Divergence of the Venom Exogene Repertoire in Two Sister Species of Turriconus

Qing Li1, Neda Barghi2,3, Aiping Lu4, Alexander E. Fedosov5, Pradip K. Bandyopadhyay6, Arturo O. Lluisma2,7, Gisela P. Concepcion2,7, Mark Yandell1,8, Baldomero M. Olivera6, and Helena Safavi-Hemami6,*

1Eccles Institute of Human Genetics, University of Utah
2Marine Science Institute, University of the Philippines-Diliman, Quezon City, Philippines
3Institute für Populationsgenetik, Vetmeduni, Vienna, 1210, Austria
4School of Life Sciences and Technology, Tongji University, Shanghai, China
5A.N. Severtsov Institute of Ecology and Evolution, Russian Academy of Science, Moscow, Russia
6Department of Biology, University of Utah
7Philippine Genome Center, University of the Philippines, Quezon City, Philippines
8USTAR Center for Genetic Discovery, University of Utah

*Corresponding author: E-mail: helena.safavi@utah.edu.

Accepted: August 22, 2017

Data deposition: This project has been deposited at GenBank under the accessions MF576542 - MF576988

Abstract

The genus Conus comprises approximately 700 species of venomous marine cone snails that are highly efficient predators of worms, snails, and fish. In evolutionary terms, cone snails are relatively young with the earliest fossil records occurring in the Lower Eocene, 55 Ma. The rapid radiation of cone snail species has been accompanied by remarkably high rates of toxin diversification. To shed light on the molecular mechanisms that accompany speciation, we investigated the toxin repertoire of two sister species, Conus andremenezi and Conus praecellens, that were until recently considered a single variable species. A total of 196 and 250 toxin sequences were identified in the venom gland transcriptomes of C. andremenezi and C. praecellens belonging to 25 and 29 putative toxin gene superfamilies, respectively. Comparative analysis with closely (Conus tribblei and Conus lenavati) and more distantly related species (Conus geographus) suggests that speciation is associated with significant diversification of individual toxin genes (exogenes) whereas the expression pattern of toxin gene superfamilies within lineages remains largely conserved. Thus, changes within individual toxin sequences can serve as a sensitive indicator for recent speciation whereas changes in the expression pattern of gene superfamilies are likely to reflect more dramatic differences in a species’ interaction with its prey, predators, and competitors.

Key words: venom evolution, speciation, conotoxins, exogenes.

Introduction

Exogenes are the subset of genes that directly mediate the biotic interactions between organisms (Olivera 2006). They are sensitively tuned to the particular ecological niche for which a given species is maximally fit. The genes that encode the components of animal venoms are an example of exogenes. When, for example, specification is associated with a change in prey preference, genes encoding venom components may experience strong selection pressure for optimal efficacy in the new prey (Olivera 2006; Chang and Duda 2016). A recent study of venom insulins has highlighted how exogene products, targeted at other organisms, appear to evolve very differently from their endogenous counterparts (Safavi-Hemami et al. 2016). In this case, exogenes were characterized by high diversity of sequence variants, thought to reflect duplication events coupled with strong positive selection. In contrast, their endogenous counterparts, due to strong purifying selection, were essentially...
identical between distantly related species. Thus, exogenes can be sensitive indicators of hidden biological differences that are not obvious from morphology or endogene sequence analyses, and offer insight into the molecular events that accompany speciation.

The venom components of cone snails (genus Conus) are one of the best-studied examples of exogenes. Conus is a large genus (~730 extant species according to WoRMS—accessed on September 28, 2016) of predatory marine mollusks that use venom for prey capture, defense, and competitive interactions (Rockel et al. 1995; Puillandre et al. 2014). According to the recent dated phylogeny (Uribe et al. 2017), the first rapid diversification of the genus Conus started in Upper Oligocene about 30–25 Ma, and was followed by a mass extinction at the Miocene–Pliocene boundary. Thus, Conus is a relatively young genus, with most extant species originated less than 10 Ma (Duda et al. 2001; Duda and Kohn 2005; Uribe et al. 2017). Several major shifts in prey preference occurred in the evolution of this genus and living cone snail species can typically be divided into those that prey on vermiform invertebrates, gastropods, or fish (Duda et al. 2001).

The venom of each species is a complex mixture of approximately 70–400 peptide toxins (Hu et al. 2012; Barghi et al. 2015a, 2015b; Phuong et al. 2016) (named conotoxins or conopeptides). Conotoxins are initially translated as precursor peptides in the epithelial cells of a specialized venom gland. Most conotoxin precursors have a well-defined primary structure: At the N-terminus, a highly conserved signal sequence, at the C-terminus, a variable mature toxin region and in between, a propeptide region (Woodward et al. 1990). The mature toxin region encodes the actual bioactive peptide injected by the cone snail into another animal. According to signal sequence conotoxins can be grouped into probably 40–50 gene superfamilies (Robinson and Norton 2014). Several of these are found in nearly every species examined and exhibit the characteristics typical of many other venom exogenes: Extraordinarily high diversity within and between species. Consistent with this observation, considerable prior work has demonstrated that conotoxins are hyperdiverse, with each species of Conus suggested to have its own distinct venom exogene repertoire, presumably shaped by species-specific interactions with prey, predators, and competitors (Olivera et al. 2014).

Conus andremenezi is a recently described cone snail (Biggs et al. 2010), and was chosen as one of “100 of our Planets Most Amazing New Species” (Wheeler et al. 2013). Conus andremenezi belongs to a group of small forms of Conus that were conventionally regarded by taxonomists as a single variable species Conus praecellens, which has recently been demonstrated to constitute a species complex (Biggs et al. 2010) (fig. 1A). Conus andremenezi and C. praecellens are distinguishable by several morphological criteria, as well as by primary sequence of mitochondrial DNA (Biggs et al. 2010). A rare aquarium recording of C. andremenezi and C. praecellens shows differences in foot and siphon coloration, potentially pointing toward different camouflage strategies (fig. 18).

The distributions of the two species overlap, in the Central Indo Pacific (Philippines, Papua New Guinea, and Solomon Islands); however, C. andremenezi is sometimes collected at greater depths than C. praecellens, based on material from several field expeditions carried out by the Museum of Natural History (Paris) (Olivera B, Bouchet P, unpublished observations).

