EVOLUTION AND EXPRESSION OF GENE FAMILIES AT THE INTERFACE OF PREDATOR-PREY INTERACTIONS

by

Dan Chang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in the University of Michigan
2012

Doctoral Committee:

Associate Professor Thomas F. Duda, Jr., Chair
Professor Lori L. Isom
Professor Earl E. Werner
Professor Jianzhi Zhang

2012

DEDICATION

This doctoral dissertation is dedicated to my parents, Xiaowu Liu and Zonglin Chang, and my husband, David Andrew Roberts.

ACKNOWLEDGMENTS

Fresh out of my undergraduate life six years ago, I set off for my adventure in the United States of America, a country that is half a globe away from my hometown. As a newcomer who knew little about graduate schools in the States, I faced countless challenges and tremendous pressure. Now at the end of my PhD study and with the completion of my dissertation, I give my sincere appreciation to those people who have helped me sail through this wonderful journey.

I feel honored and grateful to have Dr. Thomas Duda, Jr. as my PhD advisor and chair of the dissertation committee. As the lead scientist in the field, he is my role model. He is very knowledgeable, patient, caring and supportive. He gave me great advice in academic studies and research, and helped me get over the culture shock. He encouraged me to develop my own projects, apply for my own funding, and handle correspondence from reviewers and editors, all of which enabled me to develop my independence and prepare for my academic career.

I have been very fortunate to have Drs. Lori Isom, Earl Werner and Jianzhi Zhang as members of my doctoral committee. It has been a great honor and fulfilling experience to work with these world-class scientists. They are extremely helpful, patient and encouraging. They provided great academic and career advice to me, and helped me broaden the scope of my research and enhance the depth of this dissertation.

I acknowledge all collaborators for their support of my dissertation research. I greatly appreciate Dr. Alan Kohn at the University of Washington for his generous support in terms of knowledge and guidance towards my research. I thank Dr. Jon-Paul Bingham at the University of Hawaii for providing me with specimens. I especially thank Dr. Alex Kerr at the University of Guam Marine Lab for his great help with my field research. I also appreciate Marielle Terbio, Carmen Kautz and Kirstie Goodman-Randall for their company and help in Guam. Dr. Jincong Tao at the Molecular Core Lab gave me useful advice for employing Real-time qPCR technologies.

Certainly, my life at Michigan would not have been glorious without my dear collegues and friends. I would like to thank Dr. Diarmaid Ó Foighil for his advice and support of my research and academic career. From him, I learned so much about writing styles, creativity and problem-solving. I thank Dr. John Burch (WGM) for his detailed introduction to molluscs and his motivating stories and wisdom. I acknowledge Dr. Janice Pappas with for her great help and suggestions on statistical analyses of my data. I thank former postdocs from the Duda Lab, Drs. Gang Chen, Taehwan Lee and Ed Remigio for their suggestions and constructive criticism of my work. I would like to thank all current and former mollusc ladies, Celia Churchill, Cindy Bick, Samatha Flowers, Jingchun Li, Paula Teichholtz, Semoya Phillips, Alex Moore and Andrea Walther for all the wonderful time we spent together. I would like to give a special acknowledgement to the undergraduate researcher Amy Olenzek in the Duda Lab for her great dedication and hard work.

Thanks to all the people in Department of Ecology and Evolutionary Biology, Museum of Zoology and Department of Statistics for their help with my PhD studies. In particular, thanks to all the staff, especially Jane Sullivan, Julia Eussen, Norah Daugherty, Robin Murrell and Fritz Paper their generous help. I appreciate my 'countless' friends at Michigan: Huateng Huang, Wenfeng Qian, Di Ma, Ya Yang, Zhi Wang, Jin Zhang, Luqia Hou, Gang Su, Di Deng and many others for their help, company and moral support.

I received financial support from the Department of Ecology and Evolutionary Biology as fellowships and a graduate student instructor, from Museum of Zoology as a curatorial research assistant, from the National Science Foundation award (IOS-0718379) to Thomas Duda, Jr. as a research assistant and three fellowships from the Rackham Graduate School. This dissertation research was funded by three Rackham EEB Block Grants, one Rackham Graduate Research Grant, two Hinsdale-Walker Fellowships and one NSF award (IOS-0718379) to Thomas Duda, Jr.

Finally I give my deep appreciation to my parents, Zonglin Chang and Xiaowu Liu, for their 28 years of support and guidance. Without their dedication and persistence I would never have reached this far. I would like to thank my beloved husband, David Roberts, for his support, patience and sacrifice.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGMENTS iii

LIST OF FIGURES vii

LIST OF TABLES x

ABSTRACT xii

CHAPTER 1 Introduction 1

References 10

CHAPTER 2 Extensive and continuous duplication facilitates rapid evolution and diversification of gene families 13

Introduction 13

Materials and methods 16

Results 24

Discussion 28

Conclusion 36

Acknowledgements 36

References 57

CHAPTER 3 Differential exploration of venome space: Plasticity of expression of venom genes 63

Introduction 63

Materials and methods 67

Results 72

Discussion 77
Conclusion 84
Acknowledgements 85
References 94
CHAPTER 4 Geographic heterogeneity in biotic interactions drives the evolution of predatory genes 98
Introduction 98
Materials and Methods 99
Results and Discussion 107
Acknowledgements 114
References 133
CHAPTER 5 Ontogenetic plasticity of feeding traits associated with dietary breadth 137
Introduction 137
Materials and Methods 140
Results 150
Discussion 156
Acknowledgements 164
References 189

LIST OF FIGURES

Figure 2.1. 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. 38
Figure 2.2. Bayesian consensus phylogeny of DNA sequences of single alleles of each putative unique A-superfamily conotoxin locus recovered from four *Conus* species. 39
Figure 2.3. Plot of estimated pairwise rates of nonsynonymous substitution (y-axis) against the pairwise synonymous divergence (x-axis) of paralogous A-superfamily conotoxin loci of four *Conus* species.....41
Figure 2.4. Plots of logarithmic transformation of estimated rates of nonsynonymous substitution (y-axis) against the pairwise synonymous divergence (x-axis) of paralogs of each species. 42
Figure 2.5. Bayesian consensus phylogeny of all A-superfamily sequences recovered from the genome of the four *Conus* species with the HKY+I+G model.....43
Figure 2.6. Duplications and gene losses inferred from simulated randomly reduced gene pools of four species. 45
Figure 2.7. Scenario of evolution of α-conotoxin subtypes inferred from ancestral sequence reconstructions. 46
Figure 2.8. Hypothetical scenario of transition from α4/7 to α4/3 conotoxin types. 47
Figure 3.1. Phylogenies of A-superfamily conotoxin genes retrieved from genomic DNA and venom duct cDNA of four *Conus* species. 86
Figure 3.2. Relative expression levels of coexpressed orthologs between species, inparalogs and individuals with species.....88
Figure 4.1. Gene trees of unique conotoxin gene sequences recovered from venom duct cDNA of *C. ebraeus* individuals at three locations constructed using maximum-likelihood and mid-point rooting. 115
Figure 4.2. Alignment of predicted amino acid sequences of alleles of four conotoxin loci of *C. ebraeus*. 117

Figure 4.3. Haplotype networks of alleles of single conotoxin loci of *C. ebraeus* at American Samoa, Guam and Hawaii and multi-locus structure analyses. 118
Figure 4.4. Bayesian consensus phylogenies constructed from sequences of a region of the mitochondrial 16S gene recovered from fecal samples here and downloaded from GenBank. ..119
Figure 4.5. Ordination/biplots of canonical correspondence analyses of diversities and geographic divergence of conotoxin genes with heterogeneities of prey items. 121
Figure 5.1. Bayesian consensus phylogenies of mitochondrial 16S gene sequences of fecal samples of *C. ebraeus* and known polychaete species. 165
Figure 5.2. Boxplots of ranges of shell lengths of *C. ebraeus* individuals consuming different types of prey. 167
Figure 5.3. Boxplots of ranges of shell lengths of *C. ebraeus* individuals consuming different types of prey that are identified with microscopic examination of fecal samples. 168
Figure 5.4. Sliding window analyses of dietary diversity. 169
Figure 5.5. Heatmaps of dietary ontogeny of *C. ebraeus* individuals. 170
Figure 5.6. Heatmaps of relative expression levels of six conotoxin genes among individuals. 171
Figure 5.7. Hierarchical clusters of *C. ebraeus* individuals based on relative levels of expression of the six conotoxin genes. 172
Figure 5.8. Clustering results with the PAM method. 173
Figure 5.9. Moving averages of relative expression levels ($-\Delta C_T$) of the six conotoxin genes among individuals of different shell lengths (window size=5mm in shell lengths). 174
Figure 5.10. Simulated moving averages (window size of 5mm shell lengths) of expression levels ($-\Delta C_T$) of six conotoxin genes. 175
Figure 5.11. Boxplots of expression levels of *C. ebraeus* individuals feeding on different prey species. 176
Figure 5.12. Boxplots of relative conotoxin gene expression levels among individuals feeding on different prey types. 177
Figure 5.13. Heatmaps of conotoxin gene expression levels among *C. ebraeus* individuals with known prey species. 179
Figure 5.14. Hierarchical clusters of *C. ebraeus* individuals with known prey species based on patterns of expression of six conotoxin genes. 180

Figure 5.15. Patterns of ontogenetic shifts of levels of expression of conotoxin genes and dietary diversities. 181
Figure 5.16. Cross-correlation of time series of conotoxin gene expression levels with dietary diversities. 182
Figure 5.17. Linear regression of specific lag of expression levels of each conotoxin locus with dietary variables (*S*, *H'* and GD) 184
Figure 5.18. Boxplots of conotoxin gene expression levels among groups of individuals classified by sexual maturity. 185

LIST OF TABLES

Table 2.1. List of 10 sets of primers that were used to amplify A-superfamily conotoxin genes from the genome of the four closely-related *Conus* species.....48

Table 2.2. GenBank accession numbers for mitochondrial COI sequences, sequences of a tubulin intron and a calmodulin intron of *C. lividus*, *C. sanguinolentus*, *C. diadema* and *C. quercinus*. .49

Table 2.3. Jukes-Cantor distances with uniform rates among sites calculated among the intron sequences of a tubulin locus and a calmodulin locus between pairs of *Conus* species. 50

Table 2.4. Time estimations of the most common ancestors from strict and relaxed clock models of COI sequences and concatenated COI-calmodulin-tubulin intron sequences of four *Conus* species. 51

Table 2.5. Number of colonies sequenced, number of unique sequences (including errors), non-artefactual sequences and putative loci recovered from the genome of each species.....52

Table 2.6. Types of pseudogenes and their predicted amino acid sequences.....53

Table 2.7. Estimates of d_S, d_N and rates of nonsynonymous substitution of orthologous loci (including co-orthologs). 54

Table 2.8. Results of test of postive selection based on analyses with the Codeml package of PAML 4.3 on toxin-coding regions of putative functional conotoxin genes. 55

Table 2.9. Reconstructed DNA sequences and respective amino acid sequences of the ancestral nodes A and B illustrated in Figure 2.5.....56

Table 3.1. Expressed A-superfamily conotoxin recovery information. 89

Table 3.2. Numbers of orthologous coexpressed loci among species (below diagonal) and their proportions in the venom duct expression profiles of each species (above diagonal). 90

Table 3.3. Community phylogenetic indices as evaluations of phylogenetic stucture of expressed genes.....91

Table 3.4. Models used to test if presently expressed genes exhibit heightened ω (d_N/d_S) values than the rest of the genes, and results of the tests. 92

Table 3.5. Results of maximum-likelihood estimations of ω values for A-superfamily genes in each *Conus* species.....93

Table 4.1. General primers for each conotoxin superfamily. 123

Table 4.2. Locus-specific primers used to genotype each locus. 124

Table 4.3. Allelic-specific primers to verify and differentiate alleles. 125

Table 4.4. Sample sizes, numbers of total and unique alleles, gene/haplotype diversities, nucleotide diversities and their standard errors (SE), and Tajima's *D* values of the five conotoxin loci at three locations. 126

Table 4.5. Pairwise Φ_{ST} values and dietary overlap indices among populations of *C. ebraeus*. .127

Table 4.6. Results of hierarchical Analysis of Molecular Variance (AMOVA) for the highly polymorphic loci (ED4, ED6 and E1) with the Tamura-Nei model. 128

Table 4.7. Putative prey species, numbers of each species and total number of prey items of each higher taxonomic level at the three locations. 129

Table 4.8. Coefficients of the slope of the fitted line in simple regression analyses of the haplotype and nucleotide diversities of five conotoxin genes and the COI gene with the diversities of prey (*H'* and genetic distance). 130

Table 4.9. Pearson , Spearman and Kendall correlation coefficients of the pairwise Φ_{ST} matrices of each of the three highly polymorphic conotoxin genes with the pairwise divergence indices of prey (PS_I and D_{ST}). 131

Table 4.10. Haplotype/gene diversity and nucleotide diversity of two O-superfamily conotoxin genes *MIL2* and *MIL3* and the mitochondrial COI gene of *C. miliaris* populations at Easter Island (abbreviated as EI), Guam and American Samoa (abbreviated 'AS'). 132

Table 5.1. Primers for the six conotoxin genes and that were utilized in Real-time qPCR. 186

Table 5.2. Fisher's exact tests of prey compositions between size classes of *C. ebraeus* individuals. 187

Table 5.3. Pairwise Φ_{ST} values among *C. ebraeus* individuals of different size classes and locations. 188

ABSTRACT

Species interactions represent fundamental ecological processes that drive the organismal evolution, but the genetic mechanism of predation traits at the interface of predator-prey interactions is still unclear. In this dissertation, I employed ecological and molecular approaches to determine patterns of evolution and expression of gene families involved in predator-prey interactions and roles of gene duplication in these processes.

Predatory marine snails of the genus *Conus* use venoms, cocktails of conotoxins, to paralyze prey, and conotoxins are encoded by many large gene families. Investigation of the evolution of A-superfamily genes revealed a dynamic of frequent expansion and contraction of this gene familiy. Extensive gene duplication facilitates rapid evolution of these genes, combinations of which lead to dramatic differences in genomic compositions of this gene family among species. Expression of this gene family is also highly variable among closely-related species. Patterns of phylogenetic distribution of expressed genes differ among species, which implies that *Conus* species differentially exploit their venome space. Intraspecific variation in allelic composition and expression of conotoxin genes are associated with changes in dietary breath rather than shifts to certain prey taxon. Patterns of geographic variation exhibited at conotoxin genes result from difference in selective forces that likely stem from geographic difference in prey compositions, because local diversity and geographic variation of conotoxin genes are positively correlated with local diversity and geographic heterogeneity of prey utilization. Similarly, ontogenetic variation of conotoxin gene expression is significantly positively correlated but out of phase with

shifts of dietary diversity, which implies that conotoxin gene regulation is evoked by shifts of dietary breadth through development. Genes associated with species interactions undergo distinct evolutionary pathways and play different roles in these interactions.

In summary, gene duplication facilitates the extensive turnover, rapid evolution and expression divergence of gene families at the interface of predator-prey interactions. Evolution and expression of genes involved in predation are adaptive to changes of prey, and conotoxin gene evolution and expression are highly associated with dietary diversity.

CHAPTER 1 INTRODUCTION

Understanding the relationship between ecological adaptation and molecular evolution has long been of great interest to biologists. Spatial and temporal variation in types and intensities of selection drive phenotypic and genetic differentiation of organisms. Phenotypic variation, the raw materials for selection and adaptation, is determined by differences in genotypes, environments and norms of reaction (Scheiner 1993; West-Eberhard 2005). Modification of genotypes originates from mutation and gene duplication, while norms of reaction determine patterns of interaction between genotypes and environment (De Jong 1990). Variation in norms of reaction under different environmental conditions results in phenotypic plasticity, changes of phenotypes without modifications of protein-coding gene sequences (Thompson 1991; Scheiner 1993; Travis 1994). Phenotypic plasticity is usually achieved by gene regulation, a mechanism that also plays an important role in adaptation (Scheiner 1993; Behera and Nanjundiah 2004). In this dissertation, I uncover the molecular mechanisms of ecological adaptations, focusing on modifications of gene sequences and transcriptional variation.

Biotic interaction serves as a primary driving force of ecological adaptation. Variation in temporal and geographic states (population dynamics and species composition) of participants is usually more frequent than environmental changes, generating strong selection pressure on the counterparts and leading to different outcomes of the interaction (Brodie, Ridenhour, and Brodie 2002; Thompson and Cunningham 2002). Multi-directional adaptations in the context of species

interactions are termed ‘coevolution’, where reciprocal selection is exerted on both sides and may lead to escalation of phenotypic changes at the interface (Yoshida et al. 2003; Thompson 2005b; Thompson 2005a). But interspecific interactions do not necessarily lead to coevolution or even arms races of participants if traits at the interface are non-heritable, pleiotropic, or not tightly coupled (Thompson 1986; Brodie, Ridenhour, and Brodie 2002; Thompson, Nuismer, and Gomulkiewicz 2002; Yoshida, Hairston, and Ellner 2004; Thompson 2005a). To assess the coupling status between participants, it is necessary to determine the evolution of traits at this molecular interface in response to change of counterparts.

Biotoxins are efficient weaponry device employed for predation and defense. Organisms gain toxicity via internal production of toxic chemicals or peptides, or acquire it through engulfing other toxic organisms or symbiotic microorganisms (Tu 1988). Venom is a subset of the realm of biotoxins and is strictly defined as toxic biosubstrates that are secreted and injected into targets. Venomous animals range from one of the most basal phyla of invertebrates, Cnidaria, to the most derived vertebrates and include many well-known animals such as snakes, spiders and scorpions. Study of venom evolution enables us to better understand the impact of biotic interactions on molecular evolution and adaptation of ecologically-relevant genes.

Conus as a study system

Predatory marine snails of the genus *Conus* contains more than 500 species widely distributed in the Tropical Pacific, Indian and Atlantic ocean (Röckel, Korn, and Kohn 1995; Duda and Kohn 2005). *Conus* species possess some unique ecological features in terms of their habitat, distribution and dietary ecology. Multiple species tend to co-occur in the same location, as

illustrated by the coexistence of 36 species on reefs in northeast Papua New Guinea (Kohn 2001). *Conus* tend to specialize more on prey species than on microhabitats: congeners with high overlap in microhabitat utilization have different prey specializations (Kohn 1971; Kohn and Nybakken 1975). Therefore, dietary specialization is essential in the ecological diversification of *Conus*.

Conus produce venoms, cocktails of neurotoxins termed conotoxins, to capture prey. Venom is synthesized in venom ducts and injected into targets through harpoon-like radular teeth (Bergh 1895; Shaw 1914; Kohn 1956; Olivera 2002). *Conus* show tremendous differences in venom composition among and within species (Conticello et al. 2001; Olivera 2002; Jakubowski et al. 2005; Davis, Jones, and Lewis 2009; Rivera-Ortiz, Cano, and Mar í2011), and each species secretes a unique mixture of possibly 100-200 peptides (Olivera 2002). Conotoxins are divided into 11 pharmacological families (α, χ, δ, ε, γ, ι, κ, μ, ω, ρ and σ) based on their function and cysteine frameworks. Each type of conotoxins is characterized by unique arrangements of cysteines and three dimensional conformation of mature toxins (Olivera 2002). Different types of conotoxins block different ion channels and neuronal receptors of the prey: for example, α-conotoxins are selective blocker of nicotinic acetylcholine receptors, δ- and μ-conotoxins target voltage-gated sodium channels but differ in the binding sites, and ω-conotoxins block calcium channels (Terlau and Olivera 2004; Ekberg, Craik, and Adams 2008; Kaas, Westermann, and Craik 2010). Members of each pharmacological family also exhibit differences in their targets and binding efficiency. For example, α-conotoxins differ in specificity and affinity of combinations of nicotinic acetylcholine receptor subtypes (Terlau and Olivera 2004; Tsetlin,

Utkin, and Kasheverov 2009), and μ-conotoxins produced by different species target different sodium channel isoforms (Terlau and Olivera 2004; Wilson et al. 2011).

Conotoxins are encoded by many large gene families (A, D, I, J, L, M, O, P, S, T, V, Y superfamilies), which are comprised of genes with conserved prepro, signal and 3' un-translated regions and highly variable toxin-coding regions (Kaas, Westermann, and Craik 2010). These gene families expand through gene duplication and their members are among the fastest evolving protein-coding genes of metazoans (Duda and Palumbi 1999). Expression of these genes is highly plastic among species (Duda and Palumbi 2004; Duda and Remigio 2008), and differential expression may have contributed to shifts of dietary niches of this genus. Conotoxin genes are subject to strong positive selection (Duda and Palumbi 1999; Puillandre, Watkins, and Olivera 2010), but the source of selection is unclear.

Prey is a plausible factor influencing evolutionary trajectories of conotoxin genes because conotoxins are primarily utilized for predation. Though conotoxins are suggested to be used for evasion of predators and interspecies communication (Olivera 2002), both scenarios are very rarely observed or rigorously tested. Interspecific differentiation and geographic variation of conotoxin genes are suggested to be related with dietary differences (Duda and Palumbi 2004; Duda 2008; Duda et al. 2009; Duda and Lee 2009), but it is unclear the impact of dietary specialization on evolution of venom genes and if members of conotoxin gene families exhibit similar patterns of association with diet.

In this dissertation, I focused on the molecular evolution of conotoxin genes and their associations with prey. I studied evolution of conotoxin genes from perspectives of sequence divergence and gene regulation (the stage of transcriptional variation), and determined patterns of variation among and within species. I raised four relevant questions and addressed them in the following chapters.

Chapter 2. What are the evolutionary pattern of conotoxin genes among species and the role of gene duplication in this process?

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, I 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. I recovered sequences of this gene family from genomic DNA of four closely related species of *Conus* and reconstructed the evolutionary history of these genes. This study is the first to directly recover conotoxin genes from *Conus* genomes to investigate the evolution of conotoxin gene families. 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 multi-gene 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*. My 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.

Chapter 3. Is conotoxin gene expression variable between species?

Regulation of gene expression plays an important role in development of phenotypic variation. Venom composition varies dramatically among and within species of predatory gastropods of genus *Conus*. In addition to genetic mechanisms associated with extensive gene family turnover and rapid evolution, patterns of conotoxin gene expression may also induce hypervariablility in inter- and intra-specific venom composition. To determine the impact of gene expression on venom differentiation among species, I described expression patterns of A-superfamily conotoxin genes of a set of four closely related *Conus* species by comparing venom duct gene transcripts with genomic profiles of this gene superfamily. I also incorporated the community phylogenetic approach to evaluate the phylogenetic organization of expressed genes. Less than 50% of A-superfamily genes are expressed in each species. These species co-express limited number of orthologous genes that exhibit different expression levels. Expression strategies differ among species: some species express phylogenetically closely-related genes (under-dispersion) while others express more distantly-related genes (over-dispersion). Differences in phylogenetic structure of gene expression among species and limited coexpression of orthologous loci show

that *Conus* species differentially exploit their venome space. Conotoxin gene expression also appears to vary within species. The ω (d_N/d_S) values of expressed genes are higher than those of the unexpressed and ancestral genes, which imply that expression exposes genes to strong positive selection and facilitates the rapid evolution and divergence of these genes. Most inparalogs are not expressed simultaneously, suggesting that expression divergence among redundant gene copies is rapidly established after gene duplication. As the first study that directly compares transcriptomic and genomic compositions of conotoxin genes of four closely-related *Conus* species, this study revealed the dramatic variation of conotoxin gene expression and differential utilization of venome space among species that is potentially widely applicable to other venomous organisms. I determined patterns of phylogenetic organization of conotoxin gene expression that are applicable to other ecologically-relevant multi-gene families involved in adaptation.

Chapter 4. Is geographic variation of allelic distribution of conotoxin genes related with dietary differences?

Biotic interactions shape the evolutionary trajectories of species. Geographic heterogeneity in the nature of these interactions creates a geographic mosaic of selection regimes that may result in the differentiation of local populations of widespread species. In predator-prey interactions, variation in traits associated with feeding ecology is correlated with variation in diets, but little is known about the genetic processes linked to this association. Here I report that patterns of geographic variation exhibited at conotoxin genes of *Conus ebraeus* are driven by geographic heterogeneity in prey utilization. Conotoxin loci show contrasting patterns of diversity and are subject to different modes and intensities of selection among local populations. These

populations also show distinct patterns of prey utilization. The diversity of conotoxin genes is positively associated with prey diversity at each locality, but effects of local selective forces operating at these genes may be episodic or overwhelmed by extensive gene flow in some cases. This work illustrates that geographic mosaics of biotic interactions drive the evolution and local adaptations of predators, processes that are widely applicable to other organisms. In addition, genes associated with species interactions undergo distinct evolutionary pathways, implying differences in their roles in these interactions.

Chapter 5. Are changes of expression of conotoxin genes and prey utilization intricately linked during ontogeny?

Characters involved in predation exhibit phenotypic and developmental variation in response to changes of prey, a mechanism that can be achieved by epigenetic variation. To assay how regulation of genes involved in predation changes in response to shifts in diet through ontogeny, I evaluated patterns of conotoxin gene expression and their association with dietary compositions of vermivorous marine snail species, *Conus ebraeus*. *Conus* species utilize venom, a cocktail of neurotoxins encoded by members of many gene families, to capture prey. I collected juveniles, subadults and adults of *C. ebraeus* from Pago Bay, Guam, identified prey species from their feces, and quantified expression levels of six conotoxin genes relative to expression levels of an endogenous β-tubulin gene. Results revealed that diets of *C. ebraeus* change through development and follow a trend from being more generalized to specialized to generalized. Expression of conotoxin genes is highly variable among individuals, but variation in gene expression is not directly related with prey taxon or sexual maturity. Average levels of expression of these genes undergo a process of increase, decrease and then increase during

ontogeny, a pattern that is significantly positively correlated with but delayed relative to shifts of dietary diversities. This implies that variation in conotoxin gene expression is facultatively affected by changes in diet, and up-regulation of conotoxin genes is concordant with broader dietary spectrum. Rather than strict canalization of gene expression in each developmental stage, expression of genes at the interface of the predator-prey interaction is plastic with changes of dietary diversity.

References

Behera N, Nanjundiah V. 2004. Phenotypic plasticity can potentiate rapid evolutionary change. *J Theor Biol.* 226:177-184.

Bergh R. 1895. Beitr ägе zur Kenntniss der Coniden. *Nova Acta Leopold.* 65:69-214.

Brodie ED, Ridenhour BJ, Brodie ED. 2002. The evolutionary response of predators to dangerous prey: Hotspots and coldspots in the geographic mosaic of coevolution between garter snakes and newts. *Evolution.* 56:2067-2082.

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.

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

De Jong G. 1990. Quantitative genetics of reaction norms. *J Evol Biol.* 3:447-468.

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

Duda TF, Palumbi SR. 1999. Molecular genetics of ecological diversification: Duplication and rapid evolution of toxin genes of the venomous gastropod Conus. *Proc Natl Acad Sci USA.* 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.* 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.

Ekberg J, Craik DJ, Adams DJ. 2008. Conotoxin modulation of voltage-gated sodium channels. *Int J Biochem Cell Biol.* 40:2363-2368.

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.

Kaas Q, Westermann J-C, Craik DJ. 2010. Conopeptide characterization and classifications: An analysis using Conoserver. *Toxicon.* 55:1491-1509.

Kohn AJ. 1956. Piscivorous gastropods of the genus *Conus*. *Proc Natl Acad Sci USA.* 42:168-171.

Kohn AJ. 1971. Diversity, utilization of resources, and adaptive radiation in shallow-water marine invertebrates of tropical Oceanic islands. *Limnol Oceanogr.* 16:332-348.

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

Kohn AJ, Nybakken JW. 1975. Ecology of *Conus* on eastern Indian-Ocean fringing reefs - diversity of species and resource utilization. *Mar Biol.* 29:211-234.

Olivera BM. 2002. *Conus* venom peptides: Reflections from the biology of clades and species. *Ann Rev Ecol Syst.* 33:25-47.

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.

Rivera-Ortiz JA, Cano H, Mar íF. 2011. Intraspecies variability and conopeptide profiling of the injected venom of *Conus ermineus*. *Peptides.* 32:306-316.

Röckel D, Korn W, Kohn AJ. 1995. Manual of the living Conidae. Vol. I, Indo-Pacific. Wiesbaden:Christa Hemmen Verlag.

Scheiner SM. 1993. Genetics and evolution of phenotypic plasticity. *Annu Rev Ecol Syst.* 24:35-68.

Shaw HQ. 1914. On the anatomy of *Conus tulipa* Linn., and *Conus textile* Linn. *Q J Microsc Sci.* 60:1-60.

Terlau H, Olivera BM. 2004. *Conus* venoms: A rich source of novel ion channel-targeted peptides. *Physiol Rev.* 84:41-68.

Thompson JD. 1991. Phenotypic plasticity as a component of evolutionary change. *Trends Ecol Evol.* 6:246-249.

Thompson JN. 1986. Constraints on arms races in coevolution. *Trends Ecol Evol.* 1:105-107.

Thompson JN. 2005a. The geographic mosaic of coevolution. Chicago: University of Chicago Press.

Thompson JN. 2005b. Coevolution: The geographic mosaic of coevolutionary arms races. *Curr Biol.* 15:R992-994.

Thompson JN, Cunningham BM. 2002. Geographic structure and dynamics of coevolutionary selection. *Nature.* 417:735-738.

Thompson JN, Nuismer SL, Gomulkiewicz R. 2002. Coevolution and maladaptation. *Integr Comp Biol.* 42:381-387.

Travis J. 1994. Evaluating the adaptive role of morphological plasticity. . *in* Wainwright Pc, and Reilly Sm, eds. Ecological morphology. Chicago: University of Chicago Press, p. 99-122.

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.

Tu AT. 1988. Marine toxins and venoms. New York: M. Dekker.

West-Eberhard MJ. 2005. Developmental plasticity and the origin of species differences. *Proc Natl Acad Sci USA.* 102:6543-6549.

Wilson MJ, Yoshikami D, Azam L, Gajewiak J, Olivera BM, Bulaj G, Zhang M-M. 2011. M-conotoxins that differentially block sodium channels Nav1.1 through 1.8 identify those responsible for action potentials in sciatic nerve. *Proc Natl Acad Sci USA.* 108:10302-10307.

Yoshida T, Hairston NG, Ellner SP. 2004. Evolutionary trade-off between defence against grazing and competitive ability in a simple unicellular alga, *Chlorella vulgaris*. *Proc R Soc Lond B.* 271:1947-1953.

Yoshida T, Jones LE, Ellner SP, Fussmann GF, Hairston NG. 2003. Rapid evolution drives ecological dynamics in a predator-prey system. *Nature.* 424:303-306.

CHAPTER 2 EXTENSIVE AND CONTINUOUS DUPLICATION FACILITATES RAPID EVOLUTION AND DIVERSIFICATION OF GENE FAMILIES

This chapter is published as a research article titled "Extensive and Continuous Duplication Facilitates Rapid Evolution and Diversification of Gene Families" in *Molecular Biology and Evolution* (volume 29, pages 2019-2029, doi:10.1093/molbev/mss068), with the coauthor Thomas F. Duda, Jr.

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 *E. coli* (Yamanaka, Fang, and Inouye 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 NSP-like gene family (Fischer et al. 2008) and P450s 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 immunoglobin (Guldner, Godelle and Galtier 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, Wuster, and Thorpe 1996; Olivera 2002; Fry et al. 2003; Fry et al. 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 1999a, 2004; Conticello et al. 2001; Duda & 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 1999a). Conotoxins are encoded by various gene families that

contain some of the fastest evolving genes of metazoans (Duda and Palumbi 1999a). 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 et al. 2009; Duda and Lee 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, Santos, and Olivera 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 million years ago, 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; Nybakken 1978; Duda, Kohn, and Palumbi 2001; Kohn 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 1999a) 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

1. Specimens and genomic DNA extraction

Specimens of *Conus lividus* collected in Hawaii, *Conus diadema* in Panama and *Conus sanguinolentus* in American Samoa, *Conus 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 ℃. 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^{TM} Mollusc DNA kit (Omega Bio-Tek, Doraville, Georgia, USA).

2. Phylogenetic relationships of four *Conus* species and molecular clock analyses

We amplified mitochondrial COI sequences from genomic DNA of our samples with the universal primers LCO1490 and HCO2198 (Folmer et al. 1994) and sequenced the 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 (nucleotide) region of the calmodulin gene with primers described in Duda and Palumbi (1999b) 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 1000 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 MCMC analyses of 20,000,000 generations (log parameters every 1000 generations), using the divergence time of 11 million years 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 1000 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.

3. 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 (Table 2.1). The primers correspond to (i) a relatively conserved sequence region downstream of a known intron position and upstream of the toxin coding region and (ii) 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 artefactual 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-artefactual 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 artefactual sequences and all artefactual sequences were eliminated from the dataset 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).

4. 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, Halldorsson, and Vernot 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 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 (Figure 2.1) and the K80+G model (equal base frequencies, κ=1.6721, α=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 1000 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.

5. 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 4.0, 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.

6. 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 HKY (Hasegawa, Kishino, and Yano 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 million years. 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 million years ago (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 million years). 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 million years (divergence time estimated from molecular clock analyses as described above).

7. Ancestral sequence reconstruction and tests of positive selection

Ancestral sequences of each node were reconstructed with the likelihood-based empirical Baysian 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-artefactual 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 1999a; Puillandre et al. 2010) using the model of one rate across the whole tree and all sites (model=0 and NSsites=0).

Results

1. 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 Table 2.2). 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 Figure 2.1). The minimum pairwise distances of sequences of the calmodulin and tubulin intron among these species with Jukes-Cantor model (Jukes and Cantor 1969) (Table 2.3) only differ at the 2nd or the 3rd 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 (Table 2.3). 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 (*lnL*=-1237.56 with clock vs *lnL*=-1236.60 without clock; Likelihood Ratio Test yielded *P-value*<0.38, *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 (Table 2.4) and the same species

topology as in Fig. 1. Uncorrelated lognormal clock estimation based on concatenated sequences revealed that *C. sanguinolentus* and *C. lividus* separated 0.3 million years ago and *C. diadema* diverged approximately 1.6 million years ago (Figure 2.1 and Table 2.4).

2. 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* (Table 2.5; GenBank accession numbers JF723384-JF723491). The neighbor-joining tree that included all sequences contained no clades comprised exclusively of potential artefactual 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* (Table 2.5), 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 2.2.

3. 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 (type II and III) (Table 2.6, Figure 2.2). 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 non-binary gene tree (Figure 2.2) with the binary species tree (Figure 2.1) 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) (Figure 2.1). Gene duplications have occurred relatively continuously throughout the evolutionary history of these four species (Figures 2.1 and 2.2). 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 (Figures 2.1 and 2.2). The time estimates of all the branches in the species tree (Figures 2.1) sum to 23.9 million years, so the overall duplication rate of this gene family averaged over all four species is 1.13 duplications per million years. In the recent 1.6 million years, the rate of gene birth is 3.71 duplications per million years. Average duplication rates are heterogeneous among species: 26.7 duplications per million years 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* within 0.3 million years, 8 in *C. diadema* within 1.6 million years and 4 in *C. quercinus* within 11 million years (Figures 2.1).

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 million years with variance of $1.69e^{-4}$ and $1.75e^{-4}$, and ranges are 0.049 to 0.123 and 0.051 to 0.113 respectively. The average death rates are 0.0037 and 0.0049 losses per gene per million years for two chains with ranges of 0 to 0.024 and 0 to 0.023 and variances of $1.55e^{-5}$ and $1.43e^{-5}$.

4. 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 2.2 (ln*L*=-836.312 with clock vs ln*L*=-728.571 without clock, *df*=79, P-value<1e-14, length of sequences=81bp, 3rd codon position was included). Average rates of nonsynonymous substitution of A-superfamily genes are high but heterogeneous among these species: 2.7% per million years for *C. lividus*, 1.8% per million years for *C. sanguinolentus*, 1.3% per million years for *C. diadema* and 0.9% per million years for *C. quercinus*. These rates are similar to those reported previously for O-superfamily conotoxin genes (Duda and Palumbi 1999a). The estimated nonsynonymous substitution rates immediately after duplication exhibited a maximum rate of substitutions of 22.9% per million years, and decreased dramatically with greater values of d_S (representing divergence time of paralogs) in quasi-exponential L-shape relationships (Figure 2.3). 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 (Figure 2.4). In addition, several orthologous

loci among *C. lividus*, *C. diadema* and *C. sanguinolentus* were identical in sequence after separation of species while their respective paralogs differed substantially (Figure 2.5 and Table 2.7). We detected very strong positive selection within the toxin-coding regions of functional conotoxin genes (d_N/d_S=1.75) (Table 2.8).

5. 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, while only two loci represent α4/4 conotoxins. The α4/6 and α4/3 types are rarely found in worm-eating species (Puillandre, Watkins, and Olivera 2010) and were exclusively recovered from the genome of *C. lividus* (Figure 2.2). Ancestral sequence reconstruction based on the genealogy in Figure 2.5 and all the aligned conotoxin sequences showed that the nodes ancestral to the gene of α4/3 type might have been pseudogenes (Figure 2.5 and Table 2.9), 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.

1. 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 (Figures 2.2 and 2.4). The average gene duplication rate estimated with the Bayesian method is about 0.08 per gene per million years. 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 million years, 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 million years and range of 0.002 to 0.02 in Lynch and Conery 2000; 0.028 per gene per million years 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 multi-gene families such as olfactory receptor (Nei, Niimura, and Nozawa 2008), vomeronasal receptor (Grus and Zhang 2004), spider venom (Binford et al. 2009), RNase gene families (Zhang, Dyer, and Rosenberg 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 million years ago, the gene duplication rate of 3.71 duplications per million years, or roughly 0.19 per gene per million years, 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 million years, 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, Graham, and Barrett 2011) and either over- or under-estimation of numbers of gene duplication events. It is possible that our approach and our conservative evaluation of non-artefactual toxin sequences failed to recover certain member(s) 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 dataset 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 (Figure 2.6A), but did not affect fine scale measurements of duplication events during short time intervals (Figure 2.6B). 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 (Figure 2.6).

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, while 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, Durand, and Farach-Colton 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, Huzurbazar, and Liberles 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 (Figure 2.2), 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 intron(s) 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).

2. 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 1999a). 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 (Figure 2.3) and the significant negative slope and intercept of the regression (Figure 2.4) 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 million years, 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 amongst 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 M 2008; Halai et al. 2009). Hence nonsynonymous substitutions within the toxin-coding region of redundant genes 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, Santos, and Olivera 1999; Tsetlin, Utkin, and Kasheverov 2009). Inferred from the genealogy (Figure 2.2), new functional α4/6 and α4/3 types emerged by duplication and divergence from the common α4/7 type (Figure 2.7). 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 million years). In addition to insertions/deletions within the duplicates, the genealogy and ancestral sequence reconstruction (Figure 2.5 and Table 2.9) 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 Figure 2.8).

3. Model of conotoxin gene family evolution

Redundant gene copies produced by gene duplication mutation can either be neutral and fixed by genetic drift (Ohno 1970) or beneficial and fixed by selection (Francino 2005; Kondrashov and Kondrashov 2006; Bergthorsson, Andersson and Roth 2007). We are unable to assess which mechanism is associated with the fixation of conotoxin gene duplicates. We also 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.

4. 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 million years ago. 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*, while the others were not observed and so were possibly lost from its genome. Meanwhile, numerous putative species-specific duplications occurred in *C. quercinus* (Figure 2.2). 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 (Figure 2.2). Such large differences in composition among species may also be induced by differential gene losses as the numbers of gene losses differ among species (Figure 2.2). 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, Hoffmann, and Storz 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 (Figure 2.4).

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; Nybakken 1978; Kohn 2001). 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.

Acknowledgements

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 an NSF award (IOS-0718379) to Thomas F. Duda, Jr.

Figure 2.1. 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 grey bars at each node represent 95% HPD (highest posterior density) of the time of separation. The scale bar at the bottom represents the estimated time of divergence in units of million years (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 *C. lividus*, green for *C. sanguinolentus*, blue for *C. diadema* and red for *C. quercinus*.

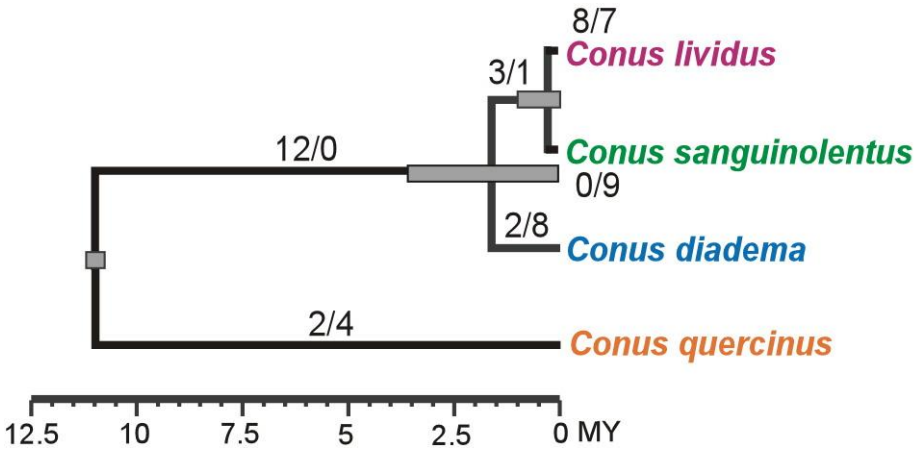


Figure 2.2. Bayesian consensus phylogeny of DNA sequences of single alleles of each putative unique A-superfamily conotoxin locus recovered from four *Conus* species. Lengths of sequences after removal of gaps are 81 nucleotides. Numbers on internal branches indicate Bayesian posterior probabilities. Color-coding scheme: at tips of the genealogy, loci of *C. lividus* are identified as purple dots, *C. diadema* as blue dots, *C. quercinus* as red dots and *C. sanguinolentus* as green dots (consistent color labels as in Figure 2.1); 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; Table 2.6) and shaded in light blue; inferred duplication events are indicated with red stars.

