

Population Genomic Analysis of a Pitviper Reveals Microevolutionary Forces Underlying Venom Chemistry

Steven D. Aird¹, Jigyasa Arora¹, Agneesh Barua¹, Lijun Qiu¹, Kouki Terada², and Alexander S. Mikheyev^{1,*}

¹Ecology and Evolution Unit, Okinawa Institute of Science and Technology Graduate University, Kunigami-gun, Okinawa-ken, Japan

²Okinawa Prefectural Institute of Health and the Environment, Biology and Ecology Group, Nanjo-shi, Okinawa, Japan

*Corresponding author: E-mail: sasha@homologo.us.

Accepted: September 27, 2017

Data deposition: Raw genomic reads and RNA-seq are available under NCBI accessions PRJNA313429 and PRJDB4386.

Abstract

Venoms are among the most biologically active secretions known, and are commonly believed to evolve under extreme positive selection. Many venom gene families, however, have undergone duplication, and are often deployed in doses vastly exceeding the LD₅₀ for most prey species, which should reduce the strength of positive selection. Here, we contrast these selective regimes using snake venoms, which consist of rapidly evolving protein formulations. Though decades of extensive studies have found that snake venom proteins are subject to strong positive selection, the greater action of drift has been hypothesized, but never tested. Using a combination of de novo genome sequencing, population genomics, transcriptomics, and proteomics, we compare the two modes of evolution in the pitviper, *Protobothrops mucrosquamatus*. By partitioning selective constraints and adaptive evolution in a McDonald–Kreitman-type framework, we find support for both hypotheses: venom proteins indeed experience both stronger positive selection, and lower selective constraint than other genes in the genome. Furthermore, the strength of selection may be modulated by expression level, with more abundant proteins experiencing weaker selective constraint, leading to the accumulation of more deleterious mutations. These findings show that snake venoms evolve by a combination of adaptive and neutral mechanisms, both of which explain their extraordinarily high rates of molecular evolution. In addition to positive selection, which optimizes efficacy of the venom in the short term, relaxed selective constraints for deleterious mutations can lead to more rapid turnover of individual proteins, and potentially to exploration of a larger venom phenotypic space.

Key words: pitvipers, selection, genetic drift, venom, population genomics.

Introduction

Understanding how natural selection leads to adaptation requires a mechanistic explanation of how alternative genotypes lead to fitness differences (Storz and Wheat 2010). Venoms, particularly proteinaceous venoms of snakes, provide an excellent model system for the study of adaptation, because venom secretions represent quantifiable phenotypes, which are under control of easily characterizable genomic loci. Venom proteins often evolve rapidly (Casewell et al. 2013; Aird et al. 2015), and phenotypic studies have shown that they are more efficacious against prey types typically encountered and consumed (Mackessy 1996; Jorge da Silva and Aird 2001; Gibbs and Mackessy 2009). The study of snake venom protein evolution, like that of many adaptive traits, has emphasized the role of positive selection on protein structure (Sunagar et al. 2014, 2015). The role of drift for protein

evolution has not been explicitly considered, except as a hypothesis to be rejected, a view that may fail to capture the complexity of selective regimes.

The strike of a snake initiates complex biochemical cascades within the prey organism that have immediate life and death consequences (Aird 2002). The venom cocktail must act quickly to immobilize the prey before it escapes or has a chance to injure the snake. Protein chemistry mediates the outcome of envenomation, and venom genes show extraordinarily high rates of evolution (Nakashima et al. 1995; Deshimaru et al. 1996). Yet, despite decades of interest in this phenomenon, still no consensus exists about the dominant forces driving venom protein sequence changes. On one hand, a large body of molecular evidence argues that venom components are subject to strong positive selection (i.e., “the adaptive hypothesis;” Nakashima et al. 1995; Kordis and

Gubensek 2000; Zupunski et al. 2003; Zhu et al. 2004; Lynch 2007; St Pierre et al. 2008; Barghi et al. 2015; Pedroso et al. 2015). On the other hand, the complexity of venoms entails pharmacological redundancy, with more venom components injected than minimally necessary to kill or otherwise immobilize the prey (i.e., the “overkill hypothesis”), predicting relaxed functional constraints upon individual venom components (Mebs 2001).

The overkill hypothesis relies on the empirical observation that some individual snakes can carry enough venom to kill up to thousands of lab mice (Broad et al. 1979). However, it has generally been dismissed as an oversimplified model of the envenomation process (Casewell et al. 2013), and in the face of strong molecular data supporting positive selection on venom components (Blumstein 2006; Barlow et al. 2009). In addition, the overkill hypothesis may not hold for a number of species of specialized snake prey or predators that have evolved resistance to venom (Perez et al. 1978; Poran et al. 1987; Heatwole and Poran 1995; Weissenberg et al. 1997; Jansa and Voss 2011; Drabeck et al. 2015; Holding, Biardi, et al. 2016; Holding, Drabeck, et al. 2016). However, the overkill hypothesis has never been empirically tested, since merely detecting positive selection on a gene does not preclude its also being subject to more relaxed selection. In addition to the overkill hypothesis, there could be other reasons for high rates of drift in venom genes following from birth-and-death evolutionary models (Nei and Rooney 2005). Consequently, most discussions of venom structure focus on adaptive explanations, whereas the extent to which neutral forces actually shape it remains unknown.

Although both the adaptive and overkill hypotheses make clear-cut predictions, it has been difficult to test them comprehensively in the same analytical framework until now. Previous studies have typically employed ratios of nonsynonymous and synonymous substitutions (d_r/d_s) to study selection. This ratio results from the joint action of both neutral (d_s) and selective forces (d_r), making it difficult to partition and to examine them separately. The advent of population genomics introduced a wide array of other proxies for detecting selection acting on genes and genomic regions (Stinchcombe and Hoekstra 2008; Ekblom and Galindo 2011). However, few approaches quantify selection and drift within the same framework. Exceptionally, McDonald–Kreitman-type tests use population-level data to examine rates of fixation of synonymous and nonsynonymous mutations between two diverging lineages, allowing additional parameters to be estimated (McDonald and Kreitman 1991; Smith and Eyre-Walker 2002). Methods based on the MK test are the only ones that can, in principle, yield an unbiased quantitative assessment of the contribution of positive selection (Eyre-Walker 2006) to venom evolution.