According to a recently published classification of cone snails (Puillandre et al. 2014), C. praecellens and C. andremenezi belong to the subgenus Turriconus, one of the early diverging lineages of the genus Conus (fig. 1A). As most of the Turriconus species, including C. andremenezi and C. praecellens, occur only in relatively deep water, direct behavioral observations are difficult to obtain, and little is known about their biology, although all early diverging clades of Conus are believed to be worm-hunting (Duda et al. 2001) and phylogenetic analysis strongly suggests that fish- and mollusk-hunting behaviors only evolved in certain more recently diverging groups of Conus (Uribe et al. 2017). Additionally, radular teeth of Turriconus bear strong barbs with no accessory process (Tucker and Tenorio 2009); such morphology is common for vermivorous Conus species, but was not observed in either fish or mollusk hunters (Olivera et al. 2015). Thus, although the possibility that members of the Turriconus clade are specialists on other prey cannot be eliminated it appears highly likely that these species are vermivorous.

In this study, we analyzed the venom composition of C. andremenezi and C. praecellens, neither of which has previously been comprehensively investigated at a molecular level. Our results provide new insight into the biology of these species and of the subgenus Turriconus in general. Moreover, comparison of the venom exogene repertoires of C. praecellens and C. andremenezi has provided insights into the genetic mechanisms associated with recent speciation. Our findings show that the expression pattern of toxin gene families remains largely conserved between these sister species, but that speciation is associated with significant diversification of individual toxin genes.

Materials and Methods

Specimen Collection, RNA Extraction, and Sequencing

All studied specimens were collected in the central Philippines. Adult specimens of C. andremenezi and C. praecellens were collected at depth about 180–250 m, off Sagod, North of Cebu Island, and the specimens of C. geographus—from the shallow water (10–25 m) off Caw-Oy, Olangano Island. Specimen identification was initially performed by morphological examination and later verified by sequence analysis of the cytochrome oxidase c subunit 1 (COI) gene (as described under phylogenetic analysis below). Venom glands of live
specimens were dissected and stored in RNAlater at −80 °C. Total RNA was isolated from the homogenized venom glands of each species using TRizol reagent following the manufacturer’s protocol (Life Technologies Corporation). Integrity of RNA was verified on a Bioanalyzer instrument (Agilent Technologies), and Illumina libraries were prepared with a mean insert size of 170 bp at Cofactor Genomics (St. Louis, MO) for C. andremenezi (specimen 1), C. praecellens (specimen 1), and C. geographus. For C. praecellens (specimen 2), strand-specific Illumina libraries with a mean insert size of 220 bp were prepared at the University of Utah High-Throughput and Genomics Core Facility. For C. andremenezi (specimen 2), mRNA was extracted from the venom gland using Dynabeads mRNA DIRECT kit (Invitrogen Dynal AS, Oslo, Norway) and a cDNA library was constructed with an average insert size of 200 bp. All libraries were sequenced on an Illumina HiSeq 2000 platform. Prior to assembly, quality filtering and adaptor trimming were performed using Fqtrim software (version 0.9.4, http://ccb.jhu.edu/software/fqtrim/) and Prinseq (version 0.20.4) (Schmieder and Edwards 2011). After preprocessing, reads shorter than 70 bp and those containing more than 5% ambiguous bases (Ns) were discarded. 58, 834, 536 (specimen 1) and 54, 177, 324 (specimen 2) reads for C. andremenezi; 70, 826, 570 (specimen 1) and 32, 695, 570 (specimen 2) reads for C. praecellens; 146, 630, 264 reads for C. geographus were used for each transcriptome assembly.

**Phylogenetic Analysis**

Phylogenetic analysis was performed using a partial sequence of the COI gene segment of mitochondrial DNA, corresponding to the standard barcode region. COI nucleotide sequences for specimens used in this study were retrieved from the assembled venom gland transcriptomes, and those from additional members of the *Turriconus* subgenus were obtained from GenBank (KJ550384, KJ550243, KJ550048, KJ549906, KJ549742, KJ550110, KJ550109, KJ550108, KJ550107, KJ550106, KJ550105, KJ550104, KJ550103, KJ550102, KJ550101, KJ549855, KJ550425, KJ549743, KJ550125, KJ549858, KJ550424, KJ550423, KJ550124, KJ550123).

Phylogenetic analyses were performed using MrBayes v.3.2.6 (Huelsenbeck and Ronquist 2001), running two parallel analyses, consisting each of six Markov chains of 15,000,000 generations with a sampling frequency of one tree each 1,000 generations, and chain temperature set to 0.02. Parameters of the substitution model were estimated during the analysis (six substitution categories, a gamma-distributed rate variation across sites, and a proportion of variable sites approximated in four discrete categories). Each codon position of the COI gene was treated as a separate partition. The first 25% of trees were omitted as burn-in prior to the construction of consensus trees. Convergence of the analysis was evaluated using Tracer 1.4.1 (http://tree.bio.ed.ac.uk/software/tracer/) to ensure that effective sample size values were all greater than 200 (default burning).
Bootstrap branch support values were obtained for an equivalent maximum-likelihood tree estimated by RAxML (Stamatakis 2006) with each codon position treated as an independent partition. Nodal support of the observed tree was estimated using the thorough bootstrapping algorithm (Felsenstein 1985) with 1,000 iterations. All phylogenetic analyses were carried out at Cipres Science Gateway (Miller et al. 2010) Kimura 2-parameter (K2P) genetic distances were calculated with MEGA 5.2.1 (Tamura et al. 2011). Conus distans, the species recovered in a sister position to all other Conus lineages by Puillandre et al. (2014), was used as an outgroup in all phylogenetic analyses.

