Expression of Venom Gene Homologs in Diverse Python Tissues Suggests a New Model for the Evolution of Snake Venom

Jacobo Reyes-Velasco,[1] Daren C. Card,[1] Audra L. Andrew,[1] Kyle J. Shaney,[1] Richard H. Adams,[1] Drew R. Schield,[1] Nicholas R. Casewell,[2] Stephen P. Mackessy,[3] and Todd A. Castoe*,[1]

[1]Department of Biology, University of Texas at Arlington

[2]Alistair Reid Venom Research Unit, Liverpool School of Tropical Medicine, Liverpool, United Kingdom

[3]School of Biological Sciences, University of Northern Colorado

*Corresponding author: E-mail: todd.castoe@uta.edu.

Associate editor: Matthew Hahn

Abstract

Snake venom gene evolution has been studied intensively over the past several decades, yet most previous studies have lacked the context of complete snake genomes and the full context of gene expression across diverse snake tissues. We took a novel approach to studying snake venom evolution by leveraging the complete genome of the Burmese python, including information from tissue-specific patterns of gene expression. We identified the orthologs of snake venom genes in the python genome, and conducted detailed analysis of gene expression of these venom homologs to identify patterns that differ between snake venom gene families and all other genes. We found that venom gene homologs in the python are expressed in many different tissues outside of oral glands, which illustrates the pitfalls of using transcriptomic data alone to define "venom toxins." We hypothesize that the python may represent an ancestral state prior to major venom development, which is supported by our finding that the expansion of venom gene families is largely restricted to highly venomous caenophidian snakes. Therefore, the python provides insight into biases in which genes were recruited for snake venom systems. Python venom homologs are generally expressed at lower levels, have higher variance among tissues, and are expressed in fewer organs compared with all other python genes. We propose a model for the evolution of snake venoms in which venom genes are recruited preferentially from genes with particular expression profile characteristics, which facilitate a nearly neutral transition toward specialized venom system expression.

Key words: origin of venom, preadaptation, ancestral expression, selective gene recruitment, toxicofera.

Introduction

Snake venoms and their evolutionary origins have received substantial attention over the past several decades (Vidal 2002; Fry et al. 2006; Casewell et al. 2014), including the evolutionary processes that have led to the toxic effects of these proteins (Casewell et al. 2013). A dominant hypothesis for the evolutionary origins of most venom toxin families involves the duplication of nontoxic genes, with subsequent neofunctionalization of gene copies to adaptively modify the structure and function of these proteins (Ivanov CP and Ivanov OC 1979; Ivanov 1981; Fujimi et al. 2003; Fry 2005; Fry et al. 2006, 2009; Tamiya and Fujimi 2006; Kini and Chinnasamy 2010; Casewell et al. 2012). Recent genome-scale resolution of this phenomenon has confirmed many of these assertions, indicating that in some cases the process of toxin gene duplication can result in expansive multilocus venom gene families, as observed in the king cobra genome (Vonk et al. 2013). Such duplication, neofunctionalization, and recruitment events appear to have occurred multiple times throughout the evolution of snakes, including multiple parallel expansion events of particular gene families in different snake lineages (Casewell et al. 2012).

There are more than 20 gene families that are traditionally considered to be "venom toxins" in squamate reptiles due primarily to their detection in venom gland secretions, and in some species, evidence for the toxicity of some of these venom components (Mackessy 2002, 2010a; Mackessy et al. 2006). The detection of expression of genes related to these "venom toxins" in venom glands or other oral glands in squamate reptiles has further become an accepted proxy for labeling such genes as "venom toxins" and the labeling of such species as "venomous" (Fry et al. 2009, 2010, 2013). Several studies, however, have shown evidence that venom genes or their homologs are expressed in tissues other than the venom or accessory venom gland of snakes and other venomous vertebrates (Rádis-Baptista et al. 2003; Whittington et al. 2008; Hargreaves et al. 2014), which calls this practice into question. Despite these inferences, there have been no comprehensive expression analyses of such "venom toxin" gene families across a broad diversity of snake organs and tissues. Thus, the degree to which venom genes or venom homolog expression in oral glands may be either a physiological default or an adaptive feature indicative

of their functional role in oral secretions remains an important yet insufficiently studied question.

Although most previous studies have focused on either gene duplication or patterns of molecular evolution of snake venoms (Fry 2005; Fox and Serrano 2008; Casewell et al. 2013), no previous studies have focused on the role that gene expression might play specifically in this venom gene recruitment process. The genes that have been targeted for recruitment into venoms appear to share certain common attributes, which support the hypothesis that successful recruitment may be linked to functional constraints of the recruited proteins (Alape-Girón et al. 1999; Fry et al. 2009). Successful recruitment of genes as venom toxins hypothetically requires a transition in which nascent venom proteins must be targeted for gene expression in specific tissues (i.e., the venom glands). Therefore understanding the evolution of expression of such genes is an essential but largely absent component for understanding their functionality, origins, and the constraints that have shaped venom repertoires. Gene expression in the venom glands of snakes has been evaluated in a number of studies (Junqueira-de-Azevedo and Ho 2002; Pahari et al. 2007; Doley et al. 2008; Fry et al. 2013; Margres et al. 2013), but due to the relative scarcity of comparative expression data for other snake tissues, venom gland gene expression is rarely viewed in the broader context of expression across diverse tissues (e.g., Hargreaves et al. 2014). It therefore remains unknown whether certain protein expression characteristics might favor their recruitment as venom toxins, or whether their expression profiles are not a relevant factor influencing recruitment.

There is uncertainty and debate over the origins of venom systems in squamate reptiles, with a common view being that a core venom system evolved a single time in the common ancestor of snakes and a clade of lizards, referred to collectively as the Toxicofera (Fry et al. 2006). This hypothesis remains controversial largely due to disagreement about what, indeed, constitutes a "venom toxin" (Terrat and Ducancel 2013) as well as a lack of apparent venom homolog expression and function in multiple large clades of Toxiceferan lizards (Fry et al. 2010, 2013). A functional definition for venom would be that it is a specialized glandular secretion which causes deleterious effects to a recipient organism when injected; this secretion is typically protein-rich and may consist of many different molecules or toxins, often representing a specialized trophic adaptation which facilitates prey handling (Mackessy 2002). However, there is continued debate of details of this definition (Nelsen et al. 2014). Current evidence indicates that a massive radiation of snakes with highly toxic venoms probably evolved after the divergence between the python and caenophidian snakes, which include, elapids, colubrids, lamprophiids, and viperids (Vidal 2002; Fry and Wüster 2004). Accordingly, recent genomic evidence from the king cobra demonstrates that many toxic venom gene families have experienced substantial duplication and divergence in the cobra relative to the python (Vonk et al. 2013). Collectively, these data indicate that the Burmese python (*Python molurus bivittatus*) may provide a system in which

to estimate patterns of gene expression prior to the expansion of highly toxic venom genes in caenophidian snakes, particularly in the highly venomous colubroid snake lineages (fig. 1). The genome and genomic resources of the nonvenomous Burmese python (Castoe et al. 2013) thereby offer a unique opportunity to study patterns of expression for genes recruited into the snake venom system within the context of a complete set of snake genes and a large set of gene expression data from diverse python tissues and organs.