In addition, this framework can simultaneously infer the proportion of substitutions driven by positive selection (α) and the selective constraint for each gene, that is, the rate at

which deleterious mutations accumulate. These values are of key importance for assessing the merits of the adaptive and overkill hypotheses, which make contrasting predictions. Namely, under adaptive evolution, venom genes should accumulate adaptive substitutions at greater rates, compared with nonvenom genes. By contrast, if venom genes are tolerant of deleterious mutations, they would experience weaker selective constraint. In addition, the McDonald–Kreitman-type framework can also estimate other parameters, such as mutation rates ($\Theta = 4N_e\mu$) for each gene. Higher mutation rates have been proposed as an alternative explanation for the more rapid evolutionary rates of venom proteins, but this hypothesis has never been comprehensively tested in a genome-wide analysis (Kini and Chan 1999; Kini and Chinnasamy 2010).

In the present study, we examined the relative contributions of adaptive selection and genetic drift to evolution of venom and housekeeping genes, the latter defined as genes expressed in the venom gland, but not secreted into the venom proteome. We focused on the elaborated venom system of a pitviper, the Taiwan habu (*Protobothrops mucrosquamatus*), comparing it with its sister species, the Sakishima habu (*Protobothrops elegans*), from which it split approximately 2.9 Ma (Guo et al. 2007; Hedges et al. 2015). We find that both positive selection and genetic drift play a role in the rapid sequence changes seen in venom genes. However, expression levels of individual venom proteins may modulate this relationship, with the most abundant proteins experiencing stronger drift.

Materials and Methods

Sample Collection

Thirty specimens of *Protobothrops mucrosquamatus* were collected at various localities throughout Okinawa, Japan: in Nago City (28) and Nakijin Village (1) in the Motobu Peninsula and in Onna Village (1), south of Nago. Venom was extracted from all specimens at Day 0 and venom glands were harvested at Days 1, 2, 4, and 8. Internal organs were also removed and frozen in liquid nitrogen, after which they were maintained at -80°C until use. Venom gland transcriptomes were created for all specimens and a reference transcriptome was created from eight specimens, with two specimens selected randomly from each of the gland harvest days (Day 1: Pm_2, Pm_5; Day 2: Pm_8, Pm_11; Day 4: Pm_30, Pm_17; Day 8: Pm_23, Pm_27) (supplementary table S1, Supplementary Material online). The experimental protocols used for handling animals in this study have been approved by the Animal Resources Section of OIST.

Genome Assembly

The reference genome was assembled from liver tissue from sample Pm_15. We prepared PCR-free Illumina TruSeq libraries

size-selected to 450 ± 25 bp, which were sequenced on a HiSeq 2500 instrument to a mean depth of $69\times$. Scaffolds were then assembled using DISCOVAR (Weisenfeld et al. 2014). We also prepared 5 and 15 kb mate-pair libraries using Illumina's Nextera protocol and sequenced them on an Illumina HiSeq 2000 to a depth of approximately $100\times$ each. Reads from mate-paired libraries were trimmed and filtered using NextClip (Leggett et al. 2014) and used to scaffold contigs using SSPACE (Boetzer et al. 2011).

Transcriptomic and Proteomic Analyses

Venom gland RNA seq libraries were prepared as described previously (Aird et al. 2013), except that ERCC92 synthetic spike-ins were added to the RNA extracts as described by Aird et al (2015) for quality control. In addition, pooled libraries were normalized using the Evrogen Trimmer-2 cDNA normalization kit and sequenced to improve genome annotation for nonvenom transcripts. Reads were mapped to predicted coding sequences using Bowtie2 (Langmead and Salzberg 2012) within the RSEM package (Li and Dewey 2011). Though expression data doesn't accurately always reflect protein levels in the venom (Casewell et al. 2014), it does in *Protobothrops* (Aird et al. 2013, 2015). Therefore, we used the number of fragments per kilobase mapped (FPKM) as a measure of protein abundance. Mapped read counts for each sample, as estimated by RSEM, can be found in supplementary table S1, Supplementary Material online. Proteomic analysis was conducted as described previously (Aird et al. 2013, 2015).

Variant Calling

Variants in the coding sequences were called using four separate algorithms: GATK, freebayes, platypus, and samtools (Li et al. 2009; McKenna et al. 2010; Garrison and Marth 2012; Rimmer et al. 2014). Variants were converted to allelic primitives using GATK, and a consensus set was called by BAYSIC using a 0.80 posterior probability (Cantarel et al. 2014). Finally, indels, variants with $>10\%$ missing data in the ingroup samples, and variants with more than two alleles were filtered out. We then classified variants as synonymous or nonsynonymous using SnpEff (Cingolani et al. 2012), and noted which were fixed versus polymorphic with respect to the outgroup.

Resequencing

Twenty *P. mucrosquamatus* samples used for RNA-seq and one *P. elegans* were selected for whole-genome resequencing. Library construction and sequencing was performed by BGI (Shenzhen, China) on a HiSeq 2000 instrument in PE150 mode, using Illumina TruSeq libraries. BGI then trimmed sequencing adaptors using in-house scripts. Raw reads were then mapped to the reference assembly using NextGenMap (Sedlazeck et al. 2013). Sample

coverages computed by GATK can be found in supplementary table S1, Supplementary Material online. Copy number variation was assessed with CNV-seq using default parameters (Xie and Tammi 2009). A gene was considered to have copy number variation if a CNV-seq window with a significant signature of copy number variation intersected an exon of the gene model.

Population Genomics

A table of fixed or polymorphic replacement and silent substitutions was used to infer population genomic parameters using two approaches: a Bayesian approach (SnIPRE) (Eilertson et al. 2012), and a maximum likelihood approach (MKTest) (Welch 2006). The two packages compute parallel measures of selection for every gene. Both of them model mutational constraint, although it is defined differently. In SnIPRE it is a regression model coefficient, with more negative values indicating that polymorphic mutations are either being fixed or eliminated at a higher rate than synonymous mutations, whereas the MKTest estimates a fraction where $1-f$ of the mutants are under strong purifying selection. In either case, larger estimated values indicate relaxed selective pressure. The overall results were qualitatively similar for the two estimates of selective constraint, but gave radically different results for serine proteases, with SnIPRE estimating relatively relaxed constraint for this gene family, and MKTest estimating stronger selective constraint than experienced by the average gene in the genome. Almost all of the MKTest's estimates of f lay on boundary values (0, 1), suggesting a relatively poor fit to the model. We include them for completeness, but are inclined to trust them less than those of SnIPRE.

Both packages also estimate mutation rates as Θ ($= N_e * \mu$). In addition, MKTest estimates the proportion of nonsynonymous divergence driven by positive natural selection, a measure of true adaptive evolution (α) (Smith and Eyre-Walker 2002). In addition, both approaches model divergence times (SnIPRE: a coefficient called τ ; MKTest: expected neutral divergence per site: $\mu * t$) and, and weakly selected beneficial mutations resulting from mildly deleterious substitutions ($\gamma = 4N_e s$). These are not presented here because they lack testable predictions and straightforward biological interpretations in this context.