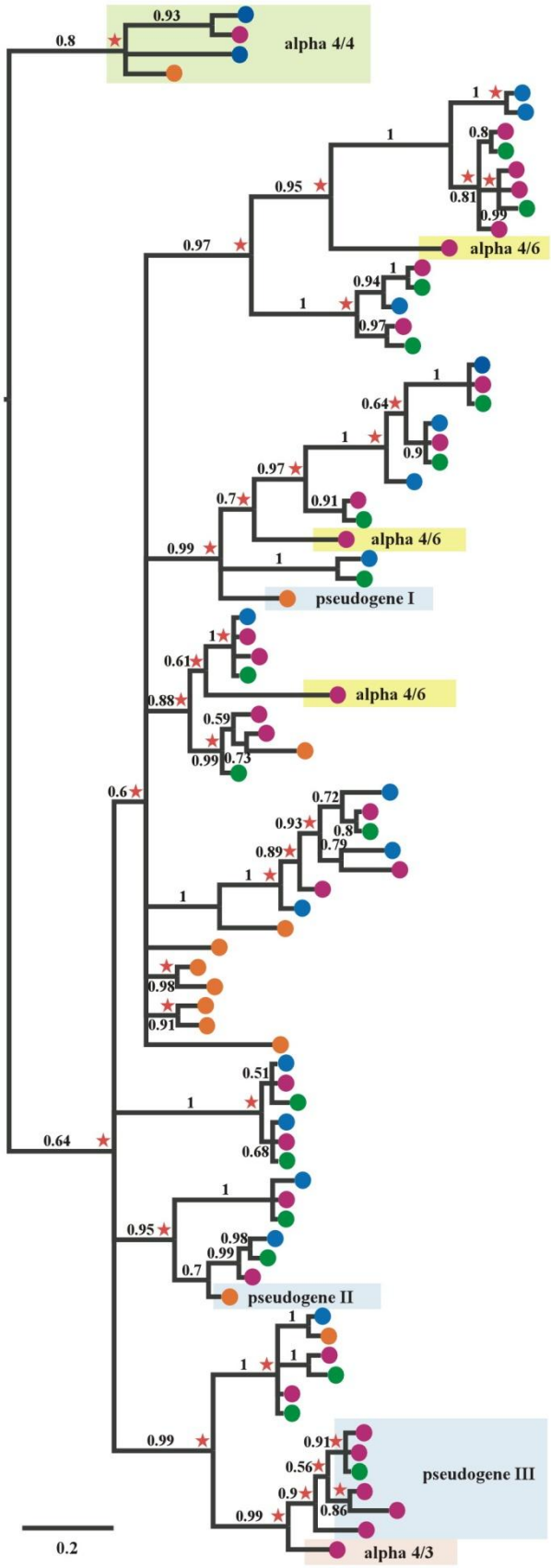

alpha 4/4
alpha 4/6
alpha 4/6
pseudogene I
alpha 4/6
pseudogene II
pseudogene III
alpha 4/3
0.2

Figure 2.3. Plot of estimated pairwise rates of nonsynonymous substitution (y-axis) against the pairwise synonymous divergence (x-axis) of paralogous A-superfamily conotoxin loci of four *Conus* species.
Purple dots represent paralogs from *C. lividus*, green dots from *C. sanguinolentus*, blue dots from *C. diadema* and red dots from *C. quercinus*.

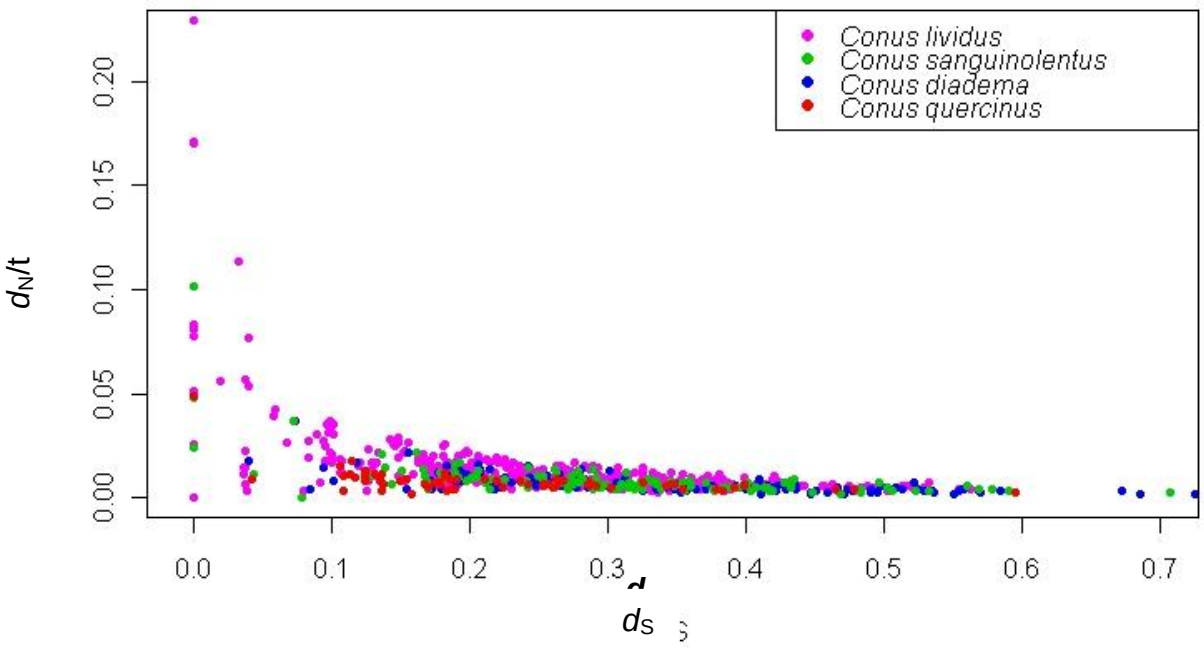


Figure 2.4. Plots of logarithmic transformation of estimated rates of nonsynonymous substitution (y-axis) against the pairwise synonymous divergence (x-axis) of paralogs of each species.

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).

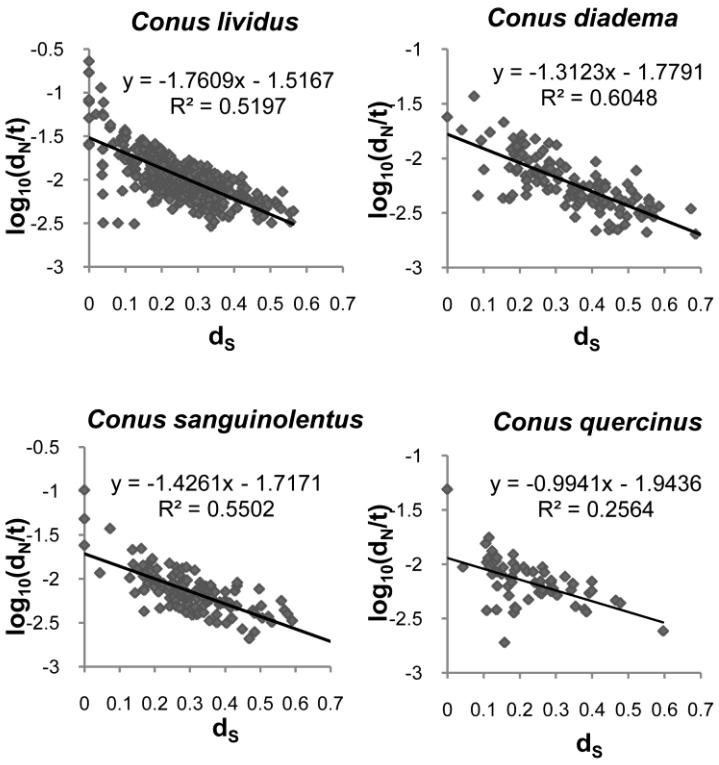


Figure 2.5. Bayesian consensus phylogeny of all A-superfamily sequences recovered from the genome of the four *Conus* species with the HKY+I+G model.
The numbers on the internal branches are Bayesian posterior probabilities. At the tips of the genealogy, sequences from *C. lividus* are labeled as '*livi*' in purple, sequences from *C. diadema* as '*diad*' in blue, sequences from *C. quercinus* as '*quer*' in red and sequences from *C. sanguinolentus* as '*sang*' in green. A sequence from *C. catus* was used to root the tree and is labeled as '*catus*'. Clades are numbered and labeled with bars to serve as examples in the estimation of rates of evolution (Table 2.5). Clades shaded in light blue represent type III pseudogenes. Sequence livi_51 shaded in yellow encodes α4/3 conopeptide. Two nodes labeled as circled A and B serve as examples in ancestral sequences reconstruction (Table 2.7).

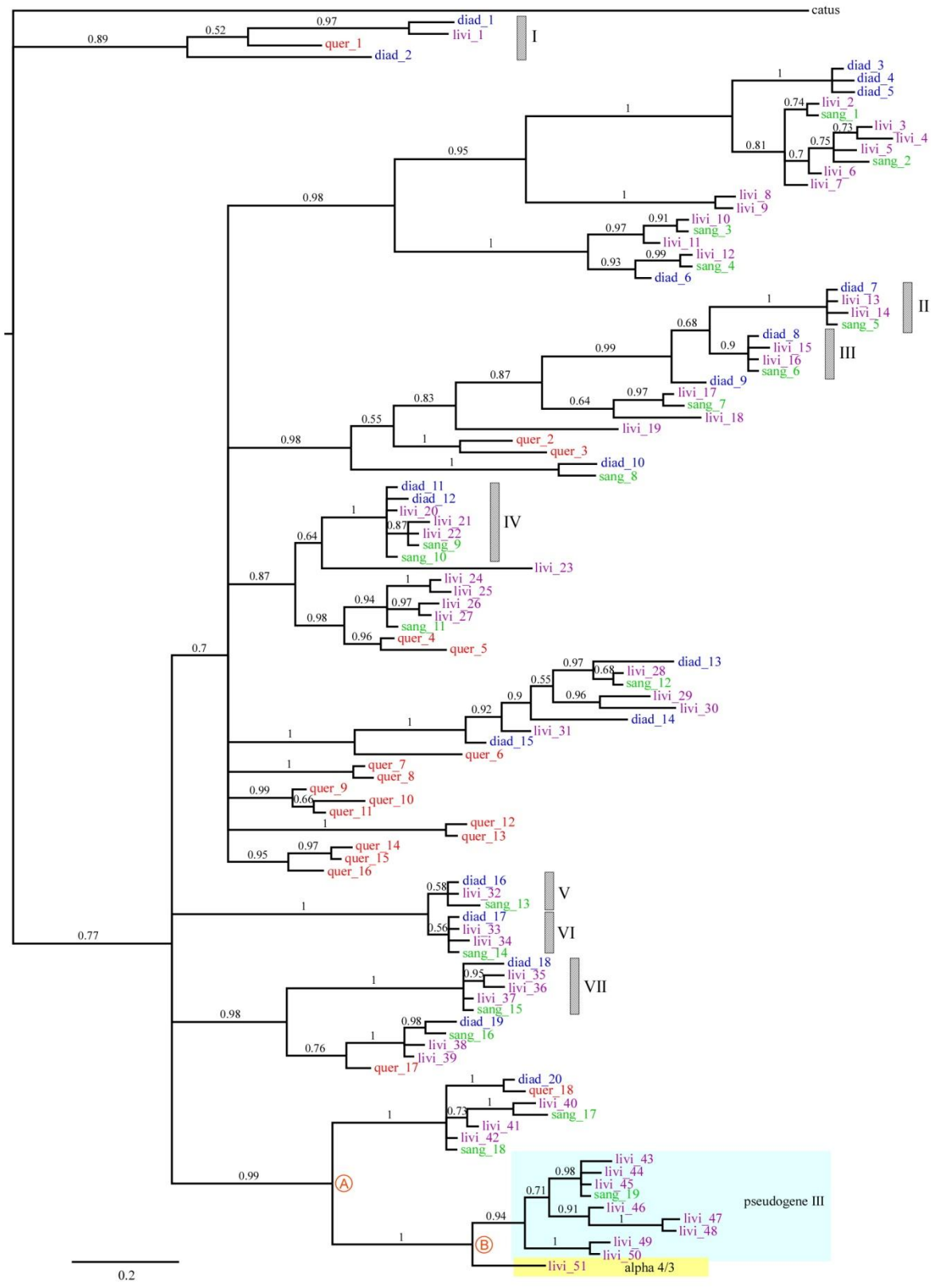
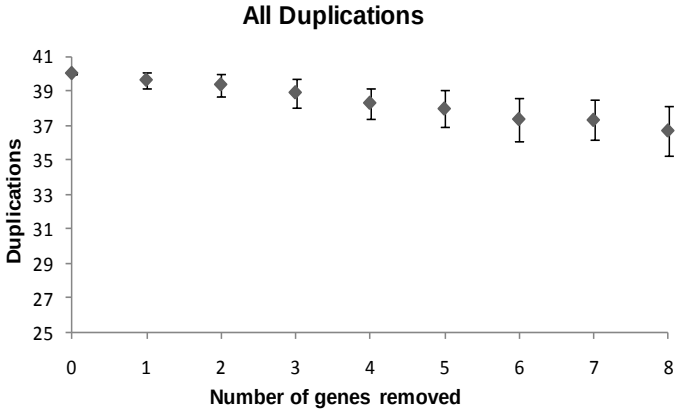

catus
diad_1
livi_1
quer_1
diad_2
I
diad_3
diad_4
diad_5
livi_2
sang_1
livi_3
livi_4
livi_5
sang_2
livi_6
livi_7
livi_8
livi_9
livi_10
sang_3
livi_11
livi_12
sang_4
diad_6
diad_7
livi_13
livi_14
sang_5
II
diad_8
livi_15
livi_16
sang_6
III
diad_9
livi_17
sang_7
livi_18
livi_19
quer_2
quer_3
diad_10
sang_8
diad_11
diad_12
livi_20
livi_21
livi_22
sang_9
sang_10
IV
livi_23
livi_24
livi_25
livi_26
livi_27
sang_11
quer_4
quer_5
diad_13
livi_28
sang_12
livi_29
livi_30
diad_14
livi_31
diad_15
quer_6
quer_7
quer_8
quer_9
quer_10
quer_11
quer_12
quer_13
quer_14
quer_15
quer_16
diad_16
livi_32
sang_13
V
diad_17
livi_33
livi_34
sang_14
VI
diad_18
livi_35
livi_36
livi_37
sang_15
VII
diad_19
sang_16
livi_38
livi_39
quer_17
diad_20
quer_18
livi_40
sang_17
livi_41
livi_42
sang_18
A
B
livi_43
livi_44
livi_45
sang_19
livi_46
livi_47
livi_48
livi_49
livi_50
pseudogene III
livi_51
alpha 4/3
0.2

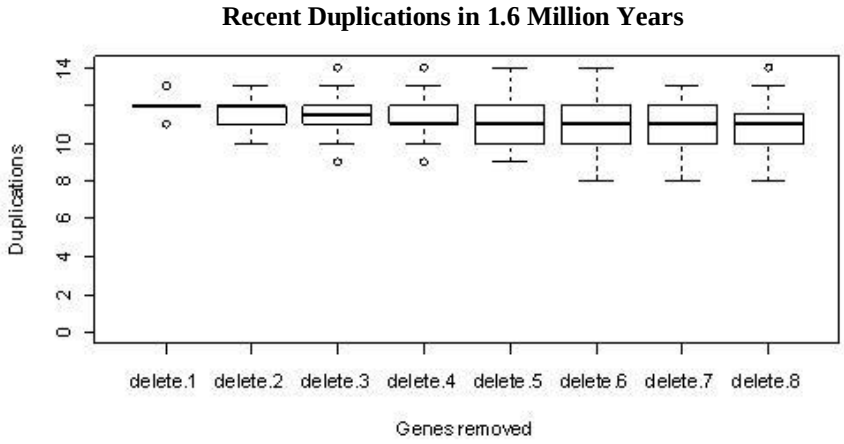
Figure 2.6. Duplications and gene losses inferred from simulated randomly reduced gene pools of four species.
The x-axis is the number of genes removed from the original dataset.

(A) All duplications. Error bars represent standard deviations.

(B) Box plot of possible duplications in the recent one million years. Proportional removal of genes from each species yielded similar results.



A



B

Figure 2.7. Scenario of evolution of α-conotoxin subtypes inferred from ancestral sequence reconstructions.
α4/4 and α4/7 types are ancestral; the α4/7 type underwent several rounds of duplication and new copies evolved into α4/6 and α4/3 subtypes. Asterisks indicate putative duplication events.

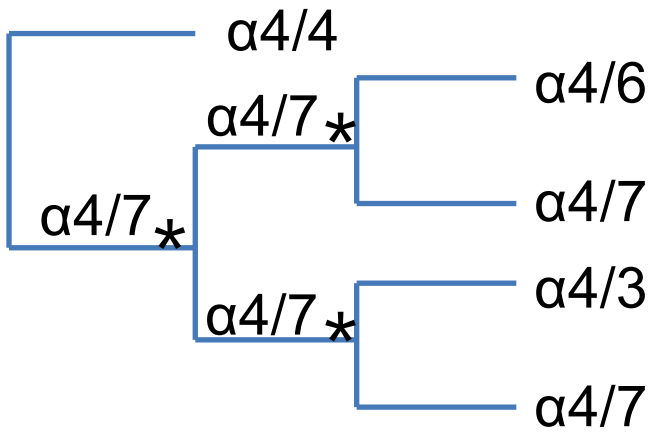


Figure 2.8. Hypothetical scenario of transition from α4/7 to α4/3 conotoxin types.
One sequence each of α4/7, intermediate type and α4/3 are illustrated in the figure as example sequences, and respective predicted amino acid sequences are shaded in light green. Cysteine (*Cys*) codons TGT and TGC and the *Cys* amino acids in the translated peptide are colored in blue. Codons that are altered are shaded in yellow. The disulfide bonds are illustrated as black and red curves above the predicted amino acid sequences connecting the cysteines. One amino acid replacement occurs in the fourth *Cys* in the backbone of the α4/7 toxin gene that prevents formation of the disulfide bond (the red curve). This change results in the ancestral sequences of pseudogene type III found in *C. lividus* and *C. sanguinolentus* (Figure 2.2; Table 2.4). Finally, the fourth codon downstream of the third *Cys* codon changes into a new *Cys* codon which restores the lost disulfide bond (in red) and produces the new α4/3 type.

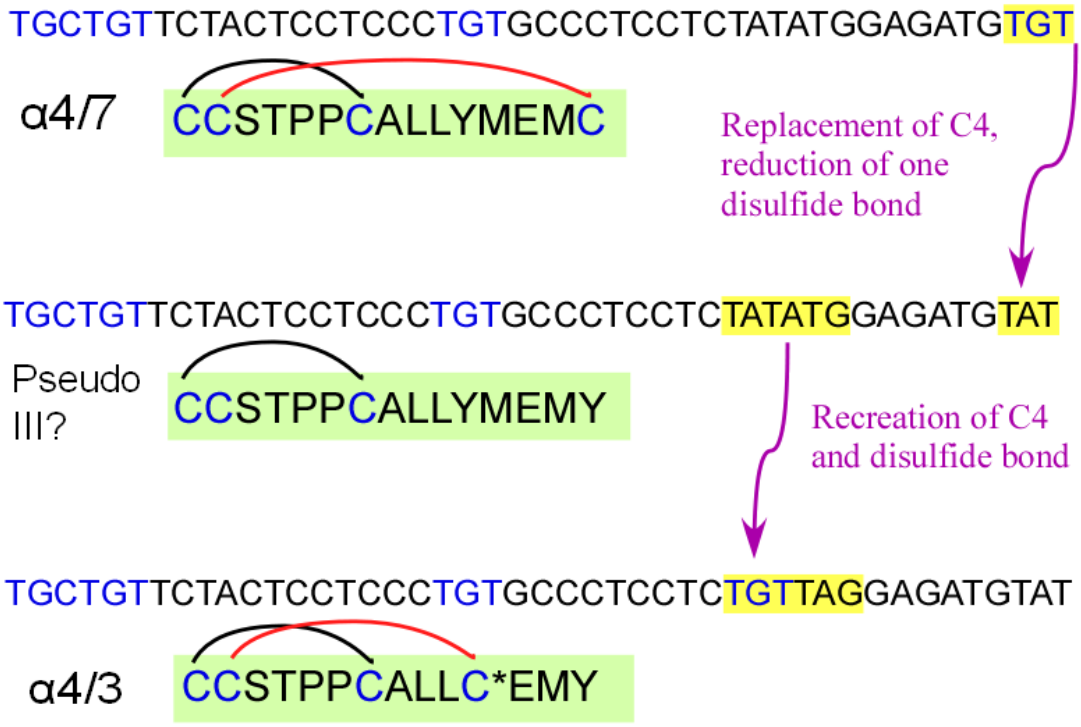


Table 2.1. List of 10 sets of primers that were used to amplify A-superfamily conotoxin genes from the genome of the four closely-related *Conus* species.

Primers	Primer sequence (5' to 3')	Primer location	Primers designed are based on
Apref3	TCGTGCATCTGATGGCAGGAATG	Prepro, downstream to intron(s)	Numerous species, about 100 sequences
Apref3a	TCGTGCATCTGATGGCAGGAATA	Prepro, downstream to intron(s)	*C. lividus, C. musicus, C. regius, C. sponsalis*; 7 sequences
Apref3b	TCGTGCATCTGATGGCAGGGATG	Prepro, downstream to intron(s)	*C. sponsalis, C. consors, C. striatus, C. purpurascens, C. catus*; 7 sequences
Apref3c	TCGTGCATCTGATGGCAGGAGTG	Prepro, downstream to intron(s)	*C. diadema, C. regius, C. imperialis, C. tulipa, C. purpurascens*; 5 sequences
Apref3d	TCGTGCATCTGATGGCAGGAAAG	Prepro, downstream to intron(s)	*C. quercinus, C. omaria, C. textile*; 4 sequences
Apref3e	TCGTGCATCTGATGGCAGGAAGG	Prepro, downstream to intron(s)	*C. aulicus, C. betulinus*; 3 sequences
Apref3f	TCGTGCATCTGATGGCAGGTATG	Prepro, downstream to intron(s)	*C. distans, C. virgo*; 3 sequences
Apref3g	TCGTGCATCTGATGGCAGGAACG	Prepro, downstream to intron(s)	*C. lividus*; 1 sequences
Apref3h	TCGTGCATCTGATGGCAGGAATC	Prepro, downstream to intron(s)	*C. betulinus*; 2 sequences
Apref3i	TCGTGCATCTGATGGCAGGAAGA	Prepro, downstream to intron(s)	*C. regius*; 2 sequences
Apref3j	TCGTGCATCTGATGGCAGGAATT	Prepro, downstream to intron(s)	*C. betulinus*; 1 sequences
CTXAR1	GTCGTGGTTCAGAGGGTCCTGG	3' UTR, pair with all the primers above	Many species, all A-superfamily genes

Table 2.2. GenBank accession numbers for mitochondrial COI sequences, sequences of a tubulin intron and a calmodulin intron of *C. lividus*, *C. sanguinolentus*, *C. diadema* and *C. quercinus*.

Species	Accession Numbers		
	COI	**Tubulin intron**	**Calmodulin intron**
C. lividus	JN831243	JN831248	JN831253
C. sanguinolentus	JN831244	JN831247	JN831251
C. diadema	JN831245	JN831249	JN831252
C. quercinus	JN831246	JN831250	JN831254

Table 2.3. Jukes-Cantor distances with uniform rates among sites calculated among the intron sequences of a tubulin locus and a calmodulin locus between pairs of *Conus* species. Length of the intron sequences of the tubulin locus is 468 base pairs (bp) and of the calmodulin locus is 210 bp after removal of gaps. The minimum of the pairwise distances are presented in the table.

Species pairs	Tubulin locus	Calmodulin locus
C. lividus – C. sanguinolentus	0.011	0
C. diadema – C. lividus	0.024	0
C. diadema – C. sanguinolentus	0.022	0
C. quercinus – C. lividus	0.044	0.049
C. quercinus – C. sanguinolentus	0.053	0.049
C. quercinus – C. diadema	0.053	0.049

Table 2.4. Time estimations of the most common ancestors from strict and relaxed clock models of COI sequences and concatenated COI-calmodulin-tubulin intron sequences of four *Conus* species.

The mean and 95% high probability density (HPD) of time estimations are inferred from results with 1000 burnin (first 1,000,000 generations). MY: million years.

Sequences	**Clock Model**	**t_{MRCA} of *C. lividus*, *C. sanguinolentus* and *C. diadema* (MY)**		**t_{MRCA} of *C. lividus* and *C. sanguinolentus* (MY)**	
		Mean	**95% HPD**	**Mean**	**95% HPD**
COI	Strict	1.6823	0.4701-2.9689	0.1118	0.0015-0.3018
	Relaxed	1.5896	0.2702-3.6378	0.0894	0.0001-0.4059
concatenated sequences (with partition)	Relaxed	1.6191	0.0498-3.5977	0.2954	0.0024-1.0091
	Strict	1.4957	0.1984-2.8548	0.2878	0.0296-0.6458
	Relaxed COI and strict calmodulin-tubulin introns	0.7052	0.0017-2.9332	0.2430	0.0002-1.1529

Table 2.5. Number of colonies sequenced, number of unique sequences (including errors), non-artefactual sequences and putative loci recovered from the genome of each species.

	C. lividus	*C. diadema*	*C. quercinus*	*C. sanguinolentus*
Colonies sequenced (error)	938	434	411	303
Unique sequences	142	76	64	51
Non-artefactual sequences	51	20	18	19
Putative loci	32	18	12	18

Table 2.6. Types of pseudogenes and their predicted amino acid sequences.
Those of the first type contained a premature stop codon upstream of the toxin-coding region. The second and the third types exhibited a non-synonymous substitution in the first or fourth cysteine residue respectively that likely lead to incomplete or improper folding of the mature toxin peptide. Cysteines are highlighted in bold, toxin-coding regions are underlined; * (stop codons) and amino acid replacements that lead to pseudogenization or dysfunction of α-conotoxins are highlighted in red.

Type	Species	Predicted Amino Acid Sequences Examples
I	*C. quercinus*	AADSKAAD*IAQTVRDP**CC**SNPS**C**AQTHPEI**C**RTLM
II	*C. quercinus*	AANNKATDLMARTVRGF**C**SDPS**C**RFRNPEL**C**DWRR*
III	*C. lividus,* *C. sanguinolentus*	AANDKTSAWTTRTVRQS**CC**ATPS**C**ARLYEKVYGRRY*

Table 2.7. Estimates of d_S, d_N and rates of nonsynonymous substitution of orthologous loci (including co-orthologs).

d_S and d_N values were calculated by the Nei-Gojobori method with Jukes-Cantor correction (Nei and Kumar 2000). Rates of nonsynonymous substitution were calculated by dividing d_N with two times the divergence time. Clade labels are consistent with labels in Figure 2.5.

Clade	Sequence pairs	Toxin Type	d_S	Toxin-coding region		
				Length (bp)	d_N	rate (per MY)
I	diad-1 vs livi-1	α4/4	0	51	0.056	1.75%
II	diad-7 vs livi-13	α4/7	0	51	0	0%
III	diad-8 vs livi-16	α4/7	0	51	0	0%
V	diad-16 vs livi-32	α4/7	0	57	0	0%
VI	diad-17 vs livi-33	α4/7	0	57	0	0%
VII	diad-18 vs livi-37	α4/7	0	60	0	0%
IV	diad-11 vs livi-20	α4/7	0	51	0	0%
IV	diad-11 vs livi-21	α4/7	0.037	51	0.029	0.91%

Table 2.8. Results of test of postive selection based on analyses with the Codeml package of PAML 4.3 on toxin-coding regions of putative functional conotoxin genes.
** indicates *P-value* <0.01.

Models	ω	Log likelihood
ω =1	1	-1180.6
ω estimated	1.75**	-1177.3

Table 2.9. Reconstructed DNA sequences and respective amino acid sequences of the ancestral nodes A and B illustrated in Figure 2.5.
Cysteines in the predicted amino acid sequences are highlighted in bold. After removal of gaps from the sequence alignment in the reconstruction process there were only four codons between the 3rd and 4th *Cys* in all the extant sequences utilized for reconstruction (except livi_51), but in reality there should be three more codons for α4/7 toxins. The 4th *Cys* (codon) is predicted to be intact in the ancestral sequence of node A but destroyed in node B. The (codon) position where the 4th *Cys* is supposed to be in the sequence of node B is underlined in both DNA and predicted amino acid sequences.

Node Label	Reconstructed DNA Sequences	Predicted Amino Acid Sequences
A	CCGCAGCCAACGACAAAGCGTCTGCC TCATGCTGTTCTACTCCTCCCTGTGCC ATGCTTTATTGTGGCACGCTGATGC	AANDKASAS**CC**STPP**C**AML Y**C**GTLM
B	CCGCAGCCAACGACAAAGCGTCTGCC TCATGCTGTGCTACTCCTTCCTGTGCC ATGCTTTAT<u>TAT</u>GGCACACTGATGC	AANDKASAS**CC**ATPS**C**AML Y<u>Y</u>GTLM

References

Akaike H. 1974. A new look at the statistical model identification. IEEE T 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 USA.* 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 USA.* 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. Molecular genetics of ecological diversification: Duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc Natl Acad Sci. USA.* 96:6820-6823.

Duda TF, Palumbi SR. 1999b. Developmental shifts and species selection in gastropods. *Proc Natl Acad Sci USA.* 96: 10272-10277.

Duda TF, Palumbi SR. 2004. Gene expression and feeding ecology: evolution of piscivory in the venomous gastropod genus *Conus*. *Proc R Soc Lond B.* 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. 2007. AChBP-targeted α-conotoxin correlates distinct binding orientations with nAChR subtype selectivity. *EMBO J.* 26:3858-3867.

Ellison M FZ, 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, and Liberles D, editors. Evolution after gene duplication: 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. 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.* 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 Monograph.* 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.

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.

McIntosh JM, Santos AD, Olivera BM. 1999. *Conus* peptides targeted to specific nicotinic acetylcholine receptor subtypes. *Annu Rev Biochem.* 68:59-88.

Nei, M. & Hughes, A. L. 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 Univ. Vol. 2, pp. 27–38.

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.

Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. Oxford: Oxford University Press.

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.* doi:10.1093/molbev/msr119.

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: 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 USA.* 105:12950-12955.

Perry GH, Dominy NJ, Claw KG et al. 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, Drummond AJ. 2007. Tracer v1.4. Available from http://beast.bio.ed.ac.uk/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, Massachusetts: 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, and Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* doi: 10.1093/molbev/msr121.

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. 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 USA.* 97:4701-4706.

CHAPTER 3 DIFFERENTIAL EXPLORATION OF VENOME SPACE: PLASTICITY OF EXPRESSION OF VENOM GENES

This chapter will be submitted for publication with the coauthor Thomas F. Duda, Jr.

Introduction

Gene regulation shapes inter- and intra-specific variation in phenotypes and affects organismal response to changes of environmental conditions (Lockhart and Winzeler 2000). Vast phenotypic and behavioral differences among species that possess highly similar genomic sequences attribute to gene regulation (King and Wilson 1975; Enard et al. 2002; Ranz et al. 2003; Khaitovich et al. 2005; Somel et al. 2008). Variation in gene expression facilitates individuality of organisms and phenotypic difference of individuals with identical genotypes (Raser and O'Shea 2005). Differences in patterns of gene expression can be viewed as differential exploitation of 'gene space', genomic regions containing protein-coding genes (Jackson, Hass Jacobus, and Pagel 2004). Messenger RNA transcripts associated with this gene space reflect the functional and adaptive roles of their protein products because mRNA synthesis, the first step of protein production, represents organismal responses to environmental perturbations in real-time (Lockhart and Winzeler 2000).

Retention and functionalization of gene duplicates are affected by variation in gene expression (Ohno 1970; Qian et al. 2010). Duplication promotes expression divergence of genes in the

genome (Li, Yang, and Gu 2005). Divergence in expression of paralogous genes is positively correlated with ages of genes (Gu et al. 2002), and is likely affected by changes of cis- and trans-regulatory elements (Zhang, Gu, and Gu 2004; Li, Yuan, and Zhang 2010; Dong, Yuan, and Zhang 2011). Closely-related paralogs show equivalent or less resemblance in patterns of expression than distantly-related genes (Oakley et al. 2005), and divergence in gene expression can be rapidly established among young duplicates (Gu et al. 2002).

Expression divergence of members of multigene families leads to interspecific differential expression (Tomanek and Somero 2002; Gu et al. 2004; Kawaura, Mochida, and Ogihara 2005; Jovelin et al. 2007). Nonetheless there is a deficit in statistical methods that can be used to assess phylogenetic structures of gene expression in each species. Here we employ community phylogenetic approaches to statistically evaluate phylogenetic patterns of expression among species, assuming divergence in expression is associated with phylogenetic disparity of genes in each species. Webb et al. (2002) proposed three types of phylogenetic community organization: phylogenetically over-dispersed communities that include a nonrandom set of distantly related species, phylogenetically under-dispersed communities that include a nonrandom set(s) of closely related species, and a default state in which the community is comprised of a phylogenetically random set of species. Here we use these three states to describe mechanisms of gene expression employed in each species. By viewing the genomic composition of the gene family as analogous to the species pool and the expressed genes as analogous to a sampling community, we can evaluate patterns of phylogenetic organizations of expressed members of the gene family in each species.

As predicted by Lluisma (2012), venomous organisms may utilize different strategies in expression of venom genes. The venome refers to proteomic compositions of venoms that are recruited from sets of genes encoding many protein families (Fry 2005). The recruiting process to final venome involves regulation of these toxin-related genes. We use the term 'venome space' to describe the combination of toxin-coding genes in the genome of each species. High variability in venom composition can stem from differences in gene expression (i.e., the differential exploitation of the venome space).

Predatory marine snails of the genus *Conus* utilize venoms that include a variety of peptide neurotoxins (conotoxins or conopeptides) to subdue prey, and conotoxins target diverse sets of ion channels and neuronal receptors (Kaas et al. 2012). Venom composition varies dramatically among and within *Conus* species (Olivera 2002; Jakubowski et al. 2005; Davis, Jones, and Lewis 2009; Rivera-Ortiz, Cano, and Mar í2011), which, in part, derives from the dynamics of conotoxin gene family evolution through extensive gene turnover, mutation and rapid evolution (Duda and Palumbi 1999; Chang and Duda 2012). Previous studies revealed the important role of differential expression in interspecific divergence of venoms based on qualitative and quantitative analyses of conotoxin gene transcripts (Duda and Palumbi 2000; Conticello et al. 2001; Duda and Palumbi 2004; Duda and Remigio 2008; Hu et al. 2012). Diversity and levels of expression of genes from each conotoxin superfamily differ among (Conticello et al. 2001) and within species (Hu et al. 2012). Species tend to avoid coexpression of orthologous loci such that closely-related species possess dramatically different gene transcripts in their venom duct cDNA (Duda and Remigio 2008), and differential expression may contribute to shifts in diet (Duda and Palumbi 2004).

Nonetheless, without knowledge of the genomic composition of venome space, it is difficult to differentiate transcriptional variation of single genes from lineage-specific gene duplication/loss, especially with the extensive turnover of conotoxin gene families (Chang and Duda 2012). Description of genomic profiles of the A-superfamily of four closely-related species *C. lividus*, *C. sanguinolentus*, *C. diadema* and *C. quercinus* by Chang and Duda (2012) (Chapter 2) provides a great opportunity to examine patterns of expression of members of this gene family in these species. A-superfamily genes encode α-conotoxins that are selective blockers of nicotinic acetylcholine receptors; α-conotoxins are characterized by their signature cysteine backbone of "**CC**$(X)_m$**C**$(X)_n$**C**" (Santos et al. 2004). A-superfamily genes possess a highly conserved prepro (a precursor region encoding a part of the prepropeptide that are cleaved from the mature toxin in the post-translation stage) and 3' untranslated regions (Santos et al. 2004; Puillandre, Watkins, and Olivera 2010), which allow us to retrieve members of this superfamily from venom duct transcripts with primers that can be designed within conserved regions.

Here we tested the hypotheses that conotoxin genes are differentially expressed among closely-related *Conus* species and that these species differentially exploit their venome space. We also evaluated if conotoxin gene expression varies within species, and examined the selectivity of expressed genes and the role of expression in evolution of gene families. We obtained expression profiles of A-superfamily from venom duct transcripts of four *Conus* species, compared the results with genomic compositions of this gene family in each species, statistically evaluated phylogenetic structures of expressed genes, and assayed patterns of expression and neutrality of gene duplicates.

Materials and methods

1. Specimens

We obtained specimens of *Conus lividus* (from Hawaii), *Conus diadema* (from Panama) and *Conus sanguinolentus* (from American Samoa) from the Mollusk Division collections at the University of Michigan Museum of Zoology. Specimens of *Conus quercinus* (from Hawaii) were provided by Jon-Paul Bingham (University of Hawaii). Venom ducts of these specimens were preserved in RNAlater (Ambion, Inc.) and stored at -20 ℃.

2. Recovery of A-superfamily genes from venom duct transcripts

We extracted mRNA from venom ducts of two individuals each of *C. lividus*, *C. diadema* and *C. quercinus* and one individual of *C. sanguinolentus*, and prepared cDNA following the protocol described in Duda and Palumbi (1999). In an attempt to recover all A-superfamily gene sequences from the venom duct transcriptome, we used a set of 'universal' primers for A-superfamily gene sequences (forward primer: 5'ATGGGCATGCGGATGATGTTCAC 3'; reverse primer: 5' GTCGTGGTTCAGAGGGTCCTGG 3') that anneal to the highly conserved prepro and 3' untranslated regions. We amplified gene sequences from venom duct cDNA of each individual, cloned PCR products, screened and sequenced expected inserts following the approach described in Chang and Duda (2012). We repeated this whole experimental procedure for each individual to help identify non-artefactual sequences (described in the next section). We generated sequence diversity curves (Duda and Remigio 2008) for each individual in each round of amplification to determine if enough inserts were sequenced to recover as many A-superfamily conotoxin gene transcripts as possible from venom duct cDNA.

3. Determination of transcribed loci

We examined sequence chromatograms in Sequencher v4.8 (Gene Codes Corporation), and manually aligned sequences in SE-AL 2.0 (Rambaut 2002) based on similarities of nucleotide and translated amino acid sequences especially the cysteine backbone of α-conotoxins as described by Chang and Duda (2012). We determined non-artefactual sequences by comparing sequences recovered from two rounds of PCR from venom duct cDNA with genomic profiles of this gene family in each species (Chang and Duda 2012; GenBank Accession Numbers JF723384-JF723491), and designated sequences recovered from both rounds of experiments or from both venom duct cDNA and genomic DNA of each species as expressed non-artefactual sequences. We constructed a neighbor-joining tree of all sequences (including artefactual sequences) with the K80 (Kimura 1980) model in PAUP 4.0 (Swofford 2002) to ensure that each major clade contains at least one non-artefactual sequence. We allocated artefactual sequences to respective groups (putative expressed alleles) represented by at least one non-artefactual sequence based on their genetic similarities and clustering patterns in the neighbor-joining tree.

4. Phylogenetic analyses of expressed genes

We performed model selection in jModelTest v0.1.1 (Posada 2008) with non-artefactual gene sequences recovered from venom duct cDNA of four species. We constructed a Bayesian consensus phylogeny of non-artefactual genes with MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001) (10,000,000 generations, 4 Markov chains, 2 runs and 25% burnin) using the best model HKY (Hasegawa, Kishino, and Yano 1985)+I and one A-superfamily gene sequence from *C. catus* to root the tree (GenBank accession number FJ868066).

5. Test of differential expression patterns among species

We developed a statistical approach based on community phylogenetic methodologies (Webb et al. 2002; Emerson and Gillespie 2008; Webb, Ackerly, and Kembel 2008) to determine phylogenetic distributions of gene expression in the venome space of each species. This approach takes into account phylogenetic signals of genomic profiles, accepts qualitative input of lists of expressed genes, and evaluates the pattern of distribution of expressed genes in the genealogy of genomic components of each species. Three possible results are over-dispersion, under-dispersion, or random distribution of expressed genes (no structure, the default state). Similar to tests of community assemblies, we treated genes in the genome of each species as 'the species pool' and expressed genes as 'the composition of species of a single community'. We used indices such as Mean Phylogenetic Distance (the average distance of pairwise comparisons of samples; MPD), Mean Nearest Phylogenetic Taxon Distance (the average distance of the most closely related samples ;MNTD), Net Relatedness Index (the standardized MPD differences between the null model and sampling community; NRI) and Nearest Taxon Index (the standardized MNTD differences between null and observed communities; NTI) (Webb et al. 2002) to evaluate the over- and under-dispersion of expression of these genes. Observed MPD and MNTD smaller than random as well as positive NRI and NTI values suggests under-dispersion of gene expression, while observed MPD and MNTD larger than random and negative NRI and NTI values suggests over-dispersion. No significant difference between observed and random MPD and MNTD as well as NRI and NTI not significantly different from zero implies lack of structure of gene expression.

To build separate genealogies for each species, we pruned the phylogeny of conotoxin genes recovered from genomic DNA of the four species (obtained from (Chang and Duda 2012)) with maximum-likelihood and HKY+G model in PAUP 4.0 (Swofford 2002). We imported the pruned genealogy of A-superfamily genes for each species into Phylocom 4.2 (Webb, Ackerly, and Kembel 2008) and a list of genes expressed in the venom duct of that species, and evaluated the phylogenetic structure of expressed genes with the *Comstruct* command. *P*-values are percentages of MPD and MNTD values obtained from 10,000 generations of random drawings of samples from the genomic profile that are smaller than observed values.

