Genome Biology and Evolution

SMBE

Comparison of the Venom Peptides and Their Expression in Closely Related *Conus* Species: Insights into Adaptive Post-speciation Evolution of *Conus* Exogenomes

Neda Barghi[1], Gisela P. Concepcion[1,2], Baldomero M. Olivera[3], and Arturo O. Lluisma[1,2,*]

[1]Marine Science Institute, University of the Philippines-Diliman, Quezon City, Philippines

[2]Philippine Genome Center, University of the Philippines, Quezon City, Philippines

[3]Department of Biology, University of Utah

*Corresponding author: E-mail: aolluisma@upd.edu.ph.

Accepted: May 26, 2015

Data deposition: Transcriptome Shotgun Assembly projects of *Conus tribblei* and *Conus lenavati* have been deposited at DDBJ/EMBL/GenBank under the accessions GCVM00000000 and GCVH00000000, respectively. The versions described in this article are the first versions: GCVM01000000 (*C. tribblei*) and GCVH01000000 (*C. lenavati*). The COI sequences of the specimens of *C. tribblei* and *C. lenavati* are available from GenBank under accessions KR107511–KR107522 and KR336542.

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

Abstract

Genes that encode products with exogenous targets, which comprise an organism's "exogenome," typically exhibit high rates of evolution. The genes encoding the venom peptides (conotoxins or conopeptides) in *Conus sensu lato* exemplify this class of genes. Their rapid diversification has been established and is believed to be linked to the high speciation rate in this genus. However, the molecular mechanisms that underlie venom peptide diversification and ultimately emergence of new species remain poorly understood. In this study, the sequences and expression levels of conotoxins from several specimens of two closely related worm-hunting species, *Conus tribblei* and *Conus lenavati*, were compared through transcriptome analysis. Majority of the identified putative conopeptides were novel, and their diversity, even in each specimen, was remarkably high suggesting a wide range of prey targets for these species. Comparison of the interspecific expression patterns of conopeptides at the superfamily level resulted in the discovery of both conserved as well as species-specific expression patterns, indicating divergence in the regulatory network affecting conotoxin gene expression. Comparison of the transcriptomes of the individual snails revealed that each specimen produces a distinct set of highly expressed conopeptides, reflecting the capability of individual snails to fine-tune the composition of their venoms. These observations reflect the role of sequence divergence and divergence in the control of expression for specific conopeptides in the evolution of the exogenome and hence venom composition in *Conus*.

Key words: conotoxin, species-specific expression pattern, diversification, conopeptide.

Introduction

The genes that encode the venom peptides in *Conus* (conopeptides or conotoxins) belong to a class of genes whose products act on exogenous targets (Olivera 2006). These genes, which may be referred to as "exogenes" and which collectively make up an organism's "exogenome" (Olivera 2006), share a common characteristic, namely their rapid evolution and diversification. The accelerated pace of evolution of conopeptide genes (Duda and Palumbi 2000; Conticello et al. 2001; Espiritu et al. 2001; Duda 2008; Puillandre et al. 2010; Chang and Duda 2012), together with the high diversity of species in *Conus sensu lato* (Puillandre et al. 2014, 2015), and the evolution of biochemical mechanisms that introduce a variety of post-translational modifications to the peptides (Dutertre et al. 2013) collectively account for the remarkable chemical diversity of conopeptides in nature.

A number of mechanisms have been proposed to explain the rapid evolution of conopeptides. Gene duplication and positive selection have been shown to be important factors (Duda and Palumbi 2000; Duda 2008; Duda and Remigio 2008; Puillandre et al. 2010; Chang and Duda 2012, 2014). Several molecular mechanisms such as recombination (Espiritu et al. 2001), hypermutation (Olivera et al. 1999; Espiritu et al. 2001), presence of error-prone DNA polymerase, and even

© The Author(s) 2015. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Genome Biology and Evolution
SMBE

special molecules protecting cysteine codons (Conticello et al. 2001) have been suggested, although solid evidence in support of these mechanisms is still lacking.

Studies have also shown that the expression of conopeptides differs among *Conus* species with different diets: Worm-, mollusk-, and fish-hunting (Conticello et al. 2001; Duda and Palumbi 2004) indicating the importance of diet as a factor influencing the evolution of venom components. Furthermore, the diet of each vermivorous *Conus* species co-occurring in a habitat is dominated by a different type of polychaete where each *Conus* species have specialized on different prey taxa (Kohn 2001). Therefore, there seems to be a dietary specialization within worm-hunting species (Duda et al. 2001; Kohn 2001). In addition, *Conus* species that feed exclusively on specific types of prey express less diverse conopeptides than those with a broader range of prey (Remigio and Duda 2008; Elliger et al. 2011).

Little is known about how gene expression patterns differ between closely related species of *Conus*. In primates, variations in gene expression patterns are known to contribute to interspecies phenotypic variation and adaptation (Romero et al. 2012). Although the expression of majority of genes in closely related species is under purifying selection (Khaitovich et al. 2006), the expression of some genes has evolved more rapidly giving rise to species-specific gene expression patterns (Enard et al. 2002). A number of studies have shown that closely related species of *Conus* having similar diet also tend to produce conopeptides with similar structure and molecular targets (Espiritu et al. 2001; Olivera and Teichert 2007). For example, a few orthologous genes of the O1- and A-superfamilies were found to be expressed among closely related species of *Conus* (Duda and Remigio 2008; Chang and Duda 2012, 2014).

A crucial limitation of these previous studies was the insufficient sampling of expressed genes or conotoxin superfamilies partly owing to the technical limitations of the traditional cDNA library-based approaches (Duda and Remigio 2008; Chang and Duda 2012, 2014). Recent studies using next generation sequencing have shown the presence of 11–36 conopeptide gene superfamilies in each species (Hu et al. 2011, 2012; Lluisma et al. 2012; Terrat et al. 2012; Dutertre et al. 2013; Lavergne et al. 2013; Robinson et al. 2014; Barghi et al. 2015). In addition, the advent of high-throughput sequencing has made possible not only the identification but also the quantification of expression levels of conopeptide transcripts. Therefore, high-throughput sequencing could be leveraged to obtain deeper insights into the expression patterns of conopeptides and identifying the orthologous genes in all expressed conopeptide gene superfamilies. Similar to those in primate species, the differential expression of orthologous conotoxin exogenes may also contribute to the species-specific gene expression patterns and differentiation of venom composition in closely related species of *Conus*.

In addition to the interspecific variation in venom composition, characterization of intraspecific variability of the venom components may also provide clues to adaptive ability of the species as a whole. High intraspecific variability in the masses of venom peptides has been observed among the individuals of worm-, mollusk- (Davis et al. 2009), and fish-hunting (Jakubowski et al. 2005; Dutertre et al. 2010; Chun et al. 2012) cone snails. It was shown recently that variable post-translational processing could account for the observed large number (thousands) of peptide fragments (Dutertre et al. 2013). The expression of toxins in individuals of venomous animals such as snakes and scorpions varies with a number of factors such as sex, age, geographic location (Daltry et al. 1998; Menezes et al. 2006; Abdel-Rahman et al. 2009), and diet (Barlow et al. 2009; Casewell et al. 2009; Gibbs et al. 2013). The factors influencing the intraspecific variation in the expression patterns of conopeptides are still poorly known, and transcriptomic analysis of the venom duct of several individuals will be essential in characterizing such variability.

This study aims to investigate the expression and diversification of conopeptides in closely related species to provide insights into the processes that contribute to inter- and intraspecific differences in the venom composition and facilitate the evolution of *Conus* exogenome. This is the first comprehensive inter- and intraspecific transcriptome analysis of the venom duct in closely related *Conus* species using next generation sequencing. In this study, the expression patterns of conopeptide gene superfamilies were evaluated in several specimens of the closely related worm-hunting species, *Conus tribblei* and *Conus lenavati*. These two species belong to the same clade (VIII) (Espiritu et al. 2001). Recently, Puillandre et al. (2014) assigned these species to the subgenus *Splinoconus*. *Conus coffeae* and *Conus glans*, members of subgenus *Leporiconus*, the most closely related clade to the subgenus *Splinoconus* (Puillandre et al. 2014), are known to prey on errant polychaetes, mostly eunicidae (Duda et al. 2001). Furthermore, *Conus papuensis* another member of the subgenus *Splinoconus* (Puillandre et al. 2015) is also known to prey on polychaetes (Tucker and Tenorio 2009). As closely related *Conus* species generally have similar diet (Espiritu et al. 2001; Puillandre et al. 2014), *C. tribblei* and *C. lenavati* may also prey on polychaetes.

Materials and Methods

Sample Collection

A total of eight specimens of *C. tribblei* (~6 cm shell length) and five specimens of *C. lenavati* (~5–8 cm shell length) were collected using tangle nets from depth of 96–136 m in Sogod, Cebu province in the Philippines. The venom duct of each specimen was dissected, stored separately in RNAlater (Ambion, Austin, TX), and kept at −20 °C. The whole body

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

tissue and the shell of each specimen were stored in 95% ethanol and kept at room temperature.

Phylogenetic Analysis

DNA was extracted from a small piece of the foot tissue of each specimen using DNeasy blood and tissue kit (Qiagen, USA) according to manufacturer's recommendations and stored at −20 °C. Using universal primers: LCO1490 and HCO2198 (Folmer et al. 1994), a fragment of cytochrome oxidase c subunit 1 (COI) gene segment of mitochondrial DNA was amplified. All PCR reactions were performed in 30 µl containing 1× Titanium Taq PCR Buffer, 0.4 mM dNTP, 0.2 mM of each primer, 0.15 µl Titanium Taq DNA polymerase, and 0.5–4 ml of DNA template (total concentration of 100 ng in each PCR reaction). The PCR amplification consisted of one cycle of 1 min at 94 °C; five cycles of 1 min at 94 °C, 1.5 min at 45 °C and 1.5 min at 72 °C; 35 cycles of 1 min at 94 °C, 1.5 min at 50 °C and 1 min at 72 °C, and a final cycle of 5 min at 72 °C (Hebert et al. 2003). Each PCR product was gel purified using GF-1 AmbiClean Kit (vivantis) and sequenced in both directions using forward and reverse primers on an ABI 3730XL Capillary sequencer (Applied Biosystems). The manual inspection and cleaning of the sequences based on the chromatograms were performed using BioEdit version 7.0.0 (Hall 1999). Multiple sequence alignment was performed on the COI sequences of *C. tribblei* and *C. lenavati* specimens and other members of the subgenus *Splinoconus* downloaded from the National Center for Biotechnology Information (NCBI) (supplementary table S1, Supplementary Material online) using MEGA 5.2 (Tamura et al. 2011). Bayesian analysis was performed using six substitution categories, a gamma-distributed rate variation across sites approximated in four discrete categories, and a proportion of invariable sites in two parallel runs using MrBayes (Ronquist et al. 2012) each consisting of six Markov chains of 5 million generations with a sampling frequency of one tree in every 10,000 generations, and the chain temperature was set to 0.02. Convergence was evaluated using Tracer1.4.1 (Rambaut and Drummond 2007) to check that all effective sample size values exceeded 200. After stabilization of the log-likelihood scores, the first 25% of trees were omitted as burn-in, and a consensus tree was calculated. MrBayes was run on the Cipres Science Gateway (https://www.phylo.org/portal2/, last accessed June 15, 2015) (Miller et al. 2010) using MrBayes 3.2.2 on XSEDE.

