Variation and evolution of toxin gene expression patterns of six closely related venomous marine snails

T. F. DUDA JR*† and E. A. REMIGIO*

*Department of Ecology and Evolutionary Biology & Museum of Zoology, University of Michigan, 1109 Geddes Avenue, Ann Arbor MI 48109, USA, †Smithsonian Tropical Research Institute, Apartado 0843-03092, Balboa, Ancón, Republic of Panama

Abstract

Venoms of predatory marine gastropods of the genus *Conus* show amazing levels of inter-specific diversity and are comprised of a cocktail of peptide neurotoxins, termed conotoxins, that are encoded by large gene families. Conotoxin gene family evolution is characterized by gene duplications and high rates of nonsynonymous substitution among paralogues; yet, what controls the differentiation of venoms among species is not clear. We compared four-loop conotoxin transcripts of six closely related *Conus* species to examine conotoxin expression patterns among species. The species examined appear to express different numbers of four-loop conotoxin loci and similarity in expression patterns does not seem to correspond with phylogenetic affinity. Moreover, several loci appear to have been independently silenced while others appear to have been revived from previously silenced states. Some loci also appear to exhibit coordinated expression patterns. These results suggest that the evolution of conotoxin expression patterns is incredibly dynamic and the differentiation of venoms of *Conus* is controlled in part by the evolution of unique conotoxin expression patterns.

Keywords: *Conus*, coordinated expression, four-loop conotoxins, gene family evolution, O-superfamily, resurrected loci

Received 20 January 2008; revision received 14 March 2008; accepted 16 April 2008

Introduction

Members of the marine gastropod genus *Conus* show tremendous differences in feeding specializations and *Conus* species predominately prey on either fish, molluscs or polychaetes (Kohn 1959). Diets are largely nonoverlapping among vermivorous species such that different species, even close relatives, typically prefer different prey (Kohn 1959, 1966, 1980, 2001; Kohn & Orians 1962; Marsh 1971; Kohn & Nybakken 1975; Leviten 1980; Reichelt & Kohn 1985; Kohn & Almasi 1993; Duda *et al*. 2001). *Conus* use a venom comprised of a plethora of small peptide neurotoxins, termed conotoxins, to subdue prey and these venoms are remarkably diverse: the venom of a single snail may contain upwards of 50 distinct conotoxins and venom composition differs dramatically among species (Olivera *et al*. 1999).

Conotoxins are encoded by large gene families (e.g. A-, I-, M-, O-, P- and T-superfamilies) and so a variety of mechanisms may contribute to the differentiation of venoms among *Conus* species. Previous analyses demonstrated that conotoxin gene family evolution is characterized by gene duplications and high rates of nonsynonymous substitutions among loci (Duda & Palumbi 1999a, 2004; Conticello *et al*. 2001). For example, analyses of nine expressed O-superfamily or 'four-loop' conotoxin loci of *Conus abbreviatus* revealed that positive selection has operated among eight of these and rates of nonsynonymous substitution are 10–20 times faster than rates reported for genes of other metazoans (Duda & Palumbi 1999a). Positive selection also enhances allelic diversity at conotoxin loci; two alleles identified at a single conotoxin locus of *Conus ebraeus* differ at nine nonsynonymous sites and yet exhibit no synonymous substitutions (Duda & Palumbi 2000). Also, major shifts in diet were accompanied by the evolution of unique conotoxin expression patterns among independently derived piscivorous *Conus* (Duda & Palumbi 2004). Based on the relationships of four-loop conotoxin transcripts of two closely related vermivorous *Conus* species, *C. abbreviatus* and *C. ebraeus*, these species apparently do not express orthologous loci (Duda & Palumbi 2000).

Gene expression levels are generally more similar among closely related taxa than among more distantly related

Correspondence: T. F. Duda Jr, Fax: + 734-763-4080; E-mail: tfduda@umich.edu

ones (Whitehead & Crawford 2006a). Nonetheless, exceptions to this pattern are known and genes that exhibit expression patterns that do not follow this trend may be linked to organismal adaptations. For example, differences in expression levels of genes of salmonids correspond better with differences in life history than with phylogeny (Giger *et al.* 2006). Furthermore, convergent changes in gene expression occurred among different species of salmonids that exhibit similar ecological adaptations (Derome & Bernatchez 2006). Also, expression patterns of 13 genes of *Fundulus heteroclitus* cannot be accounted for by phylogeny but instead show association with environmental conditions (Whitehead & Crawford 2006b). Within humans, expression levels of some genes are significantly different from those of other primates despite similarities in expression patterns of these genes among nonhuman primates (Gilad *et al.* 2006). Because conotoxin peptides serve a primary role in capturing prey, differences in conotoxin expression patterns among species may be related to differences in the feeding ecologies of *Conus* species.

Do differences in venom composition reflect the evolution of unique conotoxin expression patterns among species? Do conotoxin expression patterns correspond with phylogenetic relationships such that closely related species exhibit more similar expression patterns than more distantly related ones? To address these questions, we examined the phylogeny and patterns of divergence of four-loop conotoxin messenger RNA (mRNA) transcripts of a set of six closely related *Conus* species to identify putative orthologous and nonorthologous loci of these species and to compare patterns of conotoxin expression among species.

Materials and methods

Specimens

We focused our analyses on four-loop conotoxin transcript sequences recovered from six members of a clade of closely related vermivorous *Conus* species (clade E1 *sensu* Duda, Kohn & Palumbi 2001; see also Duda & Kohn 2005): *Conus abbreviatus*, *C. aristophanes*, *C. coronatus*, *C. ebraeus*, *C. judaeus* and *C. miliaris*. Specimens of *C. aristophanes*, *C. coronatus* and *C. miliaris* were collected from reefs in American Samoa. Specimens of *C. ebraeus* were collected from intertidal sites in the Gulf of Chiriqui, Panama. Specimens of *C. judaeus* were collected from reef flats in Okinawa. Venom ducts were preserved in RNAlater (Ambion, Inc.) and transported on ice to the laboratory.

Conotoxin sequences

Crude mRNA was extracted and complementary DNA (cDNA) was constructed as reported previously (Duda & Palumbi 1999a). Amplifications were performed on cDNA of at least two individuals of the above species using the same primers (TOX1 = CATCGTCAAGATGAAACTGACGTG, TOX2 = CACAGGTATGGATGACTCAGG) that were used previously to amplify four-loop conotoxin transcripts of *C. abbreviatus* and *C. ebraeus* from Hawaii (Duda & Palumbi 1999a, 2000). These primers amplify approximately 250 bp of four-loop conotoxin transcripts that include about 220 bp of coding region (130 bp of the 'prepro' region of the conotoxin molecule that is cleaved from the mature conotoxin peptide during processing plus approximately 90 bp that encode the mature conotoxin peptide) and 30 bp of the 3′ untranslated region. The primers have been used previously to amplify 10 putative four-loop conotoxin loci from the two vermivorous *Conus* species listed above (Duda & Palumbi 1999a, 2000) and 10 other loci from five piscivorous *Conus* species (Duda & Palumbi 2004). Sequences of both primers show 100% identity to regions of four-loop conotoxin transcripts recovered from species that presumably diverged more than 30 million years ago based on previous phylogenetic analyses (Duda & Kohn 2005). Because these primers have been used to amplify four-loop conotoxins from a variety of phylogenetically disparate *Conus* species (T.F. Duda, unpublished results), they presumably amplify a large subset of four-loop conotoxin loci. Thus, utilization of these primers permits a reasonably broad survey of expression patterns of this assemblage of loci among closely related *Conus* species.

Amplification products generated with the TOX1/TOX2 primers were cloned using a TA cloning kit (Invitrogen) and at least 18 vector inserts were sequenced from amplification products of single individuals with M13 (TGTAAAA CGACGGCCAGT) and M13rev (CAGGAAAC-AGCTATGACC) vector primers. To ensure that adequate numbers of inserts were sequenced to detect most of the expressed conotoxin loci that TOX1/TOX2 amplify, we generated sequence diversity curves that compared the cumulative number of distinct sequence variants obtained and the number of sequenced inserts for each individual and each species. We sequenced additional inserts when sequence diversity curves from individuals did not reach plateaus and examined expression patterns of additional individuals when the species-based curves did not reach plateaus.

We used the absence of sequences to imply that particular loci were not expressed, but the failure to detect particular sequences could result if a transcript is rare (i.e. is expressed at low levels) or if loci exhibit sequence differences within priming sites among species. We did not use locus-specific amplifications or other approaches to specifically test if transcript sequences were missed because they were rare. Thus, we cannot be certain that absence of sequences is due to the lack of expression of particular loci and so we consider both possibilities (i.e. nonexpressed locus and low levels of expression) in our interpretations of conotoxin expression

patterns. Nonetheless, to determine whether the inability to recover certain transcripts may have resulted from amplification bias or amplification failure, we produced a cDNA library from venom duct mRNA of an individual of *C. ebraeus* from Hawaii, one of two species we examined for which only two putative conotoxin loci were recovered (see Results). We prepared the library using slightly modified protocols of the CloneMiner cDNA Library Construction kit and TA cloning kit (Invitrogen) to permit cloning of cDNA from small quantities of mRNA. Specifically, the cDNA-adapter complex formed with a ligation reaction was size fractionated and ethanol precipitated. Then, the resulting pellet was resuspended in TE (pH 8.0), column purified (QIAGEN) using two washes of guanidine hydrochloride (Amresco), and eluted with distilled water. Because initial mRNA abundance was low, we amplified the purified cDNA-adapter complex with a set of attB primers that are identical to regions of the adaptor sequences (GGGACAACTTTGTACAAAAAAGTTGG and CACAACTTTGTACAAGAAAGTTGG-GT) using the following conditions: 40 cycles of 94 °C, 30 s; 52 °C, 30 s; and 72 °C, 30 s. We then ligated the polymerase chain reaction (PCR)-amplified cDNA-adapter with a PCR 2.1 vector and transformed the plasmid into chemically competent *Escherichia coli*. We sequenced approximately 100 vector inserts from this library, examined sequences to identify putative four-loop conotoxin transcripts and compared these with transcripts recovered from amplifications from cDNA of this species using the TOX1 and TOX2 primers.