6. Differential expression among and within species

We quantified absolute levels of expression of each allele in each individual with counts of sequenced colonies containing inserts of that expressed allele and its respective artefactual sequences, and quantified levels of expression of each locus by combining counts of all alleles of that locus. We tested independence of levels of conotoxin gene expression between two individuals of the same species of *C. lividus*, *C. diadema* and *C. quercinus* by Fisher's Exact Tests with *fisher.test* function in R v2.15.0 (R Development Core Team 2012). *P*-values were estimated by Monte Carlo simulations of 1,000,000 generations.

To standardize levels of expression among species and to eliminate bias of sample sizes, we estimated relative expression of a locus by dividing counts of each locus expressed by each individual with total counts of colonies of that individual. Similarly, we calculated relative expression of each locus of each species by dividing total counts of that locus with total counts of colonies sequenced for that species.

7. Estimation of ω (d_N/d_S) of expressed genes

We used a maximum-likelihood approach and the branch-site model implemented in the Codeml package of PAML 4.3 (Yang 2007) to test the neutrality of expressed conotoxin genes. We used this approach to determine if ω (d_N/d_S ratio) values of branches leading to currently expressed genes are significantly different from ω of the remainder branches in the A-superfamily genealogy. Three types of pseudogenes have been found in the genomic DNA of these species, and functionality of type III pseudogenes is still unclear (Chang and Duda 2012). We excluded type I and II pseudogenes and a short sequence (livi_51, a α4/3 type conotoxin) from analyses to incorporate more information of toxin-coding regions. We examined toxin-coding regions for two sets of genes (with or without type III pseudogenes) and phylogenies pruned with the approach described above. We set one ω rate across the whole tree as the null model and proposed three alternative models. The first model assumes that branches leading to expressed genes exhibit a different ω value from that of branches leading to unexpressed and ancestral gene sequences (ω_2 for terminal branches of expressed loci, ω_1 for the rest of the branches). The second alternative model assumes the opposite (ω_2 for the terminal branches of unexpressed genes, ω_1 for the rest of the branches). The third model assumes that branches leading to expressed, unexpressed and ancestral genes respectively exhibit different ω values (ω_1 for ancestral branches, ω_2 for terminal branches of unexpressed genes, ω_3 for terminal branches of expressed genes). We also used a full model permitting variable ω values for each branch in the genealogy. *P*-values were estimated with likelihood-ratio tests of the null model with alternative models.

8. Expression divergence of gene duplicates

We investigated the relationships between expression divergence of conotoxin genes and time of divergence and rates of evolution of these genes. Divergence time between paralogous genes is represented by the number of synonymous substitutions per synonymous site (d_S) between pairs of paralogs, while rates of evolution are approximated with ω (d_N/d_S) (d_N: the number of non-synonymous substitutions per non-synonymous site). We estimated pairwise d_S (based on prepro and toxin-coding regions) and d_N values (based on toxin-coding regions) of A-superfamily genes in the genome of each *Conus* species in MEGA v5.05 using the Nei-Gojobori method with Jukes-Cantor correction (Nei and Gojobori 1986). For gene pairs with d_S=0, we arbitrarily converted these zero-value d_S estimate to 0.004 for estimation of ω because the synonymous substitution rate is estimated to be 0.004 per million years (Chang and Duda 2012). Previous studies designated expression divergence of gene duplicates in yeast as fold-changes of expression levels in microarray analyses (Gu et al. 2002; Oakley et al. 2005), but this approach is not applicable to our dataset which was obtained from enriching and sequencing of genes from cDNA libraries. Instead we divided patterns of expression divergence of pairs of paralogous conotoxin genes into three categories: category 1 includes cases where both paralogs are unexpressed, category 2 includes cases where only one gene is expressed and the other is not, and category 3 includes cases where both genes are expressed. We compared d_S and ω values among three categories and tested if the mean between/among categories are identical with t-tests and ANOVA in R v2.15.0. All scripts used in this study are available upon request.

Results

1. Percentages of expression

Conotoxin genes of all four species are partially expressed in venom ducts. We sequenced 487, 167, 135 and 112 colonies from two individuals each of *C. lividus*, *C. diadema*, *C. quercinus* and one individual of *C. sanguinolentus* (Table 3.1). After identification and elimination of artefactual sequences, we determined 18, 3, 4 and 5 putative alleles for each species. All the putative alleles recovered from venom duct cDNA were retrieved from the genomic DNA of each species previously (Chang and Duda 2012). Comparison of these alleles with putative A-superfamily loci identified from genomic DNA of these species (Chang and Duda 2012) determined these expressed alleles as representatives of 13 loci in *C. lividus*, three in *C. diadema*, three in *C. quercinus* and five in *C. sanguinolentus*. In comparison with sizes of the A-superfamily of each species (32 genes in *C. lividus*, 18 in *C. diadema*, 12 in *C. quercinus* and 18 in *C. sanguinolentus*), 40.6% of genes in *C. lividus*, 16.7% in *C. diadema*, 25.0% in *C. quercinus* and 27.8% in *C. sanguinolentus* are expressed in venom ducts (Table 3.1).

2. Diversity of expressed genes

Out of a total of 24 loci expressed in four *Conus* species, 22 appear functional because translated amino acid sequences of these genes represent putatively potent α-conotoxins. Among the three types of pseudogenes found in the genomes of these species by Chang and Duda (2012), only two loci (with three unique alleles) of type III pseudogenes with a nonsynonymous substitution in the fourth cysteine codon position of the cysteine backbone are expressed exclusively in *C. lividus* (Figure 3.1), while other pseudogene types do not show evidence of expression. A-superfamily genes in these species encode four types of α-conopeptides (α4/4, α4/7, α4/6 and α4/3; Chang and Duda 2012), among which genes of the α4/7 type dominate both genomic and transcriptomic compositions of the venome space (Figure 3.1). One of the three loci of the α4/6

type and the only locus of the α4/3 type that is exclusively found in *C. lividus* are expressed. An orthologous locus of the α4/4 type characterized from genomes of *C. diadema*, *C. quercinus* and *C. lividus* is present in expression profiles of *C. diadema* and *C. quercinus* but not *C. lividus* (Figure 3.1).

3. Interspecific differential expression

Species exhibit limited coexpression of orthologous conotoxin genes. Interspecific divergence of conotoxin gene transcripts in venom duct cDNA libraries can be represented by numbers of orthologous loci coexpressed by more than one species (Duda and Remigio 2008). Limited numbers of orthologous loci are coexpressed among *Conus* species examined here, and no orthologous genes are expressed simultaneously by more than two species (Table 3.2). *C. lividus* does not coexpress any orthologous gene with *C. diadema* or *C. quercinus*, while *C. diadema* only expressed one orthologus locus concurrently with *C. sanguinolentus* (diad_10 and sang_8) or *C. quercinus* (diad_1 and quer_1) (Figure 3.1; Table 3.2). Only two orthologous genes are present in venom duct transcripts of the sister species *C. lividus* and *C. sanguinolentus* (livi_2 and sang_1; livi_10, livi_11 and sang 3; Figure 3.1; Table 3.2), even though these two species diverged less than 0.3 million years ago and may actually represent genetically differentiated populations of *C. sanguinolentus* (Chang and Duda 2012; Duda et al. 2012). Sequences of these orthologs are identical (i.e., sequence livi_2 is the same as sang_1 and livi_10 is the same as sang_3), suggesting recent divergence of these species. Moreover, the only four orthologous genes coexpressed between species exhibit heterogeneity in levels of expression in each species (Figure 3.2A).

Estimation of phylogenetic indices of expression revealed contradicting patterns of conotoxin gene expression among species. MPD values smaller than random, positive NRI values, MNTD values smaller than random and positive NTI detected for *C. lividus* and *C. sanguinolentus* (Table 3.3) are signals of phylogenetic under-dispersion of expressed genes in these species. *C. sanguinolentus* exhibits a stronger pattern than *C. lividus* because of the significance of its MNTD value (Table 3.3). On the other hand, genes expressed by *C. diadema* and *C. quercinus* are phylogenetically overdispersed, as demonstrated by larger MPD and MNTD values than random and negative NRI and NTI (Table 3.3). But significance is only reached for MPD and MNTD values of genes expressed in *C. diadema* (Table 3.3).

4. Intraspecific variation in expression

Diversity and levels of expression differ significantly between individuals of *C. lividus* and *C. quercinus* but not in C. *diadema*. Out of a total of 13 loci expressed in *C. lividus*, one individual expressed only seven loci while the other transcribed eight (Figure 3.2C). The only two genes expressed by both individuals (livi_10 and livi_45) exhibit different levels of expression between individuals, and this pattern is also observed for the two loci expressed by both individuals of *C. quercinus* (Figure 3.2C). Fisher's exact tests revealed significant difference in expression between individuals of *C. lividus* and *C. quercinus* (*P*-values for each species are less than 0.0001). Nonetheless, individuals of *C. diadema* exhibited no substantial difference in either diversity or levels of expression of conotoxin genes (Figure 3.2C; Fisher's exact test *P*-value=0.217).

5. ω (d_N/d_S) values of contemporarily expressed genes

Currently expressed genes exhibit a higher ω value than unexpressed and ancestral genes. For both gene sets examined, the first alternative model with two ω rates (ω2 for expressed terminal branches and ω1 for the rest of branches in the genealogy) is significantly better than the null model with only one ω value across the whole phylogeny, and ω2 is much larger than ω1 (Table 3.4). Assigning three variables of ω to the genealogy (ω_1 for ancestral branches, ω_2 for terminal branches of unexpressed genes, ω_3 for terminal branches of expressed genes) showed no significant improvement in likelihood scores, but expressed genes still maintain a larger ω value (Table 3.4). Moreover, when expressed terminal branches are forced to share the same ω as the ancestral branches, the ω value of expressed genes is still larger than that of the temporally non-expressed terminal branches, though no significant improvement of the model is detected (Table 3.4). These results consistently revealed heightened ω values of branches leading to expressed genes, which still holds when we examined this pattern for genes in individual species (Table 3.5).

6. Expression divergence of conotoxin genes related with time and rates of evolution

Three categories of expression divergence (i.e. both genes are unexpressed, only one gene is expressed, and both are expressed) of pairwise comparisons of members of A-superfamily in each *Conus* species did not exhibit prominent differences in d_S and d_N/d_S (except for changes of d_S values in *C. lividus*). Average d_S and d_N/d_S values are almost identical among three categories for *C. diadema, C. sanguinolentus* and *C. quercinus*, and ANOVA analyses did not reveal any significance of difference. We combined pairs of genes of categories 1 and 3 (both genes are either unexpressed or expressed simultaneously) into a group of 'no expression divergence', and viewed category 2 (only one gene is expressed) as a group of 'expression divergence'; t-tests

revealed no significant difference in d_S and d_N/d_S between groups. For *C. lividus*, ANOVA analyses and t- show no difference in average d_N/d_S values among categories or between groups. But average d_S values for categories 2 and 3 are significantly smaller than category 1 (ANOVA results: estimated difference of mean d_S between category 1 and 2 is -0.051, *P*-value<0.0001; estimated difference of mean d_S between category 1 and 3 is -0.095, *P*-value<0.0001); results of t-tests between the two groups of expression defined here are not significant (*P*-value=0.0796). Similarly, average d_N/d_S value for category 2 is significantly higher than category 1 (ANOVA: estimated difference between categories 2 and 1 is 2.554, *P*-value=0.03).

No concordant patterns of expression were detected between paralogous genes from lineage-specific duplications (defined as inparalogs by Koonin (2005)): most inparalogs are either non-expressed or expressed at different levels. Four genes recovered from venom duct transcripts of *C. lividus* (livi_24 and livi_26; livi_46 and livi_47) and two genes from *C. sanguinolentus* (sang_3 and sang_4) represent three sets of inparalogs that are expressed simultaneously (Figure 3.1), while no inparalogs were detected in venom duct cDNA of *C. diadema* and *C. quercinus*. Moreover, relative expression levels differ vastly between inparalogs that are expressed contemporaneously (Figure 3.2B).

Discussion

We investigated patterns of inter- and intra-specific variation in expression of A-superfamily conotoxin genes in venom ducts of four closely-related *Conus* species, and explored strategies of gene expression in each species. Results revealed a remarkable pattern of partial and differential expression of conotoxin genes, and expressed genes are either clustered phylogenetically in

certain species or more dispersed in other species. Our study demonstrates that plasticity of gene expression, combined with modification of toxin-coding gene sequences, has led to tremendous differences in venom composition among and within species.

1. Partial and differential expression of conotoxin genes among species

A-superfamily genes in the genome are partially expressed at low rates (less than 50%), implying that more than a half of conotoxin genomic composition does not contribute to the production of mature conotoxins. This phenomenon, in part, can be related to the functional fates of these genes. For example, as is observed for type I and II pseudogenes, genes that are unexpressed may be pseudogenized or in the process of pseudogenization, but this scenario is unlikely to be applicable to all genes because the majority of unexpressed genes appear to encode functional α-conotoxins. Alternatively, conotoxin genes may perform different roles during ontogeny such that some genes are up-regulated or exclusively expressed in the juvenile/subadult stage while others are highly expressed only in adults (see Chapter 5), and here we only captured expressed genes from adults.

Conotoxin genes are differentially regulated among species. There is little to no overlap in expressed genes among species, even between sister species that diverged very recently. Limited coexpression of orthologous genes among species has also been inferred for other *Conus* species (Duda and Palumbi 1999; Duda and Remigio 2008), implying that differential expression of conotoxin genes among species is a prevalent mechanism in generating venom diversity in *Conus*.

Conus species employ different strategies in exploiting their venome space. We developed an approach modified from community phyogenetic methods to examine the pattern of conotoxin gene expression in each species, a procedure that incorporates the phylogenetic relationships of gene members and complete information of genomic and transcriptomic profiles of these genes. Results revealed that *C. sanguinolentus* and *C. lividus* preferably express phylogenetically under-dispersed genes, while *C. diadema* and *C. quercinus* tend to more fully explore their venome space (i.e., exhibit over-dispersion) (Table 3.3). Phylogenetic distances of conotoxin genes possibly represent functional divergence of mature toxins if functional difference is positively correlated with genetic distances. From the gene duplication perspective, under-dispersion of gene expression suggests that genes originating from recent duplications are more likely expressed than paralogs that are distantly related. In this sense, the two sister species, *C. lividus* and *C. sanguinolentus*, tend to express genes emerged from relatively recent duplications and synthesize functionally similar conotoxins. Especially for *C. lividus*, many expressed genes belong to those clades that are composed of genes from multiple rounds of recent or lineage-specific duplications (Figure 3.1). Nonetheless, genes expressed by *C. diadema* and *C. quercinus* appear to have originated from more ancestral duplications and encode mature toxins serving different functions. Different patterns of conotoxin gene expression among species may also be affected by the number of genes expressed in each species. Conotoxin gene expression in *C. quercinus* and *C. diadema* are overdispersed and the absolute numbers of expressed genes are coincidently less than those of *C. lividus* and *C. sanguinolentus*, while the opposite pattern was observed for the latter two species (Table 3.1). These patterns are possibly explained by fundamental requirements of functional diversity of venom in each species: it is more important

for genes to be functionally diverse if only a few genes are expressed, while a large number of expressed genes allow more opportunities of fine-tuning subfunctions.

Interspecific differentiation of expression is driven by drift and/or selection (Whitehead and Crawford 2006), but the significantly non-random structures of gene expression in individual species (Table 3.3) and lack of coexpression of orthologous genes between species (Figure 3.2A) suggest that variation in conotoxin gene expression is unlikely accounted for solely by drift. Because conotoxin genes are known to be subject to strong positive selection (Duda and Palumbi 1999; Puillandre, Watkins, and Olivera 2010; Chang and Duda 2012), we posit that regulation of conotoxin genes is also affected by positive selection. Selection pressure likely stems from difference in dietary compositions among species, because conotoxins are primarily utilized for predation. Previous studies have demonstrated that allelic variation of conotoxin genes is positively correlated with dietary diversity (Duda et al. 2009) (also see Chapter 4), and suggested that extensity of gene turnover is possibly associated with dietary spectrum of each species (Chang and Duda 2012). *C. lividus* and *C. sanguinolentus* possess broader diets than the other two species (Chang and Duda 2012), which is possibly related to differences in numbers of expressed genes in venom ducts of these species. Interspecific divergence in prey types potentially shapes the development of expression strategies of each species.

Expressed genes are exposed to strong positive selection. Currently expressed genes possess a significantly larger ω than that of the unexpressed and ancestral genes, and this pattern still holds (both are larger than one) when we forced ancestral genes to share the same ω value as expressed genes (Table 3.4 and 3.5). This indicates that expression plays an important role in evolution of

conotoxin genes by differentially regulating exposure of genes to selection. The lower rates of ω for unexpressed genes imply that these genes are unexpressed permanently. Otherwise, selection may be highly variable through time (e.g. in the ontogenetic process) such that genes are switched off temporarily are subject to different levels/types of selection.

2. Intraspecific variation in conotoxin gene expression

Individuals of the same species also exhibit variation in conotoxin gene expression, which, combined with allelic divergence, could have led to the hypervariability of venom compositions within species (Jakubowski et al. 2005; Davis, Jones, and Lewis 2009; Rivera-Ortiz, Cano, and Mar í2011). Individuals of *C. lividus* and *C. quercinus* show significant differences in levels of expression of A-superfamily genes. Though such a pattern is not detected for *C. diadema*, more individuals need to be examined to rule out the possibility of differential expression within this species. Intraspecific variation in conotoxin gene expression may be temporal and affected by ecological factors such as variation in dietary specialization (a pattern of interaction that is confirmed in Chapter 5). Moreover, the inherent stochasticity of gene expression (Raser and O'Shea 2005) could affect venom composition among individuals even if they possess the same sets of genes. Gene duplication can also drive intra-specific variation in gene expression if gene copy numbers differ among individuals.

3. Expression divergence of paralogous genes

Because no significant differences in d_S or ω (d_N/d_S) were detected among categories of expression in three of the four *Conus* species, expression divergence of conotoxin genes did not show any association with divergence time or rates of evolution of these genes. As an exception,

the average d_S value of *C. lividus* is significantly smaller for genes that are differentially expressed than unexpressed genes. This implies that paralogous genes that are differentially expressed are relatively younger than pairs of paralogs that are expressed or unexpressed simultaneously. Expression divergence is also associated with heighted rates of evolution of these genes: average ω values of differentially expressed genes are significantly larger than those of unexpressed genes.

Previous studies present contradicting results concerning the association between expression divergence and sequence difference of coding regions (time): positive correlations were detected in model organisms such as yeast (Gu et al. 2002; Zhang, Gu, and Gu 2004) and human (Makova and Li 2003), but not in *Arabidopsis thaliana* (Haberer et al. 2004). Our results demonstrate that relationships between expression divergence and divergence time indeed differ among organisms, but revealed an opposite pattern of association for *C. lividus*, which is likely affected by the extensive duplication of this gene family in this species. Gene duplication heightens probabilities of expression divergence of paralogous genes (Li, Yang, and Gu 2005), but expression divergence and sequence distances are only coupled within a short timeframe after duplication (Gu et al. 2002; Makova and Li 2003; Oakley et al. 2005). Here we found that inparalogs (*C. lividus*, *C. sanguinolentus* and *C. diadema* who are very young species) are either not coexpressed, or coexpressed at different levels (Figure 3.1and 3.2B). Expression divergence is established for inparalogs and recent paralogs, supporting the notion proposed by Gu et al. (2002) that expression divergence can be rapidly fixed in recent gene duplicates. Differential expression contributes to the eventual retention of gene duplicates because unexpressed (or lowly expressed) redundant gene copies cannot be eliminated by purifying selection and, combined with positive

selection, potentially facilitates rapid evolution and neofunctionalization of these genes. Admittedly, our approach did not incorporate information of expression levels, and the arbitrary division of genes into categories of expression divergence may deteriorate the real pattern of evolution of conotoxin gene expression.

4. Implication for future studies of venom genes and other gene families

The high plasticity observed in expression of conotoxin genes among and within species elicits many implications for studies of venom evolution. First, examination of venom genes in the genome provides a more complete and accurate picture of the evolutionary history of these genes. Inferring evolutionary patterns based on information extracted from venom duct cDNA may severely underestimate the extensity of gene turnover. Second, venom genes that are not expressed should not be deemed as non-functional or negligible, because these genes may be temporally down-regulated, or differentially expressed among individuals. Admittedly, our investigation of expression of conotoxin genes via the approach of enrichment, cloning and sequencing of venom duct cDNA library is not high-throughput enough to completely identify all expressed genes especially the ones that are lowly expressed. The enrichment step with PCR amplification can further exaggerate differences in inferred levels of expression between highly and lowly expressed genes. Intraspecific variation in expression potentially impedes the unbiased examination of interspecific variation, if variation in expression is more extensive within species than among species.

Gene duplication accelerates expression divergence between species compared to single genes (Gu et al. 2004), a mechanism that involves both small-scale duplication and whole genome

duplication (Guan, Dunham, and Troyanskaya 2007; Ha, Kim, and Chen 2009). Genes and gene duplicates that are differentially regulated between species are affected by ecological adaptation (Gu et al. 2004; Whitehead and Crawford 2006), and conotoxin gene families are exemplar because of their essential roles in predator-prey interactions. Other ecologically-relevant gene families may exhibit expression divergence between species that are similar to the pattern determined for members of A-superfamily, and more studies are needed to examine this phenomenon at the scale of whole gene families.

Conclusion

We demonstrated partial and differential expression of venom genes among and within species, and confirmed the hypothesis that species differentially explore their venome space by preferential expression of phylogenetically similar or distant-related genes. Expressed genes are subject to strong positive selection pressure, and expression divergence of duplicate genes is established at an early stage. Extensive gene duplication and selection facilitate variation in gene expression and rapid evolution, combinations of which lead to interspecific divergence of venom composition. The inter-specific difference in expression of conotoxin gene families observed here are applicable to other multigene families, especially ones that are related to ecological adaptation. Our approach of determination of phylogenetic structure of expression can be widely utilized for investigation of patterns of gene family expression between species.

Acknowledgements

We thank JP Bingham at University of Hawaii for providing venom duct samples of *C. quercinus*. This study is supported by EEB Block Grant from Rackham Graduate School of University of Michigan to D. Chang and an NSF research grant ((IOS-0718379) awarded to T. F. Duda.

Figure 3.1. Phylogenies of A-superfamily conotoxin genes retrieved from genomic DNA and venom duct cDNA of four *Conus* species.
Bayesian consensus phylogeny of putative allele sequences of all genes in the genomic DNA of four species (termed 'genome phylogeny') was constructed with complete deletion and the HKY+I+G model (left). Bayesian consensus phylogeny of putative allele sequences expressed in venom ducts of four species (termed 'expression phylogeny') was constructed with complete deletion and the HKY+I model (right). Posterior probabilities are labeled at each node. Sequences that are expressed are shaded in yellow in both trees, and putative duplication events are labeled with red asterisks in the genome phylogeny.

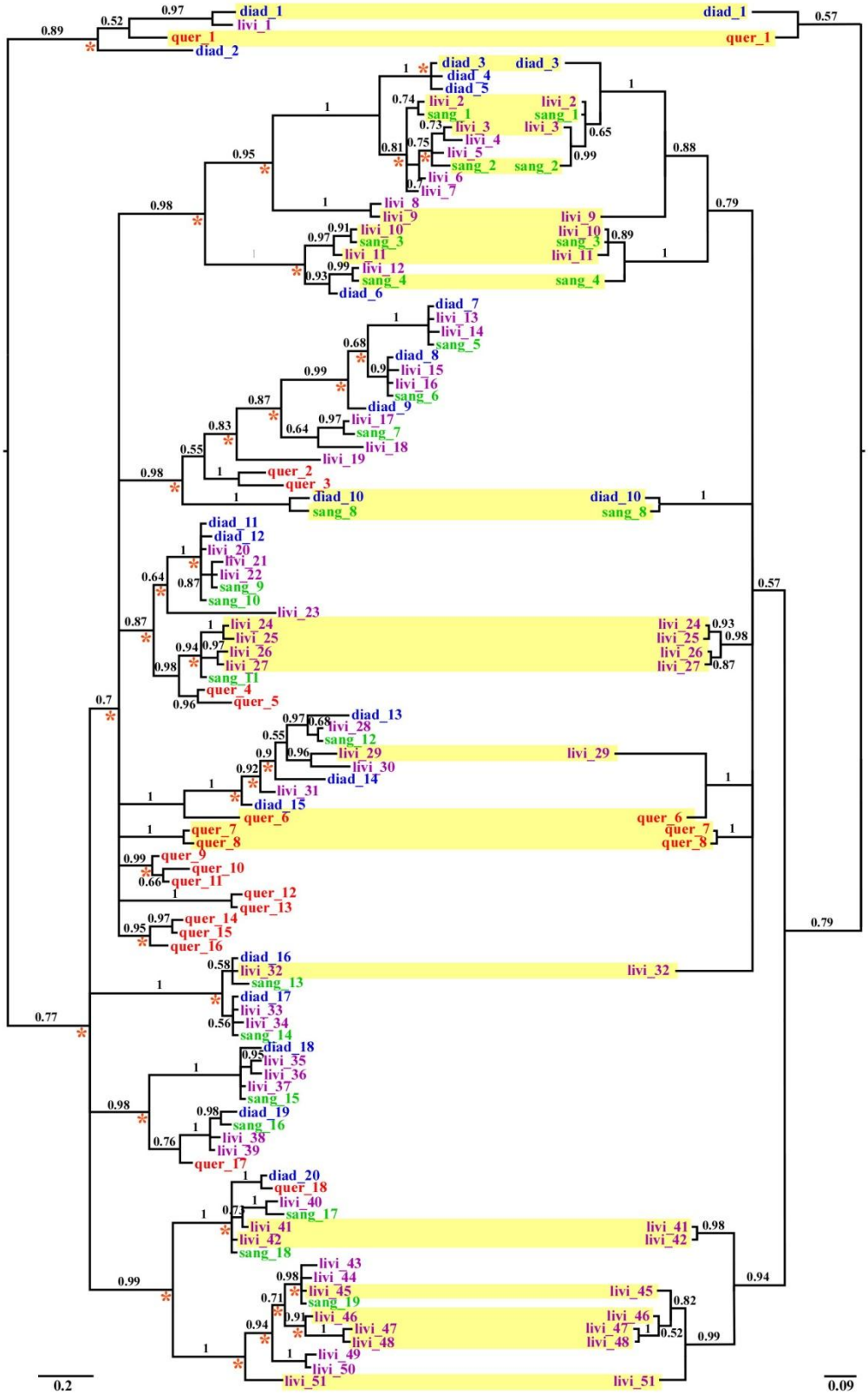

diad_1
livi_1
quer_1
diad_2
diad_3
diad_4
diad_5
livi_2
sang_1
livi_3
livi_4
livi_5
sang_2
livi_6
livi_7
livi_8
livi_9
livi_10
sang_3
livi_11
livi_12
sang_4
diad_6
diad_7
livi_13
livi_14
sang_5
diad_8
livi_15
livi_16
sang_6
diad_9
livi_17
sang_7
livi_18
livi_19
quer_2
quer_3
diad_10
sang_8
diad_11
diad_12
livi_20
livi_21
livi_22
sang_9
sang_10
livi_23
livi_24
livi_25
livi_26
livi_27
sang_11
quer_4
quer_5
diad_13
livi_28
sang_12
livi_29
livi_30
diad_14
livi_31
diad_15
quer_6
quer_7
quer_8
quer_9
quer_10
quer_11
quer_12
quer_13
quer_14
quer_15
quer_16
diad_16
livi_32
sang_13
diad_17
livi_33
livi_34
sang_14
diad_18
livi_35
livi_36
livi_37
sang_15
diad_19
sang_16
livi_38
livi_39
quer_17
diad_20
quer_18
livi_40
sang_17
livi_41
livi_42
sang_18
livi_43
livi_44
livi_45
sang_19
livi_46
livi_47
livi_48
livi_49
livi_50
livi_51
0.2
0.09

Figure 3.2. Relative expression levels of coexpressed orthologs between species, inparalogs and individuals with species.
Names of loci recovered from *C. lividus* start with L, *C. diadema* with D, *C. quercinus* with Q and *C. sanguinolentus* with S and the numbers after the letters correspond to the numbers in the sequence labels in Figure 3.1.

(A) Relative expression of orthologous conotoxin loci coexpressed by two *Conus* species. Dark and light grey bars represent orthologous loci expressed by both species.

(B) Relative expression levels of coexpressed inparalogs that are generated by lineage-specific gene duplications. Dark and light grey bars represent the paralogous genes expressed within each species.

(C) Variation in expression of conotoxin genes between two individuals of single species of *C. lividus*, *C. diadema* and *C. quercinus*. The two individuals are represented by dark grey and white bars.

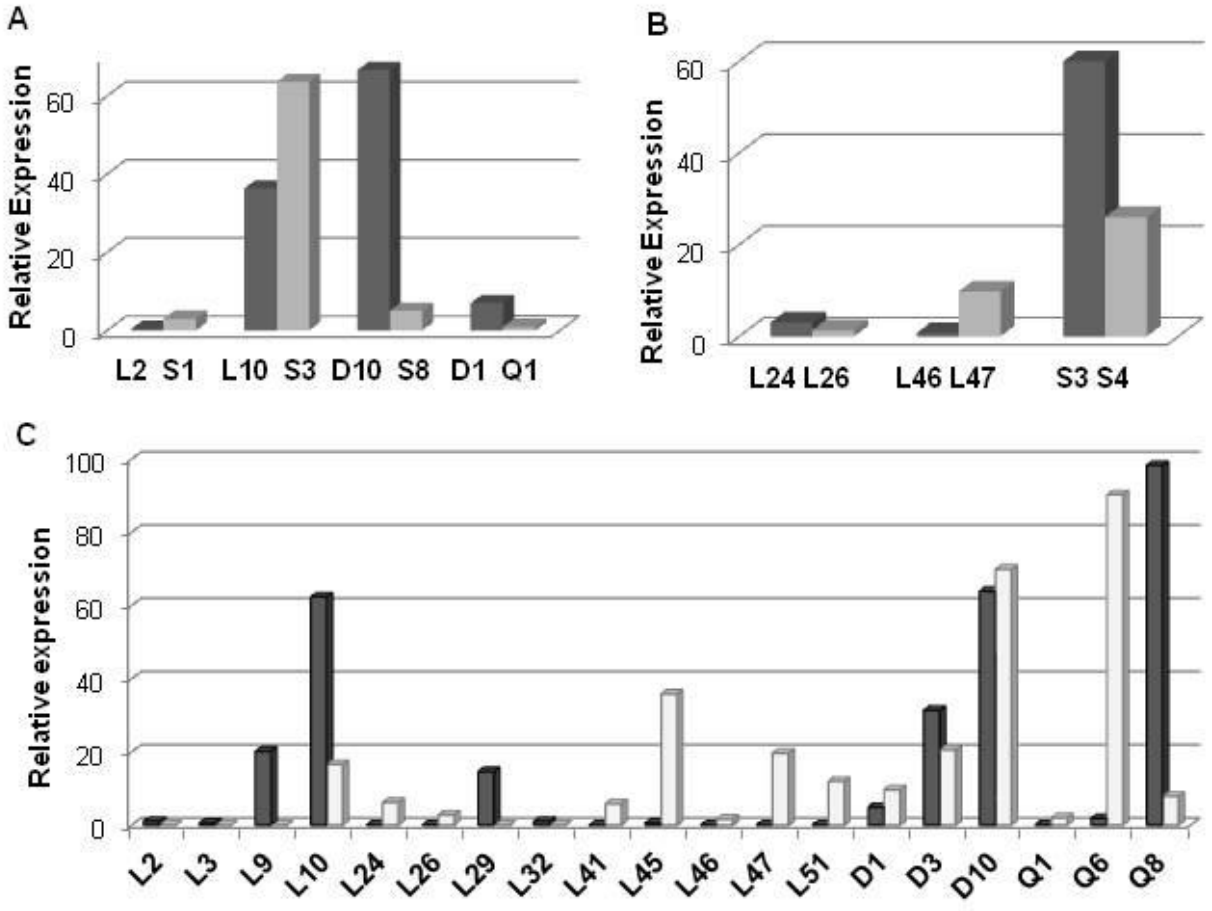


Table 3.1. Expressed A-superfamily conotoxin recovery information.
Numbers of colonies screened and sequenced, putative A-superfamily gene sequences, unique sequences, non-artefactual alleles and loci recovered from venom duct transcripts of *Conus* species.

	C. lividus	***C. diadema***	***C. quercinus***	***C. sanguinolentus***
Colonies sequenced	487	167	135	112
A-superfamily sequences	459	156	107	100
Unique sequences	66	26	17	26
Alleles	18	3	4	5
Loci	13	3	3	5
Percentage (species)	40.62	16.70	25.00	27.78

Table 3.2. Numbers of orthologous coexpressed loci among species (below diagonal) and their proportions in the venom duct expression profiles of each species (above diagonal). The number before the forward slash in each cell is the percentage (%) of coexpressed loci in the species of the row label of the cell, and the number after the slash is the percentage in the species of the respective column label.

	C. lividus	***C. diadema***	***C. quercinus***	***C. sanguinolentus***
C. lividus	-	0 / 0	0 / 0	15.4 / 40.0
C. diadema	0	-	33.3 / 33.3	33.3 / 20.0
C. quercinus	0	1	-	0 / 0
C. sanguinolentus	2	1	0	-

Table 3.3. Community phylogenetic indices as evaluations of phylogenetic stucture of expressed genes.
The mean phylogenetic distance (MPD), the net relatedness index (NRI), the mean nearest phylogenetic taxon index (MNTD) and the nearest taxon index (NTI) were estimated for each species. 10,000 generations of simulations of random sampling of the phylogenetic tree of each species were performed and *P*-values were determined by percentages of the random samples smaller or larger than observation.

Species	**MPD**	**MPD random**	**NRI**	**MNTD**	**MNTD random**	**NTI**
C. lividus	0.597 $^{P=0.466}$	0.599	0.058	0.193 $^{P=0.337}$	0.213	0.431
C. sanguinolentus	0.503 $^{P=0.056}$	0.612	1.678	0.186 $^{P=0.023}$	0.415	2.231
C. diadema	0.850 $^{P=0.016}$	0.572	-2.281	0.758 $^{P=0.021}$	0.487	-1.902
C. quercinus	0.506 $^{P=0.098}$	0.352	-1.361	0.410 $^{P=0.120}$	0.289	-1.168

Table 3.4. Models used to test if presently expressed genes exhibit heightened ω (d_N/d_S) values than the rest of the genes, and results of the tests.
Gene list describes the two sets of genes used in these tests (with or without type III pseudogenes). Ln(*L*) stands for log-likelihood of each model. *P*-values were estimated by Likelihood Ratio Tests. Definitions of ω variables in each model are described in the methods section.

Gene List	Model	ω	Ln(*L*)	*P-value*
No type I, II pseudo, livi_51	Null: One rate	ω= 1.731	-1140.559	-
	Alternative: Two rates	ω_1=1.503, ω_2= 7.953	-1138.528	0.044
	Alternative: Two rates reversed	ω1=1.836, ω2=1.582	-1139.822	0.225
	Alternative: Three rates	ω1=2.406, ω2=1.765, ω3=7.967	-1138.408	0.116
No pseudo, livi_51	Null: One rate	ω= 1.645	-1091.238	-
	Alternative: Two rates	ω_1= 1.418, ω_2= 7.813	-1089.095	0.038
	Alternative: Two rates reversed	ω1=1.642 , ω2=1.653	-1091.238	1.000
	Alternative: Three rates	ω1=1.332, ω2=1.651, ω3=7.824	-1088.990	0.106

Table 3.5. Results of maximum-likelihood estimations of ω values for A-superfamily genes in each *Conus* species.
Gene sets with and without type III pseudogenes were both used for *C. lividus* and *C. sanguinolentus*. The null model was one ω rate across the whole phylogeny while the alternative model is one ω for extantly expressed genes while another one for non-expressed genes. *P*-values were determined by Likelihood Ratio Tests of log-likelihoods of null and alternative models with one degree of freedom.

Gene Set	Model	ω	Ln(L)	*P*-value
C. diadema	Null: One rate	ω= 0.679	-572.480	0.073
	Alternative: Two rates	ω_1=0.624, ω_2= 999	-570.874	
C. lividus, no livi_51	Null: One rate	ω= 1.429	-730.687	0.008
	Alternative: Two rates	ω_1= 1.181, ω_2= 999	-727.245	
C. lividus, no livi_51, no pseudo	Null: One rate	ω= 1.285	-682.324	0.008
	Alternative: Two rates	ω_1= 1.069, ω_2= 999	-678.802	
C. quercinus, no pseudo	Null: One rate	ω= 1.379	-300.619	0.428
	Alternative: Two rates	ω_1= 1.790, ω_2= 0.939	-300.305	
C. sanguinolentus	Null: One rate	ω= 1.082	-685.390	0.041
	Alternative: Two rates	ω_1= 0.981, ω_2= 999	-683.310	
C. sanguinolentus, no pseudo	Null: One rate	ω= 0.970	-642.142	0.032
	Alternative: Two rates	ω_1= 0.862, ω_2= 999	-639.839	

References

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

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.

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

Dong D, Yuan Z, Zhang Z. 2011. Evidences for increased expression variation of duplicate genes in budding yeast: From cis- to trans-regulation effects. *Nucleic Acid Res.* 39:837-847.

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, Palumbi SR. 1999. Molecular genetics of ecological diversification: Duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc Natl Acad Sci USA.* 96:6820-6823.

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.* 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.

Duda TF, Terbio M, Chen G, Phillips S, Olenzek AM, Chang D, Morris DW. 2012. Patterns of population structure and historical demography of *Conus* species in the tropical pacific. *Am Malacol Bull.* 30:175-187.

Emerson BC, Gillespie RG. 2008. Phylogenetic analysis of community assembly and structure over space and time. *Trends Ecol Evol.* 23:619-630.

Enard W, Khaitovich P, Klose J et al. 2002. Intra- and interspecific variation in primate gene expression patterns. *Science.* 296:340-343.

Fry BG. 2005. From genome to "Venome": Molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. *Genome Res.* 15:403-420.

Gu Z, Nicolae D, Lu HHS, Li W-H. 2002. Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet.* 18:609-613.

Gu Z, Rifkin SA, White KP, Li W-H. 2004. Duplicate genes increase gene expression diversity within and between species. *Nat Genet.* 36:577-579.

Guan Y, Dunham MJ, Troyanskaya OG. 2007. Functional analysis of gene duplications in *Saccharomyces cerevisiae*. *Genetics.* 175:933-943.

Ha M, Kim E-D, Chen ZJ. 2009. Duplicate genes increase expression diversity in closely related species and allopolyploids. *Proc Natl Acad Sci USA.* 106:2295-2300.

Haberer G, Hindemitt T, Meyers BC, Mayer KFX. 2004. Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of *Arabidopsis*. *Plant Physiol.* 136:3009-3022.

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.

Hu H, Bandyopadhyay P, Olivera B, Yandell M. 2012. Elucidation of the molecular envenomation strategy of the cone snail *Conus geographus* through transcriptome sequencing of its venom duct. *BMC Genomics.* 13:284.

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

Jackson S, Hass Jacobus B, Pagel J. 2004. The gene space of the soybean genome. *in* Wilson R, Stalker H, and Brummer E, eds. Legume crop genomics. Champaign: AOCS Press, p. 187-193.

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.

Jovelin R, He X, Amores A, Yan Y-l, Shi R, Qin B, Roe B, Cresko WA, Postlethwait JH. 2007. Duplication and divergence of Fgf8 functions in teleost development and evolution. *J Exp Zool B.* 308B:730-743.

Kaas Q, Yu R, Jin A-H, Dutertre S, Craik DJ. 2012. Conoserver: Updated content, knowledge, and discovery tools in the conopeptide database. *Nucl Acids Res.* 40:D325-D330.

Kawaura K, Mochida K, Ogihara Y. 2005. Expression profile of two storage-protein gene families in hexaploid wheat revealed by large-scale analysis of expressed sequence tags. *Plant Physiol.* 139:1870-1880.

Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Pääbo S. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science.* 309:1850-1854.

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.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science.* 188:107-116.

Koonin EV. 2005. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet.* 39:309-338.

Li J, Yuan Z, Zhang Z. 2010. Revisiting the contribution of cis-elements to expression divergence between duplicated genes: The role of chromatin structure. *Mol Biol Evol.* 27:1461-1466.

Li W-H, Yang J, Gu X. 2005. Expression divergence between duplicate genes. *Trends in genetics : TIG.* 21:602-607.

Lluisma AO, Milash BA, Moore B, 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.

Lockhart DJ, Winzeler EA. 2000. Genomics, gene expression and DNA arrays. *Nature.* 405:827-836.

Makova KD, Li W-H. 2003. Divergence in the spatial pattern of gene expression between human duplicate genes. *Genome Res.* 13:1638-1645.

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* 3:418-426.

Oakley TH, Gu Z, Abouheif E, Patel NH, Li W-H. 2005. Comparative methods for the analysis of gene-expression evolution: An example using yeast functional genomic data. Pp. 40-50.

Ohno S. 1970. Evolution by gene duplication. Berlin:Springer-Verlag.

Olivera BM. 2002. *Conus* venom peptides: Reflections from the biology of clades and species. *Ann Rev Ecol Syst.* 33:25-47.

Posada D. 2008. jModelTest: Phylogenetic model averaging. *Mol Biol Evol.* 25:1253-1256.

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.

Qian W, Liao B-Y, Chang AY-F, Zhang J. 2010. Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends Genet.* 26:425-430.

Rambaut A. 2002. Se-Al sequence alignment editor. Version 2.0.A11. Oxford: University of Oxford.