Transcriptome Assembly, Annotation, and Postprocessing
Venom gland transcriptomes for C. andremenezi (both specimens) and C. praecellens (specimen 1) and C. geographus were assembled with Trinity (version 2.0.5) using the paired-end mode, wherein kmer size used to build the De Bruijn Graphs was 32, minimum percent identity for two paths to be merged into single paths was 99, maximum allowed difference encountered between path sequences to combine them was 1, maximum internal gap length allowed between path sequences to combine them was 3, and the minimum output contig length was 75 (Grabherr et al. 2011). Trinity settings were optimized for sequence read length and library type according to simulation results; RNAseq reads were simulated using the published Conus victoriae venom gland transcriptome (Robinson et al. 2014) with the same read length and quality obtained for RNAseq data sets sequenced in this study. Systematic, serial modifications of the assembly parameters and batch assemblies were performed to optimize assemblies. The true positive rate, defined as the percentage of C. victoriae toxins that were correctly assembled, was used as the criteria of optimality. Assembly parameters resulting in the highest true positive rate were used. A bash script for this procedure is available online (git@github.com/qingl0331/turri_conus.git). Because the second specimen of C. praecellens (specimen 2) was sequenced with longer (125 nt instead of 101 nt) and stranded reads, different assembly parameters were used in Trinity (version 2.0.5): Paired-end mode, wherein the kmer size used to build De Bruijn Graphs was 31, sequence library type was set to reverse forward, minimum kmer coverage was 10, and the minimum glue was 10. Transcriptome assemblies were annotated using NCBI-BLASTX against a combined ConoServer (Kaas et al. 2012) and UniProtKB database (release April 2015). Recently published conotoxin sequences that were not yet uploaded into UniProt at the time of analysis (Hu et al. 2012; Barghi et al. 2015a, 2015b; Phuong et al. 2016) were also added to our in-house database. Contigs with significant similarity (e-value $1 \times 10^{-4}$) to known conotoxins were designated as putative conotoxin sequences. Contigs with no blast annotation were translated into six frames and predicted open reading frames that contained a start and stop codon and a total length of 50–200 amino acids were selected for further analysis. Sequences that did not have a BlastX hit but fulfilled all of the following criteria were designated as putative conotoxins: Contained an N-terminal signal sequence (predicted using SignalP [Petersen et al. 2011]), were abundantly expressed in at least one specimen examined in this study (>10,000 mapped reads), and shared high similarity with a sequence expressed in another specimen examined in this study (BlastP e-value $< 1 \times 10^{-10}$). The rational for implementing these filtering criteria is that a putative new toxin gene family should not be described based on a single putative toxin sequence. Identification of a similar but distinct sequence in another specimen and/or species increases the likelihood of a sequence to be indeed a member of a novel toxin gene family.

All specimens sequenced in this study were multiplexed with other cone snail species on the same Illumina flowcell (see supplementary table 1, Supplementary Material online, for number of species multiplexed per lane) and examined for cross-contamination resulting from index misassignment (Kircher et al. 2012; Sloan et al. 2013). Conotoxin sequences with low abundance in one species (ratio of single read count/total read count $< 10^{-5}$) but highly similar (>95% sequence identity) to a high abundant one in another species (200-fold higher) were considered as cross-contamination resulting from index misassignment. This ensured correct sequence assignments, however, could have resulted in the removal of a small number of genuine sequences that are present at very low expression levels (and are therefore unlikely to play a significant role in envenomation). Next, any erroneous duplicate transcripts resulting from the assembly process were removed. Finally, a subset of raw reads was remapped to the conotoxin data set using the map-to-reference tool in the Geneious package (version 8.1.7 [Kearse et al. 2012]), in order to identify variants overlooked by the assembler and to remove misassembled contigs.

Read Counts for Expression Analysis
To determine approximate relative expression levels, reads were mapped back to the final conotoxin data set with Bowtie2 (Langmead and Salzberg 2012). Mapped reads that were shared between $>2$ sequences were divided equally. For comparative analysis with members of the Splinoconus clade, Illumina reads for the venom glands of Conus tribblei and Conus lenavati were obtained from GenBank (SRR1803937, SRR1803938, SRR1803939, SRR1803940, SRR1803941, SRR1803942) and mapped back to conotoxin sequences expressed in these two species (Barghi et al. 2015a) with Bowtie2 as described above.

Comparative Analysis
All-by-all NCBI-BLASTN searches were carried out to identify reciprocal best-hit (rbh) pairs among the inter-intraspecific
conotoxin nucleotide sequences. The rbh pairs were compared using NCBI-BLASTX to our in-house conotoxin database in order to classify rbh pairs as either conotoxin or nonconotoxin rbhs. ClustalW alignment was then used to globally align the rbh pairs (Larkin et al. 2007). The alignment results were parsed to obtain the percent identity of each aligned rbh pair. The program codeml from PAML 4.7 (Yang 2007) was used to estimate synonymous and nonsynonymous substitutions with model = 2 (fix omega), over all codon positions in an alignment of rbh pairs of both the conotoxin and nonconotoxin sequences. Analyses were conducted in runmode = -2. Rbh pairs with branches showing dS < 0.01 were discarded as inaccurate dN/dS estimation. Next, the average dN/dS ratio was calculated by averaging the values from all pairwise comparisons of toxin and nontoxin sequences.

Principal Component Analysis

Principal component analysis (PCA) was performed on relative expression levels of conotoxin superfamilies (as determined with Bowtie2 [Langmead and Salzberg 2012]) in each of the specimens analyzed. Expression levels were normalized to total read counts after trimming. Principal components were calculated using the pcomp package in R (R Core Team 2013).

Naming of Conotoxin Precursors

Conotoxin precursor sequences identified in C. andremenezi (prefix Amz) and C. praecellens (prefix Ps) were named according to the conventional conotoxin nomenclature (Walker et al. 1999) (where the species is represented by one to three letters, cysteine framework by an Arabic numeral and, following a decimal, order of discovery by a second numeral), with slight modifications. The gene superfamily was provided as a prefix (Robinson et al. 2014), and sequences that likely represent allelic variants rather than being encoded at distinct loci (i.e., only differ by 1–3 nucleotides) were given the same name but a small numeral was added as a suffix. Putative new superfamilies were named by the first five amino acids of their predicted signal peptides (e.g., putative MWSGK or put.MWSGK).

Data Availability

All C. andremenezi and C. praecellens conotoxin prepropeptide sequences from this study are provided in the Supplementary Material online and have been deposited at GenBank (accession numbers MF576542–MF576988).