Phylogenetic analyses of conotoxin sequences

Sequences were edited and aligned by eye with SEQUENCHER (Gene Codes Corp.). Phylogenetic relationships and patterns of nucleotide substitutions were examined independently for each species to identify distinct sequences with potential PCR-induced errors, putative allelic variants and putative inparalogues (i.e. loci that originated via lineage-specific duplication after a speciation event *sensu* Koonin 2005). Sequences with PCR-induced errors were defined as sequences from a single individual that showed just one or two base differences from another sequence recovered from that individual or were comprised of regions of two or more other sequences obtained from that individual (i.e. PCR-induced recombinants). Allelic variants and inparalogues were defined as distinct sequences that possessed nucleotide substitutions relative to other sequences that did not seem to be due to PCR-induced errors and that showed greatest identity to other sequences identified from the species from which they were derived. All presumed PCR-induced artefacts as well as all but one of the allelic variants or inparalogues were excluded from analyses. The particular allelic variants or inparalogues that were included were those that showed greatest identity

to sequences from other species. Removal of the additional sequence variants from our analyses should not affect estimates of levels of interlocus divergence or the ability to identify putative orthologous loci. Nonetheless, exclusion of inparalogues will cause us to underestimate identity of expression patterns among species. Although resultant trees are not presented, we constructed phylograms that incorporated the excluded sequences. All excluded sequences (except for the suspected PCR-induced recombinant sequences) clustered tightly with other putative loci from the species from which they were recovered; suspected PCR-induced recombinants typically grouped at the bases of clades that contained one of the sequences from which the artefact sequence was presumably recombined.

We aligned sequences that appeared to be free of PCR-induced errors and only one putative inparalogue or one allele of each polymorphic locus with four-loop conotoxin sequences representing nine putative loci previously reported from *C. abbreviatus* [GenBank Accession nos AF090041 (A1), AF090035 (A2), AF089995 (A3), AF089988 (A4), AF089997 (A6), AF089983 (A7), AF090055 (A9); Duda & Palumbi 1999a] and one locus from *C. ebraeus* [GenBank Accession no. AF174268 (E1); Duda & Palumbi 2000]. Phylograms were produced with PAUP* (Swofford 2002) using neighbour-joining methods with maximum-likelihood distances based on a best-fit model of nucleotide substitution as determined with MODELTEST version 3.7 (Posada & Crandall 1998) and using maximum parsimony.

Determination of orthology and paralogy from transcript sequences

Resolving orthology and paralogy is difficult particularly if only transcript sequences are compared, yet it is essential for making inferences about the conotoxin expression patterns of these six *Conus* species. We used several criteria for determining orthology and paralogy of loci. Similar methods were used to identify paralogy and orthology of duplicated genes in frogs (Chain & Evans 2006). First, we examined phylograms constructed from transcript sequences to identify sets of similar sequences from different species. Sequences of different species that occur in clades with strong bootstrap support were considered to represent orthologues. We also compared the topology of clades within the transcript phylogram to a tree estimated from mitochondrial sequence data [a region of the cytochrome oxidase I (COI) gene]. In some cases, species may express orthologous counterparts: transcripts that represent loci of different species cluster together in subclades and the topologies of these subclades resemble the topology of the COI tree (i.e. the transcript tree contains a number of subgroupings that each bear resemblance to the COI tree) (e.g. locus 1 in Fig. 1). In other cases, species may not express the same loci: transcripts from different species do
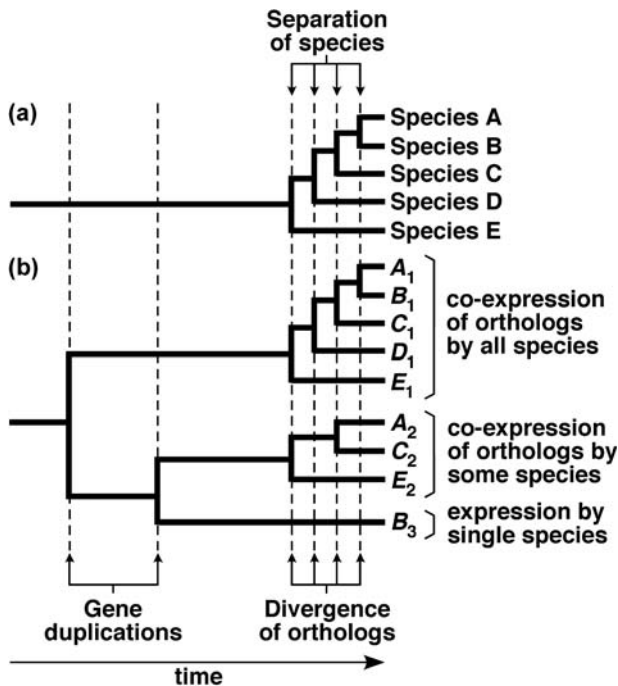
Fig. 1 Criteria used to identify orthologous and paralogous loci. (a) Hypothetical species tree of five species. (b) Hypothetical gene tree of a gene family constructed from expression data (i.e. mRNA transcript sequences) of these five species. The scales of the gene tree and species tree are identical and reflect timings of separation of loci and species, respectively. For transcripts to be considered representative of orthologous loci, topology of clades of sequences resembles the topology of the species tree and levels of divergence among sequences in clade correspond with levels of divergence among species. Loci that are paralogous do not meet these criteria. In this example, one locus is expressed by all species (e.g. locus 1); others are expressed by only a subset of species (e.g. locus 2: species A, C and E) or by a single species (e.g. locus 3: species B).

not cluster together tightly (e.g. locus 3 in Fig. 1). Although the mitochondrial locus used likely evolves at a different rate (i.e. more quickly) than nuclear gene sequences, the COI sequences provide greater resolution for determining relationships of closely related *Conus* species than other genes (see Duda & Kohn 2005; Duda & Rolán 2005).

Second, we examined levels of synonymous divergence among putative conotoxin loci and a presumably neutrally evolving intron of a calmodulin locus for determination of orthology and paralogy. Although we are aware that synonymous substitution rates can vary among species or even among loci within a species, synonymous divergence is a reasonable proxy, and is often used as an estimate, of time since divergence (Shiu *et al.* 2006). Hence, we expect that orthologous conotoxin loci exhibit lower levels of synonymous divergence than outparalogues (i.e. loci that originated via duplication preceding a speciation event *sensu* Koonin 2005) and show similar levels of divergence

as the intron sequences from the calmdoulin locus. Most if not all of the nine conotoxin loci previously recovered from *C. abbreviatus* likely originated from gene duplications that occurred before the separation of the species examined here (see Duda & Palumbi 1999a; Duda & Kohn 2005). Thus, synonymous divergence of these loci also provides a reference for determining orthology and paralogy. Accordingly, we expect that outparalogues exhibit similar or greater levels of synonymous divergence as the nine conotoxin loci of *C. abbreviatus*.

To provide support for our assignments of orthology and paralogy, we attempted to identify genomic sequences of loci that are expressed in some species from species that are presumably not expressing these loci. Specifically, we used a primer designed in the beginning of the mature toxin coding region (GTTGCACGCCTCCCGGTGGA) (primer sits just upstream of the second cysteine codon and is downstream from a suspected intron position in the prepro region) and the TOX2 primer previously mentioned to amplify four-loop conotoxin loci from genomic DNA of *C. abbreviatus*. Amplifications were performed under the following conditions: 40 cycles of 94 °C, 30 s; 45 °C, 30 s; and 72 °C, 30 s. We cloned and sequenced the amplification products as described above. Recovered sequences were examined and compared to transcript sequences with SEQUENCHER (Gene Codes Corp.) and PAUP* (Swofford 2002).

Molecular evolution of four-loop conotoxins

We estimated levels of synonymous (d_S) and nonsynonymous (d_N) divergence along branches of the resultant conotoxin phylogram with maximum-likelihood methods using the 'free-ratio' model with the CODEML module of PAML (Yang 1997). Although this model is not useful for testing hypotheses about positive selection and is very parameter rich, it assumes an independent ratio of d_N to d_S for each branch in a tree. We also used PAML to calculate pairwise estimates of d_N and d_S to compare levels of synonymous and nonsynonymous divergence among putatively orthologous and nonorthologous conotoxin loci. Estimates of d_S and d_N were calculated for 61 codons present in the coding region of the transcript (prepro plus mature toxin coding regions) and for 29 codons of the mature toxin coding region, respectively. Because the prepro region is generally conserved and presumably subject to strong purifying selection (Duda & Palumbi 1999a; Conticello *et al.* 2001), we only examined codons with the mature toxin coding region of the transcripts. For estimates of d_S values, we included codons within the prepro region to increase the number of synonymous sites examined. Codon positions that included gaps in some of the sequences and the codons immediately following the codon that encodes the terminal cysteine residue of the mature toxin were excluded; the latter region was removed from analyses because alignment of

the sequences in this region required insertions/deletions that affected reading frame and because stop codons occurred further upstream in some sequences relative to those of other sequences.