We then used the R statistical package (R core team 2015) to conduct a range of statistical analyses on the population genomic coefficient estimates. Given the relatively small number of venom genes, and the frequently nonnormal distribution of the coefficient estimates, we used nonparametric statistics, namely Spearman's rank correlation, and Kruskal-Wallis test adjusted for multiple comparisons using the method of Siegel and Castellan (Siegel and Castellan 1988). When comparing the fraction of adaptive substitutions acting on gene classes, we eliminated any genes with negative values of α , which is defined only for the range between 0 and 1,

and negative values are produced by sampling error or violations of the model (Eyre-Walker 2002).

Mapping Evolutionary Changes in Structure of Phospholipases A₂ and Serine Proteases

For each analysis, sequences of venom components from the reference genome were codon-aligned using MAFFT (Kato et al. 2002), and sites under selection were analyzed using the REL algorithm implemented in the HyPhy package on the DataMonkey server (Kosakovsky Pond and Frost 2005; Pond and Frost 2005).

A consensus sequence of amino acids was formed from the codon alignment and used as input for developing a PLA₂ model in SWISS-MODEL (Arnold et al. 2006; Kiefer et al. 2009). The suitable model was selected and edited using PyMOL (Delano 2002). Sites on the protein structure were color-coded based on the REL results, where a Bayes Factor of 50 or more for $dN > dS$, representing negative selection, was colored blue, whereas a Bayes Factor of 50 or more for $dN < dS$, representing positive selection, was colored red.

Data Accessibility

Genome annotation is available on NCBI (*Protobothrops mucrosquamatus* Annotation Release 100). Raw genomic reads and RNA-seq are available under NCBI accessions PRJNA313429 and PRJDB4386, respectively. Snakemake files (Köster and Rahmann 2012) for genome assembly and variant calling, as well as a script for R-based analysis and intermediate data are in a MySQL database available at <https://github.com/mikhayev/mucrosquamatus-selection>, last accessed September 29, 2017. A virtual machine necessary to reproduce the R-based analyses is provided at <http://mybinder.org/repo/mikhayev/mucrosquamatus-selection>, last accessed September 29, 2017. Note: the virtual machine relies on a third-party service, and its availability is not guaranteed. Therefore, we include a Dockerfile necessary to build such a virtual machine locally using the Docker platform. Finally, a human-readable version of the analysis leading to the principal results and generating draft versions of all the figures is available at <https://mikhayev.github.io/mucrosquamatus-selection/>, last accessed September 29, 2017.

Results

Genome Assembly

We produced the first genome assembly of a pitviper, *P. mucrosquamatus*. It was 1.63 gigabases long (contig and scaffold N50, 22 and 424 kb, respectively). The genome was annotated using RNA-seq data from 32 individual snakes, producing an annotation with 20,122 protein-coding genes. We also resequenced 20 additional individual *P. mucrosquamatus* genomes and a *P. elegans* genome as an outgroup (supplementary table S1, Supplementary Material online).

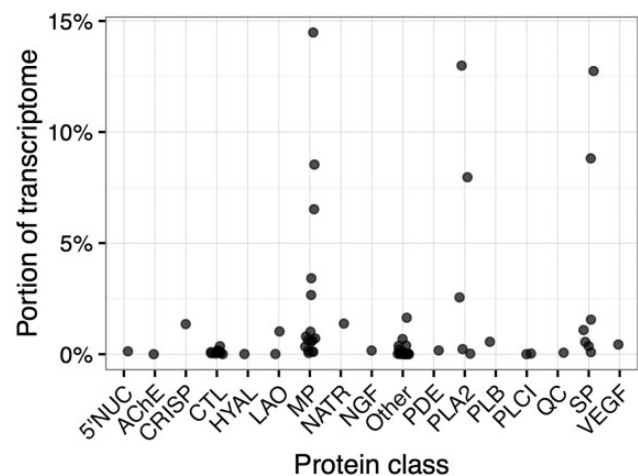


Fig. 1.—A small number of genes, belonging to just three families, dominates the venom of *P. mucrosquamatus*. Transcript proportions are correlated with proteomic abundance (Aird et al. 2013, 2015). Abbreviations: 5'NT 5'-nucleotidase, AChE acetylcholinesterase, CRISP cysteine-rich secretory proteins, CTL C-type lectin-like proteins, HYAL hyaluronidase, LAO L-amino acid oxidase, MP metalloprotease, NATR C-type natriuretic peptide, NGF nerve growth factor, PDE phosphodiesterase, PLA2 phospholipase A₂, PLB Phospholipase B, PLCI phospholipase A₂ inhibitor, QC glutaminyl cyclase, SP serine protease, VEGF vascular endothelial growth factor.

These data were used to calculate the number of fixed and replacement synonymous and nonsynonymous substitutions between the two species, which were used to compute estimates of selective constraint and positive selection for every polymorphic gene.

Venom Composition

We detected 76 proteins in the venom of *P. mucrosquamatus* using mass spectrometry. However, these venoms were dominated by a few phospholipase A₂, serine protease, and metalloprotease genes (fig. 1), with the rest of the proteins, including other members of these three families, playing minor roles. Both snakes had strongly correlated patterns of gene expression, suggesting conservation of their regulatory machinery ($r_s = 0.73$, $N = 55$, $P < 2 \times 10^{-16}$, supplementary, fig. S1, Supplementary Material online). Thus, both species should experience similar effects of gene expression on selection. To the best of our knowledge, there have been no formal dietary studies of either species; however, anecdotal reports and our own field observations suggest that both are dietary generalists, preying upon frogs, lizards, birds, rodents, and shrews. That being said, percentages of each prey type may vary locally with availability, making it is impossible to draw any firm conclusions about how venom composition in either species may be related to diet.

