–200 venom components, the ensemble of cone snails were, until recently, estimated to produce between 50,000 and 100,000 different toxins (Menez et al. 2006; Olivera 2006). Recent studies, however, clearly demonstrate that this figure is an underestimation, probably by a factor of ten or so, with several new species described every year, more venom components detected in each sample using evolving technologies such as mass spectrometry (Biass et al. 2009; Ueberheide et al. 2009; unpublished results) and NextGen sequencing (Hu et al. 2011; Terrat et al.

Electronic supplementary material The online version of this article (doi:10.1007/s00239-012-9507-2) contains supplementary material, which is available to authorized users.

N. Puillandre (✉) · D. Koua · P. Favreau · R. Stöcklin
Atheris Laboratories, Case postale 314, Bernex, 1233 Geneva,
Switzerland
e-mail: nicolas.puillandre@atheris.ch

N. Puillandre · B. M. Olivera
Department of Biology, University of Utah,
257 South 1400 East, Salt Lake City, UT 84112, USA

D. Koua
Swiss Institute of Bioinformatics, 1206 Geneva, Switzerland

2011) or combinations thereof (Violette et al. 2012), and marked intra-species and even intra-specimen variations in venom composition (Davis et al. 2009; Dutertre et al. 2010; Jakubowski et al. 2005). It is now estimated that the number of cone snail venom components exceeds one million.

An important characteristic of conopeptides which makes them attractive for drug development is their high selectivity for molecular targets that span a broad range of therapeutic applications (Gayler et al. 2005; Leary et al. 2009; Molinski et al. 2009). So far, the conopeptide MVIIA (SNX-111, Prialt, or Ziconotide) from *Conus magus* (the magician cone) that selectively blocks Cav2.2 N-type voltage-gated calcium channels has been approved for the treatment of severe chronic pain (McGivern 2007; Miljanich 2004) and there are more promising drug candidates in the pipeline (e.g., see Favreau et al. 2012; Han et al. 2008a; Lewis 2012). The potential of this rich source of pharmacological products has stimulated a race for the discovery of new toxins. From the traditional bioactivity-guided identification, lead discovery efforts have evolved toward modern structure-driven characterization (venom peptidomics and proteomics, venom gland transcriptomics, targeted genomics, structure–function studies) and biocomputing-assisted analyses (proprietary databases and bioinformatic tools) (Daly and Craik 2009; Favreau and Stöcklin 2009; Koua et al. 2012; Laht et al. 2011). In addition, phylogenetic approaches have recently emerged as an effective way to quickly identify divergent lineages that are likely to have evolved with different functional characteristics. This approach to identify these previously uncharacterized conopeptides is referred to as concerted discovery (Conticello et al. 2001; Duda and Remigio 2008; Olivera 2006; Puillandre and Holford 2010).

However, despite the effectiveness of phylogenetic approaches in concerted discovery, the technique is rarely used for the classification of conopeptides (but see Aguilar et al. 2009; Conticello et al. 2001; Wang et al. 2008; Zhangsun et al. 2006). Several statistical methods for conopeptide classification, such as Mahalanobis (Lin and Li 2007) or BLAST and Euclidian distances among others (Mondal et al. 2006) have been described; however, most of these approaches are primarily designed for classification of new sequences rather than for testing the current classification (i.e., checking the validity of each known group by a blind-exploratory approach). Conopeptide precursors are characterized by a typical structural organization consisting of a highly conserved signal region, followed by a more variable pro-region and a hyper-variable mature toxin containing a few conserved amino acids such as the cysteine residues required for disulfide bonds. Conopeptides are mainly named and classified according to three properties: first, they are characterized by their signal

sequence, this short sequence (~20 amino-acids) is highly conserved, and has been used to define superfamilies; second, mature toxins structural families are characterized depending on their pattern of cysteines (the Cys-pattern), for example, the mature toxin can include a variable number of cysteines (most commonly 4 or 6), and their respective position can vary (4 cysteines can be organized as C–C–C–C or CC–C–C where “–” represents a variable number of amino-acids); finally, several conopeptides have also been characterized according to their molecular targets, referred to hereafter as “functional families,” and also previously termed “pharmacological families.”

In a recent paper, Kaas et al. (2010) reviewed the structure, function, and diversity of conopeptides on the ConoServer database (www.conoserver.org). In particular, they proposed that “the ‘gene superfamily’ classification scheme focuses on evolutionary relationships between conopeptides”, while the two other classification schemes (cysteine framework and function) do not. Their underlying hypothesis was that similarities in the Cys-pattern or function might have arisen by convergence. While we fully agree with this statement, we also argue that it could serve as a rationale to assess the congruence between the current gene superfamily classification and the evolution of the corresponding multigenic system, and to accurately demonstrate that convergence phenomena are common in conopeptide structure and function.

Here, we review the current superfamily classification of conopeptides by analyzing all the signal sequences available in GenBank using a phylogenetic approach to check: (i) if all the defined superfamilies correspond to homogeneous groups; and (ii) if all the GenBank signal sequences belong to a known superfamily. This study seeks to provide a “rationale” for a phylogenetic classification of conopeptides and to clarify their current classification, thus complementing the work initiated by Kaas et al. (2010).

Materials and Methods

Sequences from GenBank

Since the signal sequences used for phylogenetic analyses (see below), are only found on complete nucleotide precursors and are not known for conopeptide discovered using proteomic approaches, all the nucleotide sequences associated with the genus *Conus* were downloaded from GenBank (www.ncbi.nlm.nih.gov). The sequences corresponding to non-coding regions, ribosomal genes, mitochondrial genes, and genes with a function that did not relate to toxin activity were removed from the dataset, thus keeping only coding genes with a potential toxin activity. Only sequences obtained from *Conus* species belonging to

the large major clade (Duda and Kohn 2005) were conserved, as a large number of the conopeptides found in species from other clades (e.g. *C. californicus*) are highly divergent and do not match with any of the currently known superfamilies (Biggs et al. 2010; www.conoserver.org). Consequently, the classification in the present analysis is relevant only for conopeptides of the large major clade species. Conopeptide superfamilies are defined by a conserved signal sequence, thus we used the Signalp 3.0 server (Bendtsen et al. 2004) to identify the signal sequence; all sequences that did not include at least 50 % of the signal region were removed, together with sequences including a stop codon. Only the signal region was used for phylogenetic analyses, as only this part of the conopeptides can be aligned within and, to some extent, between superfamilies.