Phylogenetic relationships of species and interspecific levels of divergence at a nuclear intron

We examined patterns of species relationships and levels of synonymous divergence at a nuclear locus of the six *Conus* species for which four-loop conotoxin sequence data were obtained. We used sequences of a fragment of the mitochondrial COI gene to estimate phylogenetic relationships of these species. Sequences of an intron located at a calmodulin locus were examined to gauge levels of synonymous divergence among species at an orthologous locus. COI and calmodulin sequences of *Conus abbreviatus*, *C. aristophanes*, *C. coronatus*, *C. ebraeus* and *C. miliaris* were obtained from GenBank (Accession nos AY588148, AY588153, AY588165, AY588175, AY588203, AF113252, AY382037, AF113268, AF113272 and AF113290). We amplified and sequenced these loci from *Conus judaeus* using methods described elsewhere (Duda & Palumbi 1999b; Duda & Rolán 2005). We included a sequence of *Conus sponsalis* (GenBank Accession no. AY588220) to root the tree and used neighbour-joining methods with maximum-likelihood distances based on a best-fit model of nucleotide substitution as determined with MODELTEST version 3.7 (Posada & Crandall 1998) to generate a phylogenetic hypothesis for the six focal species listed above. Levels of support for branches were estimated with bootstrapping procedures with PAUP*. We used maximum-likelihood methods with the BASEML module of PAML (Yang 1997) to calculate levels of divergence of the calmodulin intron sequences along branches based on the topology of the COI tree as well as to estimate pairwise genetic distances among sequences.

Patterns and evolution of conotoxin expression

We estimated indices of similarity in expression of four-loop conotoxins among species based on the proportion of loci that show similar expression patterns among species in relation to the total number of loci identified from the six species investigated. As noted above, exclusion of inparalogues from analyses would cause an overestimation of these indices.

The evolution of conotoxin gene expression was also examined with MACCLADE 4.0 (Maddison & Maddison 1992). Specifically, changes in expression of putative loci were determined by tracing the expression of conotoxin loci in relation to the hypothesized phylogeny of the six species reconstructed from the COI data. First, we assumed that all loci expressed by the six species occurred in the common ancestor of these species. Because unexpressed

loci would presumably accumulate a variety of mutations that could render their gene products nonfunctional, transitions from expression to nonexpression of a locus were assumed to be irreversible in this case. Furthermore, although the failure to detect particular conotoxin transcripts could be explained by lack of expression of these loci, it could also result from gene loss, something that is definitely irreversible. Second, we examined patterns of change in expression under the assumptions that transitions are reversible to account for the possibility of reactivation of formerly silenced loci and that no gene loss has occurred.

Results

Four-loop conotoxin sequences

Examination of 507 sequences recovered from two individuals each of *Conus aristophanes*, *C. coronatus*, *C. ebraeus* and *C. judaeus* and 10 of *C. miliaris* revealed 111 distinct sequences. Many of the distinct sequence variants appeared to represent PCR-induced artefacts, and in some cases, either allelic variants of polymorphic loci or inparalogues. Exclusion of the putative artefacts and all but one of the allelic variants or inparalogues and inclusion of sequences of the nine putative loci of *Conus abbreviatus* (A1–A9) and one previously described locus of *C. ebraeus* (E1) gave 39 sequences that presumably represent distinct loci of these species (GenBank Accession nos DQ644543–DQ644549 and EF108267–EF108288). These include an additional sequence of *C. ebraeus*, seven sequences from *C. aristophanes*, seven sequences from *C. coronatus*, two sequences from *C. judaeus* and 12 sequences from *C. miliaris*. In most cases, putative loci were identified from all individuals of a species that were examined. However, in a few cases, loci appeared to be differentially expressed within species, including one locus of *C. ebraeus* (E2) that was never detected from cDNA of individuals of this species from Hawaii. Nucleotide and predicted amino acid alignments of all sequences is available as Supplementary material (Figs S1 and S2). In all cases, sequence diversity curves generated for each individual ultimately reached plateaus; this result implied that sufficient numbers of inserts were examined to permit a reasonable assessment of the expression patterns of these individuals. We surveyed expression patterns of 10 individuals of *C. miliaris* because the sequence diversity curve generated from data combined from the first two individuals examined failed to reach a plateau.

Venom duct transcripts that were obtained from a cDNA library of an individual of *C. ebraeus* were similar to sequences of various conotoxin superfamilies and an assortment of 'house-keeping' genes or showed no identity to sequences in GenBank. Only two of the conotoxin sequences recovered were members of the O-superfamily (i.e. four-loop conotoxins). One of these sequences is identical to sequences
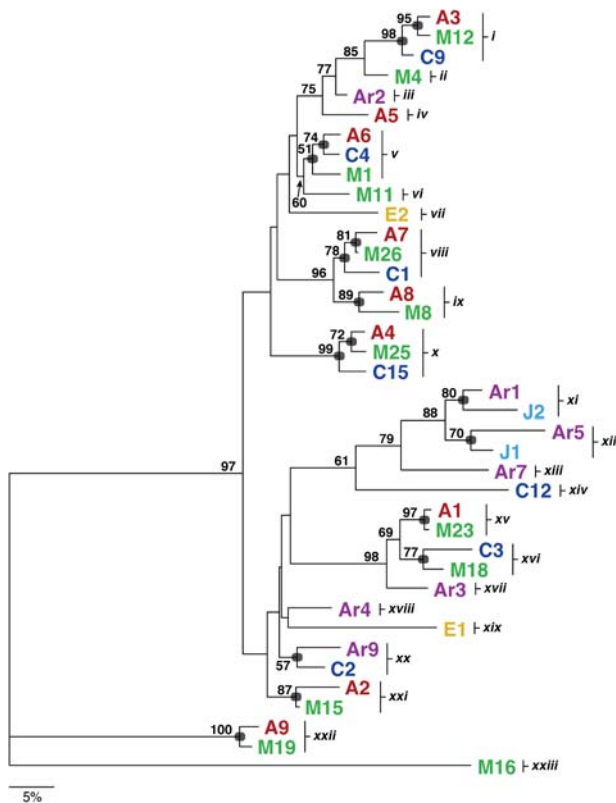
Fig. 2 Phylogram of 39 putative four-loop conotoxin loci of *Conus abbreviatus*, *C. aristophanes*, *C. coronatus*, *C. ebraeus*, *C. judaeus* and *C. miliaris* reconstructed with neighbour-joining based on maximum-likelihood distances calculated with comparison of entire transcript sequences. Bootstrap scores for branches are shown at nodes. Loci are identified as follows: A, *C. abbreviatus*; Ar, *C. aristophanes*; C, *C. coronatus*; E, *C. ebraeus*; J, *C. judaeus*; and M, *C. miliaris*; arbitrary numerical identifiers are given for each putative locus. Clades of sequences that appear to represent orthologous loci and single sequences that may represent loci that are only expressed by single species are labelled with roman numerals (i–xxiii). Nodes at the bases of clades that contain putative orthologous loci are indicated with grey ovals.

of locus E1 that were previously obtained from *C. ebraeus* via amplification with the TOX1 and TOX2 primers (Duda & Palumbi 2000) and that was examined here. The other four-loop conotoxin sequence appears to be a delta-conotoxin based on similarity of this sequence to those of known delta-conotoxins. Delta-conotoxins though are quite divergent from the particular set of four-loop conotoxins that can be amplified with the TOX1 and TOX2 primers and so the absence of this sequence in screens of amplification products produced with these primers is not surprising. None of the sequenced transcripts were identical to sequences of locus E2, the other locus identified from *C. ebraeus*, but this was expected because the cDNA library was constructed from an individual of this species that was

collected at Hawaii and individuals from this site have not been observed to express locus E2 (see above).

Phylogenetic analyses of conotoxin sequences

Based on results from MODELTEST, the TrNef + G model was used to calculate maximum-likelihood distances [rate matrix: (A–C) = 1.0000, (A–G) = 1.9398, (A–T) = 1.0000, (C–G) = 1.0000, R(e) (C–T) = 1.2029, (G–T) = 1.0000; gamma distribution shape parameter = 0.4634]. Previous analyses of four-loop conotoxin sequences A1–A9 of *C. abbreviatus* and E1 of *C. ebraeus* as well as published four-loop conotoxin sequences of other species of *Conus* reveal that sequence A9 of *C. abbreviatus* is basal relative to the other sequences of this species (Duda & Palumbi 1999a, 2004). Because one of the sequences of *C. miliaris* (M19) groups strongly with sequence A9 and another sequence from this species (M16) is roughly equidistant from the clade containing A9 + M19 as well as from all other sequences, conotoxin phylograms were rooted to sequences A9, M19 and M16.

The phylograms of the 39 four-loop conotoxin transcripts of six closely related *Conus* species that were constructed using maximum parsimony (tree not shown) and neighbour joining (Fig. 2) were quite similar and exhibited identical topologies for clades with strong bootstrap support. Differences among trees included rearrangements of branches that were poorly supported in both trees. Although discussion of loci presented below refers to relationships of loci resolved by the neighbour-joining tree, the following statements are also consistent with interpretation of the topology of the maximum parsimony tree. We also examined gene trees constructed while excluding the region of the transcript that encodes the mature conotoxin sequences and the topologies of these trees were comparable to the topology presented in Fig. 2.