Structural Organization of Venom Gene Clusters

Protobothrops mucrosquamatus venom includes four highly diversified gene families: C-type lectin-like proteins (CTLs),

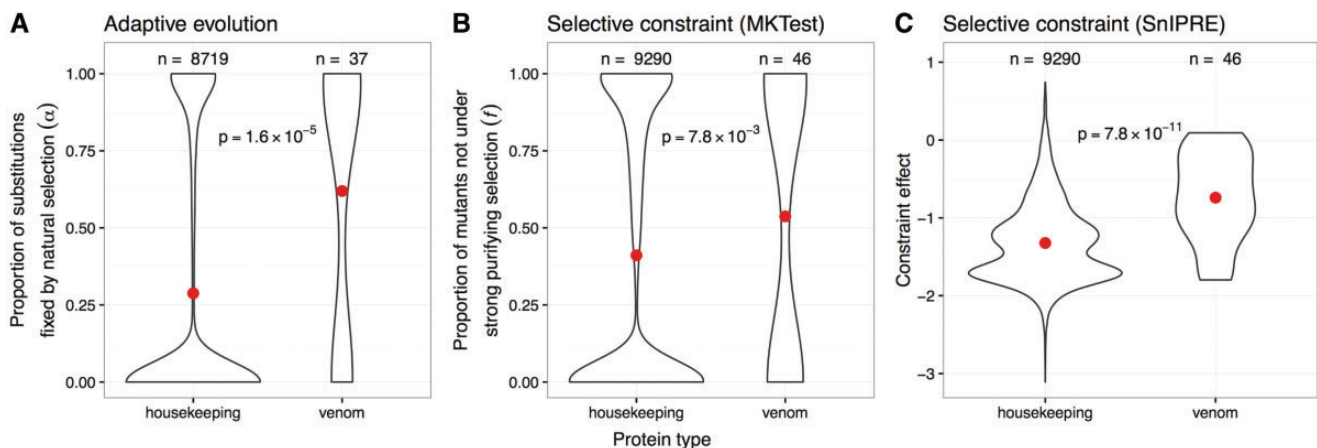


FIG. 2.—Venom proteins show higher rates of adaptive evolution (A), and relaxed selective constraint (B, C) compared with housekeeping genes. Violin plots summarize the distributions of the parameter estimates, with means given by red dots. Statistical significance is computed using Kruskal–Wallis tests. The proportion of adaptive substitutions fixed by natural selection refers to those in relatively recent evolutionary time, since the divergence of the *P. mucrosquamatus* and *P. elegans*, without distinguishing between the possible modes of natural selection (e.g., episodic vs. continuous). Selective constraint was estimated with two software packages that use different values for their coefficients, but in both cases higher values mean lower selective constraint. These data sets support both adaptive and overkill hypotheses, respectively.

metalloproteases, serine proteases, and phospholipases A_2 . The latter two families are located in tandem arrays of duplicated genes, each on its own scaffold (NW_015386730.1 and NW_015387341.1). Whether CTLs and metalloproteases have similar organization is not certain, given the fragmented nature of the assembly, but half of the genes in each family reside on just a few scaffolds. Previous work has shown that phospholipases A_2 in the Okinawa habu (*Protobothrops flavoviridis*) also occur in a single gene cluster (Ikeda et al. 2010), though the copy number and orientation of genes differs from that reconstructed by the present study for *P. mucrosquamatus*. Two phospholipase A_2 genes found in *P. mucrosquamatus* (LOC107291353 and LOC107291356) appear to be missing from the closely related *P. elegans* genome. There also appears to be copy number variation within many *P. mucrosquamatus* venom genes (supplementary fig. S2, Supplementary Material online). These findings suggest that venom gene clusters may be structurally polymorphic, with genes being duplicated and lost both within a species and between closely related species.

Do Venom Genes Mutate More Rapidly?

Differences in evolutionary rates can also potentially result from unequal rates of background mutation (Kini and Chinnasamy 2010). However, we found no difference in mutation rates (Θ) between venom and nonvenom genes (SnIPRE $P=0.059$, MKTest $P=0.73$), suggesting that drift and selection are the major forces behind venom protein sequence change.

Effects of Gene Expression on Venom Protein Evolution

There was no effect of average expression level of venom proteins on the proportion of substitutions driven by positive

selection ($r_s = 0.017$, $P=0.92$). However, there was evidence that selective constraints on the most abundant toxins are diminished (SnIPRE $r_s = 0.016$, $P=0.016$; MKTest $r_s = 0.28$, $P=0.075$; fig. 3). Relaxed constraints lead to accumulation of deleterious mutations, and an observed increase in the d_r/d_s ratio.

Discussion

Selective Forces Acting on Venom

Adaptive Evolution

The role of adaptive selection in snake venom evolution is uncontroversial, and we also find that venom genes have a greater proportion of substitutions driven by positive selection (α), than the rest of the genome (fig. 2A). Overall, there was weaker support for selection acting on individual venom protein classes (supplementary fig. S2, Supplementary Material online), largely because per-class sample sizes are significantly smaller. Sequence-based analysis of the dominant venom components, the phospholipases A_2 and serine proteases, identified multiple sites under positive selection, particularly on the exteriors of these proteins, which suggest accelerated evolution of functional components (Kini and Chan 1999) (supplementary fig. S3, Supplementary Material online). Strong adaptive selection is consistent with a wide suite of ecological observations, further supporting the adaptive value of snake venoms. For example, snake venoms tend to be more effective for subduing their preferred prey (Silva Jr and Aird 2001; Gibbs and Mackessy 2009) and even closely related taxa with divergent diets employ strongly divergent venoms (Sanz et al. 2006). Perhaps most dramatically, the adoption of a fish egg-only diet by the sea snake, *Aipysurus eydouxii*, has led to relaxed selection on its venom, and a

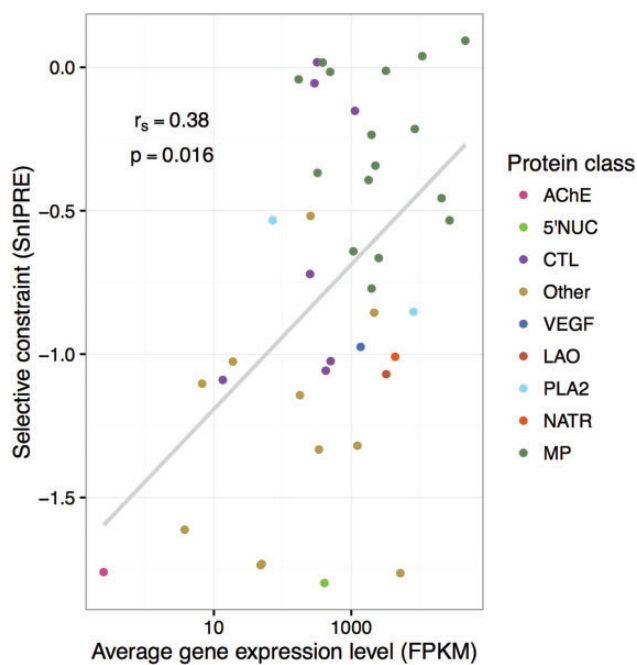


Fig. 3.—More abundant venom components experience increasingly relaxed selective constraint. As in figure 2, higher values of the selective constraint coefficient, indicate less effective elimination of deleterious mutations. MKTest results were consistent with those of SnIPRE, but marginally nonsignificant ($P=0.075$). Because estimates of mutational constraint differed widely between SnIPRE and MKTest, they are excluded from this analysis, though including them does not qualitatively change results. Abbreviations as in figure 1. These findings suggest that selective constraint can be modulated by protein expression level.

variety of dysfunctional mutations in its phospholipase A_2 genes (Li et al. 2005). It is therefore likely that positive selection is the dominant driver of snake venom evolution.

