Contrasting Modes and Tempos of Venom Expression Evolution in Two Snake Species

Mark J. Margres,* James J. McGivern,* Margaret Seavy,* Kenneth P. Wray,*

Jack Facente,⁺ and Darin R. Rokyta*,¹

*Department of Biological Science, Florida State University, Tallahassee, Florida 32306, and [†]Agritoxins, Saint Cloud, Florida 34771

ABSTRACT Selection is predicted to drive diversification within species and lead to local adaptation, but understanding the mechanistic details underlying this process and thus the genetic basis of adaptive evolution requires the mapping of genotype to phenotype. Venom is complex and involves many genes, but the specialization of the venom gland toward toxin production allows specific transcripts to be correlated with specific toxic proteins, establishing a direct link from genotype to phenotype. To determine the extent of expression variation and identify the processes driving patterns of phenotypic diversity, we constructed genotype-phenotype maps and compared range-wide toxin-protein expression variation for two species of snake with nearly identical ranges: the eastern diamondback rattlesnake (*Crotalus adamanteus*) and the eastern coral snake (*Micrurus fulvius*). We detected significant expression variation in *C. adamanteus*, identified the specific loci associated with population differentiation, and found that loci expressed at all levels contributed to this divergence. Contrary to expectations, we found no expression variation in *M. fulvius*, suggesting that *M. fulvius* populations are not locally adapted. Our results not only linked expression variation at specific loci to divergence in a polygenic, complex trait but also have extensive conservation and biomedical implications. *C. adamanteus* is currently a candidate for federal listing under the Endangered Species Act, and the loss of any major population would result in the irrevocable loss of a unique venom phenotype. The lack of variation in *M. fulvius* has significant biomedical application because our data will assist in the development of effective antivenom for this species.

NATURAL selection can be a powerful force driving rapid diversification within species and is predicted to lead to local adaptation through the increase in frequency of mutations in gene-regulatory or protein-coding regions (Stern 2000; Hoekstra and Coyne 2007; Carroll 2008; Muller 2007). Expression variation at single loci has produced adaptive phenotypic divergence in the beaks of Darwin's finches (Abzhanov *et al.* 2004), coat color in mice (Manceau *et al.* 2011), and mimicry in butterflies (Reed *et al.* 2011). Most traits, however, are products of poorly characterized developmental pathways involving many loci. Many of the fundamental features of evolving systems, such as evolvability, epistasis, pleiotropy, and basic variational properties (Rokyta *et al.* 2008, 2009,

Copyright © 2015 by the Genetics Society of America doi: 10.1534/genetics.114.172437

2011b; Wager 2008; Chou *et al.* 2011; Woods *et al.* 2011; Hill and Zhang 2012), result from the relationship between genotype and phenotype (Stadler *et al.* 2001; Hansen 2006), but the ability to study this relationship directly in polygenic traits is rare. Therefore, linking gene-regulatory changes to adaptive evolution in polygenic, complex phenotypes remains a challenge (Romero *et al.* 2012; Savolainen *et al.* 2013).

Snake venoms are complex cocktails of ~40–100 proteinaceous toxins (Boldrini-França *et al.* 2010; Calvete *et al.* 2010; Pavlicev *et al.* 2011; Rokyta *et al.* 2011a, 2012, 2013; Durban *et al.* 2013; Margres *et al.* 2013) that collectively function in prey capture, digestion, and defense. Although most quantitative traits are the products of developmental pathways where gene-regulatory changes may have effects mediated through complex interaction networks, toxin expression variation directly changes the phenotype because relative amounts of venom components determine, in part, the venom's efficacy. The specialization of the venom gland toward toxin production allows the use of transcriptomics to identify a large number of loci that contribute to the venom phenotype (Rokyta *et al.* 2011a, 2012, 2013; Margres *et al.* 2013), and proteomic

Manuscript received September 18, 2014; accepted for publication November 4, 2014; published Early Online November 11, 2014.

Supporting information is available online at http://www.genetics.org/lookup/suppl/ doi:10.1534/genetics.114.172437/-/DC1.

¹Corresponding author: Department of Biological Science, Florida State University, 319 Stadium Drive, 4058 King Life Sciences, Tallahassee, FL 32306. E-mail: drokyta@bio.fsu.edu

techniques can verify the secretion of the proteins produced by these loci (Margres *et al.* 2014). Correlating specific transcripts with specific toxic proteins in the context of whole venom establishes a direct link from genotype to phenotype, allowing us to link expression variation at specific loci to variation in a complex trait. Expression is typically measured at the mRNA level (Rokyta *et al.* 2011a, 2012, 2013; Margres *et al.* 2013). Venom, however, is a secretion, and protein expression can be measured directly by reversed-phase HPLC (RP-HPLC), permitting accurate quantification of populationlevel phenotypic variation.

Venomous snakes and their prey represent a powerful system for studying the process of local adaptation (Biardi et al. 2005). Snake venoms function solely following injection into another organism, and prey provide geographically variable selective pressures through their own genetic variation and variation in local species compositions and abundances. Selection can vary among interacting populations as a result of genotype-by-genotype-by-environment interactions, producing phenotypic variation because locally beneficial traits are not expected to become fixed at the species level (Thompson 2005). Although intraspecific variation in venom composition has been documented previously (Chippaux et al. 1991; Daltry et al. 1996; Creer et al. 2003; Mackessy 2010; Sunagar et al. 2014), comparisons have been based on clinical symptoms of human patients (Chippaux et al. 1991), suffered from small sample sizes (Straight et al. 1991; Sunagar et al. 2014), and/or employed qualitative assessments of venom appearance or protein presence or absence (Straight et al. 1991; Daltry et al. 1996; Creer et al. 2003; Boldrini-França et al. 2010; Mackessy 2010). These comparisons effectively characterized venom differences at the protein-family level, but paralogs can have vastly different functions, and expression variation may be locus specific rather than gene-family specific. Therefore, the construction of a genotype-phenotype map would improve the resolution of such comparisons by characterizing venom at the locus level (Margres et al. 2014).

To determine whether snake venoms exhibit geographic variation as predicted for a locally adapted trait and identify which loci contributed to phenotypic divergence, we used our genotype-phenotype map approach and proteomically compared toxin expression across multiple populations for two species with nearly identical ranges in the southeastern United States: the eastern diamondback rattlesnake (Crotalus adamanteus) and the eastern coral snake (Micrurus fulvius). Comparative studies of codistributed species can improve our understanding of the processes driving patterns of phenotypic diversity, and comparing trait distributions for independent species across the same landscape will show whether selection or other factors, such as drift, are responsible for the extent of the observed phenotypic variation (Gomulkiewicz et al. 2007). C. adamanteus is the largest species of rattlesnake and consumes small mammals, with rats, squirrels, and rabbits comprising most of its diet (Klauber 1997). C. adamanteus is historically native to seven states in the southeastern United States but recently has been extirpated

from Louisiana, is endangered in North Carolina (Palmer and Braswell 1995), and is currently under consideration for listing as threatened under the Endangered Species Act (United States Fish and Wildlife Service 2012). Identifying phenotypically distinct populations for this species of conservation concern is critical to effective species management. M. fulvius is also native to the forests of the southeastern coastal plain and is primarily ophiophagous, preying on small snakes and other squamates (Jackson and Franz 1981). M. fulvius envenomations can be lethal, and antivenom is currently unavailable in the United States (Norris et al. 2009). Intraspecific venom variation can limit the efficacy of antivenoms (Chippaux et al. 1991; Casewell et al. 2014; Sunagar et al. 2014), and the extent of variation within M. fulvius is currently unknown. We sought to characterize the variation within this species, assisting in the development of an effective antivenom.

Materials and Methods

Transcriptome assembly and analysis

The raw paired-end reads passing the Illumina quality filter were merged if their 3' ends overlapped, as described previously (Rokyta et al. 2012, 2013; Margres et al. 2013). To eliminate reads corresponding to extremely high-abundance transcripts, we used the Extender program (Rokyta et al. 2012) with 1,000 merged reads as seeds to attempt to generate complete transcripts using only the merged reads. Extension of seeds required an overlap of 100 nucleotides, Phred scores of at least 30 at each position in the extending read, and an exact match in the overlapping region. We performed a reference-based assembly against the 3,031 nontoxins previously annotated for C. horridus (Transcriptome Shotgun Assembly accession number GAAZ01000000; Rokyta et al. 2013) with NGen version 11 using both the merged and unmerged reads and a minimum match percentage of 93% for C. adamanteus and 85% for M. fulvius. Consensus sequences were retained if they had at least 10x coverage over the entire coding sequence; regions outside the coding sequence with less than 10x coverage were removed. Toxin sequences were clustered into groups with <1% nucleotide divergence in their coding sequences, and duplicate nontoxin sequences were eliminated following alignment of the final transcripts with NGen. We used one representative from each toxin cluster and all the unique nontoxins to filter the corresponding reads in a reference-based transcriptome assembly in NGen with a minimum match percentage of 98% using only the merged reads. The unfiltered reads then were used in a de novo transcriptome assembly in NGen with the defaultminimum match percentage of 93%, retaining only contigs comprised of at least 100 reads.