RNA-Sequencing and Transcriptome Assembly

The mRNA extraction was performed individually on the venom ducts of three specimens of *C. tribblei* (*C. tribblei*_0317_1 [trib1, female], *C. tribblei*_0317_3 [trib3, male], and *C. tribblei*_0106_21 [trib21, male]) and three specimens of *C. lenavati* (*C. lenavati*_0317_1 [lena1, male], *C. lenavati*_0317_2 [lena2, female], and *C. lenavati*_0317_3 [lena3, male]) (fig. 1) using Dynabeads mRNA DIRECT kit (Invitrogen Dynal AS Oslo, Norway). For a more efficient extraction, each venom duct was homogenized using 0.5-mm Zirconia/Silica beads (Biospec Products, Inc.) in a bead beater (Precellys, Berlin Technologies), and the rest of the protocol was performed according to the manufacturer's recommendations. A cDNA library was constructed for each of the specimens, fragments with the average insert size of 200 bp were selected, and the libraries were multiplexed and paired-end (PE) sequenced using Illumina HiSeq 2000.

The quality of 90-bp long PE reads generated for each specimen was evaluated using FastQC software v0.10.1 (www.bioinformatics.babraham.ac.uk/projects/fastqc/, last accessed June 15, 2015). Trimming of the low quality bases (<20 phred score) at 3′- and 5′-ends of the reads and removal of the low quality reads (reads in which less than 90% of bases had phred score ≥ 20) were performed using FASTX-Toolkit version 0.0.6 (Pearson et al. 1997), and the reads with no pairs were excluded using an in-house Python script. For each species, the sequencing reads from the three specimens were combined into a single data set and *de novo* assembled using Trinity (Grabherr et al. 2011) with default settings to generate the "reference assembly."

Conopeptide Identification and Superfamily Classification

The conopeptide sequences in NCBI's nr protein database, UniProtKB/Swiss-Prot database (UniProt Consortium 2015), and ConoServer (Kaas et al. 2012) were downloaded and pooled in a data set. The putative conopeptides identified by Barghi et al. (2015) were also added, the redundant sequences of the data set were removed, and the remaining sequences were formatted into the "reference conopeptide database" using formatdb software of BLAST (Basic Local Alignment Search Tool) 2.2.29+ (Altschul et al. 1990). To identify putative conopeptides, the sequences in the reference assembly of *C. tribblei* and *C. lenavati* were searched against the reference conopeptide database using BLASTX. In each BLASTX result, the sequences with significant hits (*e* value $< 10^{-5}$) were translated according to the identified reading frame and were manually inspected. The mature regions of the conopeptides were predicted using ConoPrec (Kaas et al. 2012), the redundant and truncated transcripts were excluded from the data set, and the good quality putative conopeptide precursors were collected into the "tribblei conopeptide data set" and "lenavati conopeptide data set."

The classification of the conopeptides in the tribblei conopeptide data set and lenavati conopeptide data set into gene superfamilies was performed as described in Barghi et al. (2015). Briefly, conopeptides were assigned into gene superfamilies based on the identification of the two highest-scoring full-length conopeptide precursor hits in the BLASTX search (in which the sequences in the reference assemblies of *C. tribblei* and *C. lenavati* were used to search for similar sequences in the reference conopeptide database). The signal regions of

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

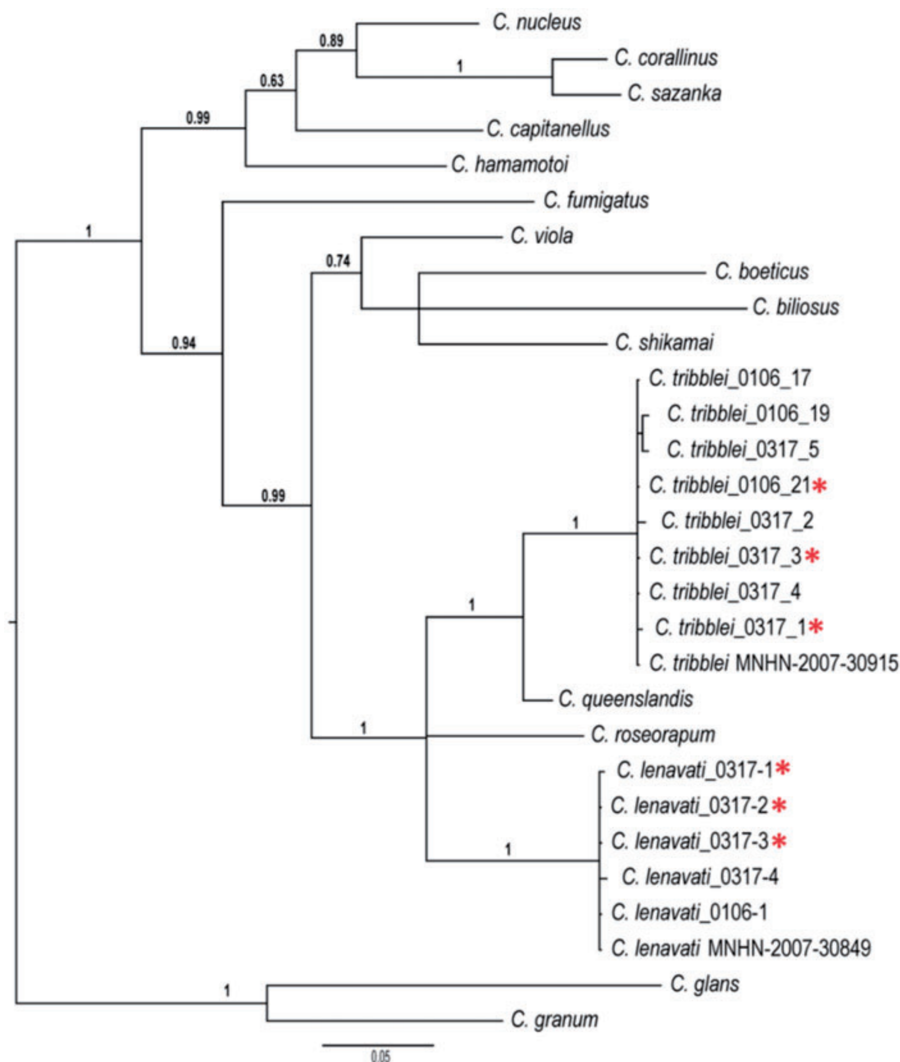

Fig. 1.—Phylogenetic relationship of the species of subgenus *Splinoconus*. The phylogenetic tree is inferred from the partial COI sequences using Bayesian analysis. Posterior probabilities are indicated for each node. The specimens of *C. tribblei* and *C. lenavati* used for the transcriptome analysis are marked with red asterisks.

the putative conopeptide precursors were predicted using SignalP 4.1 (Petersen et al. 2011), and percentage sequence identity (PID) between the signal region of each putative conopeptide precursor and the highly conserved signal sequence of the known conopeptide gene superfamilies was computed using MatGAT 2.02 (Campanella et al. 2003). Originally, the PID value of 76 was chosen as the threshold for superfamily assignment (Kaas et al. 2010). However, some conopeptide gene superfamilies have less conserved signal regions. Therefore, based on the average PID of members of each superfamily, specific PID threshold values were set for different conopeptide gene superfamilies (Barghi et al. 2015). If the PID of a conopeptide's signal region was above the threshold of a superfamily, the conopeptide was assigned to that gene superfamily. Otherwise, the conopeptide was classified as a new conopeptide group. The assigned name of each new group was "SF-" plus Arabic numbers such as "01." Multiple sequence alignment was performed on the precursors of the putative conopeptides and the reference conopeptide sequences using ClustalX version 2.1 (Larkin et al. 2007) followed by manual refinement using BioEdit.

Comparison of Conopeptide Identity

For each conopeptide gene superfamily identified in both *C. tribblei* and *C. lenavati*, the conopeptides' mature regions were aligned using ClustalX, the pairwise distances were computed using MEGA, and the PID was computed for every conopeptide pair using the following formula:

$$\text{percentage sequence identity} = \frac{\text{identical residues} \times 2}{\text{sum of the length of a conopeptide pair}}.$$

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

Genome Biology and Evolution
SMBE

For every conopeptide in *C. tribblei*, the conopeptide in *C. lenavati* with the highest PID (and belonging to the same superfamily) was chosen as the most similar match, and the same criterion was applied for choosing the most similar match for the conopeptides in *C. lenavati*.

Test of Positive Selection for Orthologous Conopeptide Genes

To identify the orthologous conopeptide genes in *C. tribblei* and *C. lenavati*, the nucleotide sequences of the conopeptide precursors of each superfamily identified in both species were aligned using ClustalX, and the best-fitted model was chosen using MEGA. The phylogenetic tree of the conopeptides of each superfamily was constructed using maximum-likelihood method based on the best-fitted model with bootstrap support value based on 500 replicates using MEGA. Those conopeptides that occurred in clades with strong bootstrap support containing only two sequences, one from *C. tribblei* and the other one from *C. lenavati*, and had identical cysteine framework were considered orthologous. For the superfamilies containing only one conopeptide in *C. tribblei* and one conopeptide in *C. lenavati*, the sequences were considered orthologs if, in addition to having identical length and cysteine framework, the PID of the mature region was above 78%. The pairwise estimates of nonsynonymous substitutions per nonsynonymous sites (d_N) and synonymous substitutions per synonymous sites (d_S) for the mature regions of the orthologous conopeptide pairs were computed using the maximum-likelihood approach implemented in PAML (runmode = −2, CodonFreq = 2) (Yang 1997).

Shannon's Diversity Index of Venom Conopeptides

Diversity of the conopeptides in the tribblei and lenavati conopeptide data sets was computed as Shannon's diversity index, H', where R is the number of conopeptide gene superfamilies and p_i is the proportion of conopeptides belonging to the ith superfamily in the data set. The evenness of conopeptide data sets was computed as Shannon's equitability, $E_{H'}$, whereas S is the data set richness based on the number of conopeptide gene superfamilies:

$$H' = -\sum_{i=1}^{R} p_i \ln(p_i) \quad E_{H'} = {}^{H'}/_{\ln S}$$

Quantification of the Transcript Expression Level

In the recent transcriptome studies of the venom duct of several *Conus* species, the expression level of each conopeptide has been determined by the total number of sequencing reads generated from each conopeptide transcript (Hu et al. 2012; Terrat et al. 2012; Dutertre et al. 2013; Jin et al. 2013; Robinson et al. 2014). However, the transcript length and the sequencing depth affect the number of reads generated from each transcript (Mortazavi et al. 2008). Therefore, for a more accurate comparison of the expression levels among conopeptides with different lengths, normalization of the number of reads by the transcript length and sequencing depth was performed (Mortazavi et al. 2008; Trapnell et al. 2010).

To determine how many of the conopeptides were expressed in each specimen of *C. tribblei* and *C. lenavati*, the sequencing reads of each specimen were aligned separately to the tribblei and lenavati reference assemblies, respectively, using Bowtie 1.0.1 (Langmead and Salzberg 2012). For each specimen, the alignment of the reads to the transcripts were visualized in Tablet (Milne et al. 2012) to ensure that reads align to the entire length of conopeptide precursor or at least to the full length of the mature region. Only those transcripts where the mature regions were entirely covered by reads were considered as "expressed." For each specimen, the number of reads representing a transcript was estimated as the maximum-likelihood abundances using expectation-maximization algorithm by RSEM (Li and Dewey 2011) using the default settings, and was represented as the "expected count." To enable comparison of the expression levels of transcripts among specimens, trimmed mean of M values (TMM) normalization factor was computed using edgeR (Robinson and Oshlack 2010). Then, the "effective library size" for each specimen was computed by normalizing the library size of each specimen (total number of aligned reads) using the TMM normalization factor. The "effective length" of each transcript which is the mean number of positions along the transcript sequence that a read may start (Li and Dewey 2011) was also computed using RSEM. Finally, the expression level of each transcript was computed as TMM-normalized Fragments Per Kilobase of transcript per Million fragments mapped (FPKM) (Trapnell et al. 2010; Haas et al. 2013) using the following formula:

$$\text{TMM} - \text{normalized FPKM} = \frac{\text{expected count} \times 10^9}{\text{effective length} \times \text{effective library size}}.$$

For each specimen, the expression level of each gene superfamily is the sum of expression levels (TMM-normalized FPKM) of the conopeptides belonging to that superfamily represented as "sum TMM-normalized FPKM." As for each species, the expression level of each gene superfamily is the average of the expression levels of that superfamily among three specimens represented as "mean TMM-normalized FPKM."

