Extensive and Continuous Duplication Facilitates Rapid Evolution and Diversification of Gene Families

Dan Chang*[,1,2] and Thomas F. Duda Jr[1,3]

[1]Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan

[2]Department of Statistics, University of Michigan

[3]Smithsonian Tropical Research Institute, Balboa, Ancón, Republic of Panama

***Corresponding author:** E-mail: changdan@umich.edu.

Associate editor: Todd Oakley

Abstract

The origin of novel gene functions through gene duplication, mutation, and natural selection represents one of the mechanisms by which organisms diversify and one of the possible paths leading to adaptation. Nonetheless, the extent, role, and consequences of duplications in the origins of ecological adaptations, especially in the context of species interactions, remain unclear. To explore the evolution of a gene family that is likely linked to species associations, we investigated the evolutionary history of the A-superfamily of conotoxin genes of predatory marine cone snails (*Conus* species). Members of this gene family are expressed in the venoms of *Conus* species and are presumably involved in predator–prey associations because of their utility in prey capture. We recovered sequences of this gene family from genomic DNA of four closely related species of *Conus* and reconstructed the evolutionary history of these genes. Our study is the first to directly recover conotoxin genes from *Conus* genomes to investigate the evolution of conotoxin gene families. Our results revealed a phenomenon of rapid and continuous gene turnover that is coupled with heightened rates of evolution. This continuous duplication pattern has not been observed previously, and the rate of gene turnover is at least two times higher than estimates from other multigene families. Conotoxin genes are among the most rapidly evolving protein-coding genes in metazoans, a phenomenon that may be facilitated by extensive gene duplications and have driven changes in conotoxin functions through neofunctionalization. Together these mechanisms led to dramatically divergent arrangements of A-superfamily conotoxin genes among closely related species of *Conus*. Our findings suggest that extensive and continuous gene duplication facilitates rapid evolution and drastic divergence in venom compositions among species, processes that may be associated with evolutionary responses to predator–prey interactions.

Key words: duplication, rapid evolution, *Conus*, species interaction, gene turnover.

Introduction

Gene duplication plays a crucial role in organismal evolution (Ohno 1970) as it facilitates increases in genetic and functional diversities (Hughes 1994; Zhang 2003), contributes to gene dosage effects (Kondrashov et al. 2002; Gevers et al. 2004; Perry et al. 2007), and can instigate reproductive isolation through the origin of Dobzhansky–Muller incompatibilities (Lynch and Conery 2000). Several works have described mechanisms of gene duplication (Zhang 2003), fates of duplicated genes (Lynch and Conery 2000; Conant and Wolfe 2008), and correlation of duplicability with factors such as adaptability and functional constraints (Conant and Wolfe 2008). Models of gene family evolution present alternative viewpoints on the neutrality of duplication and functional fates of gene duplicates (Innan and Kondrashov 2010). In particular, gene duplication has been proposed to be adaptive when organisms are confronted with ecological stress because it leads to either dosage benefits or neofunctionalization of duplicated copies (Kondrashov et al. 2002). Gene duplication has been found in many genes that are involved in ecological adaptation to abiotic changes, such as the *Dca* gene that is involved in adaptation to lower temperature in *Drosophila* (Arboleda-Bustos and Segarra 2011) and several members of *CspA* gene family for cold shock in *Escherichia coli* (Yamanaka et al. 1998). Gene duplication is also associated with many types of species interactions and plays an important role in the generation of genetic diversity in the context of biotic changes, such as the insect nitrile-specifier protein-like gene family (Fischer et al. 2008) and P450 genes (Wen et al. 2006) that are associated with adaptations to cope with the chemical defenses of plants and major histocompatibility complex (Burri et al. 2010) and immunoglobulin (Guldner et al. 2004) gene families that are involved in host–pathogen interactions.

The role of gene duplication in predator–prey interactions can be investigated in systems in which the traits associated with the interactions can be characterized genetically. Many taxa, including cnidarians, numerous arthropod species, conoidean gastropods, and various vertebrates, use venoms to capture prey or defend against predators, and in most cases, these venoms contain peptide neurotoxins that directly target various ion channels and cell receptors (Daltry et al. 1996; Olivera 2002; Fry et al. 2003, 2006; Moran et al. 2008; Binford et al. 2009). Gene duplication and positive selection have been

documented for genes expressed in venoms of a variety of venomous taxa, including cone snails (Duda and Palumbi 1999b, 2004; Conticello et al. 2001; Duda and Remigio 2008; Puillandre et al. 2010), spiders (Binford et al. 2009), and snakes (Fry et al. 2003; Juarez et al. 2008). Neurotoxic peptides in the venoms of predatory marine snails *Conus* (i.e., conotoxins) are utilized primarily for prey capture (Kohn 1959; Olivera 2002), and thus, the evolution of conotoxins is presumably driven by the evolution of resistance in prey (i.e., represents a coevolutionary arms race) and/or shifts in prey utilization patterns (Duda and Palumbi 1999b). Conotoxins are encoded by various gene families that contain some of the fastest evolving genes of metazoans (Duda and Palumbi 1999b). Previous studies on allelic variants of conotoxin genes and dietary divergence among geographical locations have shown strong associations between venom diversity and dietary specializations (Duda and Lee 2009; Duda et al. 2009). Nonetheless, because most past studies of conotoxin evolution have relied on analyses of expressed conotoxin genes (i.e., mRNA sequences), little is known about the frequencies and patterns of gene duplication and loss or the effects of these phenomena on the evolution of conotoxin gene families. Our study effectively fills this gap through examination of conotoxin gene sequences recovered from genomic DNA. Our genome-based investigation of conotoxin gene family evolution represents a large advance from previous studies that relied on venom transcripts because *Conus* species do not appear to express orthologous gene copies (Duda and Remigio 2008).

We determined sequences of A-superfamily conotoxin genes of four closely related, worm-eating *Conus* species: *Conus lividus*, *Conus diadema*, *Conus quercinus*, and *Conus sanguinolentus*. These genes encode α-conotoxin peptides that are selective inhibitors of nicotinic acetylcholine receptors (McIntosh et al. 1999). The α-conotoxins are distinguished from other conotoxin types by their particular cysteine backbone that occurs in the pattern of "C1C2$(X)_n$C3$(X)_n$C4," with various numbers of amino acids (denoted as $(X)_n$) between the second and third (C2 and C3) and third and fourth (C3 and C4) cysteine residues (Santos et al. 2004). Miocene fossil records suggest that *C. lividus* and *C. quercinus* diverged about 11 Ma, and phylogenetic studies indicate that *C. lividus*, *C. diadema*, and *C. sanguinolentus* diverged more recently (Duda and Kohn 2005), a situation that enables us to evaluate the rate of gene turnover across distinct time intervals. These four species are broadly distributed in the Indo-West Pacific (*C. lividus*, *C. sanguinolentus*, and *C. quercinus*) or Eastern Pacific (*C. diadema*) and exhibit distinct dietary characteristics (Kohn 1968, 2001; Nybakken 1978; Duda et al. 2001). Here we reconstructed the evolutionary history of A-superfamily conotoxin genes from these species, estimated rates of gene duplication and gene losses, evaluated the trajectories of rates of evolution after duplication, and predicted the functional fates of these genes. Based on past observations of high rates of evolution of conotoxin genes (Duda and Palumbi 1999b) and strong differences in expression profiles of conotoxins among closely related species (Duda and Palumbi 2004), we predict that rates of gene turnover are highly elevated within *Conus* and that increases in gene copy number facilitate the rapid evolution of conotoxin genes as well as the divergence of venom compositions among species.