To increase our chances of identifying all toxin sequences, we performed four additional *de novo* assemblies for each species. Three assemblies were performed with NGen with a minimum match percentage of 98% using 1 million, 5 million, and 10 million reads. We opted for high stringency for these assemblies to attempt to differentiate among paralogs and varied the number of reads because we found that some high-abundance transcripts were difficult to assemble with too many reads (*i.e.*, >1 million), apparently because of low levels of unspliced transcripts. The fourth additional *de novo* assembly used the Extender program as earlier on 1,000 new random reads, and we only used this assembly to identify snake venom metalloproteinases, which were difficult for other methods to assemble. After combining the results of all the preceding assemblies and eliminating duplicates, we performed one final round of read filtering of the merged reads, followed by a *de novo* assembly of the unfiltered reads as earlier, keeping only contigs comprising >1,000 reads. This step was included to ensure that we missed no toxin sequences with appreciable expression.

To identify and eliminate potentially chimeric sequences in our toxin databases for both species, we screened for evidence of recombination within each toxin family with GARD (Pond et al. 2006). We used the general reversible model of sequence evolution and gamma-beta rates. If we found a signal for recombination resulting in significantly different tree topologies for different regions of the alignments based on KH tests, we performed a reference-based assembly with NGen version 11 with a minimum match percentage of 98% and the autotrim parameter set to false, using the toxin coding sequences as references and 10 million merged reads. Such high-stringency alignments facilitate the identification of chimeric sequences by producing either multimodal or extremely uneven coverage distributions, particularly in combination with our long, merged reads. Suspect sequences were confirmed to be chimeras of other sequences in our toxin database before removal.

Sequences were identified by means of nucleotide 6-frame translation-protein (blastx) searches against the NCBI nonredundant (nr) protein database with a minimum E-value of 10^{-4} and retaining only 10 hits. De novo assembled transcripts were retained and annotated only if they had complete protein-coding sequences. Putative toxins were identified by searching their blastx match descriptions for toxinrelated keywords, as described previously (Rokyta et al. 2012, 2013; Margres et al. 2013). The final set of unique transcripts for each species was generated by combining the results from all assemblies and eliminating duplicates by means of an NGen assembly and a second, more stringent assembly in SeqMan Pro. Final transcript abundances were estimated by means of a reference-based transcriptome assembly with NGen with a minimum match percentage of 95% using only the coding sequences of transcripts. Signal peptides for toxins were identified via SignalP analyses (Bendtsen et al. 2004). Putative toxins were named with a toxin class abbreviation, a number indicating cluster identity, and a lowercase letter indicating the particular member of a cluster.

Mass spectrometry analysis

Chromatographic separation and tandem mass spectrometry of the 21 RP-HPLC peaks previously identified in *C. adamanteus* were performed as described previously (Margres *et al.* 2014; Figure 1A). Four additional peaks were identified following RP-HPLC analysis of all *C. adamanteus* venom samples. These peaks (1b, 15b, 20b, and 20c) were isolated from a venom sample collected from KW1127, an adult female (127-cm snout-vent length, 136-cm total length) from Ca1 (Figure 2), as described earlier (Figure 1B). Twenty-two RP-HPLC peaks were collected from 400 μ g of *M. fulvius* venom on an analytical column (Figure 1C), as described previously (Margres *et al.* 2014). These peaks were isolated from a venom sample collected from the adult female (62-cm snout-vent length, 68.5-cm total length; Mf2 in Figure 2) used in the transcriptome analysis of Margres *et al.* (2013).

An externally calibrated Thermo Scientific LTQ Orbitrap Velos nLC-ESI-LIT was used for tandem mass spectrometry on the four additional C. adamanteus peaks (1b, 15b, 20b, and 20c) and 22 M. fulvius peaks. A 2-cm, 100-µm inside diameter trap column was followed by a 10-cm analytical column of 75 µm inside diameter (SC001 and SC200 Easy Columns, C-18 AQ packing, Thermo Scientific, Waltham, MA). Separation was carried out using EASY-nLC II with a continuous vented column configuration. A $2-\mu l$ (200-ng) sample was aspirated into a 20-µl loop and loaded onto the trap. The flow rate was set to 300 nl/min for separation on the analytical column, and a 1-h linear gradient from 0 to 45% acetonitrile in 0.1% formic acid was run. The LC effluent was directly nanosprayed into the mass spectrometer, which was operated in a data-dependent mode under direct control of the Xcalibur software. Ten data-dependent collisional-induced-dissociation (CID) MS/MS scans per full scan were performed. Each sample was run in triplicate.

Tandem mass spectra were extracted by Proteome Discoverer version 1.4.0.288. Charge-state deconvolution and deisotoping were not performed. Sequest version 1.4.0.288 (Thermo Fisher Scientific, San Jose, CA) was used to search the transcriptome databases containing 4,186 entries for C. adamanteus and 1,352 entries for M. fulvius with signal peptides removed and assuming the digestion enzyme trypsin, allowing one missed digestion site. Oxidation of methionine (variable) and carbamidomethylation of cysteine (variable) were specified as potential modifications. For C. adamanteus, the 21 RP-HPLC peaks identified in Margres et al. (2014) were analyzed with a fragment-ion mass tolerance of 0.80 Da and a parent-ion tolerance of 2.0 Da. The four new C. adamanteus peaks identified in this study and the 22 RP-HPLC peaks identified in M. fulvius were analyzed using Sequest and X! Tandem version Cyclone 2010.12.01.1 (The GPM, thegpm.org) using the transcriptome databases as described earlier. Because of the increased sensitivity of the Orbitrap, a fragment-ion mass tolerance of 0.50 Da and a parent-ion tolerance of 10.0 ppm were specified. Glutamic acid and glutamine conversions to pyroglutamate and deamination of N-termini were specified in X! Tandem as variable modifications. Scaffold version 4.3.2 (Proteome Software, Portland, OR) was used to validate MS/MS-based peptide and protein identifications for both species. Peptide identifications were



Figure 1 RP-HPLC profiles of *C. adamanteus* and *M. fulvius* venom samples. (A) Margres *et al.* (2014) identified 21 major peaks in the RP-HPLC analysis of 100 μ g of venom from a juvenile female *C. adamanteus* (72.4-cm snout-vent length, 77.5-cm total length) from population Ca2 in Figure 2. (B) We identified 25 major peaks in the RP-HPLC analysis of 100 μ g of venom from an adult female *C. adamanteus* (127-cm snout-vent length, 136-cm total length) from population Ca1 in Figure 2. Peaks 1b, 15b, 20b, and 20c were collected for mass spectrometry analysis. (C) We identified 22 major peaks in the RP-HPLC analysis of 25 μ g of venom from an adult female *M. fulvius* (62.0-cm snout-vent length), 68.5-cm total length) from population Mf2 used in the transcriptomic analysis of Margres *et al.* (2013).

accepted if they could be established at greater than 95% probability by the Scaffold local FDR algorithm, contained at least two identified peptides, possessed unique peptide evidence, and had a minimum of 20% sequence coverage. To



Figure 2 Large-scale sampling of *C. adamanteus* and *M. fulvius*. We collected venom samples from 65 adult *C. adamanteus* from seven putative populations and 49 adult *M. fulvius* from eight putative populations. AR, Apalachicola River; Ca, *C. adamanteus*; Mf, *M. fulvius*; SMR, Saint Mary's River; SR, Suwannee River.

identify the major toxins within each peak, only proteins with >4% of the total spectral matches within each peak were reported.

Reversed-phase high-performance liquid chromatography

RP-HPLC was performed on a Beckman System Gold HPLC (Beckman Coulter, Fullerton, CA) equipped with Beckman 32 Karat Software version 8.0 for peak quantification, as described by Margres *et al.* (2014), for 100 μ g of total protein for all *C. adamanteus* samples. For all *M. fulvius* samples, approximately 25 μ g of total protein was injected onto a Jupiter C18 narrow-bore column, 250 \times 2 mm (Phenomenex, Torrance, CA), using the standard solvent system, as described previously (Margres *et al.* 2014). After 5 min at 5% B, a linear gradient to 45% B was run over 120 min at a flow rate of 0.2 ml/min. Peaks were quantified as described by Margres *et al.* (2014).