Phylogenetic Analysis

Aligning signal sequences between highly divergent conopeptides (i.e., belonging to different superfamilies) is arduous, and homology hypotheses are doubtful. Thus, sequences were translated to amino acids and automatically aligned using two different algorithms: Muscle (Edgar 2004 www.ebi.ac.uk/Tools/msa/muscle) and ClustalW (<http://clustalw.ddbj.nig.ac.jp/top-e.html>). Best model of evolution for these two datasets was selected using Modelgenerator V.85 (Keane et al. 2006) following the corrected Akaike Information Criterion (with four discrete gamma categories) and used to reconstruct phylogenetic trees. The best model of evolution identified by Modelgenerator was JTT + G (Jones Taylor Thornton model, implemented under the name “Jones model” in MrBayes—Jones et al. 1992) for both datasets. Bayesian analyses were performed by running two parallel analyses in MrBayes (Huelsenbeck et al. 2001), each consisting of eight Markov chains of 30,000,000 generations each with a sampling frequency of one tree every ten thousand generations. The number of swaps was set to 5, and the chain temperature at 0.02. A neighbor-joining tree obtained with MEGA5 (Tamura et al. 2011) was used as starting tree. Convergence of the parameters was evaluated using Tracer 1.4.1 (Rambaut and Drummond 2007), and analyses were terminated when ESS values were all superior to 200. A consensus tree was then calculated after omitting the first 25 % trees as burn-in.

As is the case for most multigenic families, the identification of an outgroup was highly problematic. No gene phylogenetically related to, and proven to be an outgroup for conopeptides has been described. Furthermore, the use of toxins from other conoidean species was not possible, as it would require that the toxins from cone snails all arose from duplication events that took place after the divergence

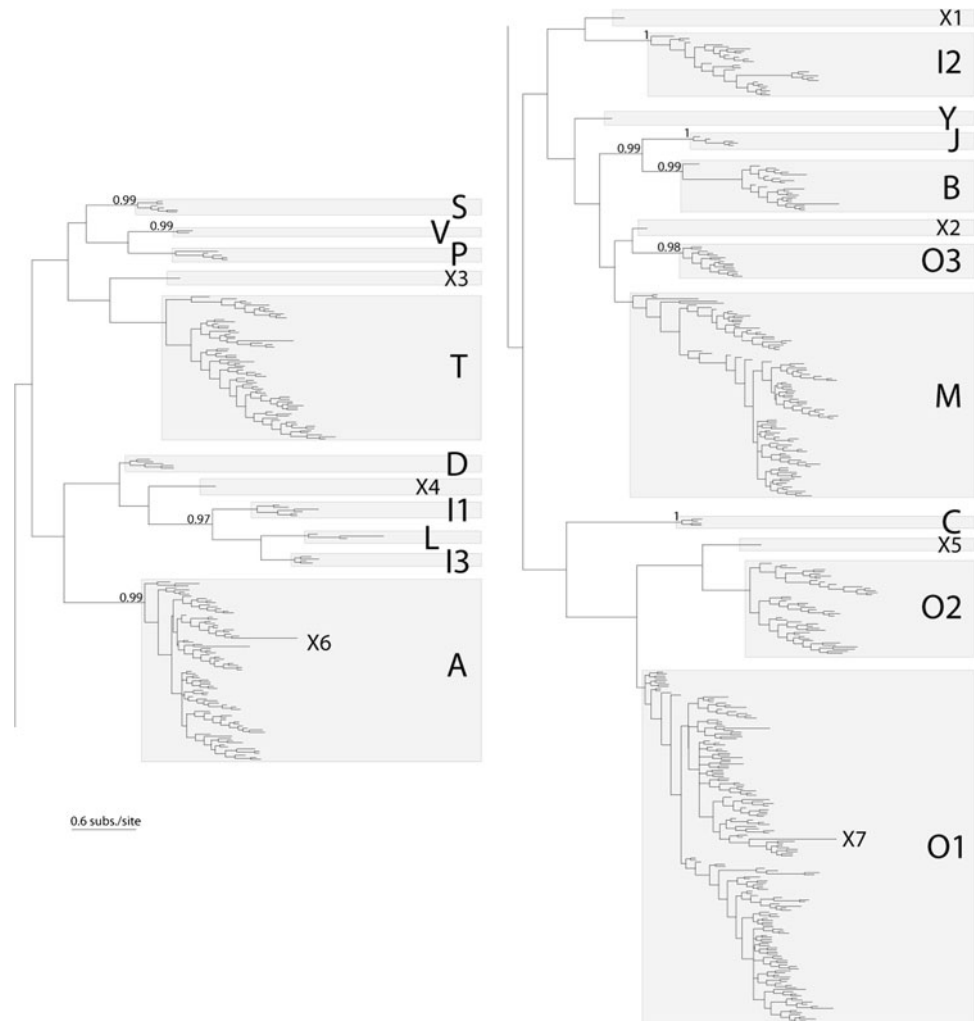
between the cone snails and other conoideans. Consequently, no outgroup was included in the analysis. This absence of an outgroup did not allow us to infer ancestor/descendant relationships.

Results

A total of 1,364 sequences potentially corresponding to conopeptides and with a signal sequence were downloaded from GenBank (performed on 1st of July, 2011). Alignments were 34 and 30 amino-acids long with Muscle and Clustal W, respectively. To limit the time of calculation for phylogenetic analysis, only one sequence per amino-acid haplotype was kept; finally, 585 sequences were retained. Overall, the phylogenetic trees obtained from the Muscle and Clustal alignments were congruent; discrepancies were not supported (posterior probabilities <0.90) and concerned phylogenetic relationships between the main clades and the position of a few highly divergent sequences (see details below). For clarity, only the phylogenetic tree based on the Clustal alignment is presented (Fig. 1) but the results obtained from the Muscle alignment, when different, are discussed.

Using information from GenBank and the literature, it was possible to link the clades defined with the bayesian analysis to known superfamilies. Most of the defined superfamilies (A, D, I1, I2, I3, J, L, O1, O3, P, S, T, V) corresponded to monophyletic groups, with some highly supported (Fig. 1). With the Muscle alignment, the O2 superfamily was included within the O1 superfamily; the superfamily Y was represented by a single sequence, and corresponded to a unique lineage in the tree. However, some superfamilies did not correspond to a monophyletic group, as they included other conopeptides (e.g., O2 included sequences of contryphans, and M included conomarphin—a result already discussed by Han et al. 2008b). Several conopeptides from GenBank did not cluster in any of the known superfamilies. These corresponded to known cysteine-poor conopeptides, contulakin, and conantokin, shown in Fig. 1 as the B and C superfamilies, respectively (the C superfamily has been previously defined by Jimenez et al. (2007)); two conoCAP sequences (FN868446.1 and FN868447.1—named X1 in the Fig. 1 and appendix 1) described by Möller et al. (2010); and sequences putatively annotated (FJ237364.1, named X2) or without annotation in GenBank (DQ359922.1, EF493183.1/EF493184.1 and DQ359921.1, named respectively X3, X4, and X5). In the Clustal alignment, two other groups of sequences, FJ375238.1/FJ375239.1/FJ375240.1 and EF208033.1 clustered in the superfamily A and O1, respectively with long branches, but corresponded to the independent lineages in the Muscle alignment (X6 and X7, respectively).