Materials and Methods

Specimens and Genomic DNA Extraction

Specimens of *C. lividus* collected in Hawaii, *C. diadema* in Panama, *C. sanguinolentus* in American Samoa, and *C. quercinus* from Hawaii provided by J-P Bingham (University of Hawaii) were deposited in the collections of the Mollusk Division of the University of Michigan Museum of Zoology. Body tissues were preserved in 95% ethanol. Venom ducts were preserved in RNAlater (Ambion, Inc.) and stored at −20 °C. We extracted genomic DNA from the foot tissue of two individuals each of *C. lividus* and *C. diadema*, venom ducts of two individuals of *C. quercinus*, and the foot tissue of one individual of *C. sanguinolentus* using the E.Z.N.A Mollusc DNA kit (Omega Bio-Tek, Doraville, GA).

Phylogenetic Relationships of Four *Conus* Species and Molecular Clock Analyses

We amplified mitochondrial cytochrome c oxidase subunit I (COI) sequences from genomic DNA of our samples with the universal primers LCO1490 and HCO2198 (Folmer et al. 1994) and sequenced the polymerase chain reaction (PCR) products in both directions at the University of Michigan DNA Sequencing Core facilities. Sequences of a calmodulin intron and a β-tubulin intron were amplified from genomic DNA of each individual with exon priming, intron crossing primers for a 262 nt region of the calmodulin gene with primers described in Duda and Palumbi (1999a) and a 523 nt intron region of the β-tubulin gene (forward primer 5′CTGCGACTGTCTGCAAGGTATCG3′ and reverse primer 5′GAATGCGTCAGCTGGAAACCTGC3′). PCR products of calmodulin and tubulin introns were ligated into TA cloning vectors, which were then transformed into competent *E. coli* using The Original TA Cloning Kit with Top 10 Competent Cells (Invitrogen). We screened colonies for expected insert sizes with vector primers and sequenced those with appropriately sized inserts. Chromatograms were examined in Sequencher version 4.8 (Gene Codes Corporation), and sequences were manually aligned in Se-Al v2.0a11 (Rambaut 2002). We performed model selection on COI sequences in jModelTest 0.1.1 (Guindon and Gascuel 2003) (number of substitution schemes = 11, including models with unequal base frequencies, invariable sites, rate variation among sites, and maximum-likelihood tree for likelihood calculations), and the best models suggested by Akaike's information criterion (Akaike 1974) and Bayesian information criterion (Schwarz 1978) were selected for phylogenetic analyses. We constructed phylogenetic trees of mitochondrial COI sequences of our samples and two outgroup species (*C. catus* and *C. lorenzianus*) (GenBank accession numbers AY588194 and AY588163) with maximum-likelihood approaches in PAUP 4.0 (Swofford 2002)

(heuristic search with the Nearest Neighbor Interchange swapping on best trees only) and 1,000 bootstrap replicates and with Bayesian methods in MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001) (10,000,000 generations, 4 Markov chains, 2 runs, and 200 absolute burnin). We measured distances of intron sequences of the calmodulin and tubulin loci among these four species using the Jukes–Cantor model (Jukes and Cantor 1969) with uniform rates. Because of the existence of potential paralogs of tubulin and calmodulin sequences in our target species, we utilized the minimum genetic distances of these sequences among species for phylogenetic and molecular clock analyses as well as to identify orthologous and paralogous conotoxin loci (see below).

To avoid the influence of outgroups on estimation of divergence time, we constructed the phylogeny from analyses of mitochondrial COI sequences of our target species only and rooted the tree with the COI sequence from *C. quercinus*. A maximum-likelihood test of the molecular clock hypothesis was performed in MEGA 5.05 (Tamura et al. 2011) using the best model for COI sequences. Bayesian estimations of divergence times of the four species under strict and relaxed molecular clock model (uncorrelated lognormal) (Drummond et al. 2006) were conducted in BEAST v1.6.1 (Drummond and Rambaut 2007) with Markov chain Monte Carlo (MCMC) analyses of 20,000,000 generations (log parameters every 1,000 generations), using the divergence time of 11 My between *C. lividus* and *C. quercinus* as the reference calibration. The XML input files for BEAST analyses were created in software BEAUti v1.6.1 included in the BEAST package (prior t_{MRCA} and the root height of four species set to lognormal distribution with mean of ln(11) and standard deviation of 0.01). Both COI sequences and concatenated COI and calmodulin and tubulin intron sequences of these four species were used to build species tree under strict and relaxed clock models (with partition of substitution models and evolution rates in separate gene regions of concatenated sequences) and to infer time of divergence of these species. Output log files from BEAST were analyzed in Tracer v1.5 (Rambaut and Drummond 2007) to evaluate convergence, and tree files were imported into TreeAnnotator v1.6.1 in the BEAST package (first 1,000 results were burnin, posterior probability limit set to 0.5, mean node heights estimation) to build the maximum clade credibility tree and summarize the time estimations.

Recovery of A-superfamily Genes from Genomic DNA and Phylogenetic Analyses

To attempt to recover all A-superfamily genes from the genomes of these species, we designed ten sets of primers based on alignments of expressed A-superfamily gene sequences of more than 100 *Conus* species (supplementary table S1, Supplementary Material online). The primers correspond to 1) a relatively conserved sequence region downstream of a known intron position and upstream of the toxin-coding region and 2) a highly conserved region of the 3′ untranslated region. We used these primers to amplify A-superfamily genes from genomic DNA of each individual, cloned the amplification products, screened the resultant colonies with M13 primers, and sequenced suspected A-superfamily gene inserts to recover as many unique conotoxin gene sequences as possible. We repeated the whole procedure to recognize putative amplification or cloning-induced artifactual sequences (Duda and Remigio 2008). Sequence diversity curves (Duda and Remigio 2008) were generated for each round of amplification of each individual to evaluate whether enough inserts were sequenced to potentially recover all A-superfamily genes from the genomes of these species.

All sequences were manually aligned in SE-AL v2.0a11 (Rambaut 2002) based on examination of nucleotides and translated amino acids among sequences (especially the conserved cysteine backbone). Sequences recovered in both rounds of experiments or from both individuals of the same species were considered to represent non-artifactual sequences. We constructed a neighbor-joining tree of all nucleotide sequences recovered with the K80 model (Kimura 1980) in PAUP 4.0 (Swofford 2002) to confirm that no distinct clade included solely artifactual sequences and all artifactual sequences were eliminated from the data set for subsequent analyses. We constructed gene trees using maximum-likelihood methods and Bayesian methods as described above. We used an A-superfamily gene sequence from *C. catus* (GenBank accession number FJ868066) to root the tree, based on the previous studies on phylogenetic relationships of *Conus* species (Duda and Kohn 2005) and the evolutionary trajectories of α-conopeptides (Puillandre et al. 2010).

Determination of Orthology and Inference of Duplication and Loss

We identified sets of sequences that exhibited clustering patterns in the genealogy that resemble the topology of the species tree. We measured synonymous divergence (d_S) among conotoxin loci with the modified Nei–Gojobori method with the Jukes–Cantor model (Nei and Kumar 2000). Any pairs of sequences with d_S not exceeding the minimum genetic distances determined from introns of the calmodulin and tubulin loci among respective species were regarded as putative orthologs, with the assumption that rates of synonymous divergence of the calmodulin and tubulin loci are roughly equivalent to that of orthologous conotoxin sequences. We used orthology and counts of alleles from each individual (i.e., no more than two alleles for one locus in each individual) to determine the number of unique loci present in each species.