Compositional analysis

Twenty-five RP-HPLC peaks per *C. adamanteus* venom sample and 22 RP-HPLC peaks per *M. fulvius* venom sample were quantified as described earlier (Figure 1). Briefly, relative amounts of individual peaks were determined by measuring the area under each peak relative to the total area of all identified peaks (Margres *et al.* 2014). According to the Lambert-Beer law, this relative amount corresponds to the percentage of total peptide bonds in the sample (McNaught

and Wilkinson 1997), and this measure has been shown to be a useful proxy for the relative amount of a specific protein by weight (Gibbs et al. 2009). However, this process generates compositional data, with each peak representing an individual component of the whole venom. Compositional data sets are subject to constant-sum constraints and are inherently biased toward negative correlations among components (Aitchison 1986). We used the robCompositions package (Templ et al. 2011) in R to transform the data prior to statistical analysis using either the centered logratio (clr) or isometric logratio (ilr) transformations (Egozcue et al. 2003; Filzmoser et al. 2009). The clr transformation takes the data from the simplex to real space, and because it retains the individual identities of the peaks, this transformation was used for visualization purposes (Figure 4 and Figure 5). The clr-transformed data, however, still suffer from a sum constraint because the components must add to zero. Therefore, we used the ilr transformation for all statistical analyses testing for differences among venom samples because this transformation does not suffer from the same sum constraint. The ilr transformation projects vectors in the simplex space into the real space using an orthonormal basis for the simplex and maintains all metric properties in the process. Zero values did occur during peak quantification, and we could not transform the data using logratios because we could not take the log of zero. Therefore, we treated zeros as trace values. Trace values represent a peak that was not completely absent but rather was not quantifiable as a result of the accuracy of the measurement process, and traces need to be positive values that are less than the smallest recorded value (i.e., 0.01% in our data set; Aitchison 1986).

We used the adonis function from the vegan package (Oksanen et al. 2013) in R and Euclidean distances to perform a permutational or nonparametric MANOVA (McArdle and Anderson 2001) on the ilr-transformed data to test for significant expression variation. The ilr transformation takes Aitchison distances in the simplex space to Euclidean distances in real space, enabling the application of standard multivariate statistical approaches (Egozcue et al. 2003; Aitchison and Egozcue 2005). Nonparametric tests do not assume that the data are normally distributed, as do traditional parametric tests (McArdle and Anderson 2001), and P-values were calculated under 100,000 permutations. To determine the sensitivity of the full nonparametric MANOVA analysis to the inclusion of trace values, as discussed earlier, we replaced all zeros in both data sets with three different values prior to normalization to percentages, transformation, and statistical analysis: 10^{-3} , 10^{-4} , and 10^{-6} . Analysis of all three trace-value data sets resulted in nearly identical results. We performed all remaining statistical analyses for both species using a trace value of 10^{-4} .

If the nonparametric MANOVA returned a significant result, we used the sample function in R to randomly permute population designations to identify which populations demonstrated significant expression variation. We used the ilr-transformed data and performed all pairwise comparisons in the context of the full model. First, the test statistic was obtained through a traditional parametric MANOVA in R. Population assignments then were permuted for the two populations of interest over 100,000 iterations, and a MANOVA was conducted following each permutation to create a null distribution of the test statistic. The test statistic calculated using the original unpermuted data then was compared with the null distribution. If the test statistic was within the one-tailed 95% C.I., the populations were not considered to exhibit significant expression variation and were combined into a single population. If the test statistic was outside the 95% C.I., the populations were considered distinct and retained.

If variation was detected, we performed a linear discriminant function analysis using the lda function in R on the ilr-transformed data to assess group membership placement probabilities across populations. The lda function tests for differences in the multivariate means, enabling group discrimination using the variables that capture the largest amount of variation in the data.

Sample collection

We collected venom samples from 65 adult C. adamanteus from seven putative populations and 49 adult M. fulvius from eight putative populations (Figure 2). Samples were collected across all months and seasons. C. adamanteus reaches sexual maturity at approximately 102-cm snout-vent length (Waldron et al. 2013), and M. fulvius becomes sexually mature at approximately 50-cm snout-vent length (Jackson and Franz 1981). All specimens used in our analyses met or exceeded these measurements. We limited our analyses to adults to avoid the potentially confounding effects of ontogenetic toxin expression variation, which has been documented previously (e.g., Mackessy 1988; Calvete et al. 2010; Durban et al. 2013). Individuals were grouped a priori into populations based on various biogeographic barriers (*i.e.*, major river drainages) and/or geographic distance. Because our randomization tests would combine populations with similar expression patterns but would not split one population with different expression patterns into multiple populations, we attempted to divide our samples into as many populations as possible while maintaining reasonable sample sizes (*i.e.*, \geq 4). Population designations were as follows: Ca1/Mf1, Florida panhandle west of the Apalachicola River; Ca2/Mf2, Florida panhandle and southern Georgia east of the Apalachicola River and west of the Suwannee River; Ca3, Brooksville Ridge of the Florida peninsula and one sample due east on the Atlantic Coastal Ridge; Mf3, northern peninsular Florida near Osceola National Forest; Ca4/Mf8, Everglades National Park; Mf4, northern peninsular Florida near Gainesville/ Ocala; Ca5, Little St. George Island; Mf5, southern region of the Brooksville Ridge of peninsular Florida; Ca6, Caladesi Island; Mf6, east coast of central Florida; Ca7, Sapelo Island, Georgia; and Mf7, southern peninsular Florida near Lake Okeechobee (Figure 2).

Results and Discussion

Constructing the genotype-phenotype map

Because the original venom-gland transcriptome assembly for C. adamanteus described by Rokyta et al. (2012) was found to contain a few chimeric toxin sequences, we performed a complete reanalysis with higher stringency and steps to eliminate chimeric sequences. The original data (NCBI Sequence Read Archive accession number SRA050594) consisted of 95,643,958 pairs of 101-nucleotide reads generated on an Illumina HiSeq 2000 with average Phred scores of 32. We merged 68,987,031 of these pairs on the basis of their 3' overlaps to give a set of reads with an average Phred score of 43 and an average length of 143 nucleotides. Transcriptome assembly and annotation resulted in 4,119 unique nontoxin transcripts with full-length coding sequences and 76 unique putative toxin transcripts (Figure 3, A and B). These 76 toxin transcripts were combined into 44 clusters with <1% nucleotide divergence in their coding sequences. Such clustering facilitates the estimation of transcript abundances and provides a better estimate of the number of toxin sequences.

The original proteomic analysis for *C. adamanteus* described by Margres *et al.* (2014) identified 40 toxic proteins that grouped into 20 unique clusters based on shared peptide evidence. Following reanalysis of the venom-gland transcriptome, we identified unique proteomic evidence for 21 of the 44 toxin clusters identified, including 15 of the 20 most highly expressed transcripts (Table 1, Supporting Information, Table S1, Figure 3B). These 21 clusters represented 10 of the 17 toxin gene families identified in the transcriptome, and we uniquely identified 22 individual transcripts (*i.e.*, a single allelic variant) in the 25 RP-HPLC peaks.

Reassembly of the venom-gland transcriptome for M. fulvius (Margres et al. 2013) yielded 1,270 unique nontoxin and 110 unique toxin transcripts. These toxin transcripts were combined into 44 clusters with <1% nucleotide divergence for analysis (Figure 3, C and D). The raw data (NCBI Sequence Read Archive accession number SRA062772) consisted of 79,573,048 pairs of 101-nucleotide reads generated on an Illumina HiSeq 2000 with an average Phred score of 32. We merged 59,414,816 of these pairs on the basis of their 3' overlaps to give composite reads with an average Phred score of 44 and an average length of 137 nucleotides. Only 1,269 of the 1,270 total unique nontoxins were included in Figure 3C because a single extremely lowabundance transcript was removed during abundance estimation because none of the 10 million reads used for estimating abundance mapped to this transcript.

We identified unique proteomic evidence for 29 of the 44 toxin clusters identified in the venom-gland transcriptome of *M. fulvius*, including 17 of the 20 most highly expressed transcripts. These 29 clusters represented 11 of the 15 toxin gene families identified in the transcriptome, and we uniquely identified 38 individual transcripts (*i.e.*, 9 allelic variants; Table S2) in the 22 RP-HPLC peaks (Figure 1C). Peptide reports for both species are listed in Table S3.

Our proteomic analyses often failed to detect proteins corresponding to the lower-abundance transcripts for both species (Figure 3, B and D), most likely reflecting a detection threshold and the large range of expression levels for toxin genes (*i.e.*, the signal for low-abundance proteins may be overwhelmed by that for high-abundance proteins). Additionally, certain protein classes expressed at appreciable levels in the venom gland may not encode true toxin proteins. These putative toxins were identified through homology with known toxins and may not encode true toxin proteins that are secreted into the venom. Venom is produced in the venom gland, but the venom gland is associated with a smaller gland downstream, the accessory gland. The function of the accessory gland is still uncertain, and Vonk et al. (2013) sequenced the venom-gland transcriptome and the accessory-gland transcriptome of the king cobra (Ophiophagus hannah). They found that most toxins are expressed at high levels only in the venom gland and at relatively low levels in the accessory gland. Lectins, however, were an exception, being much more highly expressed in the accessory gland than in the venom gland (Vonk et al. 2013). We identified 12 clusters of C-type lectins in the venom-gland transcriptome of C. adamanteus but detected only proteomic evidence for four of these proteins. C-type lectins were much less abundant in M. fulvius. We identified two clusters of C-type lectins in the *M. fulvius* transcriptome and verified one of the two transcript products proteomically. Vonk et al. (2013) argued that lectins may be deactivated in the venom gland. Although we found proteomic evidence for some of these proteins in the venoms of both species, certain lectin transcripts may not be true toxins or at least not significant components of the venoms.