Fig. 1 Bayesian phylogenetic tree (midpoint rooting) obtained from the Clustal alignment of the signal sequences of conopeptides from GenBank. Posterior probabilities (when >0.9) are provided for each node. *Gray boxes* are used to visualize the superfamilies. The B and C superfamilies respectively correspond to the contulakins and conantokins. The lineages X1–X7 potentially correspond to previously unrecognized superfamilies (see details in the text)



Function and cysteine pattern were not clade-specific; conopeptides with the same function or cysteine pattern were found in different clades. In addition, sixteen new (i.e., not numbered with Roman numbers) cysteine patterns were identified; however, most of them certainly correspond to anecdotic mutations of the canonical framework in a given family (i.e., C–CC–C–C, C–C–CC–C–CC, and C–CC–C–C, found in the O1 superfamily, differ from the pattern VI/VII by only one mutation), while others may represent a new Cys-pattern number (e.g., the Cys-pattern C–C–C–CC–C, found in the three members of the X6 group). The results are summarized in Table 1 (full details are provided in Appendix 1).

Table 2 lists the number of conopeptides found in each superfamily and their distribution among the 71 *Conus* species. The superfamilies A, M, and O1 were the largest, each containing at least 39 species, followed by the superfamilies T and I2. *Conus characteristicus*, *C. imperialis*, and *C. litteratus* each express conopeptides belonging to more than 10 different superfamilies in their venom;

however, it was difficult to know if this result reflects a higher conopeptide diversity in comparison to other species, or is due to a greater sampling effort in these species. All the superfamilies present in more than 10 *Conus* species (A, B, I2, M, O1, O2, and T) were found in mollusk, worm, and fish-hunting species.

Discussion

An Updated Classification of Conopeptides

Overall, the molecular phylogeny, based on more than 1,300 conopeptides signal sequences extracted from GenBank, strongly supports the current superfamily classification based on phenetic resemblances, as established in ConoServer. But, this relative congruency between phylogenetic and phenetic classifications is not surprising given the relative conservation of the signal sequence within superfamilies compared with between superfamilies, and

Table 1 Number of sequences found in each superfamily, with list of cysteine patterns identified and known function in each superfamily

Superfamily		Cysteine			Known function	
ID	# of sequences	ID	Pattern	# of sequences		
A	153	I	CC-C-C	119	α, κ, ρ	
		II	CCC-C-C-C	3		
		IV	CC-C-C-C-C	25		
		VI/VII	C-C-CC-C-C	1		
		XIV	C-C-C-C	3		
			C	1		
		CC-C-C-C	1			
B	41		0	38	Conantokin	
			C-C	3		
C	4		0	1	Contulakin	
			C-C	3		
D	13	XX	C-CC-C-CC-C-C-C-C	5	α	
			C-CC-C-CC-C-C-C	1		
			C-C-C-CC-C-C-C-C-C	7		
I1	6	XI	C-C-CC-CC-C-C	6	ι	
I2	45	XI	C-C-CC-CC-C-C	35	κ	
			XII	C-C-C-C-CC-C-C		9
				C-C-CC-CC-C		1
I3	7	XI	C-C-CC-CC-C-C	7		
J	12	XIV	C-C-C-C	12	$\alpha + \kappa$	
L	4	XIV	C-C-C-C	3	α	
			C-C-C	1		
M	193		0	3	$\alpha, \kappa, \mu, \text{conomarphin}$	
		II	CCC-C-C-C	1		
		III	CC-C-C-CC	172		
		IV	CC-C-C-C-C	4		
		IX	C-C-C-C-C-C	1		
		XVI	C-C-CC	1		
		XIX	C-C-C-CCC-C-C-C-C	1		
			C	1		
			C-C	2		
			CC-C-C-C	1		
			CC-C-C-CC-C	2		
			C-CC-C-C-C	4		
		O1	625			0
VI/VII	C-C-CC-C-C			613		
	C-C-C			1		
	C-C-CC-C			1		
	C-CC-C-C			4		
	C-C-C-C-C			1		
	C-C-CC-C-CC			1		

Table 1 continued

Superfamily		Cysteine			Known function
ID	# of sequences	ID	Pattern	# of sequences	
O2	67	VI/VII	C–C–CC–C–C	51	γ , contryphan
			C–C	7	
		XV	C–C–CC–C–C–C	9	
O3	25	VI/VII	C–C–CC–C–C	25	bromosleeper
P	7	XIV	C–C–C–C	2	
		IX	C–C–C–C–C	5	
S	7	VIII	C–C–C–C–C–C–C–C–C	7	σ , α
T	140		0	12	ε , ζ , τ
		X	CC–CXPC	4	
		V	CC–CC	121	
			C–C	2	
			CC–CCC	1	
V	2	XV	C–C–CC–C–C–C	2	
X1	2		C–C–C–C–C–C	2	conoCAP
X2		III	CC–C–C–CC	1	
X3	1		0	1	
X4	2		C–C–C–C–C–C–C–C–C–CC–C–C–C–C–C–C–C–C–C	2	
X5	1	VIII	C–C–C–C–C–C–C–C–C	1	
X6	3		C–C–C–CC–C	3	
X7	1	VIII	C–C–C–C–C–C–C–C–C	1	
Y	1	XVII	C–C–CC–C–CC–C	1	

the phylogenetic tree reflects these differences. However, the phylogenetic approach also revealed several new features, the most striking of which is the presence of deeply divergent lineages that, until now, were not included in the conotoxin superfamily classification. There are two main explanations for this result. First, the conopeptide superfamily classification reviewed by Kaas et al. (2010) includes only what is traditionally referred to as “cysteine-rich” conotoxins [i.e., conopeptides with at least two disulfide bridges in the mature sequence as defined by Norton and Olivera (2006)], thus excluding the conopeptides with two cysteines and linear conopeptides also broadly present in the venom (unpublished results). However, although the authors noted that “in future, all disulfide-poor conopeptides will probably have to be attributed to a superfamily,” they refrained from doing so because of the low number of cysteine-poor conopeptides with precursor sequences in ConoServer (21). In GenBank, we identified more than 50 such sequences and included them

in the current analysis. The signal sequences of cysteine-poor conopeptides do not cluster separately from the conotoxins; some of them share highly similar signals with known superfamilies (contryphan with O2 and conomorphin with M), therefore, their exclusion from the superfamily classification is not phylogenetically justified. We identified two additional superfamilies, B and C, for conantokins and contulakins, respectively, one of which (C) has been proposed previously (Jimenez et al. 2007). Second, including non-annotated sequences from GenBank in the dataset helped to identify several independent lineages in the tree (X1–X7). The level of divergence of their respective signal sequences with the signals of other superfamilies was equivalent to the level of divergence between known superfamilies, and they thus deserve recognition as new superfamilies. However, as these independent lineages are represented by only one, two or three sequences, and because some of them may not exhibit toxin activity (even if they were all found in venom ducts