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science.* 300:1742-1745.

Raser JM, O'Shea EK. 2005. Noise in gene expression: Origins, consequences, and control. *Science.* 309:2010-2013.

Rivera-Ortiz JA, Cano H, Mar íF. 2011. Intraspecies variability and conopeptide profiling of the injected venom of *Conus ermineus*. *Peptides.* 32:306-316.

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.

Somel M, Creely H, Franz H, Mueller U, Lachmann M, Khaitovich P, P ääbo S. 2008. Human and chimpanzee gene expression differences replicated in mice fed different diets. *PLoS ONE.* 3:e1504.

Swofford. 2002. PAUP*: Phylogenetic analysis using parsimony (and other methods) 4.0b10. *Sunderland, Massachusetts: Sinauer Associates.*

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Http://www.R-project.org/.

Tomanek L, Somero GN. 2002. Interspecific- and acclimation-induced variation in levels of heat-shock proteins 70 (hsp70) and 90 (hsp90) and heat-shock transcription factor-1 (HSF1) in congeneric marine snails (genus *Tegula*): Implications for regulation of hsp gene expression. *J Exp Biol.* 205:677-685.

Webb CO, Ackerly DD, Kembel SW. 2008. Phylocom: Software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics.* 24:2098-2100.

Webb CO, Ackerly DD, McPeek MA, Donoghue MJ. 2002. Phylogenies and community ecology. *Annu Rev Ecol Syst* 33:475-505.

Whitehead A, Crawford DL. 2006. Variation within and among species in gene expression: Raw material for evolution. *Mol Ecol.* 15:1197-1211.

Yang Z. 2007. Paml 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586-1591.

Zhang Z, Gu J, Gu X. 2004. How much expression divergence after yeast gene duplication could be explained by regulatory motif evolution? *Trends Genet.* 20:403-407.

CHAPTER 4 GEOGRAPHIC HETEROGENEITY IN BIOTIC INTERACTIONS DRIVES THE EVOLUTION OF PREDATORY GENES

This chapter will be submitted for publication with coauthors Amy O. Olenzek and Thomas F. Duda, Jr.

Introduction

Biotic interactions shape the evolutionary trajectories of species (Van Valen 1973; Paterson et al. 2010). Selection from geographic heterogeneity in the composition and strength of species interactions drives divergence of traits at the interface of these interactions (Thompson 2005). Such patterns of divergence may reflect coevolutionary responses, as suggested for defenses of fruit flies to parasitoid wasps (Kraaijeveld and Godfray 1999) and resistance of garter snakes to toxic newts (Brodie, Ridenhour, and Brodie 2002). Variation in feeding traits, as illustrated by beaks of Darwin's finches (Schluter and Grant 1984), gill rakers of alewives (Palkovacs and Post 2008) and sticklebacks (Gross and Anderson 1984), radular teeth and drilling behaviors of marine snails (Andrade and Solferini 2006; Sanford and Worth 2010), and venoms of snakes (Daltry, Wuster, and Thorpe 1996; Mackessy et al. 2006; Gibbs et al. 2011), are associated with characteristics of feeding resources. Such associations may be genetically based (Smith 1993; Palkovacs and Post 2008; Sanford and Worth 2010), but few studies have examined the genetic determinants due to limited knowledge of genes responsible (Thompson 2005).

Conotoxin peptides of *Conus* are expressed by members of many gene superfamilies (A, D, I, J, L, M, O, P, S, T, V, Y) and target ion channels and neuronal receptors (Kaas, Westermann, and Craik 2010). Conotoxin genes exhibit remarkably rapid rates of evolution and are subject to extensive gene turnover and selection (Chang and Duda 2012). The wealth of knowledge of venoms and ecology of *Conus* enables us to directly target genes that operate at the molecular interface of predator-prey interactions. *Conus ebraeus*, a vermivorous species that is widely distributed in the Indo-West Pacific (IWP), shows no evidence of population structure in this region based on analyses of mitochondrial gene COI sequences (all pairwise Φ_{ST} values are less than 0.05 and non-significant) (Duda and Lessios 2009; Duda et al. 2012). Nonetheless, significant differences in allelic frequencies among populations at a single conotoxin locus imply that this gene has been affected by selection (Duda et al. 2009). Is this pattern of variation apparent at other conotoxin loci? Do conotoxin genes differ in their patterns of variation, suggesting different roles for their products in species interactions? Most importantly, are patterns of variation in venom composition driven by geographic heterogeneity in prey utilization? We posit that geographic differences in prey utilization drive the evolution of conotoxins and the diversity of conotoxin genes is positively associated with prey diversity.

Materials and Methods

1. Specimens and fecal samples

Specimens of *Conus ebraeus* were collected from Guam in 2008, Hawaii in 2009 and American Samoa in 2009. Specimens were deposited in the Mollusk Division collections at the University of Michigan Museum of Zoology. Body tissues were preserved in 95% ethanol and venom ducts

were preserved in RNAlater (Ambion, Inc.) and stored at -20 or -80 ℃. Fecal samples were collected and preserved following the approach of Duda et al. (2009).

2. cDNA preparation and characterization of members of each gene family

We extracted mRNA from venom ducts of 31 individuals of *C. ebraeus* from Hawaii, 39 individuals from Guam and 15 individuals from American Samoa. We synthesized cDNA following the procedure described by Duda and Palumbi (1999). We utilized general primers designed in conserved regions of A (one set of primers that should theoretically amplify all components of this gene family), I (one set of primers), M (two sets of primers MPr1 and MPr2 for divergent classes of genes from this superfamily) and O-superfamilies (one set of primers for divergent classes of genes from this superfamily that putatively encode δ-conotoxins) (Table 4.1) to amplify members of these gene families from venom duct transcripts of one to five individuals at each location. We ligated PCR products into vectors and transformed these into competent cells using The Original TA Cloning Kit with Top 10 Competent Cells (Invitrogen). We screened colonies and sequenced amplification products of expected target sizes at the University of Michigan DNA Sequencing Core. We examined sequence chromatograms in Sequencher version 4.8 (Gene Codes Corporation) and manually aligned sequences with Se-Al v2.0a11 (Rambaut 2002) based on similarity of nucleotide and predicted amino acid sequences and consistency in the structure of the cysteine backbone of each superfamily. We constructed maximum-likelihood phylogenies of sequences (including suspected artefactual sequences representing polymerase or cloning errors) belonging to each conotoxin gene family with MEGA 5.05 (Tamura et al. 2011), using the best model selected by jModelTest 0.1.1 (Posada 2008), complete deletion of gaps (12-34 nucleotides out of a total sequence length of 180-278

nucleotides) in the alignments, NNI (Nearest-Neighbor Interchange) branch-swapping approach and bootstrap analyses of 100 replicates. Putative loci were determined from the major clades, based on phylogenetic relationships of all recovered sequences and the criterion that the average within-clade distances are much smaller than distances among clades.

3. Individual genotyping

We designed locus-specific primers for each putative locus that exhibited allelic variation and genotyped individuals from all three locations through amplifications with locus-specific primers (Table 4.2) and direct sequencing of products. We determined genotypes of each individual at each locus by examining resultant chromatograms. Allelic sequences were identified from sequences recovered earlier via cloning or from chromatograms of putative homozygous individuals that contained no double peaks. Sequences of new alleles that were not recovered through cloning or from homozygotes were determined by subtracting peaks of known alleles from chromatograms with double peaks. For certain alleles that could not be distinguished in these manners (i.e., identity of alleles contributing to double peaks in chromatograms could not be confirmed), we designed additional allele-specific primers and utilized the same genotyping approach described above to identify them (Table 4.3).

Data for locus E1 include individuals at Hawaii and Guam previously reported by Duda et al. (2009), individuals at all three locations collected for this study, and additional individuals collected at American Samoa in 2000 with the same approach as described by Duda et al. (2009). These data were pooled together because no temporal shifts of allelic compositions were detected at locus E1.

4. Population analyses of single locus

We aligned alleles of each locus with Sequencher (version 4.8) using the contig assembly tool and assembly parameters set to 95%. Allelic divergence and patterns of variation among locations were examined and visualized in the form of statistical parsimony networks with TCS 1.21 (Clement, Posada, and Crandall 2000). The 3' untranslated regions (including the stop codon) of sequences of each locus were removed for population analyses. We calculated molecular diversity and gene diversity indices with Arlequin version 3.1 (Excoffier, Laval, and Schneider 2005) using the best substitution model (i.e., the Tamura-Nei model (Tamura and Nei 1993)) selected for each locus as determined with jModelTest 0.1.1 (Posada 2008). The first polymorphic site of the common alleles of locus ED4 was not sequenced from rare alleles due to the location where one of the locus specific primers was designed. To take into account the first polymorphic site that differentiates alleles ED5 and ED40, we included all sites with less than 50% missing information for population analyses of locus ED4. We set the missing level to 0.05 for the other four loci to exclude sites that contain more than 5% missing information among sequences. To verify the validity of our assumption that we were characterizing alleles of single loci, we performed exact tests of Hardy-Weinberg equilibrium with allelic compositions of each locus in each population by 100,000 Markov chain steps in Arlequin version 3.1(Excoffier, Laval, and Schneider 2005); significance cutoff was determined after correction for multiple tests (Lessios 1992). Population divergence was examined in Arlequin version 3.1 with pairwise F-statistics. Significance of results was evaluated by 10,100 random permutations from the pooled dataset of all three populations. We performed hierarchical AMOVA (Excoffier, Smouse, and Quattro 1992) for each locus with all three possible hierarchical groupings (Table 4.6) and

compared levels of genetic variance among groups and within groups across the three options. We tested for the neutrality of each locus at each location by estimating Tajima's *D* (Tajima 1989) and Fu's F_S (Fu 1997) values in Arlequin version 3.1; significance were determined by the percentage of values estimated from 10,000 simulations that are less than or equal to the observed values. Fu and Li's *D** and *F** (Fu and Li 1993) were computed in DNASP v5 (Librado and Rozas 2009) with complete deletion of gaps in the aligned gene sequences; significance was evaluated with empirical distributions.

5. Multi-locus population data analyses

We utilized 30 individuals from Hawaii, 29 from Guam and 14 from American Samoa that were genotyped at four loci (ED4, ED6, ED20 and EA4) for multivariate data analyses. Information from locus E1 was not utilized because the individuals that were genotyped for this locus are not the same as those genotyped for the other loci. Tests of linkage disequilibrium were performed on each pair of the four loci in each population with GENEPOP (Raymond and Rousset 1995; Rousset 2008); significance was determined with likelihood ratio tests. Genotypic data of the four loci (including missing data) from each individual were pooled for clustering analyses with Structure 2.3.3 (Hubisz et al. 2009). We utilized an admixture model and correlated allelic frequencies model with default priors. We ran the MCMC analyses for 100,000 steps for K=2, 3 and 4 (K= number of clusters), removed the first 10,000 results as burnin and examined convergence of F_{ST} values and α. We compared estimated (log probability of the data, *ln*(*Prob*(data))) across analyses of different K to determine the most likely clustering pattern of individuals based on these loci.

6. Identification of prey and estimation of dietary diversity

We identified prey species from fecal samples of *C. ebraeus* individuals with the DNA barcoding approach as described by Duda et al. (Duda et al. 2009). We aligned 16S gene sequences recovered from fecal samples and polychaete sequences downloaded from GenBank (accession numbers labeled in the names of sequences in Figure 4.9) and performed model selection and phylogenetic analyses of these sequences. We constructed a maximum-likelihood phylogeny with the NNI branch-swapping approach and the best model selected by jModelTest v0.1.1 with complete deletion of missing data in MEGA 5.05 (Tamura et al. 2011). Because the putative prey species were members of two taxonomic groups within Polychaeta (order Eunicida and family Nereididae of order Phyllodocida), we separated these 16S gene sequences into two datasets composed exclusively of sequences of putative Eunicida species and Nereididae species. Bayesian consensus phylogenies were constructed in MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001) with these sequences (5,000,000 generations, two runs, four chains and 25% burnin) and the best models selected for each dataset in jModelTest v0.1.1. Prey species were determined based on the clustering patterns of fecal sequences with sequences of polychaetes from GenBank.

We used Shannon-Wiener's index (Shannon 1948) (*H'*) and mean genetic distances to quantify dietary diversity at each location. We estimated proportional similarity indices (PS_I (Whittaker 1952), Pianka's overlap indices (Pianka 1974), and a measure of phylogenetic disparity of prey items D_{ST} that is analogous to measures of Φ_{ST}) to quantify the extent of geographic differentiation in diet. Mean genetic distances were estimated with the K80 model (Kimura 1980) in MEGA5.05 and D_{ST} values were estimated by F-statistics in Arlequin version 3.1 with the

Tamura-Nei distance model. We evaluated significance of PS_I and Pianka's overlap index values through a Monte Carlo simulation approach that randomizes prey items recovered for paired samples based on pooled frequencies of prey from these samples and calculates PS_I and Pianka's overlap index values for the random samples. The analysis compares observed PS_I and Pianka's overlap index values to the distribution of these values calculated from 10,000 simulated datasets constructed using the same sample sizes as the original data sets and assuming a null hypothesis that the samples are not independent. *P*-values were determined from the number of values that are less than or equal to the values observed for the original data.

7. Test of association between variation of venom genes and dietary heterogeneity

To test whether the gradient of diversity of conotoxin genes among locations can be explained by dietary variables, we employed canonical correspondence analysis (CCA) (Ter Braak 1986). CCA is a multivariate statistical tool for exploration of correlative patterns of a set of variables (Ter Braak 1986) (i.e., inter-population diversities of conotoxin genes and prey in this study). We constructed two contingency tables with five conotoxin genes as column variables, three locations as row variables and the gene/nucleotide diversities as inputs of each cell; we built another contingency table with three locations as row variables and Shannon-Wiener's indices (*H*') and mean genetic distances of prey items as column variables. Canonical correspondence analyses of diversities of conotoxin genes with diversities of local prey items were performed with the *cca* function in the package *vegan* (Oksanen et al. 2006) in R v2.15.0 (R Development Core Team 2012). We estimated proportions of the total eigenvalues explained by each dimension and constructed biplots.

Every CCA ordination/biplot contains four axes, with two showing relative gradients of dependent variables (i.e. gene/nucleotide diversities of conotoxin genes) and the other two showing percentages of variation explained by the independent/explanatory variables (i.e. *H'* and genetic distance of prey items). The relative positions and distances among populations represent their similarities in the gradient of dependent variables (diversities of conotoxin genes). Vectors of the two explanatory variables (*H'* and genetic distance of prey items) point to their higher values; angles of the vectors convey the relative correlations between the dietary variables; and lengths of vectors represent the proportion of covariance of diversities of conotoxin genes explained by dietary variables. If vectors of two dietary variables point to the incremental gradient of diversities of conotoxin genes among the populations, gene/nucleotide diversities of conotoxin genes are positively correlated with the dietary diversities.

As a control, we estimated the nucleotide diversity of mitochondrial COI gene sequences of populations of *C. ebraeus* at Guam, American Samoa and Hawaii presented by Duda and Lessios (Duda and Lessios 2009) with the Tamura-Nei model with Arlequin version 3.1. We also obtained estimates of haplotype diversities of the COI gene from Duda and Lessios (Duda and Lessios 2009). We estimated coefficients of simple linear regressions of haplotype and nucleotide diversity of the COI gene with *H'* and genetic distances of diets at each location in R v2.15.0; we compared these values with values of the same coefficients estimated for the five conotoxin genes.

Canonical correspondence analyses were also performed with a contingency table of pairwise Φ_{ST} values of five conotoxin genes among locations (American Samoa-Guam, American Samoa-

Hawaii and Guam-Hawaii) as dependent variables and a contingency table of pairwise PS_I and D_{ST} of prey compositions among locations as independent/explanatory variables. Negative Φ_{ST} values were converted to zero. The interpretation of the CCA ordination/biplot is described previously. Vectors of pairwise PS_I and D_{ST} point to their higher values; and lengths of vectors represent the proportion of covariance of pairwise Φ_{ST} explained by these two dietary variables. If the dietary vector points to the incremental gradient of pairwise Φ_{ST} among locations, pairwise divergence of conotoxin genes is positively related with prey differentiation. To support the CCA results, correlation coefficients of pairwise Φ_{ST} values of the highly polymorphic conotoxin genes and pairwise PS_I, Pianka's overlap index and D_{ST} of prey species among locations were computed with Pearson (parametric) (Rodgers and Nicewander 1988), Spearman (Spearman 1910) and Kendall (Kendall 1948) methods (non-parametric) in R v2.15.0. Correlation coefficients of conotoxin loci ED20, EA4 and the mitochondrial COI gene were not estimated, because their Φ_{ST} values are essentially zero for every pairwise comparison. R scripts for the statistical analyses used in this study are available upon request.

Results and Discussion

1. Geographic variation of conotoxin genes and modes of selection

To identify conotoxin genes expressed by *C.* ebraeus, we investigated the diversity of multiple conotoxin superfamilies in venom duct cDNA of a few individuals collected at Hawaii, Guam and American Samoa. We recovered 30 unique sequences (GenBank Accession numbers JX177103 - JX177132) out of 144 colonies sequenced representing three putative A-superfamily loci, 45 sequences (which potentially encode encode δ -conotoxins, GenBank accession numbers JX177236 - JX177277) out of 146 colonies sequenced representing four O-superfamily loci, 22

unique sequences (GenBank accession numbers JX177133 - JX177161) out of 131 colonies sequenced representing two I-superfamily loci, and 74 sequences (GenBank accession numbers JX177162 - JX177235) out of 223 colonies sequenced representing at least seven M-superfamily loci (Figure 4.1).

Among all putative conotoxin loci identified, we successfully determined genotypes of individuals from Guam, American Samoa and Hawaii at five conotoxin loci: four O-superfamily loci (ED4, ED6, ED20 and E1) and one A-superfamily locus (EA4) (Table 4.4, Figure 4.1; GenBank accession numbers FJ804530-FJ804536, FJ834437 and JX177278- JX177299). Genotypes of certain loci in some individuals could not be obtained with our amplification approach, a result that we interpret to have resulted from lack of expression of these genes in some individuals (e.g., see (Duda and Lee 2009)). In addition, chromatograms of all individuals were interpreted to contain at most two unique sequences (i.e., no more than two alleles were detected from single individuals). These loci show no evidence of deviation from Hardy-Weinberg equilibrium or linkage disequilibrium.

Loci ED4, ED6 and E1 ('highly polymorphic loci') in general exhibited much higher levels of nucleotide and gene diversity than loci ED20 and EA4 ('conservative loci') (Table 4.4). Locus ED4 includes nine alleles in which four alleles (40, 5, 4 and 9) were identical to individual sequences obtained through cloning, while the other five alleles (a1, a2, a4, a5 and a6) were inferred from chromatograms. Substitutions at the upstream seven sites are associated with seven amino acid replacements in the putative mature toxins (Figure 4.2). Alleles 5 and 40 may encode the same mature conotoxin because the only site that differentiates the two alleles represents an

amino acid replacement at the putative toxin cleavage site. Locus ED6 possesses seven putative alleles. All nine segregating sites reside in the toxin coding region; substitutions at these sites lead to six amino acid changes in the mature conotoxins (Figure 4.2). Nine putative alleles for locus E1 possess 13 polymorphic sites upstream of the stop codon that give rise to 10 amino acid replacements (nine in the mature toxin region) (Figure 4.2). Locus ED20 contains only three alleles: 20, A1 and A2, among which only allele 20 was identified previously through cloning. Substitutions of the two segregating sites of this locus are nonsynonymous and result in two amino acid replacements in the mature toxin (Figure 4.2). Allele A2 is a putative null allele based on the presence of a premature stop codon at the fourth *Cys* codon position in the toxin coding region.

Analyses of the five conotoxin genes revealed significant geographic differentiation and strongly contrasting patterns of variation among loci and locations (Table 4.4 and 4.5; Figure 4.3). Three highly polymorphic loci possessed fewer alleles and lower gene and nucleotide diversities at Hawaii, while levels of diversity at Guam and American Samoa were equivalent (Table 4.4). Allelic frequencies of the highly polymorphic loci also differ among locations (Figure 4.3A-C). Based on pairwise Φ_{ST} values, the population at Hawaii is genetically differentiated at these loci, while the other populations show no divergence from each other (Table 4.5). Results from hierarchical Analysis of Molecular Variance (AMOVA) (Excoffier, Smouse, and Quattro 1992) support this interpretation (Table 4.6); the pattern is also robust when genes are analyzed jointly (Figure 4.3F).

The distinct allelic composition and frequencies of the highly polymorphic loci at Hawaii result from selection rather than recent population expansion because these loci show contrasting patterns of variation as revealed from neutrality tests (Table 4.4). Moreover, alleles of each locus exhibit an overwhelming prevalence of nonsynonymous substitutions in the toxin coding region (Figure 4.2). Modes of selection also differ considerably among locations and loci: purifying selection is most prevalent at Hawaii (except locus E1) while diversifying selection occurs predominantly at other locations (Table 4.4).

Alternatively, the low levels of diversity and absence of structure at the conservative loci (Table 4.4 and 4.5), however, may reflect the historical demography of these populations, selective sweeps and/or recent gene duplication events that gave rise to these loci. Nonetheless, the lack of variation of the conservative loci contrasts with the high levels of diversity at COI (Duda and Lessios 2009) and thus is unlikely to have resulted solely from demographic processes. Locus ED20 may have experienced a recent selective sweep based on the consistently negative values estimated from neutrality tests, but EA4 appears to be neutral (Table 4.4).

2. Geographic variation of dietary specialization

Populations of *C. ebraeus* exhibit substantial differences in dietary specializations and an overwhelming degree of geographic heterogeneity in their interactions with prey (Table 4.5 and 4.7; GenBank accession numbers JX177300-JX177352, FJ804537-FJ804572 and FJ907334-FJ907342). Our phylogenetic approach revealed a total of 11 putative prey species from the annelid Order Eunicida and Family Nereididae (Order Phyllodocida) (Figure 4.4). Putative *Palola* species (Order Eunicida) were determined based on the individual clades in the species

tree and classifications proposed by Schulze (2006) (Figure 4.4). Other sequences that have not been reported from identified polychaetes were arbitrarily treated as species based on their clustering patterns in the phylogeny and were assigned new names (e.g. *Palola* AX1, AX2, AX3) (Figure 4.4). The majority of prey items represent *Palola* species and very few are Nereids (Figure 4.4; Table 4.7). The population at Hawaii possesses the most distinct diet with the lowest diversity and most uneven composition of prey species (Table 4.5 and 4.7). Diets at Guam and American Samoa show similar levels of diversity but a very limited overlap in prey species utilized (Table 4.5 and 4.7). This pattern may result from a heterogeneous distribution of prey species on spatial and/or temporal scales, geographic variation in feeding preferences of *C. ebraeus*, and/or other factors (e.g. competition) that limit access to particular prey in certain locations.

3. Association between geographic variation of conotoxin genes and dietary specialization

Patterns of variation of conotoxin genes are highly influenced by prey heterogeneity. As revealed by canonical correspondence analyses (CCA) (Ter Braak 1986) and regression, diversity of conotoxin genes is positively correlated with dietary diversity at each location (Figure 4.5A-B), a pattern that contrasts with the lack of association between dietary diversity and variation at COI (Table 4.8). Populations at American Samoa and Guam are completely isolated from the population at Hawaii by the first dimension (CCA1) which represents more than 75% of the total variance (Figure 4.5A-B). Such a pattern of isolation is mostly contributed by differences in prey diversities among populations, because of the consistency in the increasing trends of dietary variables and diversities of conotoxin genes (Figure 4.5A-B). Similarly, the geographic variation of conotoxin genes, especially between Hawaii and the other two populations, is highly affected

by prey divergence among populations. The directions and lengths of vectors of dietary variables in the CCA ordination (Figure 4.5C) reveal that geographic divergence of conotoxin genes (Φ_{ST}) is positively associated with prey heterogeneity (D_{ST}) and inversely related with prey similarities (PS_I), results that are supported by regression (Table 4.9).

The positive association of local allelic and nucleotide diversities of conotoxin genes and prey diversity shows that increased diversity of certain venom components is beneficial for capturing diverse sets of prey. The positive association of conotoxin gene and prey diversity for *C. ebraeus* is also exhibited by another *Conus* species. *Conus miliaris* underwent ecological release at Easter Island and consumes more diverse prey at this location than elsewhere in the IWP (Kohn 1978). Gene and nucleotide diversities of two conotoxin genes (*MIL2* and *MIL3*) at Easter Island are higher than those at Guam and American Samoa, despite similar levels of diversity at COI (as calculated from data of Duda and Lee (2009); Table 4.10). This phenomenon is likely to be demonstrated by other venomous taxa as well. For example, snakes employ different envenomation strategies towards different prey (Hayes et al. 2002) and prey species differ in their responses to venoms of different snake (Barlow et al. 2009; Gibbs and Mackessy 2009) and spider (Binford 2001) species. Enhanced variation at particular venom genes may enable predators to better accommodate the temporal and spatial variation of prey, explore new ecological niches and reduce intraspecific competition.

Selection from geographic heterogeneity in predator-prey interactions facilitates divergence of conotoxin genes among populations, a pattern that is most evident for the population at Hawaii . Conotoxin genes at Hawaii possess unique allelic compositions, are subject to distinct selection

forces, and are driven by complete divergence in prey utilization (Figure 4.3 and 4.5; Table 4.4, 4.5 amd 4.7). Though Daltry et al. (1996) detected this relationship among populations of a Malayan pitviper species, debates arose over the universal applicability of this pattern and the existence of confounding factors (Sasa 1999). For *C. ebraeus*, the positive association between geographic differentiation of conotoxin genes and divergence in prey utilization (Figure 4.5C; Table 4.9), however, is not a universal phenomenon. Local populations of *C. ebraeus* at American Samoa and Guam show considerable differences in prey utilization but exhibit no differentiation at the highly polymorphic loci (Table 4.5 and 4.7; Figure 4.3). We propose that this resulted from more intense selection regimes at Hawaii than at other locations, possibly as a consequence of episodic limited availabilities of resources at this locality, a phenomenon that accounts for selection on beak morphologies of Galapagos finches (Grant and Grant 2002). Alternatively, gene flow counteracted the impacts of selection more effectively at American Samoa and Guam than at Hawaii because of lower levels of gene flow associated with the Hawaii population (that are not apparent from examination of COI sequences). Similar to the peripheral speciation mechanism presented by Mayr (1963), local selection pressures generate more prominent effects at the edges of the distribution of *Conus* species because gene flow involving these locations is lower than in the center of its distribution (Duda et al. 2012). Results from analyses of patterns of variation of *C. miliaris*, in which the most isolated and peripheral population at Easter Island exhibits the highest levels of differentiation at conotoxin genes and COI (Duda and Lee 2009), also supports this notion.

The contrasting patterns of variation illustrated by different conotoxin loci imply that the functional roles of these genes' products and/or the evolution of prey defense systems differ.

This phenomenon was also detected in venoms of pitviper species (Daltry, Wuster, and Thorpe 1996; Creer et al. 2003). Hence, some venom genes may track divergent targets and undergo adaptive divergence, while others track conserved targets and do not. These results illustrate that study of evolutionary patterns of multiple loci and populations is essential for understanding the origins of ecological adaptations at the interface of predator-prey interactions.

Acknowledgements

We acknowledge Mike Hadfield (University of Hawaii); Alex Kerr, Barry Smith, Marielle Terbio, Chris Rosario, Cabrini Rivera and Jonathan Lim (University of Guam Marine Lab); and Doug Fenner, Lucy Jacob and Maloy Sabater (Department of Marine and Wildlife Resources, American Samoa) for their assistance with obtaining specimens and feces of *C. ebraeus* at respective locations. We thank Janice Papas (University of Michigan) for suggestions and assistance with statistical analyses. We thank Jianzhi Zhang and Taehwan Lee (University of Michigan) and Gang Chen (University of Rhode Island) for their valuable comments and suggestions on our manuscript. This study is funded by an NSF grant (IOS-0718379) to T. F. D.

Figure 4.1. Gene trees of unique conotoxin gene sequences recovered from venom duct cDNA of *C. ebraeus* individuals at three locations constructed using maximum-likelihood and mid-point rooting.
Major clades labeled with grey bars are of putative single loci and numbers on internal branches are bootstrap values of major clades (except for I-superfamily).

(A) Gene tree of 30 A-superfamily sequences recovered from two individuals at American Samoa, three at Guam and two at Hawaii, constructed with the Tajima 3-parameter (Tamura 1992) +G model. Sequences within clades differed at between one and five nucleotides (nt) (out of a total of 169-185 nt); sequences among clades differed at between 23 and 39 nt (out of a total of 160 nt).

(B) Gene tree of 45 unique O-superfamily sequences obtained from two individuals at American Samoa, five at Guam and two at Hawaii, constructed with the Tamura-Nei (Tamura and Nei 1993) +I model. Sequences within clades differed at between one and ten nt (out of a total of 266-278 nt); sequences among clades differed at between 21 and 61 nt (out of a total of 266 nt).

(C) Gene tree of 22 unique I-superfamily sequences from two individuals at American Samoa, five at Guam and one at Hawaii, constructed with the HKY model. Sequences within clade EI2 differed at between one and 23 nt (out of a total of 229 nt); sequences between the two clades differed at between 43 and 64 nt (out of a total of 226 nt).

(D) Gene tree of 67 M-superfamily sequences from three individuals at American Samoa, six at Guam and two at Hawaii (amplified with the primer set MPr2 (Table 4.1)) constructed with the Tamura-Nei+G model. These sequences fell into more than six major clades. Sequences within clades (except clade 'M1') differed at between one and nine nt out of 217-233 nt while sequences among the six clades differed at between 29 and 59 nt out of 214 nt. Sequences of clade 'M1' differ at maximum of 20 nt, indicating the possibility that these sequences represent two loci. Out of the 37 colonies sequenced from two individuals at American Samoa, one at Guam and two at Hawaii, we only recovered seven sequences with the primer set MPr1 (Table 4.1; GenBank accession numbers JX177162 - JX177168). These sequences differed at a maximum of two nt (out of a total of 233 nt) and represent one putative locus.

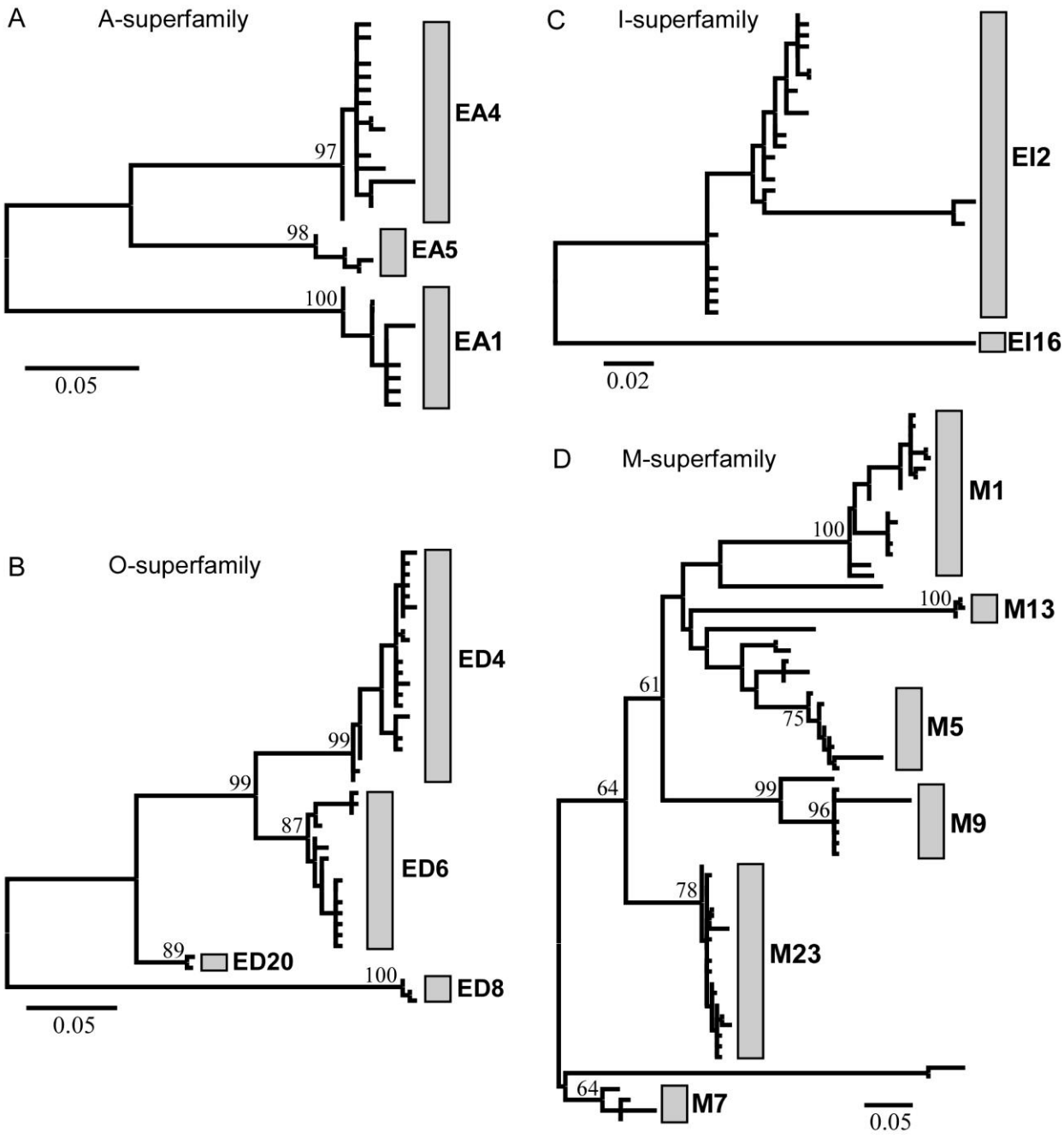

A
A-superfamily
EA4
97
98
EA5
100
EA1
0.05
C
I-superfamily
EI2
EI16
0.02
B
O-superfamily
ED4
99
99
87
ED6
89
ED20
100
ED8
0.05
D
M-superfamily
M1
100
100
M13
61
75
M5
64
99
96
M9
78
M23
64
M7
0.05

Figure 4.2. Alignment of predicted amino acid sequences of alleles of four conotoxin loci of *C. ebraeus*.

The cysteine backbone of each predicted peptide is highlighted in bold; amino acid replacements among alleles are highlighted in grey. *: stop codon. Because one of the locus-specific primers of locus ED4 could only be designed in the region where this site occurs, we do not have sequence data for all individuals at the first polymorphic site. The nucleotide composition of the first three segregating sites are not known for allele 6a because the allele-specific primer for locus ED6 (Table 4.3) occurred in this region and allele 6a is inferred from the sequence chromatogram obtained with this primer set.

Locus	Allele	Predicted Amino Acid Sequences
ED4	40	K G **C** V Q T G G S **C** P S T T G **C** **C** N G L **C** N V D K **C** T * F
	5	K R **C** V Q T G G S **C** P S T T G **C** **C** N G L **C** N V D K **C** T * F
	4	K R **C** V Q T G S S **C** P S T T G **C** **C** S G L **C** N V N K **C** T * L
	9	K G **C** V Q T G G S **C** P S T N G **C** **C** N G L **C** N V N R **C** A * L
	a1	- - - - - T G G S **C** P S T T G **C** **C** N G L **C** N V D K **C** T * L
	a2	- - - - - T G G S **C** P S T T G **C** **C** S G L **C** N V N K **C** T * F
	a4	- - - - - T G S S **C** P S T T G **C** **C** S G L **C** N V N K **C** A * F
	a5	- - - - - T G S S **C** P S T T G **C** **C** N G L **C** N V N K **C** T * F
	a6	- - - - - T G G S **C** P S T T G **C** **C** N G L **C** N V N K **C** T * F
ED6	6	A **C** V N R G D P **C** Q R T V R **C** **C** S R R **C** G I N G **C**
	7	R **C** V N S G D P **C** Q R T V R **C** **C** S R R **C** G I N G **C**
	29	A **C** V N R G D P **C** Q R T V R **C** **C** S R R **C** S I N G **C**
	32	A **C** V N R G D P **C** Q R T V R **C** **C** S R R **C** G V N G **C**
	38	A **C** I N S G D P **C** Q R T V R **C** **C** S R R **C** G V N S **C**
	39	A **C** V N R G D P **C** Q R T V R **C** **C** S R R **C** G V N S **C**
	6a	- - - - R G D P **C** Q R T V R **C** **C** S R R **C** S I N S **C**
E1	E1a	T H S G G A **C** N S H D Q **C** **C** N A F **C** D T A T R T **C** V
	E1b	T D S G G A **C** N S H D Q **C** **C** N E F **C** S T A T R T **C** I
	E1bii	T D S G G A **C** N S H D Q **C** **C** N E F **C** S T A T R T **C** I
	E1c	T R S G G A **C** Y S H N Q **C** **C** D D F **C** S T A T S T **C** V
	E1d	T R S G G A **C** N S H T Q **C** **C** D D F **C** S T A T S T **C** I
	E1e	T R S G G A **C** Y S H N Q **C** **C** D D F **C** S T A T S T **C** I
	E1f	T H S G G A **C** N S H D Q **C** **C** A N F **C** R K A T S T **C** M
	E1g	T R S G G A **C** N S H T Q **C** **C** D H F **C** S T A T S T **C** I
	E1h	T R S G G A **C** N S H D Q **C** **C** A N F **C** R K A T S T **C** M
ED20	20	K G E P **C** N S S V P **C** **C** S G I **C** G Y F N **C** A * L S C R D
	A1	K G E P **C** N W S V P **C** **C** S G I **C** G Y F N **C** A * L S C R D
	A2	K G E P **C** N S S V P **C** **C** S G I * G Y F N **C** A * L S C R D
EA4	2	R I A L I A T R E **C** **C** A N P Q **C** W G K N **C** R * R R * *
	4	R I A L I A T R E **C** **C** A N P Q **C** W A K N **C** R * R R C Y

Figure 4.3. Haplotype networks of alleles of single conotoxin loci of *C. ebraeus* at American Samoa, Guam and Hawaii and multi-locus structure analyses.

(A-E) Haplotype networks of locus (A) ED4, (B) ED6, (C) E1, (D) ED20 and (E) EA4. Haplotypes are illustrated as circles; blue: American Samoa, red: Guam and green: Hawaii. Hypothetical haplotypes are shown as small white circles. Pie diagrams indicate allelic frequencies of haplotypes at each location; areas of circles are proportional to the overall frequencies of each allele combined from all three locations.

(F) Bar plots of results of clustering analyses of four loci (ED4, ED6, ED20 and EA4) with K=2 and K=3 (K: number of clusters). Hypothetical clusters are illustrated with different colors in each plot. AS: American Samoa. Samples pooled from all three locations are more likely divided into two clusters (K=2, log probability of the data ln(Prob(data)) =-332.3; K=3, ln(Prob(data)) =-352.4; K=4, ln(Prob(data)) =-367.7), with the population at Hawaii completely isolated from Guam and American Samoa samples. Further separation of clusters (K>2) only divides samples from Guam and American Samoa irrespective of source.

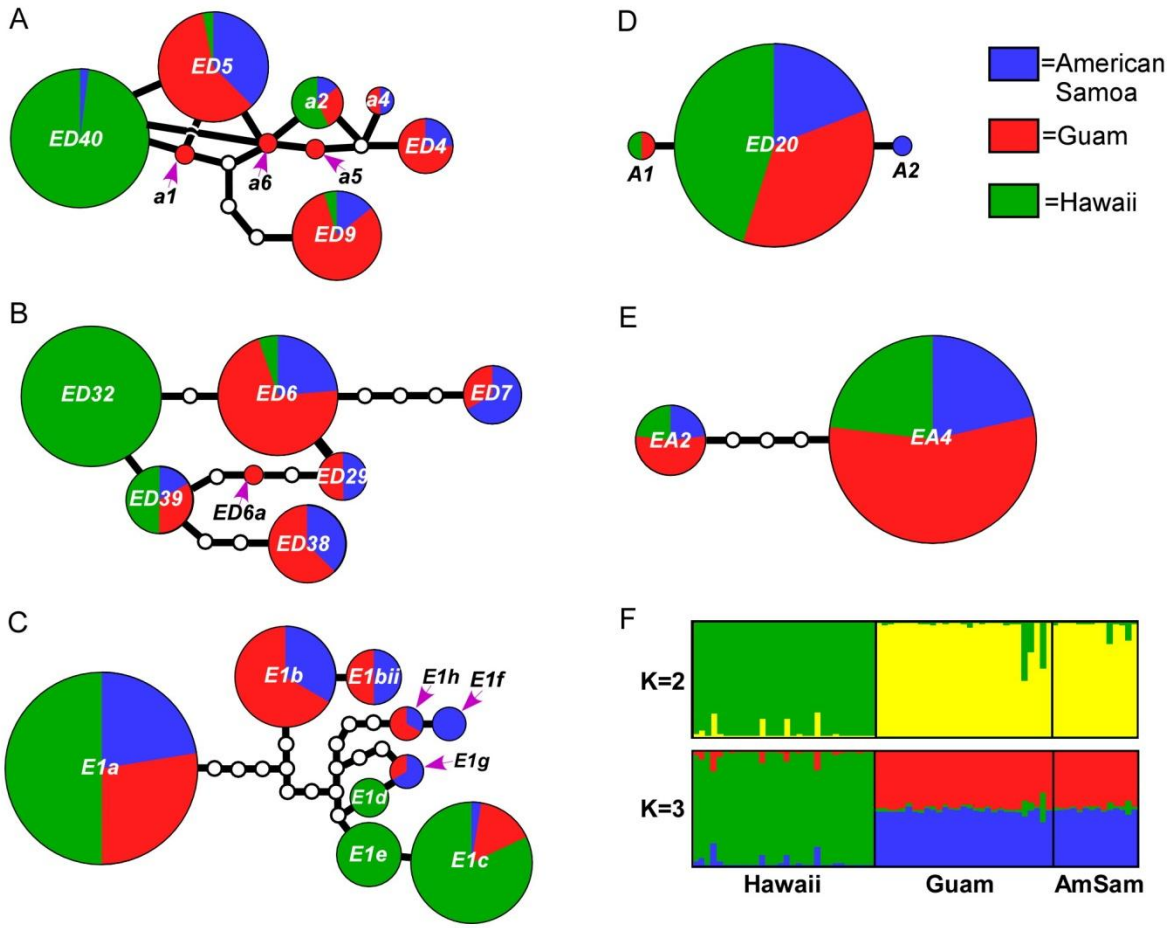


Figure 4.4. Bayesian consensus phylogenies constructed from sequences of a region of the mitochondrial 16S gene recovered from fecal samples here and downloaded from GenBank.
Posterior probabilities are labeled at nodes of major clades. Sequences obtained from *C. ebraeus* fecal samples (GenBank accession numbers JX177300-JX177352, FJ804537-FJ804572 and FJ907334-FJ907342) are highlighted in bold. Names of fecal sequences include the location and the number of identical samples from each location if identical sequences were obtained from more than one individual. GenBank accession numbers of downloaded sequences are included in the names of sequences. Classification of putative prey species are labeled in blue next to the clades. Am Sam: American Samoa.