The Overkill Hypothesis and the Role of Drift in Venom Proteome Evolution

We likewise find that venom components experience less selective constraint than housekeeping genes (fig. 2B and C and supplementary fig. S4, Supplementary Material online). Although this finding may, at first, seem to contradict the higher levels of adaptive evolution, it is important to remember that selective pressures may vary across the secondary and tertiary structure of a given protein (supplementary fig. S5, Supplementary Material online), and that, in principle, some parts of the protein can experience positive selection, whereas other parts are free to evolve neutrally. Consequently, relaxed selective constraints do not imply the absence of positive selection, but rather its relative strength. Indeed, numerous studies, and our data from *P. mucrosquamatus*, show that major venom components in snakes are under strong selection (Kordis et al. 1998; Kordis and Gubensek 2000; Zupunski et al. 2003; Juárez et al. 2008; Barlow et al. 2009; Casewell et al. 2011; Vonk et al. 2013; Malhotra et al. 2015).

That being said, reduced selective constraint on venom proteins supports predictions of the overkill hypothesis, and in turn, has three interesting and underexplored consequences for venom proteins: 1) increased rates of pseudogenization, 2) higher levels of standing variation, and 3) venom evolution may happen neutrally during vicariant speciation.

First, accumulation of deleterious mutations could lead to pseudogenization. Indeed, there is evidence that this has occurred in the phospholipase A_2 cluster of *P. flavoviridis*, which contains two pseudogenes and three active genes (Ikeda et al. 2010). Furthermore, allopatrically isolated populations of this species differ in the presence or absence of pseudogenes (Chijiwa et al. 2000; Ikeda et al. 2010). Although the authors attribute such interisland variability to selection, based on our findings, it also seems possible that at least some of the variation is due to neutral mutation and drift, which would be particularly strong in small isolated populations. Furthermore, as pseudogenized genes no longer produce functional proteins, their expression should be strongly selected against, ultimately eliminating them from the venom, whereas other proteins take their place. Consequently, drift, possibly followed by negative selection on pseudogenized genes, can lead to accelerated turnover of individual venom components. Indeed, gene loss was found to be a key factor in the evolution of phospholipase A_2 genes in rattlesnakes (Dowell et al. 2016).

Second, higher levels of variability associated with relaxed selective constraint may expand the phenotypic space available to snake venoms. The concept of phenotypic space is analogous to Hutchinson's concept of the niche, as an N-dimensional hypervolume (Hutchinson 1957). In venomous phenotypic space, each venom constituent multiplied by its concentration could be conceptualized as a dimension in N-dimensional phenotypic space, such that the incorporation or deletion of venom constituents, or changes in their abundance, would add, delete, lengthen, or shorten dimensions. It seems highly likely that for any given snake species, there could be multiple biochemical strategies that might be equally efficacious, particularly if it feeds on multiple prey types. Nonetheless, members of a venomous snake species, in essence, represent elements of a living array that performs what a pharmaceutical company might call "pharmaceutical lead optimization." That is, many toxin variants and their combinations are screened continuously and simultaneously across populations.

Although positive selection promotes local optima, genetic drift may allow populations to cross fitness valleys, and potentially leads to new venom compositions (Wright 1932). For example, previously dominant toxins could accumulate deleterious mutations and become less abundant, whereas more minor components could acquire beneficial mutations and increase in concentration. This could result from either changes in their regulatory sequences, or from changes in copy number (Aird et al. 1991; Malhotra 2015). Although such evolutionary dynamics would be difficult to conclusively

demonstrate, relative abundance of venom components can indeed change relatively rapidly. For instance, a phospholipase A_2 comprising 73.4% of the *P. elegans* transcriptome, is barely present in venom of its closest relatives, *P. flavoviridis* (<0.05%) (Aird et al. 2015) and also in *P. mucrosquamatus* (present study, supplementary fig. S1, Supplementary Material online). Even within a species, there can be significant geographic variation in protein abundance (Mebs and Kornalik 1984; Chijiwa et al. 2003), as demonstrated by studies of venom composition among littermates (Chippaux et al. 1982; Pintor et al. 2011). Future work focused on phylogenetic reconstruction of ancestral states of proteins, and their expression levels, should provide important insights into the interplay between expression, selection, and venom evolution, as has been done for a range of other animals (Morandin et al. 2016).

Third, the observation that venoms are more susceptible to the effects of genetic drift than the rest of the genome, has important implications for venom differences between species. Vicariance effects, and in particular founder effects, will change effective population sizes and further exacerbate the effects of drift on venom components, possibly leading to significant amounts of differentiation between populations or nascent species. Such regional variation is found in Okinawan *Protobothrops*, which differs in venom composition between islands (Chijiwa et al. 2003). Many of these islands are small and isolated, likely harboring small populations of snakes. Even more recent work on the eastern diamondback rattlesnake (*Crotalus adamanteus*) has shown that populations vary greatly in copy number at toxin-encoding loci, and that selection for increased expression, rather than sequence diversity has driven duplications (Margres et al. 2017). Therefore, neutral divergence should be considered when comparing possible adaptive differences between species.

Effect of Expression on Venom Evolution

Such an increase has, in fact, recently been documented by us (Aird et al. 2015). At that time, we interpreted the higher d_f/d_s rate as evidence of stronger adaptive selection on the most abundant proteins, whereas the present analysis suggests that the opposite is true, highlighting the advantages of the McDonald–Kreitman framework over traditional d_f/d_s statistics.

At first glance, the relatively lower strength of selection on more abundant venom components seems counterintuitive. However, we propose a model to explain this pattern, based on several additional observations. First, in order to assure rapid immobilization of prey, snakes inject an optimal amount of venom, which is several times more than the minimally necessary dose (Allon and Kochva 1974; Hayes 1995; Herbert and Hayes 2008). Second, studies of *P. flavoviridis* venom found that the venom fraction containing hemorrhagic metalloproteases, second only to phospholipases in abundance (Aird et al. 2013), showed roughly twice the

lethality of crude venom (Ohsaka 1960; Takahashi and Ohsaka 1970). However, Aird (2002) suggested that immobilization via hypotension and paralysis, rather than lethality, is the objective of envenomation, and by extension, the trait upon which selection acts. In envenomated mice, immobilization time is roughly 3–5× shorter than the time to death (Herbert and Hayes 2008). Therefore, if excess venom is injected to rapidly immobilize prey, concentrations of the most abundant venom components may indeed reach ‘overkill’ levels. Thus, an excess of the most abundant components could lead to relaxed selective constraints upon them, since deleterious mutations could be compensated by higher concentration.