Table 2 Number of conopeptides in each superfamily and species

Species	Prey	A	B	C	D	II	I2	I3	J	L	M	O1	O2	O3	P	S	T	V	X1	X2	X3	X4	X5	X6	X7	Y	Occurrence	
<i>achatinus</i>	F	4														1											2	
<i>aurisiacus</i>	F	2								5	1																	3
<i>ballatus</i>	F	4								10	8																	3
<i>cattus</i>	F	4									29																	2
<i>circumcisus</i>	F	4								1	4																	3
<i>consors</i>	F	8	1							3	10																	4
<i>ermineus</i>	F	4								1	4																	3
<i>geographus</i>	F	3	5	1						2	5					1	2											7
<i>lynceus</i>	F									1																		1
<i>magus</i>	F	5								5	13	1																4
<i>monachus</i>	F	4									3																	2
<i>obscurus</i>	F	3	5																									2
<i>ochroleucus</i>	F		2																									1
<i>parius</i>	F					1																						1
<i>purpurascens</i>	F	4								1	9						1											4
<i>radiatus</i>	F	6	3			2				4	3		1			1												7
<i>stercusmuscarum</i>	F	6								3	6																	3
<i>striatus</i>	F	10					1			3	50						1											5
<i>striolatus</i>	F	1									8																	2
<i>sulcatus</i>	F	8	2																									2
<i>tulipa</i>	F	3								2	2																	2
<i>ammitralis</i>	M	2								5	3																	3
<i>aulicus</i>	M	2								3	3																	3
<i>aureus</i>	M									2																		1
<i>bandanus</i>	M	2																										1
<i>dalii</i>	M									3	8																	1
<i>episcopatius</i>	M	1				1				3	2																	2
<i>gloriamaris</i>	M		2							3	6		1		1		4											4
<i>marmoreus</i>	M	4					2			14	15	5					12					2						6
<i>omaria</i>	M									4	6																	7
<i>pennaceus</i>	M	4								6	13	3	2				16											2
<i>textile</i>	M	5					2			15	29	24			1	1	18											6
<i>victoriae</i>	M	3																										8
<i>abbreviatus</i>	W										98																	1
<i>arenatus</i>	W		1							2	31	8				8												1
<i>aristophanes</i>	W										8																	5
<i>betulinus</i>	W	8				1	1	1		10	1	2				1												1
																												7

Table 2 continued

Species	Prey	A	B	C	D	II	I2	I3	J	L	M	O1	O2	O3	P	S	T	V	X1	X2	X3	X4	X5	X6	X7	Y	Occurrence
<i>capitaneus</i>	W						1				1	2	2													4	
<i>characteristicus</i>	W	2	5					3			6	3	1	1		3	5							1	1	11	
<i>coronatus</i>	W										1	7														2	
<i>distans</i>	W	1									2	5														3	
<i>ebraeus</i>	W											29														1	
<i>eburneus</i>	W	2	1	2			4	1	1	1	6	2		1												9	
<i>emaciatius</i>	W						3				1	1	1													3	
<i>ferrugineus</i>	W								2		1	1														2	
<i>figulinus</i>	W										5															1	
<i>flavidus</i>	W										1															1	
<i>generalis</i>	W										2	2														2	
<i>imperialis</i>	W	4	3	1	2	6					2	7	1	1	1	2	2			1	1	3				13	
<i>judaeus</i>	W										2															1	
<i>leopardus</i>	W	8									5	9				4										4	
<i>litteratus</i>	W	4	2	2	7	6	6	1	3	3	11	8	5	3	3	16	16			1	1					15	
<i>lividus</i>	W	3									1	84	2	1		3	3									6	
<i>miles</i>	W	1			1	1					6	2	2			1										6	
<i>miliaris</i>	W										2	18														2	
<i>musculus</i>	W										1															1	
<i>mustelinus</i>	W				2																					1	
<i>planorbis</i>	W								4		1															1	
<i>pulicarius</i>	W	4						3			3	6			6	6										2	
<i>quercinus</i>	W	6	3								4	3			2	2										5	
<i>rattus</i>	W										2	4														5	
<i>regius</i>	W																									2	
<i>sponsalis</i>	W	2										14			1											1	
<i>spurius</i>	W						10				1						8									2	
<i>tessulatus</i>	W										14	6	2	4			11									3	
<i>ventricosus</i>	W										6	12	9	5			15									5	
<i>vexillum</i>	W				2		2				1	8	3													5	
<i>vilpepinii</i>	W																		2							1	
<i>viola</i>	W										5															1	
<i>virgo</i>	W	2	2			4					5	3	2			4	1									1	
<i>vitulinus</i>	W	3	1			2		2		2																8	
Occurrence		39	16	4	5	4	14	3	5	2	50	51	17	9	5	5	21	2	1	1	1	2	1	1	1	6	

Feeding types: F fish-hunting species, M mollusc-hunting species, W worm-hunting species

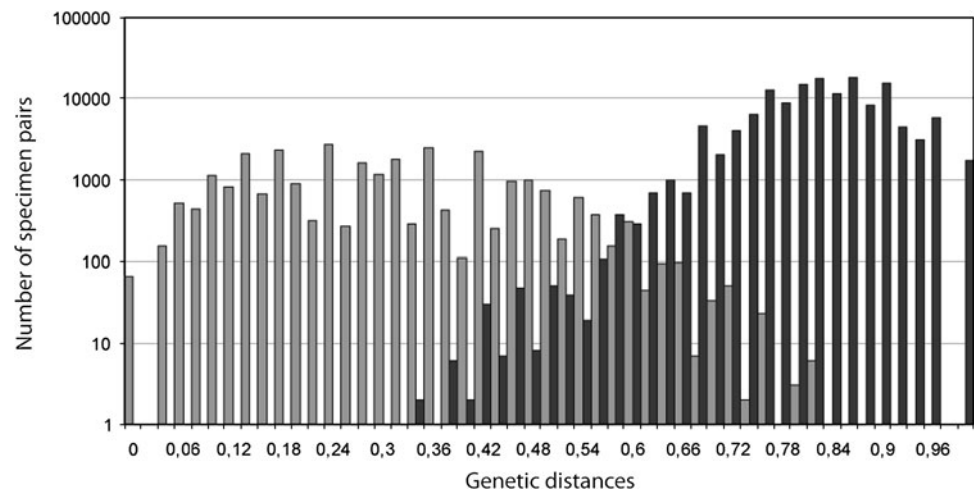
of cone snails), we refrained from proposing new superfamily names, and only provided temporary names (X1–X7). It should also be borne in mind that many other conopeptides have been described in the literature, some of which have been given formal names (conkunitzin, conolysin, conomap, conophysin, conopressin, conorfamide, and conorphan). Because their signal sequences are not represented as nucleotides in GenBank, they were not included in the analysis. However, a search in the protein database of GenBank retrieved two complete precursors of Conkunitzin, with highly similar signal sequences (POC1X2.1 and POCY85.1) and a local BLAST search (performed using BioEdit—Hall 1999) of the dataset used for the phylogenetic analyses revealed that the conkunitzin signals were unique, and probably represent a new superfamily. Finally, if most of the superfamily-level clades are highly supported, most of the inter-superfamily nodes are not, preventing any reliable conclusion concerning the phylogenetic relationships at this level.