Functional Annotation

The functional annotation of the reference assemblies of *C. tribblei* and *C. lenavati* was performed using the pipeline version of BLAST2GO software (B2G4Pipe) (Götz et al. 2008). The transcripts of the reference assemblies were searched

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

Genome Biology and Evolution
SMBE

against the UniProtKB/Swiss-Prot database using BLASTX, and the results containing the significant hits (*e* value $< 10^{-5}$) were loaded to the pipeline. The GO terms for each blast hit were extracted from the local B2G MySQL database and assigned to each transcript. After refinement of the annotation results by GO-Slim function, the plot of transcripts GO classifications was constructed using Web Gene Ontology Annotation Plot (WEGO) software (Ye et al. 2006).

Correlation Analysis of Transcripts' Expression Levels

The Pearson's correlation coefficient was computed for the expression levels of conopeptides and gene superfamilies among specimens of either *C. tribblei* or *C. lenavati* using "TMM-normalized FPKM" and "sum TMM-norm FPKM" values, respectively. In order to validate the correlation in the expression patterns of conopeptide superfamilies between *C. tribblei* and *C. lenavati*, the expression levels of conopeptide superfamilies, mean TMM-normalized FPKM, were used for the Pearson's correlation coefficient test. Additionally, to confirm the correlation of expression levels of the orthologous conopeptide genes between *C. tribblei* and *C. lenavati*, the mean expression level of each orthologous conopeptide was computed among specimens of each species, and these values were used for the correlation test. In all the correlation analyses, the expression level values were transformed to logarithmic scale (log2), and the significance levels were calculated using Bonferroni correction.

Results

Taxonomic Identification of the Specimens and Their Phylogenetic Relationships

The phylogenetic reconstruction of the relationship of COI sequences of members of the subgenus *Splinoconus* (Puillandre et al. 2014) (supplementary table S1, Supplementary Material online) showed that *C. tribblei*, *Conus queenslandis*, *Conus roseorapum,* and *C. lenavati* clustered together in a clade (fig. 1). Specifically, *C. tribblei* and *C. lenavati* formed two distinct branches of this clade whereas all the specimens of each species were grouped together.

Transcriptome Sequencing and Assembly

The mRNA sequencing of *C. tribblei* specimens generated 29,138,176 (trib1), 23,917,406 (trib3), and 29,916,262 (trib21) good quality PE reads, whereas the mRNA sequencing of *C. lenavati* specimens generated 32,335,757 (lena1), 28,637,629 (lena2), and 33,027,986 (lena3) good quality PE reads. The pooled reads of three *C. tribblei* specimens (82,971,844 PE reads) and three specimens of *C. lenavati* (94,001,372 PE reads) with the average size of 81 bp were *de novo* assembled using Trinity into "tribblei reference assembly" and "lenavati reference assembly," respectively.

The tribblei reference assembly contained 327,700 transcripts with an average length of 489 bp, whereas 298,481 transcripts with an average length of 476 bp were identified in the lenavati reference assembly (supplementary table S2*a*, Supplementary Material online). The size of the assembled transcriptome was 160.27 Mb in *C. tribblei* and 142.14 Mb in *C. lenavati*. The N50 of the transcripts was 567 bp in the tribblei reference assembly and 547 bp in the lenavati reference assembly. Around 64% of the transcripts in both assemblies were 201–400 bp long and the length of 23–24% of the transcripts was 401–800 bp (supplementary table S2*b*, Supplementary Material online). However, the maximum transcript length was 21,724 and 16,659 bp in the tribblei and lenavati reference assemblies, respectively.

Conopeptide Diversity

A total of 100 unique putative conopeptide precursors belonging to 39 gene superfamilies were identified in the tribblei conopeptide data set whereas the lenavati conopeptide data set contained 132 novel putative conopeptide precursors that were classified into 40 gene superfamilies (fig. 2*a*). A total of 55 conopeptides in the tribblei conopeptide data set were previously discovered in *C. tribblei* (accession GCJM00000000 in DDBJ/EMBL/GenBank; Barghi et al. 2015), whereas the rest of the identified conopeptides were new. The majority of the identified conopeptides was classified into the known gene superfamilies (fig. 2*a*). Moreover, some new conopeptide groups indentified in recent transcriptome studies of the *Conus* venom duct were also discovered (fig. 2*a*); several sequences of the tribblei and lenavati conopeptide data sets showed high similarity to SF-mi2 and SF-mi4 in *Conus miles* (Jin et al. 2013), R, W, and Y2 in *Conus marmoreus* (Lavergne et al. 2013), U superfamily in *Conus victoriae* (Robinson et al. 2014), and A-like, G-like, N-like, Y2-like, and SF-01 groups in *C. tribblei* (Barghi et al. 2015) (supplementary fig. S1, Supplementary Material online). In addition, conopeptides with signal region PID values below the threshold for any conopeptide gene superfamily were classified into conopeptide groups SF-02 to SF-06 (fig. 3). Furthermore, seven putative conopeptides in the tribblei conopeptide data set and five conopeptides in the lenavati conopeptide data set showed similarity to the "divergent" superfamilies: Divergent MKFPLLFISL, divergent M---L-LTVA, divergent MSTLGMTLL-, and divergent MSKLVILAVL (fig. 2*a*). Divergent superfamilies refer to the conopeptides identified in early divergent species, *Conus californicus* and *Conus distans*. The conopeptide sequences identified in *C. tribblei* and *C. lenavati* are shown in supplementary figure S1, Supplementary Material online.

In both species, roughly half of the identified conopeptide precursors belonged to few superfamilies, and the most prominent superfamilies in both species were the con-ikot-ikot family and the M- and O1-, and O2-superfamilies (fig. 2*a*). Although the majority of gene superfamilies was represented

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

Genome Biology and Evolution
SMBE

by more than one conopeptide, 19 and 16 superfamilies had only one sequence in the tribblei and lenavati conopeptide data sets, respectively. A total of 32 conopeptide gene superfamilies and groups were represented in both species. However, a few superfamilies such as the D-, J-, K- and Y2-superfamilies, the "divergent MSKLVILAVL" and "divergent MSTLGMTLL-" superfamilies, and the N-like group were only found in *C. tribblei*, whereas the T-, A-, E- and C-superfamilies, SF-05, SF-06, SF-mi2 and SF-mi4 groups were identified only in *C. lenavati*. On the other hand, the cysteine frameworks identified in the majority of conopeptide gene superfamilies were similar between the tribblei and lenavati conopeptide data sets (supplementary table S3, Supplementary Material online). However, several conopeptide superfamilies such as the O1-, O2-, L-, M-, H- and S-superfamilies, the con-ikot-ikot family, and the Y2-like and SF-04 groups exhibited some cysteine patterns that were exclusively observed in either the tribblei or lenavati conopeptide data sets.

The diversity of putative conopeptides as indicated by the number of gene superfamilies was higher in *C. tribblei* and *C. lenavati* than in other previously studied species (table 1). Subsequently, both *C. tribblei* and *C. lenavati* had Shannon's diversity index of 3.30, higher than in other species which ranges from 1.29 (*Conus bullatus*) to 2.35 (*C. victoriae*) (table 1). Additionally, specimens of *C. tribblei* and *C. lenavati* showed Shannon's diversity indices ranging from 3.16 to 3.30

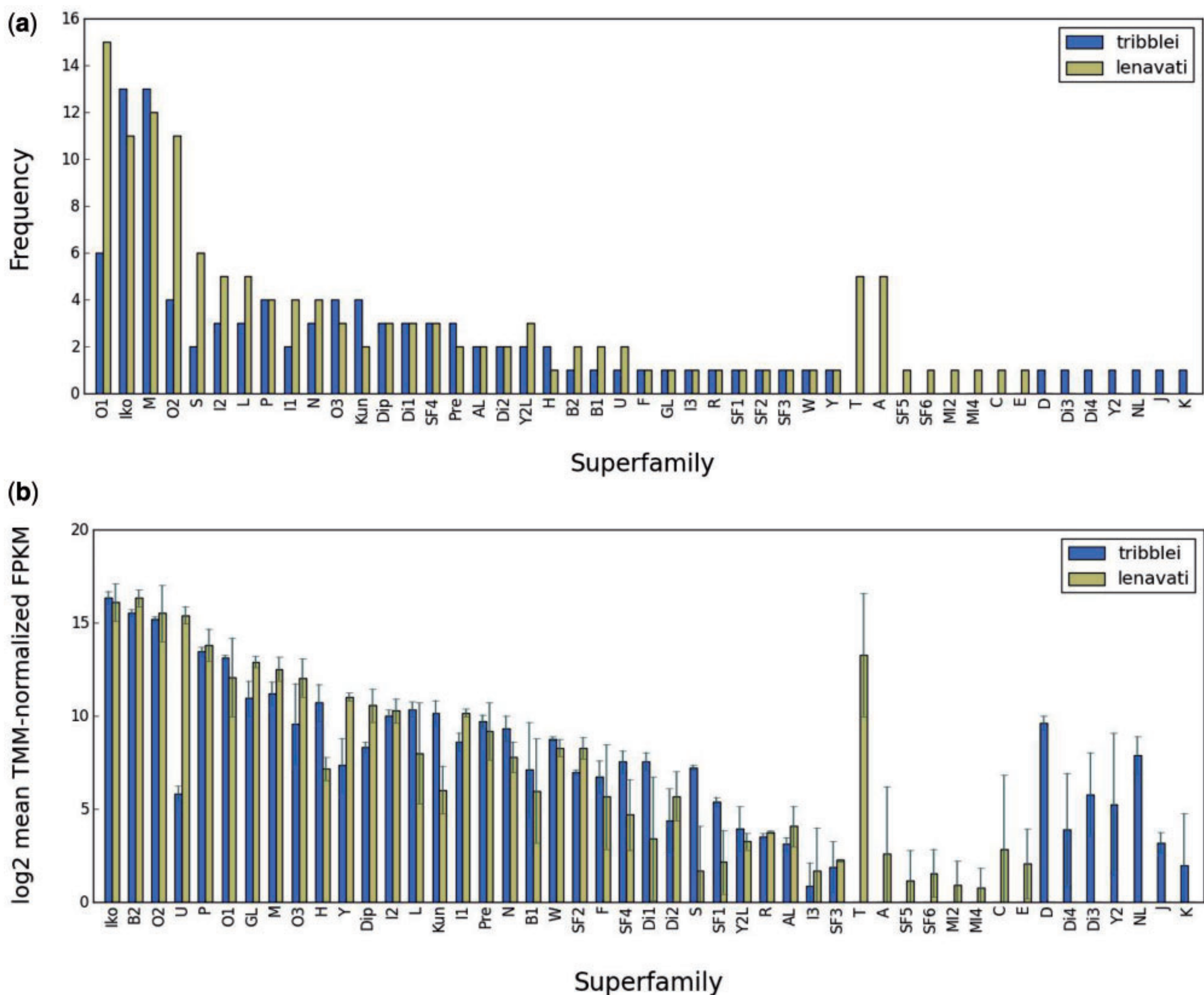

Fig. 2.—(a) The frequencies and (*b*) expression patterns of conopeptide gene superfamilies in *C. tribblei* and *C. lenavati*. The standard deviation ($n = 3$) bars are shown, and the expression values are log2-transformed. The superfamily abbreviations are: Iko, con-ikot-ikot; Dip, conodipine; Kun, conkunitzin; Pre, conopressin/conophysin; Di1, divergent MKFPLLFISL; Di2, divergent M---L-LTVA; Di3, divergent MSTLGMTLL-; Di4, divergent MSKLVILAVL; SF1, SF-01; SF2, SF-02; SF3, SF-03; SF4, SF-04; SF5, SF-05; SF6, SF-06; MI2, SF-mi2; MI4, SF-mi4; NL, N-like; GL, G-like; AL, A-like; Y2L, Y2-like.