Results

Phylogenetic Analysis of Turriconus Resolves C. andremenezi and C. praecellens as Two Sister Species

The Turriconus segment of the COI tree (highlighted in red) is characterized by the overall high posterior probability (PP) and bootstrap (B) support values, and distinctive well-supported terminal clades (PP value 0.99 or 1; fig. 2) corresponding to Conus acutangulus, Conus excelsus, C. praecellens, and C. andremenezi; Conus mininexclusus is represented by a single sequence. Conus praecellens and C. andremenezi show sister relationship, with the grouping of C. andremenezi–C. praecellens being highly supported in both Bayesian analysis (PP = 0.99) and RaxML (B = 92). Conus mininexclusus forms a sister group to C. andremenezi–C. praecellens, and the resulting three-species clade is also well supported (PP = 1, B = 99). In the Splinoconus subtree (PP = 1, B = 100; shown in green), the well-supported clad corresponding to C. tribblei shows sister relationship to the C. queenslandis with a moderate support (PP = 0.96, B = 88), and the highly supported C. lenavati clad is a sister group to the C. tribblei–C. queenslandis clad.

The mean K2P genetic distance among specimens of C. andremenezi is 1.9% and among specimens of C. praecellens is 0.9%. The mean K2P genetic distance between specimens of C. praecellens and C. andremenezi is 5.4%. Within-group K2P genetic distance in the analyzed data set ranges from 0% to 3.4%, whereas the minimal K2P value for between-group comparison is 4.8%. This difference corresponds to the so-called “barcode gap” marking the difference in within- and among-species comparisons, and is overall consistent with within- and among-species genetic distances in other groups of Conoidea (Puillandre et al. 2009; Fedosov and Puillandre 2012).

Similarities in Toxin Number and Superfamily Expression Patterns within the Turriconus Clade

We analyzed venom component transcripts from five venom gland transcriptomes: Two specimens of C. andremenezi and C. praecellens, from the Turriconus subgenus, and one specimen of C. geographus from the distantly related subgenus Gastridium that served as a reference. An in-depth analysis of the venom components of C. geographus has been provided elsewhere (Hu et al. 2012; Dutertre et al. 2014; Safavi-Hemami et al. 2014). Information on RNA sequencing and assembly statistics for these transcriptomes is provided in supplementary table 1, Supplementary Material online. Additionally, published data for two species of the Splinoconus clade (C. tribblei and C. lenavati) were interrogated. Comparison between the two specimens of the same species within Turriconus allowed us to estimate intra-specific variation in venom composition, whereas comparison between C. andremenezi and C. praecellens made it possible
to examine interspecific variation between these two sister species. Finally, comparison of \textit{C. andremenezi} and \textit{C. praecellens} to \textit{C. tribblei} and \textit{C. lenavati} from the closely related subgenus \textit{Splinoconus} (Barghi et al. 2015a) and \textit{C. geographus}, a distantly related cone snail of the \textit{Gastriidium} clade with distinct prey preferences (\textit{C. geographus} feeds on fish), provided relative scale by which to measure the magnitude of interspecific variation in \textit{Turriconus}.

A number of general similarities exist across all transcriptomes examined (fig. 3): Each species’ venom repertoire is dominated by less than five conotoxin superfamilies, which typically account for over 70% of all of the sequences found. Multiple members of some superfamilies were observed, whereas other superfamilies were represented by a single toxin in the transcriptome of a species (table 1, supplementary table 2, Supplementary Material online).

In the venom gland transcriptome of \textit{C. andremenezi} (specimen 1), a total of 128 distinct conotoxin transcripts were identified. These could be grouped into 24 conotoxin gene superfamilies. Nevertheless, the transcriptome was...
dominated by only a few conotoxin gene superfamilies, namely the O1δ-like- (O1d), T-, M-, and P-superfamilies (in descending order of relative expression level) in terms of both expression and number of isoforms (fig. 3, table 1). Analysis of O1-superfamily sequences revealed two distinct gene classes within the O1 family, one of which shares high sequence similarity with so-called δ-conotoxins (Heinemann and Leipold 2007). Thus, this class was named the O1d superfamily whereas all other O1 sequences were grouped into the O1 superfamily.

The venom gland transcriptome of a second specimen of C. andremenezi closely mirrored that of the first specimen: A total of 107 distinct conotoxin transcripts were identified belonging to 23 conotoxin superfamilies. High similarities were also observed in the expression of the most abundant toxin superfamilies: The O1d was the most highly expressed
superfamily followed by the T-, P-, and M-superfamily (in descending order of relative expression level) (fig. 3, table 1). Together the two specimens of *C. andremenezi* expressed 196 distinct sequences of which 77 were likely to be allelic variants (defined as differing by 1–3 nucleotides).

A total of 155 and 149 distinct conotoxin transcripts representative of 28 and 29 conotoxin gene superfamilies were identified in the venom gland transcriptomes of *C. praecellens* specimens 1 and 2, respectively. The majority of the identified transcripts represented the same conotoxin superfamilies expressed in *C. andremenezi*: O1d-, T-, P-, and M-superfamilies (fig. 3, table 1). Notably, however, the relative expression ranks of these superfamilies are different: Although the P-superfamily is most highly expressed in *C. praecellens*, the O1d superfam-