The original results presented herein raise several issues concerning the classification and nomenclature of the conopeptides and, more generally, of the genes that belong to multigenic systems. The updated classification system we propose is based on a phylogenetic reconstruction that guarantees the identification of sequence clusters that share a common ancestor. However, such phylogenetic trees cannot help in deciding which clades deserve a superfamily-level ranking and which ones do not. One common solution is to rely on a threshold of genetic distances, but the analyses of the genetic distances (calculated as the number of differences) between all the conopeptide signal sequences revealed that the distribution of genetic distances within superfamilies of conopeptides largely overlaps with the distribution of genetic distances between superfamilies (Fig. 2). This overlap can be linked to the high level of homoplasy found in conopeptides, making two conopeptides from different clades having, by chance,

a relatively low genetic distance, or to the fact that two previously defined superfamilies would actually correspond to only one. This is the case of the L and I3 superfamilies, separated by genetic distances comprised between 0.38 and 0.69 that would, in most cases, correspond to within superfamily genetic distances.

Consequently, it is not possible to rely only on a genetic threshold to define superfamilies for conotoxins. A threshold of 0.6, roughly corresponding to the gap between the two distributions of genetic distances (Fig. 2), would lead to the division of the M-superfamily into numerous superfamilies (indeed, Wang et al. 2008 proposed to divide the M-superfamily in M1 and M2), and to the grouping of the superfamilies I1, I3, and L in a single one. However, our approach is aimed at offering a complementary guidance to help, in the future, deciding if a conotoxin or a group of conotoxins deserve a superfamily name: (i) since the minimum genetic distance between superfamilies is 0.32, this distance should be the minimum distance between the potential new superfamily(ies) and all the others; (ii) the new superfamily(ies) should correspond to an independent lineage, i.e., it should not cluster in any of the superfamily clades previously defined; (iii) the molecular target of the new conotoxin(s) should ideally be identified, to avoid naming conopeptides that would not be functional; (iv) the structure (cysteine pattern) and/or function should be different from the most closely related superfamilies in terms of genetic distances and/or phylogenetic relationships. All these criteria apply to the B and C superfamilies (genetic distances with other superfamilies >0.3, these two lineages are independent and monophyletic, their molecular targets are identified—Mena et al. 1990, Craig et al. 1999—, and their cysteine framework are different from their respective sister-groups), justifying the attribution of new superfamily names. We followed the traditional nomenclature of conopeptide superfamilies, i.e., a Roman capital letter. As the number of Roman letter is

Fig. 2 Pairwise distribution of genetic distances (p distances) calculated with MEGA5 using the Clustal alignment. Genetic distances between sequences from the same superfamily are shown in gray, genetic distances between sequences from different superfamily in black



limited, some superfamilies have been named with a Roman letter followed by an Arabic number (e.g., I1, I2, I3, O1, O2, and O3) when several superfamilies share a common cysteine framework or molecular target. Because of the potentially high number of unknown superfamilies of conopeptides, we have no doubt that the nomenclature based on both Roman letters and Arabic numbers will become the reference rule.

The first and fourth criteria also apply to the seven “X” lineages (Fig. 1), but the second applies to only 5 of them (two clustered within the A and O1 superfamilies with the muscle alignment) and the third to none of them. We propose to name such potential superfamilies of conopeptides that currently do not meet all the criteria but could in the future with the X Roman letter, followed by an Arabic number, waiting for either to be fully recognized as a separate superfamily or as belonging to an existing one.

Evolution of the Conopeptides

The phylogenetic analysis clearly confirms that most of the defined superfamilies include conopeptides with different cysteine frameworks and functions. Conversely, similar cysteine frameworks and functions are found in different superfamilies, suggesting that a given cysteine framework or function can appear several times independently, probably as a result of convergent evolution. The multiple apparitions of the same framework and function during conotoxin evolution are probably linked to the extremely rapid diversification of the genes. Several molecular mechanisms have been proposed as being responsible for this high rate of diversification. Pi et al. (2006) suggested that alternative splicing, unequal crossing-over or exon shuffling could explain this diversity. Olivera et al. (1999) proposed two other mechanisms: the lack of a mismatch repair system, at least in the hypervariable part of the sequence (the mature toxin); and recombination mechanisms. Several other hypotheses, such as a high rate of duplication, followed by a strong diversifying selection on the newly created gene copies that could lead to the rapid appearance of several structurally and functionally highly divergent genes, have been also proposed and tested by different authors (Duda and Palumbi 1999, 2000; Conticello et al. 2000, 2001; Espiritu et al. 2001; Duda and Remigio 2008; Chang and Duda 2012). All these molecular mechanisms, together with observed differences in the expression pattern between species, maybe linked to episodes of gene silencing and reactivation (“Lazarotoxins”, Conticello et al. 2001; Duda and Palumbi 2004; Duda 2008), could favor the rapid diversification of *Conus* species, by allowing them to envenomate and feed on new prey and thus colonize new niches (Duda and Lee 2009).

A phylogenetic approach could be very useful to identify divergent conopeptides with potentially different functions, even if they share a common structural framework. For example, the cysteine framework IV, found in the A-superfamily, is already linked to two different functions (α A—Hopkins et al. 1995 and κ A—Craig et al. 1998). However, conotoxins, described by Conticello et al. (2001), with the same framework, belong to the M-superfamily, suggesting that these IV-conotoxins that are structurally convergent with the IV-conotoxins in a different superfamily, could exhibit a completely different function. A similar strategy could also apply within each superfamily, where not only the signal sequence, but also the propeptide and mature regions can be aligned, and could reveal divergent lineages with as yet uncharacterized functions (e.g., see Aguilar et al. 2009; Puillandre et al. 2010; Wang et al. 2008; Zhangsun et al. 2006).