In this study, we use the python genome and tissue-specific expression data to investigate the origins of venom genes in highly venomous caenophidian snakes and to assess the validity of defining genes as "venom toxins" based solely on evidence of gene expression detected in the oral glands of squamates. As a first step toward addressing these goals, we conducted thorough analyses to identify the relationships between python genes and known venom genes from caenophidian snakes and other squamate reptiles, and we provide new evidence for the orthology and patterns of gene expansion in snake venom gene families. We used these estimates of gene orthology together with python gene expression data to address two related questions: 1) Are there inherent characteristics of gene expression for venom gene homologs that may have predisposed them for recruitment as venoms? 2) Are venom gene homologs uniquely expressed or particularly abundant in python oral glands, such as the rictal gland?

Results

Estimates of Python Gene Homology to Known Venom Genes

We were able to confidently identify the homologous gene (or genes) in the python for 20 of the 24 venom gene families analyzed (table 1 and supplementary table S1, Supplementary Material online). We identified a single orthologous gene in the python for 15 of the venom gene families, whereas two homologs were found for *cystatin, metalloproteinase, phospholipase A_2 (PLA$_2$), serine proteinase*, and *veficolin*. In the case of *PLA$_2$*, however, we found two separate clades of venom genes, each with a single ortholog in the python. Our analyses resulted in the identification of a total of 25 homologs for 20 gene families (supplementary figs. S1–S20, Supplementary Material online). Phylogenetic inferences of orthology of python venom homologs in relation to known venom genes were strongly supported for 19 gene families (>95% posterior probability; supplementary figs. S1–S20, Supplementary Material online). Only the python orthologs for *exendin* had posterior support below this threshold, with 92% posterior support. In *bradykinin potentiating peptide/natriuretic peptide (BPP)* and *sarafotoxin*, orthologous sequences could not be confidently identified by phylogenetic analyses; these genes appear to have many domain insertions and deletions yielding poor alignments and it is known that *sarafotoxin* presents a unique structure which is very distinct from its putative ancestral endothelin protein (Takasaki et al. 1992; Ducancel et al. 1993). The other genes for which a python homolog could not be inferred with confidence from
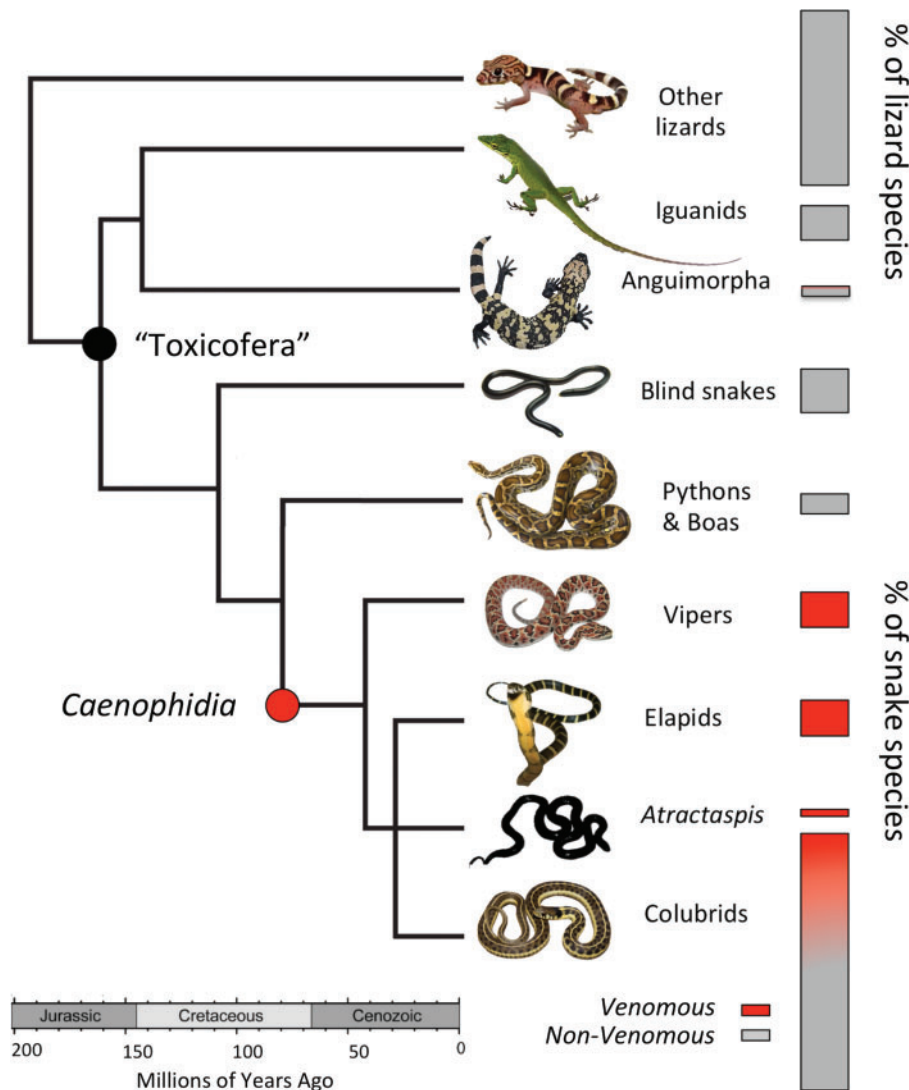
Fig. 1. Phylogenetic tree showing lizard and snake relationships and the distribution of venomous species. The black circle refers to the "Toxicofera," which includes all snakes and some lizards, and the red circle represents the Caenophidia, which contains all known deadly venomous snakes. The percentage of venomous colubrid snakes is an approximation.

phylogenetic analyses were *crotamine* and *waprin*. Several studies have found homologous sequences for these genes in nonvenomous reptiles with either low posterior support or when no reptilian outgroups were included, which we believe may result in a biased inference of gene relationships (Fry 2005; Fry et al. 2006; Vonk et al. 2013). Given an absence of quality alignments for these four gene families, we instead used protein similarity (based on the best tBLASTx hit) to estimate the probable homolog in the python for subsequent analyses. In total, further analyses therefore included 29 gene homologs for 24 gene families. Due to the controversy surrounding resolution of what qualities define a protein as a venom toxin, we also repeated all analyses including only venom protein families known to have well-defined toxic and/or cytotoxic properties (supplementary table S2, Supplementary Material online). In this case only four gene families were included: *Three-finger toxins (3FTs), metalloproteinase, serine proteinase,* and *PLA2*.

Comparison of Expression Profiles of Python Venom Homologs across Tissues

Twenty of the 29 venom homologs identified in the python show at least some level of gene expression in the python rictal gland (fig. 2A and B). Four venom homologs (*3FTs, C-type lectin, veficolin I,* and *vespryn*) show their highest levels of expression in the rictal gland. Of these, *C-type lectin* is expressed at levels that are orders of magnitude higher in the rictal gland than in any other tissues surveyed (1,000–10,000 counts per million [CPM]), whereas *3FTs, vespryn,* and *veficolin* orthologs are expressed at intermediate to high levels (100–1,000 CPM). All of the venom homologs that show expression in the rictal gland, however, show some level of expression in other python tissues. Two venom homologs, 5'-nucleotidase and *cobra venom factor,* show very high levels of expression in the liver (1,000–10,000 CPM) and phosphodiesterase is found expressed at similar levels in the small intestine. Five venom homologs are expressed at intermediate to

Table 1. Venom Gene Families Used in This Study and the Number of Orthologs Estimated in the Python and Other Snake Genomes.