The conotoxin gene tree exhibits several well-resolved clusters of sequences (Fig. 2). In total, 12 clades that include 28 sequences are defined that contain sequences from more than one species but not more than one sequence from a single species: four of these clades contain sequences of *C. abbreviatus*, *C. coronatus* and *C. miliaris*, four contain sequences of *C. abbreviatus* and *C. miliaris*, two contain sequences of *C. aristophanes* and *C. judaeus*, one contains sequences of *C. coronatus* and *C. miliaris* and one contains sequences of *C. aristophanes* and *C. coronatus* (Fig. 2, Table 1). These 12 sets of sequences presumably represent orthologous loci that are co-expressed by two or more species. The remaining 11 sequences likely represent loci that are uniquely expressed by just one of these six species.

COI data and phylogenetic analyses of species

A 634-bp fragment of the mitochondrial COI gene was obtained from *C. judaeus* (GenBank Accession no. EF108266)

Table 1 Expression patterns of four-loop conotoxins of closely related vermivorous *Conus* species. Identities (percentage) of expression patterns among species are indicated above diagonal based on the proportions of loci that are similarly expressed. Numbers of putative loci that are co-expressed among species are indicated below diagonal; numbers that are uniquely expressed in only one species are indicated along diagonal. Species abbreviations are those as listed in Fig. 2. Numbers of putative loci of each species are shown in parentheses next to species' abbreviations

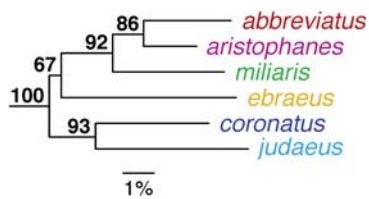|	A	Ar	M	E	C	J
A (9)	1	31.8	81.8	54.5	68.2	54.5
Ar (7)	0	4	18.2	63.6	50.0	81.8
M (12)	8	1	3	40.9	63.6	40.9
E (2)	0	0	0	2	63.6	86.4
C (7)	4	1	4	0	1	63.6
J (2)	0	2	0	0	0	0

Fig. 3 Phylogram of *Conus abbreviatus*, *C. aristophanes*, *C. coronatus*, *C. ebraeus*, *C. judaeus* and *C. miliaris* reconstructed with neighbour-joining of maximum-likelihood distances calculated from a region of the mitochondrial COI gene. Bootstrap scores for branches are shown at nodes.

and combined with published COI sequences of *C. abbreviatus*, *C. aristophanes*, *C. coronatus*, *C. ebraeus*, *C. miliaris* and *C. sponsalis* (the sequence from the latter species was used to root the tree). Based on results from MODELTEST, the HKY + G model was used to calculate maximum-likelihood distances (frequencies of bases: A = 0.2403, C = 0.1749, G = 0.2213 and T = 0.3635; transition/transversion ratio = 22.1516; gamma distribution shape parameter = 0.1208). A phylogram built with neighbour-joining methods using maximum-likelihood distances among sequences and bootstrap values from 1000 replicates were determined with PAUP* (Swofford 2002) and are presented in Fig. 3. Relationships inferred from this tree suggest that *C. abbreviatus* and *C. aristophanes* as well as *C. coronatus* and *C. judaeus* are sister species and that *C. miliaris* separated from the lineage that gave rise to the two former species.

Topological comparisons of the conotoxin and COI gene trees

We compared the topology of the phylogeny reconstructed from analyses of the COI sequence data (Fig. 3) to the topologies of clades present in the conotoxin phylogram (Fig. 2) to assess orthology and paralogy of the conotoxin loci. Of the 12 conotoxin subclades that were identified, topological comparisons were only possible for the four clades that contain loci from more than two species (putative orthologous loci i, v, viii and x; Fig. 2). These clades contain sequences from *C. abbreviatus*, *C. coronatus* and *C. miliaris*. The identity in topology of three of the four clades (loci i, viii and x; Fig. 2) to that inferred from the COI phylogeny implies that the sequences represent orthologous loci. The topology of the fourth clade (locus v, Fig. 2) though differs from expectations of species relationships based on the COI data and so the locus recovered from *C. miliaris* (M1) may be a paralogue of loci A6 and C4 of *C. abbreviatus* and *C. coronatus*. However, levels of synonymous divergence among sequences A6, C4 and M1 are consistent with expectations that these represent orthologous loci (see below).

Examination of the topology of the remainder of the conotoxin phylogram in consideration of the patterns of species relationships shown in the tree derived from COI sequence data reveals that orthologous counterparts of loci identified from particular species or sets of species are not expressed by other species. For example, locus xxii contains sequences from *C. abbreviatus* and *C. miliaris* but no loci from the other four species examined show affinity to these sequences (Fig. 2). Several other loci or clades of loci exhibit this pattern too, including loci vii, x, xiv, xxiii (Fig. 2). Also, relationships among loci indicate that many of these loci are paralogues. For example, the topology of the clade that contains loci xv, xvi and xvii (Fig. 2) can best be explained by the paralogy of these loci and that they originated via gene duplication in the common ancestor of the six species examined.

Nuclear intron sequence data and patterns of sequence divergence among species

A 266-bp fragment of a calmodulin gene that contains a 210-bp intron was obtained from *C. judaeus* (GenBank Accession no. EF108265) and aligned with published sequences of this locus from *C. abbreviatus*, *C. aristophanes*, *C. coronatus*, *C. ebraeus* and *C. miliaris*. While a phylogeny constructed with these sequences was not completely resolved, it did not contradict the topology exhibited in the COI tree. We excluded exon sequences and estimated levels of divergence of the intron sequence with the BASEML module of PAML (Yang 1997). Pairwise estimates of Kimura 2-parameter distances (Kimura 1980) among intron sequences ranged from 0.015 to 0.042 (average = 0.028). Also, levels of sequence divergence (based on Kimura 2-parameter distances) were calculated for each branch of the tree constructed from analyses of the COI data. Divergence among branches connected by nodes ranged
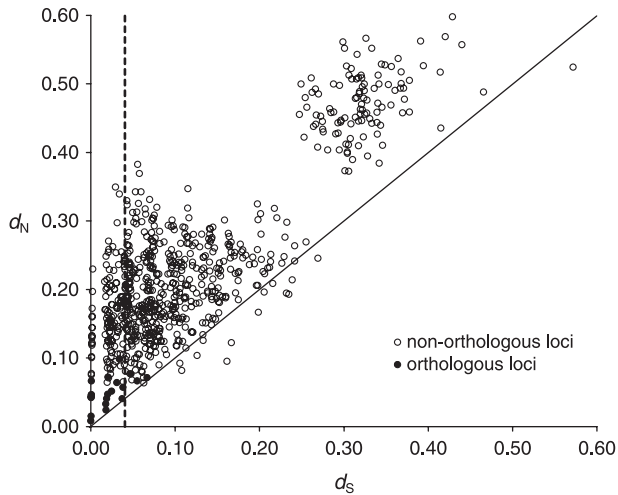
Fig. 4 Comparison of pairwise estimates of synonymous (d_S) and nonsynonymous (d_N) divergence among putative loci of six closely related *Conus* species. Comparisons among putative orthologous loci are indicated with filled circles while comparisons among putative nonorthologous loci are indicated with open circles. A dashed line at $d_S = 0.042$ indicates the maximum genetic distance observed between calmodulin intron sequences of the *Conus* species examined. The diagonal line corresponds to $d_N = d_S$ or $d_N/d_S = 1$; points above the diagonal indicate comparisons in which $d_N > d_S$ ($n = 726$) while those below indicate comparisons in which $d_N < d_S$ ($n = 15$).

from 0.016 for one of the nodes that directly connects two terminal branches (the node connecting the sister species *C. abbreviatus* and *C. aristophanes*) to 0.041 for a node that occurs deeper in the tree.

Patterns of divergence of four-loop conotoxin loci

Pairwise estimates of synonymous divergence (d_S) among members of the 12 sets of sequences that presumably represent orthologous loci range between 0 and 0.067 and all except three pairwise distances are less than 0.042 (i.e. the maximum pairwise distance of orthologous nuclear loci among species based on examination of the calmodulin intron sequences) (Fig. 4). Pairwise estimates of nonsynonymous divergence (d_N) range from 0.008 to 0.098 (Fig. 4) with a mean d_N to d_S ratio of 2.34. Pairwise estimates of d_S and d_N among sequences that presumably represent nonorthologous loci range from 0.001 to 0.572 and from 0.047 to 0.598, respectively (Fig. 4) with a mean d_N to d_S ratio of 1.99. Of the 61 codons that were examined, on average 45.2 positions were considered to be synonymous sites (range: 35.5–59) and 137.8 positions were considered to be nonsynonymous sites (range: 124–147.5).