Furthermore, our identification of numerous new cysteine frameworks among the GenBank sequences was also surprising. Even if some of them may be non-functional genes (pseudogenes), others could correspond to novel protein structures. A few publications demonstrated that even toxins with odd numbers of cysteines can be functional, for example with two 5-Cys toxins forming a functional dimer or bioactive polymers of the 13-Cys “Con-ikot-ikot” peptide from *Conus striatus* (Quinton et al. 2009, Walker et al. 2009). Our findings challenge the traditional view where conotoxins are characterized by a limited number of cysteine frameworks: by exploring new evolutionary pathways, the apparition of novel cysteine frameworks may also participate in the hyper-diversification of the conotoxins. In addition, this raises the question of the total number of cysteine patterns one could expect to find among cone snail toxins. It is possible to predict the theoretic number of cysteine patterns that could exist. If we limit the exercise to the 2, 4, and 6 cysteine patterns and exclude those with more than two consecutive cysteines, 20 different frameworks can be proposed (C–C*, CC, CC–C*, CC–CC*, C–CC–C, C–C–CC*, C–C–C–C*, CC–CC–CC, CC–CC–C–C, CC–C–CC–C, CC–C–C–CC*, CC–C–C–C–C*, C–CC–CC–C, C–CC–C–CC, C–CC–C–C–C*, C–C–CC–CC, C–C–CC–C–C*, C–C–C–CC–C*, C–C–C–C–CC, C–C–C–C–C–C*). Ten of these frameworks (marked with an *) can be found in GenBank. Given the extreme capacity of the conopeptides to evolve and the apparent lack of evolutionary constraints (as illustrated by the multiple apparitions of identical frameworks during their evolution), there is no reason that all these theoretical patterns will not be found in the future. It could be argued that mechanical constraints would prevent the existence of some cysteine patterns; for example, it could be unfavorable to have a disulfide bridge between two adjacent cysteines. However, despite this we found a short mature toxin

in the venom of one cone snail with a disulfide bridge between adjacent cysteines (unpublished results). The peptide has been reproduced by protein synthesis, confirming this finding.

Conus and Conoidea Toxin Diversity

The diversity of conotoxins in the venom of several *Conus* species (Table 2) confirms that most species are able to express a variety of conotoxins, as widely reported in literature (e.g., Olivera 2002). Furthermore, our results also suggest that *Conus* diet (fish, mollusk, and worm) is not correlated with differences in venom composition at the superfamily level. If differences exist, as suggested in the literature (e.g., Conticello et al. 2001; Kaas et al. 2010), they most likely occur at the species and intra-superfamily levels. Furthermore, phylogenetic analyses suggest that, at least, the worm- and fish-hunting species are not monophyletic, as these two diets appeared independently several times during the *Conus* evolution (Duda and Palumbi 2004; Espiritu et al. 2001; Kraus et al. 2011). Thus, differences in the venom composition should not be sought between the three diet groups, but between the monophyletic clades defined within these three groups (Duda and Palumbi 2004).

Diversity of the marine snail toxins is not limited to species included in the large major clade of *Conus*. Recent analyses in other conoidean taxa suggest that toxin hyperdiversity is not the privilege of the *Conus* large major clade. *C. californicus*, which is highly divergent from all the other *Conus* species (Duda and Kohn 2005), showed a high diversity of toxins in its venom and several of them were thought to correspond to new superfamilies (Biggs et al. 2010; www.conoserver.org/?page=classification&type=genesuperfamilies). To a lesser extent, species in the small major clade of *Conus*, may also contain several novel conotoxins, as suggested by an original Cys-pattern (XIII) found in the species *C. delessertii* (Aguilar et al. 2005). In addition to the family Conidae, original toxins have already been reported in several other species of Conoidea, such as *Polystira albida* (Lopez-Vera et al. 2004; Rojas et al. 2008), *Gemmula periscelida* (Lopez-Vera et al. 2004), *G. speciosa*, *G. sogodensis*, *G. diomedea*, *G. kieneri* (Heralde et al. 2008), *Lophiotoma olangoensis* (Watkins et al. 2006), *Terebra subulata* (Imperial et al. 2003), *Hastula hectica* (Imperial et al. 2007) and *Crassispira cerithina* (Cabang et al. 2011). Furthermore, taxonomic surveys (Bouchet et al. 2009) and phylogenetic analyses (Puillandre et al. 2011) suggest that the superfamily Conoidea actually comprises a number of deeply divergent clades, whose species diversity is currently largely underestimated. Presently, around 4,500 species have been described, but the group is believed to include more than

10,000 species (Bouchet et al. 2009). Even if the venom apparatus has been lost in several lineages of Conoidea (e.g., Fedosov 2007; Fedosov and Kantor 2008; Holford et al. 2009; Medinskaya and Sysoev 2003), these findings suggest that the conotoxin diversity characterized so far represents only a small part. If the level of diversity across all conoidean species is similar to that found in those already investigated, the number of toxins produced by this single superfamily could be as high as ten millions.

Acknowledgments We are grateful to the European Commission for financial support. This study has been performed as a part of the CONCO cone snail genome project for health (www.conco.eu) within the 6th Framework Program (LIFESCIHEALTH-6 Integrated Project LSHB-CT-2007, contract number 037592). We are also grateful to Frédérique Lisacek from the Swiss Institute of Bioinformatics for ongoing help. We would like to thank Dr Ron Hogg of OmniScience SA for editorial support.

Conflict of interest The authors declare that they have no conflict of interest.

References

- Aguilar MB, Lopez-Vera E, Ortiz E, Becerril B, Possani LD, Olivera BM, de la Heimer Cotera EP (2005) A novel conotoxin from *Conus delessertii* with posttranslationally modified lysine residues. *Biochemistry* 44:11130–11136
- Aguilar MB, Chan de la Rosa RA, Falcon A, Olivera BM, de la Heimer Cotera EP (2009) Peptide pal9a from the venom of the turrid snail *Polystira albida* from the Gulf of Mexico: purification, characterization, and comparison with P-conotoxin-like (framework IX) conoidean peptides. *Peptides* 30:467–476
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: signalP 3.0. *J Mol Biol* 340(4): 783–795
- Biass D, Dutertre S, Gerbault A, Menou J-L, Offord R, Favreau P, Stöcklin R (2009) Comparative proteomic study of the venom of the piscivorous cone snail *Conus consors*. *J Proteomics* 72:210–218
- Biggs JS, Watkins M, Puillandre N, Ownby JP, Lopez-Vera E, Christensen S, Moreno KJ, Bernaldez J, Licea-Navarro A, Showers Corneli P, Olivera BM (2010) Evolution of *Conus* peptide toxins: analysis of *Conus californicus* Reeve, 1844. *Mol Phylogenet Evol* 56:1–12
- Blunt JW, Copp BR, Keyzers RA, Munro MH, Prinsep MR (2012) Marine natural products. *Nat Prod Rep* 29:144–222
- Bouchet P, Lozouet P, Sysoev AV (2009) An inordinate fondness for turrids. *Deep Sea Res II* 56:1724–1731
- Cabang AP, Imperial JS, Gajewiak J, Watkins M, Showers Corneli P, Olivera BM, Concepcion GP (2011) Characterization of a venom peptide from a crassispirid gastropod. *Toxicon* 58:672–680
- Chang C, Duda TF (2012) Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. *Mol Biol Evol*. Advance access
- Conticello SG, Pilpel Y, Glusman G, Fainzilber M (2000) Position-specific codon conservation in hypervariable gene families. *Trends Genet* 16:57–59
- Conticello SG, Gilad Y, Avidan N, Ben-Asher E, Levy Z, Fainzilber M (2001) Mechanisms for evolving hypervariability: the case of conopeptides. *Mol Biol Evol* 18:120–131