Venom Gene Family	Python	Caenophidian Snakes		
		Cobra	Vipers	Rattlesnake
3FT	1	25	—	5
5'-Nucleotidase	1	1	1	—
Acetylcholinesterase	1	0	—	—
AVIToxin	1	1	—	—
BPP	1[a]	1	>1	1
C-type lectin	1	11	>6	3
Cobra Venom Factor	1	5	—	—
CRISp	1	6	1	2
Crotamine/Crotasine	1[a]	—	—	—
Cystatin	2	2	2	—
Exendin	1	1	—	—
Exonuclease	1	0	—	—
Hyaluronidase	1	0	1	—
LAAO	1	1	1	1
Metalloproteinase	2	8	>11	6
Nerve Growth Factor	1	1	1	1
Phosphodiesterase	1	0	1	1
PLA$_2$ I (Viperids)	1	1	>2	1
PLA$_2$ II (Elapids)	1	8	—	—
Sarafotoxin	1[a]	—	—	—
Serine Proteinase	2	5	>3	12
Veficolin	2	2	—	—
VEGF	1	1	1	2
Vespryn	1	1	—	—
Waprin	1[a]	—	—	1

NOTE.—Gene numbers are based on the following citations: Python (this study), cobra (Vonk et al. 2013), vipers (Casewell et al. 2009, 2014), rattlesnake (Pahari et al. 2007). Gene numbers for the Cobra are based on the complete genome sequence; estimates for vipers and the rattlesnake are based on venom gland transcriptome data and may represent a lower bound. LAAO, L-amino acid oxidase; VEGF, vascular endothelial growth factor.

[a]Orthologs represent venom genes where homology could not be inferred by gene trees.

high levels across all of the sampled tissues (fig. 2): 5'-nucleotidase, exonuclease, metalloproteinase A, phosphodiesterase, and PLA$_2$ I. Eighteen of the 29 homologs are expressed in at least half of all tissues samples, but only four of them are expressed at medium to high levels (100–1,000 CPM) in most tissues (fig. 2). In contrast, ten python venom orthologs are expressed in only seven tissues or less and at low levels (<100 CPM). Thus, although the majority of venom homologs are expressed in the rictal gland, other tissues demonstrate similar or higher levels of expression of these same genes, and the brain, small intestine and kidney had more venom homologs being expressed than the rictal gland (fig. 2).

Statistical Enrichment Analysis of Python Venom Gene Homolog Expression

Comparison of expression patterns between all other python genes versus python venom homologs indicates that python venom homolog expression is statistically different from the

patterns observed for all other genes. Venom homologs tend to be expressed at lower levels (0–1 CPM) more frequently than expected, and are less commonly expressed at intermediate levels (10–100 CPM; fig. 3A). Very similar patterns of deviation from the complete set of genes are observed when only genes with known cytotoxic activity are compared with all other python genes (fig. 3B).

To address the question of whether venom homologs tend to be expressed in more or in fewer tissues compared with all python genes, we used multiple expression levels as cutoff values for "presence" in a tissue because it is unclear what level of expression might be physiologically relevant. At the lowest threshold for presence (>1 CPM), venom homologs were enriched for higher frequencies of presence in a single tissue, and their presence was substantially underrepresented in many tissues (fig. 4A). The trend of venom homologs to be present at greater than expected frequencies in a single tissue was also found at higher thresholds of >10 CPM, >100 CPM (fig. 4B and C), and >1,000 CPM (data not shown). Finally, we asked whether the variation in venom homolog expression across tissues was significantly different than that of all python genes, and found that python venom homologs tended to show greater variation in expression levels across tissues, based on the standard error in expression levels across tissues (fig. 4D).

Discussion

Our findings provide broad evidence that there are one or two venom gene orthologs present per venom gene family in the python genome. These gene families appear to have undergone varying degrees of duplication and diversification in highly venomous caenophidian snake lineages (including elapids, viperids, and others) and in several cases, result in large multilocus gene families that encode many related toxins. The python belongs to a lineage that is the sister group to the caenophidian snakes, which appears to have diverged from caenophidian snakes prior to the expansion and diversification of major venom gene families (table 1 and fig. 1). These findings have two important ramifications. First, they suggest that regardless of when venom systems may have initially evolved in squamate reptiles, either a single time in the ancestor of the Toxicofera (Fry et al. 2006) or independently in caenophidian snakes and lizards (Kochva 1978), substantial venom gene family expansion and diversification is unlikely to have occurred in snakes prior to the caenophidian lineage (Casewell et al. 2012; Vonk et al. 2013). The availability of additional genomes from basally diverging snake lineages (e.g., blindsnakes) would be valuable to test this hypothesis further, as it is possible that instead the python secondarily lost many copies of venom genes that were duplicated early in snake or toxiceroferan evolution. However, this alternative hypothesis seems unlikely, as it would require that the python would have independently lost numerous copies of at least seven different venom gene families (see table 1). Second, our results indicate that the python provides a reasonable and valuable approximation of ancestral gene expression patterns prior to major venom gene recruitment in caenophidian snakes. Thus, patterns of venom gene homolog expression
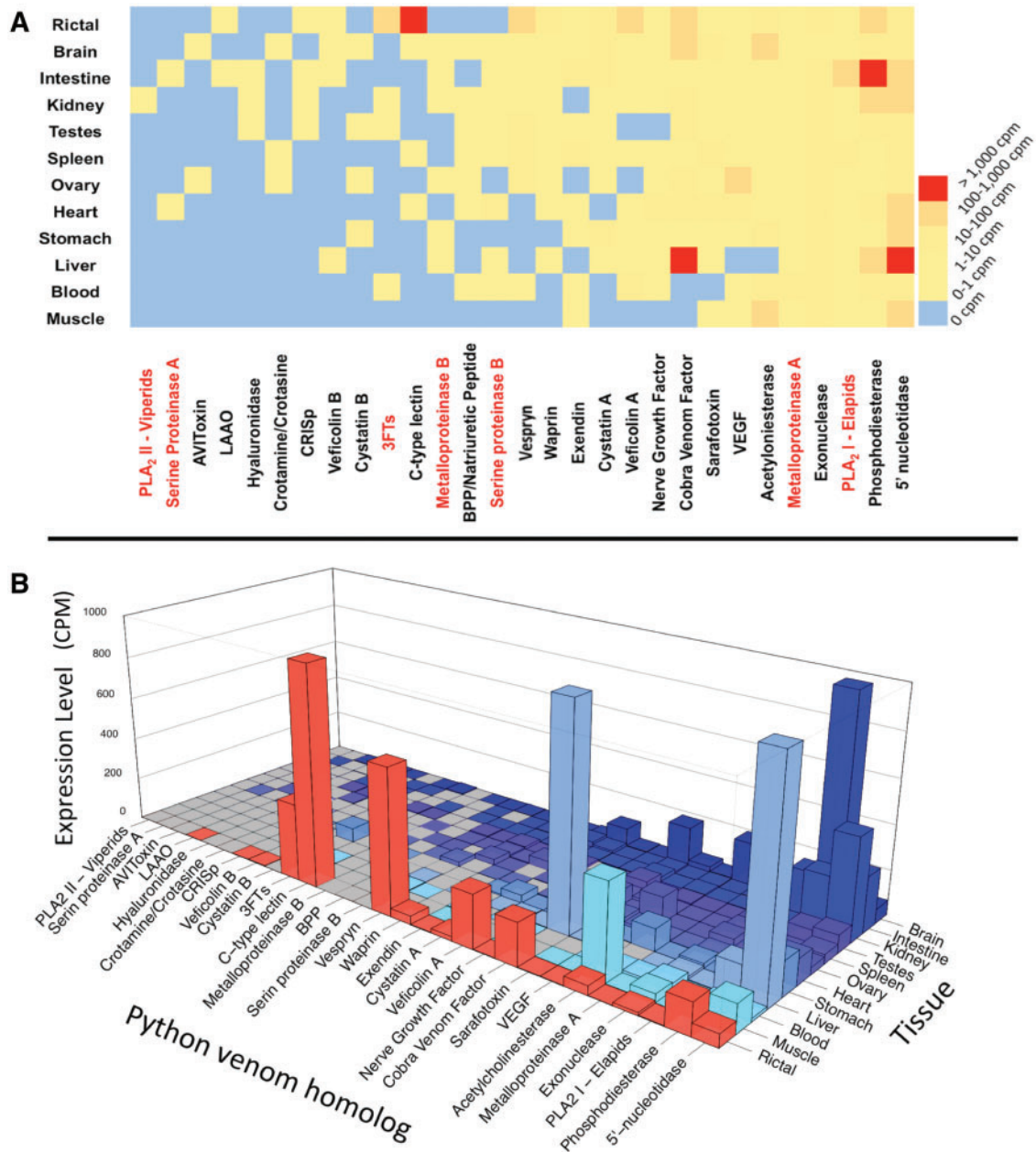
Fig. 2. Expression profiles for python venom gene homologs across tissues. (A) Heatmap of gene expression profiles shown as CPM on a log$_{10}$-scale. Names of genes with known toxicity are in red. (B) Python venom gene homolog expression with expression levels is shown in CPM. Note that the y axis (expression level) is truncated to 1,000 CPM. LAAO, L-amino acid oxidase; NGF, nerve growth factor; VEGF, vascular endothelial growth factor.