However, can higher concentrations rescue mutated proteins? Indirect insights into this question come from site-directed mutagenesis studies that look at the activity of venom components. Because such studies typically target sites of likely biochemical importance, and make nonrandom alterations of the protein, they explore the more extreme end of the protein fitness landscape. However, site-directed mutagenesis studies generally find that, except for mutations in the active site, which tend to be strongly deleterious, most other mutations have either no effect, or relatively mild effects on enzymatic properties of the venom (e.g., Trémeau et al. 1995; Dennis et al. 1993; Zhang et al. 1997). Such decreases in catalytic or noncatalytic activity can, in principle, be rescued by increased protein concentration. Conversely, if the levels of the protein were high in the first place, mutations may be masked as long as the decrease in activity is relatively modest.

Analytical Caveats

Whole-genome inference of population-level parameters is difficult, and subject to a number of important assumptions. First, changes in demography may affect parameter estimates. Although this could affect absolute values of the parameters, our major conclusions should not be affected, since we are making comparisons within a genome. Another important assumption is the independence of genes. Since venom genes are found in large clusters, they may be effectively linked, possibly biasing results. However, the different venom classes are most likely unlinked, and the omnibus tests presented in figure 2, should be relatively robust. Unfortunately, linkage at the family level affects all previous studies of selection acting on venom genes as well. Finally, copy number variation may affect some parameter estimates by possibly inflating the extent of polymorphism existing in some genes. Filtering out any genes that display evidence of copy number variation still gives a higher overall level of adaptive substitution in venom genes ($P=0.013$), and a significantly lower selective constraint by one of the two software packages (SnIPRE $P=0.022$; MKTest $P=0.76$). Thus, while copy number variation can potentially skew some parameter

estimates, it is unlikely to be the driving force behind the statistical signal.

Another point that bears mention, is that the specimens used in this study came from populations introduced to Okinawa, and may represent only a subset of genetic diversity found in the native range populations. This will reduce number of polymorphic loci in the genome, and introduce a level of noise due to stochastic sampling from the source population. However, while these factors may change the absolute estimates of selective constraint and adaptive evolution, they should not affect the relative estimates within a genome, that is between venom and nonvenom components (fig. 2), or between different venom components (fig. 3).

Conclusions

An important characteristic of snake venoms is functional redundancy among components, resulting in multiple impacts upon critical prey systems, such as blood pressure regulation (Aird 2002; Aird et al. 2016). Redundancy is achieved both by employment of highly specific toxins that target particular receptors, ion channels, etc., and also through delivery and release of secondary messengers and modulators, such as purine nucleosides and polyamines, that affect multiple systems (Aird 2002). Individual venom components experience a wide variety of selective pressures, ranging from positive selection at certain sites, to purifying selection at others, to neutrality (supplementary fig. S5, Supplementary Material online). Our data suggest that selective regimes of venom components result from a complex adaptive landscape dominated by positive and negative selection, but also affected by varying degrees of genetic drift. In particular, changes in population size or population substructure, which affect genetic drift, may also affect trajectories of venom evolution. Although we have focused on understanding evolutionary forces acting on venom protein composition since the separation of two snake species, the increasing availability of genomes and genomic tools for other species will soon show whether these patterns may be more broadly applicable to other snakes, or perhaps even to all venomous organisms that employ peptidyl/proteinaceous secretions.

Supplementary Material

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

Acknowledgments

We are grateful to Miquel Graul Lopez for help depositing sequences to DDBJ, and to Alejandro Villar Briones for carrying out the mass spectrometric analysis. We thank Anita Malhotra and Ivan Koludarov for comments on the manuscript draft and Brian Fry for general feedback. We thank

the OIST Sequencing Center for making the mate-paired libraries and for sequencing the genomic shotgun libraries. This work was provided by subsidy funded to the Okinawa Institute of Science and Technology Graduate University.

Literature Cited

- Aird SD. 2002. Ophidian envenomation strategies and the role of purines. *Toxicon* 40(4): 335–393.
- Aird SD, et al. 2013. Quantitative high-throughput profiling of snake venom gland transcriptomes and proteomes (*Ovophis okinavensis* and *Protobothrops flavoviridis*). *BMC Genomics* 14: 790.
- Aird SD, et al. 2015. Snake venoms are integrated systems, but abundant venom proteins evolve more rapidly. *BMC Genomics* 16: 647.
- Aird SD, Kruggel WG, Kaiser II. 1991. Multiple myotoxin sequences from the venom of a single prairie rattlesnake (*Crotalus viridis viridis*). *Toxicon* 29(2): 265–268.
- Aird SD, Villar Briones A, Roy MC, Mikheyev AS. 2016. Polyamines as snake toxins and their probable pharmacological functions in envenomation. *Toxins* 26: E279.
- Allon N, Kochva E. 1974. The quantities of venom injected into prey of different size by *Vipera palaestinae* in a single bite. *J Exp Zool.* 188(1): 71–75.
- Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22(2): 195–201.
- Barghi N, Concepcion GP, Olivera BM, Lluisma AO. 2015. Comparison of the venom peptides and their expression in closely related *Conus* species: Insights into adaptive post-speciation evolution of *Conus* exogenomes. *Genome Biol Evol.* 7(6): 1797–1814.
- 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 Biol Sci.* 276(1666): 2443–2449.
- Blumstein DT. 2006. The Multipredator Hypothesis and the evolutionary persistence of antipredator behavior. *Ethology* 112(3): 209–217.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4): 578–579.
- Broad AJ, Sutherland SK, Coulter AR. 1979. The lethality in mice of dangerous Australian and other snake venom. *Toxicon* 17(6): 661–664.
- Cantarel BL, et al. 2014. BAYSIC: a Bayesian method for combining sets of genome variants with improved specificity and sensitivity. *BMC Bioinformatics* 15: 104.
- Casewell NR, et al. 2014. Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. *Proc Natl Acad Sci U S A.* 111(25): 9205–9210.
- Casewell NR, Wagstaff SC, Harrison RA, Renjifo C, Wüster W. 2011. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. *Mol Biol Evol.* 28(9): 2637–2649.
- Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG. 2013. Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol. Evol* 28(4): 219–229.
- Chijiwa T, et al. 2003. Interisland evolution of *Trimeresurus flavoviridis* venom phospholipase A2 isozymes. *J Mol Evol.* 56(3): 286–293.
- Chijiwa T, et al. 2000. Regional evolution of venom-gland phospholipase A2 isoenzymes of *Trimeresurus flavoviridis* snakes in the southwestern islands of Japan. *Biochem J.* 347(Pt 2): 491–499.
- Chippaux JP, Boche J, Courtois B. 1982. Electrophoretic patterns of the venoms from a litter of *Bitis gabonica* snakes. *Toxicon* 20(2): 521–522.
- Cingolani P, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6(2): 80–92.