We constructed a Bayesian consensus phylogeny with DNA sequences of single alleles representing each unique locus (a "reduced" gene tree) as described above. Reconciliation of the species tree and reduced gene tree was performed in Notung 2.6 (Durand et al. 2006; Vernot et al. 2007) to estimate all possible gene duplication and loss events and the timings of these events. To avoid overestimation of turnover caused by poorly supported clades (clades with posterior probabilities less than 0.9), we

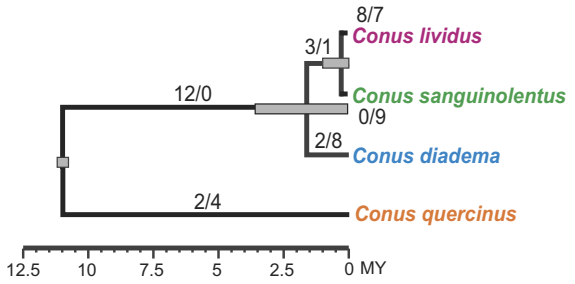

Fig. 1. The species tree of four *Conus* species and estimated times of divergence from relaxed molecular clock analyses of concatenated sequences of a region of the mitochondrial COI gene, a tubulin intron, and a calmodulin intron. The gray bars at each node represent 95% highest posterior density (HPD) of the time of separation. The scale bar at the bottom represents the estimated time of divergence in units of My. Estimated numbers of gene duplication (before the forward slash) and gene loss (after the forward slash) events are indicated on each branch of the linearized species tree. Color-coding scheme of species names: purple for *Conus lividus*, green for *Conus sanguinolentus*, blue for *Conus diadema*, and red for *Conus quercinus*.

utilized Notung to produce resolved alternative topologies and reported the minimum estimation of duplications and losses. Pseudogenes were identified based on the presence of premature stop codons or nonsynonymous substitutions in any of the four cysteine codons recovered from the four species.

To verify the inference of duplications/losses from the reconciliation analyses, we performed Bayesian rates estimation of duplications and losses in PrIME-GSR 1.0 (Åkerborg et al. 2009). This approach reconstructs and reconciles gene trees simultaneously with prior knowledge of the species tree, substitution model, molecular clock model of gene sequences, and gene duplication/loss process. We utilized the species topology and branch times of the relaxed molecular clock analyses of concatenated sequences (fig. 1) and the K80 + *G* model (equal base frequencies, $\kappa = 1.6721$, $\alpha = 0.813$), one of the best models selected for conotoxin gene sequences. We set the relaxed clock model of gene sequences to independent identical lognormal and the prior duplication/loss rates to 0.8 and performed MCMC analyses in two parallel chains of 700,000 generations each (logging every 1,000 generations). We analyzed the output file with the PERL script "mcmc_analysis.pl" included in the PrIME program and exported results of posterior probabilities and rates parameters. Convergence tests of two chains were performed with *geweke.diag* and *heidel.diag* functions implemented in CODA package in R (Plummer et al. 2006). Mean and ranges of the duplication rates and loss rates were summarized after removal of results of the first 500,000 generations.

Simulations of Sampling Effects on Inference of Gene Duplications

Limited by the unavailability of a complete *Conus* genome and our experimental approach, it is possible that we failed to identify some A-superfamily genes even though sequence diversity curves were saturated. It is unclear whether missing certain "essential" gene or gene combinations would affect the pattern and estimation of gene duplications and thus limit our evaluation of the effects of gene birth. To evaluate the impact of possible incomplete sampling on our estimates of gene duplication, we conducted several simulations. First, we randomly selected a set of genes from the gene pool of the four species, deleted these genes from the gene tree in PAUP, reconciled the pruned gene tree with the species tree with Notung, and estimated the overall duplication events and duplications after separation of *C. diadema*. The whole process was automated in PERL. We repeated the trial for 100 random combinations of excluded sequences (removal of single gene is exhaustive and we evaluated effects of every unique gene removal trial). In addition to random removal from genes of four species, we also conducted trials in which a proportion of genes were randomly removed from each species.

Estimation of Rates of Evolution of Contoxin Gene Paralogs and Orthologs

We tested a strict molecular clock hypothesis for conotoxin genes with the same approach as described above using the reduced Bayesian consensus tree and Hasegawa–Kishino–Yano (HKY) (Hasegawa et al. 1985) + *I* + *G* model. We calculated d_S of the prepro and toxin-coding regions and d_N of toxin-coding regions of paralogous loci of each species using the modified Nei–Gojobori method with Jukes–Cantor correction (Nei and Kumar 2000) in MEGA 4 (Tamura et al. 2007). Based on the assumption that the rate of synonymous substitution is constant, we estimated the synonymous substitution rate to be 0.004 per My. Then we calculated the time of separation of each pair of paralogs by dividing the pairwise d_S with 0.004 and estimated the rate of evolution by dividing pairwise d_N values with corresponding estimates of time of divergence. Because the genus *Conus* dates back to 55 Ma (Kohn 1990), we calculated mean rates of evolution of each species by averaging rates of the pairs with $d_S \leq 0.2$ (representing 50 My). We calculated d_N and d_S of identified orthologous loci and estimated the rates of nonsynonymous substitution of orthologous loci of *C. diadema* and *C. lividus* by dividing d_N by two times 1.6 My (divergence time estimated from molecular clock analyses as described above).

Ancestral Sequence Reconstruction and Tests of Positive Selection

Ancestral sequences of each node were reconstructed with the likelihood-based empirical Bayesian approach implemented in the Baseml package of PAML 4.3 (Yang 2007) with our Bayesian consensus genealogy, aligned conotoxin gene sequences, and the model utilized to build the genealogy (HKY + *I* + *G*; no clock). We tested positive selection with the maximum-likelihood method of the Codeml package of PAML 4.3 (Yang 2007). We used the Bayesian consensus gene topology and the alignment of suspected non-artifactual conotoxin sequences; we

excluded pseudogenes and one short toxin sequence of the α4/3 type. Models of $d_N/d_S = 1$ and d_N/d_S estimated were tested on the toxin-coding region where signatures of positive selection were detected previously for conotoxin genes (Duda and Palumbi 1999b; Puillandre et al. 2010) using the model of one rate across the whole tree and all sites (model = 0 and NSsites = 0).

Results

Species Tree and Dates of Separation

We obtained mitochondrial COI sequences and sequences of a calmodulin and tubulin intron from *C. lividus*, *C. diadema*, *C. quercinus*, and *C. sanguinolentus* (GenBank accession numbers in supplementary table S2, Supplementary Material online). The best substitution model is the HKY + I model (base frequencies $A = 0.2232$, $C = 0.1643$, $G = 0.2231$, $T = 0.3894$, ti/tv = 33.9489, proportion of invariate sites = 0.7110). Maximum-likelihood analysis of these COI sequences with the same model yielded the same topology for the four ingroup species as the Bayesian consensus phylogeny (ingroup topology shown in fig. 1). The minimum pairwise distances of sequences of the calmodulin and tubulin intron among these species with Jukes–Cantor model (Jukes and Cantor 1969) (supplementary table S3, Supplementary Material online) only differ at the second or the third decimal place from other models. As anticipated from analyses of the COI sequences, sequences of the calmodulin and tubulin introns of *C. quercinus* are most diverged from sequences of the other three species, but calmodulin sequences provided no resolution for the latter species (supplementary table S3, Supplementary Material online). Test of a strict molecular clock hypothesis of COI sequences through comparisons of maximum-likelihood scores of trees with and without molecular constraints accepted the null hypothesis of one evolution rate across the whole tree ($\ln L = -1{,}237.56$ with clock vs. $\ln L = -1{,}236.60$ without clock; likelihood ratio test yielded P value < 0.38, degrees of freedom [df] = 2). We selected the HKY model for calmodulin intron sequences and the HKY + I model for COI and tubulin intron sequences to fulfill the requirements of the BEAST software (Drummond and Rambaut 2007) used for molecular clock analyses. Analyses based on COI sequences only and concatenated COI–tubulin–calmodulin sequences with strict and relaxed clock models yielded relatively consistent time estimations (supplementary table S4, Supplementary Material online) and the same species topology as in figure 1. Uncorrelated lognormal clock estimation based on concatenated sequences revealed that *C. sanguinolentus* and *C. lividus* separated 0.3 Ma and *C. diadema* diverged approximately 1.6 Ma (fig. 1 and supplementary table S4, Supplementary Material online).