in the python may provide evidence for biases in the processes of venom gene recruitment in caenophidian snakes related to patterns of expression of ancestral venom gene homologs.

With the increasing availability of transcriptome sequencing, it has become common for researchers to sequence the transcriptome of venom glands or other oral glands of squamate reptiles and other venomous taxa (Casewell et al. 2009; Whittington et al. 2010; Fry et al. 2013; Vonk et al. 2013). Based on such data, it has also become common to identify transcripts of genes with sequence similarity to known venom toxins, to define these as "venom toxin" transcripts, and in some cases even classify a particular species as "venomous"

(Fry et al. 2009, 2010, 2013). Here we compared gene expression of python venom homologs in the rictal gland, an oral gland, with that of other python tissues. We find that although the rictal gland does indeed show expression of many venom homologs, these homologs are also expressed at comparable levels in many other tissues. In some limited cases, such venom homologs are expressed at remarkably high levels in particular organs or tissues (fig. 2). For example, brain, liver, and intestinal tissue all show moderate to high levels for several venom homologs.

Our results, including multiple examples of venom homolog expression across many tissues, argue against the adaptive and functional relevance of simply observing such transcripts
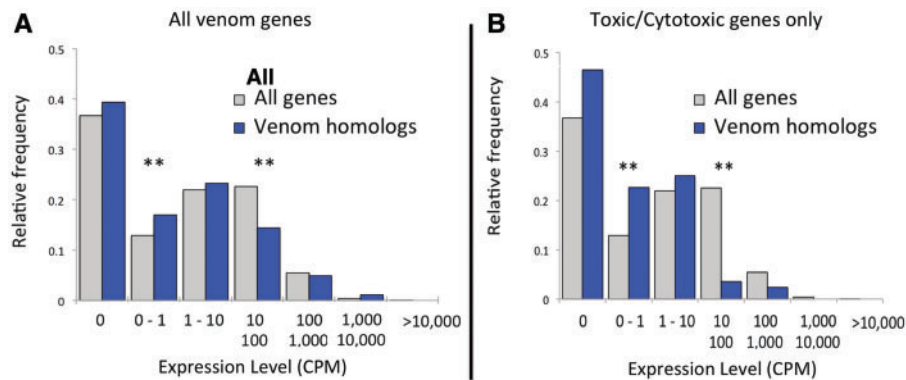
Fig. 3. Relative frequencies of genes observed at different expression levels calculated across all tissues. Results are shown for (*A*) all venom gene homologs, and (*B*) venom gene homologs that are known to be cytotoxic only. Asterisks represent expression-level bins where the difference between venom homologs and all genes is statistically significant (Fisher's exact test, *P* value < 0.05).
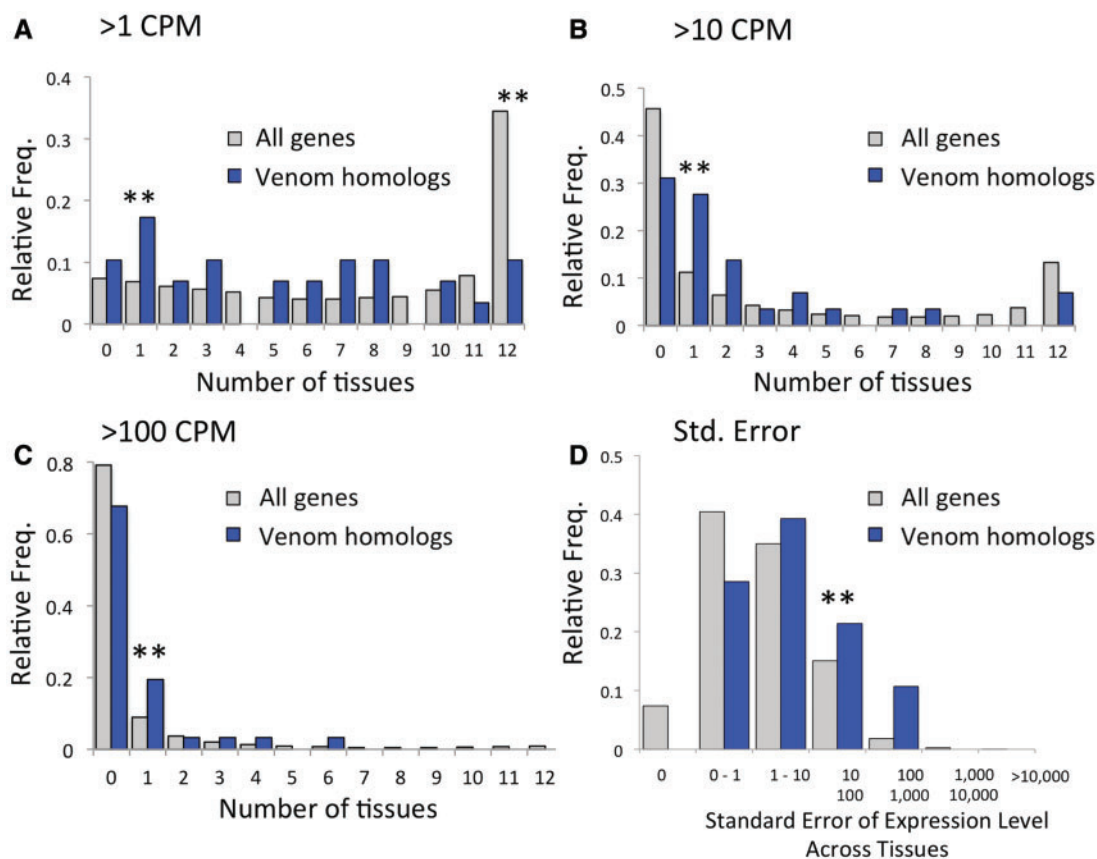