(A) Phylogeny of sequences of Eunicida species constructed with the HKY+I+G model, rooted with the outgroup *Armandia bilobata*.

(B) Phylogeny of sequences of Nereididae species constructed with the GTR+G model, rooted with the outgroup *Ctenodrilus serratus*.

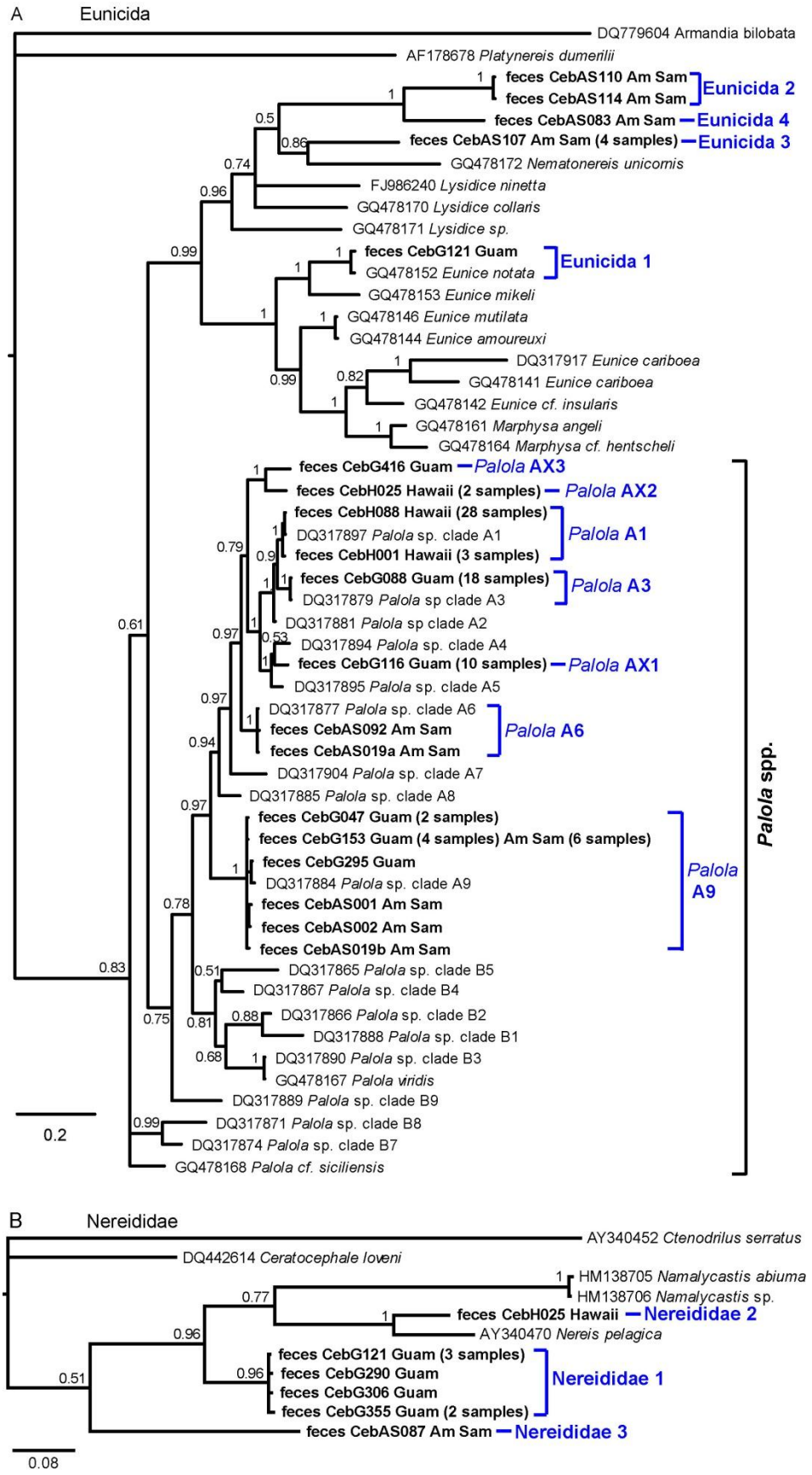
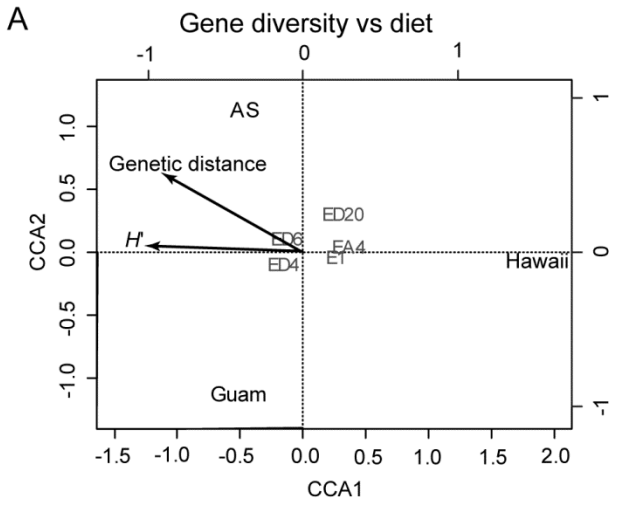
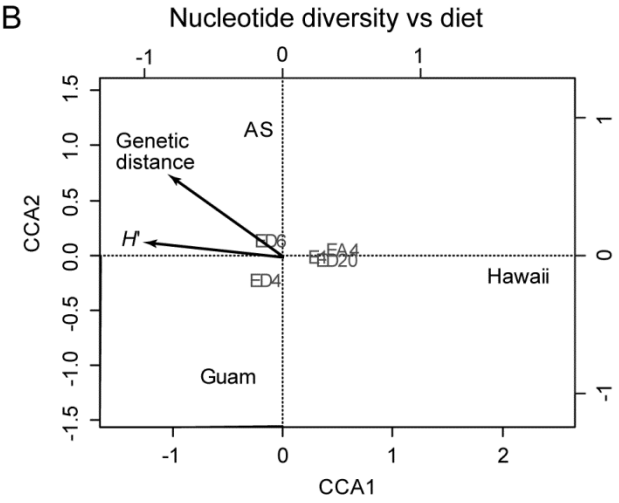

A
Eunicida
DQ779604 Armandia bilobata
AF178678 Platynereis dumerilii
feces CebAS110 Am Sam
feces CebAS114 Am Sam
Eunicida 2
feces CebAS083 Am Sam
Eunicida 4
feces CebAS107 Am Sam (4 samples)
Eunicida 3
GQ478172 Nematonereis unicornis
FJ986240 Lysidice ninetta
GQ478170 Lysidice collaris
GQ478171 Lysidice sp.
feces CebG121 Guam
GQ478152 Eunice notata
Eunicida 1
GQ478153 Eunice mikeli
GQ478146 Eunice mutilata
GQ478144 Eunice amoureuxi
DQ317917 Eunice cariboea
GQ478141 Eunice cariboea
GQ478142 Eunice cf. insularis
GQ478161 Marphysa angeli
GQ478164 Marphysa cf. hentscheli
feces CebG416 Guam
Palola AX3
feces CebH025 Hawaii (2 samples)
Palola AX2
feces CebH088 Hawaii (28 samples)
DQ317897 Palola sp. clade A1
feces CebH001 Hawaii (3 samples)
Palola A1
feces CebG088 Guam (18 samples)
DQ317879 Palola sp clade A3
Palola A3
DQ317881 Palola sp clade A2
DQ317894 Palola sp. clade A4
feces CebG116 Guam (10 samples)
Palola AX1
DQ317895 Palola sp. clade A5
DQ317877 Palola sp. clade A6
feces CebAS092 Am Sam
feces CebAS019a Am Sam
Palola A6
DQ317904 Palola sp. clade A7
DQ317885 Palola sp. clade A8
feces CebG047 Guam (2 samples)
feces CebG153 Guam (4 samples) Am Sam (6 samples)
feces CebG295 Guam
DQ317884 Palola sp. clade A9
feces CebAS001 Am Sam
feces CebAS002 Am Sam
feces CebAS019b Am Sam
Palola A9
Palola spp.
DQ317865 Palola sp. clade B5
DQ317867 Palola sp. clade B4
DQ317866 Palola sp. clade B2
DQ317888 Palola sp. clade B1
DQ317890 Palola sp. clade B3
GQ478167 Palola viridis
DQ317889 Palola sp. clade B9
DQ317871 Palola sp. clade B8
DQ317874 Palola sp. clade B7
GQ478168 Palola cf. siciliensis
0.2
B
Nereididae
AY340452 Ctenodrilus serratus
DQ442614 Ceratocephale loveni
HM138705 Namalycastis abiuma
HM138706 Namalycastis sp.
feces CebH025 Hawaii
Nereididae 2
AY340470 Nereis pelagica
feces CebG121 Guam (3 samples)
feces CebG290 Guam
feces CebG306 Guam
feces CebG355 Guam (2 samples)
Nereididae 1
feces CebAS087 Am Sam
Nereididae 3
0.08

Figure 4.5. Ordination/biplots of canonical correspondence analyses of diversities and geographic divergence of conotoxin genes with heterogeneities of prey items.
These analyses depict and evaluate the contribution of dietary variables to the patterns of variation of conotoxin genes among populations of *C. ebraeus*. Arrows/vectors represent independent dietary variables and are drawn from the centroid of the dispersion of populations. The dependent variables (conotoxin genes at three locations) are labeled with their names and positions indicating their relationships. Dashed lines are horizontal and vertical lines crossing the centroid. The bottom and bottom-left axes represent the 1st (CCA1) and 2nd dimensions (CCA2); the top and top-right axes demonstrate percentages of covariance explained by independent variables. AS: American Samoa.

(A-B) (A) Gene diversities and (B) nucleotide diversities of five conotoxin genes with diversities of prey items (*H*' and genetic distance) at the three locations. Populations are largely discriminated by CCA1. CCA1 in (A) explains 87.9% of the total variance, while CCA1 in (B) explains 77.5%. The vectors of dietary variables point to the same direction as the incremental trend of diversities of conotoxin genes (i.e. American Samoa and Guam > Hawaii), showing that diversities of conotoxin genes and diets are positively related.

(C) Pairwise Φ_{ST} values of conotoxin genes with PS_I and D_{ST} values of prey items among locations. CCA1 explains 99.8% of total constrained eigenvalues while CCA2 explains only 0.2%. Pairwise comparisons of populations are almost completely discriminated at CCA1. ASG: comparison of American Samoa and Guam, GH: Guam and Hawaii, and ASH: American Samoa and Hawaii. Data from locus EA4 were not included because Φ_{ST} values are zero after data conversion and are not informative in analyses of gradients.


A
Gene diversity vs diet
AS
Genetic distance
H'
ED20
ED6
EA4
ED4
ET
Hawaii
Guam
CCA1
CCA2


B
Nucleotide diversity vs diet
AS
Genetic distance
H'
ED6
EA4
ED20
ED4
Hawaii
Guam
CCA1
CCA2

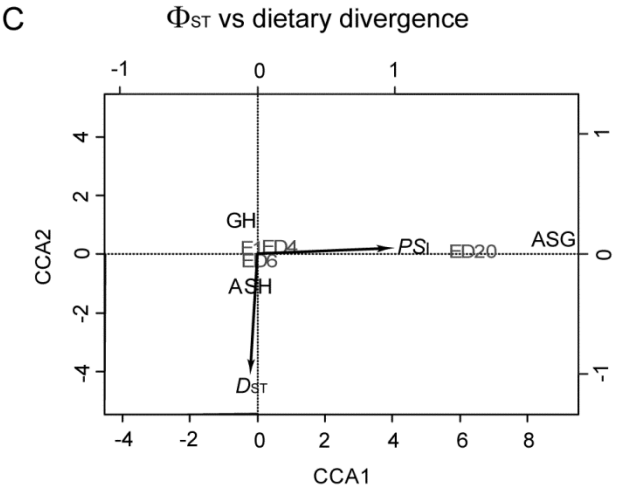

C
ΦST vs dietary divergence
GH
ED4
PSI
ED20
ASG
ED6
ASH
DST
CCA1
CCA2

Table 4.1. General primers for each conotoxin superfamily.
3'UTR: 3' untranslated region.

Conotoxin superfamily	Toxin type	Primer location	Primer sequences
A	α-conotoxin	Prepro	5'ATGGGCATGCGGATGATGTTCAC 3'
		3'UTR	5' GTCGTGGTTCAGAGGGTCCTGG 3'
O	δ-conotoxin	Prepro	5'CATCACCAAGATGAAACTGACGTG 3'
		3'UTR	5' GCGCCAATCAAAGATCAAGCC 3'
M (primer set MPr1)	μ-conotoxin	Prepro	5' CATGATGTCTAAACTGGGAGT 3'
		3'UTR	5' GCAAATCTGAAGGAGACTGCAATC 3'
M (primer set MPr2)		Prepro	5' GTTGAAAATGGGAGTGGTGCT 3'
		3'UTR	5' ATGATATCAACAAACGCTGTCGTTG 3'
I	-	Prepro	5' ATGATGTTTCGATTGACGTCAGTCAG 3'
		3'UTR	5' ACGTCAGGCTTGAGTTATCTGCC 3'

Table 4.2. Locus-specific primers used to genotype each locus.

Loci	Location	Primer Sequences (5' to 3')
ED4	Forward	GTAAAACGACGGCCAGTATTGCACCAGAAAAGATGCGTACAG
	Reverse	CAGGAAACAGCTATGACCGCGCCAATCAAAGATCAAGCC
ED6	Forward	GTAAAACGACGGCCAGTAATTGCACCAGAAAAGATGCRTAAAC
	Reverse	CAGGAAACAGCTATGACCGCGCCAATCAAAGATCAAGCC
ED20	Forward	GTAAAACGACGGCCAGTTAAATTGCACGAGAAATCATGCATTAG
	Reverse	CAGGAAACAGCTATGACCGCGCCAATCAAAGATCAAGCC
EA4	Forward	GTAAAACGACGGCCAGTGATCGCTCTGATCGCCACACGC
	Reverse	CAGGAAACAGCTATGACCTGGAGTAGCAGCGTCTTCAACG

Table 4.3. Allelic-specific primers to verify and differentiate alleles.

Locus	Location	Primer Sequences	Purpose
ED4	Forward	5'GTAAAACGACGGCCAGTCATCAG CAAGATGAAACTGAC3'	differentiate allele 40 from allele 5, verified by sequencing
	Reverse	5'CAGGAAACAGCTATGACCCGATG GACACGAACCACCCGTC3'	
ED6	Forward	5'GTAAAACGACGGCCAGTGCACCA GAAAGCATGCGTAAACAG3'	Differentiate 7+39 allele pairs and 6+38 pairs, verified by sequencing
	Reverse	5'CAGGAAACAGCTATGACCGCGCC AATCAAAGATCAAGCC3'	

Table 4.4. Sample sizes, numbers of total and unique alleles, gene/haplotype diversities, nucleotide diversities and their standard errors (SE), and Tajima's *D* values of the five conotoxin loci at three locations.

American Samoa is abbreviated 'AS'. Tajima's *D* values were estimated for each locus and population with the infinite-allele model and *P*-values were evaluated by estimation of percentage of values in 10,100 simulations that are smaller than observed values. *D* values that are significant based on 0.05 significance levels are labeled with asterisks and highlighted in bold (* represents $P<0.05$, ** represents $P<0.01$, *** represents $P<0.001$). Results of other neutrality tests exhibited consistent patterns as Tajima's *D*.

Locus	Location	Sample size	Alleles (unique)	Gene Diversity (SE)	Nucleotide Diversity (SE)	Tajima's *D*
ED4	**AS**	10	6 (0)	0.632 (0.113)	0.017 (0.010)	0.466
	Guam	24	8 (3)	0.714 (0.041)	0.028 (0.017)	**2.205***
	Hawaii	28	4 (0)	0.201 (0.070)	0.005 (0.004)	**-1.649***
ED6	**AS**	13	5 (0)	0.785 (0.041)	0.053 (0.030)	**1.805***
	Guam	24	6 (1)	0.638 (0.064)	0.041 (0.024)	1.108
	Hawaii	30	3 (1)	0.242 (0.070)	0.011 (0.009)	-0.187
E1	**AS**	21	7 (1)	0.678 (0.064)	0.025 (0.014)	0.837
	Guam	29	6 (0)	0.682 (0.041)	0.023 (0.013)	0.842
	Hawaii	48	4 (2)	0.620 (0.031)	0.022 (0.012)	**3.216*****
ED20	**AS**	11	2 (1)	0.091 (0.081)	0.001 (0.002)	-1.162
	Guam	25	2 (0)	0.040 (0.038)	0.001 (0.002)	**-1.103***
	Hawaii	20	2 (0)	0.050 (0.047)	0.001 (0.002)	**-1.124***
EA4	**AS**	14	2 (0)	0.198 (0.092)	0.004 (0.004)	-0.477
	Guam	36	2 (0)	0.178 (0.056)	0.003 (0.004)	-0.225
	Hawaii	15	2 (0)	0.186 (0.088)	0.004 (0.004)	-0.537

Table 4.5. Pairwise Φ_{ST} values and dietary overlap indices among populations of *C. ebraeus*.
Summary statistics of dietary overlap include proportional similarities indices (PS_I) (Whittaker 1952), estimates of the phylogenetic disparity of prey species among samples (D_{ST} values that are analogous to Φ_{ST}), numbers of prey species shared among samples and total numbers of prey species identified in combined samples of each comparison. Values with associated *P*-values less than 0.05 are labeled with asterisks (* represents *P*-value <0.001) and highlighted in bold. AS: American Samoa.

Comparison	Φ_{ST} values of conotoxin genes					Dietary overlap		
	ED4	ED6	E1	ED20	EA4	PS_I	D_{ST}	Shared species (total)
AS-Guam	-0.009	0.008	-0.008	0.012	-0.025	**0.182***	**0.198***	1 (10)
Hawaii-AS	**0.270***	**0.427***	**0.177***	0.007	-0.035	**0.000***	**0.505***	0 (10)
Hawaii-Guam	**0.226***	**0.349***	**0.167***	-0.022	-0.024	**0.000***	**0.245***	0 (8)

Table 4.6. Results of hierarchical Analysis of Molecular Variance (AMOVA) for the highly polymorphic loci (ED4, ED6 and E1) with the Tamura-Nei model.
Three types of grouping were tested for each locus: H, (G, A) represents grouping of Guam with American Samoa; G, (H, A) represents grouping of Hawaii and American Samoa; A, (H, G) represents grouping of Hawaii and Guam. Percentage of variation among groups, percentage of variation among populations within groups, F_{SC}, F_{ST} and F_{CT} were estimated and presented for each grouping scheme. Significance of F_{SC}, F_{ST} and F_{CT} values was evaluated by 10,100 random permutations. The negative percentage of covariance among groups may result from the linear restriction of the model and large variations within groups. Results showed that levels of variance among groups for the H, (G, A) grouping is much larger than levels of variance among populations within groups, F_{CT} is large, and the *P*-value is the smallest among the three groupings.

Locus	Grouping	Variation among groups (%)	Variation among pops within groups (%)	Variation within pops (%)	F_{SC}	F_{ST}	F_{CT}
ED4	H, (G,A)	20.20	0.34	79.46	$0.004^{P=0}$	$0.205^{P=0}$	$0.202^{P=0.33}$
	G, (H,A)	-1.75	18.58	83.17	$0.183^{P=0}$	$0.168^{P=0}$	$-0.017^{P=0.66}$
	A, (H,G)	-13.45	24.34	89.11	$0.215^{P=0}$	$0.109^{P=0}$	$-0.134^{P=1}$
ED6	H, (G,A)	30.98	1.83	67.19	$0.03^{P=0}$	$0.33^{P=0}$	$0.31^{P=0.34}$
	G, (H,A)	-20.90	44.68	76.22	$0.37^{P=0}$	$0.24^{P=0}$	$-0.21^{P=1}$
	A, (H,G)	-7.29	32.54	74.75	$0.30^{P=0}$	$0.25^{P=0}$	$-0.07^{P=0.67}$
E1	H, (G,A)	17.53	-0.65	83.12	$-0.01^{P=0.01}$	$0.17^{P=0}$	$0.18^{P=0.33}$
	G, (H,A)	-7.01	18.81	88.20	$0.18^{P=0}$	$0.12^{P=0}$	$-0.07^{P=1}$
	A, (H,G)	-5.82	17.32	88.50	$0.16^{P=0}$	$0.12^{P=0}$	$-0.06^{P=0.67}$

Table 4.7. Putative prey species, numbers of each species and total number of prey items of each higher taxonomic level at the three locations.
The summary statistics of local prey diversity are Shannon-Wiener's indices (*H'*) (Shannon 1948) and mean genetic distances. American Samoa is abbreviated 'AS'. *Palola* species A1-A9 were described previously by Schultz (2006) (also see Figure 4.9).

Prey	**Guam**	**AS**	**Hawaii**
Eunicida (total)	(37)	(18)	(33)
Palola spp. (total)	(36)	(11)	(33)
Palola AX1	10	--	--
Palola AX2	--	--	2
Palola AX3	1	--	--
Palola A1	--	--	31
Palola A3	18	--	--
Palola A6	--	2	--
Palola A9	7	9	--
Other spp. (total)	(1)	(7)	(0)
Eunicida 1	1	--	--
Eunicida 2	--	2	--
Eunicida 3	--	4	--
Eunicida 4	--	1	--
Nereididae (total)	(7)	(1)	(1)
Nereididae 1	7	--	--
Nereididae 2	--	--	1
Nereididae 3	--	1	--
Total prey items	44	19	34
H'	1.46	1.47	0.35
Mean genetic distance	0.15	0.22	0.04

Table 4.8. Coefficients of the slope of the fitted line in simple regression analyses of the haplotype and nucleotide diversities of five conotoxin genes and the COI gene with the diversities of prey (*H'* and genetic distance).

Haplotype diversities of the mitochondrial COI gene for populations at Hawaii, Guam and American Samoa are nearly equivalent (0.963 at Hawaii, 0.978 at Guam, 0.947 at American Samoa; retrieved from Duda and Lessios (2009)). The Guam population exhibits slightly higher nucleotide diversity (0.009 at Guam, 0.006 at American Samoa and Hawaii; estimated with Tamura-Nei model from the COI gene sequences reported in Duda and Lessios (2009)). Linear regressions of measures of diversity at COI with dietary diversity (*H'* and genetic distance) showed lack of correlation between the mitochondrial marker and prey. Diversities of conotoxin loci ED20 and EA4 do not show correlations with diets, but positive relationships were detected at loci ED4, ED6 and E1.

Locus	Haplotype diversity vs *H*'	Haplotype diversity vs genetic distance	Nucleotide diversity vs *H*'	Nucleotide diversity vs genetic distance
ED4	0.423	2.597	0.016	0.079
ED6	0.422	3.069	0.032	0.237
E1	0.054	0.344	0.001	0.016
ED20	0.014	0.199	0	0
EA4	0.002	0.054	0	-0.001
COI	-0.001	-0.069	0.001	0.002

Table 4.9. Pearson , Spearman and Kendall correlation coefficients of the pairwise Φ_{ST} matrices of each of the three highly polymorphic conotoxin genes with the pairwise divergence indices of prey (PS_I and D_{ST}).
Coefficients of Φ_{ST} with Pianka's overlap index are identical to those with PS_I.

Locus	PS_I			D_{ST}		
	Pearson	**Spearman**	**Kendall**	**Pearson**	**Spearman**	**Kendall**
ED4	-0.999	-0.866	-0.817	0.727	1.000	1.000
ED6	-0.985	-0.866	-0.817	0.746	1.000	1.000
E1	-0.999	-0.866	-0.817	0.655	1.000	1.000

Table 4.10. Haplotype/gene diversity and nucleotide diversity of two O-superfamily conotoxin genes *MIL2* and *MIL3* and the mitochondrial COI gene of *C. miliaris* populations at Easter Island (abbreviated as EI), Guam and American Samoa (abbreviated 'AS'). Standard deviations of indices are presented in parentheses. Distances among haplotypes are calculated with respective models used in Duda and Lee (2009): K80 (Kimura 1980) model for locus *MIL2*, Jukes-Cantor (Jukes and Cantor 1969) model for locus *MIL3*, and Tamura-Nei model for the COI gene.

Locus	Gene/Haplotype Diversity (Standard Deviation)			Nucleotide Diversity (Standard Deviation)		
	EI	**Guam**	**AS**	**EI**	**Guam**	**AS**
MIL2	0.635 (0.043)	0.271 (0.084)	0.381 (0.094)	0.015 (0.009)	0.010 (0.007)	0.014 (0.009)
MIL3	0.747 (0.036)	0.594 (0.070)	0.631 (0.064)	0.021 (0.001)	0.017 (0.002)	0.017 (0.001)
COI	0.961 (0.014)	1.000 (0.017)	0.979 (0.016)	0.008 (0.004)	0.010 (0.005)	0.010 (0.005)

References

Andrade SCS, Solferini VN. 2006. The influence of size on the radula of *Littoraria angulifera* (gastropoda: Littorinidae). *Malacologia.* 49:1-5.

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 B.* 276:2443-2449.

Binford GJ. 2001. Differences in venom composition between orb-weaving and wandering Hawaiian *Tetragnatha* (Araneae). *Biol J Linn Soc.* 74:581-595.

Brodie ED, Ridenhour BJ, Brodie ED. 2002. The evolutionary response of predators to dangerous prey: Hotspots and coldspots in the geographic mosaic of coevolution between garter snakes and newts. *Evolution.* 56:2067-2082.

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

Clement M, Posada D, Crandall KA. 2000. TCS: A computer program to estimate gene genealogies. *Mol Ecol.* 9:1657-1659.

Creer S, Malhotra A, Thorpe RS, Stocklin RS, Favreau PS, Hao Chou WS. 2003. Genetic and ecological correlates of intraspecific variation in pitviper venom composition detected using matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF-MS) and isoelectric focusing. *J Mol Evol.* 56:317-329.

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

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

Duda TF, Lessios HA. 2009. Connectivity of populations within and between major biogeographic regions of the tropical pacific in *Conus ebraeus,* a widespread marine gastropod. *Coral Reefs.* 28:651-659.

Duda TF, Palumbi SR. 1999. Molecular genetics of ecological diversification: Duplication and rapid evolution of toxin genes of the venomous gastropod *Conus. Proc Natl Acad Sci USA.* 96:6820-6823.

Duda TF, Terbio M, Chen G, Phillips S, Olenzek AM, Chang D, Morris DW. 2012. Patterns of population structure and historical demography of *Conus* species in the tropical Pacific. *Am Malacol Bull.* 30:175-187.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform.* 1:47-50.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial-DNA restriction data. *Genetics.* 131:479-491.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics.* 147:915-925.

Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics.* 133:693-709.

Gibbs HL, Mackessy SP. 2009. Functional basis of a molecular adaptation: Prey-specific toxic effects of venom from *Sistrurus* rattlesnakes. *Toxicon.* 53:672-679.

Gibbs HL, Sanz L, Chiucchi JE, Farrell TM, Calvete JJ. 2011. Proteomic analysis of ontogenetic and diet-related changes in venom composition of juvenile and adult dusky pigmy rattlesnakes (*Sistrurus miliarius barbouri*). *J Proteomics.* 74:2169-2179.

Grant PR, Grant BR. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. *Science.* 296:707-711.

Gross HP, Anderson JM. 1984. Geographic variation in the gillrakers and diet of European threespine sticklebacks, *Gasterosteus aculeatus*. *Copeia.* 1984:87-97.

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.

Hayes W, Herbert S, Reeling G, Gennaro J. 2002. Factors that influence venom expenditure by vipers and other snakes during predatory and defensive contexts. *in* Schuett Gw, Höggren M, and Greene Hw, eds. Biology of the vipers. Traverse City: Biological Sciences Press, p. 207-234.

Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour.* 9:1322-1332.

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

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. *in* Munro Hn, ed. Mammalian protein metabolism. New York: Academic Press, p. 21-123.

Kaas Q, Westermann J-C, Craik DJ. 2010. Conopeptide characterization and classifications: An analysis using Conoserver. *Toxicon.* 55:1491-1509.

Kendall MG. 1948. Rank correlation methods. Oxford, England:Griffin.

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. 1978. Ecological shift and release in an isolated population - *Conus miliaris* at Easter Island. *Ecol Monogr.* 48:323-336.

Kraaijeveld AR, Godfray HCJ. 1999. Geographic patterns in the evolution of resistance and virulence in *Drosophila* and its parasitoids. *Am Nat.* 153 (S5):S61–S74.

Lessios HA. 1992. Testing electrophoretic data for agreement with Hardy-Weinberg expectations. *Mar Biol.* 112:517-523.

Librado P, Rozas J. 2009. Dnasp v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 25:1451-1452.

Mackessy SP, Sixberry NA, Heyborne WH, Fritts T. 2006. Venom of the brown treesnake, *Boiga irregularis*: Ontogenetic shifts and taxa-specific toxicity. *Toxicon.* 47:537-548.

Mayr E. 1963. Animal species and evolution. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Oksanen J, Kindt R, Legendre P, O'Hara R. 2006. Vegan: Community ecology package.

Palkovacs EP, Post DM. 2008. Eco-evolutionary interactions between predators and prey: Can predator-induced changes to prey communities feed back to shape predator foraging traits? *Evol Ecol Res.* 10:699-720.

Paterson S, Vogwill T, Buckling A et al. 2010. Antagonistic coevolution accelerates molecular evolution. *Nature.* 464:275-278.

Pianka ER. 1974. Niche overlap and diffuse competition. *Proc Nat Acad Sci USA.* 71:2141-2145.

Posada D. 2008. jModelTest: Phylogenetic model averaging. *Mol Biol Evol.* 25:1253-1256.

Rambaut A. 2002. Se-Al sequence alignment editor. Version 2.0.A11. Oxford: University of Oxford.

Raymond M, Rousset F. 1995. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *J Hered.* 86:248-249.

Rodgers JL, Nicewander WA. 1988. Thirteen ways to look at the correlation coefficient. *Am Stat.* 42:59-66.

Rousset F. 2008. Genepop'007: A complete reimplementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resour.* 8:103-106.

Sanford E, Worth DJ. 2010. Local adaptation along a continuous coastline: Prey recruitment drives differentiation in a predatory snail. *Ecology.* 91:891-901.

Sasa M. 1999. Diet and snake venom evolution: Can local selection alone explain intraspecific venom variation? *Toxicon.* 37:249-252.

Schluter D, Grant PR. 1984. Ecological correlates of morphological evolution in a Darwin's finch, *Geospiza difficilis*. *Evolution.* 38:856-869.

Schulze A. 2006. Phylogeny and genetic diversity of palolo worms (*Palola*, Eunicidae) from the tropical north Pacific and the Caribbean. *Biol Bull.* 210:25-37.

Shannon CE. 1948. A mathematical theory of communication. *Bell Syst Tech J.* 27:379-423, 623-656.

Smith TB. 1993. Disruptive selection and the genetic basis of bill size polymorphism in the African finch *Pyrenestes*. *Nature.* 363:618-620.

Spearman C. 1910. Correlation calculated from faulty data. *Brit J Psychol, 1904-1920.* 3:271-295.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 123:585-595.

Tamura K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol Biol Evol.* 9:678-687.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 10:512-526.

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.

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Ter Braak CJF. 1986. Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology.* 67:1167-1179.

Thompson JN. 2005. The geographic mosaic of coevolution. Chicago: The University of Chicago Press.

Van Valen L. 1973. A new evolutionary law. *Evol Theor.* 1:1-30.

Whittaker RH. 1952. A study of summer foliage insect communities in the Great Smoky Mountains. *Ecol Monogr.* 22:1-44.

CHAPTER 5 ONTOGENETIC PLASTICITY OF FEEDING TRAITS ASSOCIATED WITH DIETARY BREADTH

This chapter will be submitted for publication with the coauthor Thomas F. Duda, Jr.

Introduction

Phenotypic variation depends upon genotypes, environmental conditions and norms of reactions (West-Eberhard 2005). Developmental plasticity, variation of phenotypes in response to ecological changes through development without alteration of coding sequences (Piersma and Drent 2003), derives in part from epigenetic variation (Scheiner 1993). Plasticity of gene expression, combined with natural selection, facilitates ultimate fixation of a specific norm of reaction, a process of ecological adaptation termed canalization (Scheiner 1993; Valena and Moczek 2012). Studies of plasticity and canalization of gene expression during ontogeny allow us to assay the role of gene regulation in ecological adaptation and to determine the relationship between gene regulation and changes of ecological variables.

Ecological characteristics of organisms such as habitat use, prey utilization and mate choice are typically modified during development. Numerous metazoans exhibit transformation of traits associated with feeding in response to dietary transitions among discrete life history stages. Examples include the bite force of slider turtles (Herrel and O' Reilly 2006), cranial

musculoskeletal system of water snakes (Vincent et al. 2007), mandible sizes of damselfishes (Frederich, Adriaens, and Vandewalle 2008), gill rakes of alewifes (MacNeill and Brandt 1990) and radula morphology of marine snails (Nybakken 1990; Kawamura, Roberts, and Yamashita 2001) that are associated with dietary changes during development. Several species of snakes and jellyfish exhibit ontogenetic shifts in venom composition (Andrade and Abe 1999; MacKessy, Williams, and Ashon 2003; Kintner, Seymour, and Edwards 2005; Mackessy et al. 2006; Alape-Giron et al. 2008; Antunes et al. 2010; Zelanis et al. 2010), and these shifts appear to be associated with changes in diets during development. For example, MacKessy et al. (2003) detected an association between changes in venom composition and diets of Pacific rattlesnakes that show a time lag (shift in venom composition that follows dietary changes), hinting a causal relationship between these two factors. Nonetheless, these findings are based on proteomic analyses and functional assays of venom, while plasticity of venom gene expression is unknown. As an exception, Zelanis et al. (2012) detected differences in diversity of venom genes in venom gland transcripts of newborn and adult snakes by cloning and sequencing venom duct cDNA libraries, but causes of such a differentiation in expression are unclear.

Here we report the evaluation of the association between ontogenetic shifts of venom gene expression and prey utilization of predatory marine snails *Conus*. Life histories of many *Conus* species include discrete stages as planktotrophic larvae, juveniles, subadults and adults (Kohn and Perron 1994). Larvae feed on phytoplankton, nanoplankton and detritus, while most *Conus* species hunt polychaetes after settlement (Kohn and Perron 1994). Dietary composition changes substantially with increase of body sizes in some species (Kohn and Nybakken 1975), and

Leviten (1976) suggests that diets of vermivorous *Conus* species generally shift from being trophic specialists as juveniles to generalists as subadults to specialists as adults.

Some *Conus* species exhibit changes in characters associated with predation through development. *Conus* use radular teeth, hollow needles with barbs, to penetrate the epidermis of prey and inject venom (Nybakken 1990). Changes in radular tooth morphology from juvenile to adult stages are likely coupled with changes in prey specialization (Nybakken 1988; Nybakken and Perron 1988; Nybakken 1990). In addition to using harpoon-like radular teeth in their feeding apparatus, cone snails synthesize and utilize venom, cocktails of neurotoxins termed conotoxins, to paralyze prey (Olivera 2002). Venom composition varies within species (Jakubowski et al. 2005; Davis, Jones, and Lewis 2009; Rivera-Ortiz, Cano, and Mar í2011), but little is known about the molecular mechanism responsible for this variation and its association with trophic resource utilization. Conotoxins are expressed by members of many large gene families (A, D, I, J, L, M, O, P, S, T, V and Y) (Kaas, Westermann, and Craik 2010). These genes undergo extensive gene duplication and rapid evolution (Duda and Palumbi 1999; Chang and Duda 2012), and their expression is highly plastic (Duda and Palumbi 2004; Duda and Remigio 2008) (also see Chapter 3). Allelic variation of conotoxin genes among populations is positively associated with local diversities and geographic differentiation of diets (Duda 2008; Duda et al. 2009) (also see Chapter 4).

We chose a vermivorous species *Conus ebraeus* as the research body, because this species is widely and abundantly distributed in the Indo-West Pacific and dietary and venom composition of this species have been investigated by several studies. For example, a population from the Eastern Indian Ocean handles different prey species among size classes: *Nereis jacksoni* (10-

25mm shell length), *Perinereis singaporiensis* (22-28mm shell length), *Palola siciliensis* (>31mm) (Kohn and Nybakken 1975). Leviten (1976) discovered a decrease in prey diversity of *C. ebraeus* with increase in body size (shell lengths larger than 11mm). Individuals at Okinawa with shell lengths smaller than 13 mm specialize on members of polychaete family Syllidae, while larger ones feed on families Eunicidae, Nereididae and Capitellidae polychaetes (Duda, Kohn, and Matheny 2009). In addition to knowledge of dietary shifts of this species, we have alignments of conotoxin gene sequences of several single loci from previous population genetic (Duda et al. 2009) (Chapter 4) and evolutionary studies (Duda and Palumbi 1999), all of which facilitate the experimental design of this study.

We specifically addressed the following questions: does conotoxin gene expression vary through development? Which genes are up-regulated or uniquely expressed in each ontogenetic stage? Is dietary composition of individuals at Guam also distinct among developmental stages? Are ontogenetic changes in conotoxin gene expression and diets intricately linked? If so, how are they related? To answer these questions, we sampled *C. ebraeus* individuals at Guam, examined the population structure of these individuals, identified prey species, quantified levels of expression of conotoxin genes of individuals representing different developmental stages, and determined the relationship between ontogenetic shifts of conotoxin gene expression and diets.

Materials and Methods

1. Specimens

We collected specimens of *Conus ebraeus* at Pago Bay, Guam in May 2010. We measured sizes of individuals (shell lengths, widths and heights) immediately after field collection. As described

by Duda et al. (2009), we placed individuals into separate cups with sea water, collected feces after their defecation, and preserved fecal samples in 95% ethanol. We determined sex and maturity of each specimen based on the presence of a penis. We preserved venom ducts in RNAlater (Ambion, Inc.) and stored them in the -20 ℃ freezer.

2. Identification of prey items

We identified prey species from fecal samples with a DNA barcoding approach described by Duda et al. (2009). We aligned sequences of the mitochondrial 16S rDNA gene retrieved from fecal samples in Se-Al 2.0 (Rambaut 2002) with sequences of putative polychaete species recovered from previous studies of annelid phylogeny and dietary studies of *C. ebraeus* (Schulze 2006; Duda et al. 2009; Chang and Duda 2012) (GenBank accession numbers listed in Figure 5.1). We classified these sequences into three groups (Eunicida, Nereididae, Syllidae) based on their phylogenetic similaries. We performed model selection in jModelTest v0.1.1 (Posada 2008) with alignments of 16S gene sequences from each group, and built Bayesian consensus phylogenies for each group with best models (10,000,000 generations, two runs, four chains, 25% burnin). We determined prey species based on phylogenetic relationships of fecal gene sequences with sequences of known or pre-defined polychaete species in phylogenies.

3. Analyses of shifts in diet

Shell lengths have been used as an approximation of ages of cone snails (Kohn and Nybakken 1975; Leviten 1976; Duda, Kohn, and Matheny 2009). To visualize the age differences of *C. ebraeus* individuals subduing specific types of prey, we constructed boxplots of shell lengths of individuals consuming different prey species and categories of higher taxonomic levels. We

performed one-way Analysis of Variance (ANOVA) of shell lengths of individuals among groups (of prey species or categories of higher taxonomic levels) with function *lm* in R v2.15.0 (R Development Core Team 2012) to determine if individuals of different sizes/developmental stages show differences in diet.

To evaluate patterns of transition in dietary composition among individuals of different sizes, we built a heatmap based on percentages of each prey species handled by individuals of a specific shell length with the *heatmap.2* function in the *gplots* package (Warnes 2012) in R v2.15.0. We chose Shannon-Weiner index (*H'*) (Shannon 1948), Simpson's index (*S*) (Simpson 1949) and average genetic distances (GD) to quantify levels of prey diversity. To calculate GD, we estimated pairwise genetic distances of 16S gene sequences of prey species with the Tamura-Nei (Tamura and Nei 1993) +G model and complete deletion in Mega 5.05 (Tamura et al. 2011), and then computed the average genetic distances. To monitor the transitional pattern of dietary diversity through ontogeny, we performed sliding-window analyses of *H'*, S and GD of dietary compositions within window sizes of 2, 3, 4, 5 and 10 mm in shell lengths. These analyses were done in R v2.15.0 with function *diversity* in package *vegan* (Oksanen et al. 2012).