---

**Table 1**

<table>
<thead>
<tr>
<th>Conus andremenezi 1</th>
<th>Conus andremenezi 2</th>
<th>Conus praecellens 1</th>
<th>Conus praecellens 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Superfamily</strong></td>
<td><strong>Relative Expression Level</strong></td>
<td><strong>Superfamily</strong></td>
<td><strong>Relative Expression Level</strong></td>
</tr>
<tr>
<td>O1d</td>
<td>22.8% (30)</td>
<td>O1d</td>
<td>36.8% (3)</td>
</tr>
<tr>
<td>T</td>
<td>19.8% (4)</td>
<td>T</td>
<td>13.9% (18)</td>
</tr>
<tr>
<td>M</td>
<td>14.3% (18)</td>
<td>P</td>
<td>11.6% (9)</td>
</tr>
<tr>
<td>P</td>
<td>9.9% (18)</td>
<td>M</td>
<td>8.1% (14)</td>
</tr>
<tr>
<td>O1</td>
<td>8% (11)</td>
<td>O1</td>
<td>8.1% (13)</td>
</tr>
<tr>
<td>B</td>
<td>4.9% (4)</td>
<td>S</td>
<td>4.9% (2)</td>
</tr>
<tr>
<td>S</td>
<td>3.9% (1)</td>
<td>I3</td>
<td>4% (2)</td>
</tr>
<tr>
<td>O2</td>
<td>3.4% (10)</td>
<td>I2</td>
<td>3.2% (10)</td>
</tr>
<tr>
<td>I1</td>
<td>2.9% (2)</td>
<td>I2</td>
<td>1.7% (2)</td>
</tr>
<tr>
<td>SF-mi1</td>
<td>2.7% (2)</td>
<td>I1</td>
<td>1.3% (1)</td>
</tr>
<tr>
<td>SF-01</td>
<td>2.6% (2)</td>
<td>V</td>
<td>0.5% (1)</td>
</tr>
<tr>
<td>put.MGGRF</td>
<td>2% (4)</td>
<td>put.MKAVA</td>
<td>0.3% (1)</td>
</tr>
<tr>
<td>I2</td>
<td>1.7% (1)</td>
<td>SF-mi1</td>
<td>0.3% (1)</td>
</tr>
<tr>
<td>V</td>
<td>1.5% (2)</td>
<td>put.MSGLR</td>
<td>0.3% (1)</td>
</tr>
<tr>
<td>PH4*</td>
<td>0.7% (1)</td>
<td>put.MWSGK</td>
<td>0.3% (1)</td>
</tr>
<tr>
<td>put.MWSGK</td>
<td>0.7% (1)</td>
<td>put.MGGRF</td>
<td>0.2% (1)</td>
</tr>
<tr>
<td>put.MKAVA</td>
<td>0.6% (2)</td>
<td>PH4*</td>
<td>0.1% (1)</td>
</tr>
<tr>
<td>B2</td>
<td>0.3% (2)</td>
<td>B2</td>
<td>0.03% (1)</td>
</tr>
<tr>
<td>I4</td>
<td>0.23% (2)</td>
<td>SF-01</td>
<td>0.01% (1)</td>
</tr>
<tr>
<td>put.MSGLR</td>
<td>0.2% (2)</td>
<td>Con-ikot-ikot</td>
<td>0.01% (2)</td>
</tr>
<tr>
<td>put.MMFLM</td>
<td>0.2% (2)</td>
<td>I4</td>
<td>0.01% (2)</td>
</tr>
<tr>
<td>Conopressina*</td>
<td>0.2% (1)</td>
<td>Conopressina*</td>
<td>0.009% (1)</td>
</tr>
<tr>
<td>B4</td>
<td>0.02% (1)</td>
<td>H</td>
<td>0.008% (1)</td>
</tr>
<tr>
<td>Conoporina*</td>
<td>0.02% (1)</td>
<td>A2</td>
<td>0.0004% (2)</td>
</tr>
<tr>
<td>A2</td>
<td>0.01% (3)</td>
<td>L</td>
<td>0.03% (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I2</td>
<td>0.02% (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A2</td>
<td>0.02% (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V</td>
<td>0.02% (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>0.02% (1)</td>
</tr>
</tbody>
</table>

**NOTE.**—Relative expression levels are provided for each superfamily, from highest to lowest expression. The number of individual sequences detected per superfamily is shown in parentheses.

*Sequences that may have derived from housekeeping proteins or serve in endogenous processes in the venom gland rather than in envenomation.

---

superfamily followed by the T-, P-, and M-superfamily (in descending order of relative expression level) (fig. 3, table 1). Together the two specimens of *C. andremenezi* expressed 196 distinct sequences of which 77 were likely to be allelic variants (defined as differing by 1–3 nucleotides).

A total of 155 and 149 distinct conotoxin transcripts representative of 28 and 29 conotoxin gene superfamilies were identified in the venom gland transcriptomes of *C. praecellens* specimens 1 and 2, respectively. The majority of the identified transcripts represented the same conotoxin superfamilies expressed in *C. andremenezi*: O1d-, T-, P-, and M-superfamilies (fig. 3, table 1). Notably, however, the relative expression ranks of these superfamilies are different: Although the P-superfamily is most highly expressed in *C. praecellens*, the O1d superfam-

Consistent with previous observations, conotoxin-encoding genes are under strong positive selection (average \( \omega = 1.243 \)) (Chang and Duda 2012; Sunagar et al. 2015) whereas other venom gland proteins remain largely conserved (average \( \omega = 0.182 \)). Pairwise dN/dS values are provided in supplementary table 3, Supplementary Material online. dN/dS ratios greatly vary between individual pairs of sequences within each superfamily, and almost all superfamilies contain sequences that appear to be under both, positive and purifying selection.

A side-by-side comparison with the venom composition of *C. tribblei* and *C. lenavati* (subgenus *Spilnoconus*), and with a distantly related fish-hunting species, *C. geographus* identified in *C. praecellens*, albeit at very low expression levels (table 1). Together the two specimens of *C. praecellens* expressed 250 distinct sequences of which 81 were likely to be allelic variants.
(subgenus Gastridium) (fig. 3), lends scale to these differences. The venom composition of these three species differs significantly from the two Turriconus members and from each other, with little overlap in the expression pattern of the most highly expressed toxin gene superfamilies (fig. 3). The total of 39 toxin gene superfamilies identified in C. tribblei and 40 in C. lenavati exceed the gene superfamily diversity in both C. andremenezi and C. praecellens. The two Splinoconus species are characterized by very high expression of the con-ikot-ikot and B2 superfamilies (fig. 3), which together account for 63% and 55% of the expressed venom toxins in C. tribblei and C. lenavati, respectively (supplementary table 2, Supplementary Material online). At the same time, only about 0.3% of identified toxins in C. andremenezi and C. praecellens represent the B2 superfamily, and the con-ikot-ikot superfamily was not detected in either Turriconus species. Similarly, the O1d superfamily, which is highly expressed in Turriconus species, was not identified in either, C. tribblei or C. lenavati. Nevertheless, the toxin gene superfamilies P-, M-, O1, and O2 are detected in venom gland transcriptomes of all four species, and although their relative expression levels differ notably between Turriconus and Splinoconus, the P-, M-, O1, and O2 are among the ten mostly highly expressed toxin gene superfamilies in all four species. Furthermore, the T-superfamily toxins that are not detected in the venom of C. tribblei are highly expressed in the venom glands of C. andremenezi and C. praecellens, as well, as in C. lenavati. The total contribution of the P-, M-, O1-, O2-, and T-toxin gene superfamilies ranges from 24% of all putative toxin transcripts in C. lenavati to as much as 71% in C. praecellens.