Not all venoms vary geographically

We conducted nonparametric MANOVAs and, consistent with predictions for a locally adapted trait, detected significant expression variation among C. adamanteus populations (P = 0.002). Despite significant variation in C. adamanteus, we found no significant expression variation in *M*. fulvius at any spatial scale $(0.100 \le P \le 0.190)$, indicating that M. fulvius may not be engaged in a coevolutionary arms race with specific prey taxa, may have experienced a species-wide selective sweep (i.e., the rapid range-wide fixation of a specific phenotype as a result of selection), or has potentially undergone a recent range expansion. Intraspecific venom variation had been described previously as a ubiquitous phenomenon within snakes that would limit the efficacy of antivenoms (Chippaux et al. 1991; Casewell et al. 2014; Sunagar et al. 2014). However, our results demonstrate that intraspecific venom variation is not a valid a priori assumption, and bites from different geographic locations may not always result in different pathologies and degrees of neutralization by antivenom. Raw, untransformed data for both species are given in Table S4.



Figure 3 Venom-gland transcriptomes of *C. adamanteus* and *M. fulvius*. (A) A total of 4,119 non-toxin-encoding and 76 toxin-encoding transcripts were identified for *C. adamanteus*. Toxins were grouped into 44 clusters based on <1% nucleotide divergence. The inset shows a magnification of the top 200 transcripts, the vast majority of which code for toxins. (B) The pie chart shows that the *C. adamanteus* venom gland is biased toward toxin production, and the histogram shows expression levels of individual toxin clusters, color coded by toxin class. Toxins detected proteomically are indicated with asterisks. (C) A total of 1,270 non-toxin-encoding and 110 toxin-encoding transcripts were identified for *M. fulvius*. Toxins were grouped into 44 clusters, and most of the highly expressed transcripts coded for toxins. (D) The pie chart shows that the *M. fulvius* venom gland is extremely biased toward toxin production, and the histogram shows expression levels of individual toxin clusters, color coded by toxin class. Toxins detected proteomically are indicated with asterisks. BPP, bradykinin-potentiating and C-type natriuretic peptides; CTL, C-type lectin; CFVII, coagulation factor VII; CRISP, cysteine-rich secretory protein; HYAL, hyaluronidase; KUN, Kunitz-type protease inhibitor; KunKun, dual-domain Kunitz-type protease inhibitor; LAAO, L-amino acid oxidase; LCN, long-chain neurotoxin; MYO, myotoxin-A; NGF, nerve growth factor; NP, natriuretic peptide; NUC, nucleotidase; PDE, phosphodiesterase; PLA₂, phospholipase A₂; SVMPII, snake venom metalloproteinase type II; SVMPIII, snake venom metalloproteinase type III; SVMPIII, snake venom metalloproteinase; VEGF, vascular endothelial growth factor; VF, venom factor.

Pervasive geographic expression variation in C. adamanteus

To determine the extent of geographic expression variation and to identify which populations demonstrated significant expression variation in *C. adamanteus*, we performed pairwise randomization tests and determined that five of the seven *a priori C. adamanteus* populations had distinct patterns of expression. The three northernmost populations (Ca1, Ca2, and Ca7; Figure 2) were combined into a single population, but all other adult populations showed unique expression patterns (Figure 4). We updated population designations and conducted the nonparametric MANOVA with five populations (P = 0.001). We used a linear discriminant function analysis to assess group membership placement probabilities for the five adult populations, and the analysis accurately placed 79.31% of Ca1/2/7 individuals (formerly populations Ca1, Ca2, and Ca7; n = 29), 56.25% of Ca3 individuals (n = 16), 25.00% of Ca4 individuals (n = 4), 72.73% of Ca5 individuals (n = 11), and 60.00% of Ca6 individuals (n = 5). The lower percentages of the group membership placement probabilities (*i.e.*, <70%) may have been a function of unequal prior probabilities owing to large differences in sample size.

We compared the highest and lowest mean clr expression levels for individual peaks among each population and identified the most variable (maximum difference >2) and conserved (maximum difference <1) loci. Although these values are arbitrary, a maximum difference >2 indicates that the highest expression of the gene in one population is more than seven times greater than the lowest expression of the gene in another population relative to the geometric mean. Conversely,

Fraction	Protein name	% TSC	Seq coverage (%)
1a	None		_
1b	LAAO	27.92	27.7
	SVMPII-2b	15.56	16.2
	MYO-1a/b	10.07	58.3
	SVMPIII-5	7.32	30.8
	PLA ₂ -1a	5.03	47.5
	PDE	4.81	6.0
	SVMPII-1b/d	4.81	15.6
2	SVMPIII-2b	4.58	15.2
2	MYO-1b	100.00	52.1
3	SVMPIII-2b	100.00	18.3
4 F	SVIVIPIII-2D	100.00	22.5
5	SVIVIPIII-ZD	100.00	10.1
7	CPISP	06.22	
/ 0		90.23	20.8
0	CRISP	19.36	17.6
Q	None	15.50	17.0
10	PLA ₂ =1a/b	100.00	91.8
10	SV/SP-2	86.39	25.7
	SVSP-5	13 61	17 5
12	SVSP-4	98.77	59.0
13	SVSP-7	76.08	55.0
15	SVSP-4	23.92	31.1
14	CTL-10	68.02	69.2
	SVSP-4	20.62	38.5
	CTL-11a/b/c	7.28	52.3
15a	CTL-10	42.02	58.1
	CTL-9	39.21	61.3
	SVSP-4	11.40	38.5
15b	CTL-10	27.63	100.0
	CTL-13f	21.48	96.0
	SVSP-7	7.33	89.0
	LAAO	7.16	70.7
	SVSP-4	6.65	85.2
	CTL-11a/b	5.02	99.2
16	LAAO	45.21	39.8
	SVSP-7	25.92	55.1
	SVSP-4	20.27	45.1
	CTL-11a/b/c	4.67	32.8
17	SVSP-4	49.68	45.1
	SVSP-7	25.98	42.9
	NUC	16.98	6.6
10		5.70	10.7
18	SVIVIPIII-40/e	98.01	29.9
19	SVIVIPIII-5	80.80 12.84	43.0
202	CIL-IId SVMDUL26	12.04 62.65	05.5
20a	SVMPIL-2c/d	17.99	30.8
	SVMPIL26	17.55	30.8
20h	SVMPII-2b	17.47	52.8
200		17.56	87.6
	CTI-10	8.07	100.0
	CTL-11a/b	5.36	99.2
	SVMPIII-2b	4.96	46.9
	CTL-13f	4.75	100.0
	SVSP-4	4.51	96.3
20c	LAAO	17.24	76.7
	SVMPII-2b	9.40	42.1
	CTL-11a/b	7.03	99.2
	SVMPII-1b/d	6.73	43.3
	CTL-10	5.09	94.0
21	SVMPII-1a/b/c/d	94.78	46.9

Table 1 Tandem mass spectrometry analysis of RP-HPLC fractions in *C. adamanteus*

Toxin identifications based on peptide spectral evidence of isolated peaks from RP-HPLC. CTL, C-type lectin; CRISP, cysteine-rich secretory protein; HYAL, hyaluronidase; LAAO, L-amino-acid oxidase; MYO, myotoxin; NUC, nucleotidase; PDE, phosphodiesterase; PLA₂, phospholipase A₂; SVMP, snake venom metalloproteinase (types II and III); SVSP, snake venom serine proteinase; TSC = total spectrum count.

172 M. J. Margres et al.

a maximum difference <1 indicates that the highest expression of the gene in one population is approximately 2.7 times greater than the lowest expression of the gene in another population relative to the geometric mean. Eight peaks (2, 4, 5, 6, 9, 12, 17, and 20a) representing seven loci and the only two unidentified proteins (peaks 6 and 9) were extremely variable and had a difference >2, three peaks (7, 10, and 15a) corresponding to five loci were conserved and had a difference <1, and 14 peaks representing 10 loci had a difference between 1 and 2.