Fig. 4. The numbers of tissues in which genes are expressed and variation in expression across tissues. In (*A*)–(*C*), different CPM values are used in different panels as thresholds for the "presence" of a gene being expressed in a given tissue; (*A*) threshold ≥ 1 CPM, (*B*) threshold ≥ 10 CPM, and (*C*) threshold ≥ 100. Asterisks represent bins where the difference between venom homologs and all genes is statistically significant (Fisher's exact test, *P* value <0.05). (*D*) Comparison of standard error in expression level across tissues for all genes and venom gene homologs. Asterisks represent bins where the difference between venom homologs and all genes is statistically significant (Fisher's exact test, *P* value <0.05).

in a given tissue, as has also been argued recently by Hargreaves et al. (2014). Expression patterns in the rictal gland (fig. 2) are intriguing, particularly with regards to *3FTx* and *C-type lectin* orthologs, which at first glance appear to be consistent with previous reports of venom production in some Australian pythons (Fry et al. 2013). Interpreting these data under the Toxicofera hypothesis would suggest that the high amplification of such genes in the rictal gland might be

an artifact of the shared evolutionary history of the venom system with other toxicoferans, with the python "venom system" presumably atrophying following a switch to using constriction for prey capture (Fry et al. 2013). However, it is important to note that even these levels observed in the python rectal gland are not particularly unusual compared with expression patterns of other toxin orthologs in various nongland tissues (fig. 2). Additionally, in the absence of

functional activity data, caution is required when attempting to extrapolate from protein toxin family (e.g., 3FTx) identification to biological activity, as many toxin family members have diverse actions, which are difficult to correlate with structure. For example, proteins with the canonical 3FTx fold and highly conserved disulfides have pharmacological activities as diverse as neurotoxins and anticoagulants (Heyborne and Mackessy 2013) to salamander pheromones (Palmer et al. 2007) and regulators of limb regeneration (Garza-Garcia et al. 2009). Thus, using such data singularly from an oral gland and reaching the conclusion that venom homolog expression represents evidence of "venom toxin" production, or "venomousness" of a species, would be baseless without additional evidence for a functional role of such gene products. An old report suggested toxicity of rictal gland material from Eryx and Uropeltis to birds (Phisalix and Caius 1918), so it is possible that the rictal gland of Python could have toxic properties, but this hypothesis has yet to be tested; further, the relevance of dosages and identity of material used is unclear. We also suggest caution when interpreting toxicity data outside of a biologically relevant context; for example, although submaxillary gland secretions of male Mus musculus were rather toxic to rats, the interpretation of this secretion as either an offensive or defensive venom is inconclusive at best, as rats quickly kill mice using mechanical means, in spite of the submaxillary glands.

Our results indicate that the probability of successful recruitment of a particular gene for use in caenophidian venom systems may have been biased by the ancestral expression pattern of that gene. Compared with all other python gene expression profiles, python venom homologs tend to be expressed at lower levels overall, expressed at moderate–high levels in fewer tissues, and show among the highest variation in expression level across tissues. These python venom homologs also tend to have higher expression in a single tissue and tend not be expressed in all tissues.

In highly venomous caenophidian snakes, recent studies have shown that highly toxic venom proteins are expressed at moderate to high levels in the venom gland and low–moderate levels in the accessory venom glands (Vonk et al. 2013), but there are only limited data on their expression levels in other tissues. What is known about their expression in diverse tissues pertains only to their presence/absence (Hargreaves et al. 2014), which substantially limits insight into their relative biological activity in those tissues, particularly as we find here that python venom homologs may be expressed at levels that span more than 4 orders of magnitude across tissues. Many caenophidian venom toxins are known to be cytotoxic (Lee 1972), to the extent that they are difficult to study in expression vectors (Brenes et al. 2010); within the caenophidian venom gland, redundant mechanisms maintain these venom toxins in a competent but inactive state (Mackessy and Baxter 2006). The expression of such genes at high (biologically active) levels in other nonvenom-related tissues would thus likely be deleterious. These data collectively indicate that during the evolutionary recruitment of such venom toxins in caenophidian snakes, the evolution of venom protein toxicity and higher levels of "venom toxin" expression

in the venom system must have been coordinated with an increase in the degree to which such a toxin's expression is confined to venom system tissues. For most venom gene families in caenophidian snakes, this process also appears to be coupled with gene duplication and neofunctionalization through accelerated point mutation (Nakashima et al. 1995; Deshimaru et al. 1996; Kordiš and Gubenšek 2000), accelerated segment switch in exons (Doley et al. 2008, 2009) and other mechanisms, resulting in a diversity of functionalities housed within a conserved protein scaffold. This functional diversification has been well documented for most of the potent functional toxins of caenophidian snake venoms, including 3FTs, PLA$_2$s, serine proteinases, and metalloproteinases (Lynch 2007; Vaiyapuri et al. 2011; Brust et al. 2013; Sunagar et al. 2013).

Based on biases in the regulatory characteristics we have identified in venom homologs in the python, we propose a stepwise model for how proto-venom genes with such regulatory characteristics might have originally been recruited into snake venom systems. We refer to this model as the stepwise intermediate nearly neutral evolutionary recruitment (SINNER) model. This model has three main steps which may or may not involve gene duplication: 1) Expression of proto-venom genes in oral secretory glands at low levels, which is favored as a default by regulatory architecture favoring low near constitutive expression, 2) switching of tissue-specific higher expression levels to target oral/venom glands, and 3) reduction in expression levels in nonvenom-related tissues that is driven by the degree of toxicity to the tissue itself. In this model, the evolution of toxicity (i.e., neofunctionalization) would be constrained by two factors: The functional requirements and expression levels of the protein in nonvenom tissues. Gene duplication would release the first of these two constraints, allowing the evolution of reduced expression in nonvenom tissues, and thus allowing the evolution of greater toxicity to prey. This SINNER model therefore implies the existence of a nearly neutral intermediate phase during which the pace of evolution of the toxicity of a venom gene product is balanced by the tissue-specificity and magnitude of its expression, and it accounts for variation in the evolution of toxicity of such venom homologs in various lineages due to differential patterns of drift and selection. The SINNER model thus successfully predicts that a large number of different gene families may exist in venom systems and possess members with different toxicity and expression levels in different lineages, as the expression of those gene families in the venom gland (or any tissue) is a physiological default to some extent. Also, different genes may occupy one of an infinite number of steps along the continuum of the recruitment model's nearly neutral intermediate landscape due to both selection and drift. For example, even though 3FTs do not constitute the main components of viperid venom, they are still expressed in viperid venom glands (Pahari et al. 2007); on the other hand, metalloproteinases and serine proteinases, both important components of viper venom, but not of elapid venom, are still expressed but at low levels in elapid venom

glands (Correa-Netto et al. 2011; Jiang et al. 2011; Margres et al. 2013).