The venom of C. geographus is constituted by relatively few (21) toxin gene superfamilies (Hu et al. 2012), with very high relative expression of the A-, B-, and O1 superfamilies toxins, followed by the S-, B2-, M- and T-superfamilies. Both A- and B-gene superfamilies are absent from the venom gland transcriptomes of Turriconus and Splinoconus species. However, the O1- and M-superfamilies are expressed by all five species, although with varying expression levels (fig. 3), the T-superfamily toxins show relatively high expression in all species with the exception of C. tribblei, and S-superfamily toxins are expressed at similar levels in Turriconus spp. and C. geographus, while showing very low expression levels in Splinoconus. The B2- and Con-ikot-ikot superfamilies that are characterized by very high expression in Splinoconus spp. and on the contrary, low in Turriconus, display intermediate expression levels in C. geographus. Expectedly, several gene families expressed in the two Turriconus species are absent (or of negligible expression) in C. geographus: P, I1, I2, V, I4, and A2.

Collectively, these data suggest rather minor differences in venom composition between the two closely related Turriconus species but pronounced differences between Turriconus spp. and members of other Conus lineages. Indeed, at least at the gene superfamily level, it appears that the notable difference between the C. andremenezi and C. praecellens venom repertoires is limited to differences in the relative expression levels of δ-like conotoxins (O1 superfamily) (Heinemann and Leipold 2007) and P-superfamily peptides (table 1). This is further illustrated by PCA based on the relative expression of conotoxin superfamilies (fig. 4). The two Turriconus species are projected closely together in Eigen PC1 and PC2 space (relative to the other three species), and the slight separation in projection is largely due to differences in the relative expression of the O1d- and P-superfamilies. The two members of the Splinoconus clade group closely together in PC1 and PC2 space and are well separated from the other three species.

**Interspecific Differences within Turriconus: The P- and O1d-Superfamilies**

In all four Turriconus individuals analyzed, P-superfamily conotoxins were both diverse and highly expressed. In contrast, P-superfamily conotoxin expression was lower in the two species of Splinoconus and not detected in the venom gland transcriptome of the fish-hunter C. geographus. In C. praecellens, the P-superfamily is by far the most highly expressed conotoxin superfamily (41–44%), with 16–27 variants identified. In both specimens of C. andremenezi P-superfamily expression levels, although generally high...
(9.9% and 11.6%), are still substantially lower than in *C. praecellens*.

Interestingly, much of the difference in the total change in frequency for the P-superfamily between the two *Turriconus* species can be accounted for by the extremely high relative expression of two unusual peptide isoforms in *C. praecellens*, Ps9.1 and Ps9.2 (14% and 24%, respectively). These peptides share the cysteine framework typical of the P-superfamily, but are substantially longer than most previously reported sequences and exhibit a remarkably high frequency of histidine residues (12 of 45 residues in Ps9.2). A putative ortholog of Ps9.2 was also detected in *C. andremenezi*, where it also shows relatively high expression (1.9%), but no ortholog was detected for Ps9.1 (fig. 5).

Another major difference between these closely related species is the high expression of the O1d superfamily in both specimens of *C. andremenezi* (~20% and 36%). In vertebrate prey *δ*-conotoxins block deactivation of sodium channels, which can lead to tetanic paralysis producing prey immobilization (Heinemann and Leipold 2007). Although

---

**Fig. 5.**—Inter- and intraspecific variation in *Conus* venom composition. (A) Sequence alignments of a selection of highly expressed conotoxins, where putative orthologs were identified in all four *Turriconus* specimens (predicted mature peptides and their names are shown); cysteines are highlighted in yellow and differences in amino acids are shown in red. *Predicted C-terminal amidation.* (B) Histograms of rbh pair counts versus percent sequence identity for all species analyzed in this study. The total number of rbh pairs is depicted on graphs.
this superfamily is also relatively highly expressed in C. praecellens (7–13%), its expression is still considerably lower than in C. andremenezi.

**A Fine-Grained Comparison—Interspecific Difference at the Sequence Level**

To investigate interspecific differences at sequence level, we performed comparative sequence analyses between the venom repertoires of each species. All-by-all NCBI-BLASTN searches identified 67 and 90 rbh pairs between the two specimens of each Turriconus species (C. andremenezi and C. praecellens, respectively) and between 53 and 57 rbh pairs when comparing the two species to one another (fig. 5B). In contrast, only between 9 and 16 rbh pairs were identified between the two Turriconus species and C. tribblei, C. lenavati, and C. geographus (fig. 5B).

Comparative analysis of conotoxin sequences between the two C. andremenezi specimens revealed that 98.5% of rbh pairs were greater than 95% identical, and of these, 46% of sequences were 100% identical (fig. 5B). Comparative analysis of conotoxin sequences between the two specimens of C. praecellens showed that 89% of rbh pairs were greater than 95% identical, and of these, 52% of sequences were 100% identical. These data are consistent with low variation between individuals of the same species of Conus.

Comparative analysis of C. praecellens conotoxin sequences to those of the two C. andremenezi specimens revealed that 68/60% of conotoxin precursors shared >95% identity, but of these, only 9/2% of sequences shared 100% identity. These data indicate that although there is overall similarity between the two species of Turriconus, considerable variation is present at individual sequence level. The low number of identical conotoxin sequences between C. andremenezi and C. praecellens suggests that despite being closely related these species have distinct venom repertoires. Comparison of conotoxin sequences from the two Splinoconus species and C. geographus to those of C. andremenezi and C. praecellens revealed that no toxin precursor shared >87% identity, highlighting the drastic differences between conotoxin sequences belonging to species from different Conus lineages.