The genotype-phenotype map allowed us to identify the specific loci and mechanism driving venom divergence among adult C. adamanteus. For example, myotoxin (peak 2) and one of the two unidentified proteins (peak 6) were expressed at their highest levels in Ca1/2/7 and Ca3 (Figure 4), the northernmost populations, and were completely absent in Ca6, a southern population on Caladesi Island (Figure 2). This expression dichotomy had been documented previously in myotoxin (Straight et al. 1991), and we demonstrated that parallel regulatory changes affecting other genes are also occurring along this latitudinal gradient. Although expressing myotoxin at a relatively high level, Ca3 downregulated seven other loci (snake venom type II metalloproteinase 2; snake venom type III metalloproteinase 2; L-amino acid oxidase; C-type lectins 10, 11, and 13; and snake venom serine proteinase 4, peak 20b) in comparison with all other populations. Snake venom serine proteinase 4 (peak 12) and snake venom type II metalloproteinase 2 and snake venom type III metalloproteinase 2 (peak 20a) were downregulated in Ca4. Snake venom serine proteinases 4 and 7, nucleotidase, and hyaluronidase (peak 17) were expressed at relatively low levels in Ca5, while snake venom type II metalloproteinase 1 (peak 21) was significantly upregulated in comparison to all other populations. The variation in expression of these loci between populations was significant and affected this highly polygenic, complex phenotype in a tractable manner (Fraser 2013), providing candidate genes for which to test the functional and evolutionary significance of this variation (e.g., promoter evolution).

Casewell *et al.* (2014) recently suggested that venom variation can be influenced not only by regulatory changes but also by postgenomic effects, including post-translational modifications to specific toxins. We detected the same toxic protein, snake venom type III metalloproteinase 2, in multiple RP-HPLC fractions (Table 1, Table S1), suggesting several different protein products following proteolysis (Casewell *et al.* 2014). This toxin was the only protein identified in peaks 4 and 5 (Table 1, Table S1). Peak 4 was expressed at its highest level in Ca4 and lowest level in Ca6, the two southernmost populations (Figure 2), while the opposite was found in peak 5. This result demonstrates that posttranslational modifications of a single gene product also can contribute to rapid venom divergence.

The pervasive expression variation we identified not only demonstrates the importance of gene regulation in the evolution of venoms but also has significant conservation



Figure 4 Statistical analyses identified extensive geographic expression variation in *C. adamanteus*. Following randomization tests, Ca1, Ca2, and Ca7 from Figure 2 were combined into a single population (represented by the white bar), but all other populations possessed unique expression patterns. The mean centered log ratio expression levels for individual peaks (excluding peak 6) for each *C. adamanteus* population are shown. Peak 6 was included in all analyses but was excluded from the figure for aesthetic purposes. Every individual in Ca6 lacked this protein, and taking the log of the extremely small trace value resulted in a large negative number. Colors correspond to Figure 2. Clr, centered log ratio.

implications. *C. adamanteus*, as a result of human persecution and habitat fragmentation, is currently a candidate for federal listing as threatened under the Endangered Species Act (United States Fish and Wildlife Service 2012). Traditional conservation genetics aims to identify and conserve genetically distinct populations based on neutral markers. Neutral markers, however, can leak across strong selective boundaries, and if selection is driving venom divergence, we might expect divergence in this phenotype to precede neutral divergence among populations. All five distinct populations identified in this study should be involved, in conjunction with sequence data and population genetics estimates, with the identification and designation of evolutionarily significant units because the loss of any of these populations would result in the loss of a unique venom phenotype.

Expression variation in both high- and low-abundance proteins

Gene expression is a metabolically costly process, and selection will drive expression to an optimal level at which its cost is balanced by its fitness benefit (Gout et al. 2010). Proteins expressed at low levels are predicted to exhibit patterns of rapid adaptive evolution and influence complex trait variation more than highly expressed proteins because highly expressed proteins are thought to evolve under strong purifying selection and have relatively conserved expression levels (Gout et al. 2010). Therefore, abundant toxins should exhibit less variation in expression as these generic proteins may perform nonspecific prey functions, while less abundant proteins may be prey specific, thus explaining their presence or absence (Gibbs et al. 2009). To determine whether variation among adult C. adamanteus was restricted to lowabundance proteins and highly expressed loci were conserved, we divided the 25 RP-HPLC peaks into low-abundance (12 peaks) and high-abundance (13 peaks) data sets based on the clr mean for each peak and conducted a nonparametric MANOVA. We detected significant expression variation in both low-abundance ($P < 10^{-3}$) and high-abundance ($P < 10^{-3}$) data sets, indicating that loci expressed at all levels contributed to geographic divergence among adult C. adamanteus.

Our results contradict theoretical predictions that highly expressed genes have more conserved expression patterns and evolve more slowly than genes expressed at lower levels (Gout *et al.* 2010), potentially as a result of the tissue-specific expression of toxin genes. The modularity of the venom system may increase the evolvability of this complex phenotype, allowing for extreme levels of expression variation in both highly and weakly expressed loci.

Conserved expression patterns in M. fulvius

Despite the lack of significant expression variation in M. fulvius, we also compared the highest and lowest mean clr expression levels for individual peaks among each population and identified the most variable (maximum difference >2) and conserved (maximum difference <1) loci, as described earlier. Mf1 (Figure 2) was excluded from this analysis because of its extremely small sample size (n = 2). Consistent with the lack of significant variation identified in our previous analysis, 11 peaks (1, 4b, 5, 6a, 6d, 7, 11, 12, 15a, 15b, and 16) representing the protein products of 22 loci were conserved and had a difference <1, 10 peaks (2a, 2b, 3, 4a, 6b, 6c, 8, 9, 10, and 13) representing 18 loci had a difference between 1 and 2, and only a single peak (peak 14) was variable and had a difference >2 (Figure 5). Peak 14 contained the protein products of five loci (phospholipases A2-2, -3, and -7 and three-finger toxins-9 and -10). Phospholipases A2-2 and -3 were the two most highly expressed transcripts in the venom-gland transcriptome, and all five loci are in the top 13 most highly expressed toxins. However, all these proteins are present in multiple peaks, again suggesting that the variation identified may be restricted to specific post-translationally modified protein products or complexes (Bohlen et al. 2011; Casewell et al. 2014).

As discussed earlier, *M. fulvius* envenomations can be lethal, and antivenom is currently unavailable in the United States (Norris *et al.* 2009). Intraspecific venom variation has been shown to limit the efficacy of antivenoms (Chippaux *et al.* 1991; Casewell *et al.* 2014; Sunagar *et al.* 2014), but our results demonstrate that intraspecific venom variation is not as pervasive as previously believed. The paucity of



Figure 5 Lack of geographic expression variation in *M. fulvius*. Most peaks and their corresponding loci were conserved across all populations, with peak 14 (5 loci) being the only peak that exhibited extensive variation (*i.e.*, maximum difference >2). The mean centered log ratio expression levels for individual peaks for each *M. fulvius* population (excluding Mf1) are shown. Mf1 was excluded because of its small sample size (n = 2). Colors correspond to Figure 2. Clr, centered log ratio.

expression variation in *M. fulvius* suggests that a single anitvenom would effectively neutralize bites from different geographic locations, providing critical information for the development of an effective antivenom.

Extremes of toxin expression variation in sympatric species

Comparative studies of codistributed species can improve our understanding of the processes driving patterns of phenotypic diversity. We compared phenotypic distributions for independent species across the same landscape to determine whether selection or other factors were responsible for the observed patterns (Gomulkiewicz et al. 2007). A direct link between the expression level of a locus and fitness has been documented in other organisms (Gout et al. 2010). This relationship should be even more pronounced in venoms because expression variation directly changes the phenotype (*i.e.*, relative amounts of venom components should determine, at least in part, the venom's efficacy). We identified extensive expression variation in C. adamanteus and found no variation in M. fulvius. These incongruent patterns suggest that phenotypic diversification in C. adamanteus was driven by selection, potentially resulting from dietary differences or local coevolutionary interactions with prey (Biardi et al. 2005), because locally beneficial traits are not expected to become fixed at the species level owing to genotype-by-genotype-by-environment interactions. The monomorphic venom phenotype in M. fulvius suggests that a recent range expansion and/or a species-wide selective sweep rather than reciprocal selection from specific prey taxa and/or dietary differences dictated the evolution of this complex trait.

Conclusion

We showed that regulatory changes at specific loci, potentially in response to coevolutionary interactions, can lead to rapid divergent phenotypic evolution in a polygenic complex trait. The lack of expression variation in *M. fulvius* demonstrated that intraspecific venom variation is not as pervasive as previously believed, provided critical information for the development of an effective and currently unavailable antivenom, and highlighted the significance of the venom differences identified in *C. adamanteus*, a species currently under consideration for listing as threatened under the Endangered Species Act (United States Fish and Wildlife Service 2012). We identified distinct populations of *C. adamanteus* based on an ecologically critical trait for this species, and the conservation of these populations will preserve unique venom phenotypes.