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

(table 2). However, the evenness of the conopeptides in *C. tribblei* and *C. lenavati* was slightly higher than other species (table 1).

Expression Patterns of Conopeptide Gene Superfamilies

The expression patterns of conopeptides gene superfamilies in *C. tribblei* and *C. lenavati* were highly variable ranging from FPKM >85,000 to <10 (fig. 2*b* and supplementary table S4, Supplementary Material online). Superfamilies such as B2, O2, and P and con-ikot-ikot family were highly expressed (FPKM > 10,000) in *C. tribblei* and *C. lenavati*. Others, like M-, O1-, O3- and I2-superfamilies and G-like group also showed high expression levels (1,000 < FPKM < 10,000) in both species. Several conopeptide groups were moderately (100 < FPKM < 1,000) expressed: F-, N-, B1- and W-superfamilies, conopressin/conophysin family, and SF-02 group. On the other hand, expression levels of "Divergent M---L-LTVA"- and R-superfamilies and A-like group in *C. tribblei* and *C. lenavati* were much lower (10 < FPKM < 100). Other conopeptide superfamilies had more variable expression levels between the two species (supplementary table S4, Supplementary Material online). Both *C. tribblei* and *C. lenavati* expressed seven to eight superfamilies that were not present in the other species (fig. 2*b*). Although, the majority of these conopeptide groups had relatively low expression levels, the T-superfamily in *C. lenavati* was highly expressed (FPKM > 10,000).

The expression patterns of all the conopeptide superfamilies were moderately correlated between *C. tribblei* and *C. lenavati* ($r = 0.608$, P value < 0.001) (fig. 4), but it was apparent that a great number of superfamilies had highly

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

SF-02

```
             1                                                                          76
Cln_SF2_1    MLFVFTVVWILTMVMIITDVTFQSTCNTDNKPSCSEDTRLCGKNNSWGNCVALCKCPNQQACTTDTDHKVQVKRGP
Ctr_SF2_1    MLFVFTVVWILTMVMIITDVTFQSTCNTDNKPSCSEDTRLCGKNNSWGNCVALCKCPNQQACTTDTDHKVQVKRGP

             77                                                       133
Cln_SF2_1    FQLTETYYTCKNVSTMSDCQSNAKAMSGTSESTYKIMCKCDDTYKPSAPTNWKFICG
Ctr_SF2_1    FQLTETYYTCKNVSTMSDCQSNAKAMSGTSESTYKIMCKCDDTYKPSAPTNWKFICG
```

SF-03

```
Cln_SF3_1    MEALTIFRLCLLVALTTSVVVSAPLNDKVSDQEGECPVGGGRNPFVLCMRACLTTSTPYQCEHEYCKHCPRGRYAIVGHS
Ctr_SF3_1    MEALTIFRLCLLVALTTSVVVSAPLNDKVSDQNGECPVGGGRNPFVLCMRACLTTSTPYLCEHEYCKNC-RGRYAGLGHS
```

SF-04

```
Ctr_SF4_2    --------VLLLISTIAALYQEGRATQRRGKTVRTMSNLLNIQKRECPYGCPVVCPNGDECCDGIYCS---QGDRRYCVGCGGGRGE
Ctr_130_T    --------VLLLISTIAALYQEGRATQRRGKTVRTMSNLLNIQKRECPYGCPVVCPNGDECCDGIYCS---QGDRRYCVGCGGGRGE
Cln_SF4_2    --------VLLLISTIAALYQEGRATQRRGKTVRTMSNLLNIQKRECPYGCPVVCPNGDECCDGIYCS---QGDRRYCVGCGGGRGE
Ctr_SF4_1    MNCLQLLLVLLLISTTTALYPDVRKTRRRGGSIKTILRLLNSEKRDRSTGCPVKCPKHTECCSGITCTYLTPGGEYYCITSGGGGGE
Cln_SF4_1    --------------TIAALYQDGRATQRRDGNIRTMSDLLNIQKRECSPDCVATCPNGNECCDGDLCVYSSALEKYFCIGCGSGGGE
Ctr_SF4_3    --------------TIAALYQDGRATQRRDGNIRTMSDLLNIQKRECSPDCVATCPNGNECCDGDLCVYSSALEKYFCIGCGSGGGE
Ctr_131_T    --------VLLLISTIAALYQDGRATQRRDGNIRTMSDLLNIQKRECSPDCVATCPNGNECCDGDLCVYSSALEKYFCIGCGSGGGE
Cln_SF4_3    --------------TITALYQVDRATRREGRTIRAMSNLLKILKRRCSPDCVPYCPKKSECCDGDVCVYSETLSKHFCVGCG-GGGE
```

SF-05

```
P01619       -------------------------------------GGLGRCIYNCMNSGGGLSFIQCKTMCY
P02828       -------------------------------------GGVGRCIYNCMNSGGGLNFIQCKTMCY
Cln_SF5_1    MGFKVALIVLVLVVMATTSALPFQFSEFAETSVGQKRGGMGQCIHDCMNSGGGLSFIQCKTMCY
```

SF-06

```
Cln_SF6_1
MLSMFAWTLMTATVVVIAERQYCPIAGQTCTFGSDLCGKEESGSCSPRCNCKNERMCSRDSDHTITVVRVFRRRRPVEERYYTCVALSGLEECSNQKALTDL
VPETRELNSVEVHCKCSSPKVYGYHMYLKGYFCGTYERS
```

Fig. 3.—The putative conopeptide precursors of new conopeptide groups. The conopeptides identified in *C. tribblei* and *C. lenavati* are shown in black. The name of each conopeptide is presented as Ctr/Cln_$_#: Ctr, *C. tribblei*; Cln, *C. lenavati*; $, superfamily (the abbreviations are indicated in fig. 2); #, arbitrary assigned number. The reference sequences are shown in green, and cysteine residues are shown in bold italic red. The signal regions are highlighted, and the mature regions are underlined. The reference sequences are Ctr_130_T, Ctr_131_T (Barghi et al. 2015), P01619, and P02828 (ConoServer database).

correlated expression (32 shared superfamilies, $r = 0.820$, P value < 0.001). Although the expression patterns of majority of these superfamilies were apparently conserved, another subset consisting of the S-, H-, and U-superfamilies, and the conkunitzin family had very different expression levels in *C. tribblei* and *C. lenavati* (fig. 4 and supplementary table S4, Supplementary Material online). There is moderate correlation between the abundance and expression of different conopeptide gene superfamilies in *C. tribblei* ($r = 0.602$, P value < 0.001) and *C. lenavati* ($r = 0.597$, P value < 0.001) (supplementary fig. S2, Supplementary Material online).

Interspecific Genetic Divergence of Conopeptide Superfamilies

Although, the total number of identified conopeptides in the lenavati conopeptide data set (132) was higher than in the tribblei conopeptide data set (100), the number of superfamilies in both species was similar (39–40 superfamilies). More than half of the conopeptides of the superfamilies shared in *C. tribblei* and *C. lenavati* had PID > 81%, and these sequences were distributed across all gene superfamilies (except for H-, S-, and F-superfamilies) (fig. 5). Additionally, 12.02% (12 superfamilies) and 8.65% (11 superfamilies) of conopeptides in *C. tribblei* and *C. lenavati* showed 61–80% and 41–60% PID, respectively. Interestingly, a large portion (19.71%) of conopeptides belonging to 12 gene superfamilies had only 21–40% PID. 2.88% of the conopeptides exhibited PID < 20%, whereas the minimum sequence identity was 10%.

Divergence of Orthologous Conopeptide Genes

A total of 67 pairs of orthologous conopeptide genes belonging to 31 conopeptide gene superfamilies and groups were identified in *C. tribblei* and *C. lenavati* (supplementary table S5, Supplementary Material online). More than half of the orthologous gene pairs (37 pairs) had d_N to d_S ratios > 1, whereas nine gene pairs showed d_N to d_S ratios < 1. Interestingly, 21 orthologous gene pairs exhibited identical mature regions. Although 15 of these identical orthologs also contained identical signal and pre regions ($d_N = d_S = 0$), six gene pairs had 1–10 amino acid differences in the prepro regions.

The conopeptides with identical mature regions belonged to 14 superfamilies (B1-, N-, O1-, O2-, P-, W- and Y-superfamilies, "divergent MKFPLLFISL" superfamily, con-ikot-ikot and conopressin/conophysin families, and A-like, Y2-like, SF-02 and SF-04 groups) (fig. 6a). The expression levels of all orthologous conopeptide gene pairs in *C. tribblei* and *C. lenavati* were moderately correlated ($r = 0.640$, P value < 0.001, data not shown), and the correlation of the expression of identical orthologous conopeptide genes was slightly lower

Table 1
Diversity of Conopeptides in the Studied *Conus* Species through Next Generation Sequencing

Species	Richness (S)	No. of Conopeptides	Shannon's Diversity Index	Evenness (E)	Sequencing Platform	References
Conus bullatus	6	30	1.29	0.72	Roche 454	Hu et al. 2011
Conus geographus	16	63	2.30	0.83	Roche 454	Hu et al. 2012
Conus pulicarius	14	82	2.02	0.77	Roche 454	Lluisma et al. 2012
Conus consors	11	61	2.04	0.85	Roche 454	Terrat et al. 2012
Conus marmoreus	26	263	2.17	0.67	Roche 454	Dutertre et al. 2013; Lavergne et al. 2013
Conus victoriae	20	117	2.35	0.78	Roche 454	Robinson et al. 2014
Conus tribblei	36	136	3.13	0.87	Roche 454, Illumina	Barghi et al. 2015
Conus tribblei[a]	39	100	3.30	0.90	Illumina	This study
Conus lenavati [a]	40	132	3.30	0.89	Illumina	This study

Note.—Richness (S) is the number of identified conopeptide gene superfamilies.
[a]The number of conopeptides of *C. tribblei* and *C. lenavati* in this study is the total nonredundant conopeptides identified in three specimens for each species.

Table 2
Diversity of Conopeptides in the Specimens of *Conus tribblei* and *Conus lenavati*

Species	Specimen	Richness (S)	No. of Conopeptides	Shannon's Diversity Index	Evenness (E)
Conus tribblei	trib1	36	87	3.23	0.90
	trib3	36	85	3.23	0.90
	trib21	38	89	3.32	0.91
Conus lenavati	lena1	31	71	3.19	0.93
	lena2	31	71	3.16	0.92
	lena3	40	125	3.30	0.90

SMBE

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

($r = 0.614$, $P = 0.003$) (fig. 6*a*). Notably, the expressions of three outliers (conopeptides of O1- and O2- and P-superfamilies) (fig. 6*a*) in the correlation test of identical orthologs were highly divergent that upon their exclusion, the correlation increased to $r = 0.924$ (P value < 0.001). The expression of orthologs having higher rate of d_S to d_N was moderately though not significantly correlated ($r = 0.749$, $P = 0.02$) (fig. 6*b*), whereas the expression of orthologous pairs under positive selection showed weak correlation ($r = 0.585$, P value < 0.001) (fig. 6*c*).