Sequences of highly expressed conotoxins, for which similar sequences were identified in all four Turriconus specimens, are shown in figure 5B. This comparison provides concrete examples of sequences from two individuals of the same species, with these compared with the most similar peptide from a sister species. The transcripts of the first set shown (T-superfamily, fig. 5Bi) have identical mature toxin sequences in the four specimens analyzed. This identity appears atypical; more common are the sets of peptides in the same superfamily (T-superfamily fig. 5Bii) where the two C. andremenezi specimens are identical, but a small number of amino acid substitutions are found in the putative C. praecellens homolog. An interesting variation of this is shown in the next group of peptides compared (O1- and P-superfamily, fig. 5Biii and iv) for which the two C. andremenezi specimens have identical sequences, and C. praecellens expresses peptides with a few amino acid substitutions. In all of these cases however, the sequences align with few gaps and identical cysteine arrangement, and the differences between C. andremenezi and C. praecellens are predominantly accounted for by amino acid substitutions. The 11-superfamily of peptides provides an example of identical sequences in C. andremenezi, with a divergent sequence in C. praecellens that involves indels (fig. 5Bv). Thus, there is considerable divergence in this example between the two species, but no amino acid substitutions between the two specimens in the same species.

**The A2-Superfamily in Turriconus**

The A-superfamily, a prominent and well-represented group of venom components in most Conus species investigated to date, is not detected in the four Turriconus specimens analyzed and also absent (or of negligible expression) in the two members of Splinoconus. In C. geographus these comprise over 10% of all transcripts. A related unexpected discovery from the analysis of the Turriconus transcriptomes was that although the A-superfamily is not detected, a distinct class of A-like sequences is present, albeit at low expression levels (0.0004–0.02%, table 1). As discussed later, similar sequences have recently been reported in other cone snail species (Bigs et al. 2008; Barghi et al. 2015a). Although these were named A-like, based on distinct sequence characteristics, we have now designated these peptides as the A2-superfamily. A2 conotoxins do not share the signal peptide that defines the A-superfamily (see fig. 6, examples of A-superfamily toxins from C. geographus); however there is some weak similarity in the propeptide region, and the encoded mature peptides exhibit the structural features that define the A-superfamily (i.e., cysteine framework I, similar intercysteine loop spacing, as well as other key residues).

**Discussion**

We have provided the first comprehensive portrait of the venom repertoire of two sister species in the Turriconus subgenus of Conus and insights into the venom divergence upon speciation.

**The Turriconus Venom Repertoire**

In general, with between 107 and 155 conotoxins derived from 21 to 29 superfamilies, the venom diversity of the two species from the Turriconus clade is consistent with what has been reported for other species of Conus including studies addressing intraspecific variations in Conus venom gland transcriptomes (Hu et al. 2012; Robinson et al. 2014, 2017; Barghi et al. 2015a, 2015b; Phuong et al. 2016). Nevertheless, these
numbers conflict with recent reports of as many as 3,303 conotoxins in *Conus epicopatus* that led authors to hypothesize that “transcriptional messiness” accounts for the rapid evolution of conotoxin encoding genes (Jin et al. 2013; Lavergne et al. 2015). We believe that these are overestimates that result from inadequate processing of next-generation sequencing data sets, as recently suggested by others (Phuong et al. 2016).

Although the overall number of conotoxins is comparable to other cone snail species, the two members of the *Turriconus* clade exhibit clear differences in toxin superfamily expression. Relative to other species of *Conus*, *Turriconus* appears to be more specialized in P-superfamily conotoxins. P-conotoxins account for almost half of total toxin number expressed in *C. praecellens* and approximately 10% in *C. andremenezi*. The specific biological activity of P-superfamily conotoxins is not well established: Very few representatives have been characterized to-date. Of these, TxIXA from *Conus textile* produces a “spasmodic phenotype” in mice when injected intracranially (Lirazan et al. 2000), nevertheless, the molecular target of TxIXA remains to be elucidated. Very high expression levels suggest a key functional role for these toxins in *Turriconus*, so future work will be directed at investigating the biological activity of these peptides and correlating findings to the biology of *Turriconus*.

Another difference is the absence (or very low expression) of A-superfamily toxins that are functionally important components of several other *Conus* venoms characterized to date, especially in species that prey on fish (Santos et al. 2004; Terrat et al. 2012), including *C. geographus*. α-conotoxins comprise the largest and most-studied group of A-superfamily toxins. These peptides potently and selectively block neuronal and muscular nicotinic acetylcholine receptors (Santos et al. 2004). α-conotoxins have been identified from diverse phylogenetic lineages of *Conus*, including the early-diverging worm-hunting *Stephanoconus* clade (Elsson et al. 2003). Thus, the absence (or very low expression) of A-superfamily toxins in *Turriconus* suggests that these peptides ceased to serve a functional role in envenomation. As noted earlier, we instead detected expression of a genetically distinct A-superfamily that would be predicted to play a similar functional role, but is expressed at relatively low levels.

Although the A2-superfamily sequences display all of the hallmarks of typical conotoxins, one puzzling finding is that they appear to be expressed at only very low relative levels. It is notable that structurally similar conotoxins had previously been characterized from *Conus pulicarius* (Biggs et al. 2008), *C. tribblei*, and *C. lenavati* (Barghi et al. 2015a), all presumably worm-hunting species. In *C. tribblei* and *C. lenavati*, A2 conotoxins are similarly expressed at low levels, whereas in *C. pulicarius*, which belongs to the more recently diverged subgenus *Puncticulus*, this gene superfamily was expressed in the salivary glands and not the venom gland, where conventional A-superfamily conotoxins were expressed. It will be of interest to further investigate whether the absence (or very low expression) of the A-superfamily correlates with presence of the A2-superfamily in other *Conus* species.

Besides these apparent differences between *Turriconus* and other *Conus* clades, several similarities could be observed. Highly expressed toxin superfamilies in *Turriconus* included the O1d-, M-, and T-superfamilies, which are found in almost all *Conus* venoms characterized to date. Thus, these gene superfamilies must play a fundamental role in envenomation and can be considered hallmarks of *Conus* venoms.