- Craig AG, Zafaralla G, Cruz LJ, Santos AD, Hillyard DR, Dykert J, Rivier J, Gray WR, Imperial J, DelaCruz RG, Sporning A, Terlau H, West PJ, Yoshikami D, Olivera BM (1998) An O-glycosylated neuroexcitatory Conus peptide. *Biochemistry* 37:16019–16025
- Craig AG, Norberg T, Griffin D, Hoeger C, Akhtar M, Schmidt K, Low W, Dykert J, Richelsoni E, Navarro V, Mazella J, Watkins M, Hillyard DR, Imperial J, Cruz LJ, Olivera BM (1999) Contulakin-G, an O-glycosylated invertebrate neurotensin. *J Biol Chem* 274:13752–13759
- Daly NL, Craik DJ (2009) Structural studies of conotoxins. *IUBMB Life* 61:144–150
- Davis J, Jones A, Lewis RJ (2009) Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. *Peptides* 30:1222–1227
- Duda TF (2008) Differentiation of venoms of predatory marine gastropods: divergence of orthologous toxin genes of closely related *Conus* species with different dietary specializations. *J Mol Evol* 67:315–321
- Duda TF, Kohn AJ (2005) Species-level phylogeography and evolutionary history of the hyperdiverse marine gastropod genus *Conus*. *Mol Phylogenet Evol* 34:257–272
- Duda JTF, Lee T (2009) Ecological release and venom evolution of a predatory marine Snail at Easter Island. *PLoS One* 4:e5558
- Duda TF, Palumbi SR (1999) Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc Natl Acad Sci* 96:6820–6823
- Duda TF, Palumbi SR (2000) Evolutionary diversification of multi-gene families: allelic selection of toxins in predatory cone snails. *Mol Biol Evol* 17:1286–1293
- Duda TF, Palumbi SR (2004) Gene expression and feeding ecology: evolution of piscivory in the venomous gastropod genus *Conus*. *Proc Royal Soc B* 271:1165–1174
- Duda TF, Remigio A (2008) Variation and evolution of toxin gene expression patterns of six closely related venomous marine snails. *Mol Ecol* 17:3018–3032
- Dutertre S, Biass D, Stöcklin R, Favreau P (2010) Dramatic intraspecific variations within the injected venom of *Conus consors*: an unsuspected contribution to venom diversity. *Toxicon* 55:1453–1462
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
- Espiritu DJD, Watkins M, Dia-Monje V, Cartier GE, Cruz LE, Olivera BM (2001) Venomous cone snails: molecular phylogeny and the generation of toxin diversity. *Toxicon* 39:1899–1916
- Favreau P, Stöcklin R (2009) Marine snail venoms: use and trends in receptor and channel neuropharmacology. *Curr Opin Pharmacol* 9:594–601
- Favreau P, Benoit E, Hocking E, Carlier L, D'hoedt D, Leipold E, Markgraf D, Schlumberger S, Cordova M, Gaertner H, Paolini-Bertrand M, Hartley O, Tytgat J, Heinemann S, Bertrand D, Boelens R, Stöcklin R, Molgo J (2012) A novel mu-conopeptide, CnIIIc, exerts potent and preferential inhibition of NaV1.2/1.4 channels and blocks neuronal nicotinic acetylcholine receptors. *Br J Pharmacol* (in press)
- Fedosov AE (2007) Anatomy of accessory rhynchodeal organs of *Vepracula vepratrica* and *Tritonoturris subbrissoides*: new types of foregut morphology in Raphitominae (Conoidea). *Ruthenica* 17:33–41
- Fedosov A, Kantor Y (2008) Toxoglossan gastropods of the subfamily Crassispirinae (Turridae) lacking a radula, and a discussion of the status of the subfamily Zemaciinae. *J Mollusc Stud* 74:27–35
- Gayler K, Sandall D, Greening D, Keays D, Polidano M, Livett B, Down J, Satkunathan N, Khalil Z (2005) Molecular prospecting for drugs from the sea. *IEEE Eng Med Biol Mag* 24:79–84
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98
- Han TS, Teichert RW, Olivera BM, Bulaj G (2008a) Conus venoms—a rich source of peptide-based therapeutics. *Curr Pharm Des* 14:2462–2479
- Han Y, Huang F, Jiang H, Liu L, Wang Q, Wang Y, Shao X, Chi C, Du W, Wang C (2008b) Purification and structural characterization of a d-amino acid-containing conopeptide, conomarphin, from *Conus marmoreus*. *FEBS J* 275:1976–1987
- Heralde FM, Imperial J, Bandyopadhyay P, Olivera BM, Concepcion GP, Santos AD (2008) A rapidly diverging superfamily of peptide toxins in venomous *Gemmula* species. *Toxicon* 51:890–897
- Holford M, Puillandre N, Terryn Y, Cruaud C, Olivera BM, Bouchet P (2009) Evolution of the Toxoglossa venom apparatus as inferred by molecular phylogeny of the Terebridae. *Mol Biol Evol* 26:15–25
- Hopkins C, Grilley M, Miller C, Shon K-J, Cruz LJ, Gray WR, Dykert J, Rivier J, Yoshikami D, Olivera BM (1995) A new family of *Conus* peptides targeted to the nicotinic acetylcholine receptor. *J Biol Chem* 270:22361–22367
- Hu H, Bandyopadhyay PK, Olivera BM, Yandell M (2011) Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. *BMC Genomics* 12:60
- Huelsenbeck JP, Ronquist F, Hall B (2001) MrBayes: bayesian inference of phylogeny. *Bioinformatics* 17:754–755
- Imperial JS, Watkins M, Chen P, Hillyard DR, Cruz LJ, Olivera BM (2003) The augertoxins: biochemical characterization of venom components from the toxoglossate gastropod *Terebra subulata*. *Toxicon* 42:391–398
- Imperial JS, Kantor Y, Watkins M, Heralde FM, Stevenson B, Chen P, Hansson K, Stenflo J, Ownby J-P, Bouchet P, Olivera BM (2007) Venomous auger snail *Hastula (Impages) hectica* (Linnaeus 1758): molecular phylogeny, foregut anatomy and comparative toxinology. *J Exp Zool* 308B:744–756
- Jakubowski JA, Kelley WP, Sweedler JV, Gilly WF, Schulz JR (2005) Intraspecific variation of venom injected by fish-hunting *Conus* snails. *J Exp Biol* 208:2873–2883
- Jimenez EC, Olivera BM, Teichert RW (2007) α C-conotoxin PrXA: a new family of nicotinic acetylcholine receptor antagonists. *Biochemistry* 46:8717–8724
- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *CABIOS* 8:275–282
- Kaas Q, Westermann JC, Craik DJ (2010) Conopeptide characterization and classifications: an analysis using ConoServer. *Toxicon* 55:1491–1509
- Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McInerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* 6:1–17
- Koua D, Brauer A, Laht S, Kaplinski L, Favreau P, Remm M, Lisacek F, Stöcklin R (2012) ConoDictor: a tool for prediction of conopeptide superfamilies. *Nucleic Acids Res* (in press)
- Kraus NJ, Showers Corneli P, Watkins M, Bandyopadhyay PK, Seger J, Olivera BM (2011) Against expectation: a short sequence with high signal elucidates cone snail phylogeny. *Mol Phylogenet Evol* 58:383–389
- Laht S, Koua D, Kaplinski L, Lisacek F, Stöcklin R, Remm M (2011) Identification and classification of conopeptides using profile Hidden Markov models. *Biochim Biophys Acta* 1824:488–492