Intraspecific Patterns of Conopeptide Expression

Most (75) of the identified conopeptide precursors in *C. tribblei* were found in all specimens, but some sequences were observed in either one or two individuals only (fig. 7*a* and supplementary table S6, Supplementary Material online). Similarly, although 34 superfamilies were identified in all specimens of *C. tribblei*, a few superfamilies were expressed in only one or two individuals (fig. 7*a* and supplementary table S6, Supplementary Material online). Interestingly, only less than half of the identified conopeptides in lenavati conopeptide data set were expressed in all three specimens (fig. 7*b* and supplementary table S6, Supplementary Material online). Specimen lena3 expressed 54 more conopeptides than the other two individuals (fig. 7*b* and supplementary table S6, Supplementary Material online). In addition, although 29 superfamilies were identified in all three individuals of *C. lenavati*, 7 superfamilies including A, C, I3, S and SF-05, SF-mi2, and SF-mi4 groups were expressed only in specimen lena3. Moreover, the number of orthologous conopeptide genes with identical mature regions among the individuals of *C. tribblei* and *C. lenavati* varied from 12 to 16 (fig. 7*c*).

The expression patterns of conopeptides among the specimens of *C. tribblei* were moderately correlated, whereas a higher correlation in the conopeptide expression was observed among the *C. lenavati* specimens (table 3). In addition, the expression levels of conopeptide gene superfamilies were highly correlated among individuals of *C. tribblei* as well as among specimens of *C. lenavati* (table 3). Despite the high correlation, the levels of expression of each conopeptide (and also each superfamily) were not the same among the specimens of each species (supplementary table S4, Supplementary Material online).

The similarity of the expression patterns of conopeptides across individuals of each species was evaluated using the ten most highly expressed conopeptides as reference (supplementary tables S7 and S8, Supplementary Material online). It was found that of the ten most highly expressed conopeptides in each specimen of *C. tribblei*, seven conopeptides were common among all individuals whereas other highly expressed conopeptides were observed in one or two specimens only (fig. 8*a*). Likewise, six of the ten most highly expressed conopeptides in each *C. lenavati* individual were similar in all

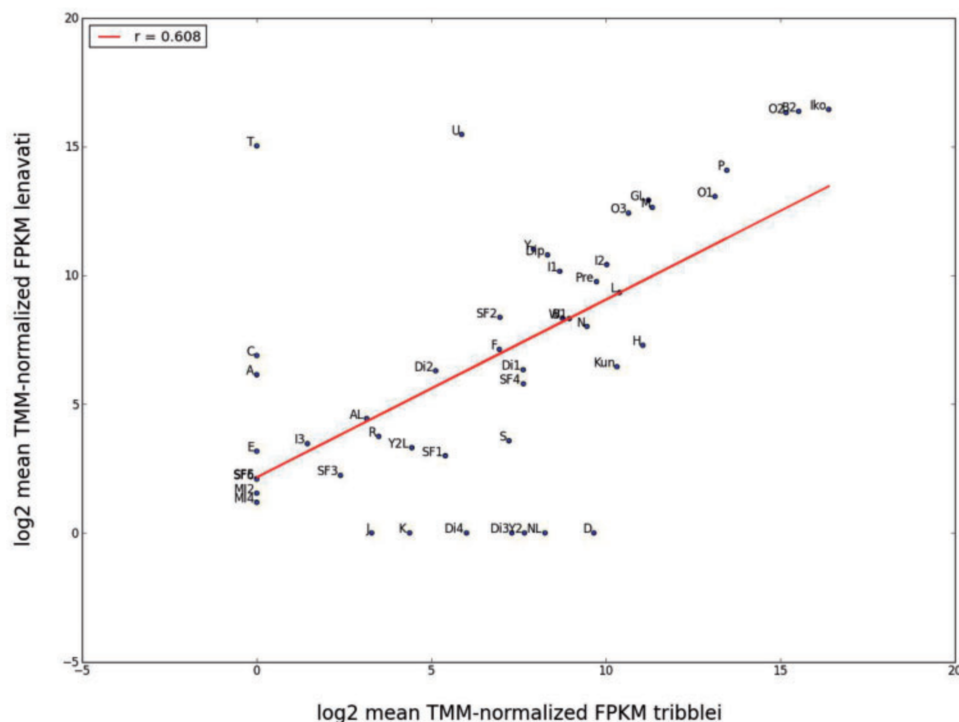

Fig. 4.—Correlation of the expression levels of conopeptide gene superfamilies in *C. tribblei* and *C. lenavati*. The Pearson's correlation coefficient (r) is 0.608 (P value < 0.001). The abbreviations of superfamily names are noted in figure 2.

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

Genome Biology and Evolution
SMBE

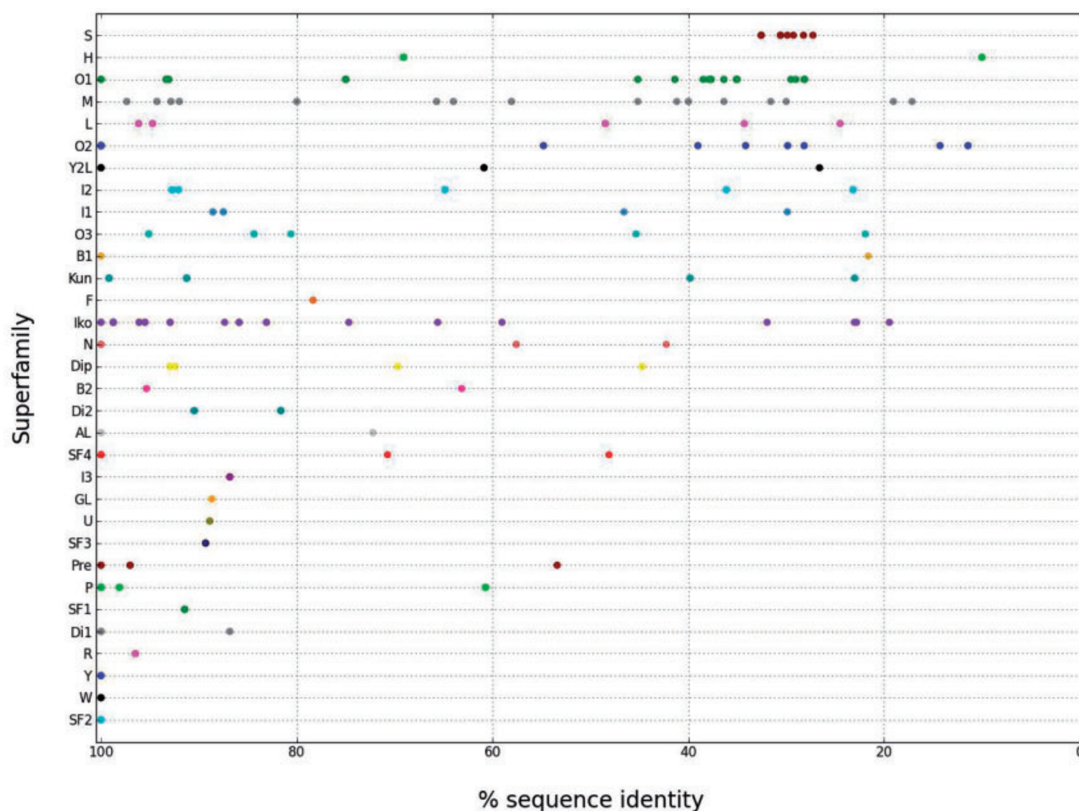

Fig. 5.—Distribution of the conopeptides in each gene superfamily based on the highest PID (the most similar match) of the conopeptides' mature regions in *C. tribblei* and *C. lenavati*. The abbreviations of superfamily names are noted in figure 2.

three individuals whereas the other most highly expressed conopeptides were different among specimens of *C. lenavati* (fig. 8*b*). The presence of a combination of highly expressed conopeptides that are common in all individuals and other conopeptides which are highly expressed in one or two specimens resulted in a distinct set of highly expressed conopeptides in each individual (fig. 8 and supplementary tables S7 and S8, Supplementary Material online). The majority of the highly expressed conopeptides in both species belonged to superfamilies and groups (O1, O2, B2, P, con-ikot-ikot, and G-like) with high expression levels (fig. 2*b*). Notably, conopeptides of several superfamilies such as O3, M, T, and U were among the highly expressed conopeptide in individuals of *C. lenavati* but not in *C. tribblei* (fig. 8 and supplementary tables S7 and S8, Supplementary Material online). Similarly, one H-superfamily conopeptide was highly expressed only in the individuals of *C. tribblei*. Interestingly, six of the ten highly expressed conopeptides in the individuals of *C. tribblei* and *C. lenavati* were orthologous conopeptide pairs identified between these species (fig. 8).

Functional Analysis of Transcriptomes

In addition to the conopeptides, a total of 41,875 and 39,516 transcripts in the tribblei and lenavati reference assemblies, respectively, exhibited high similarity to the proteins in the UniProtKB/Swiss-Prot database. Gene ontology terms were assigned to 34,876 transcripts in *C. tribblei*, and also to 32,778 transcripts in *C. lenavati* data sets. In both species, transcripts having binding and catalytic activities comprised a high percentage of the GO terms in the molecular function category, whereas cellular and metabolic processes and biological regulation were the most prominent categories in the biological process category in *C. tribblei* and *C. lenavati* (supplementary fig. S3, Supplementary Material online). Also, cell and organelle were the most abundant terms in the cellular component category in the tribblei and lenavati data sets.