Some venom genes are also known to produce multiple splice variants (Ducancel et al. 1993; Cousin et al. 1998; Siigur et al. 2001), and it is relevant to consider how these alternative transcripts may contribute to evolution under the SINNER model. If alternative splicing were capable of producing toxic and nontoxic peptides from the same gene, this would decrease the relative role of gene duplication, and would also increase the number of evolutionarily labile features that could act to shift venom toxin expression toward venom gland specificity. Specifically in the case of alternative splice variants, evolution could act on small interfering RNAs and spliceosomal components, in addition to promoter/enhancer/repressor regulatory elements, to accomplish venom gland targeting of toxic peptides; thus, alternative splicing may act to increase the evolvability and rate of progression of genes across the continuum of the SINNER model.

Although we have developed the SINNER model of gene functional recruitment based on snake venom genes, many aspects of this model may apply equally well to other instances of evolutionary co-option of genes that involve duplication and sub-/neo-functionalization. Particularly when there is selection for novel tissue function (e.g., salivary-to-venom gland function), genes that are essentially constitutively expressed in many tissues at low levels and at higher levels in a small number of tissues may be important "raw material" for shifting tissue function through co-option of these genes in a variety of biological circumstances. It is likely that the SINNER model of gene co-option and recruitment may also fit the evolution of venom systems in other animals, and comparative analysis of gene expression across diverse tissues and venomous and nonvenomous sister lineages will be important for evaluating the explanatory power of this model in these systems. One prediction of the SINNER model is that a venom repertoire should contain a diverse collection of gene families, some of which are expressed as a physiological default, and some will be intermediate on the spectrum between high secretion level, venom system specificity, and toxicity, and thus will not be particularly toxic. For example, even though 3FTs do not constitute the main toxic components of viperid venom, they are still expressed in viperid venom glands (Pahari et al. 2007). Similarly, metalloproteinases and serine proteinases, both important functional components of viper venom, but not of elapid venom, are expressed at low levels in elapid venom glands (Correa-Netto et al. 2011; Jiang et al. 2011; Margres et al. 2013). Some of the most common venom components include CRISp (cysteine-rich secretory protein), waprin/kunitz, hyaluronidases, serine proteases, and PLA$_2$, among many others (Fry et al. 2009), and even though the same venom protein families can be found across venoms of several animal phyla, their unique patterns of expression, functionality, and toxicity can vary considerably among species (e.g., Kreil 1995; Ma et al. 2010; Whittington et al. 2010; Ruder et al. 2013; Undheim et al. 2014), which is consistent with predictions from the SINNER model.

Similar to our study, a recent study also found evidence for the presence of venom homologs and known venom genes in diverse tissues of nonvenomous and venomous snakes, respectively (Hargreaves et al. 2014). Based on these data, the authors argue for a shift in the otherwise broadly accepted model of venom gene duplication and recruitment, and suggesting instead that this process be viewed as "restriction" rather than recruitment because venom genes do not appear to be targeted de novo to venom glands but instead are "restricted" to venom systems over evolutionary time. Their conclusions do share some aspects of our SINNER model in that venom genes are not likely de novo targeted to the venom gland, but instead undergo a spectral evolutionary transition toward venom gland-specific targeting. Analysis of next-generation RNA-sequencing data to measure expression is so highly sensitive to extremely rarely expressed transcripts, however, that their use of a "presence–absence" detection of venom-related transcripts is potentially misleading and is capable of detecting transcripts far below the levels at which they will produce physiologically relevant biologically active proteins. Thus, future work examining organism-wide patterns of venom gene expression should carefully consider the relative frequencies of venom homologs in the context of estimating patterns of expression across tissues to differentiate between biologically relevant expression levels and extremely rare transcripts due, for example, to slightly "leaky" promoters.

As additional genomic and transcriptomic information becomes available for snakes, particularly different lineages of highly venomous caenophidian snakes as well as in more basally diverging lineages of snakes (e.g., Scolecophidian blindsnakes) and toxicoferan lizards, it will be interesting to further test the SINNER model for snake venom gene recruitment, and the hypothesis that venom gene expansion occurred "late," in the caenophidian lineage. Such diverse sampling across the toxicoferan tree is ultimately required to more definitively determine how evolution has shaped tissue expression patterns of venom homologs in the development of squamate venom systems. The SINNER model, our data from the python, other evidence for venom homolog expression in multiple nonvenom gland tissues in other venomous and nonvenomous snakes (Rádis-Baptista et al. 2003; Whittington et al. 2008; Hargreaves et al. 2014), and the lack of evidence for toxicity or function of multiple venom components relevant for prey capture (Lavin 2010; Ahmed et al. 2012; Fry et al. 2012), collectively suggest that a strict and static definition of a gene family as representing "venom toxins" is inaccurate. Instead, these data indicate that a set of venom gland (or other oral gland) secretions may represent a collection of proteins that span the full continuum of stages in the evolution of toxicity and functionality as venoms, some of which may be present largely due to random processes rather than selection for function as venom. Thus, the definition of proteins as "venom toxins" based solely on homology in the absence of functional evidence of toxic effects on prey (or other functional advantages for prey handling) may be misleading. Accordingly, our results indicate the need for a critical re-evaluation of the criteria required to

consider a protein a "venom toxin" across the tree of life, not only in snakes. We suggest that such criteria should incorporate more direct evidence for the toxicity or function of such proteins in prey handling.

Materials and Methods

BLAST Analyses to Identify Python Gene Homologs of Known Venom Genes

We studied a total of 24 venom gene families (Mackessy 2002, 2010a, 2010b), which we obtained examples from GenBank (table 1 and supplementary table S2, Supplementary Material online). These 24 venom gene families represent the vast majority of known squamate venoms, and the only one with available DNA sequences. To identify homologous genes in other lineages, we blasted each venom gene to the complete protein coding sequences (CDSs) of the human, anole lizard, Burmese python, and king cobra using tBLASTx. CDS files were obtained from Ensembl (Flicek et al. 2014) and from recently published snake genomes (Castoe et al. 2013; Vonk et al. 2013). From each BLAST search, we retained the top three hits for each taxon based on its E value (E value <1e-05), and the top three hits based on bit scores (bit score > 70). If neither criterion was met, we retained the highest E value hit and the gene with the highest bit score for each queried species. To increase phylogenetic resolution, we included additional sequences from several other vertebrate species from GenBank, and sampling used previously (Vonk et al. 2013).