**Fig. 6.**—The A2-superfamily sequences identified in the venom gland transcriptomes of *C. praecellens* and *C. andremenezi*. For comparison, another A2-superfamily sequence identified in the salivary gland of *C. pulicarius* PuSG1.1 (Uniprot: P0C8U6), and two A-superfamily sequences, GID and (Uniprot: P60274) GI (Uniprot: P01519) from *C. geographus* are shown. Signal peptides are underlined in purple, predicted mature peptides are underlined in black.

**Speciation and Rapid Divergence of Conotoxin Exogenes**

Conotoxins are among the fastest evolving gene products known in nature (Chang and Duda 2012; Sunagar et al. 2015). Considerable prior work has demonstrated that at the sequence level, each species of *Conus* appears to have its own distinct repertoire of conotoxins. Individuals of the same species would be expected to share a near identical venom repertoire, with only allelic differences among genes evident (Duda et al. 2009; Chang et al. 2015). The comparison of the venom repertoires of two individuals of the
same species is also clearly consistent with expectations—approximately half of the identified rbh pairs matched with 100% identity, whereas the other half demonstrated >90% identity, suggesting reasonable levels of allelic polymorphism of conotoxin genes within the species population (Chang et al. 2015). Our findings on intraspecific variation of venom repertoire are largely consistent with the recent studies on multiple individuals of two closely related species C. tribblei and C. lenavati (Barghi et al. 2015a). The expression level of conotoxin gene superfamilies was similar across all three specimens of C. tribblei analyzed, and two of the three specimens of C. lenavati also showed low levels of intraspecific variation in conotoxin gene superfamily expression. Interestingly, the third C. lenavati specimen showed some substantial differences. Numerous factors might contribute to this observed difference, and the authors noted that this specimen was smaller than the two others examined, suggesting that age may be a factor contributing to venom composition as reported by others (Safavi-Hemami et al. 2011).

Consistent with published data, closely related species often share similar but very rarely identical conotoxin sequences. Nevertheless, interpretation of the differences in toxin expression patterns between two species of Conus cannot be solely addressed by the degree of phylogenetic relatedness between these two species (Phuong et al. 2016). Differences in the ecology between these species should ultimately be taken into consideration, as this provides insights into the different ecological niches’ requirements that determined the adaptation of these species upon speciation. In this respect, syntopic species constitute a perfect model for addressing divergence of toxin expression patterns.

In addition to the two Turriconus species analyzed in this study, the Splinocorus species, C. tribblei and C. lenavati (Barghi et al. 2015a), were sampled from the same geographic locality off North Cebu in the Philippines. As all four species occur in (and were sampled from) the same locality and habitat, it is reasonable to expect that the venom gland transcriptomes of species in these two pairs are not greatly affected by adaptations to different environmental variables. Furthermore, as muddy bottom communities at depths of 180–250 m are more homogeneous at small spatial scale, and feature very limited diversity of microhabitats comparing to shallow water coral reef ecosystems (Gray 2002), we hypothesize that the main evolutionary force driving diversification in both species pairs were differential biotic interactions, primarily adaptation to different prey and predator species. This is why exogenes that are presumed to be sensitively tuned to mediate particular biological interactions following speciation may be expected to diversify rapidly in the case of both Turriconus and Splinocorus.

Indeed, we have shown that less than 10% of the venom exogene repertoire is identical between C. praecellens and C. andremenezi. Only a very small fraction of rbh pairs were 100% identical (2–9%). On the other hand, over half (68–78%) shared >90% identity, reflecting the close relationship between these two species. Although at conotoxin superfamily level, the venoms of C. andremenezi and C. praecellens are highly similar, there are two apparent differences: The much higher relative expression of several O1d conotoxins in C. andremenezi and very high relative expression of P-superfamily peptides in C. praecellens. In additional to these differences, it should be noted that within each superfamily there exists a large repertoire of species-specific sequences and structural scaffolds. Together, these data demonstrate that beyond the relatively subtle differences evident at the level of conotoxin gene superfamily expression, there are substantial differences present at the sequence level between the two sister species. The demonstrated notable divergence of the venom composition between C. andremenezi and C. praecellens further highlights that exogenes are a sensitive indicator of hidden biological differences.

Our comparison of the two species of Turriconus and C. tribblei, C. lenavati revealed no sequence similarity exceeding 87% across the entire conotoxin repertoire. Nevertheless, certain similarity can be noted at the gene superfamily level, with a set of 4–5 toxin gene superfamilies showing comparable and generally high relative expression levels in the four species. Therefore, at the gene superfamily level, our results highlight the role of the species-specific toxin expression patterns that are closely comparable among closely related species, but retain some recognizable features even in comparison of distantly related lineages.

Several studies have now demonstrated that diet is a major contributor to venom evolution in cone snails. Conotoxin gene diversity appears to be highest in species that prey on the widest diversity of prey (Phuong et al. 2016; Phuong and Mahardika 2017). Our results reveal that the overall diversity of conotoxin transcripts and gene superfamilies does not significantly change upon speciation in the two species of Turriconus but demonstrate variations of expression levels of individual gene families (i.e., the P- and O1d-superfamilies). This finding indicates that changes in dietary breadth may not have contributed to speciation in Turriconus. Our recent study on weaponized insulins expressed in the venoms of several cone snail species suggests that toxin gene expression may correlate with prey preference and hunting strategies (Safavi-Hemami et al. 2016). In this scenario, speciation in Turriconus may have been accompanied by changes in the expression levels of toxin genes evolved to target different types of prey and/or to induce different physiological endpoints in their prey. Although this is an intriguing hypothesis, more dense taxon sampling combined with behavioral studies is required to correlate the role of prey preference and predation strategy in shaping toxin expression and to disentangle its contribution to speciation in this genus.

**Supplementary Material**

Supplementary data are available at Genome Biology and Evolution online.
Acknowledgments

The specimens used in this study were obtained in conjunction with a collection trip supported in part by an ICBG Grant No. 1U01TW008163. We would like to thank Noel Saguil for help with sample collection. This work was supported in part by National Institutes of Health Grants GM 48677 (to B.M.O.) and GM 099939 (to M.Y. and P.K.B.), an International Outgoing Fellowship Grant from the European Commission (CONBIOS 330486 to H.S.-H.), and the National Natural Science Foundation of China (Grant No. 31500626 to A.L.).

Literature Cited


**Associate editor:** Mandë Holford