Conotoxin Gene Tree

We sequenced 938, 434, 411, and 303 cloned products from two individuals each of *C. lividus*, *C. diadema*, and *C. quercinus* and one individual of *C. sanguinolentus*. After evaluation and elimination of putative polymerase, cloning, and sequencing errors from results of both rounds of experiments, we identified 51 unique putative A-superfamily conotoxin sequences from *C. lividus*, 20 from *C. diadema*, 18 from *C. quercinus*, and 19 from *C. sanguinolentus* (supplementary table S5, Supplementary Material online; GenBank accession numbers JF723384–JF723491). The neighbor-joining tree that included all sequences contained no clades comprised exclusively of potential artifactual sequences. Saturation of sequence diversity curves implied that sequencing of additional products was unlikely to uncover additional unique sequences. Based on determinations of orthology and counts of alleles at each locus in each individual, we determined that these sequences represented 32 A-superfamily loci from *C. lividus*, 18 from *C. diadema*, 18 from *C. sanguinolentus*, and 12 from *C. quercinus* (supplementary table S5, Supplementary Material online), including several polymorphic loci. Both maximum-likelihood and Bayesian methods with the HKY + *I* + *G* model produced identical topologies of these sequences. Because the presence of allelic variants can lead to overestimation of duplication events, we selected sequences representing single alleles of each locus to build the reduced gene tree. The Bayesian consensus phylogeny is illustrated in figure 2.

Duplication and Loss

Based on examination of predicted amino acid sequences, a total of 13 sequences of eight loci from three *Conus* species appeared to represent pseudogenized gene copies, and these putative pseudogenes are of three distinct types: premature stop codon (type I) and destruction of cysteine backbone at different cysteine positions (types II and III) (supplementary table S6, Supplementary Material online, fig. 2). Additional gene losses may not be observable or identifiable because we may not have been able to sample them with the approach used. Reconciliation of the nonbinary gene tree (fig. 2) with the binary species tree (fig. 1) yielded 44 duplications and 39 gene losses. Reconciliation with alternative consideration revealed a minimum of 38 duplications and 29 gene losses (including eight putative pseudogene sequences) (fig. 1). Gene duplications have occurred relatively continuously throughout the evolutionary history of these four species (figs. 1 and 2). Since the divergence of *C. lividus*, *C. diadema*, and *C. sanguinolentus*, A-superfamily conotoxin genes underwent 13 rounds of duplication, and one locus exhibited up to four rounds of duplication within this time frame (figs. 1 and 2). The time estimates of all the branches in the species tree (fig. 1) sum to 23.9 My, so the overall duplication rate of this gene family averaged over all four species is 1.13 duplications per My. In the recent 1.6 My, the rate of gene birth is 3.71 duplications per My. Average duplication rates are heterogeneous among species: 26.7 duplications per My for *C. lividus*, 1.25 for *C. diadema*, 0 for *C. sanguinolentus*, and 0.18 for *C. quercinus*. Similarly, the frequencies of inferred gene losses are different among these species: 7 losses in *C. lividus* and 9 in *C. sanguinolentus*

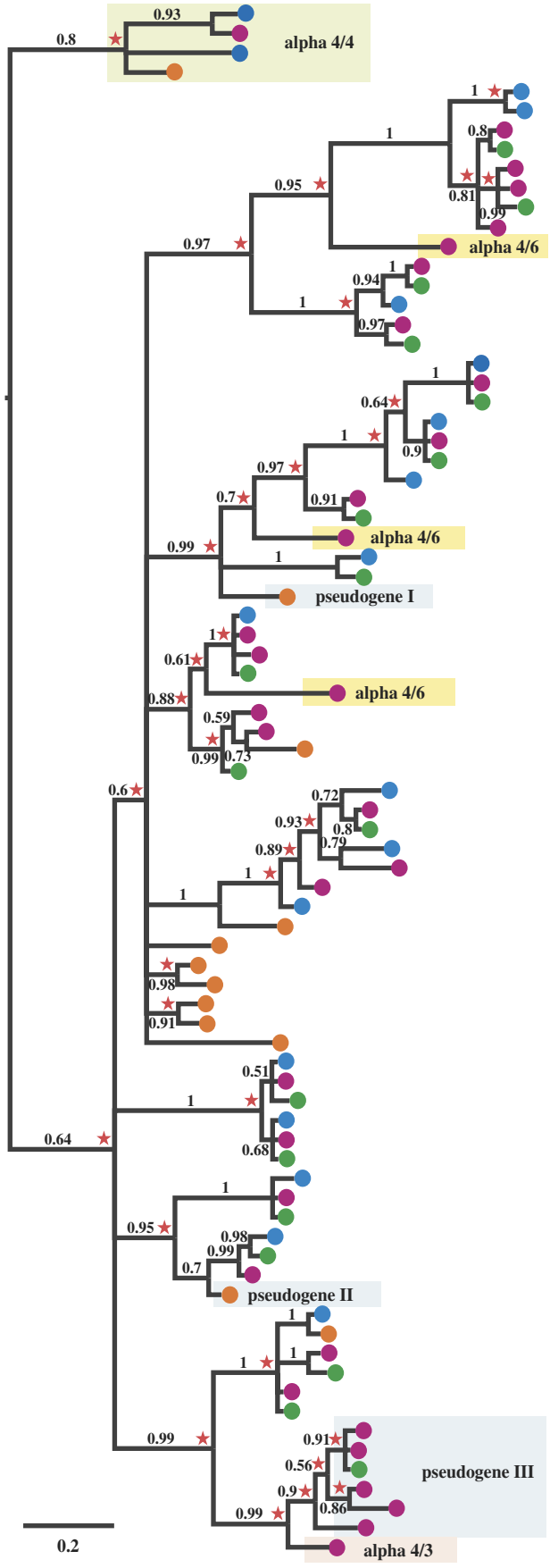

FIG. 2. Bayesian consensus phylogeny of DNA sequences of single alleles of each putative unique A-superfamily conotoxin locus recovered from four *Conus* species. Lengths of sequences after removal of gaps are 81 nt. Numbers on internal branches indicate Bayesian posterior probabilities. Color-coding scheme: at tips of the genealogy, loci of *Conus lividus* are identified as purple dots, *Conus diadema* as blue dots, *Conus quercinus* as red dots, and *Conus sanguinolentus* as green dots (consistent color labels as in fig. 1); sequences of the α4/4 type are shaded in green, α4/6 in yellow, and α4/3 in pink; pseudogenes are labeled according to types (type I, II, or III; supplementary table S6, Supplementary Material online) and shaded in light blue; inferred duplication events are indicated with red stars.

within 0.3 My, 8 in *C. diadema* within 1.6 My, and 4 in *C. quercinus* within 11 My (fig. 1).