Acknowledgments

The authors thank Karalyn Aronow, Richard Bartlett, Jeffrey Fobb, Nathanael Herrera, Pierson Hill, Chip Leavine, Charles Looney, Jacob Loyacano, John Malone, Moses Michelson, Michael Rochford, Joe Pfaller, Jordan Sirosky, Jim Mendenhall, and Lora Smith and Jennifer Howze at the Joseph W. Jones Ecological Research Center at Ichuaway for help in acquiring venom samples. The authors also thank Megan Lamb, Danielle Jones, and Jennifer Wanat with the Florida Department of Environmental Protection and the Apalachicola River National Estuarine Research Reserve, Bradley Smith and Shelley Stiaes of the United States Fish and Wildlife Service and St. Vincent National Wildlife Refuge, Dorset Hurley and the Sapelo Island National Estuarine Research Reserve, Brett Williams and the Department of the Air Force, and Peter Krulder and Carl Calhoun at Caladesi Island State Park for access to field sites. We thank the College of Medicine Translational Science Laboratory at Florida State University for their assistance in collecting and analyzing the mass spectrometry data and Brian Inouye for his recommendations regarding statistical analyses. We also thank the laboratory of Lisle Gibbs at Ohio State University for providing comments. Samples were collected under the following permits: Florida Fish and Wildlife Conservation Commission (FWC) LSSC-13-00004 and LSSC-09-0399, Eglin Air Force Base 96 SFS/ SFOIRP, Everglades National Park EVER-2012-SCI-0053, Florida Department of Environmental Protection permit number 04101310, and St. Vincent National Wildlife Refuge permit number 41650-2012-08. Sample collection was approved by the Florida State University Institutional Animal Care and Use Committee (IACUC) under protocol 0924. Annotated transcriptome sequences were submitted to the GenBank Transcriptome Shotgun Assembly database under accession number GBEX01000000 for *C. adamanteus* and GBEW01000000 for *M. fulvius*. This work was supported by the National Science Foundation (DEB-1145987 to D.R.R.).

LITERATURE CITED

- Abzhanov, A., M. Protas, B. Grant, P. Grant, and C. Tabin, 2004 Bmp4 and the morphological variation of beaks in Darwin's finches. Science 305: 1462–1465.
- Aitchison, J., 1986 The Statistical Analysis of Compositional Data, Chapman and Hall, London.
- Aitchison, J., and J. J. Egozcue, 2005 Compositional data analysis: where are we and where should we be heading? Math. Geol. 37: 829–850.
- Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak, 2004 Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340: 783–795.
- Biardi, J. E., D. C. Chien, and R. G. Coss, 2005 California ground squirrel (*Spermophilus beecheyi*) defenses against rattlesnake venom digestive and hemostatic toxins. J. Chem. Ecol. 31: 2501–2518.
- Bohlen, C. J., A. T. Chesler, R. Sharif-Naeini, K. F. Medzihradszky, S. Zhou *et al.*, 2011 A heteromeric Texas coral snake toxin targets acid-sensing ion channels to produce pain. Nature 479: 410–414.
- Boldrini-França, J., C. Corrêa-Netto, M. M. S. Silva, R. S. Rodrigues, P. D. L. Torre *et al.*, 2010 Snake venomics and antivenomics of *Crotalus durissus* subspecies from Brazil: assessment of geographic variation and its implication on snakebite management. J. Proteomics 73: 1758–1776.
- Calvete, J. J., L. Sanz, P. Cid, P. de la Torre, M. Flores-Díaz et al., 2010 Snake venomics of the Central American rattlesnake *Crotalus simus* and the South American *Crotalus durissus* complex points to neurotoxicity as an adaptive paedomorphic trend along *Crotalus* dispersal in South America. J. Proteome Res. 9: 528–544.
- Carroll, S., 2008 Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134: 25– 36.
- Casewell, N., S. Wagstaff, W. Wüster, D. Cook, F. Bolton *et al.*, 2014 Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. Proc. Natl. Acad. Sci. USA 111: 9205–9210.
- Chippaux, J., V. Williams, and J. White, 1991 Snake venom variability: methods of study, results and interpretation. Toxicon 29: 1279–1303.
- Chou, H.-H., H.-C. Chiu, N. F. Delaney, D. Segrè, and C. J. Marx, 2011 Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science 332: 1190–1192.
- Creer, S., A. Malhotra, R. S. Thorpe, R. Stöcklin, P. Favreau *et al.*, 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, J. C., W. Wüster, and R. S. Thorpe, 1996 Diet and snake venom evolution. Nature 379: 537–540.
- Durban, J., A. Perez, L. Sanz, A. Gomez, F. Bonilla et al., 2013 Integrated "omics" profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, *Crotalus simus simus*. BMC Genomics 14: 234.
- Egozcue, J. J., V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal, 2003 Isometric logratio transformations for compositional data analysis. Math. Geol. 35: 279–300.
- Filzmoser, P., K. Hron, and C. Reimann, 2009 Principal component analysis of compositional data with outliers. Environmetrics 20: 621–632.
- Fraser, H., 2013 Gene expression drives local adaptation in humans. Genome Res. 23: 1089–1096.
- Gibbs, H. L., L. Sanz, and J. J. Calvete, 2009 Snake population venomics: proteomics-based analyses of individual variation reveals significant gene regulation effects on venom protein expression in *Sistrurus* rattlesnakes. J. Mol. Evol. 68: 113–125.
- Gomulkiewicz, R., D. Drown, M. Dybdahl, W. Godsoe, S. Nuismer *et al.*, 2007 Dos and don'ts of testing the geographic mosaic theory of coevolution. Heredity 98: 249–258.
- Gout, J.-F., D. Kahn, and L. P. P.-G. C. Duret, 2010 The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. PLoS Genet. 6: e1000944.
- Hansen, T. F., 2006 The evolution of genetic architecture. Annu. Rev. Ecol. Evol. Syst. 37: 123–157.
- Hill, W., and X.-S. Zhang, 2012 Assessing pleiotropy and its evolutionary consequences: pleiotropy is not necessarily limited, nor need it hinder the evolution of complexity. Nat. Rev. Genet. 13: 296.
- Hoekstra, H., and J. Coyne, 2007 The locus of evolution: evo devo and the genetics of adaptation. Evolution 61: 995–1016.
- Jackson, D., and R. Franz, 1981 Ecology of the eastern coral snake (*Micrurus fulvius*) in northern peninsular Florida. Herpetologica 37: 213–228.
- Klauber, L. M., 1997 Rattlesnakes: Their Habits, Life Histories, and Influence on Mankind, Ed. 2. University of California Press, Berkeley, CA.
- Mackessy, S. P., 1988 1988 Venom ontogeny in the Pacific rattlesnakes *Crotalus viridis helleri* and *C. v. oreganus*. Copeia No. 1: 92–101.
- Mackessy, S. P., 2010 Evolutionary trends in venom composition in the western rattlesnakes (*Crotalus viridis* sensu lato): toxicity vs. tenderizers. Toxicon 55: 1463–1474.
- Manceau, M., V. Domingues, R. Mallarino, and H. Hoekstra, 2011 The developmental role of Agouti in color pattern evolution. Science 331: 1062–1065.
- Margres, M. J., K. Aronow, J. Loyacano, and D. R. Rokyta, 2013 The venom-gland transcriptome of the eastern coral snake (*Micrurus fulvius*) reveals high venom complexity in the intragenomic evolution of venoms. BMC Genomics 14: 531.
- Margres, M. J., J. J. McGivern, K. P. Wray, M. Seavy, K. Calvin *et al.*, 2014 Linking the transcriptome and proteome to characterize the venom of the eastern diamondback rattlesnake (*Crotalus adamanteus*). J. Proteomics 96: 145–158.
- McArdle, B., and M. Anderson, 2001 Fitting multivariate models to community data: a comment on distance-based redundancy analysis. Ecology 82: 290–297.
- McNaught, A., and A. Wilkinson, 1997 Compendium of Chemical Terminology: IUPAC Recommendations, Vol. 1669. Blackwell Science, Oxford, UK.
- Muller, G., 2007 Evo-devo: extending the evolutionary synthesis. Nat. Rev. Genet. 8: 943–949.
- Norris, R. L., R. R. Pfalzgraf, and G. Laing, 2009 Death following coral snake bite in the United States—first documented case (with ELISA confirmation of envenomation) in over 40 years. Toxicon 53: 693–697.