Discussion

High Diversity of Expressed Conopeptide Genes

The venom of *C. tribblei* and *C. lenavati* had the highest number of conopeptide gene superfamilies and the most diverse venom complement observed in any *Conus* species thus far. Moreover, the conopeptide diversity even in the individuals of *C. tribblei* and *C. lenavati* was higher than in other *Conus* species indicating that the high conopeptide diversity and the venom complexity were consistent across the specimens. These observations also suggest that although using

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

Genome Biology and Evolution
SMBE

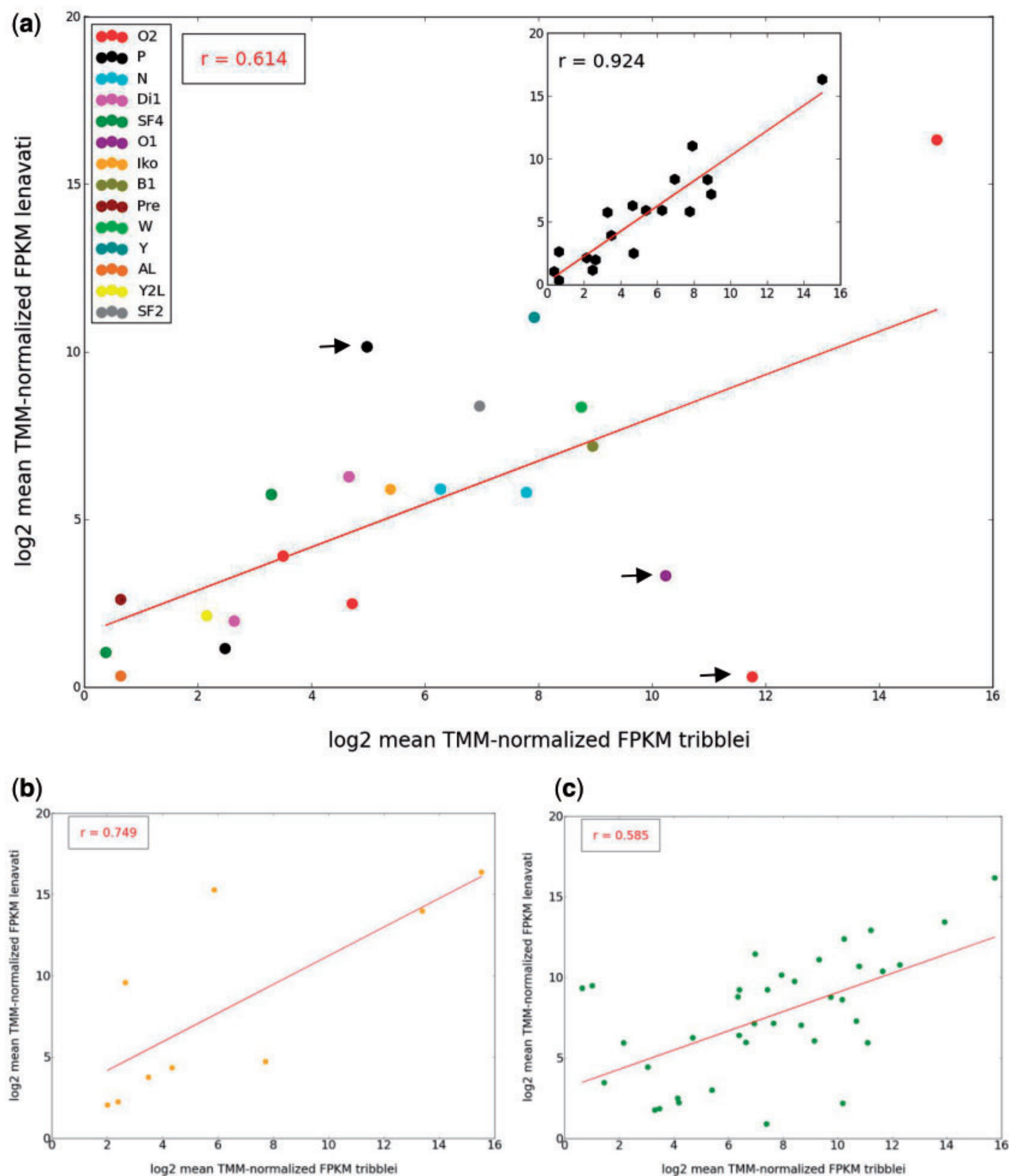

FIG. 6.—The expression of orthologous conopeptide pairs in *C. tribblei* and *C. lenavati*. (*a*) The correlation of the expression levels of identical orthologous conopeptide pairs ($r = 0.614$, $P = 0.003$). The correlation of the expression of orthologs pairs except for the three conopeptides with divergent expression patterns (marked with black arrows) ($r = 0.924$, P value < 0.001) is shown in insert. The correlation of orthologous conopeptide pairs (*b*) showing d_N to d_S ratios < 1 ($r = 0.749$, $P = 0.02$) and (*c*) d_N to d_S ratios > 1 ($r = 0.585$, P value < 0.001). The abbreviations of superfamily names are noted in figure 2.

pooled sample of several individuals in the transcriptome studies may increase the total number of identified conopeptides, it may not significantly affect the inferred diversity of the conopeptide superfamilies. The diversity of the conopeptide genes in cone snails reflects dietary specialization and prey types (Duda and Palumbi 2004; Duda et al. 2009); species with more exclusive prey types produce less diverse conopeptides (Remigio and Duda 2008). The remarkably high diversity of conopeptide gene superfamilies in *C. tribblei* and *C. lenavati* may thus indicate a relatively wide range of prey targets for these closely related species. In addition, identification of divergent superfamilies that have only been reported for the generalist feeder *C. californicus* (Elliger et al. 2011) is consistent with the hypothesis that the high diversity of conopeptides in *C. tribblei* and *C. lenavati* may be an adaptation to a wide taxonomic prey field. The conopeptide diversity in *C. tribblei*, *C. lenavati,* and other *Conus* species is compared in table 1; Although these estimates may to some degree be

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

Genome Biology and Evolution
SMBE

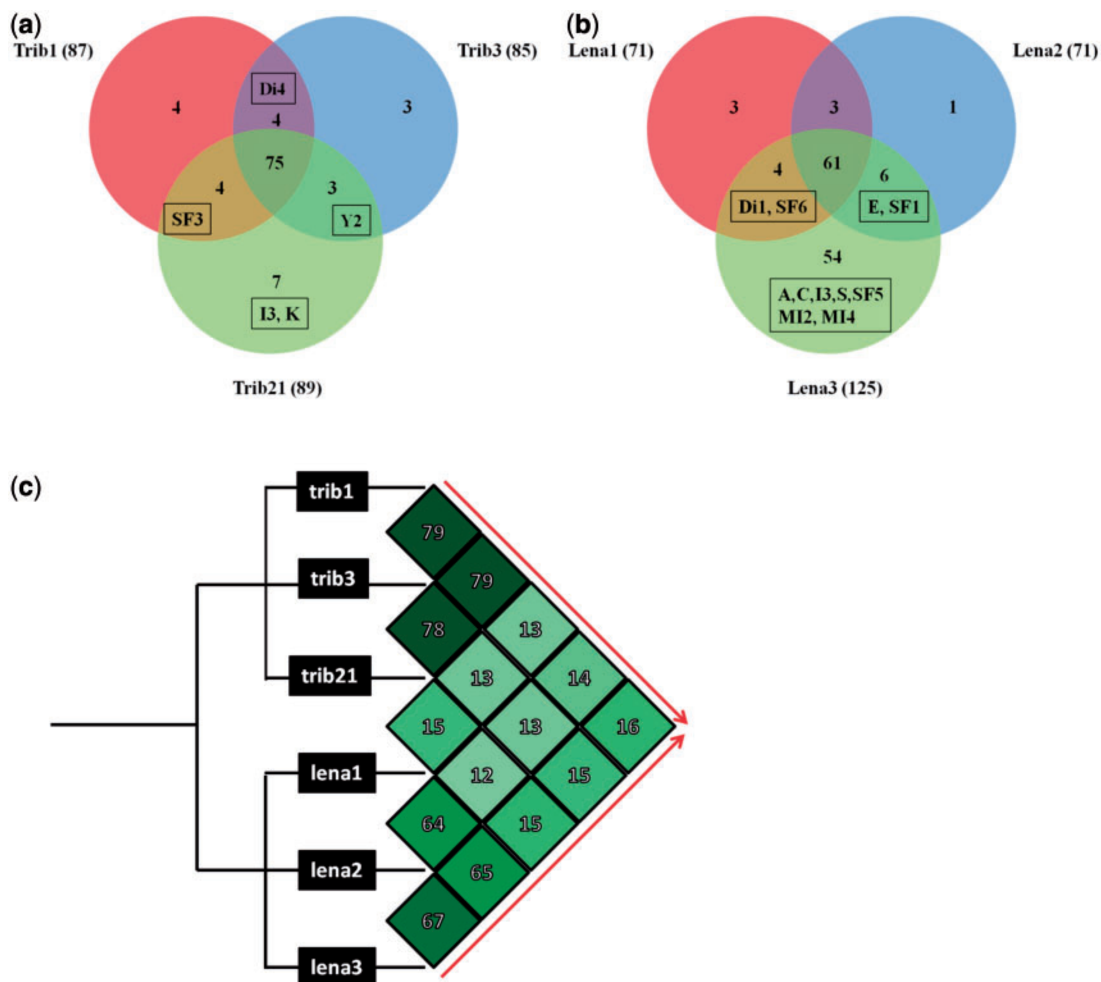

Fig. 7.—Shared and distinct conopeptides and gene superfamilies for specimens of (*a*) *C. tribblei* and (*b*) *C. lenavati*, (*c*) the number of identical conopeptide (considering only the mature regions) among the specimens of *C. tribblei* and *C. lenavati*. The total number of conopeptides in each specimen is shown in parenthesis. The superfamilies that are found in only one or two specimens are shown in box. The abbreviations of superfamily names are noted in figure 2. The number of identical conopeptides between a pair of specimens is shown in each diamond. To find the diamond corresponding to one pair of specimens, the lines extended from the dark boxes containing the specimen's name should be followed. The diamond, in which the columns leading from a pair of individuals intersect, shows the number of identical mature regions. For example, the diamond at the intersection of the two red arrows corresponds to trib1 and lena3. The phylogenetic relationship of the specimens shown here is not based on the actual Bayesian tree. The intensity of the colors of the diamonds corresponds to the number of identical conopeptides.

subject to inaccuracies (owing to such factors as sequencing platforms and bioinformatics tools used), the data clearly indicate the relatively high diversity of conopeptides in *C. tribblei* and *C. lenavati*.

The conopeptide gene superfamilies expressed in *C. tribblei* and also in *C. lenavati* had variable frequencies (fig. 2*a*). Similar observations were previously reported (Hu et al. 2011, 2012; Lluisma et al. 2012; Terrat et al. 2012; Dutertre et al. 2013; Lavergne et al. 2013; Robinson et al. 2014; Barghi et al. 2015). Surprisingly, the computed Shannon's evenness indices did not reflect such variability (tables 1 and 2). This incongruence may have arisen because except for a few predominant superfamilies, 19 and 16 conopeptide groups in *C. tribblei* and *C. lenavati*, respectively, were each represented by only one sequence. Hence, despite the high diversity of gene superfamilies, the frequencies of most of the conopeptide superfamilies were similar.

Patterns of Sequence Divergence in Conopeptide Superfamilies

The members of different conopeptide gene superfamilies have diverged differently after separation of closely related species *C. tribblei* and *C. lenavati* so that several divergence patterns have emerged (fig. 5). At one end, the divergence of

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

Table 3
Correlation of the Expression Levels of Conopeptides and Gene Superfamilies among Specimens of *Conus tribblei* and *Conus lenavati*

Conus tribblei			
	trib1	trib3	trib21
trib1	1	0.751	0.644
trib3	0.884	1	0.682
trib21	0.835	0.833	1
Conus lenavati			
	lena1	lena2	lena3
lena1	1	0.913	0.798
lena2	0.936	1	0.824
lena3	0.858	0.883	1

Note.—The Pearson's correlation coefficient for the conopeptides is shown in the cells above the diagonal and for the gene superfamilies below the diagonal. For all the correlation coefficients (r), the P value is less than 0.001.

conopeptides of some gene superfamilies with identical or highly similar conopeptides has occurred slowly. At the other end, few superfamilies (S and H) had diverged more. Between these two groups, several superfamilies (O1, O2, M, con-ikot-ikot, and conodipine) with high frequencies (fig. 2*a*) showed a wide range of divergence, from identical conopeptides to the conopeptides showing as low as 11% PID between *C. tribblei* and *C. lenavati*. Similarly, the sequence identity of O1-superfamily conotoxins in *Conus abbreviatus* and *Conus miliaris* ranged from 51.66% to 96.3% (Duda 2008).