Tree-based estimates of synonymous (d_S) and nonsynonymous divergence (d_N) among conotoxin transcript

sequences were determined from analyses of an average of 60 synonymous sites within the prepro and mature toxin coding regions and 120 nonsynonymous sites within the mature toxin coding region, respectively (Fig. 5). Similar to values obtained from pairwise comparisons among pairs of putative orthologous loci, estimates of d_S range between 0 and 0.051 for these sets of loci. From the tree-based estimates of d_S, we also calculated levels of synonymous divergence among putative paralogous loci for each species. Assuming that these sequences represent paralogous loci that originated by duplication in the most recent common ancestor of these species, they should show levels of synonymous divergence greater than 0.042. Only 10 of 80 pairwise estimates of d_S among putative paralogues of *C. abbreviatus*, *C. aristophanes*, *C. coronatus*, *C. ebraeus* and *C. judaeus* show values less than 0.042; other values range from 0.042 to 0.426 (values derived from Fig. 3). For *C. miliaris*, 20 of 60 estimates of d_S among putative paralogues are less than 0.042 and the remainder of the values range between 0.042 and 0.481 (values derived from Fig. 3).

Unexpressed locus from genomic DNA

We obtained a sequence from genomic DNA of *C. abbreviatus* (GenBank Accession no. EU404113) that is nearly identical to transcript Ar7 from *C. aristophanes*. The sequence consists of 114 bp, including 72 bp of the mature toxin coding region and 42 bp of the 3′ untranslated region. The sequence from genomic DNA of *C. abbreviatus* and Ar7 from mRNA of *C. aristophanes* differed at only three positions: two nonsynonymous sites and one synonymous site. Based on the similarity of these sequences, the sequence recovered from *C. abbreviatus* presumably represents an orthologue of Ar7 that is not expressed by *C. abbreviatus*.

Patterns and evolution of conotoxin expression

Five of the six species examined express at least one four-loop conotoxin locus that is not expressed by the other six species; the two putative loci expressed by *C. judaeus* are apparently orthologous to two loci expressed by *C. aristophanes* (Fig. 2, Table 1). In total, 28 of 39 loci are co-expressed by two or more species; the remaining 11 loci are uniquely expressed by only one species. The percentage of loci co-expressed by more than one species ranges from 0% (loci of *C. ebraeus*) to 100% (loci of *C. judaeus*) for each species. *Conus aristophanes* and *C. miliaris* express four and three loci, respectively, that are not expressed by any of the other species. Per cent similarity in expression patterns based on the frequency of loci that show similar expression patterns among species ranges from 18.2% (*C. aristophanes* and *C. miliaris*) to 86.4% (*C. ebraeus* and *C. judaeus*) (Table 1). Because estimates of identity in expression pattern are based on the proportion of loci that are similarly expressed,
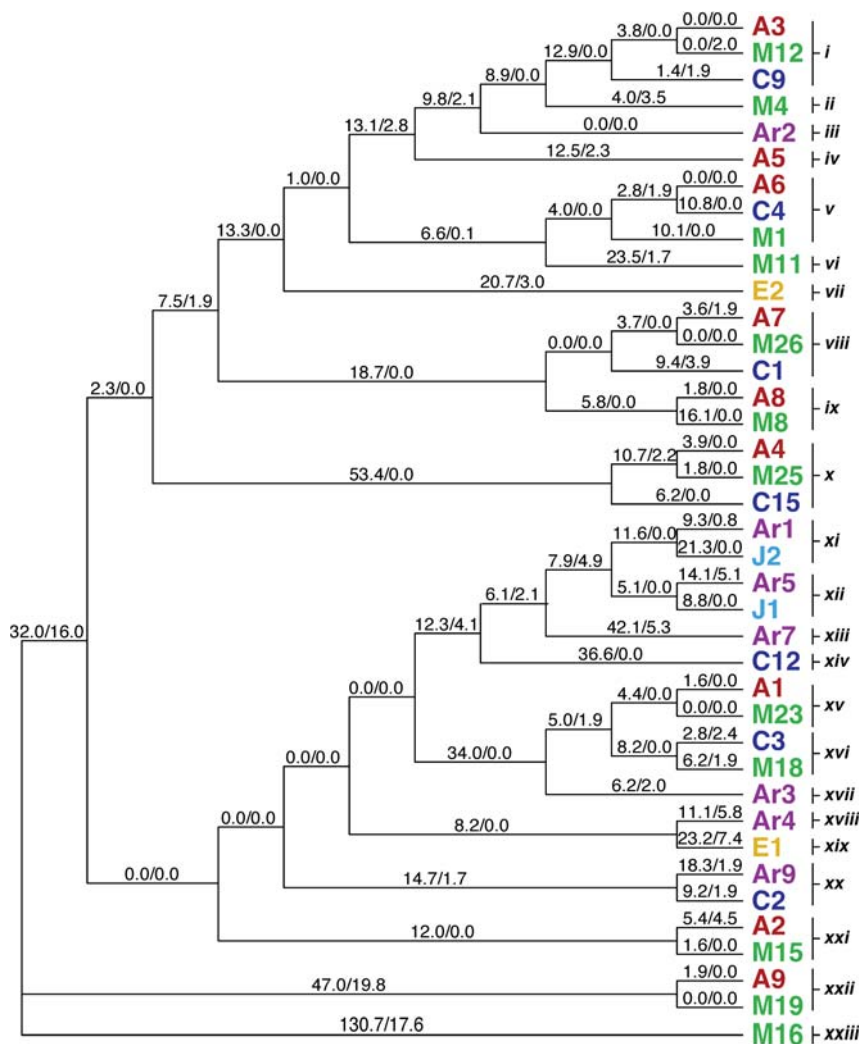
Fig. 5 Tree-based estimates of synonymous (d_S) and nonsynonymous (d_N) divergence among branches of the conotoxin phylogram as calculated with the free-ratio model in PAML (Yang 1997). Values of d_N were estimated from comparison of aligned codons of the mature toxin coding region (first number before slash) and values of d_S were estimated from comparison of aligned codons of both the prepro and mature toxin coding regions (second number after slash) (values are reported as X 100 for presentation purposes). Loci are identified as in Fig. 2. Roman numerals correspond to those given in Fig. 2.

Phylogram branch values and terminal taxa:

- A3, M12, C9 — i (3.8/0.0; 0.0/0.0; 0.0/2.0; 1.4/1.9; 12.9/0.0)
- M4 — ii (4.0/3.5; 8.9/0.0)
- Ar2 — iii (0.0/0.0; 9.8/2.1)
- A5 — iv (12.5/2.3; 13.1/2.8)
- A6, C4, M1 — v (2.8/1.9; 0.0/0.0; 10.8/0.0; 10.1/0.0; 4.0/0.0; 6.6/0.1)
- M11 — vi (23.5/1.7)
- E2 — vii (20.7/3.0)
- A7, M26, C1 — viii (3.7/0.0; 3.6/1.9; 0.0/0.0; 9.4/3.9; 0.0/0.0; 18.7/0.0)
- A8, M8 — ix (1.8/0.0; 16.1/0.0; 5.8/0.0)
- A4, M25, C15 — x (3.9/0.0; 1.8/0.0; 6.2/0.0; 10.7/2.2; 53.4/0.0)
- Ar1, J2 — xi (9.3/0.8; 21.3/0.0; 11.6/0.0)
- Ar5, J1 — xii (14.1/5.1; 8.8/0.0; 5.1/0.0; 7.9/4.9)
- Ar7 — xiii (42.1/5.3; 6.1/2.1)
- C12 — xiv (36.6/0.0; 12.3/4.1)
- A1, M23 — xv (1.6/0.0; 0.0/0.0; 4.4/0.0)
- C3, M18 — xvi (2.8/2.4; 6.2/1.9; 8.2/0.0; 5.0/1.9)
- Ar3 — xvii (6.2/2.0; 34.0/0.0)
- Ar4 — xviii (11.1/5.8; 8.2/0.0)
- E1 — xix (23.2/7.4)
- Ar9, C2 — xx (18.3/1.9; 9.2/1.9; 14.7/1.7)
- A2, M15 — xxi (5.4/4.5; 1.6/0.0; 12.0/0.0)
- A9, M19 — xxii (1.9/0.0; 0.0/0.0; 47.0/19.8)
- M16 — xxiii (130.7/17.6)

Internal branches: 1.0/0.0; 13.3/0.0; 7.5/1.9; 2.3/0.0; 0.0/0.0; 0.0/0.0; 0.0/0.0; 0.0/0.0; 32.0/16.0

including loci that are co-expressed as well as those that are not expressed, large values not only identify species that co-express many of the same loci (e.g. *C. abbreviatus* and *C. coronatus*, 68.2%; *C. abbreviatus* and *C. miliaris*, 81.8%; and *C. coronatus* and *C. miliaris*, 63.6%; *C. aristophanes* and *C. judaeus*, 81.8%), but also species that do not express a number of loci in common (e.g. *C. aristophanes* and *C. ebraeus*, 63.6%; *C. coronatus* and *C. ebraeus*, 63.6%; *C. coronatus* and *C. judaeus*, 63.6%; *C. ebraeus* and *C. judaeus*, 86.4%) (Table 1).

In total, 74 transitions (all homoplasious) from expression to nonexpression of loci must have occurred if these transitions are irreversible or if gene loss occurred (Fig. 6a). However, three sets of loci show similar expression patterns in two or more species: loci i, v, viii and x ('ACM') that are co-expressed by *C. abbreviatus*, *C. coronatus* and *C. miliaris*; loci ix, xv, xxi and xxii ('AM') that are co-expressed by *C. abbreviatus* and *C. miliaris*; and loci xi and xii ('ArJ') that are co-expressed by *C. aristophanes* and *C. judaeus* (Fig. 2). If expression of these sets of loci is controlled by a single factor

(e.g. linked or shared regulation), transitions from expression to nonexpression of these sets would represent single instead of multiple steps; physical proximity of these locus sets and deletion of the chromosomal region that contains these sets could also represent single steps. Under this scenario, 52 transitions (also all homoplasious) are necessary to account for the patterns of expression among species. Alternatively, if transitions from expression to nonexpression are reversible, 39 steps are required to explain current expression patterns and all but 11 transitions are homoplasious (Fig. 6b). As above, if transitions in expression of sets of loci that show similar expression patterns among two or more species are considered to represent single events, only 22 steps are necessary.