- Oksanen, J., F. Blanchet, R. Kindt, P. Legendre, P. Minchin, *et al.*, 2013 vegan: Community Ecology Package. R package version 2.0-7. Available at: http://cran.r-project.org/.
- Palmer, W. M., and A. L. Braswell, 1995 *Reptiles of North Carolina*, University of North Carolina Press, Chapel Hill.
- Pavlicev, M., E. Norgard, G. Fawcett, and J. Cheverud, 2011 Evolution of pleiotropy: epistatic interaction pattern supports a mechanistic model underlying variation in genotype-phenotype map. J. Exp. Zool. B Mol. Dev. Evol. 316: 371–385.
- Pawlowsky-Glahn, V., and A. Buccianti, 2011 robCompositions: an R-package for robust statistical analysis of compositional data, pp. 341–355 in Compositional Data Analysis: Theory and Applications, edited by M. Templ, K. Hron, and P. Filzmoser. John Wiley & Sons, Chichester, UK.
- Pond, S., D. Posada, M. Gravenor, C. Woelk, and S. Frost, 2006 Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. 23: 1891–1901.
- Reed, R., R. Papa, A. Martin, H. Hines, B. Counterman *et al.*, 2011 Optix drives the repeated convergent evolution of butterly wing pattern mimicry. Science 333: 1137–1141.
- Rokyta, D. R., C. J. Beisel, P. Joyce, M. T. Ferris, C. L. Burch et al., 2008 Beneficial fitness effects are not exponential for two viruses. J. Mol. Evol. 67: 368–376.
- Rokyta, D. R., Z. Abdo, and H. A. Wichman, 2009 The genetics of adaptation for eight microvirid bacteriophages. J. Mol. Evol. 69: 229–239.
- Rokyta, D. R., K. P. Wray, A. R. Lemmon, E. M. Lemmon, and S. B. Caudle, 2011a A high-throughput venom-gland transcriptome for the eastern diamondback rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes. Toxicon 57: 657–671.
- Rokyta, D. R., P. Joyce, S. B. Caudle, C. Miller, C. J. Beisel *et al.*, 2011b Epistasis between beneficial mutations and the phenotypeto-fitness map for a ssDNA virus. PLoS Genet. 7: e1002075.
- Rokyta, D. R., A. R. Lemmon, M. J. Margres, and K. Aronow, 2012 The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genomics 13: 312.
- Rokyta, D. R., K. P. Wray, and M. J. Margres, 2013 The genesis of an exceptionally deadly venom in the timber rattlesnake (*Crotalus horridus*) revealed through comparative venom-gland transcriptomics. BMC Genomics 14: 394.

- Romero, I., I. Ruvinsky, and Y. Gilad, 2012 Comparative studies of gene expression and the evolution of gene regulation. Nat. Rev. Genet. 13: 505–516.
- Savolainen, O., M. Lascoux, and J. Merila, 2013 Ecological genomics of local adaptation. Nat. Rev. Genet. 14: 807–820.
- Stadler, B. M. R., P. F. Stadler, G. P. Wagner, and W. Fontana, 2001 The topology of the possible: formal spaces underlying patterns of evolutionary change. J. Theor. Biol. 213: 241– 274.
- Stern, D. L., 2000 Perspective: evolutionary developmental biology and the problem of variation. Evolution 54: 1079–1091.
- Straight, R. C., J. L. Glenn, T. B. Wolt, and M. C. Wolfe, 1991 Regional differences in content of small basic peptide toxins in the venoms of *Crotalus adamanteus* and *Crotalus horridus*. Comp. Biochem. Physiol. B 100: 51–58.
- Sunagar, K., E. Undheim, H. Scheib, E. Gren, C. Cochran et al., 2014 Intraspecific venom variation in the medically significant southern Pacific rattlesnake (*Crotalus oreganus helleri*): biodiscovery, clinical and evolutionary implications. J. Proteomics 99: 68–83.
- Thompson, J. N., 2005 *The Geographic Mosaic of Coevolution*, University of Chicago Press, Chicago.
- United States Fish and Wildlife Service, 2012 Endangered and threatened wildlife and plants: 90-day finding on a petition to list the eastern diamondback rattlesnake as threatened. Fed. Regist. 77: 27403–27411.
- Vonk, F., N. Casewell, C. Henkel, A. Heimberg, H. Jansen *et al.*, 2013 The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. Proc. Natl. Acad. Sci. USA 110: 20651–20656.
- Wager, A., 2008 Robustness and evolvability: a paradox resolved. P. Roy. Soc. Lond. B Biol. Sci. 275: 91–100.
- Waldron, J. L., S. M. Welch, S. H. Bennett, W. G. Kalinowsky, and T. A. Mousseau, 2013 Life history constraints contribute to the vulnerability of a declining North American rattlesnake. Biol. Conserv. 159: 530–538.
- Woods, R. J., J. E. Barrick, T. F. Cooper, U. Shrestha, M. R. Kauth et al., 2011 Second-order selection for evolvability in a large *Escherichia coli* population. Science 331: 1433–1436.

Communicating editor: J. J. Bull

GENETICS

Supporting Information http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.172437/-/DC1

Contrasting Modes and Tempos of Venom Expression Evolution in Two Snake Species

Mark J. Margres, James J. McGivern, Margaret Seavy, Kenneth P. Wray, Jack Facente, and Darin R. Rokyta

> Copyright © 2015 by the Genetics Society of America DOI: 10.1534/genetics.114.172437

Fraction	Protein name	Exclusive UPC	Exclusive USC	TSC	% TSC	Seq Coverage (%)
1a	None	_	_	_	—	_
1b	LAAO	26	35	122	27.92	27.7
	SVMPII-2b	12	17	68	15.56	16.2
	MYO-1a/b	4	5	44	10.07	58.3
	SVMPIII-5	14	17	32	7.32	30.8
	PLA_2-1a	6	8	22	5.03	47.5
	PDE	4	9	21	4.81	6.0
	SVMPII-1b/d	6	8	21	4.81	15.6
	SVMPIII-2b	7	8	20	4.58	15.2
2	MYO-1b	3	4	586	100.00	52.1
3	SVMPIII-2b	10	15	816	100.00	18.3
4	SVMPIII-2b	11	18	1073	100.00	22.5
5	SVMPIII-2b	7	9	287	100.00	16.1
6	None	_	_	_	_	—
7	CRISP	14	18	1403	96.23	79.2
8	SVSP-5	5	5	579	80.64	30.8
	CRISP	3	3	139	19.36	17.6
9	None	_	_	_	—	_
10	PLA_2-1a/b	3	3	4040	100.00	91.8
11	SVSP-2	2	2	146	86.39	25.7
	SVSP-5	1	1	23	13.61	17.5
12	SVSP-4	7	9	2335	98.77	59.0
13	SVSP-7	4	6	706	76.08	55.1
	SVSP-4	3	4	222	23.92	31.1
14	CTL-10	2	2	1270	68.02	69.2
	SVSP-4	3	3	385	20.62	38.5
	CTL-11a/b/c	3	3	136	7.28	52.3
15a	CTL-10	1	1	1198	42.02	58.1
	CTL-9	1	1	1118	39.21	61.3
	SVSP-4	3	3	325	11.40	38.5
15b	CTL-10	26	55	1420	27.63	100.0
	CTL-13f	26	66	1104	21.48	96.0
	SVSP-7	26	37	377	7.33	89.0
	LAAO	51	101	368	7.16	70.7
	SVSP-4	17	32	342	6.65	85.2
	CTL-11a/b	12	22	258	5.02	99.2

Table S1: Tandem mass spectrometry analysis of the reversedphase high-performance liquid chromatography fractions in *Crotalus adamanteus*.

Fraction	Protein name	Exclusive UPC	Exclusive USC	TSC	% TSC	Seq Coverage (%)
16	LAAO	10	12	368	45.21	39.8
	SVSP-7	4	6	211	25.92	55.1
	SVSP-4	4	4	165	20.27	45.1
	CTL-11a/b/c	2	2	38	4.67	32.8
17	SVSP-4	4	4	392	49.68	45.1
	SVSP-7	2	2	205	25.98	42.9
	NUC	2	2	134	16.98	6.6
	HYAL	2	2	45	5.70	10.7
18	SVMPIII-4d/e	14	16	4627	98.01	29.9
19	SVMPIII-3	17	19	1707	80.86	43.6
	CTL-11a	4	4	271	12.84	63.3
20a	SVMPIII-2b	13	15	1553	62.65	35.9
	SVMPII-2c/d	1	2	446	17.99	30.8
	SVMPII-2b	1	4	433	17.47	30.8
20b	SVMPII-2b	48	91	648	17.38	52.8
	LAAO	80	138	472	12.66	87.6
	CTL-10	26	48	301	8.07	100.0
	CTL-11a/b	14	28	200	5.36	99.2
	SVMPIII-2b	31	63	185	4.96	46.9
	CTL-13f	26	54	177	4.75	100.0
	SVSP-4	16	33	168	4.51	96.3
20c	LAAO	78	145	691	17.24	76.7
	SVMPII-2b	36	62	377	9.40	42.1
	CTL-11a/b	16	26	282	7.03	99.2
	SVMPII-1b/d	l 21	36	2	6.7 6.7	3 43.3
	CTL-10	21	36	2	5.0	9 94.0
21	SVMPII-1a/b/c	z/d 14	17	3	141 94.7	78 46.9

Table S1: Toxin identifications based on peptide spectral evidence of isolated peaks from reversed-phase high-performance liquid chromatography (RP-HPLC). Variable abbreviations: UPC = Unique Peptide Count; USC = Unique Spectrum Count; TSC = Total Spectrum Count. Toxin abbreviations: CTL, C-type lectin; CRISP, Cysteine-rich secretory protein; HYAL, Hyaluronidase; LAAO, L-amino-acid oxidase; MYO, Myotoxin; NUC, Nucleotidase; PDE, Phosphodiesterase; PLA₂, Phospholipase A₂; SVMP, Snake venom metalloproteinase (types II and III); SVSP, Snake venom serine proteinase.