To verify the estimates of parameters in Notung reconciliation, we performed the Bayesian analyses of the rates of gene gain and losses in PrIME-GSR. Tests of convergence of MCMC analyses of 700,000 generations with 450,000 burnin indicated that both chains of analyses converged. The mean birth rates of each chain are 0.081 and 0.078 per gene per My with variance of $1.69e^{-4}$ and $1.75e^{-4}$ and ranges are 0.049–0.123 and 0.051–0.113, respectively. The average death rates are 0.0037 and 0.0049 losses per gene per My for two chains with ranges of 0–0.024 and 0–0.023 and variances of $1.55e^{-5}$ and $1.43e^{-5}$.

Evolution Rate and Positive Selection

A test of a strict molecular clock with maximum-likelihood method and HKY + I + G model rejected the null hypothesis of equal evolution rates across the reduced conotoxin genealogy illustrated in figure 2 ($\ln L = -836.312$ with clock vs. $\ln L = -728.571$ without clock, df = 79, P value $< 1 \times 10^{-14}$ length of sequences = 81 bp, third codon position was included). Average rates of nonsynonymous substitution of A-superfamily genes are high but heterogeneous among these species: 2.7% per My for *C. lividus*, 1.8% per My for *C. sanguinolentus*, 1.3% per My for *C. diadema*, and 0.9% per My for *C. quercinus*. These rates are similar to those reported previously for O-superfamily conotoxin genes (Duda and Palumbi 1999b). The estimated nonsynonymous substitution rates immediately after duplication exhibited a maximum rate of substitutions of 22.9% per My and decreased dramatically with greater values of d_S (representing divergence time of paralogs) in quasi-exponential L-shape relationships (supplementary fig. S1, Supplementary Material online). Normalization of nonsynonymous substitution rates with $\log_{10}$ transformation and linear regression of the transformed data against d_S showed that the decreasing patterns were significant in each species (fig. 3). In addition, several orthologous loci among *C. lividus*, *C. diadema*, and *C. sanguinolentus* were identical in sequence after separation of species, whereas their respective paralogs differed substantially (supplementary fig. S2 and table S7, Supplementary Material online). We detected very strong positive selection within the toxin-coding regions of functional conotoxin genes ($d_N/d_S = 1.75$) (supplementary table S8, Supplementary Material online).

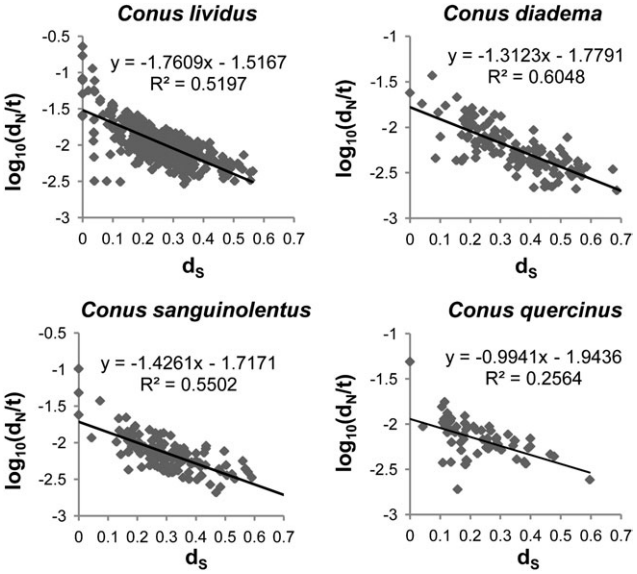

Fig. 3. Plots of logarithmic transformation of estimated rates of nonsynonymous substitution (*y* axis) against the pairwise synonymous divergence (*x* axis) of paralogs of each species. The fitted trend lines of linear regression are shown in each plot with equations and R^2 values labeled. All intercepts and negative slopes are significant (P value < 0.001).

Function and Ancestral Sequences Reconstruction

Genes recovered from our study encode four types of α-conotoxins: α4/4, α4/7, α4/6, and α4/3. The α4/7 type is the most common conopeptide, and genes for this peptide occurred in all four species, whereas only two loci represent α4/4 conotoxins. The α4/6 and α4/3 types are rarely found in worm-eating species (Puillandre et al. 2010) and were exclusively recovered from the genome of *C. lividus* (fig. 2). Ancestral sequence reconstruction based on the genealogy in supplementary figure S2 (Supplementary Material online) and all the aligned conotoxin sequences showed that the nodes ancestral to the gene of α4/3 type might have been pseudogenes (supplementary fig. S2 and table S9, Supplementary Material online), a possible intermediate stage from the α4/7 to the α4/3 type.

Discussion

We recovered conotoxin genes from genomic DNA of four closely related vermivorous *Conus* species, reconstructed the evolutionary history of these genes, and estimated numbers of duplication and loss, rates of gene birth, and nonsynonymous substitution rates among paralogs and orthologs of this gene family. This represents the first thorough study of conotoxin genes recovered exclusively from genomic DNA of *Conus* species. Our results revealed a remarkable pattern of quite extensive gene turnover, rapid evolution, and diversification of these genes within relatively recent evolutionary time.

Turnover of A-superfamily Genes

A-superfamily conotoxin genes appear to evolve in a "birth-and-death" pattern, a model of gene family evolution presented by Nei and Hughes (1992), but they do so in an extreme manner. Immediately after gene duplication mutation, redundant gene copies go through a phase of fixation in the population (Innan and Kondrashov 2010). The duplication and fixation phases cannot be considered separately in our case, so duplication of A-superfamily genes here is regarded as the product of both duplication mutation and fixation of gene duplicates. Gene duplication appears to have occurred relatively continuously throughout the evolutionary history of these four species, but with asymmetrical bursts of duplications among lineages (figs. 1 and 2). The average gene duplication rate estimated with the Bayesian method is about 0.08 per gene per My. If we assume that the size of the A-superfamily in the common ancestor of the four species analyzed was approximately 20 genes (mean of the numbers of unique loci of the four extant species), the average overall duplication rate estimated from reconciliation analyses is roughly 0.06 duplications per gene per My, which is essentially similar to the Bayesian estimation of duplication rate. This rate of gene duplication is about three times greater than the highest average rates estimated from several eukaryotic genomes (average 0.01 per gene per My and range of 0.002–0.02 in Lynch and Conery 2000; 0.028 per gene per My in yeast, 0.0014 in *Drosophila*, and 0.024 in *C. elegans* in Gu et al. 2002) and at least two times greater than the highest rates determined for multigene families such as olfactory receptor (Nei et al. 2008), vomeronasal receptor (Grus and Zhang 2004), spider venom (Binford et al. 2009), RNase gene families (Zhang et al. 2000), and pancrustacean eye development and phototransduction genes (Rivera et al. 2010). Since the divergence of *C. lividus*, *C. diadema*, and *C. sanguinolentus* 1.6 Ma, the gene duplication rate of 3.71 duplications per My, or roughly 0.19 per gene per My, is at least two times as high as the overall rate of this gene family, signifying an acceleration of gene birth, most of which is contributed by *C. lividus* and the common ancestor of *C. lividus* and *C. sanguinolentus*. The overall rate of gene duplication may be as high as the rate estimated from the recent 1.6 My but is difficult to prove because such extensive gene turnover may have eliminated traces of ancestral gene duplications.