Phylogenetic Analysis to Identify Gene Homologs

We conducted first-pass alignments of translated amino acid sequences using Muscle (Edgar 2004). Once aligned, sequences were converted to nucleotides, and nucleotide-level alignments were used for all subsequent analyses. We estimated best-fit models of nucleotide evolution using PartitionFinder (Lanfear et al. 2012). We inferred phylogenies in MrBayes version 3.2.1 (Ronquist et al. 2012). For each gene, we ran two simultaneous analyses of 10^7 generations and sampled the chain every 10^3 generations. We confirmed mixing and convergence using Tracer V.1.5 (Rambaut and Drummond 2007) and discarded the first 10% of all runs as burn-in. After first-pass analysis, we identified nonhomologous sequences as those with extremely long branches and very low posterior support (<50%), and these sequences were removed from alignments, alignments were reoptimized, and we estimated new phylogenetic trees based on these revised alignments.

Analysis of Gene Expression Data from the Python

We used all gene expression data available for the Burmese python (Castoe et al. 2013). Where available, expression data from multiple individuals were combined per tissue for all analyses. We normalized read counts using TMM normalization in edgeR (Robinson et al. 2010) and converted read counts to CPM. We used our phylogenetic estimates for each of the 24 venom gene families to identify venom gene homologs in the python (table 1), and we used the term

homolog to refer to multiple situations, including evidence of orthology (including 1:1 orthology) and other instances where our best estimate is based on a BLAST-based hit. We categorized patterns of gene expression in several ways and compared these patterns between python venom homologs and the complete python gene set. We assigned all python genes to one of seven different log-scale categories based on their normalized expression levels in a given tissue: 1) CPM = 0; 2) CPM = 0–1; 3) CPM = 1–10; 4) CPM = 10–100; 5) CPM = 100–1,000; 6) CPM = 1,000–10,000; and 7) CPM ≥ 10,000. We compared the pattern of expression levels between venom gene homologs and all other python genes using a Fisher's exact test. For each gene, we also calculated the mean and variance in expression level across all tissues combined and tested for differences between venom homologs and all genes using Fisher's exact tests (supplementary table S5, Supplementary Material online). Because it is unclear what level of gene expression might be biologically relevant, we used multiple thresholds of CPM read counts for "presence" of a gene in a given tissue: 1) CPM > 1; 2) CPM > 10; 3) CPM > 100; 4) CPM > 1,000; and 5) CPM > 10,000 (fig. 4). Significant differences between venom homologs and all other genes were tested using Fisher's exact tests.

Supplementary Material

Supplementary figures S1–S20 and tables S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

References

Ahmed M, Latif N, Khan R, Ahmad A, Rocha J, Mazzanti C, Bagatini M, Morsch V, Schetinger M. 2012. Enzymatic and biochemical characterization of *Bungarus sindanus* snake venom acetylcholinesterase. *J Venom Anim Toxins Incl. Trop. Dis.* 18:236–243.

Alape-Girón A, Persson B, Cederlund E, Flores-Díaz M, Gutiérrez JM, Thelestam M, Bergman T, JoÈrnvall H. 1999. Elapid venom toxins: multiple recruitments of ancient scaffolds. *Eur J Biochem.* 259: 225–234.

Brenes O, Munoz E, Roldan-Rodriguez R, Diaz C. 2010. Cell death induced by *Bothrops asper* snake venom *metalloproteinase* on endothelial and other cell lines. *Exp Mol Pathol.* 88:424–432.

Brust A, Sunagar K, Undheim EA, Vetter I, Yang DC, Caswell NR, Jackson TN, Koludarov I, Alewood PF, Hodgson WC. 2013. Differential evolution and neofunctionalization of snake venom *metalloprotease* domains. *Molecular & Cellular Proteomics.* 12: 651–663.

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

Casewell NR, Huttley GA, Wuster W. 2012. Dynamic evolution of venom proteins in squamate reptiles. *Nat Commun.* 3:1066.

Casewell NR, Wagstaff SC, Wüster W, Cook DN, Bolton FM, King SI, Pla D, Sanz L, Calvete JJ, Harrison RA. 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, Wuster W, Vonk FJ, Harrison RA, Fry BG. 2013. Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol Evol.* 28: 219–229.

Castoe TA, de Koning APJ, Hall KT, Card DC, Schield DR, Fujita MK, Ruggiero RP, Degner JF, Daza JM, Gu W, et al. 2013. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc Natl Acad Sci U S A.* 110:20645–20650.

Correa-Netto C, Junqueira-de-Azevedo IDM, Silva DA, et al. 2011. Snake venomics and venom gland transcriptomic analysis of Brazilian coral snakes, *Micrurus altirostris* and *M. corallinus. J Proteomics.* 74: 1795–1809.

Cousin X, Bon S, Massoulié J, Bon C. 1998. Identification of a novel type of alternatively spliced exon from the *acetylcholinesterase* gene of *Bungarus fasciatus*. Molecular forms of *acetylcholinesterase* in the snake liver and muscle. *J Biol Chem.* 273:9812–9820.

Deshimaru M, Ogawa T, Nakashima K, Nobuhisa I, Chijiwa T, Shimohigashi Y, Fukumaki Y, Niwa M, Yamashina I, Hattori S. 1996. Accelerated evolution of crotalinae snake venom gland *serine proteases. FEBS Lett.* 397:83–88.

Doley R, Mackessy SP, Kini RM. 2009. Role of accelerated segment switch in exons to alter targeting (ASSET) in the molecular evolution of snake venom proteins. *BMC Evol Biol.* 9:146.

Doley R, Tram NN, Reza MA, Kini RM. 2008. Unusual accelerated rate of deletions and insertions in toxin genes in the venom glands of the pygmy copperhead (*Austrelaps labialis*) from Kangaroo island. *BMC Evol Biol.* 8:70.

Ducancel F, Matre V, Dupont C, Lajeunesse E, Wollberg Z, Bdolah A, Kochva E, Boulain J-C, Menez A. 1993. Cloning and sequence analysis of cDNAs encoding precursors of *sarafotoxins*. Evidence for an unusual "rosary-type" organization. *J Biol Chem.* 268:3052–3055.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.

Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S. 2014. Ensembl 2014. *Nucleic Acids Res.* 42:D749–D755.

Fox JW, Serrano SM. 2008. Exploring snake venom proteomes: multifaceted analyses for complex toxin mixtures. *Proteomics* 8:909–920.

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

Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, Nevalainen TJ, Norman JA, Lewis RJ, Norton RS. 2012. The structural and functional diversification of the Toxicofera reptile venom system. *Toxicon* 60:434–448.

Fry BG, Undheim EA, Ali SA, Jackson TN, Debono J, Scheib H, Ruder T, Morgenstern D, Cadwallader L, Whitehead D, et al. 2009. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu Rev Genomics Hum Genet.* 10: 483–511.

Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SR, Kuruppu S, Fung K, Hedges SB, Richardson MK, et al. 2013. Squeezers and leaf-cutters: differential diversification and degeneration of the venom system in toxiceran reptiles. *Mol Cell Proteomics.* 12: 1881–1899.