Discussion

From analyses of 517 four-loop conotoxin sequences recovered from cDNA, we identified 39 putative conotoxin
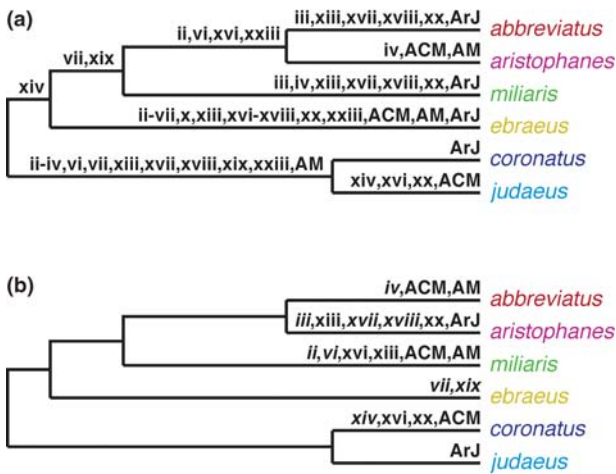
Fig. 6 Hypothesized transitions in expression of four-loop conotoxins among *Conus abbreviatus*, *C. aristophanes*, *C. coronatus*, *C. ebraeus*, *C. judaeus* and *C. miliaris*. Cladograms are based on the relationships of species inferred from the COI sequence data (see Fig. 3); for illustration purposes branch lengths are arbitrary and not drawn to scale. Transitions in expression of particular four-loop conotoxin loci to account for patterns of gene expression of the six species are indicated with names of loci on branches (locus names correspond to labels given in Figs 1 and 4; ACM, loci i, v, viii and x that are co-expressed by *C. abbreviatus*, *C. coronatus* and *C. miliaris*; AM, loci ix, xv, xxi and xxii that are co-expressed by *C. abbreviatus* and *C. miliaris*; ArJ, loci xi and xii that are co-expressed by *C. aristophanes* and *C. judaeus*). (a) Transitions from expression to nonexpression are irreversible and the common ancestor of the six species expressed ancestral copies of all loci currently expressed by these species. Labels on branches indicate loci that have been silenced in a particular lineage. (b) Transitions in expression are reversible and the common ancestor of the six species may not have expressed ancestral copies of all loci that these species currently express. Labels on branches indicate loci that have been reactivated in a particular lineage; transitions that are unique to a particular lineage are indicated with labels in italics. An equally parsimonious explanation for expression patterns of locus set ACM is that the ancestral copies of these loci were expressed by the most recent common ancestor of these species and that descendant loci were independently silenced in *C. aristophanes*, *C. ebraeus* and *C. judaeus*. Similarly, an equally parsimonious explanation for the expression patterns of locus sets ACM and AM for *C. abbreviatus*, *C. aristophanes* and *C. miliaris* is that ancestral copies of these loci were expressed by the most recent common ancestor of these species and that expression of descendant loci was only maintained in *C. abbreviatus* and *C. miliaris*.

loci from six *Conus* species that presumably represent 23 outparalogues (*sensu* Koonin 2005). Twelve sets of loci exhibit patterns of relationships and show levels of synonymous divergence that suggest that members of these sets represent transcripts of orthologous loci. The remaining 11 putative outparalogues appear to be loci that are uniquely expressed by single species.

We interpreted the absence of orthologous counterparts as evidence that species do not express the same conotoxin loci. The objective of the combined phylogenetic and synonymous divergence approach was to identify orthologous loci and loci that did not satisfy our criteria of orthology were deemed paralogues. Application of such criteria to identify orthologues could cause orthologous loci to be misidentified as paralogues in some cases. Furthermore, we may not have recovered particular transcripts from some species because these transcripts are expressed at low levels, their expression is phenotypically plastic, or our primers fail to amplify them due to substitutions within priming sites. For example, without more extensive surveys of conotoxin transcripts from additional individuals including representatives of all developmental stages and from a variety of locations, we cannot reject the hypotheses that some of the loci we considered to be silenced are actually expressed at low levels or are differentially expressed within species. However, because different species express different numbers of putative loci and a putative unexpressed orthologous counterpart was recovered from genomic DNA, we suspect that most if not all of the putative orthologues and paralogues were correctly identified. Moreover, we assume that amplification bias is not responsible for the failure to detect certain transcripts because our survey of sequences from a venom duct cDNA library of *Conus ebraeus* detected the same transcripts that were identified using the amplification approach. Also, conotoxin expression patterns were largely consistent among individuals of the species we examined and so phenotypic plasticity in expression does not appear to be a major concern. Further confirmation of our assignments of orthology/paralogy and our interpretations of conotoxin expression patterns will require genome mapping and additional examination of conotoxin gene sequences recovered from genomic DNA and mRNA.

Phylogenetic relationships and conotoxin expression patterns of species

Similar to results from analyses of gene expression in other taxa in which expression patterns are not related to species relationships (e.g. salmonids, Derome & Bernatchez 2006; Giger *et al.* 2006; *Fundulus*, Whitehead & Crawford 2006b; primates, Gilad *et al.* 2006), our results suggest that differences in expression of conotoxins in *Conus* cannot be explained by phylogeny. The two species that most recently separated, *Conus abbreviatus* and *Conus aristophanes* (see Fig. 3), express 16 of the 39 putative conotoxin loci that were identified (Table 1), yet none of these loci appear to be orthologous (Figs 2 and 4). COI and calmodulin data suggest that *C. abbreviatus* and *C. aristophanes* shared a most recent common ancestor with *Conus miliaris* (Fig. 3). While *C. abbreviatus* and *C. miliaris* co-express eight loci and each

species uniquely expresses one and four other loci, respectively, *C. aristophanes* and *C. miliaris* co-express just one locus out of the 19 loci that are expressed by these species (Table 1). Although *C. abbreviatus* and *C. miliaris* display one of the largest identities in expression (81.8%), the two lowest identities were observed for comparisons between *C. abbreviatus* and *C. aristophanes* (31.8%) and *C. aristophanes* and *C. miliaris* (18.2%) (Table 1). These results suggest that the eight loci co-expressed by *C. abbreviatus* and *C. miliaris* were silenced in or lost by *C. aristophanes* or that these genes were independently reactivated in the two former species. Alternatively, if our approach failed to recover conotoxin transcripts that were expressed at low levels, the loci that were recovered from *C. abbreviatus* and *C. miliaris* but not from *C. aristophanes* may be expressed at higher levels in these two former species than they are in *C. aristophanes*. Thus, unless the orthologous counterparts from *C. aristophanes* are considerably divergent and hence unrecoverable or unrecognizable as orthologues of the eight loci that are putatively co-expressed by *C. abbreviatus* and *C. miliaris*, conotoxin expression patterns do not correspond with species relationships independent of whether the loci of *C. aristophanes* are expressed at low levels or not expressed at all.

Based on species relationships (Fig. 2), co-expression of four conotoxin loci by *C. abbreviatus*, *C. coronatus* and *C. miliaris*, two loci by *C. aristophanes* and *C. judaeus*, and one locus by *C. aristophanes* and *C. coronatus* implies that the common ancestor of these species, which is presumably the most recent common ancestor of the six species examined (see Fig. 2), possessed the ancestral loci that gave rise to the loci currently expressed by these species. Moreover, comparison of levels of synonymous divergence between putative paralogous loci and levels of divergence of orthologous loci (i.e. most putative paralogous conotoxin loci are more divergent than intron sequences of an orthologous nuclear locus from these species) implies that most loci likely arose from gene duplication events before the separation of the lineages that gave rise to these species. Thus, if the expression patterns of these four-loop conotoxins resemble those of other venom components, differences in venom composition among *Conus* species result largely from differences in expression of members of the various large conotoxin gene families.

Duda & Palumbi (2004) reported previously that piscivorous and vermivorous *Conus* species express unique sets of four-loop conotoxin loci despite results from phylogenetic analyses that show that piscivory evolved more than once in *Conus* and that some piscivores are more closely related to vermivores than they are to other piscivores. Duda & Palumbi (2004) proposed that these expression patterns evolved convergently in relation to diet. Similarity in conotoxin expression patterns of some of the species examined here may also have resulted from the convergent evolution of conotoxin gene expression.

Lazarus loci

The observed conotoxin expression patterns of the six species examined could be explained by the retained expression of some loci in some species and the independent silencing or gene loss of other loci in other species, but this requires that the common ancestor of these species expressed the 23 loci that are currently expressed by its descendants. Alternatively, patterns of expression could be explained by the independent activation of previously silenced loci in particular lineages and the common ancestor of these species did not express all 23 loci. Based on the average (6.5) and maximum (12) number of four-loop conotoxin loci recovered from the six species we examined, it seems more plausible that the common ancestor of these species did not express all 23 of the loci.