Conserved and Species-Specific Expression Patterns of Conopeptide Genes

In both *C. tribblei* and *C. lenavati*, the expressions of different gene superfamilies had order-of-magnitude differences. Although half of the venom transcriptome consisted of transcripts from few highly expressed superfamilies, the majority of the conopeptide groups had moderate to low expression levels. Similarly, high variations in the expression patterns of different superfamilies have been observed in other species of *Conus* (Hu et al. 2011, 2012; Terrat et al. 2012; Dutertre et al. 2013; Robinson et al. 2014). High correlation (r = 0.938) between the frequency and the expression level of conopeptide superfamilies in *C. marmoreus* was observed (Dutertre et al. 2013). On the contrary, the frequency and the expression of different superfamilies in *C. tribblei* and *C. lenavati* were only moderately correlated (supplementary fig. S2, Supplementary Material online). This is because although O2-superfmily and con-ikot-ikot family were predominant components of the venom in terms of both the number of conopeptides and also the expression levels, other superfamilies such as P and B2 had extremely high expression levels despite containing only 1–4 conopeptides.

Comparison of the expression levels of conopeptide gene superfamilies between the two species revealed the presence of both conserved and species-specific expression patterns. The expression levels of superfamilies common in both species were highly correlated whereas several conopeptide superfamilies appeared to be species-specific and are only expressed in either *C. lenavati* or *C. tribblei*. The complement of gene superfamilies identified in each *C. tribblei* specimen was a relatively robust representative of the superfamilies characteristically expressed in this species. Of the 36 gene superfamilies identified in the transcriptome analysis of the pooled mRNA from the venom ducts of 20 *C. tribblei* specimens (Barghi et al. 2015), 32 were identified in all the three individuals of *C. tribblei* in this study, and 3 gene superfamilies were expressed in either 1 or 2 specimens. Moreover, none of the species-specific superfamilies of *C. lenavati* (T, A, C, E, SF-05, SF-06, SF-mi2, and SF-mi4) was identified in the pooled sample transcriptome of *C. tribblei* (Barghi et al. 2015). Whether similar observation is relevant for *C. lenavati* requires an analysis of the transcriptome of a pooled sample of *C. lenavati*. The conopeptides identified in this transcriptome analysis represent the genes expressed at a specific time and in a specific developmental stage of the organism. The sequencing depth in this study has allowed detection of a wide range of expression level (10–85,000 FPKM). Hence, if some conopeptides were not detected in this study, they must have been expressed at extremely low level. Furthermore, some conopeptide genes might be present in the genome but not expressed in the specific individuals sequenced in this study.

In general, the diet of each worm-hunting cone snail is dominated by a different polychaete, and even *Conus* species coexisting in a habitat specialize on different prey types (Kohn 2001). The distribution of these species in the Philippines is not well documented, but both species used in this study were collected in deep waters at the same location in Eastern Cebu, Sogod. *Conus tribblei* was collected at lower depth (100 m) than *C. lenavati* which is usually found at the depth of 130 m. Hence the evolution of species-specific expression patterns in *C. tribblei* and *C. lenavati* may indicate their divergence in range of conopeptide targets (prey, predators, or competitors) in their habitat.

Divergent Expression of Orthologous Conopeptide Genes

The expression pattern of identical (hence, orthologous) conopeptide genes in *C. tribblei* and *C. lenavati* reveals strong conservation of regulatory modes in the two species, although instances of divergence were noted (fig. 6*a*). On the other hand, the conopeptides under positive selection showed greater dissimilarity in expression in *C. tribblei* and *C. lenavati* and showed only slight correlation (fig. 6*c*). It appears that the conservation of sequence in the orthologous conopeptides of *C. tribblei* and *C. lenavati* is correlated with conservation of expression pattern. Our results elucidated that even among

SMBE

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

Genome Biology and Evolution

SMBE

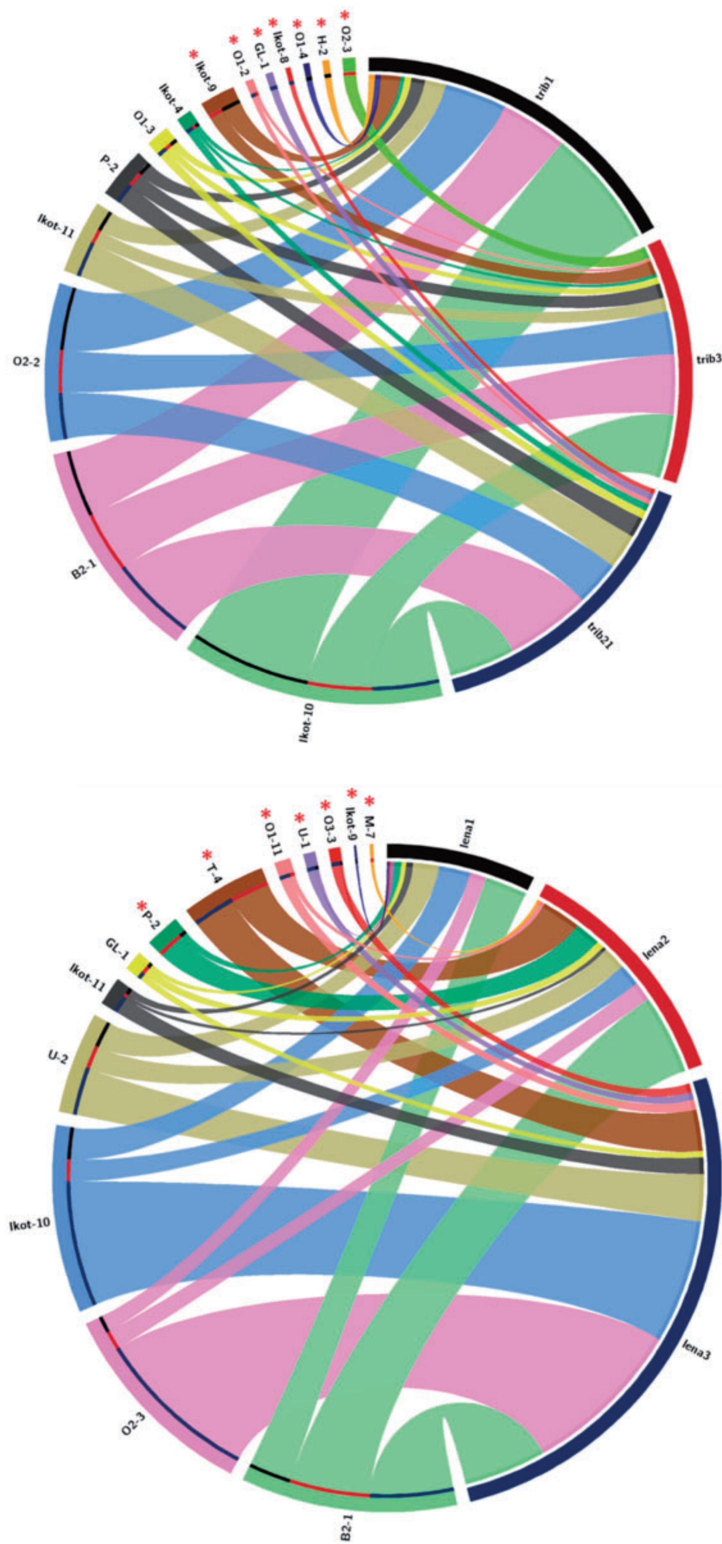

Fig. 8.—The ten most highly expressed conopeptides among the individuals of (*a*) *C. tribblei* and (*b*) *C. lenavati*. The thickness of ribbons corresponds to the expression level of conopeptides. The nomenclature of the conopeptide names is noted in figure 3, and the abbreviations of superfamily names are noted in figure 2. The conopeptides that are present in the highly expressed conopeptides of only one or two individuals are marked with red asterisks. The figures were constructed using Circos (Krzywinski et al. 2009). The orthologous conopeptide pairs are: B2-1, GL-1, P-2, Ikot-10, Ikot-11, Ikot-8 (*C. tribblei*), and Ikot-9 (*C. lenavati*).

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

Genome Biology and Evolution
SMBE

the orthologous conopeptide pairs, the divergence in pattern of expression can contribute to differences and uniqueness of venom composition in each species.

This study discovered a large number of orthologous conopeptide pairs (a total of 67) belonging to a wide range (31) of gene superfamilies in *C. tribblei* and *C. lenavati*. Duda and Remigio (2008) identified 12 sets of orthologous O1-superfamily genes among closely related worm-hunting *Conus* species. No orthologous pair was identified between recently diverged sister species *C. abbreviatus* and *Conus aristophanes*, eight pairs were identified in *C. abbreviatus* and *C. miliaris* which is separated from the lineage of *C. abbreviatus* and *C. aristophanes* (Duda and Remigio 2008). Additionally, few orthologs of A-superfamily genes were expressed in four closely related worm-hunting species (Chang and Duda 2012, 2014). Identification of very few orthologous conopeptide genes in the previous studies (Duda and Remigio 2008; Chang and Duda 2012, 2014) could be partly explained by the fact that only 1–2 conopeptide gene superfamilies were investigated. It must also be mentioned that because the previous studies (Duda and Remigio 2008; Chang and Duda 2014) used the traditional cDNA library method, the accuracy of the measurement of gene expression could have been limited, and some conopeptides with low expression levels could not have been detected.

Intraspecific Variation of Conopeptide Expression

The intraspecific variation of the number of conopeptides among the specimens of *C. tribblei* was low. This observation may be due to the similar size of *C. tribblei* individuals (~6 cm), which implies their similar developmental stage. A more pronounced intraspecific variation was observed among the specimens of *C. lenavati* mainly because of the expression of significantly higher number of conopeptides and the exclusive expression of several superfamilies in one specimen, lena3. This specimen (shell length 5 cm) was smaller than the other two specimens (8 cm) which may indicate an earlier developmental stage. Ontogenic changes in the expression of conopeptides have been previously documented in *Conus* (Safavi-Hemami et al. 2011). However, further analysis of the conopeptide expression in individuals of different developmental stages is needed to corroborate our results. The observed correlation of the conopeptide expression patterns among individuals of *C. tribblei* and of *C. lenavati* (table 3) indicates species-characteristic mechanism for regulating conopeptide expression, but the observed quantitative differences in levels of gene expression (supplementary table S4, Supplementary Material online) and specimen-specific set of highly expressed conopeptides (fig. 8) suggest that the expression levels of a subset of conopeptides can also be independently regulated by individual snails. All the previous studies describing intraspecific variation in *Conus* only studied the peptide masses in the venom (Jakubowski et al. 2005; Davis et al. 2009; Dutertre et al. 2010; Chun et al. 2012). Despite the variability in the peptide masses among individuals, the venom profile of the individuals for each species remains consistent (Jakubowski et al. 2005; Dutertre et al. 2010), and even the predominant peptides remain the same over time (Chun et al. 2012). These observations suggest that each species apparently evolves species-specific regulatory mechanisms for its venom peptides.