- Delano WL. 2002. The PyMOL Molecular Graphics System. Available from: <http://www.pymol.org>, last accessed September 29, 2017.
- Dennis MS, Carter P, Lazarus RA. 1993. Binding interactions of kistrin with platelet glycoprotein IIb-IIIa: analysis by site-directed mutagenesis. *Proteins* 15(3): 312–321.
- Deshimaru M, et al. 1996. Accelerated evolution of crotalinae snake venom gland serine proteases. *FEBS Lett.* 397(1): 83–88.
- Dowell NL, et al. 2016. The deep origin and recent loss of venom toxin genes in rattlesnakes. *Curr Biol.* 26(18): 2434–2445.
- Drabek DH, Dean AM, Jansa SA. 2015. Why the honey badger don't care: Convergent evolution of venom-targeted nicotinic acetylcholine receptors in mammals that survive venomous snake bites. *Toxicon* 99: 68–72.
- Eilertson KE, Booth JG, Bustamante CD. 2012. SniPRE: selection inference using a Poisson random effects model. *PLoS Comput Biol.* 8(12): e1002806.
- Eklom R, Galindo J. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 107(1): 1–15.
- Eyre-Walker A. 2002. Changing effective population size and the McDonald-Kreitman test. *Genetics* 162(4): 2017–2024.
- Eyre-Walker A. 2006. The genomic rate of adaptive evolution. *Trends Ecol Evol.* 21(10): 569–575.
- Garrison E, Marth G. 2012. FreeBayes. arXiv preprint 1207.3907 [q-bio.GN]. Available from: <http://arxiv.org/abs/1207.3907>, last accessed September 29, 2017.
- Gibbs HL, Mackessy SP. 2009. Functional basis of a molecular adaptation: prey-specific toxic effects of venom from *Sistrurus rattlesnakes*. *Toxicon* 53(6): 672–679.
- Guo P, Malhotra A, Li PP, Pook CE, Creer S. 2007. New evidence on the phylogenetic position of the poorly known Asian pitviper *Protobothrops kaulbacki* (Serpentes: Viperidae: Crotalinae) with a re-description of the species and a revision of the genus *Protobothrops*. *Herpetol J.* 17: 237–246.
- Hayes WK. 1995. Venom metering by juvenile prairie rattlesnakes, *Crotalus v. viridis*: effects of prey size and experience. *Anim Behav.* 50(1): 33–40. 17.
- Heatwole H, Poran NS. 1995. Resistances of sympatric and allopatric eels to sea snake venoms. *Copeia* 1995(1): 136–147.
- Hedges SB, Marin J, Suleski M, Paymer M, Kumar S. 2015. Tree of life reveals clock-like speciation and diversification. *Mol Biol Evol.* 32(4): 835–845.
- Herbert SS, Hayes WK. 2008. Venom expenditure by rattlesnakes and killing effectiveness in rodent prey: do rattlesnakes expend optimal amounts of venom? In: Hayes, WK, Beaman, KR, Cardwell, Bush, SP, editors. *The biology of rattlesnakes*. Loma Linda: Loma Linda University Press, CA. p. 221–228.
- Holding ML, Biardi JE, Gibbs HL. 2016. Coevolution of venom function and venom resistance in a rattlesnake predator and its squirrel prey. *Proc Biol Sci.* 283(1829): 20152841.
- Holding ML, Drabek DH, Jansa SA, Gibbs HL. 2016. Venom resistance as a model for understanding the molecular basis of complex coevolutionary adaptations. *Integr Comp Biol.* 56(5): 1032–1043.
- Hutchinson GE. 1957. Concluding remarks. *Cold Spring Harb Symp Quant Biol.* 22: 415–427.
- Ikeda N, et al. 2010. Unique structural characteristics and evolution of a cluster of venom phospholipase A2 isozyme genes of *Protobothrops flavoviridis* snake. *Gene* 461(1-2): 15–25.
- Jansa SA, Voss RS. 2011. Adaptive evolution of the venom-targeted vWF protein in opossums that eat pitvipers. *PLoS One* 6(6): e20997.
- Juárez P, Comas I, González-Candelas F, Calvete JJ. 2008. Evolution of snake venom disintegrins by positive Darwinian selection. *Mol Biol Evol.* 25(11): 2391–2407.
- Katoh K, Misawa K, Kuma K-I, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30(14): 3059–3066.
- Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. 2009. The SWISS-MODEL repository and associated resources. *Nucleic Acids Res.* 37(Database issue): D387–D392.
- Kini RM, Chan YM. 1999. Accelerated evolution and molecular surface of venom phospholipase A2 enzymes. *J Mol Evol.* 48(2): 125–132.
- Kini RM, Chinnasamy A. 2010. Nucleotide sequence determines the accelerated rate of point mutations. *Toxicon* 56(3): 295–304.
- Kordis D, Bdoolah A, Gubensek F. 1998. Positive Darwinian selection in *Vipera palaestinae* phospholipase A2 genes is unexpectedly limited to the third exon. *Biochem Biophys Res Commun.* 251(2): 613–619.
- Kordis D, Gubensek F. 2000. Adaptive evolution of animal toxin multigene families. *Gene* 261(1): 43–52.
- Kosakovskiy SL, Frost SDW. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 22(5): 1208–1222.
- Köster J, Rahmann S. 2012. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* 28(19): 2520–2522.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4): 357–359.
- Leggett RM, Clavijo BJ, Clissold L, Clark MD, Caccamo M. 2014. NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. *Bioinformatics* 30(4): 566–568.
- Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12: 323.
- Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16): 2078–2079.
- Li M, Fry BG, Kini RM. 2005. Putting the brakes on snake venom evolution: the unique molecular evolutionary patterns of *Aipysurus eydouxi* (Marbled sea snake) phospholipase A2 toxins. *Mol Biol Evol.* 22(4): 934–941.
- Lynch VJ. 2007. Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. *BMC Evol Biol.* 7: 2.
- Mackessy SP. 1996. Characterization of the major metalloprotease isolated from the venom of the northern pacific rattlesnake, *Crotalus viridis oreganus*. *Toxicon* 34(11–12): 1277–1285.
- Malhotra A. 2015. Mutation, duplication, and more in the evolution of venomous animals and their toxins. In: Gopalakrishnakone P, Malhotra A, editors. *Evolution of venomous animals and their toxins*. *Toxinology*. Netherlands: Springer. p. 1–11.
- Malhotra A, Creer S, Harris JB, Thorpe RS. 2015. The importance of being genomic: Non-coding and coding sequences suggest different models of toxin multi-gene family evolution. *Toxicon* 107(Pt B): 344–358.
- Margres MJ, Bigelow AT, Lemmon EM, Lemmon AR, Rokytka DR. 2017. Selection to increase expression, not sequence diversity, precedes gene family origin and expansion in rattlesnake venom. *Genetics* 206(3): 1569–1580.
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351(6328): 652–654.
- McKenna A, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20(9): 1297–1303.
- Mebs D. 2001. Toxicity in animals. *Trends in evolution?* *Toxicon* 39(1): 87–96.
- Mebs D, Kornalik F. 1984. Intraspecific variation in content of a basic toxin in eastern diamondback rattlesnake (*Crotalus adamanteus*) venom. *Toxicon* 22(5): 831–833.
- Morandin C, et al. 2016. Comparative transcriptomics reveals the conserved building blocks involved in parallel evolution of diverse phenotypic traits in ants. *Genome Biol.* 17: 43.
- Nakashima K, et al. 1995. Accelerated evolution in the protein-coding regions is universal in crotalinae snake venom gland phospholipase A2 isozyme genes. *Proc Natl Acad Sci U S A.* 92(12): 5605–5609.