Incomplete sampling of paralogous genes in the gene family may lead to incorrect placements of duplication events on the species tree (Ness et al. 2011) and either over- or underestimation of numbers of gene duplication events. It is possible that our approach and our conservative evaluation of non-artifactual toxin sequences failed to recover certain members of the gene family or that recent gene duplicates are too similar to be recognized as such. We simulated scenarios of incomplete sampling of conotoxin loci by treating our current data set as reference and randomly removing up to 10% of loci. Results implied that the pattern of extensive gene gain revealed from our study was not affected by potential incomplete sampling. Simulations of both proportional removal of sequences of each species and random removal of sequences of all four species combined showed that failure to recover additional genes

would have lead to an underestimation of the overall number of duplication events (supplementary fig. S3a, Supplementary Material online) but did not affect fine-scale measurements of duplication events during short time intervals (supplementary fig. S3b, Supplementary Material online). The variance of estimates slightly increased with more genes removed, but no outlier was detected that would have significantly altered the magnitude of gene turnover (supplementary fig. S3, Supplementary Material online).

The overall rates of gene duplication estimated through Bayesian and reconciliation approaches were remarkably very similar, but rates of gene loss differed considerably. Results from the reconciliation analysis implied that the A-superfamily has maintained its size over time, whereas results from the Bayesian approach suggested that the gene family has undergone constant expansion. Reconciliation of the gene trees and species tree (topology only) utilizes maximum parsimony, and its optimization weighs heavily on minimization of gene duplications (Chen et al. 2000; Vernot et al. 2007). On the other hand, the Bayesian approach models gene gain and loss based on a species tree with known branch lengths (Åkerborg et al. 2009). The rates of gene loss are modeled as being constant through time which is incompatible with neofunctionalization of gene duplicates or selection (Eulenstein et al. 2010) and which may not be applicable to the A-superfamily. The discrepancies in estimates of gene loss from these two approaches may also be induced by lack of resolution near the root of the gene tree (fig. 2), which may impact the estimation of these rates.

Even though we were unable to evaluate the duplication and fixation phases separately, the rate of duplication mutation alone (i.e., not including rates of duplicate fixation) of these conotoxin genes is likely to be much higher than our estimate of gene duplication (which includes the rate of duplicate fixation) because some duplicated genes may not have been fixed after duplication and because some duplicates may not have diverged in sequence and so are unrecognizable in the genomes of these species. Nothing is known about the mechanism of conotoxin gene duplication, but based on the presence of highly conserved regions of the toxin prepropeptide as well as introns and untranslated regions of conotoxin genes, the process of gene duplication is more likely to be due to unequal crossing-over than retroposition (Zhang 2003). Locations of these genes may also affect the rate of gene duplication, and we predict that these genes and members of other conotoxin gene families that are evolving rapidly are predominantly clustered within regions of the genome that are prone to extremely high rates of unequal crossing-over, as is suspected for other gene-rich gene families (Zhang 2003).

Evolution of A-superfamily Genes

The rates of evolution of the A-superfamily conotoxin genes are comparable with those observed for O-superfamily conotoxin genes (Duda and Palumbi 1999b). Results from the strict molecular clock test indicated that the genes analyzed exhibit heterogeneous rates of evolution, and the average nonsynonymous substitution rates differed slightly among species. The semi-L-shape pattern of the nonsynonymous substitution rates of paralogs against their divergence time (supplementary fig. S1, Supplementary Material online) and the significant negative slope and intercept of the regression (fig. 3) imply that rates of evolution decrease significantly immediately after duplication and then gradually stabilize at a plateau. The nonsynonymous substitution rate immediately after duplication exhibited a maximum rate of substitution of 22.9% per My, suggesting that duplication facilitates the rapid evolution of these genes. The evolution of recent gene duplicates may be asymmetrical such that heightened rates of evolution occur only within copies that are relaxed from selection.

Based on the strong signals of positive selection, we posit that duplicated gene copies have undergone neofunctionalization. Because all recovered sequences exhibit similarity to a variety of sequences of A-superfamily gene transcripts (i.e., they cluster among the breadth of A-superfamily conotoxin sequences recovered from venom duct mRNAs that are reported in GenBank), we presume that these sequences represent A-superfamily conotoxin genes and that they do not represent genes with other functions (e.g., descendants of the ancestral genes from which conotoxin genes were coopted). In addition, we are unaware of any studies that have indicated that A-superfamily–like genes are expressed or have alternative functions outside of the venom delivery system of *Conus*. Furthermore, mutagenesis studies of conotoxins have shown that modification of single amino acids of the mature toxin alters the peptide's functional specificity and binding efficiency (Dutertre et al. 2007; Whiteaker et al. 2007; Ellison et al. 2008; Halai et al. 2009). Hence, nonsynonymous substitutions within the toxin-coding region of redundant gene copies likely affect the functions of the expressed products, possibly in terms of their utility in prey capture. Gene duplication and the subsequent evolution of the duplicates appear to have increased the functional diversity of conotoxins and may have led to functional shifts of some genes. α-Conotoxin peptides comprise several distinct types (α3/5, α4/3, α4/4, α4/6, and α4/7) that are distinguished by the number of amino acids that occur between the second and third and third and fourth cysteine residues; each type targets certain subsets of muscle/neuronal receptors (McIntosh et al. 1999; Tsetlin et al. 2009). Inferred from the genealogy (fig. 2), new functional α4/6 and α4/3 types emerged by duplication and divergence from the common α4/7 type (supplementary fig. S4, Supplementary Material online). The α4/6 and α4/3 types were only recovered from *C. lividus*, a pattern that implies that these functional types emerged from fairly recent duplications (<0.3 My). In addition to insertions/deletions within the duplicates, the genealogy and ancestral sequence reconstruction (supplementary fig. S2 and table S9, Supplementary Material online) suggest that the shift from the α4/7 type to α4/3 type may have resulted from "disulfide-bond reshuffling," a phenomenon proposed by Zhang (2007) and observed in

RNase A genes of primates (hypothetical process illustrated in supplementary fig. S5, Supplementary Material online).

Model of Conotoxin Gene Family Evolution

Redundant gene copies produced by gene duplication mutation can be either neutral and fixed by genetic drift (Ohno 1970) or beneficial and fixed by selection (Francino 2005; Kondrashov and Kondrashov 2006; Bergthorsson et al. 2007). We are unable to assess which mechanism is associated with the fixation of conotoxin gene duplicates. We also are not aware if increases in toxin dosage improve predation efficiency or if amplification of secondary functions of these genes is beneficial to cope with ecological shifts and varying stress. Based on the high rates of gene turnover and the rapid evolution of conotoxin genes we observed, we posit that extensive gene duplication events create redundant gene copies for rare though beneficial mutations to occur and hence dramatically increase the frequency at which gene duplicates become fixed and new adaptive genotypes arise. Strong positive selection leading to neofunctionalization of duplicate genes may dramatically enhance the rapid fixation of advantageous genotypes and contribute to the rapid evolution of conotoxin genes. Neutral or disadvantageous copies may be pseudogenized or lost from the genome in the fate determination stage. This scenario is similar to predictions of the adaptive radiation model of gene family evolution: rapid bursts of duplication, strong selection on paralogs, and eventual pseudogenization of some gene copies (Francino 2005), even though the neutrality of duplication is debatable.