We also used F-statistics to evaluate the genetic differentiation of dietary compositions among groups of individuals, treating each group (e.g. individuals with shell lengths between 10-15mm, 11-16mm, etc.) as a single population. We constructed populations with the sliding window approach (5mm range of shell lengths for each window), and estimated pairwise Φ_{ST} values (D_{ST} or the phylogenetic disparity of prey items among samples) of 16S gene sequences of prey items among populations in Arlequin 3.1 (Excoffier, Laval, and Schneider 2005) with the Tamura-Nei

distance model. *P*-values were estimated from Monte Carlo simulations of 10,100 replicates. We built a heatmap of absolute D_{ST} values along the gradient of shell lengths with the approach described earlier.

Based on changes of slopes of each point in the plot of results from sliding window analyses (of 5mm in shell lengths) of dietary diversities (*H'*, *S* and *GD*) and the significance of genetic differences (D_{ST}), we defined ranges of shell lengths of small, medium and large groups based on patterns of shifts in dietary diversity. We tested if the three groups show differences in dietary compositions by Fisher's exact tests (Fisher 1954) with *fisher.test* function in R, and *P*-values were determined with 100,000 generations of simulation for each test.

4. Quantification of conotoxin gene expression

We extracted messenger RNA from venom ducts of 60 *C. ebraeus* individuals (shell lengths ranging from 7mm to 26 mm) and prepared cDNA following the approach described previously (Duda and Palumbi 1999). We selected six conotoxin genes from pools of putative single genes identified from population studies of this species in Indo-West Pacific (Duda et al. 2009) (also see Chapter 4): locus E1 (O-superfamily locus that putatively encodes ω-conotoxin), locus EA1 and EA4 (A-superfamily loci that putatively encode α-conotoxin), locus ED4, ED8 and ED20 (O-superfamily loci that putatively encode δ-conotoxin). Expression of these genes are putatively ontogenetically related because these gene sequences are differentially recovered from cDNA samples of individuals of different sizes in initial screenings of these cDNA. Contamination of genomic DNA in the cDNA of these individuals can inflate levels of expression of conotoxin genes measured by Real-time qPCR. To avoid the impact of 'genomic carryover' on

quantification of expression levels, we designed sets of primers that span an intron position of these genes. The target gene region spanning the intron is too long to be amplified from the genomic DNA with our approach, and gene transcripts in venom duct cDNA are preferably amplified in the Real-time PCR. We designed locus-specific reverse primers annealing to the toxin-coding region downstream to the intron(s) (Table 5.1), and paired them with general forward primers that anneal to the conservative prepro region upstream to the intron(s) and are specific to members of a conotoxin superfamily (Table 5.1). We tested specificity of these sets of primers for individuals with known genotypes.

We used Real-time qPCR with SYBR Green chemistry to quantify levels of expression of conotoxin genes. To ensure the same amount of venom duct gene transcripts to be used across rounds of Real-time qPCR for all genes so that expression levels quantified in each PCR run are comparable, we added Tris buffer to each sample to a total volume of 175 μ L, and aliquoted equal volume of cDNA samples for each run of qPCR. We chose a β-tubulin gene as the endogenous control and estimated abundance of its gene transcript in venom duct cDNA of individuals with Real-time qPCR and primers specific for this locus (forward primer 5'ACAGCAGCTACTTTGTTGAATGGAT3' and reverse primer 5'CAGTGTACCAATGGAGGAAAGCC3'). We performed all Real-time qPCR runs in an ABI Prism 7500 machine at the Molecular Biology Core Laboratory at University of Michigan School of Dentistry. To reduce noise and avoid potential errors, in each run we prepared three identical PCR samples for each individual and used average results of the three samples as the estimated C_T value of that individual. The PCR procedure involves ten minutes of initial denaturing at 95℃ and 40 cycles of amplification (denaturing: 95℃ for 15 seconds; annealing:

54 ℃ for 30 seconds; elongation: 72 ℃ for 35 seconds where fluorescent signals were collected). We added a dissociation stage (95 ℃ for 15 seconds, 60 ℃ for one minute, and 95 ℃ for 15 seconds for each sample) at the end of each run to evaluate the specificity of amplification. The dissociation stage measures temperatures at which amplified products renature, and can be used to detect presence of non-specific PCR products if multiple temperatures of renaturation are detected.

We quantified abundance of conotoxin gene transcripts in the venom duct cDNA of *C. ebraeus* individuals with Real-time qPCR using locus-specific reverse primers and general forward primers (Table 5.1). Each round of qPCR was performed on three replicated samples of each individual, and cycles were the same as that of the β-tubulin locus except that the annealing temperature for locus E1 and ED4 was 60 ℃. To ensure the similarity in efficiencies of primers of conotoxin genes with primers of the β-tubulin gene, we made 1/5 and 1/25 dilutions of cDNA samples of up to 12 individuals and compared efficiencies of these primers with the approach described by Schmittgen and Livak (2008). We used the comparative C_T method (Schmittgen and Livak 2008) to estimate levels of expression of these conotoxin genes relative to the endogenous β-tubulin gene with the assumption that levels of expression of the endogenous gene are invariable among individuals. The C_T value of any sample labeled as 'undertermined' in each qPCR was converted to 40 (C_T of 40 means no amplified products was detected). We estimated ΔC_T values of each conotoxin gene relative to the endogenous gene by subtracting average C_T values of conotoxin genes among three replicates of each individual with that of the β-tubulin gene, and calculated relative levels of expression of conotoxin genes with the formula $2^{-\Delta C_T}$ (Schmittgen and Livak 2008).

5. Assessment of population structure

We determined patterns of ontogenetic shifts of dietary specialization and conotoxin gene expression by investigating patterns of variation of these variables among *C. ebraeus* individuals of different sizes/ages. But this approach assumes that our samples represent the local population at Guam. To test if the 60 individuals of *C. ebraeus* sampled here are cohorts self-recruited locally instead of migrants from other regions of Indo-West Pacific, we tested if specimens of different size classes exhibited any structure of difference at conotoxin locus E1. Hawaiian population exhibits significant difference in allelic variation from other populations in the Indo-West Pacific at locus E1 (Duda et al. 2009)(also see Chapter 4), and investigation of structure of individuals at this locus allows us to find out if certain cohorts of individuals sampled here migrated from Hawaii. We genotyped locus E1 for these individuals by PCR amplification with E1 primers described in (Duda et al. 2009) (primer TOX1: 5'CATCGTCAAGATGAAACTGACGTG3', and primer TOX2: 5'CACAGGTATGGATGACTCAGG3') and Sanger sequencing at University of Michigan Sequencing Core facility. We estimated pairwise Φ_{ST} values at locus E1 in Arlequin 3.1 with the Tamura-Nei distance model among groups of individuals classified by size (traditional classification of size classes for *Conus* based on shell lengths: <10mm small, 10-20mm medium and >20mm large) and adults (>20mm in shell lengths) at American Samoa, Guam and Hawaii retrieved from previous studies (Chapter 4). We also estimated pairwise Φ_{ST} values among individuals of each sliding-window (window size of 5mm in shell lengths). We determined significance of results by 10,100 repeats of random permutations.

6. Analyses of patterns of conotoxin gene expression

To normalize the data of conotoxin gene expression for statistical analyses, we used $-\Delta C_T$ as an approximation to the log transformation of levels of expression of conotoxin genes relative to the β-tubulin gene (log ($2^{-\Delta C_T}$)). We constructed a heatmap with *C. ebraeus* individuals as row variables, conotoxin loci as column variables and absolute and scaled values of $-\Delta C_T$ as input, and plotted dendograms for row and column variables with the same approach described previously. This heatmap presents a visualization of the dataset with patterns of color changes and enables us to identify samples of low quality. We determined individual samples that grouped separately from others in the column dendogram and exhibited no amplification for β-tubulin gene or conotoxin genes as cDNA samples of low quality (because of failed amplification from these samples in several Real-time PCR runs). We measured Euclidean distances of relative levels of expression of conotoxin genes among samples with the function *daisy* in package 'cluster' (Maechler et al. 2012) in R v2.15.0, and performed hierarchical clustering analyses with single linkage, complete linkage, average linkage and Wald's method (Struyf, Hubert, and Rousseeuw 1997) with the function *agnes* in the package *cluster* to identify potential hierarchical structures of conotoxin gene expression. We selected the best model (Wald's method) based on differences in agglomerative coefficients (a measure of quality of clustering) (Rousseeuw 1986) among models. We examined the frequency distribution of shell lengths of individuals within each cluster/group to identify differences between/among clusters. We also used a non-hierarchical clustering approach PAM (Partitioning Around Medoids) (Struyf, Hubert, and Rousseeuw 1997) with the function *pam* in package *cluster* to evaluate the consistency of the structure.

To identify the intrinsic trend of relative levels of expression of conotoxin genes through development, we employed a moving average approach (or sliding window analyses of different window sizes) that calculates average $-\Delta C_T$ values of each conotoxin gene among individuals of a specific size range (e.g. range of 5mm in shell lengths). To evaluate the impact of variation in sample sizes on patterns of expression, we constructed reduced datasets by randomly drawing two samples from all individuals with the same shell length and pooling them together. We repeated this drawing for 100 times and plotted moving averages (window size of 5mm in shell lengths) of these simulated datasets with the same approach described here.

7. Analyses of the association of shifts of diets and venom

We determined the relationship between conotoxin gene regulation and dietary specialization, using samples of individuals with known prey species (identified in the previous section). To visualize patterns of conotoxin gene expression among different types of prey, we built boxplots of $-\Delta C_T$ of each conotoxin gene of individuals with the same prey species and categories of higher taxonomic levels. One-way ANOVAs were performed in R to test if expression levels differ among groups (for example, groups classified based on prey species).

For visual examination of the possible association between venom and diets through ontogeny, we built heatmaps of expression levels of the six conotoxin genes among individuals with known diets using the function *heatmap.2* as described earlier. We performed hierarchical clustering analysis with the Wald's method on this 'reduced' dataset (compared to the complete dataset of 60 *C. ebraeus* individuals) with function *agnes* (as described earlier) to determine the structure and evaluate size and dietary differences among clusters. We tested if dietary composition differs

between clusters with Fisher's Exact Tests, and estimated *P*-values with Monte Carlo simulation of 100,000 replicates.

To compare patterns of ontogenetic changes of conotoxin gene expression and dietary composition, we centered and standardized results of sliding window analyses (5mm in shell lengths) of $-\Delta C_T$ values of all conotoxin genes as well as dietary diversities (*H'*, *S* and *GD*), and superimposed these series on a single plot. We treated results of sliding-window analyses of expression levels of each conotoxin gene and variable of dietary diversity along the gradient of shell lengths as independent time series. We tested if two time series of conotoxin gene expression are positively correlated with a lead or lag of time series of dietary diversities (cross-correlations) with the *ccf* function, and verified the significance of results with the linear regression model (function *lm*) in R.

8. Relationships of diets and expression of conotoxin genes with sexual maturity

We identified the sex of *C. ebraeus* individuals based on the presence of a discernible penis, and determined the shell length of the smallest individual with an identifiable penis as the stage of sexual maturity. For individuals without a penis (putative females), we designated individuals smaller than the cutoff shell length as immature juveniles and larger ones as mature females. We compared dietary compositions and relative expression levels of the six conotoxin genes among the three sex or maturity groups (immature juveniles, mature males and mature females) with boxplots. We tested the similarity of dietary composition among groups with Fisher's exact tests (function *fisher.test*) and estimated *P*-values from 1,000 repeats of bootstrap analyses. We used one-way ANOVA (function *lm*) and t-statistics (function *t.test*) to test the similarity of conotoxin

gene expression among/between the sex/maturity groups. All R scripts used in this study are available upon request.

Results

1. Identification of prey species

We collected 151 fecal samples containing remains of polychaetes from 243 *C. ebraeus* individuals, among which 86 samples were identified via microscopic examination to be of familily Nereididae, six of family Syllidae, one of family Terebellidae, 27 of the genus *Palola*, 22 samples of genera other than the genus *Palola* in order Eunicida. The rest of samples are identifiable with this approach because of limited morphological remain in the fecal samples. We successfully obtained 16S gene sequences from 77 fecal samples, among which 54 fecal samples (of 54 individuals) represent annelids based on BLAST analyses of the 16S gene sequences, while sequences of other samples were blasted to be human or bacteria contamination. Initial phylogenetic investigation classified these prey species into two annelid orders: Eunicida and Phyllodocida (families Nereididae and Syllidae). Based on phylogenetic similarities of these sequences with sequences of known annelid species (Figure 5.1) and species defined in previous studies of diets of *Conus ebraeus* (see dissertation chapter 4), we determined that these fecal samples represent six Eunicida species (three species of genus *Palola*), three Nereididae species and two Syllidae species (Figure 5.1). Two inferred species, *Palola* A3 and *Palola* A9, were previously observed by Schulze (2006). Five inferred polychaete species, *Palola* AX1, Eunicida 1, Eunicida 2, Eunicida 3 and Nereididae 1, were found in studies of diets of *C. ebraeus* adults at Guam and American Samoa (see Chapter 4). Nereididae 1 is most frequently preyed upon, followed by *Palola* species (see sample sizes in Figure 5.2A). Three species have never been

found in studies of diets of *C. ebraeus*: Syllidae 1, Syllidae 2, Nereididae 4 and Nereididae 5 (Figure 5.1), but were rarely targeted (Figure 5.2A).

2. Ontogenetic shifts of diet

Several of the prey species are consumed by a wide size range of *C. ebraeus* (Figure 5.2A). These include the dominant prey types (Nereididae and *Palola* species). Nereididae species are predominantly consumed by mid-size individuals (around 11-17 mm in shell lengths), while *Palola* species are hunted mostly by adults (larger than 17mm) (Figure 5.2A). The rare prey species Eunicida 1, 2 and 3 and Syllidae 1 and 2 are mostly consumed by small individuals (smaller than 11mm in shell lengths) (Figure 5.2A). The heterogeneity of the size distribution of individuals consuming each type of prey is still prominent when the pattern is examined at the inferred generic level of prey (Figure 5.2B), but disappears when we evaluated the prey categories at high prey taxonomic categories (Figure 5.2C). One-way ANOVA analyses demonstrate that these prey species and higher taxonomic levels are targeted by *C. ebraeus* individuals of significantly different shell lengths: *P*-value for groups divided by prey species (Figure 5.2A) is 0.011; for groups divided by higher taxonomic levels (Figure 5.2B), *P*-value is less than 0.0001. Even though this pattern of difference in size ranges of individuals feeding on each type of prey is determined from the 54 individuals with known prey species, initial determination of the 151 fecal samples to putative order/genus of annelids with the microscopic examination approach yielded a consistent pattern (Figure 5.3) as shown in Figure 5.2B.

The diversity of prey items utilized differs among classes of individuals representing different developmental stages. Sliding window analyses of several variables of dietary diversities within

certain size ranges of *C. ebraeus* individuals revealed a pattern of an initial decrease followed by an increase in dietary diversity (Figure 5.4). Small individuals (less than 11mm in shell lengths) exhibit the broadest dietary spectrum, medium sized ones (11-17mm) specialize on Nereididae species, and large ones (larger than 17mm) prey on both Nereididae and *Palola* species (Figure 5.4; Figure 5.5A). Medium individuals defined based on shifts in diets (shell lengths of 11-17mm) exhibit significantly different diet in comparison to individuals of other size ranges, as illustrated by sliding window analyses (window size of 5mm in shell lengths) of pairwise D_{ST} values of the 16S gene sequences of fecal samples (Figure 5.5B) and Fisher's exact tests of prey species among the three size classes (Table 5.2). Even based on traditional classification of size classes of *Conus* individuals (juvenile:<10mm, subadults: 10-20mm, adults: >20mm), we detected the lowest dietary diversity for subadults.

3. Lack of population structure at locus E1

We successfully obtained genotypic information of locus E1 from venom duct cDNA of 49 *C. ebraeus* individuals. These genotypes represent combinations of six alleles (E1a, E1b, E1bii, E1c, E1g and E1h) that are identical to alleles described by Duda et al. (2009) (also see Chapter 4; GenBank accession numbers FJ804530-FJ804532, FJ804535, FJ804536 and FJ834437). Estimation of pairwise Φ_{ST} values does not reveal any structure at this locus among small (<=10mm), medium (10-20mm) and large individuals (>20mm), with non-significant Φ_{ST} values close to zero (Table 5.3). Sliding window analyses (of 5mm shell lengths) did not reveal any structure either (very small Φ_{ST} values non-significant from 0). Allelic composition of each size class is significantly different from that of the Hawaiian population (Table 5.3). Allelic

differences between individuals of different sizes and adults at American Samoa are close to significance based on the 5% significance level, and Φ_{ST} values are fairly small (Table 5.3).

4. Ontogenetic shifts of conotoxin gene expression

Initial exploration of conotoxin gene expression levels with the heatmap approach revealed that two individuals do not show any evidence of expression at the six conotoxin genes (C_T values are undertermined in Real-time PCRs). Another sample did not show any evidence of expression at the β-tubulin locus (C_T value is 'undetermined'). These results imply that cDNA samples of these individuals were of bad quality or that some of these specific reactions failed and so we eliminated them from the subsequent analyses.

Individuals of different sizes examined exhibit large variation in expression at the six conotoxin loci (Figure 5.6). Expression levels reach a maximum in individuals with shell lengths of 9-13mm and then gradually decrease in larger ones. Hierarchical clustering analyses divided individuals into two major groups (Figure 5.7A): a cluster (cluster 1) with individuals of a relatively even size distribution (Figure 5.7B) and a cluster (cluster 2) composed exclusively of individuals with extreme sizes (small and large) (Figure 5.7C). Classification of individuals with the non-hierarchical PAM method also produced similar results, with large difference in size distributions of individuals between two clusters (Figure 5.8). Members of cluster 2 identified by both hierarchical and non-hierarchical approaches mostly exhibit higher levels of conotoxin gene expression than those of cluster 1, a pattern that implies that average expression levels among small and large individuals are higher than those of medium ones.

As revealed from results of clustering analyses, average expression levels of all conotoxin genes (except locus EA1) undergo an initial decrease and then increase process through development (Figure 5.9). Expression levels of these genes reaches are highest in small individuals and the lowest in medium size individuals, and then begin to increase in adults (but do not reach the same levels as in small individuals; Figure 5.9). This general pattern of variation in expression is unlikely the consequence of variation in sample sizes of individuals of different shell lengths, because simulations demonstrated the robustness of this trend of variation of expression, especially the decrease among juveniles and subadults (Figure 5.10). Relative levels and patterns of expression also differ among loci: these loci occupy different color spectrum in the heatmap of absolute $-\Delta C_T$ values (Figure 5.6A), and the column dendogram separates loci EA1 and ED20 from the other loci (Figure 5.6). Average expression levels at locus EA1 decrease in medium individuals, but does not show the same increasing pattern in adults as exhibited in the other genes (Figure 5.9).

5. Ontogenetic shifts in conotoxin gene expression and diet

Conotoxin gene expression levels are not directly associated with diets. A total of 35 individuals with diet data were used. Direct comparison of expression levels among groups of individuals assembled based on identification of prey species with boxplots, heatmaps and one-way ANOVA approaches did not reveal any significant differences or trends, except for locus ED20 (*P*-value= 0.006) (Figure 5.11 and 5.13). No significant differences in conotoxin gene expression levels (including locus ED20) were detected among groups determined based on higher taxonomic levels of their prey (Figure 5.12). The hierarchical clustering approach divided these

individuals into two major clusters that exhibit no significant difference in prey utilization (*P*-value of the Fisher's exact test is 0.270) (Figure 5.14).

Patterns of variation in conotoxin gene expression and dietary diversities through development are in the shape of a semi-upward parabola (a trend of decrease and then increase) along the gradient of shell lengths, but their timings of changes are not coincident (Figure 5.15). Dietary changes in ontogeny lead changes in conotoxin gene expression by the amount of time equivalent to the growth time of one or two mm in shell lengths (Figure 5.15). Such a pattern is confirmed by the significantly positive correlation coefficients in cross-correlation analyses and linear regression, analyses applied to determine if dietary diversity of individuals is correlated with conotoxin gene expression levels of individuals one or two mm larger (Figure 5.16 and 5.17). Locus EA1 is an exception. The down-regulation of this locus in medium individuals is eminent, but expression levels remain minimal among mature adults (Figure 5.13 and 5.15B). Though regulation of locus EA1 seems to lag changes in dietary diversity, we did not detect any significantly positive correlations between these two series (Figure 5.16B).

6. Relationship of dietary composition and venom expression with sexual maturity

The smallest individual among our samples that exhibits a discernible penis has a shell length of 12mm. We thus considered all individuals larger than this size as sexually mature adults, and determined individuals larger than 12mm and without a penis as mature females. Individuals smaller than 12mm in shell lengths were considered to be immature juveniles. Among 37 individuals with records of sexual identification (i.e. with or without a penis) and diet, we determined 17 individuals as mature females, 5 as mature males and the rest as immature

juveniles. Fisher's exact tests revealed no significant differences in dietary compositions among groups classified by sex and the stage of development (mature males vs mature females: *P*-value=0.484; mature males vs juveniles: *P*-value=0.214; mature females vs juveniles: *P*-value=0.524). A total of 57 individuals possess records of sex identification and conotoxin gene expression levels, which include 24 mature females, 6 mature males and 27 immature juveniles (Figure 5.18A). Levels of expression of conotoxin genes are highly variable within each group, but do not differ among juveniles, mature males and females (Figure 5.18), a result that is supported by non-significant results of t-tests or ANOVA after correction for multiple tests.

Discussion

As the first study to assay the association between conotoxin gene expression and dietary specialization through development, we discovered an intriguing relationship between conotoxin gene regulation and dietary shifts. Large variation in levels of expression among individuals of each size class reveals a lack of canalization of conotoxin gene regulation within developmental stages. Conotoxin gene expression appears to be related to changes in dietary breadth rather than types of prey, and dietary shifts occur before shifts in levels of expression of these genes. Our results demonstrate that regulation of conotoxin genes does in fact change during development, and these changes may be affected by changes in resource utilization.

One explanation for variation in expression levels among different size classes of individuals is that these size classes represent genetically distinct cohorts from different locations. Nonetheless, *C. ebraeus* individuals sampled here appear to represent a local population rather than a mixture of cohorts of local and immigrant individuals. Populations of *C. ebraeus* in the Indo-West Pacific

appear to be panmictic based on analyses of sequences of a region of the mitochondrial COI gene (Duda et al. 2012), but population at Hawaii shows significant differences in allelic composition at locus E1 from other populations in Indo-West Pacfiic (Duda et al. 2009)(also see Chapter 4). Though planktonic larvae of *C. ebraeus* can make long-distance dispersal (Kohn and Perron 1994), lack of structure at locus E1 among individuals of different sizes and adults at Hawaii (and possibly American Samoa) implies that these individuals unlikely include migrants from Hawaii or American Samoa (Table 5.3); instead, they are more likely to be recruited locally at Guam.

1. Ontogenetic shifts of dietary specializations

C. ebraeus individuals at Guam undergo several dietary shifts through development: from trophic generalists to specialists and then to generalists (Figure 5.4). This contrasts with the general pattern of dietary shifts suggested by Levinten (1976). Contrary to his conclusion that small individuals (<10mm in shell lengths) are trophic specialists, juveniles of *C. ebraeus* at Guam possess the most broad dietary breadth that spans more than three families representing two orders of annelids. The phenomenon that subadults specializing on one type of prey for a limited time (growth time of about 5mm of shell lengths) contradicts the previous notion that medium individuals are trophic generalists. Individuals larger than 17mm in shell length tend to regain an increased dietary breadth but do not feed on polychaete species that are consumed by small individuals (e.g. Syllidae and some Eunicid species). We were not able to evaluate the dietary specialization of individuals larger than 30mm in shell lengths because of lack of samples in this size range. But it is more likely for large adults to prey on multiple species of the genus

Palola than to specialize on a specific species if the trend of increase of dietary breadth observed in adults still holds as snails grow bigger (Figure 5.4 and 5.5).

The discrepancy of our results with previous observations may be affected by temporal variation in prey availability and/or the different experimental approaches utilized. Compared to the traditional approach of prey identification solely based on microscopic examination of gut contents, the DNA barcoding approach used here improves accuracy of species identification and provides additional phylogenetic information for the incomplete taxonomical records of annelids, especially when the pattern of dietary shifts is only prominent at the species/genera level of prey (Figure 5.2). Differences in strategies of resource utilization employed by individuals of different sizes/ages can be triggered by both external factors such as intraspecific competition, microhabitat differentiation among developmental stages, and variation in body volume of different polychaete species. Intrinsic factors associated with feeding efficiency, such as development of radular teeth and venom potency, can affect their ability to subdue certain types of prey as well.

2. Ontogenetic changes of conotoxin gene expression

C. ebraeus individuals exhibit extensive variability in expression levels of six conotoxin genes relative to the estimated expression level of the β-tubulin gene (Figure 5.6). The relative expression levels of conotoxin genes among individuals differ at up to six orders of magnitude: from about 1/1000 to 1000-fold relative to the β-tubulin gene (Figure 5.6). For each conotoxin gene, variation in relative expression levels among individuals is on the magnitude of about 10^4 to $3x10^5$. But standard deviations of C_T values of these genes among triplicate samples of the

same individual can be as high as 6, which equates to about 64-fold (2^6) variation in expression levels within an individual. Thus it is possible that the 10^4 -10^5 magnitudes of differences in relative expression represent experimental noise, and the real fold-changes are not so extreme.

Sliding-window/moving average analyses smoothed out variation in expression levels of these genes and revealed a prominent pattern of change in the regulation of these genes during development. Conotoxin genes are more frequently up-regulated in small and large individuals of *C. ebraeus* than in medium sized individuals (Figure 5.7 and 5.8). Small ones seem to possess the most diverse expressed genes and the highest abundance of conotoxin gene transcripts. Expression levels in mature adults are not comparable to those in small individuals, even though five out of the six genes are upregulated in adults relative to the medium sized individuals (Figure 5.6 and 5.9).

Though the six conotoxin genes are all up-regulated among juveniles and down-regulated by subadults, loci ED20 and EA1 are almost exclusively expressed at high levels by small individuals. Among medium and large individuals, expression levels of locus ED20 relative to the endogenous gene can be as low as 0.001, and locus EA1 exhibits large decreases in levels of expression. This suggests that these two genes are exclusively up-regulated during early stages of development (Figure 5.6A). Admittedly there may be other conotoxin genes that are exclusively expressed or up-regulated in every stage of development, because each *Conus* individual can produce 100-200 different conotoxins (Olivera 2002) and we only investigated patterns of expression of six genes. The vast variation of venom compositions within species of *Conus* observed previously (Jakubowski et al. 2005; Davis, Jones, and Lewis 2009; Rivera-Ortiz, Cano,

and Mar í2011) may stem from the temporal and intraspecific variation in conotoxin gene expression, if abundance of mature toxins in the venom is positively correlated with abundance of conotoxin gene transcripts rather than accumulation of gene products through time.

The inconsistency in patterns and relative levels of expression among the six conotoxin genes in this study implies that these genes possess independent regulatory mechanisms (Figure 5.6 and 5.13). New conotoxin genes emerge from gene duplication (Duda and Palumbi 1999), but it is unclear if paralogs are regulated in a consistent manner. Variation in expression of different conotoxin genes within the same individual and patterns of ontogenetic shifts of expression of these genes demonstrated that conotoxin genes are regulated differentially, even for members of the same gene family.

3. Association between conotoxin gene expression and diets

Regulation of conotoxin gene expression does not appear to show a direct relationship with types of prey that are consumed by individuals of different size classes. Conotoxin gene expression levels are not significantly associated with specific prey categories, nor there is any linear relationship between them (Figure 5.11 and 5.12). We did not detect any significant difference in dietary composition or expression levels of six conotoxin genes among juveniles, mature females and males, implying that neither sexual maturity nor gender (among adults) plays a role in ontogenetic shifts of conotoxin gene expression and diet (Figure 5.8).

Average expression levels instead are significantly positively correlated with shifts in dietary diversity, with changes in conotoxin gene expression levels lagging behind dietary changes (Figure 5.16 and 5.17). Frank (1969) found that growth rates of *C. miliaris* exhibit a logarithmic

relationship with shell lengths; during the first year of growth, shell length (up to 15mm) is a linear approximation of growth rates. The lead of dietary shifts over changes of conotoxin expression reported here (i.e. 1-2mm of shell lengths) represents about 25 to 50 days of growth time if we assume growth rates of *C. ebraeus* and *C. miliaris* are identical. Nevertheless, increases in shell sizes of *Conus* species can be abrupt, most likely related to recent feeding bouts (personal communication by Alan Kohn). Thus the difference in timing of dietary shifts and conotoxin gene regulation may be negligible.

It is unclear if such a correlation is a mere coincidence or signifies the plasticity of conotoxin gene expression. The ontogenetic shift of conotoxing gene expression is affected by changes in dietary breadth but does not respond immediately to these changes. The up-and-down regulation of conotoxin genes may be facultatively evoked by prey, and this 'adjusting with prior experience' strategy cannot be verified with our approach of quantifying expression and diets in the same sample concurrently. Environmental-induced morphological variation often exhibits some time delay relative to environmental changes (Palumbi 1984; Alstyne 1988; Padilla and Adolph 1996; Starck 1999), and this phenotypic plasticity is only advantageous when the delay is small (Padilla and Adolph 1996). The induced morphological variation may result from gene expression (Landry et al. 2006; Lopez-Maury, Marguerat, and Bahler 2008; Richter, Haslbeck, and Buchner 2010; Yampolsky, Glazko, and Fry 2012), and timings of gene regulation differ among genes and organisms. For example, increase of expression of heat-shock protein genes in yeasts that deter the heat shock occurs almost immediately after heat exposure, but regulation other genes that are involved in adaptation lag behind (Richter, Haslbeck, and Buchner 2010). To understand plasticity of conotoxin gene expression and its role in delay of phenotypic changes

relative to dietary shifts, it is important to assay the regulatory mechanisms of conotoxin genes and the signaling pathways used by venom duct cells, which are still unknown. However, we cannot rule out the possibility that shifts of conotoxin gene expression through development represent a systematic process rather than being plastic with dietary changes.

We postulate that high toxicity and diverse venom composition are beneficial to capture more diverse prey by small individuals, while venoms are fine-tuned for more specialized diets among mid-size and large individuals. This is very similar to the pattern of venom ontogeny observed in snakes: higher toxicity of venoms of juveniles ensures immobility of prey, while venoms produced by adult snakes are more associated with predigestive functions (Andrade and Abe 1999; Saldarriaga et al. 2003; Mackessy et al. 2006). We did not detect any strict canalization of conotoxin gene expression in the developmental process. Seasonality in prey availability and prey choice may affect regulation of these genes within each size group. For example, Gibbs et al. (2011) performed proteomic analyses of venom milked from juvenile and adult rattlesnakes feeding on different prey, and detected larger variation of venomous peptides among adults than in juveniles. Though our study detected an opposite pattern of variation in conotoxin gene expression, experimental manipulation of these snails may reveal more information about the plasticity of gene expression.

4. Relationships between conotoxin diversification and prey utilization

Results from Chapters 4 and 5 revealed that conotoxin gene evolution and expression are highly associated with dietary specialization. Geographic variation of conotoxin genes is driven by heterogeneity in types and extensity of selection among locations, which likely stems from

geographic differentiation of dietary compositions. Shifts in conotoxin gene expression through development are likely evoked by changes of prey diversity. Genetic diversity and expression of conotoxin genes are in a positive frequency-dependent relationship with prey species, i.e. higher diversity of prey species instigates higher allelic/nucleotide diversities and up-regulation of conotoxin genes. Here we propose a coevolutionary relationship (bidirectional adaptations) between the functional specificity of conotoxins and the defense mechanism of prey species. For example, genes encoding sodium channels, targets of μ- and δ-conotoxins, vary among animal groups (Zakon 2012). Invertebrates generally possess two sodium channel genes while in vertebrates many copies of sodium channel genes emerge from duplications (Widmark et al. 2011; Zakon 2012), but sodium channel genes are fairly variable among invertebrate species (unpublished data in Duda Lab). Moreover, evolution of sodium channels in garter snakes and pufferfishes is adaptive to levels of tetrodotoxin (Geffeney et al. 2005; Venkatesh et al. 2005; Jost et al. 2008). Therefore, it is highly possible that genes encoding ion channels and neuronal receptors are highly variable among prey species and are evolving in response to changes of venom diversity and toxicity in *Conus*. On the other hand, function of conotoxin genes is highly sensitive to non-synonymous mutations: replacement of a single amino acid in the mature toxin potentially affects the specificity and binding efficiency of conotoxins (Terlau and Olivera 2004; Dutertre et al. 2007; Whiteaker et al. 2007). Driven by the increase of diversity of conotoxin targets in the prey, *Conus* species refine their venom through extensive gene duplication, rapid evolution and frequent regulation of gene expression to increase the efficacy of predation. The possibility and patterns of coevolution between *Conus* species and their prey require further ecological, functional, genetic and evolutionary studies of conotoxins and their targets.

Acknowledgements

We thank Alex Kerr, Susanne Wilkins, Marielle Terbio, Carmen Kautz and Kirstie Goodman-Randall at University of Guam Marine Lab for their assistance with field collections. We acknowledge Jincong Tao in the Molecular Core Lab at University of Michigan School of Dentistry for his help with the Real-time PCR. We thank the Department of Fishery and Wildlife Sciences at Guam for the collection permits, and the Department of Fishery and Wildlife Sciences at Hawaii for the custom clearance. This project is funded by two Hinsdale-Walker Fellowships awarded to D. Chang by University of Michigan Museum of Zoology.

Figure 5.1. Bayesian consensus phylogenies of mitochondrial 16S gene sequences of fecal samples of *C. ebraeus* and known polychaete species.
Bayesian posterior probabilities are labeled at nodes of major clades. Sequences downloaded from GenBank are labeled with their respective accession numbers. Sequences obtained in this study are highlighted in bold. Putative prey species are labeled in blue next to the sequence name.

(A) Phylogeny of species of Order Eunicida with GTR (Tavaré 1986) +I+G model.

(B) phylogeny of species of Family Nereididae with GTR+G model.

(C) phylogeny of species of Family Syllidae with GTR+G model.

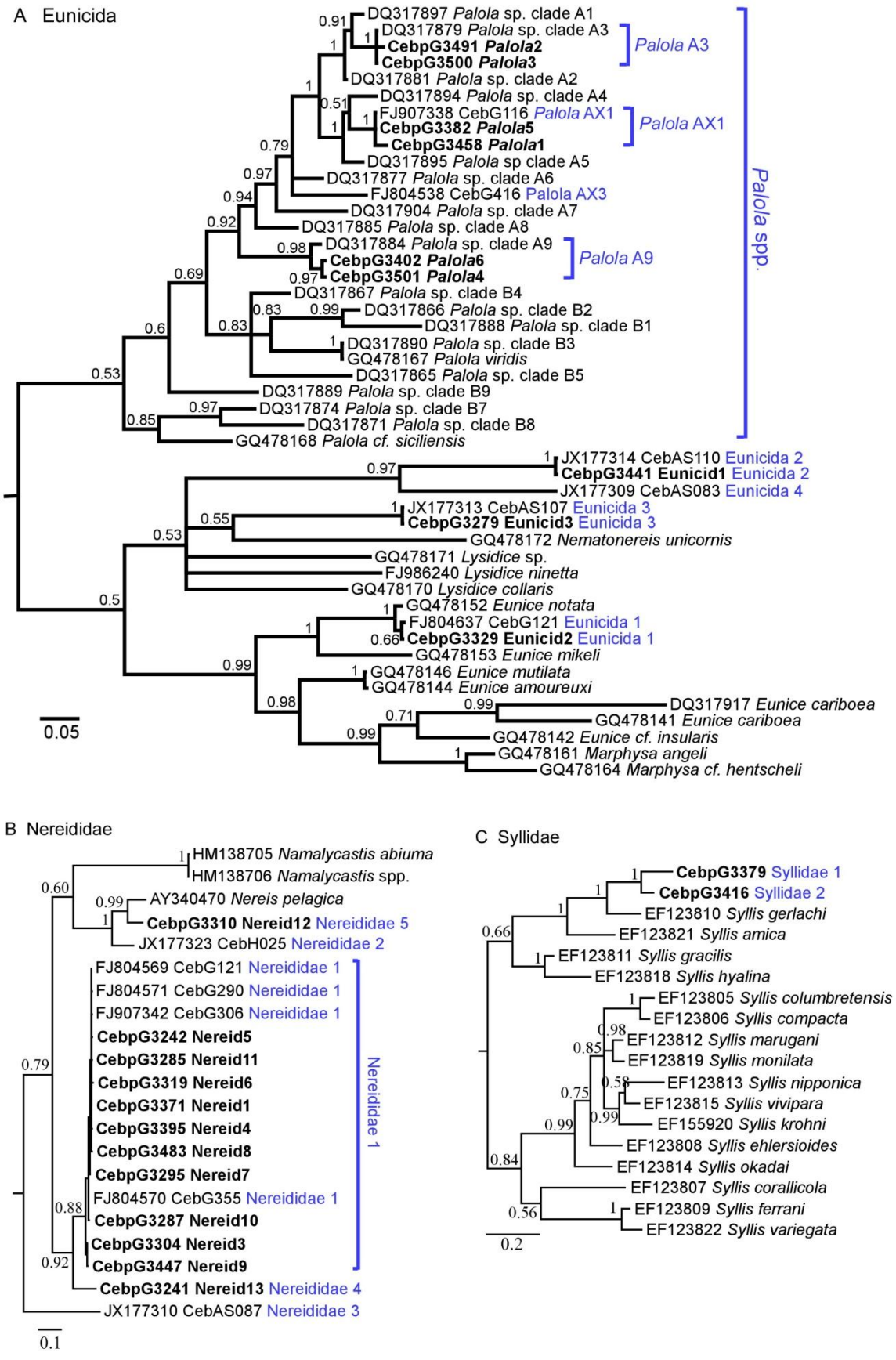

A Eunicida
DQ317897 Palola sp. clade A1
DQ317879 Palola sp. clade A3
CebpG3491 Palola2
CebpG3500 Palola3
Palola A3
DQ317881 Palola sp. clade A2
DQ317894 Palola sp. clade A4
FJ907338 CebG116 Palola AX1
CebpG3382 Palola5
CebpG3458 Palola1
Palola AX1
DQ317895 Palola sp clade A5
DQ317877 Palola sp. clade A6
FJ804538 CebG416 Palola AX3
DQ317904 Palola sp. clade A7
DQ317885 Palola sp. clade A8
DQ317884 Palola sp. clade A9
CebpG3402 Palola6
CebpG3501 Palola4
Palola A9
Palola spp.
DQ317867 Palola sp. clade B4
DQ317866 Palola sp. clade B2
DQ317888 Palola sp. clade B1
DQ317890 Palola sp. clade B3
GQ478167 Palola viridis
DQ317865 Palola sp. clade B5
DQ317889 Palola sp. clade B9
DQ317874 Palola sp. clade B7
DQ317871 Palola sp. clade B8
GQ478168 Palola cf. siciliensis
JX177314 CebAS110 Eunicida 2
CebpG3441 Eunicid1 Eunicida 2
JX177309 CebAS083 Eunicida 4
JX177313 CebAS107 Eunicida 3
CebpG3279 Eunicid3 Eunicida 3
GQ478172 Nematonereis unicornis
GQ478171 Lysidice sp.
FJ986240 Lysidice ninetta
GQ478170 Lysidice collaris
GQ478152 Eunice notata
FJ804637 CebG121 Eunicida 1
CebpG3329 Eunicid2 Eunicida 1
GQ478153 Eunice mikeli
GQ478146 Eunice mutilata
GQ478144 Eunice amoureuxi
DQ317917 Eunice cariboea
GQ478141 Eunice cariboea
GQ478142 Eunice cf. insularis
GQ478161 Marphysa angeli
GQ478164 Marphysa cf. hentscheli
0.05
B Nereididae
HM138705 Namalycastis abiuma
HM138706 Namalycastis spp.
AY340470 Nereis pelagica
CebpG3310 Nereid12 Nereididae 5
JX177323 CebH025 Nereididae 2
FJ804569 CebG121 Nereididae 1
FJ804571 CebG290 Nereididae 1
FJ907342 CebG306 Nereididae 1
CebpG3242 Nereid5
CebpG3285 Nereid11
CebpG3319 Nereid6
CebpG3371 Nereid1
CebpG3395 Nereid4
CebpG3483 Nereid8
CebpG3295 Nereid7
FJ804570 CebG355 Nereididae 1
CebpG3287 Nereid10
CebpG3304 Nereid3
CebpG3447 Nereid9
Nereididae 1
CebpG3241 Nereid13 Nereididae 4
JX177310 CebAS087 Nereididae 3
0.1
C Syllidae
CebpG3379 Syllidae 1
CebpG3416 Syllidae 2
EF123810 Syllis gerlachi
EF123821 Syllis amica
EF123811 Syllis gracilis
EF123818 Syllis hyalina
EF123805 Syllis columbretensis
EF123806 Syllis compacta
EF123812 Syllis marugani
EF123819 Syllis monilata
EF123813 Syllis nipponica
EF123815 Syllis vivipara
EF155920 Syllis krohni
EF123808 Syllis ehlersioides
EF123814 Syllis okadai
EF123807 Syllis corallicola
EF123809 Syllis ferrani
EF123822 Syllis variegata
0.2

Figure 5.2. Boxplots of ranges of shell lengths of *C. ebraeus* individuals consuming different types of prey.

(A-C) Prey categories include annelid (A) species, (B) genera and (C) orders. Types of prey are labeled on the X-axis. n: sample size.

(D) Size distribution of *C. ebraeus*: overall distribution of shell lengths of individuals with known diet.