- Leary D, Vierros M, Hamon G, Arico S, Monagle C (2009) Marine genetic resources: a review of scientific and commercial interest. *Mar Policy* 33:183–194
- Lewis RJ (2012) Discovery and development of the χ -conopeptide class of analgesic peptides. *Toxicon* 59(4):524–528
- Lin H, Li Q-Z (2007) Predicting conotoxin superfamily and family by using pseudo amino acid composition and modified Mahalanobis discriminant. *Biochem Biophys Res Commun* 354:548–551
- Lopez-Vera E, de la Heimer Cotera EP, Maillo M, Riesgo-Escovar JR, Olivera BM, Aguilar MB (2004) A novel structure class of toxins: the methionine-rich peptides from the venoms of turrid marine snails (Mollusca, Conoidea). *Toxicon* 43:365–374
- McGivern JG (2007) Ziconotide: a review of its pharmacology and use in the treatment of pain. *Neuropsychiatr Dis Treat* 3:69–85
- Medinskaya AI, Sysoev A (2003) The anatomy of *Zemacies excelsa*, with a description of a new subfamily of Turridae (Gastropoda, Conoidea). *Ruthenica* 13:81–87
- Mena EE, Gullak MF, Pagnozzi MJ, Richter KE, Rivier J, Cruz LJ, Olivera BM (1990) Conantokin-G: a novel peptide antagonist to the N-methyl-D-aspartic acid (NMDA) receptor. *Neurosci Lett* 118:241–244
- Menez A, Stocklin R, Mebs D (2006) Venomics' or: the venomous systems genome project. *Toxicon* 47:255–259
- Miljanich GP (2004) Ziconotide: neuronal calcium channel blocker for treating severe chronic pain. *Curr Med Chem* 11:3029–3040
- Molinski TF, Dalisay DS, Lievens SL, Saludes JP (2009) Drug development from marine natural products. *Nat Rev Drug Discov* 8:69–85
- Möller C, Melaun C, Castillo C, Díaz ME, Renzelman CM, Estrada O, Kuch U, Lokey S, Mari F (2010) Functional hypervariability and gene diversity of cardioactive neuropeptides. *J Biol Chem* 285:40673–40680
- Mondal S, Bhavna R, Babu RM, Ramakumar S (2006) Pseudo amino acid composition and multi-class support vector machines approach for conotoxin superfamily classification. *J Theor Biol* 243:252–260
- Norton RS, Olivera BM (2006) Conotoxins down under. *Toxicon* 48:780–798
- Olivera BM (2002) *Conus* venom peptides: reflections from the biology of clades and species. *Annu Rev Ecol Syst* 33:25–47
- Olivera BM (2006) *Conus* peptides: biodiversity-based discovery and exogenomics. *J Biol Chem* 281:31173–31177
- Olivera BM, Walker C, Cartier GE, Hooper D, Santos AD, Schoenfeld R, Shetty R, Watkins M, Bandyopadhyay PK, Hillyard DR (1999) Speciation of cone snails and interspecific hyperdivergence of their venom peptides. Potential evolutionary significance of introns. *Ann NY Acad Sci* 870:223–237
- Pi C, Liu J, Peng C, Liu Y, Jiang X, Zhao Y, Tang S, Wang L, Dong M, Chen S, Xu A (2006) Diversity and evolution of conotoxins based on gene expression profiling of *Conus litteratus*. *Genomics* 88:809–819
- Puillandre N, Holford M (2010) The Terebridae and teretoxins: combining phylogeny and anatomy for concerted discovery of bioactive compounds. *BMC Chem Biol* 10:7
- Puillandre N, Watkins M, Olivera BM (2010) Evolution of *Conus* peptide genes: duplication and positive selection in the A-superfamily. *J Mol Evol* 70:190–202
- Puillandre N, Kantor Y, Sysoev A, Couloux A, Meyer C, Rawlings T, Todd JA, Bouchet P (2011) The dragon tamed? A molecular phylogeny of the Conoidea (Mollusca, Gastropoda). *J Mollusc Stud* 77:259–272
- Quinton L, Gilles N, De Pauw E (2009) TxXIIIa, an atypical homodimeric conotoxin found in the *Conus textile* venom. *J Proteomics* 72:219–226
- Rambaut A, Drummond AJ (2007) Tracer v1.4. Available from <http://beast.bio.ed.ac.uk/Tracer>
- Rojas A, Feregrino A, Ibarra-Alvarado C, Aguilar MB, Falcon A, de la Heimer Cotera EP (2008) Pharmacological characterization of venoms obtained from Mexican toxoglossate gastropods on isolated guinea pig ileum. *J Venom Anim Toxins Incl Trop Dis* 14:497–513
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739
- Terrat Y, Biass D, Dutertre S, Favreau P, Remm M, Stöcklin R, Piquemal D, Ducancel F (2011) High-resolution picture of a venom gland transcriptome: case study with the marine snail *Conus consors*. *Toxicon* 59:34–46
- Ueberheide BM, Fenyo D, Alewood PF, Chait BT (2009) Rapid sensitive analysis of cysteine rich peptide venom components. *Proc Natl Acad Sci* 106:6910–6915
- Violette A, Leonardi A, Piquemal D, Terrat Y, Biass D, Dutertre S, Noguier F, Ducancel F, Stöcklin R, Krizaj I, Favreau P (2012) Recruitment of glycosyl hydrolase proteins in a cone snail venomous arsenal: further insights into biomolecular features of *Conus* venoms. *Mar Drugs* 10:258–280
- Walker CS, Jensen S, Ellison M, Matta JA, Lee WY, Imperial JS, Duclos N, Brockie PJ, Madsen DM, Isaac JT, Olivera BM, Maricq AV (2009) A novel *Conus* snail polypeptide causes excitotoxicity by blocking desensitization of AMPA receptors. *Curr Biol* 19:900–908
- Wang Q, Jiang H, Hana Y-H, Yuan DD, Chi C-W (2008) Two different groups of signal sequence in M-superfamily conotoxins. *Toxicon* 51:813–822
- Watkins M, Hillyard DR, Olivera BM (2006) Genes expressed in a Turrid venom duct: divergence and similarity to conotoxins. *J Mol Evol* 62:247–256
- Zhangsun D, Luo S, Wu Y, Xiaopeng Z, Hu Y, Xie L (2006) Novel O-superfamily conotoxins identified by cDNA cloning from three vermivorous *Conus* species. *Chem Biol Drug Des* 68:256–265