Venom Evolution

Rates of duplication of A-superfamily conotoxins are asymmetrical among species, even between populations of *C. lividus* and *C. sanguinolentus* that appear to have shared a recent common ancestor about 0.3 Ma. High rates of turnover and the rapid evolution of A-superfamily conotoxin genes cause large divergence in the composition of this gene family among closely related *Conus* species. Out of the six ancestral orthologous genes of the four species examined, two genes were pseudogenized in *C. quercinus*, whereas the others were not observed and so were possibly lost from its genome. Meanwhile, numerous putative species-specific duplications occurred in *C. quercinus* (fig. 2). Gene duplication events after separation of *C. diadema* and species-specific duplications in *C. diadema* and *C. lividus* also contributed to divergence of the composition of the A-superfamily of these two species (fig. 2). Such large differences in composition among species may also be induced by differential gene losses as the numbers of gene losses differ among species (fig. 2). Without information about the structure and distribution of these genes in the genome, we cannot completely rule out the possibility of ancient duplication and lineage-specific losses, a pattern found in many genes such as *tuf* genes in eubacteria (Lathe and Bork 2001) and globin genes in mammals (Opazo et al. 2008). But the simultaneous gain and loss patterns should be more probable because the size of A-superfamily in *C. lividus* is much larger than in its close relatives and the ages (d_S) of paralogs are relatively continuous (fig. 3).

Sources of selection and the correlation of gene duplication and diversification of conotoxin genes with ecological adaptations in this study remain unknown. The evolution and diversity of conotoxin genes have been suggested to correlate with prey specializations in various studies of *Conus* species at both interspecific and population levels (Duda and Palumbi 2004; Duda and Lee 2009; Duda et al. 2009). Our target species prey on diverse sets of marine worms and exhibit different geographical distributions (Kohn 1968, 2001; Nybakken 1978). Hence, it is possible that the differences in patterns of evolution of A-superfamily among species are correlated with prey diversities or prey availability in different geographical boundaries. This hypothesis could be verified with future studies of functional assays of conotoxins on different types of prey and direct tests of patterns of conotoxin gene family evolution and dietary shifts.

Conclusion

Our study revealed that A-superfamily conotoxin genes of *Conus* species possess heightened rates of gene turnover coupled with enhanced rates of evolution. The extensive gene turnover appears to have facilitated vast diversification of the composition of the A-superfamily among species and presumably enabled functional shifts of peptides expressed in the venoms of these species, a condition that may be compelled by dietary shifts or the origins of resistance in prey. Increases in gene copy number likely create additional targets of opportunity for beneficial mutations, enhance the efficacy of positive selection, and may eventually lead to the origin of novel gene functions. In this sense, continuous radiation of gene families facilitates the diversification and rapid evolution of genes that are associated with predator–prey interactions. Such extensive turnover of conotoxin genes affects the ability to reconstruct the long-term evolutionary patterns of these genes, and so it is critical to examine the evolutionary histories and relationships of these genes over short time intervals.

Supplementary Materials

Supplementary tables S1–S9 and figures S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org).

Acknowledgments

We thank Jon-Paul Bingham (University of Hawaii) for providing *C. quercinus* specimens. We thank David A. Roberts for his help with simulations. We acknowledge Jianzhi Zhang, Alex S. Kondrashov, Taehwan Lee, and Celia K. C. Churchill for their suggestions on the manuscript. We thank the three anonymous reviewers and the associate editor Todd Oakley for their valuable suggestions and comments. This work was supported by funds from the University of Michigan Rackham Graduate School and

a Rackham/Department of Ecology and Evolutionary Block Grant awarded to Dan Chang and a National Science Foundation award (IOS-0718379) to Thomas F. Duda Jr. The authors declare no competing financial interests.

References

Akaike H. 1974. A new look at the statistical model identification. *IEEE Trans Automat Contr.* 19:716–723.

Åkerborg Ö, Sennblad B, Arvestad L, Lagergren J. 2009. Simultaneous Bayesian gene tree reconstruction and reconciliation analysis. *Proc Natl Acad Sci U S A.* 106:5714–5719.

Arboleda-Bustos CE, Segarra C. 2011. The *Dca* gene involved in cold adaptation in drosophila melanogaster arose by duplication of the ancestral regucalcin gene. *Mol Biol Evol.* 28:2185–2195.

Bergthorsson U, Andersson DI, Roth JR. 2007. Ohno's dilemma: evolution of new genes under continuous selection. *Proc Natl Acad Sci U S A.* 104:17004–17009.

Binford GJ, Bodner MR, Cordes MHJ, Baldwin KL, Rynerson MR, Burns SN, Zobel-Thropp PA. 2009. Molecular evolution, functional variation, and proposed nomenclature of the gene family that includes sphingomyelinase D in sicariid spider venoms. *Mol Biol Evol.* 26:547–566.

Burri R, Salamin N, Studer RA, Roulin A, Fumagalli L. 2010. Adaptive divergence of ancient gene duplicates in the avian MHC class II B. *Mol Biol Evol.* 27:2360–2374.

Chen K, Durand D, Farach-Colton M. 2000. A hybrid micro-macroevolutionary approach to gene tree reconstruction. *J Comput Biol.* 7:429–447.

Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet.* 9:938–950.

Conticello SG, Gilad Y, Avidan N, Ben-Asher E, Levy Z, Fainzilber M. 2001. Mechanisms for evolving hypervariability: the case of conopeptides. *Mol Biol Evol.* 18:120–131.

Daltry JC, Wuster W, Thorpe RS. 1996. Diet and snake venom evolution. *Nature* 379:537–540.

Drummond A, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7:214.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e88.

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

Duda TF, Kohn AJ. 2005. Species-level phylogeography and evolutionary history of the hyperdiverse marine gastropod genus *Conus*. *Mol Phylogenet Evol.* 34:257–272.

Duda 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, Lee T. 2009. Ecological release and venom evolution of a predatory marine snail at Easter Island. *PLoS One* 4:e5558.

Duda TF, Palumbi SR. 1999a. Developmental shifts and species selection in gastropods. *Proc Natl Acad Sci U S A.* 96:10272–10277.

Duda TF, Palumbi SR. 1999b. Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc Natl Acad Sci U S A.* 96:6820–6823.

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.

Durand D, Halldorsson BV, Vernot B. 2006. A hybrid micro evolutionary approach to gene tree reconstruction. *J Comput Biol.* 13:320–335.

Dutertre S, Ulens C, Buttner R, et al. (13 co-authors). 2007. AChBP-targeted α-conotoxin correlates distinct binding orientations with nAChR subtype selectivity. *EMBO J.* 26:3858–3867.

Ellison MFZ, Park AJ, Zhang X, Olivera BM, McIntosh JM, Norton RS. 2008. α-RgIA, a novel conotoxin that blocks the α9α10 nAChR: structure and identification of key receptor binding residues. *J Mol Biol.* 377:1216–1227.

Eulenstein O, Huzurbazar S, Liberles DA. 2010. Reconciling phylogenetic trees. In: Dittmar K, Liberles D, editors. Evolution after gene duplication. Hoboken (NJ): John Wiley & Sons, Inc., p. 185–206.

Fischer HM, Wheat CW, Heckel DG, Vogel H. 2008. Evolutionary origins of a novel host plant detoxification gene in butterflies. *Mol Biol Evol.* 25:809–820.

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.

Francino MP. 2005. An adaptive radiation model for the origin of new gene functions. *Nat Genet.* 37:573–578.

Fry BG, Vidal N, Norman JA, et al. (14 co-authors). 2006. Early evolution of the venom system in lizards and snakes. *Nature* 439:584–588.