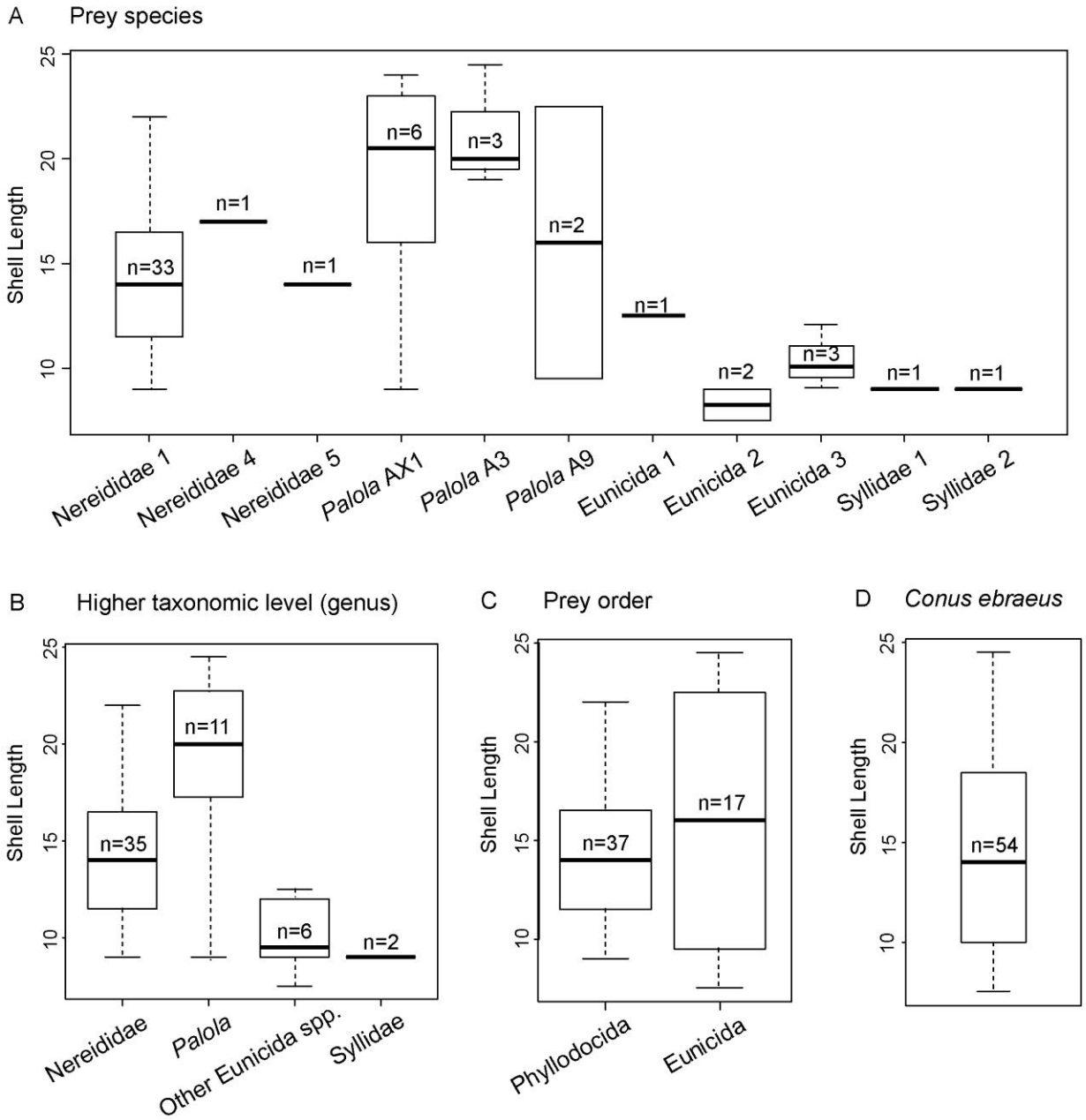


Figure 5.3. Boxplots of ranges of shell lengths of *C. ebraeus* individuals consuming different types of prey that are identified with microscopic examination of fecal samples.

(A) Prey identified at the genus level with the approach of microscopic examination. Types of prey are labeled on the X-axis. n: sample size.

(B) Size distribution of *C. ebraeus*: overall distribution of shell lengths of individuals with known diets.

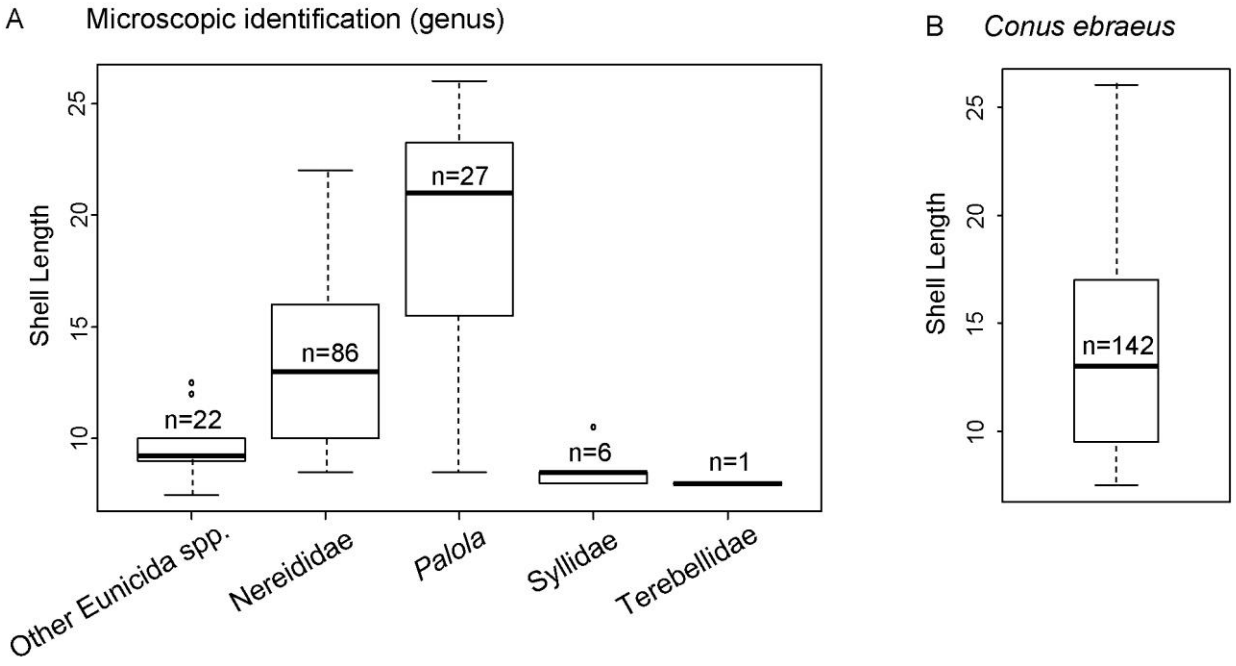


Figure 5.4. Sliding window analyses of dietary diversity.

(A-F) (A-B) Shannon's index, (C-D) Simpson's index, (E-F) average genetic distance (GD) of prey compositions. X-axis is average shell lengths of *C. ebraeus* individuals in the sliding-window analyses. Gray dashed lines are the fitted curves of polynomial regression.

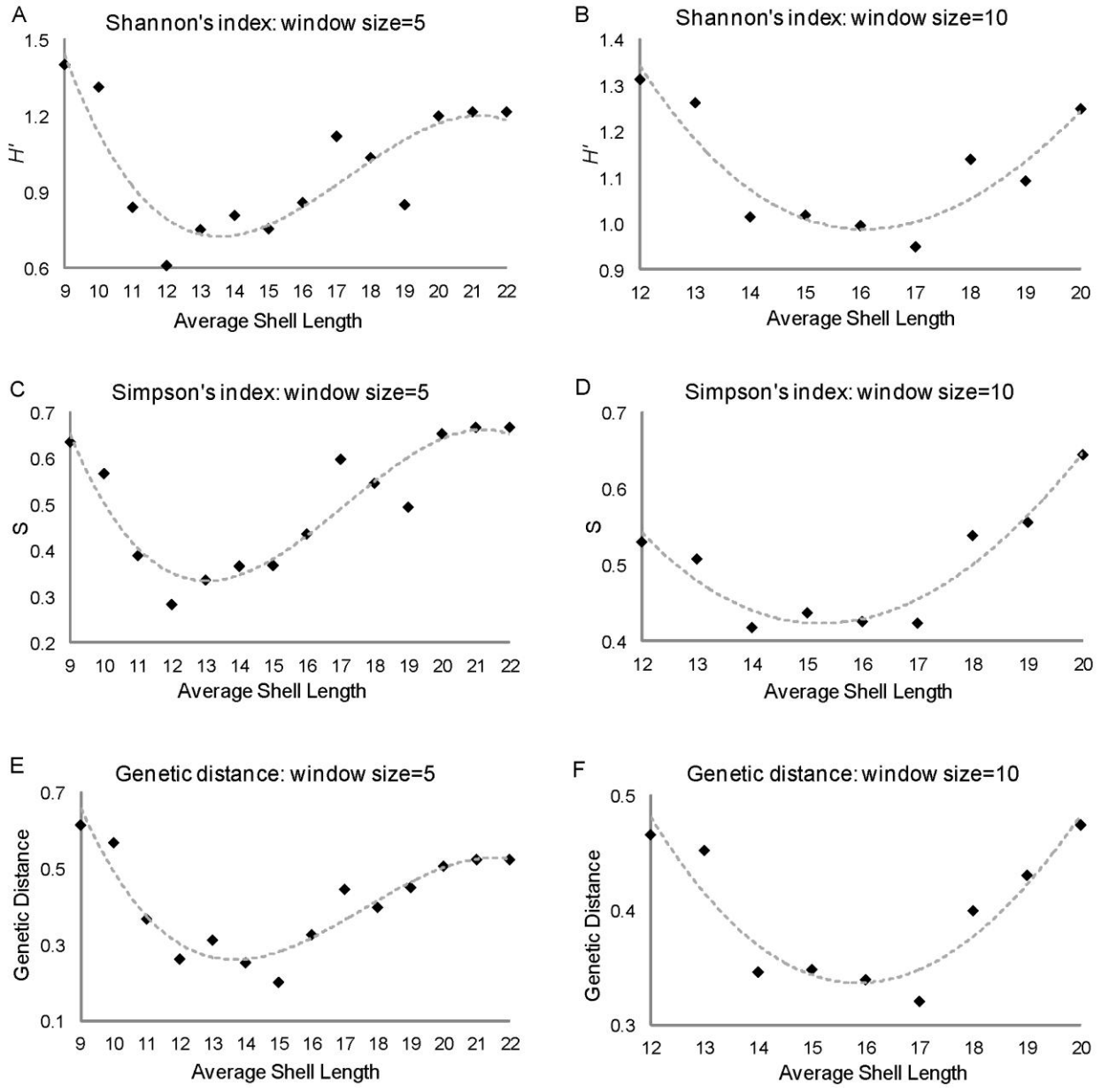


Figure 5.5. Heatmaps of dietary ontogeny of *C. ebraeus* individuals.

(A) Heat map of frequencies of prey species consumed by *C. ebraeus* individuals of the same size (shell lengths rounded to integers).

(B) Heat map of pairwise D_{ST} values calculated with the Tamura-Nei model among size classes of sliding-window analyses (window size=5). *P*-values are estimated with simulations of 10,100 replicates, and significant results (*P*-value<0.05) are labeled with asterisks in the cells.

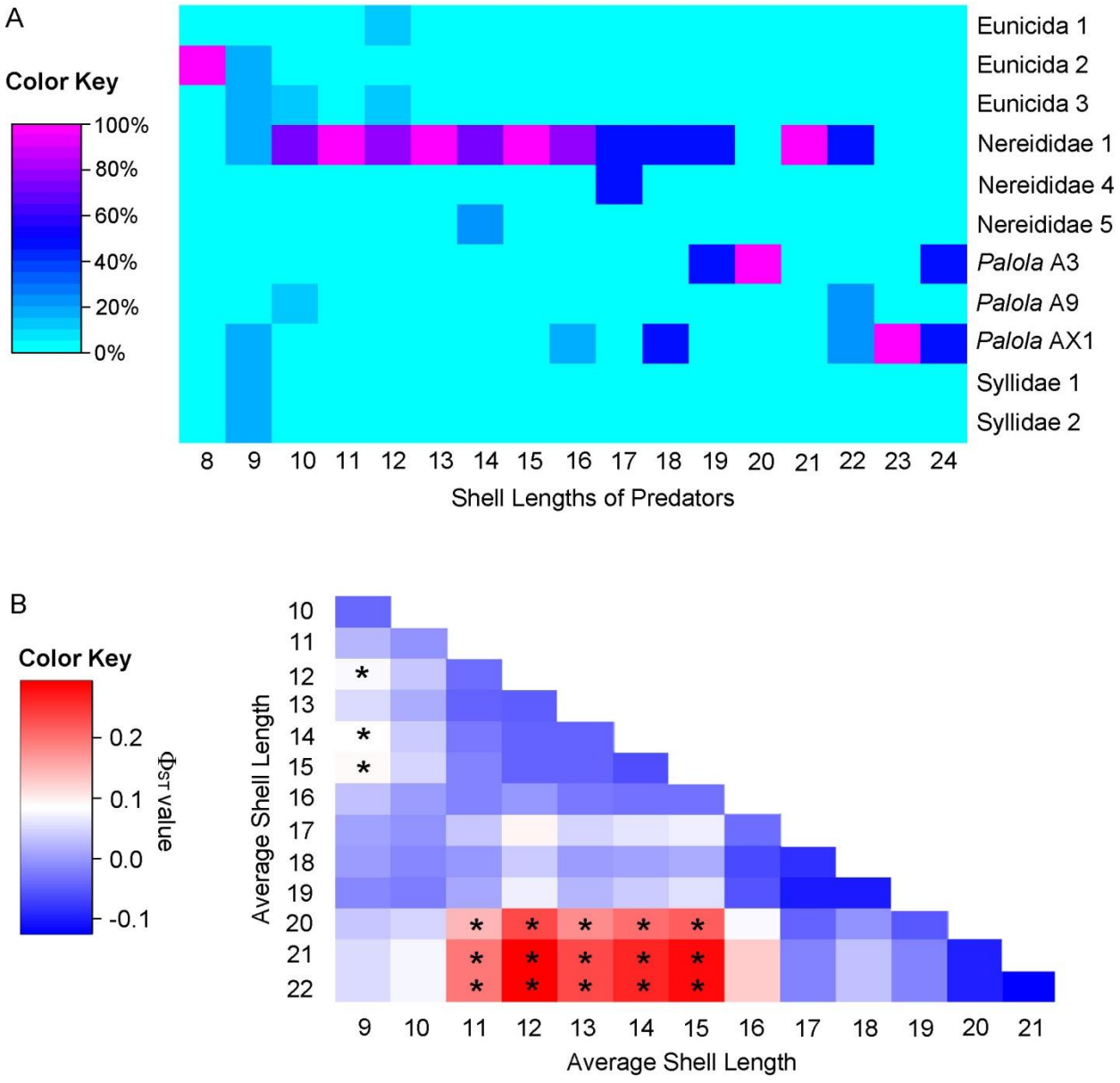


Figure 5.6. Heatmaps of relative expression levels of six conotoxin genes among individuals. Row variables are individuals with different shell lengths, and column variables are six conotoxin loci. The column dendogram is illustrated above the column labels.

(A-B) Heatmaps with inputs of each cell as (A) absolute $-\Delta C_T$ values between the specific conotoxin locus and the β-tubulin locus and (B) centered and standardized $-\Delta C_T$ values (Z-score) for each conotoxin locus.

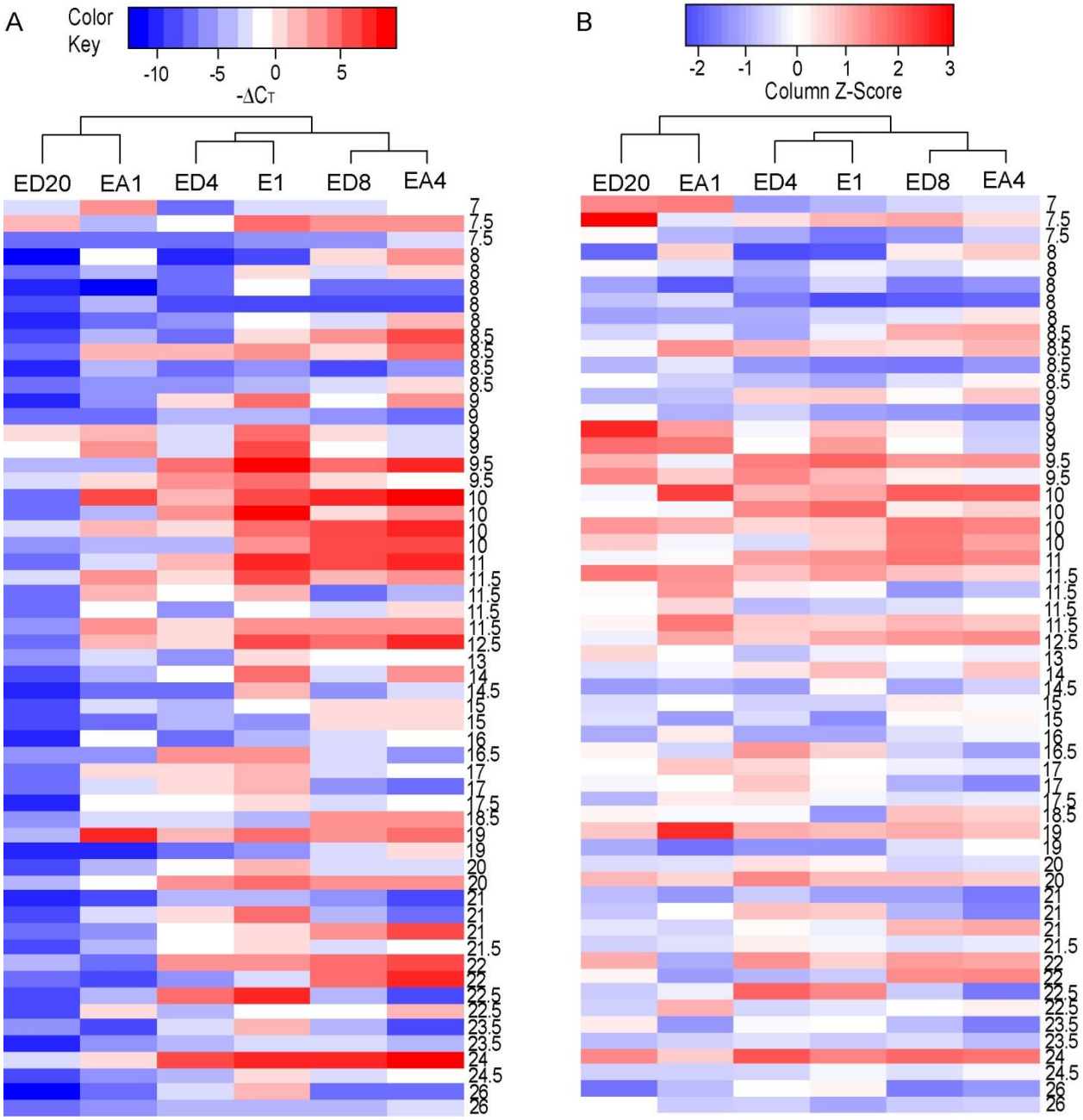


Figure 5.7. Hierarchical clusters of *C. ebraeus* individuals based on relative levels of expression of the six conotoxin genes.

(A) Dendogram of clusters classified with Wald's method. The two major clusters are labeled at the top nodes.

(B) Size frequency distribution of individuals within the first cluster defined in (A).

(C) Size frequency distribution of individuals within the second cluster defined in (A).

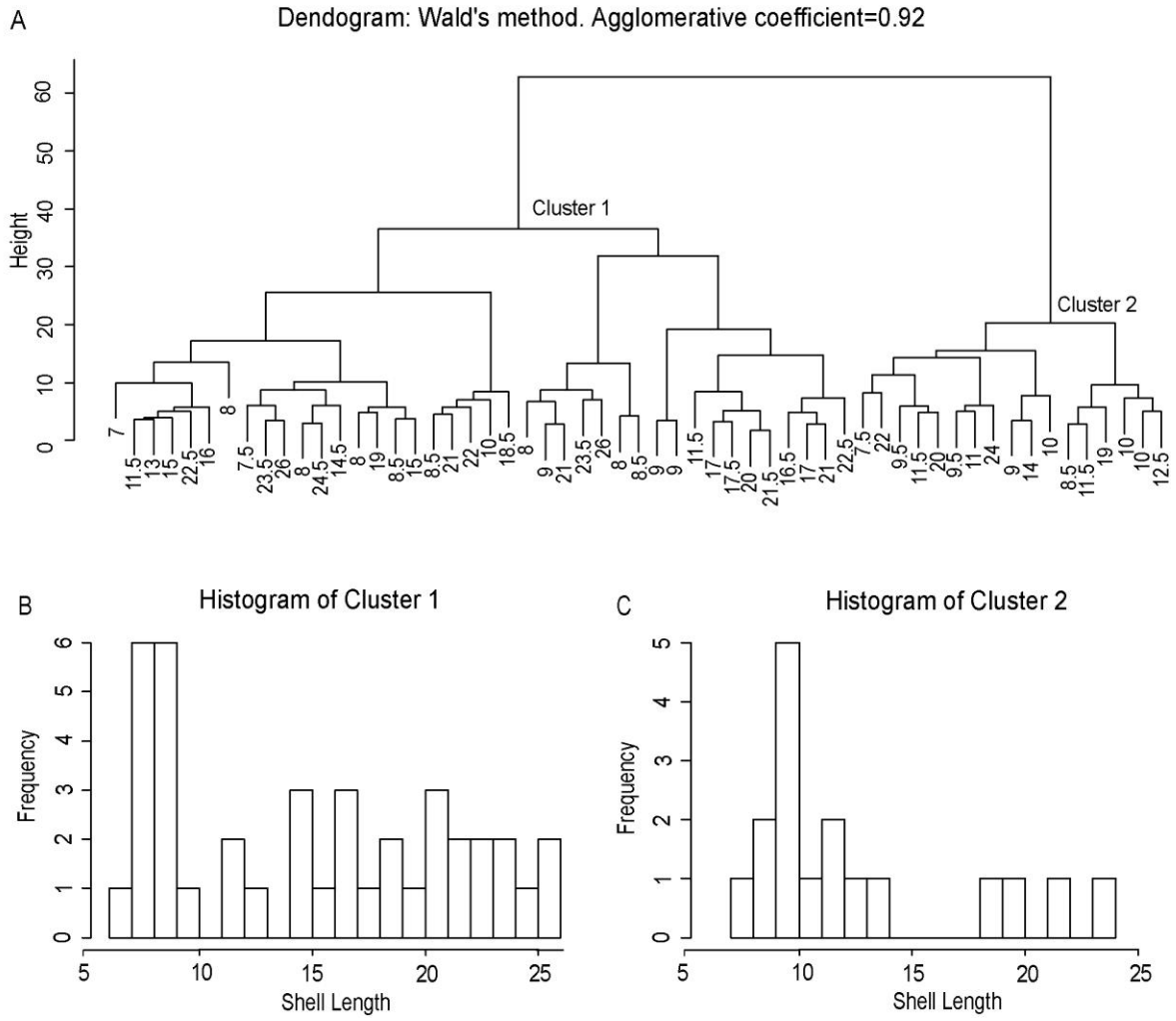


Figure 5.8. Clustering results with the PAM method.
$-\Delta C_T$ of all six conotoxin genes were used for the analysis after removal of three outliers.

(A-B) Patterns of distribution of individuals in (A) two clusters and (B) three clusters. Each cluster is represented by different symbols and circled by blue curves. The first two components explained 74.03% of variability.

(C-D) Size frequency distributions of the two clusters defined in panel A: (C) cluster 1, (D) cluster 2.

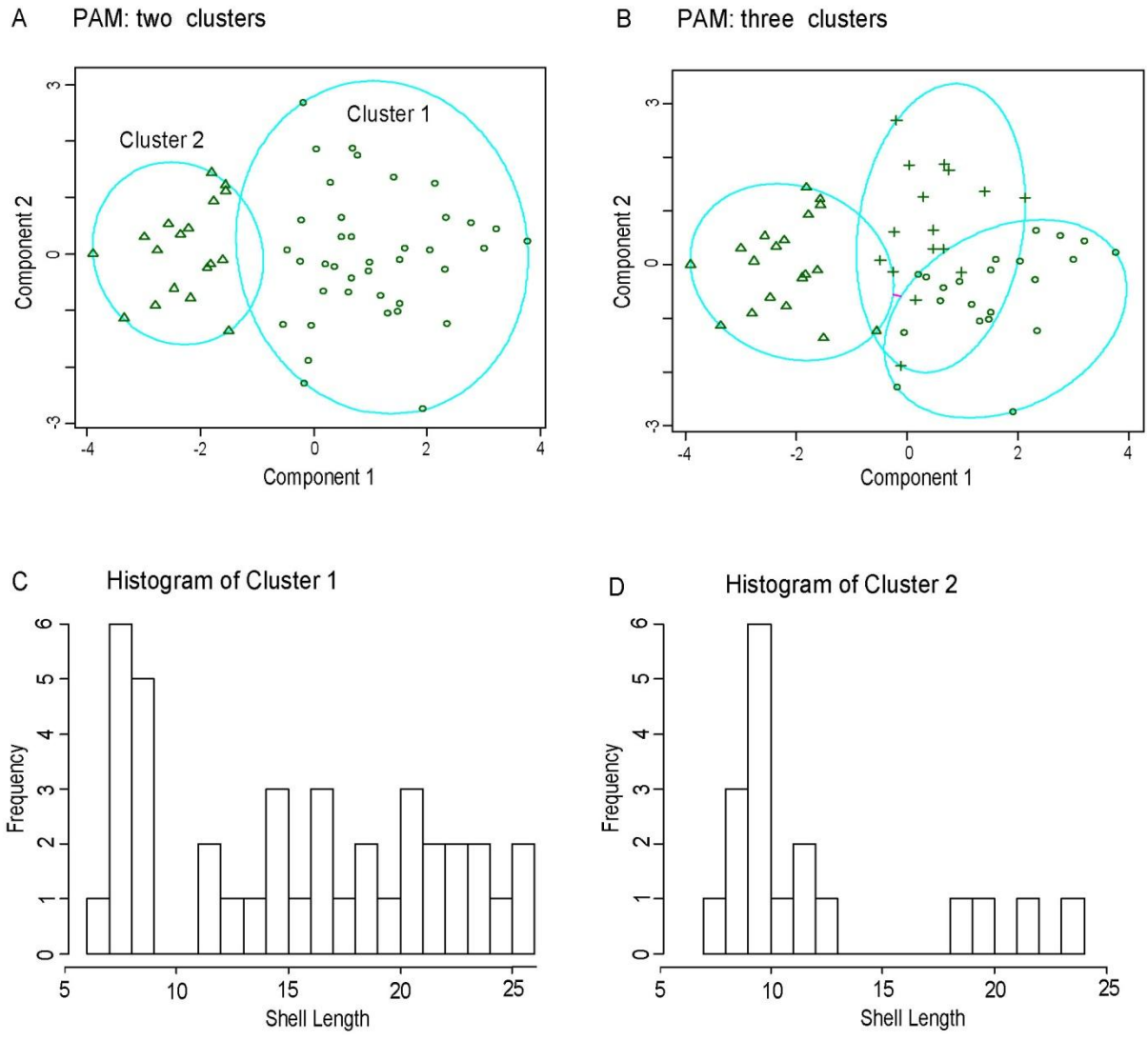


Figure 5.9. Moving averages of relative expression levels ($-\Delta C_T$) of the six conotoxin genes among individuals of different shell lengths (window size=5mm in shell lengths).

(A-F) (A) Locus E1, (B) locus EA1, (C) locus EA4, (D) locus ED4, (E) locus ED8, and (F) locus ED20.

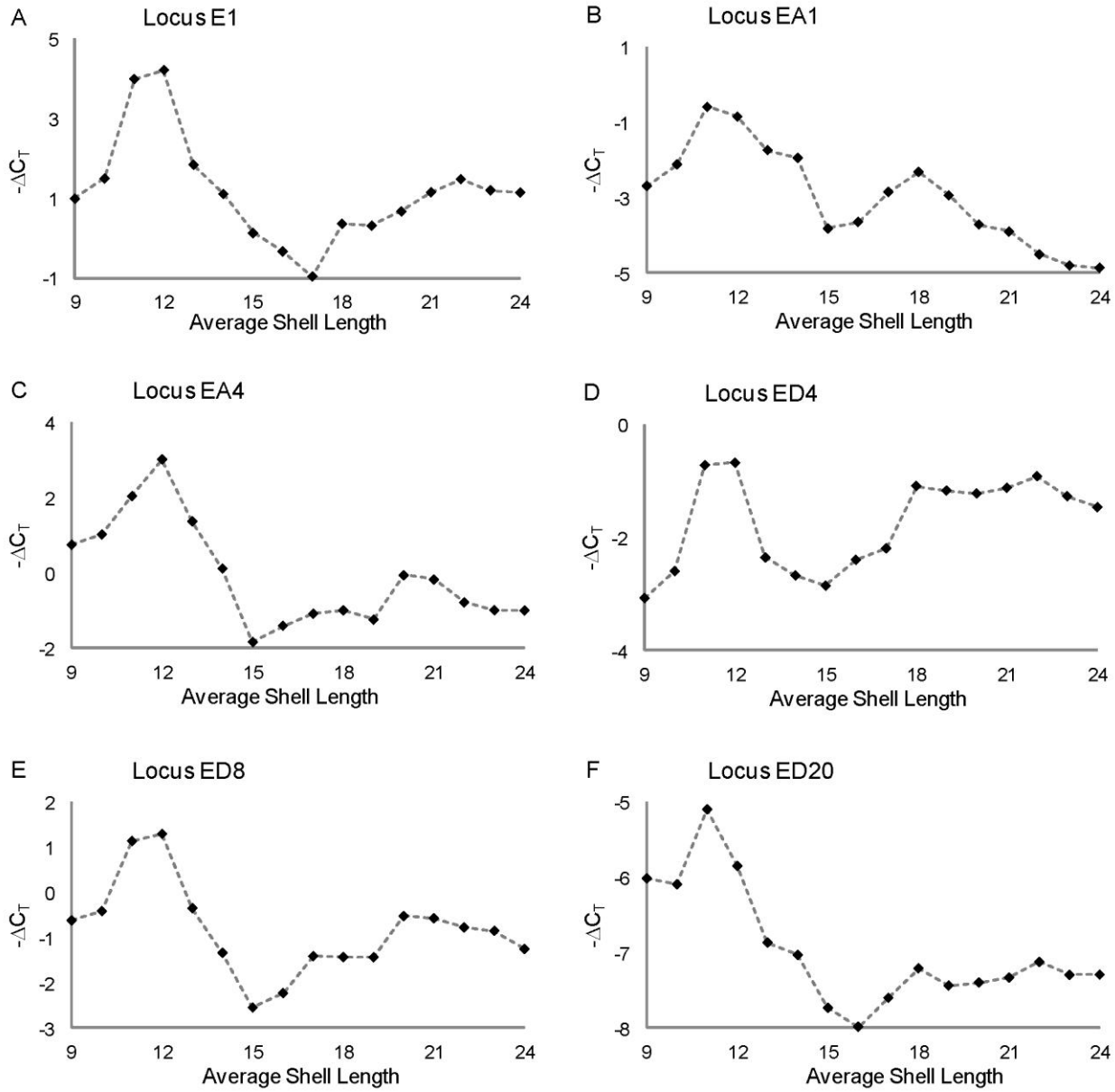


Figure 5.10. Simulated moving averages (window size of 5mm shell lengths) of expression levels ($-\Delta C_T$) of six conotoxin genes.
The simulation process involved randomly sampling two individuals of each shell size from pools of individuals of the same shell length, pooling these samples of different sizes together as a new dataset, and estimating moving averages for each locus through development. The whole procedure was repeated for 100 times and all results for each locus were plotted in the same figure with different colors among generations of simulation. (A) Locus E1. (B) Locus EA1. (C) Locus EA4. (D) Locus ED4. (E) Locus ED8. (F) Locus ED20.

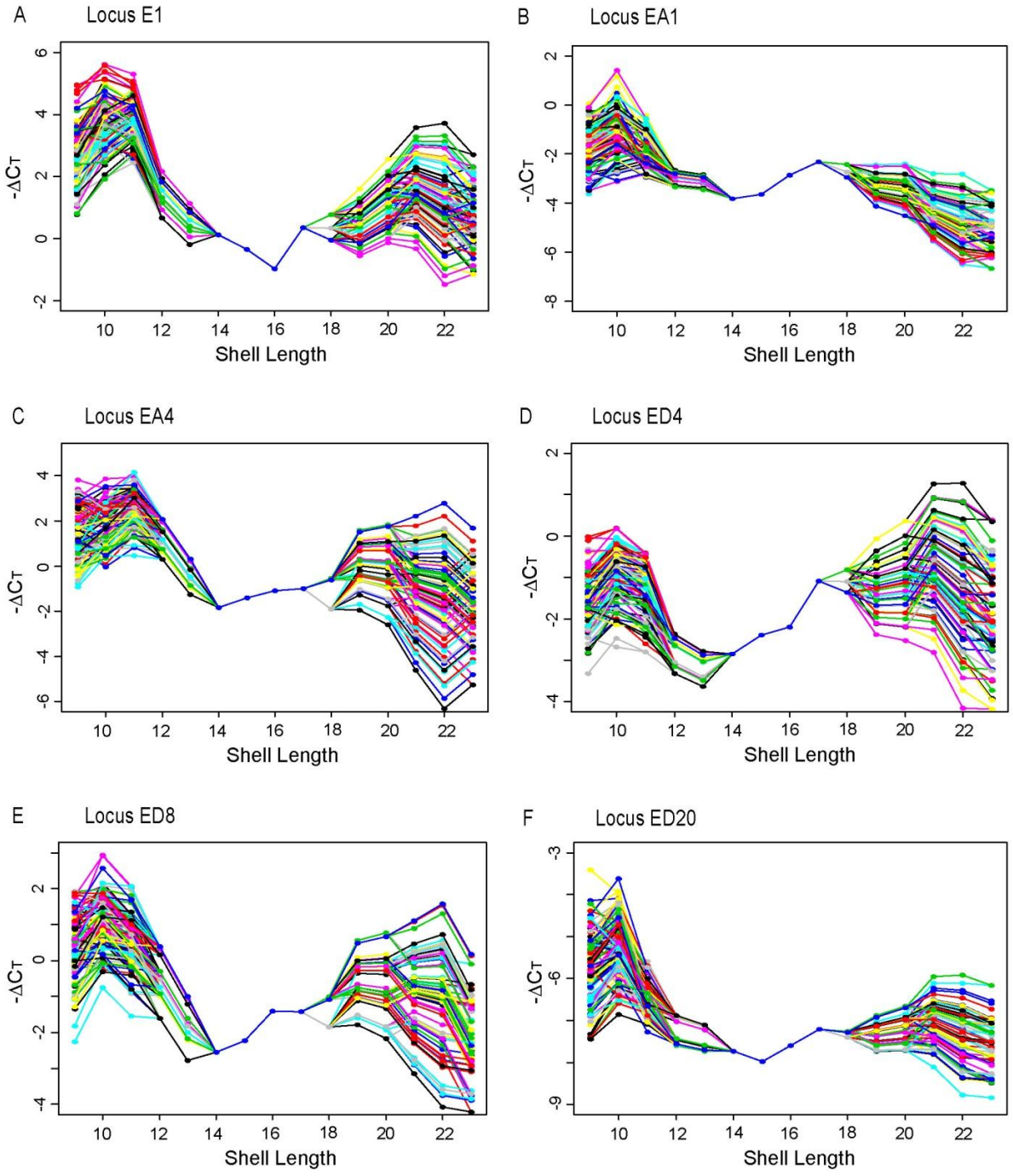


Figure 5.11. Boxplots of expression levels of *C. ebraeus* individuals feeding on different prey species.

(A-F) Conotoxin loci examined are (A) locus E1, (B) locus EA1, (C) locus EA4, (D) locus ED4, (E) locus ED8 and (F) locus ED20. Sample sizes of each category are labeled in (A).

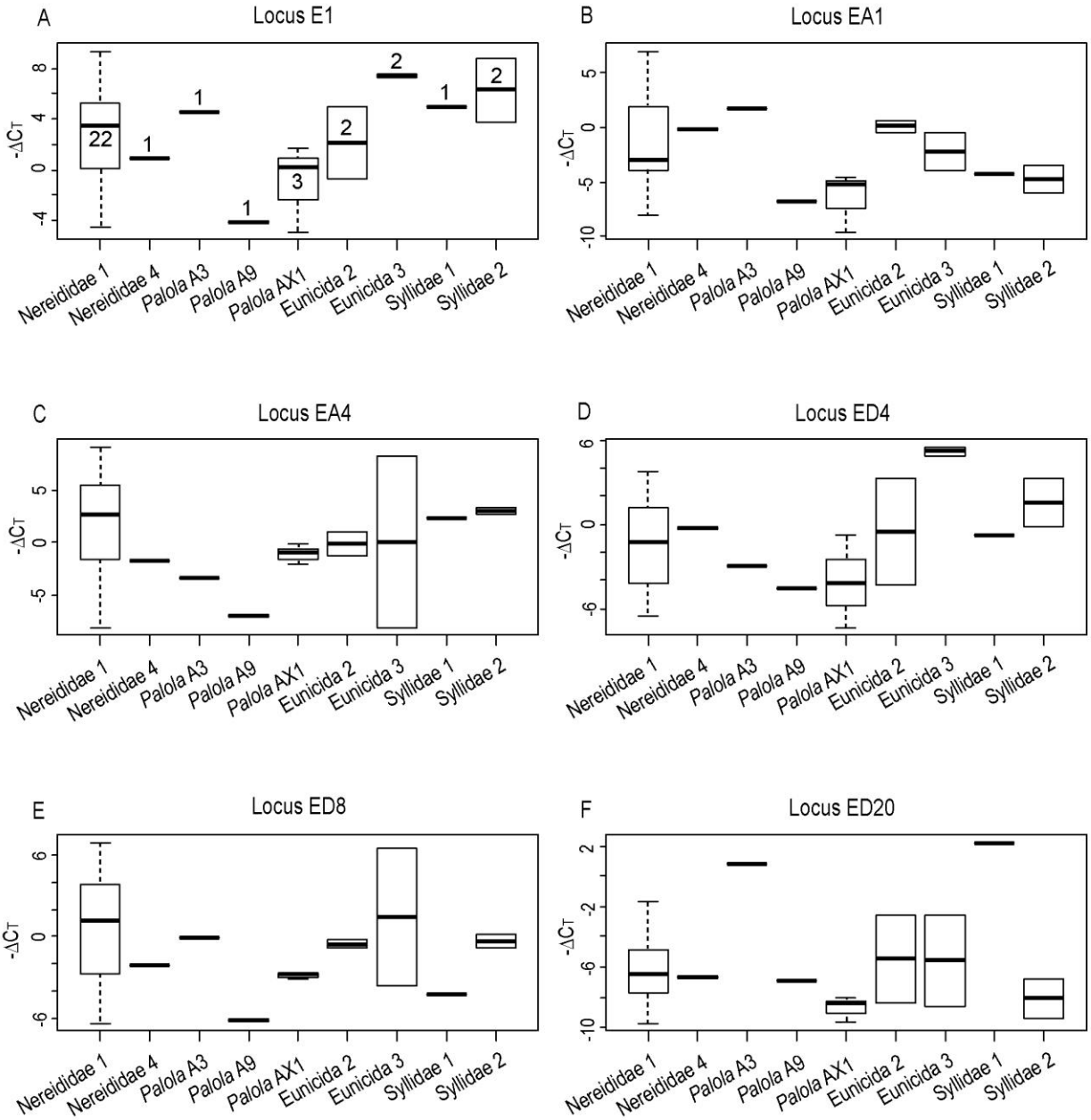


Figure 5.12. Boxplots of relative conotoxin gene expression levels among individuals feeding on different prey types.
Other Eunicida spp.: all Eunicida species except *Palola*. Eunicida: all Eunicida species. Non-Nereididae: prey species that are not of family Nereididae.

(A-F) Conotoxin loci examined are (A) locus E1, (B) locus EA1, (C) locus EA4, (D) locus ED4, (E) locus ED8 and (F) locus ED20. The sample sizes of each category are labeled next to the average lines of boxplots in panel A and are the same for all panels.