Based on analyses of the number of steps necessary to explain the evolution of the observed expression patterns (Fig. 6), one explanation for the differences in these patterns is that some loci that are currently expressed by the species were not expressed by their most recent common ancestor and that previously silenced genes were reactivated in particular species. Because silenced genes are not exposed to selection, they should degrade because of mutations that alter reading frame or otherwise significantly change their function, and hence, these genes should ultimately become pseudogenes. However, gene death may fail to occur if the time frame over which a gene is silenced is brief. Marshall *et al*. (1994) estimated that unexpressed loci can endure periods of inactivation up to about 10 million years without acquiring mutations that cause loss of function. The probability that a silenced gene survives point mutations and frameshift mutations is dependent on the rate at which these events occur, sensitivity of the gene's function to nucleotide substitutions, the time interval over which the gene is silenced and length of the gene's coding sequence, with smaller genes having a lower probability of loss of function than larger ones (Marshall *et al*. 1994). The small size of the exon regions of conotoxin genes (e.g. the coding region of four-loop conotoxin genes contains less than 270 bp) may permit these genes to retain function after long periods of inactivation (i.e. for periods greater than 10 million years). Based on the fossil record and rates of evolution of calmodulin intron sequences of *Conus*, the species examined here probably originated from a common ancestor within about the last 10 million years (Duda & Kohn 2005) and so it is possible that conotoxin genes that were silenced in the common ancestor of these species did not degrade over this time.

Thus, patterns of conotoxin expression may be explained by the resurrection of 'Lazarus' conotoxin loci and interspecific differences in venom composition of *Conus* species hinge on the expression of particular members of incredibly diverse conotoxin gene families. Based on conotoxin

expression patterns of a set of four phylogenetically disparate *Conus* species, Conticello *et al.* (2001) hypothesized that distinct venoms of *Conus* species arise via conotoxin gene duplication, hypermutation and shifts in expression of particular loci, including enhanced expression of 'lazarotoxins' that were previously expressed at low levels. Although analyses of our data provide no clear evidence of hypermutation, our results support the other aspects of this hypothesis. While the term 'Lazarus effect' has been applied to species that exhibit large gaps in the fossil record (Flessa & Jablonski 1983) in reference to the biblical story of the resurrection of Lazarus by Jesus, 'Lazarus loci' refers to silenced genes that have been reactivated after a period of dormancy. The birth-and-death model of gene family evolution states that gene duplication creates new gene family members and that while some genes are maintained in the genome, others suffer gene *death* through deleterious mutations or gene loss (Nei *et al.* 1997). Reactivation of silenced genes may also play an important role in gene family evolution and, as discussed by Marshall *et al.* (1994), represent an important means by which homoplasious conditions emerge among taxa. Starbäck *et al.* (2000) posited that resurrection of loci could explain patterns of expression of a neuropeptide receptor gene in mammals. However, the length of time that these genes would have been silenced before reactivation and patterns of sequence divergence among loci imply that independent inactivation of genes in particular lineages is a more plausible hypothesis to explain the expression patterns of this gene. Thus, if our assignments of orthology and paralogy are correct and the failure to detect orthologous counterparts in species reflects gene silencing and not an inability to recover transcript sequences of lowly expressed genes, gene reactivation appears to be an important aspect of the evolution of conotoxin gene expression patterns and could apply to the evolution of gene expression in other organisms as well, although we are unaware of examples of this in other taxa.

However, genes that are not expressed during a particular developmental stage or in a particular tissue may be immortal because they are expressed at other stages or in other tissues and selection prevents their decay. For example, the evolution of unique expression patterns of snail and slug transcription factors in vertebrates is best explained by the hypothesis that the genes that were reactivated in certain tissues of particular taxa did not degenerate because they serve functions in other tissues or developmental stages (Locascio *et al.* 2002). The *Conus* individuals examined in this study were adults and conotoxin expression patterns of juveniles or subadults of these species are unknown, but examination of expression patterns of these early life-history stages may reveal that conotoxin loci are differentially expressed during ontogeny (see Duda & Palumbi 2004). Because venoms are utilized primarily for capturing prey and some *Conus* species show differences in diet during

development (Nybakken & Perron 1988), different sets of conotoxin loci may be expressed during different life-history stages. The presumed primary function of conotoxins is to paralyse prey and so it is unlikely that conotoxin genes are expressed in tissues other than those involved with venom production. Nonetheless, we are not aware of any studies that have examined conotoxin expression patterns in different tissue types of *Conus*. Furthermore, conotoxins may be differentially expressed within or among populations of species or expression may be phenotypically plastic such that the expression patterns we observed are specific to the particular individuals or populations we sampled. Indeed, locus E2 from *C. ebraeus* was not detected previously from individuals from Hawaii but was expressed by individuals from Panama. Also, based on liquid chromatography/mass spectrometry analyses of peptides recovered from the venoms of *Conus textile*, Jones *et al.* (1995) showed that different individuals of this species express unique venom components. However, recent analyses of expressed four-loop conotoxins of more than 75 individuals of *C. ebraeus* from populations in the western, central and eastern Pacific did not detect any additional loci than those reported here (T.F. Duda, unpublished data). Also, expression patterns were largely consistent among individuals of the six species we examined. Thus, although the consistency of expression patterns of individuals of the species we examined supports our interpretations, we cannot rule out the possibilities that loci we interpreted to be silenced are actually expressed during particular circumstances or conotoxin expression is phenotypically plastic.

Coordinated expression patterns among species

The most parsimonious explanation for the observed conotoxin expression patterns is that several sets of putative orthologous loci exhibit coordinated expression. Although 11 loci appear to be uniquely expressed by only one of the six species examined, 12 sets of putative orthologous loci appear to be expressed by two or more species and 10 of these sets show coordinated expression patterns among species. *Conus abbreviatus* and *C. miliaris* apparently co-express eight loci and four of these loci also appear to be co-expressed by *C. coronatus*; none of the other three species express these loci. While similar expression patterns of *C. abbreviatus* and *C. miliaris* may be related to the phylogenetic affinity of these species, *C. aristophanes* is more recently separated from *C. abbreviatus* and yet does not appear to express any loci in common with *C. abbreviatus* or *C. miliaris*. Furthermore, *C. aristophanes* and *C. judaeus* apparently express two loci in common and neither of these loci appear to be expressed by *C. abbreviatus* (i.e. the sister of *C. aristophanes*) or the other species examined. These apparently nonrandom expression patterns suggest that sets of loci were reactivated or silenced/lost in concert and not independently.

Coordinated expression of loci may occur because genes are linked to a common regulatory region, expression is controlled by the same regulatory factors or the gene products of these loci act synergistically and selection favours their coordinated expression. The two sets of loci that appear to be co-expressed by *C. aristophanes* and *C. judaeus* cluster together uniquely in the conotoxin phylogram (Fig. 2). This relationship suggests that these two sets of loci share a common ancestral locus and arose from a duplication event in the common ancestor of these species. If these genes are arranged in tandem, they could reside near a single promoter region that regulates transcription of both loci. Lack of expression of these genes in some species too could be explained by deletion of the region that contains these genes. However, the relationships of the sets of loci that are presumably co-expressed by *C. abbreviatus*, *C. coronatus* and *C. miliaris* do not fit this pattern and these loci presumably arose from unique duplication events. For example, locus viii that appears to be co-expressed by *C. abbreviatus*, *C. coronatus* and *C. miliaris* clusters with locus ix that is co-expressed by *C. abbreviatus* and *C. miliaris* (Fig. 2). The relationships of these loci imply that a single duplication event in the common ancestor of these species gave rise to these genes, but these loci do not exhibit coordinated expression.

To explain the considerable diversity of venom components of *Conus*, Olivera and co-workers (1991) proposed that multiple diverse conotoxin molecules are required to effectively block particular receptor targets of prey. According to their '*Conus* toxin macrosite' model, cell-surface receptor macrosites include a site for normal ligand binding as well as a number of 'microsites'. Different conotoxin peptides preferentially bind to different microsites and so receptors are most effectively blocked by multiple sets of conotoxin peptides. *Conus* species may therefore exhibit coordinated expression of particular conotoxin loci because the gene products of these locus sets work in concert to block particular receptors. On the contrary, different conotoxins may act synergistically by targeting different receptors in prey and complete neuromuscular paralysis may only be possible with particular combinations of conotoxins (Terlau *et al.* 1996). For example, several studies of snake venom cytotoxins and phospholipases have shown that activities of cytotoxins are boosted when they occur with phospholipase (Condrea *et al.* 1964; Louw & Visser 1978; Harvey *et al.* 1983; Bougis *et al.* 1987; Chaim-Matyas *et al.* 1995). Studies of conotoxin gene regulation and analyses of the functions of their gene pro-ducts in prey will illuminate the factors responsible for the observed coordinated expression patterns.

References

Bougis PE, Marchot P, Rochat H (1987) In vivo synergy of cardiotoxin and phospholipase A2 from the elapid snake *Naja mossambica mossambica*. *Toxicon*, **25**, 427–431.

Chaim-Matyas A, Borkow G, Ovadia M (1995) Synergism between cytotoxin P4 from the snake venom of *Naja nigricollis nigricollis* and various phospholipases. *Comparative Biochemistry and Physiology — Part B: Biochemistry and Molecular Biology*, **110**, 83–89.

Chain FJJ, Evans BJ (2006) Multiple mechanisms promote the retained expression of gene duplicates in the tetraploid frog *Xenopus laevis*. *PLoS Genetics*, **2** (6), 478–490.