Fry BG, Wüster W, Kini RM, Brusic V, Khan A, Venkataraman D, Rooney AP. 2003. Molecular evolution and phylogeny of elapid snake venom three-finger toxins. *J Mol Evol.* 57:110–129.

Gevers D, Vandepoele K, Simillion C, Van de Peer Y. 2004. Gene duplication and biased functional retention of paralogs in bacterial genomes. *Trends Microbiol.* 12:148–154.

Grus WE, Zhang J. 2004. Rapid turnover and species-specificity of vomeronasal pheromone receptor genes in mice and rats. *Gene* 340:303–312.

Gu Z, Cavalcanti A, Chen F, Bouman P, Li W. 2002. Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Mol Biol Evol.* 19:256–262.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:696–704.

Guldner E, Godelle B, Galtier N. 2004. Molecular adaptation in plant hemoglobin, a duplicated gene involved in plant–bacteria symbiosis. *J Mol Evol.* 59:416–425.

Halai R, Clark RJ, Nevin ST, Jensen JE, Adams DJ, Craik DJ. 2009. Scanning mutagenesis of α-conotoxin Vc1.1 reveals residues crucial for activity at the α9α10 nicotinic acetylcholine receptor. *J Biol Chem.* 284:20275–20284.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* 22:160–174.

Huelsenbeck JP, Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.

Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B Biol Sci.* 256:119–124.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet.* 11:97–108.

Juarez P, Comas I, Gonzalez-Candelas F, Calvete JJ. 2008. Evolution of snake venom disintegrins by positive Darwinian selection. *Mol Biol Evol.* 25:2391–2407.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro HN, editor. Mammalian protein metabolism. New York: Academic Press. p. 21–123.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* 16:111–120.

Kohn AJ. 1959. The ecology of *Conus* in Hawaii. *Ecol Monogr.* 29:47–90.

Kohn AJ. 1968. Microhabitats abundance and food of *Conus* on atoll reefs in Maldive and Chagos islands. *Ecology* 49:1046–1062.

Kohn AJ. 1990. Tempo and mode of evolution in *Conidae*. *Malacologia* 32:55–67.
Kohn AJ. 2001. Maximal species richness in *Conus*: diversity, diet and habitat on reefs of northeast Papua New Guinea. *Coral Reefs* 20:25–38.
Kondrashov FA, Kondrashov AS. 2006. Role of selection in fixation of gene duplications. *J Theor Biol.* 239:141–151.
Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. 2002. Selection in the evolution of gene duplications. *Genome Biol.* 3:research0008. 0001–0008.0009.
Lathe WC, Bork P. 2001. Evolution of *tuf* genes: ancient duplication, differential loss and gene conversion. *FEBS Lett.* 502:113–116.
Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
McIntosh JM, Santos AD, Olivera BM. 1999. *Conus* peptides targeted to specific nicotinic acetylcholine receptor subtypes. *Annu Rev Biochem.* 68:59–88.
Moran Y, Weinberger H, Sullivan JC, Reitzel AM, Finnerty JR, Gurevitz M. 2008. Concerted evolution of sea anemone neurotoxin genes is revealed through analysis of the *Nematostella vectensis* genome. *Mol Biol Evol.* 25:737–747.
Nei M, Hughes AL. 1992. Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. In: Tsuji K, Aizawa M, Sasazuki T, editors. 11th Histocompatibility Workshop and Conference. Oxford: Oxford University. Vol. 2p. 27–38.
Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. Oxford: Oxford University Press.
Nei M, Niimura Y, Nozawa M. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet.* 9:951–963.
Ness RW, Graham SW, Barrett SCH. 2011. Reconciling gene and genome duplication events: using multiple nuclear gene families to infer the phylogeny of the aquatic plant family Pontederiaceae. *Mol Biol Evol.* 28:3009–3018.
Nybakken J. 1978. Population characteristics and food resource utilization of *Conus* in the galapagos islands. *Pac Sci.* 32:271–280.
Ohno S. 1970. Evolution by gene duplication. Berlin (Germany): Springer-Verlag.
Olivera BM. 2002. *Conus* venom peptides: reflections from the biology of clades and species. *Annu Rev Ecol Syst.* 33:25–47.
Opazo JC, Hoffmann FG, Storz JF. 2008. Differential loss of embryonic globin genes during the radiation of placental mammals. *Proc Natl Acad Sci U S A.* 105:12950–12955.
Perry GH, Dominy NJ, Claw KG, et al. (13 co-authors). 2007. Diet and the evolution of human amylase gene copy number variation. *Nat Genet.* 39:1256–1260.
Plummer M, Best N, Cowles K, Vines K. 2006. CODA: convergence diagnosis and output analysis for MCMC. *R News* 6:7–11.
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. 2002. SE-AL v. 2.0a11 [Internet]: sequence alignment program [cited 2012 Feb 23]. Available from: http://tree.bio.ed.ac.uk/software/seal/
Rambaut A, Drummond AJ. 2007. Tracer v1.4. [Internet]. [cited 2012 Feb 23]. Available from: http://tree.bio.ed.ac.uk/software/tracer/
Rivera SR, Pankey MS, Plachetzki DC, Villacorta C, Syme AE, Serb JM, Omilian AR, Oakley TH. 2010. Gene duplication and the origins of morphological complexity in pancrustacean eyes, a genomic approach. *BMC Evol Biol.* 10:123.
Santos AD, McIntosh JM, Hillyard DR, Cruz LJ, Olivera BM. 2004. The A-superfamily of conotoxins—structural and functional divergence. *J Biol Chem.* 279:17596–17606.
Schwarz GE. 1978. Estimating the dimension of a model. *Ann Stat.* 6:461–464.
Swofford. 2002. PAUP*: phylogenetic analysis using parsimony (and other methods) 4.0b10. Sunderland (MA): Sinauer Associates.
Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol.* 24:1596–1599.
Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.
Tsetlin V, Utkin Y, Kasheverov I. 2009. Polypeptide and peptide toxins, magnifying lenses for binding sites in nicotinic acetylcholine receptors. *Biochem Pharmacol.* 78:720–731.
Vernot B, Stolzer M, Goldman A, Durand D. 2007. Reconciliation with non-binary species trees. Computational Systems Bioinformatics: CSB2007. London: Imperial College Press. p. 441–452.
Wen Z, Rupasinghe S, Niu G, Berenbaum MR, Schuler MA. 2006. CYP6B1 and CYP6B3 of the black swallowtail (*Papilio polyxenes*): adaptive evolution through subfunctionalization. *Mol Biol Evol.* 23:2434–2443.
Whiteaker P, Christensen S, Yoshikami D, Dowell C, Watkins M, Gulyas J, Rivier J, Olivera BM, McIntosh JM. 2007. Discovery, synthesis, and structure activity of a highly selective α7 nicotinic acetylcholine receptor antagonist. *Biochemistry* 46:6628–6638.
Yamanaka K, Fang L, Inouye M. 1998. The CspA family in *Escherichia coli*: multiple gene duplication for stress adaptation. *Mol Microbiol.* 27:247–255.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol.* 18:292–298.
Zhang J. 2007. Disulfide-bond reshuffling in the evolution of an ape placental ribonuclease. *Mol Biol Evol.* 24:505–512.
Zhang J, Dyer KD, Rosenberg HF. 2000. Evolution of the rodent eosinophil-associated RNase gene family by rapid gene sorting and positive selection. *Proc Natl Acad Sci U S A.* 97: 4701–4706.