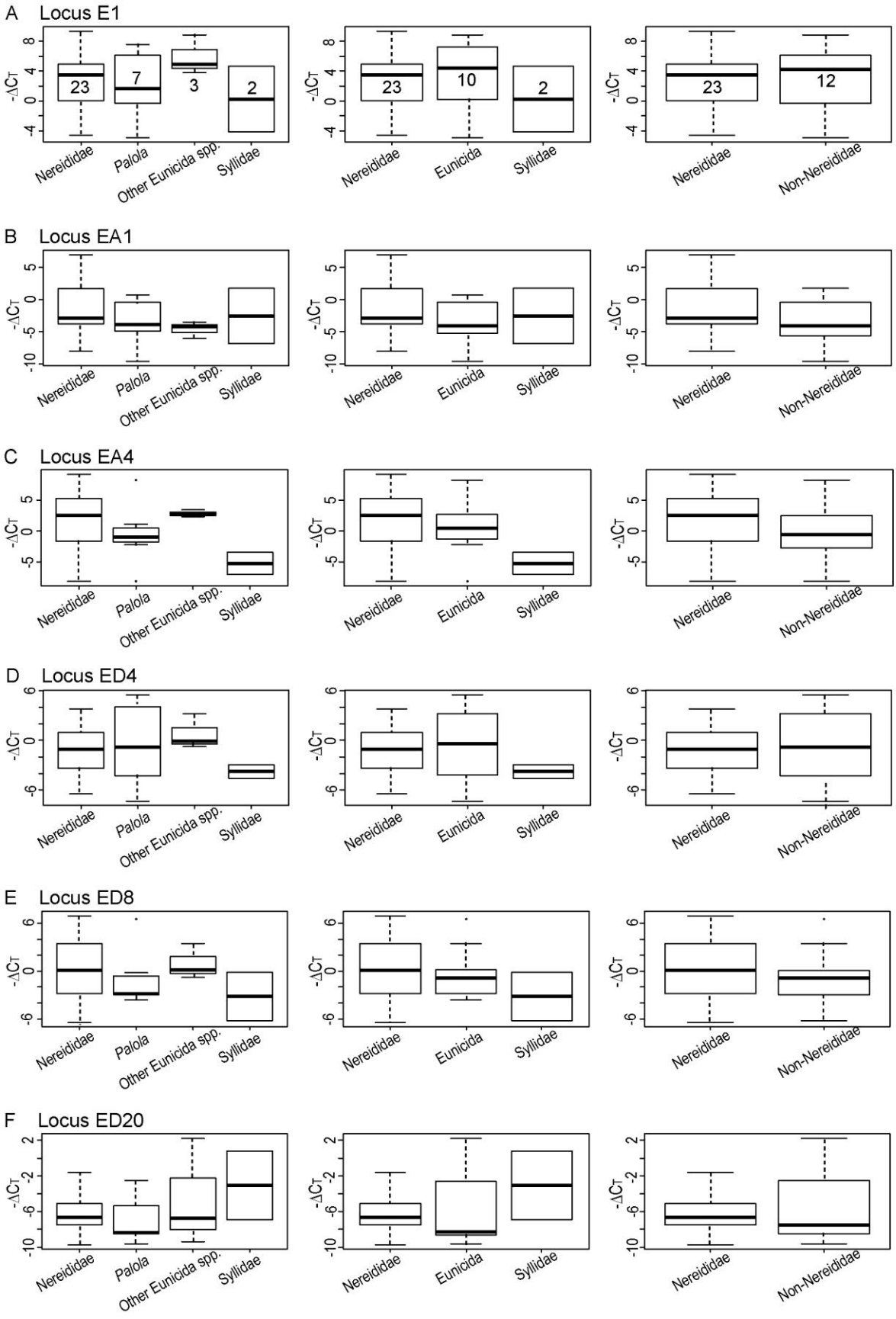

A Locus E1
B Locus EA1
C Locus EA4
D Locus ED4
E Locus ED8
F Locus ED20
-ΔCT
Nereididae
Palola
Other Eunicida spp.
Syllidae
Eunicida
Non-Nereididae
23
7
3
2
10
12

Figure 5.13. Heatmaps of conotoxin gene expression levels among *C. ebraeus* individuals with known prey species.
The row variables are individuals and the column variables are conotoxin genes. The input of each cell is (A) absolute and (B) scaled and centered $-\Delta C_T$ values between each conotoxin gene and the β-tubulin locus. Row labels on the left are shell lengths (mm) of each individual; those on the right are prey species. The column dendogram is illustrated above column labels.

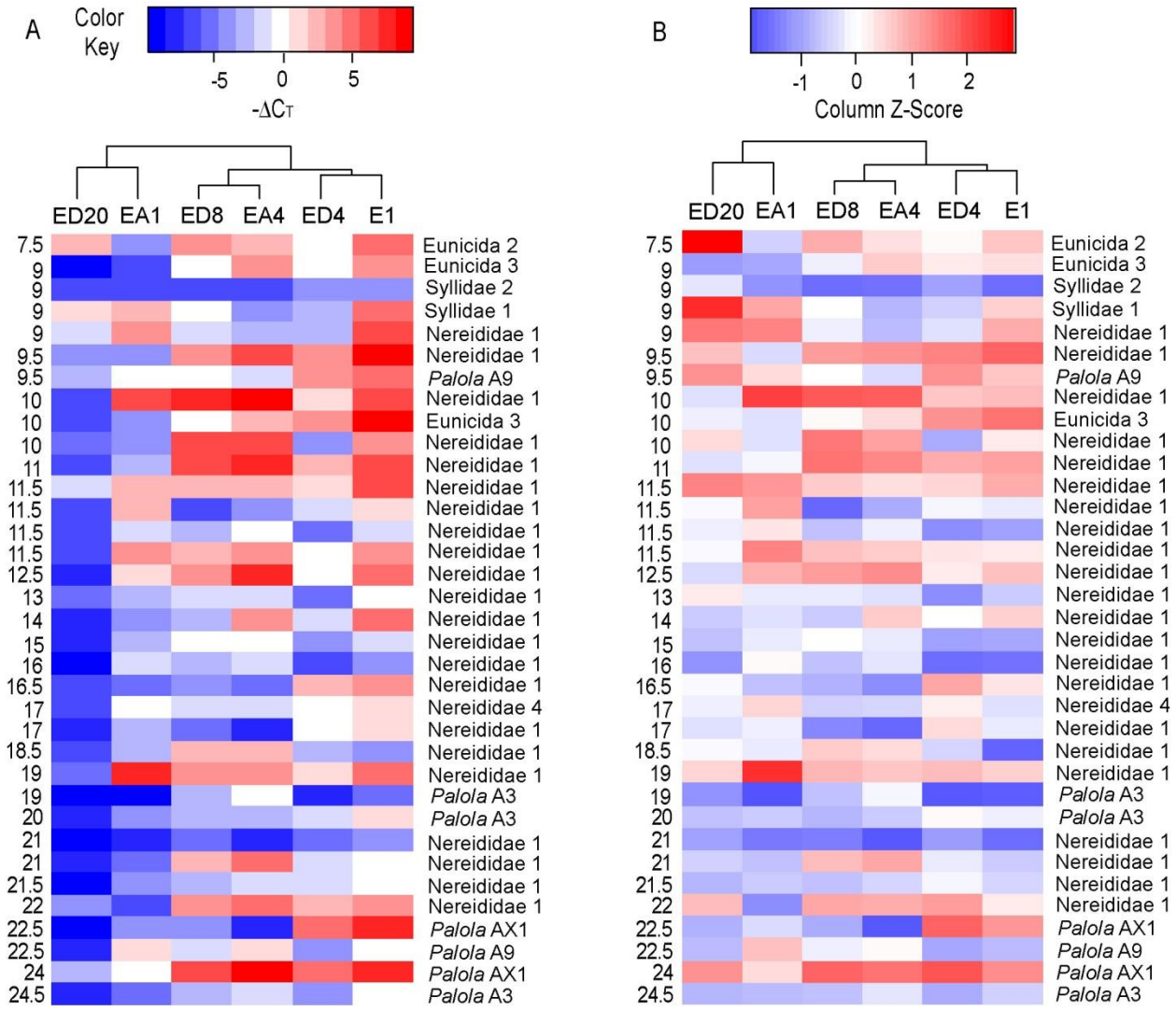


Figure 5.14. Hierarchical clusters of *C. ebraeus* individuals with known prey species based on patterns of expression of six conotoxin genes.

(A) Dendogram of individuals with the Wald's method. Shell lengths and prey species of each individual are labeled at the tip of each branch. The two major clusters are labeled at basal nodes.

(B-C) Size frequency distribution of shell lengths of individuals in (B) cluster 1 and (C) cluster 2 as defined in (A).

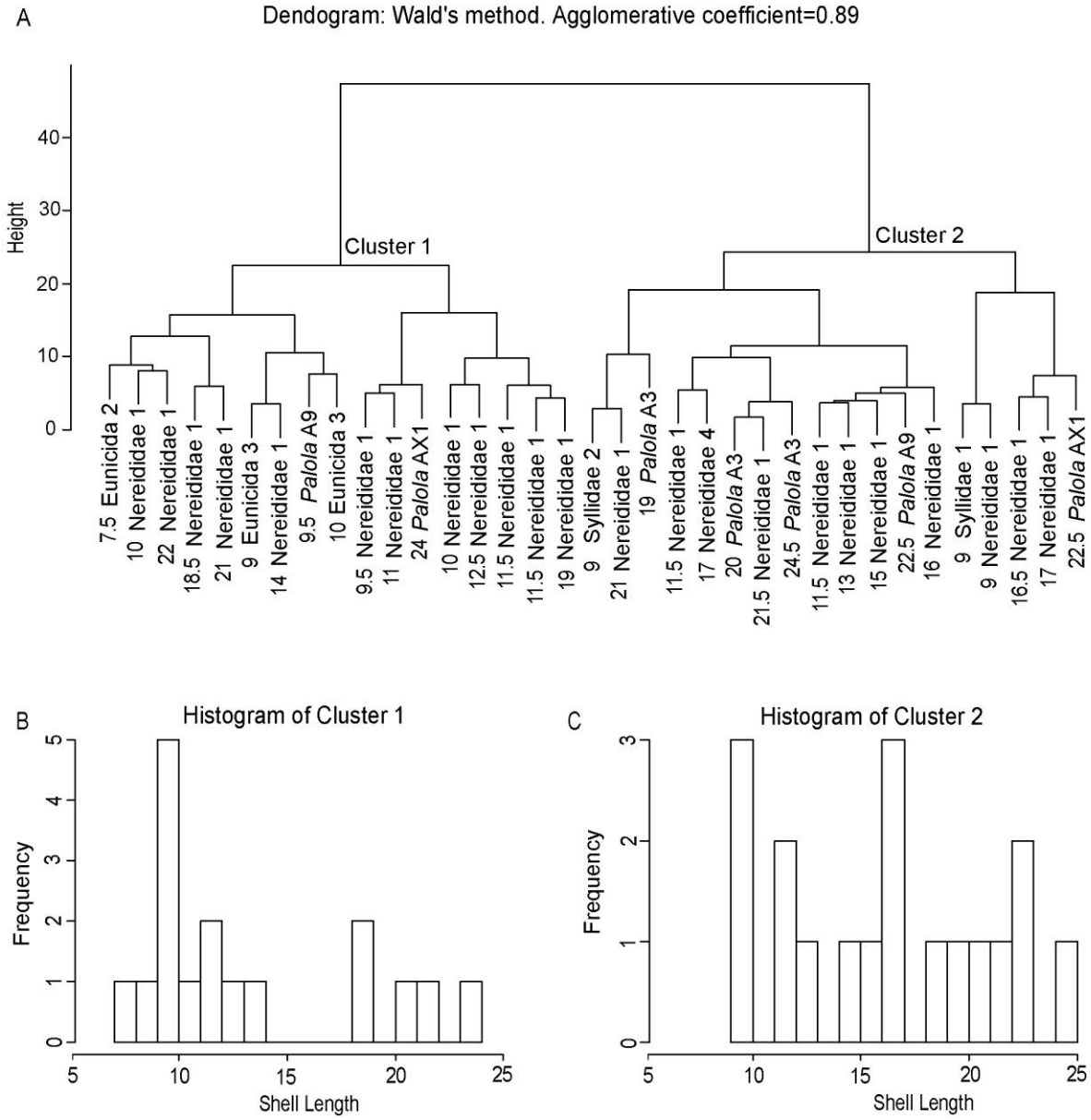


Figure 5.15. Patterns of ontogenetic shifts of levels of expression of conotoxin genes and dietary diversities.
Average levels of expression of the six conotoxin genes and dietary diversities are calculated with the sliding window approach (window size=5mm), centered and standardized. Dietary variables include Shannon's index (*H'*), Simpson's index (*S*) and average genetic distances (GD). (A-B) Plots of standardized dietary variables and relative levels of expression of (A) loci E1, EA4, ED4 and ED8 or (B) loci EA1 and ED20 against shell lengths.

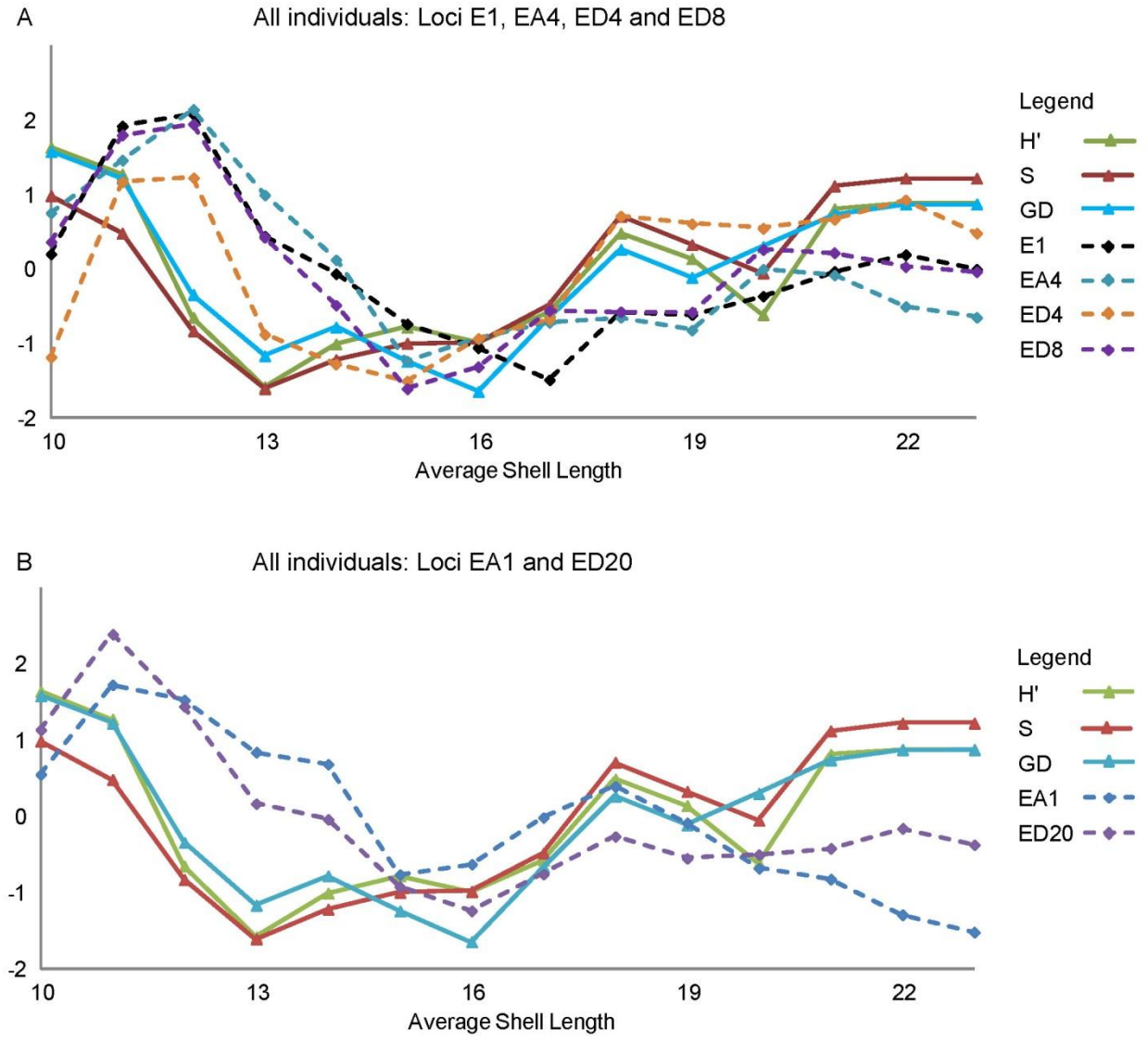


Figure 5.16. Cross-correlation of time series of conotoxin gene expression levels with dietary diversities.
The time trajectory is represented by increase of shell lengths. Dietary variables include Shannon's index (*H'*), Simpson's index (*S*) and average genetic distances of 16S gene sequences of prey (GD). In each plot, Y-axis is the correlation coefficient of two series, X-axis is the lag in shell lengths of dietary variables compared to conotoxin gene expression levels, and blue dashed lines are 95% confidence intervals. Any vertical line exceeding the 95% confidence intervals signifies a significant positive or negative correlation.

(A-F) Cross-correlations of dietary variables with relative levels of expression of (A) locus E1, (B) locus EA1, (C) locus EA4, (D) locus ED4, (E) locus ED8 and (F) locus ED20.

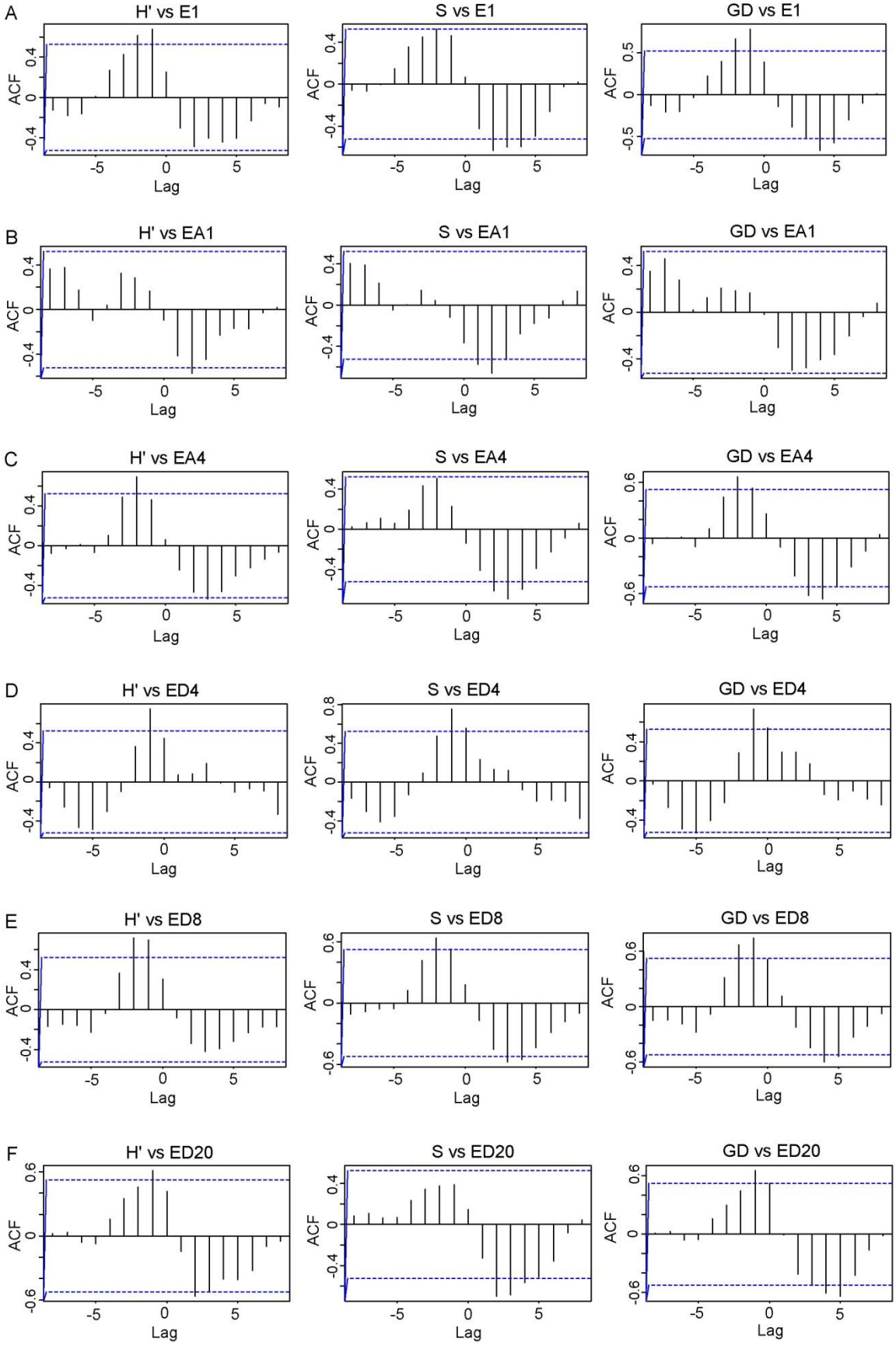

A
H' vs E1
S vs E1
GD vs E1
B
H' vs EA1
S vs EA1
GD vs EA1
C
H' vs EA4
S vs EA4
GD vs EA4
D
H' vs ED4
S vs ED4
GD vs ED4
E
H' vs ED8
S vs ED8
GD vs ED8
F
H' vs ED20
S vs ED20
GD vs ED20
ACF
Lag

Figure 5.17. Linear regression of specific lag of expression levels of each conotoxin locus with dietary variables (*S*, *H'* and GD).
Equations and R^2 values of linear regression are labeled next to the fitted trend line. All slope variables are significant based on 5% significance levels except locus ED20 against S in (E). Both axes are centered and standardized.

(A-E) Linear regression of lag of (A) locus E1, (B) locus EA4, (C) locus ED4, (D) locus ED8 and (E) locus ED20 against dietary variables.

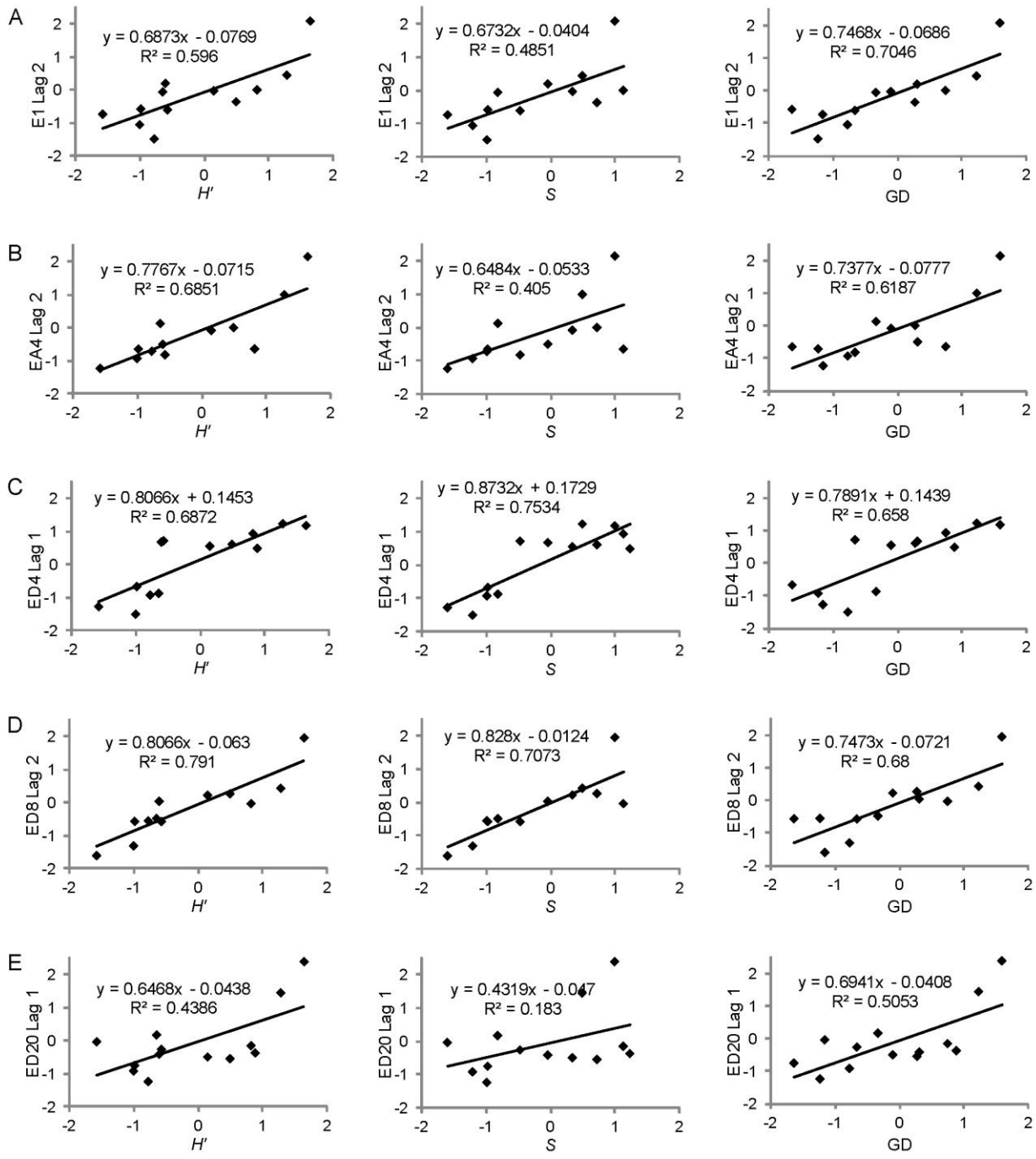


Figure 5.18. Boxplots of conotoxin gene expression levels among groups of individuals classified by sexual maturity.
Female and male: mature females and males. Juvenile: immature individuals (<12mm in shell lengths).

(A) Size distributions of individuals among three groups. Sample sizes (n) are labeled next to each group.

(B-G) Conotoxin gene expression levels among groups. Sample sizes of each group are the same as in panel (A). (B) locus E1; (C) locus EA1; (D) locus EA4; (E) locus ED4; (F) locus ED8; (G) locus ED20.

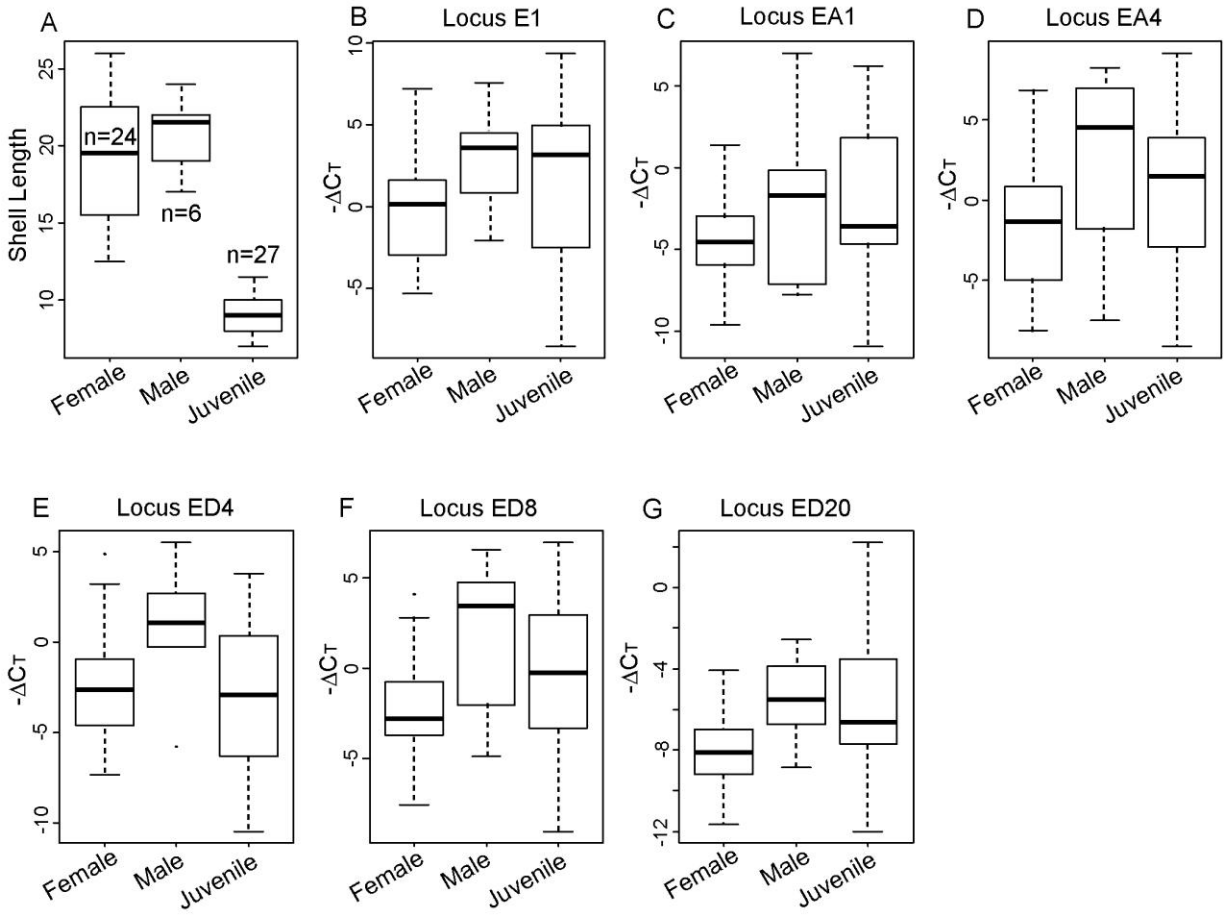


Table 5.1. Primers for the six conotoxin genes and that were utilized in Real-time qPCR. Putative conotoxin type encoded by each locus is labeled in parentheses of the 'superfamily' category. 'Toxin': toxin-coding region. The IUPAC nomenclature code (Johnson 2010) is used in primer sequences.

Type	Superfamily	Locus	Location	Primer Sequence (5' to 3')
Forward primer	O (ω-conotoxin)	E1	Prepro	CATCGTCAAGATGAAACTGACGTG
	O (δ-conotoxin)	ED4, ED8, ED20	Prepro	CATCACCAAGATGAAACTGAC
	A (α-conotoxin)	EA1, EA4	Prepro	ATGGGCATGCGGATGATGTTCAC
Reverse primer	O (ω-conotoxin)	E1	Toxin	ATCACGAAAGGGAAATATCAGGCG
	O (δ-conotoxin)	ED4	Toxin	CATTACATAAGCCATTGCAGCATCC
		ED8	Toxin	CAACTAGAGGCAGACGTGGAAAAG
		ED20	Toxin	AGCTCAACTAGGCGCAGTTGAAAT
	A (α-conotoxin)	EA1	Toxin	GGGTCCTGGAGCATCAGCCTTTA
		EA4	Toxin	TAKCAGCGTCTTCAACGACAATTC

Table 5.2. Fisher's exact tests of prey compositions between size classes of *C. ebraeus* individuals.
Small: individuals with shell lengths less than 11mm. Medium: individuals with shell lengths between 11mm and 17mm. Large: individuals with shell lengths larger than 17mm. *P*-values were estimated by the Monte Carlo simulation of 100,000 replicates.

Null hypothesis	**Alternative hypothesis**	***P*-value**
Small = Medium = Large	Small ≠ Medium ≠ Large	0.001
Small = Medium	Small ≠ Medium	0.024
Medium = Large	Medium ≠ Large	0.004
Small = Large	Small ≠ Large	0.111

Table 5.3. Pairwise Φ_{ST} values among *C. ebraeus* individuals of different size classes and locations.

P-values were estimated by 10,100 repeats of permutations. Guam 2008: adult samples (>20mm in shell lengths) collected at Guam in 2008. All data at Guam (2008), Hawaii and American Samoa were obtained from Chapter 4. Small: shell lengths less than 10mm; medium: shell lengths between10 to 20mm; large: shell lengths larger than 20mm. All individuals: all *C. ebraeus* samples collected at Guam in 2010 in this study. Sample sizes (n) of each group/population are listed in the row and column labels. Am Sam: American Samoa.

	Medium (n=22) (*P*-value)	**Large (n=13)** (*P*-value)	**Guam 2008 (n=29)** (*P*-value)	**Hawaii (n=48)** (*P*-value)	**Am Sam (n=21)** (*P*-value)
Small (n=14)	-0.026 (0.964)	0.009 (0.276)	0.014 (0.197)	0.287 (0.000)	0.034 (0.096)
Medium (n=22)	-	0.005 (0.320)	0.023 (0.100)	0.280 (0.000)	0.042 (0.047)
Large (n=13)	-	-	0.016 (0.198)	0.187 (0.000)	0.053 (0.042)
All individuals (n=49)	-	-	0.021 (0.065)	0.260 (0.000)	0.047 (0.016)

References

Alape-Giron A, Sanz L, Escolano J, Flores-Diaz M, Madrigal M, Sasa M, Calvete JJ. 2008. Snake venomics of the lancehead pitviper *Bothrops asper*: geographic, individual, and ontogenetic variations. *J Proteome Res.* 7:3556-3571.

Alstyne KLV. 1988. Herbivore grazing increases polyphenolic defenses in the intertidal brown alga *Fucus distichus*. *Ecology.* 69:655-663.

Andrade DV, Abe AS. 1999. Relationship of venom ontogeny and diet in *Bothrops*. *Herpetologica.* 55:200-204.

Antunes TC, Yamashita KM, Barbaro KC, Saiki M, Santoro ML. 2010. Comparative analysis of newborn and adult *Bothrops jararaca* snake venoms. *Toxicon.* 56:1443-1458.

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

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 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, Matheny AM. 2009. Cryptic species differentiated in *Conus ebraeus*, a widespread tropical marine gastropod. *Biol Bull.* 217:292-305.

Duda TF, Palumbi SR. 1999. Molecular genetics of ecological diversification: Duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc Natl Acad Sci USA.* 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.* 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.

Duda TF, Terbio M, Chen G, Phillips S, Olenzek AM, Chang D, Morris DW. 2012. Patterns of population structure and historical demography of *Conus* species in the tropical pacific. *Am Malacol Bull.* 30:175-187.

Dutertre S, Ulens C, Buttner R et al. 2007. AChBP-targeted α-conotoxin correlates distinct binding orientations with naChR subtype selectivity. *EMBO J.* 26:3858-3867.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform.* 1:47-50.

Fisher RA. 1954. Statistical methods for research workers (biological monographs and. Manuals no. V). Edinburgh: Oliver and Boyd.

Frank PW. 1969. Growth rates and longevity of some gastropod mollusks on the coral reef at Heron Island. *Oecologia.* 2:232-250.

Frederich B, Adriaens D, Vandewalle P. 2008. Ontogenetic shape changes in Pomacentridae (Teleostei, Perciformes) and their relationships with feeding strategies: A geometric morphometric approach. *Biol J Linn Soc.* 95:92-105.

Gibbs HL, Sanz L, Chiucchi JE, Farrell TM, Calvete JJ. 2011. Proteomic analysis of ontogenetic and diet-related changes in venom composition of juvenile and adult dusky pigmy rattlesnakes (*Sistrurus miliarius barbouri*). *J Proteomics.* 74:2169-2179.

Herrel A, O' Reilly JC. 2006. Ontogenetic scaling of bite force in lizards and turtles. *Physiol Biochem Zool.* 79:31-42.

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.

Johnson AD. 2010. An extended IUPAC nomenclature code for polymorphic nucleic acids. *Bioinformatics.* 26:1386-1389.

Kaas Q, Westermann J-C, Craik DJ. 2010. Conopeptide characterization and classifications: An analysis using Conoserver. *Toxicon.* 55:1491-1509.

Kawamura T, Roberts RD, Yamashita Y. 2001. Radula development in abalone *Haliotis discus hannai* from larva to adult in relation to feeding transitions. *Fisheries Sci.* 67:596-605.

Kintner AH, Seymour JE, Edwards SL. 2005. Variation in lethality and effects of two Australian chirodropid jellyfish venoms in fish. *Toxicon.* 46:699-708.

Kohn AJ, Nybakken JW. 1975. Ecology of *Conus* on Eastern Indian-Ocean fringing reefs - diversity of species and resource utilization. *Mar Biol.* 29:211-234.

Kohn AJ, Perron FE. 1994. Life history and biogeography patterns in *Conus*. New York:Oxford University Press.

Landry CR, Oh J, Hartl DL, Cavalieri D. 2006. Genome-wide scan reveals that genetic variation for transcriptional plasticity in yeast is biased towards multi-copy and dispensable genes. *Gene.* 366:343-351.

Leviten PJ. 1976. The foraging strategy of vermivorous conid gastropods. *Ecol Monogr.* 46:157-178.

Lopez-Maury L, Marguerat S, Bahler J. 2008. Tuning gene expression to changing environments: From rapid responses to evolutionary adaptation. *Nat Rev Genet.* 9:583-593.

Mackessy SP, Sixberry NA, Heyborne WH, Fritts T. 2006. Venom of the brown treesnake, *Boiga irregularis*: Ontogenetic shifts and taxa-specific toxicity. *Toxicon.* 47:537-548.

MacKessy SP, Williams K, Ashon KG. 2003. Ontogenetic variation in venom composition and diet of *Crotalus oreganus concolor*: A case of venom paedomorphosis? *Copeia.*769-782.

MacNeill DB, Brandt SB. 1990. Ontogenetic shifts in gill-raker morphology and predicted prey capture efficiency of the alewife, *Alosa pseudoharengus*. *Copeia.* 1990:164-171.

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. 2012. Cluster: Cluster analysis basics and extensions.

Nybakken J. 1988. Possible ontogenetic change in the radula of *Conus patricius* of the Eastern Pacific. *Veliger.* 31:222-225.

Nybakken J. 1990. Ontogenic change in the *conus* radula, its form, distribution among the radula types, and significance in systematics and ecology. *Malacologia.* 32:35-54.

Nybakken J, Perron F. 1988. Ontogenetic change in the radula of *Conus magus* (Gastropoda). *Mar Biol.* 98:4.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos PM, Stevens HH, Wagner H. 2012. Vegan: Community ecology package. R package version 2.0-3. Http://cran.R-project.Org/package=vegan.

Olivera BM. 2002. *Conus* venom peptides: Reflections from the biology of clades and species. *Ann Rev Ecol Syst.* 33:25-47.

Piersma T, Drent J. 2003. Phenotypic flexibility and the evolution of organismal design. *Trends Ecol Evol.* 18:228-233.

Posada D. 2008. jModelTest: Phylogenetic model averaging. *Mol Biol Evol.* 25:1253-1256.

Rambaut A. 2002. Se-al sequence alignment editor. Version 2.0.A11. Oxford: University of Oxford.

Richter K, Haslbeck M, Buchner J. 2010. The heat shock response: Life on the verge of death. *Mol Cell.* 40:253-266.

Rivera-Ortiz JA, Cano H, Mar íF. 2011. Intraspecies variability and conopeptide profiling of the injected venom of *Conus ermineus*. *Peptides.* 32:306-316.

Rousseeuw PJ. 1986. A visual display for hierarchical classificiation. *in* Diday E, Escoufier Y, Lebart J, Pages J, Schektman Y, and Tomassone R, eds. Data analysis and informatics 4. North-Holland, Amsterdam, p. 743-748.

Saldarriaga MMa, Otero R, Núñez V, Toro MF, Díaz A, Gutiérrez JMa. 2003. Ontogenetic variability of *Bothrops atrox* and *Bothrops asper* snake venoms from Colombia. *Toxicon.* 42:405-411.

Scheiner SM. 1993. Genetics and evolution of phenotypic plasticity. *Annu Rev Ecol Syst.* 24:35-68.

Schmittgen TD, Livak KJ. 2008. Analyzing real-time pcr data by the comparative C_T method. *Nat Protocols.* 3:1101-1108.

Schulze A. 2006. Phylogeny and genetic diversity of palolo worms (*Palola*, Eunicidae) from the tropical North Pacific and the Caribbean. *Biol Bull.* 210:25-37.

Shannon CE. 1948. A mathematical theory of communication. *Bell Syst Tech J.* 27:379-423, 623-656.

Simpson EH. 1949. Measurement of diversity. *Nature.* 163:688.

Starck JM. 1999. Phenotypic flexibility of the avian gizzard: Rapid, reversible and repeated changes of organ size in response to changes in dietary fibre content. *J Evol Biol.* 202:3171-3179.

Struyf A, Hubert M, Rousseeuw P. 1997. Clustering in an object-oriented environment. *J Stat Softw.* 1:1-30.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 10:512-526.

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.

Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *in* Miura Rm, ed. Lectures on mathematics in the life sciences. Providence, Rhode Island: American Mathematical Society, p. 57-86.

Terlau H, Olivera BM. 2004. *Conus* venoms: A rich source of novel ion channel-targeted peptides. *Physiol Rev.* 84:41-68.

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Valena S, Moczek AP. 2012. Epigenetic mechanisms underlying developmental plasticity in horned beetles. *Genet Res Int.* 2012:14.

Venkatesh B, Lu SQ, Dandona N, See SL, Brenner S, Soong TW. 2005. Genetic basis of tetrodotoxin resistance in pufferfishes. *Curr Biol.* 15:2069-2072.

Vincent SE, Moon BR, Herrel A, Kley NJ. 2007. Are ontogenetic shifts in diet linked to shifts in feeding mechanics? Scaling of the feeding apparatus in the banded watersnake *Nerodia fasciata*. *J Exp Biol.* 210:2057-2069.

Warnes GR. 2012. gplots: Various R programming tools for plotting data, R package version 2.11.0. http://cran.R-project.Org/package=gplots.

West-Eberhard MJ. 2005. Developmental plasticity and the origin of species differences. *Proc Natl Acad Sci USA.* 102:6543-6549.

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 alpha 7 nicotinic acetylcholine receptor antagonist. *Biochemistry.* 46:6628-6638.

Widmark J, Sundström G, Ocampo Daza D, Larhammar D. 2011. Differential evolution of voltage-gated sodium channels in tetrapods and teleost fishes. *Mol Biol Evol.* 28:859-871.

Yampolsky LY, Glazko GV, Fry JD. 2012. Evolution of gene expression and expression plasticity in long-term experimental populations of *Drosophila melanogaster* maintained under constant and variable ethanol stress. *Mol Ecol.* 21:4287-4299.

Zakon HH. 2012. Adaptive evolution of voltage-gated sodium channels: The first 800 million years. *Proc Natl Acad Sci USA.*

Zelanis A, Andrade-Silva D, Rocha MM, Furtado MF, Serrano SMT, Junqueira-de-Azevedo ILM, Ho PL. 2012. A transcriptomic view of the proteome variability of newborn and adult *Bothrops jararaca* snake venoms. *PLoS Negl Trop Dis.* 6:e1554.

Zelanis A, Tashima AK, Rocha MMT, Furtado MF, Camargo ACM, Ho PL, Serrano SMT. 2010. Analysis of the ontogenetic variation in the venom proteome/peptidome of *Bothrops jararaca* reveals different strategies to deal with prey. *J Proteome Res.* 9:2278-2291.