- Nei M, Rooney AP. 2005. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet.* 39: 121–152.
- Ohsaka A. 1960. Fractionation of habu snake venom by chromatography on CM-cellulose with special reference to biological activities. *Jpn J Med Sci Biol* 13: 199–205.
- Pedroso A, Matioli SR, Murakami MT, Pidde-Queiroz G, Tambourgi DV. 2015. Adaptive evolution in the toxicity of a spider's venom enzymes. *BMC Evol Biol.* 15: 290.
- Perez JC, Haws WC, Garcia VE, Jennings BM. 3rd. 1978. Resistance of warm-blooded animals to snake venoms. *Toxicon* 16(4): 375–383.
- Pintor AFV, Winter KL, Krockenberger AK, Seymour JE. 2011. Venom physiology and composition in a litter of Common Death Adders (*Acanthophis antarcticus*) and their parents. *Toxicon* 57(1): 68–75.
- Pond SLK, Frost SDW. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21(10): 2531–2533.
- Poran NS, Coss RG, Benjamini E. 1987. Resistance of California ground squirrels (*Spermophilus beecheyi*) to the venom of the northern Pacific rattlesnake (*Crotalus viridis oreganus*): a study of adaptive variation. *Toxicon* 25(7): 767–777.
- R core team. 2015. R: A language and environment for statistical computing. Vienna, Austria: R core team.
- Rimmer A, et al. 2014. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet.* 46(8): 912–918.
- Sanz L, Gibbs HL, Mackessy SP, Calvete JJ. 2006. Venom proteomes of closely related *Sistrurus* rattlesnakes with divergent diets. *J Proteome Res.* 5(9): 2098–2112.
- Sedlazeck FJ, Rescheneder P, von Haeseler A. 2013. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* 29(21): 2790–2791.
- Siegel S, Castellan NJ Jr. 1988. Non parametric statistics for the behavioural sciences. New York: MacGraw Hill Int.
- Silva Jr, N, Aird SD. 2001. Prey specificity, comparative lethality and compositional differences of coral snake venoms. *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol.* 128(3): 425–456.
- Smith NGC, Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. *Nature* 415(6875): 1022–1024.
- Stinchcombe JR, Hoekstra HE. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100(2): 158–170.
- Storz JF, Wheat CW. 2010. Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution* 64(9): 2489–2509.
- St Pierre L, et al. 2008. Common evolution of waprin and kunitz-like toxin families in Australian venomous snakes. *Cell Mol Life Sci.* 65(24): 4039–4054.
- Sunagar K, et al. 2014. Deadly innovations: unraveling the molecular evolution of animal venoms. In: *Venom genomics and proteomics*. The Netherlands: Springer. p. 1–23.
- Sunagar K, Moran Y, Hoekstra HE. 2015. The rise and fall of an evolutionary innovation: contrasting strategies of venom evolution in ancient and young animals. *PLoS Genet.* 11(10): e1005596.
- Takahashi T, Ohsaka A. 1970. Purification and some properties of two hemorrhagic principles (HR2a and HR2b) in the venom of *Trimeresurus flavoviridis*; complete separation of the principles from proteolytic activity. *Biochim Biophys Acta* 207(1): 65–75.
- Trémeau O, et al. 1995. Genetic engineering of snake toxins. The functional site of Erabutoxin a, as delineated by site-directed mutagenesis, includes variant residues. *J Biol Chem.* 270(16): 9362–9369.
- Vonk FJ, et al. 2013. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc Natl Acad Sci U S A.* 110(51): 20651–20656.
- Weisenfeld NI, et al. 2014. Comprehensive variation discovery in single human genomes. *Nat Genet.* 46(12): 1350–1355.
- Weissenberg S, Bouskila A, Dayan T. 1997. Resistance of the common spiny mouse (*Acomys cahirinus*) to the strikes of the Palestine saw-scaled viper (*Echis coloratus*). *Isr J Zool.* 43: 119.
- Welch JJ. 2006. Estimating the genomewide rate of adaptive protein evolution in *Drosophila*. *Genetics* 173(2): 821–837.
- Wright S. 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Available from: <http://www.esp.org/books/6th-congress/facsimile/contents/6th-cong-p356-wright.pdf>, last accessed September 29, 2017.
- Xie C, Tammi MT. 2009. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics* 10: 80.
- Zhang Y, et al. 1997. *Trimeresurus stejnegeri* snake venom plasminogen activator. Site-directed mutagenesis and molecular modeling. *J Biol Chem.* 272(33): 20531–20537.
- Zhu S, Bosmans F, Tytgat J. 2004. Adaptive evolution of scorpion sodium channel toxins. *J Mol Evol.* 58(2): 145–153.
- Zupunski V, Kordis D, Gubensek F. 2003. Adaptive evolution in the snake venom Kunitz/BPTI protein family. *FEBS Lett.* 547(1–3): 131–136.

Associate editor: Mandë Holford