Condrea E, De Vries A, Mager J (1964) Hemolysis and splitting of human erythrocyte phospholipids by snake venom. *Biochimica et Biophysica Acta*, **84**, 60–73.

Conticello SG, Gilad Y, Avidan N, Ben-Asher E, Levy Z, Fainzilber M (2001) Mechanisms for evolving hypervariability: the case of conopeptides. *Molecular Biology and Evolution*, **18**, 120–131.

Derome N, Bernatchez L (2006) The transcriptomics of ecological convergence between 2 limnetic coregonine fishes (Salmonidae). *Molecular Biology and Evolution*, **23**, 2370–2378.

Duda TF Jr, Kohn AJ (2005) Species-level phylogeography and evolutionary history of the hyperdiverse marine gastropod genus *Conus*. *Molecular Phylogenetics and Evolution*, **34**, 257–272.

Duda TF Jr, Kohn AJ, Palumbi SR (2001) Origins of diverse feeding ecologies within *Conus*, a genus of venomous marine gastropods. *Biological Journal of the Linnean Society*, **73**, 391–409.

Duda TF Jr, Palumbi SR (1999a) Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proceedings of the National Academy of Sciences, USA*, **96**, 6820–6823.

Duda TF Jr, Palumbi SR (1999b) Developmental shifts and species selection in gastropods. *Proceedings of the National Academy of Sciences, USA*, **96**, 10272–10277.

Duda TF Jr, Palumbi SR (2000) Evolutionary diversification of multigene families: allelic selection of toxins in predatory cone snails. *Molecular Biology and Evolution*, **17**, 1286–1293.

Duda TF Jr, Palumbi SR (2004) Gene expression and feeding ecology: evolution of piscivory in the venomous gastropod genus *Conus*. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 1165–1174.

Duda TF Jr, Rolán E (2005) Explosive radiation of Cape Verde *Conus*, a marine species flock. *Molecular Ecology*, **14**, 267–272.

Flessa KW, Jablonski D (1983) Extinction is here to stay. *Paleobiology*, **9** (4), 315–321.

Giger T, Excoffier L, Day PJR *et al.* (2006) Life history shapes gene expression in salmonids. *Current Biology*, **16** (8), R281–R282.

Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP (2006) Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature*, **440**, 242–245.

Harvey AL, Hider RC, Khader F (1983) Effect of phospholipase A on actions of cobra venom cardiotoxins on erythrocytes and skeletal muscle. *Biochimica et Biophysica Acta*, **728**, 215–221.

Jones A, Bingham JP, Gehrmann J *et al*. (1995) Isolation and characterization of conopeptides by high-performance liquid chromatography combined with mass spectrometry and tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, **10**, 138–143.

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Kohn AJ (1959) The ecology of *Conus* in Hawaii. *Ecological Monographs*, **29** (1), 47–90.

Kohn AJ (1966) Food specialization in *Conus* in Hawaii and California. *Ecology*, **47**, 1041–1043.

Kohn AJ (1980) Abundance, diversity, and resource use in an assemblage of *Conus* species in Enewetak Lagoon. *Pacific Science*, **34** (4), 359–369.

Kohn AJ (2001) Maximal species richness in *Conus*: diversity, diet and habitat on reefs of northeast Papua New Guinea. *Coral Reefs*, **20**, 25–38.

Kohn AJ, Almasi KN (1993) Comparative ecology of a biogeographically heterogeneous *Conus* assemblage. In: *Proceedings of the Fifth International Marine Biological Workshop: the Marine Flora and Fauna of Rottnest Island, Western Australia* (eds Wells FE, Walker DI, Kirkman H, Lethbridge R), pp. 509–521. Western Australia Museum, Perth, Western Australia.

Kohn AJ, Nybakken JW (1975) Ecology of *Conus* on eastern Indian Ocean fringing reefs: diversity of species and resource utilization. *Marine Biology*, **29**, 211–234.

Kohn AJ, Orians GH (1962) Ecological data and the classification of closely related species. *Systematic Zoology*, **11**, 119–127.

Koonin EV (2005) Orthologs, paralogs, and evolutionary genomics. *Annual Review in Genetics*, **39**, 309–338.

Leviten PJ (1980) The foraging strategy of vermivorous conid gastropods. *Ecological Monographs*, **46**, 157–178.

Locascio A, Manzanares M, Blanco MJ, Nieto MA (2002) Modularity and reshuffling of *Snail* and *Slug* expression during vertebrate evolution. *Proceedings of the National Academy of Sciences, USA*, **99** (26), 16841–16846.

Louw AI, Visser RC (1978) The synergism of cardiotoxins and phospholipase A2. *Biochimica et Biophysica Acta*, **512**, 163–171.

Maddison WP, Maddison DR (1992) MACCLADE version 3: Analysis of Phylogeny and Character Evolution. 398 pp (book) + 900K (Computer Program). Sinauer & Associates, Sunderland, Massachusetts.

Marsh H (1971) Observations on the food and feeding of some vermivorous *Conus* on the Great Barrier Reef. *Veliger*, **14** (1), 45–53.

Marshall CR, Raff EC, Raff RA (1994) Dollo's law and the death and resurrection of genes. *Proceedings of the National Academy of Sciences, USA*, **91**, 12283–12287.

Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences, USA*, **94**, 7799–7806.

Nybakken J, Perron F (1988) Ontogenetic change in the radula of *Conus magus* (Gastropoda). *Marine Biology*, **98**, 239–242.

Olivera BM, Rivier J, Scott JK, Hillyard DR, Cruz LJ (1991) Minireview: conotoxins. *Journal of Biological Chemistry*, **266** (33), 22067–22070.

Olivera BM, Walker C, Cartier GE *et al*. (1999) Speciation of cones snails and interspecific hyperdivergence of their venom peptides. *Annals of the New York Academy of Sciences*, **870**, 223–237.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Reichelt RE, Kohn AJ (1985) Feeding and distribution of predatory gastropods on some Great Barrier Reef platforms. *Proceedings of the Fifth International Coral Reef Congress*, **5**, 191–196.

Shiu SH, Byrnes JK, Pan R, Zhang P, Li WH (2006) Role of positive selection in the retention of duplicate genes in mammalian genomes. *Proceedings of the National Academy of Sciences, USA*, **103**, 2232–2236.

Starbäck P, Wraith A, Eriksson H, Larhammar D (2000) Neuropeptide Y receptor gene y6: multiple deaths or resurrections? *Biochemical and Biophysical Research Communications*, **277**, 264–269.

Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods*), Version 4. Sinauer & Associates, Sunderland, Massachusetts.

Terlau H, Shon K-J, Grilley M, Stocker M, Stühmer W, Olivera BM (1996) Strategy for rapid immobilization of prey by a fish-hunting marine snail. *Nature*, **381**, 148–151.

Whitehead A, Crawford DL (2006a) Variation within and among species in gene expression: raw material for evolution. *Molecular Ecology*, **15**, 1197–1211.

Whitehead A, Crawford DL (2006b) Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences, USA*, **103** (14), 5425–5430.

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences*, **13**, 555–556.

Tom Duda is an Assistant Professor in the Department of Ecology and Evolutionary Biology and Assistant Curator of molluscs in the UMMZ as well as a Research Associate at the Smithsonian Tropical Research Institute. Dr. Duda uses molecular evolutionary analyses to investigate the evolution of *Conus* venoms and the relationship between conotoxin gene family evolution and the origins of feeding specialisations in this group. Dr. Duda is also interested in the diversification of taxa, particularly marine molluscs, and uses phylogenetic and population genetic approaches to examine the evolutionary history and biogeography of these taxa. Ed Remigio has worked on several research projects concerning the evolution of *Conus* as a postdoctoral research associate in the Mollusk Division of the University of Michigan Museum of Zoology (UMMZ). His research interests include the application of molecular genetic analyses for phylogenetic inference, systematics, and conservation biology of gastropods, as well as other invertebrates.

Supplementary material

The following supplementary material is available for this article:

Fig. S1 Interleaved alignment of four-loop conotoxin transcript sequences of *Conus abbreviatus*, *C. aristophanes*, *C. coronatus*, *C. ebraeus*, *C. judaeus* and *C. miliaris*. Loci are identified as follows: A, *C. abbreviatus*; Ar, *C. aristophanes*; C, *C. coronatus*; E, *C. ebraeus*; J, *C. judaeus;* and M, *C. miliaris*; arbitrary numerical identifiers are given for each putative locus. Sequences are arranged according to their position in the tree (Fig. 2). Regions of the transcripts are colour coded as follows: yellow, prepro coding region; red/turquoise, toxin coding region (cysteine codons are red); stop codon, dark yellow; 3′ untranslated region, gray.

Fig. S2 Interleaved alignment of predicted amino acid sequences of four-loop conotoxin transcript sequences of *Conus abbreviatus*, *C. aristophanes*, *C. coronatus*, *C. ebraeus*, *C. judaeus* and *C. miliaris*.

Loci are identified as in Fig. S1. Sequences are arranged according to their position in the tree (Fig. 2). Regions of the peptide are colour coded as follows: yellow, prepro coding region; red/turquoise, toxin coding region (cysteine codons are red); stop codon, dark yellow. Amino acids following the terminal cysteine residue of the toxin coding region are given as they occur and are not based on alignments of nucleotide sequences because of the presence of insertions/deletions that affect reading frame (see Fig. S1).

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/
10.1111/j.1365-294X.2008.03804.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.