Evolution of *Conus* Exogenomes

Identification of positive selection in the conopeptide orthologs and the presence of several divergence patterns among the conopeptide superfamilies specifically those exhibiting very high divergence between *C. tribblei* and *C. lenavati* (fig. 5) provide further evidence that multiple molecular mechanisms synergistically act to cause differentiation of venom composition between species and drive the evolution of *Conus* exogenome. These extend previous observations on the molecular evolution and extremely high diversification rate of conopeptides through gene duplication and positive selection in piscivorous and vermivorous species (Duda and Palumbi 2000; Duda 2008; Duda and Remigio 2008; Puillandre et al. 2010; Chang and Duda 2012, 2014). The role of altered gene expression in evolution has been demonstrated, for example in morphological and behavioral changes in mammals (Enard et al. 2002). Divergence of gene expression and the protein sequence divergence both contribute to the adaptive evolution (Nuzhdin et al. 2004; Lemos et al. 2005; Khaitovich et al. 2006). Modifications of the expression level of toxin transcripts together with the diversity of isoforms in each toxin family were shown to account for the differentiation of the venom affecting prey specificity in species of vipers (Casewell et al. 2009). In this study, we showed that divergence in the control of conopeptide expression is another factor that may contribute to the differentiation of *Conus* venom composition. The divergent expression patterns of positively selected conopeptide orthologs in *C. tribblei* and *C. lenavati* (fig. 6) and the species-specific expression patterns of some superfamilies (figs. 2*b* and 4) contribute to the presence of unique venom complement in these species. Our results identified the divergence in the conopeptide expression as an important adaptation strategy in these two species; further studies will be necessary to uncover the molecular mechanisms of conotoxin gene regulation and exogenome evolution in *Conus.*

Conclusion

The results of this study highlight the role of sequence divergence, positive selection, and selective changes in the patterns of gene expression in the generation of conopeptide diversity and modification of venom composition over evolutionary time, and how these processes reflect the evolution of *Conus* exogenomes. These genome-level processes are among the factors that underlie the capacity of *Conus* species

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

Genome Biology and Evolution
SMBE

to exploit their taxonomic prey field or adapt to changing biotic conditions (e.g., changes in prey availability, or type of predators or competitors in their environment). However, other related processes (e.g., gene duplication, hypermutation, recombination, or chromosome-level mutations) are likely to play a significant role in this evolutionary process. To obtain deeper insights into the evolution of *Conus* exogenomes, it will be important to investigate how the processes described in this study and those related processes interact.

Supplementary Material

Supplementary figures S1–S3 and tables S1–S8 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The specimens used in this study were obtained in conjunction with a collection trip supported in part by ICBG grant #1U01TW008163. This study was supported by the Emerging Interdisciplinary Research Program of the University of the Philippines System through the Philippine Genome Center to AOL. The data analysis was carried out using the High-Performance Computing Facility of the Advanced Science and Technology Institute and the Philippine e-Science Grid, Diliman, Quezon City. The authors thank Noel Saguil for the help in sample collection and Alexander Fedosov for the help in the phylogenetic analysis.

Literature Cited

Abdel-Rahman MA, Omran MAA, Abdel-Nabi IM, Ueda H, McVean A. 2009. Intraspecific variation in the Egyptian scorpion *Scorpio maurus palmatus* venom collected from different biotopes. Toxicon 53:349–359.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Barghi N, Concepcion GP, Olivera BM, LLuisma AO. 2015. High conopeptide diversity in *Conus tribblei* revealed through analysis of venom duct transcriptome using two high-throughput sequencing platforms. Mar Biotechnol. 17(1):81–98.

Barlow A, Pook CE, Harrison RA, Wüster W. 2009. Coevolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution. Proc R Soc Lond B Biol Sci. 276:2443–2449.

Campanella JJ, Bitincka L, Smalley J. 2003. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics 4:29.

Casewell NR, Harrison RA, Wüster W, Wagstaff SC. 2009. Comparative venom gland transcriptome surveys of the saw-scaled vipers (Viperidae: *Echis*) reveal substantial intra-family gene diversity and novel venom transcripts. BMC Genomics 10:564.

Chang D, Duda TF. 2012. Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. Mol Biol Evol. 29:2019–2029.

Chang D, Duda TF. 2014. Application of community phylogenetic approaches to understand gene expression: differential exploration of venom gene space in predatory marine gastropods. BMC Evol Biol. 14:123.

Chun JBS, et al. 2012. Cone snail milked venom dynamics—a quantitative study of *Conus purpurascens*. Toxicon 60:83–94.

Conticello SG, et al. 2001. Mechanisms for evolving hypervariability: the case of conopeptides. Mol Biol Evol. 18:120–131.

Daltry JC, Wüster W, Thorpe RS. 1998. Intraspecific variation in the feeding ecology of the crotaline snake *Calloselasma rhodostoma* in Southeast Asia. J Herpetol. 32:198–205.

Davis J, Jones A, Lewis RJ. 2009. Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. Peptides 30:1222–1227.

Duda TF. 2008. Differentiation of venoms of predatory marine gastropods: divergence of orthologous toxin genes of closely related *Conus* species with different dietary specializations. J Mol Evol. 67:315–321.

Duda TF, Chang D, Lewis B, Lee T. 2009. Geographic variation in venom allelic composition and diets of the widespread predatory marine gastropod *Conus ebraeus*. PLoS One 4(7):e6245.

Duda TF, Kohn AJ, Palumbi SR. 2001. Origins of diverse feeding ecologies within *Conus*, a genus of venomous marine gastropods. Biol J Linn Soc. 73:391–409.

Duda TF, Palumbi SR. 2000. Evolutionary diversification of multigene families: allelic selection of toxins in predatory cone snails. Mol Biol Evol. 17:1286–1293.

Duda TF, Palumbi SR. 2004. Gene expression and feeding ecology: evolution of piscivory in the venomous gastropod genus *Conus*. Proc R Soc Lond B Biol Sci. 271:1165–1174.

Duda TF, Remigio EA. 2008. Variation and evolution of toxin gene expression patterns of six closely related venomous marine snails. Mol Ecol. 17:3018–3032.

Dutertre S, Biass D, Stöcklin R, Favreau P. 2010. Dramatic intraspecimen variations within the injected venom of *Conus consors*: an unsuspected contribution to venom diversity. Toxicon 55:1453–1462.

Dutertre S, et al. 2013. Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom. Mol Cell Proteomics. 12:312–329.

Elliger CA, et al. 2011. Diversity of conotoxin types from *Conus californicus* reflects a diversity of prey types and a novel evolutionary history. Toxicon 57:311–322.

Enard W, et al. 2002. Intra- and Interspecific variation in primate gene expression patterns. Science 296:340–343.

Espiritu DJ, et al. 2001. Venomous cone snails: molecular phylogeny and the generation of toxin diversity. Toxicon 39:1899–1916.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol. 3:294–299.

Gibbs AL, Sanz L, Sovic MG, Calvete JJ. 2013. Phylogeny-based comparative analysis of venom proteome variation in a clade of rattlesnakes (*Sistrurus* sp.) PLoS One 8:e67220.

Götz S, et al. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36:3420–3435.

Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol. 29:644–654.

Haas BJ, et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protocols. 8:1494–1512.

Hall TA. 1999. BioEdit, a user-friendly biological sequence alignment editor and analysis program for Windows 95, 98, NT. Nucleic Acids Symp Ser. 41:95–98.

Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. Proc R Soc Lond B. 270:313–321.

Hu H, Bandyopadhyay PK, Olivera BM, Yandell M. 2011. Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. BMC Genomics 12:60.

Hu H, Bandyopadhyay PK, Olivera BM, Yandell M. 2012. Elucidation of the molecular envenomation strategy of the cone snail *Conus geographus*

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015

Genome Biology and Evolution
SMBE

through transcriptome sequencing of its venom duct. BMC Genomics 13:284.

Jakubowski JA, Kelley WP, Sweedler JV, Gilly WF Schulz JR. 2005. Intraspecific variation of venom injected by fish-hunting *Conus* snails. J Exp Biol. 208:2873–2883.

Jin A, et al. 2013. Transcriptomic messiness in the venom duct of *Conus miles* contributes to conotoxin diversity. Mol Cell Proteomics. 12(12):3824–3833.

Kaas Q, Westermann J, Craik DJ. 2010. Conopeptide characterization and classifications: an analysis using ConoServer. Toxicon 55:1491–1509.

Kaas Q, Yu R, Jin A, Dutertre S, Craik DJ. 2012. ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. Nucleic Acids Res. 40:D325–D330.

Khaitovich P, Enard W, Lachmann M, Pääbo S. 2006. Evolution of primate gene expression. Nat Rev Genet. 7:693–702.

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

Krzywinski MI, et al. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19:1639–1645.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357–360.

Larkin MA, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948.

Lavergne V, et al. 2013. Systematic interrogation of the *Conus marmoreus* venom duct transcriptome with ConoSorter reveals 158 novel conotoxins and 13 new gene superfamilies. BMC Genomics 14:708.

Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL. 2005. Evolution of proteins and gene expression levels are coupled in *Drosophila* and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. Mol Biol Evol. 22(5):1345–1354.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. Bioinformatics 12:323.

Lluisma AO, Milash BA, Moore M, Olivera BM, Bandyopadhyay PK. 2012. Novel venom peptides from the cone snail *Conus pulicarius* discovered through next-generation sequencing of its venom duct transcriptome. Mar Genomics. 5:43–51.

Menezes MC, Furtado MF, Travaglia-Cardoso SR, Camargo ACM, Serrano SMT. 2006. Sex-based individual variation of snake venom proteome among eighteen *Bothrops jararaca* siblings. Toxicon 47:304–312.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), 2010 Nov 14; New Orleans, LA. p. 1–8.

Milne I, et al. 2012. Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform. 14:193–202.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-seq. Nat Methods. 5:621–628.

Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM. 2004. Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. Mol Biol Evol. 21(7):1308–1317.

Olivera BM, et al. 1999. Speciation of cone snails and interspecific hyperdivergence of their venom peptides. Ann N Y Acad Sci. 870:223–237.

Olivera BM. 2006. *Conus* peptides: biodiversity-based discovery and exogenomics. J Biol Chem. 281:31173–31177.

Olivera BM, Teichert RW. 2007. Diversity of the neurotoxin *Conus* peptides, a model for concerted pharmacological discovery. Mol Interv. 7:251–260.

Pearson WR, Wood T, Zhang Z, Miller W. 1997. Comparison of DNA sequences with protein sequences. Genomics 46:24–36.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 8:785–786.

Puillandre N, Duda TF, Meyer C, Olivera BM, Bouchet P. 2015. One, four or 100 genera? A new classification of the cone snails. J Mollus Stud. 81(1):1–23.

Puillandre N, et al. 2014. Molecular phylogeny and evolution of the cone snails (Gastropoda, Conoidea). Mol Phylogenet Evol. 78:290–303.

Puillandre N, Watkins M, Olivera BM. 2010. Evolution of *Conus* peptide genes: duplication and positive selection in the A-superfamily. J Mol Evol. 70:190–202.

Rambaut A, Drummond AJ. 2007. Tracer v1.4. Available from: http://beast.bio.ed.ac.uk/Tracer

Remigio EA, Duda TF. 2008. Evolution of ecological specialization and venom of a predatory marine gastropod. Mol Ecol. 17:1156–1162.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11:R25.

Robinson SD, et al. 2014. Diversity of conotoxin gene superfamilies in the venomous snail, *Conus victoriae*. PLoS One 9:e87648.

Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. Nat Rev Genet. 13:505–516.

Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61(3):539–542.

Safavi-Hemami H, et al. 2011. Embryonic toxin expression in the cone snail *Conus victoriae*: primed to kill or divergent function? J Biol Chem 286(25):22546–22557.

Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731–2739.

Terrat Y, et al. 2012. High-resolution picture of a venom duct transcriptome: case study with the marine snail *Conus consors*. Toxicon 59:34–46.

Trapnell C, et al. 2010. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 28:511–515.

Tucker JK, Tenorio MJ. 2009. Systematic classification of recent and fossil Conoidean gastropods. Hackenheim (Germany): Conchbooks.

UniProt Consortium. 2015. UniProt: a hub for protein information. Nucleic Acids Res. 43:D204–D212.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13:555–556.

Ye J, et al. 2006. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. 34(Web service issue):293–297.

Associate editor: Michael Purugganan

Downloaded from http://gbe.oxfordjournals.org/ by guest on July 19, 2015