Table S2: Tandem mass spectrometry analysis of the reversed-phase high-performance liquid chromatography fractions in *Micrurus ful*vius.

Fraction	Protein name	Exclusive UPC	Exclusive USC	TSC	$\%~{\rm TSC}$	Seq Coverage (%)
1	3FTx-8a	1	16	147	47.88	85.2
	KUN-4	2	3	65	21.17	81.4
	KUN-1	3	3	57	18.57	73.3
	KUN-5	3	3	18	5.86	49.2
2	KUN-5	5	5	97	48.50	72.9
	3FTx-5	9	17	50	25.00	59.0
	KUN-4	1	1	21	10.50	72.9
	3FTx-8a	3	4	12	6.00	67.2
	KUN-2a	3	5	10	5.00	52.5
3	LCN-2a	4	6	113	57.36	80.3
	KUN-5	3	3	22	11.17	49.2
	3FTx-5	7	10	20	10.15	59.0
	KUN-2a	3	6	13	6.60	72.9
	KUN-3	3	7	12	6.09	74.6
	3FTx-9b	2	3	12	6.09	81.0
4	LCN-2a	4	6	37	45.68	56.1
	3FTx-9b	2	5	16	19.75	81.0
	KUN-3	2	6	9	11.11	50.8
	KUN-2a	4	5	9	11.11	76.3
	KUN-5	2	2	4	4.94	44.1
5	3FTx-9b	3	18	627	39.43	81.0
	KUN-6a/b	6	16	221	13.90	84.7
	3FTx-9c	1	3	185	11.64	58.7
	LCN-1	7	16	182	11.45	65.2
	KUN-3	5	20	152	9.56	81.4
	KUN-2a	6	14	100	6.29	89.8
6	3FTx-6b	5	11	291	21.91	89.7
	3FTx-11a	6	12	269	20.26	71.6
	3FTx-9b	3	16	218	16.42	81.0
	3FTx-11c	1	5	167	12.58	65.7
	KUN-3	5	18	99	7.45	81.4
	3FTx-13	4	8	91	6.85	100.0
7	PLA_2-4b	3	8	946	28.05	82.9
	PLA_2-6b	7	15	598	17.73	81.3
	PLA_2-3b	4	8	378	11.21	89.4
	PLA ₂ -2aa	1	1	285	8.45	69.9

Fraction	Protein name	Exclusive UPC	Exclusive USC	TSC	% TSC	Seq Coverage (%)
	PLA_2-2v	1	1	278	8.24	82.1
	3FTx-11c	3	9	206	6.11	77.6
	PLA_2-7a	7	10	173	5.13	76.8
8	3FTx-9b	3	16	263	34.07	81.0
	PLA_2-7a	1	22	147	19.04	87.2
	3FTx-11a	6	11	118	15.28	67.2
	3FTx-6b	4	8	49	6.35	89.7
	KUN-3	3	12	46	5.96	57.6
	3FTx-13	3	7	44	5.70	84.5
9	PLA_2-8b	7	15	829	38.43	88.8
	PLA_2-4b	2	3	389	18.03	74.8
	PLA_2-3b	4	8	155	7.19	89.4
	PLA_2 -2ad	1	1	130	6.03	65.0
	PLA_2-2h	2	3	122	5.66	70.7
	PLA_2-6b	5	12	120	5.56	75.6
	3FTx-11c	3	6	98	4.54	65.7
10	3FTx-9b	3	16	204	50.50	81.0
	3FTx-11a	5	9	80	19.80	59.7
	PLA_2-7a	7	9	29	7.18	76.8
	KUN-3	3	8	19	4.70	57.6
11	PLA_2-2m	2	2	1991	19.80	88.6
	PLA_2-2v	1	1	1718	17.09	88.6
	PLA_2 -2ad	1	1	1707	16.98	61.8
	PLA_2-3b	5	17	1484	14.76	98.2
	PLA_2-2r	1	1	1238	12.31	86.2
	PLA_2 -2aa	2	3	720	7.16	77.2
12	PLA ₂ -2h	3	9	1000	20.69	77.2
	PLA_2-3b	4	8	800	16.55	96.5
	PLA_2 -2aa	4	15	734	15.19	77.2
	PLA_2-2m	2	2	721	14.92	88.6
	PLA_2-2v	3	4	525	10.86	88.6
	PLA_2 -7a	4	5	291	6.02	63.2
	PLA_2-6b	1	1	288	5.96	35.0
	PLA_2-4b	1	1	288	5.96	51.2
13	PLA_2-2h	3	8	302	26.24	70.7
	PLA_2 -2aa	4	10	197	17.12	70.7
	PLA_2-3b	2	3	159	13.81	89.4
	3FTx-10a	9	12	152	13.21	78.8
	PLA_2-2v	1	1	79	6.86	78.0
	3FTx-9b	3	10	63	5.47	81.0

Fraction	Protein name	Exclusive UPC	Exclusive USC	TSC	% TSC	Seq Coverage $(\%)$
14	PLA_2-2h	3	8	126	24.75	87.0
	PLA_2-3b	4	4	97	19.06	89.4
	PLA_2 -2aa	3	7	94	18.47	87.0
	3FTx-9b	3	11	69	13.56	54.0
	3FTx-10a	3	4	32	6.29	37.9
	PLA_2-7a	3	4	24	4.72	52.8
15a	3FTx-9b	2	9	81	33.89	54.0
	PLA_2-2h	3	7	56	23.43	70.7
	PLA_2 -2aa	2	2	38	15.90	70.7
	3FTx-11a	5	6	20	8.37	46.3
	PLA_2-7a	3	3	16	6.69	52.8
15b1	CTL-2a	21	42	165	21.91	97.8
	PLA_2-3b	2	9	151	20.05	89.4
	PLA_2-2h	2	3	127	16.87	70.7
	PLA_2 -2aa	2	2	78	10.36	70.7
	3FTx-10a	4	5	45	5.98	37.9
	NGF-1b	6	12	41	5.44	36.4
	PLA_2-4b	3	3	38	5.05	74.8
	PLA_2 -7a	5	6	35	4.65	53.6
15b2	PLA_2-3b	1	1	43	33.08	76.1
	PLA_2-2h	1	1	31	23.85	64.2
	CTL-2a	8	10	24	18.46	74.5
	NGF-1b	4	5	14	10.77	31.1
	PLA_2 -7a	3	3	10	7.69	36.0
	3FTx-10a	3	3	8	6.15	37.9
15b3	CTL-2a	16	34	106	22.94	78.8
	PLA_2-2h	3	5	80	17.32	70.7
	PLA_2-2v	1	1	73	15.80	82.1
	PLA_2-3b	2	7	65	14.07	89.4
	PLA_2 -2aa	2	2	40	8.66	70.7
	NGF-1b	5	8	28	6.06	33.3
	PLA_2 -7a	3	3	22	4.76	42.4
16	SVMPIII-1	41	93	1341	82.52	54.8
$108 \min^*$	SVMPIII-1	32	67	755	27.44	55.3
	PDE	67	150	559	20.32	69.9
	NUC-1c	2	6	346	12.58	78.7
	NUC-1a	1	1	336	12.21	78.6
	SVMPIII-2a	1	1	237	8.62	31.1
114 min*	LAAO-1a	2	4	1004	32.89	81.7
	LAAO-1c	2	3	995	32.59	81.7

Fraction	Protein name	Exclusive UPC	Exclusive USC	TSC	% TSC	Seq Coverage (%)
	SVMPIII-1	23	45	335	10.97	45.4
	PDE	33	50	157	5.14	52.5
122 min^*	LAAO-1a	2	3	377	32.28	70.9
	LAAO-1c	2	2	373	31.93	70.9
	SVMPIII-1	19	27	151	12.93	41.8
	PLB	21	30	76	6.51	36.2
	NUC-1a	1	1	61	5.22	58.9

Table S2: Toxin identifications based on peptide spectral evidence of isolated peaks from reversed-phase high-performance liquid chromatography (RP-HPLC). Peaks that were only analyzed by mass spectrometry (i.e., not included in population comparisons due to extremely low abundance) are indicated by an asterisk and labeled with the retention time in minutes. RP-HPLC peak 15b, a single peak in population-level analyses, contained three peaks that were analyzed independently by means of mass spectrometry. These peaks are indicated by numbers above. Variable abbreviations: UPC = Unique Peptide Count; USC = Unique Spectrum Count; TSC = Total Spectrum Count. Toxin abbreviations: CTL, C-type lectin; KUN, Kunitz-type protease inhibitor; LAAO, L-amino acid oxidase; LCN, Long-chain neurotoxin; NGF, Nerve growth factor; NUC, Nucleotidase; PDE, Phospholiesterase; PLA₂, Phospholipase A₂; PLB, Phospholipase B; SVMP, Snake venom metalloproteinase; 3FTx, Three-finger toxin.

Tables S3-S4

Available for download as Excel files at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.172437/-/DC1

- **Table S3** Peptide reports for both species
- Table S4
 Raw, untransformed data for both species