Fry BG, Winter K, Norman JA, Roelants K, Nabuurs RJ, van Osch MJ, Teeuwisse WM, van der Weerd L, Mcnaughtan JE, Kwok HF. 2006. Early evolution of the venom system in lizards and snakes. *Nature* 439:584–588.

Fry BG, Winter K, Norman JA, Roelants K, Nabuurs RJ, van Osch MJ, Teeuwisse WM, van der Weerd L, Mcnaughtan JE, Kwok HF. 2010. Functional and structural diversification of the Anguimorpha lizard venom system. *Mol Cell Proteomics.* 9:2369–2390.

Fry BG, Wüster W. 2004. Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. *Mol Biol Evol.* 21:870–883.

Fujimi T, Nakajyo T, Nishimura E, Ogura E, Tsuchiya T, Tamiya T. 2003. Molecular evolution and diversification of snake toxin genes, revealed by analysis of intron sequences. *Gene* 313:111–118.

Garza-Garcia A, Harris R, Esposito D, Gates PB, Driscoll PC. 2009. Solution structure and phylogenetics of *Prod1*, a member of the three-finger protein superfamily implicated in salamander limb regeneration. *PLoS One* 4:e7123.

Hargreaves AD, Swain MT, Hegarty MJ, Logan DW, Mulley JF. 2014. Restriction and recruitment-gene duplication and the origin and evolution of snake venom toxins. *Genome Biol Evol.* 6: 2088–2095.

Heyborne WH, Mackessy SP. 2013. Identification and characterization of a taxon-specific *three-finger toxin* from the venom of the Green Vinesnake (*Oxybelis fulgidus*; family Colubridae). *Biochimie* 95: 1923–1932.

Ivanov CP, Ivanov OC. 1979. The evolution and ancestors of toxic proteins. *Toxicon* 17:205–220.

Ivanov OC. 1981. The evolutionary origin of toxic proteins. *Toxicon* 19: 171–178.

Jiang Y, Li Y, Lee W, Xu X, Zhang Y, Zhao R, Zhang Y, Wang W. 2011. Venom gland transcriptomes of two elapid snakes (*Bungarus multicinctus* and *Naja atra*) and evolution of toxin genes. *BMC Genomics* 12:1.

Junqueira-de-Azevedo IdL, Ho PL. 2002. A survey of gene expression and diversity in the venom glands of the pitviper snake *Bothrops insularis* through the generation of expressed sequence tags (ESTs). *Gene* 299: 279–291.

Kini RM, Chinnasamy A. 2010. Nucleotide sequence determines the accelerated rate of point mutations. *Toxicon* 56:295–304.

Kochva E. 1978. Oral glands of the Reptilia. In: Gans C, editor. Biology of the Reptilia, Vol. 8. London: Academic Press. p. 43–161.

Kordiš D, Gubenšek F. 2000. Adaptive evolution of animal toxin multigene families. *Gene* 261:43–52.

Kreil G. 1995. Hyaluronidases—a group of neglected enzymes. *Protein Sci.* 4:1666–1669.

Lanfear R, Calcott B, Ho SY, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol.* 29:1695–1701.

Lavin MF, Earl S, Birrell G, St. Pierre L, Guddat L, de Jersey J, Masci P. 2010. Snake venom nerve growth factors. In: SP M, editor. Handbook of venoms and toxins of reptiles. Boca Raton (FL): CRC Press. p. 377–392.

Lee C. 1972. Chemistry and pharmacology of polypeptide toxins in snake venoms. *Annu Rev Pharmacol.* 12:265–286.

Lynch VJ. 2007. Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. *BMC Evol Biol.* 7:2.

Ma Y, Zhao Y, Zhao R, Zhang W, He Y, Wu Y, Cao Z, Guo L, Li W. 2010. Molecular diversity of toxic components from the scorpion *Heterometrus petersii* venom revealed by proteomic and transcriptome analysis. *Proteomics* 10:2471–2485.

Mackessy SP. 2002. Biochemistry and pharmacology of colubrid snake venoms. *Toxin Rev.* 21:43–83.

Mackessy SP, editor. 2010a. The field of reptile toxinology: snakes, lizards and their venoms. In: Handbook of venoms and toxins of reptiles. Boca Raton (FL): CRC Press. p. 3–23.

Mackessy SP. 2010b. Evolutionary trends in venom composition in the Western Rattlesnakes (*Crotalus viridis* sensu lato): toxicity vs. tenderizers. *Toxicon* 55:1463–1474.

Mackessy SP, Baxter LM. 2006. Bioweapons synthesis and storage: the venom gland of front-fanged snakes. *Zool Anz.* 245: 147–159.

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

Margres MJ, Aronow K, Loyacano J, Rokyta DR. 2013. The venom-gland transcriptome of the eastern coral snake (*Micrurus fulvius*) reveals high venom complexity in the intragenomic evolution of venoms. *BMC Genomics* 14:1–18.

Nakashima K-I, Nobuhisa I, Deshimaru M, 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.* 92:5605–5609.

Nelsen DR, Nisani Z, Cooper AM, Fox GA, Gren EC, Corbit AG, Hayes WK. 2014. Poisons, toxungens, and venoms: redefining and classifying toxic biological secretions and the organisms that employ them. *Biol Rev.* 89:450–465.

Pahari S, Mackessy SP, Kini RM. 2007. The venom gland transcriptome of the Desert Massasauga Rattlesnake (*Sistrurus catenatus edwardsii*): towards an understanding of venom composition among advanced snakes (Superfamily Colubroidea). *BMC Mol Biol.* 8:115.

Palmer CA, Hollis DM, Watts RA, Houck LD, McCall MA, Gregg RG, Feldhoff PW, Feldhoff RC, Arnold SJ. 2007. Plethodontid modulating factor, a hypervariable salamander courtship pheromone in the three-finger protein superfamily. *FEBS J.* 274:2300–2310.

Phisalix M, Caius R. 1918. L'extension de la fonction venimeuse dans l'ordre entière des ophidiens et son existence chez des familles ou elle n'avait pas été soupçonnée jusqu'içi. *Journal de Physiologie et de Pathologie Générale* 17:923–964.

Rádis-Baptista G, Kubo T, Oguiura N, Svartman M, Almeida T, Batistic RF, Oliveira EB, Vianna-Morgante ÂM, Yamane T. 2003. Structure and chromosomal localization of the gene for *crotamine*, a toxin from the South American rattlesnake, *Crotalus durissus terrificus*. *Toxicon* 42:747–752.

Rambaut A, Drummond A. 2007. Tracer v. 1.5. Computer program and documentation distributed by the authors. [cited 2014 Oct 20]. Available from: http://beast.bio.ed.ac.uk/Tracer.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61:539–542.

Ruder T, Sunagar K, Undheim E, Ali SA, Wai T, Low D, Jackson T, King GF, Antunes A, Fry BG. 2013. Molecular phylogeny and evolution of

the proteins encoded by coleoid (cuttlefish, octopus, and squid) posterior venom glands. *J Mol Evol.* 76:192–204.

Siigur E, Aaspõllu A, Siigur J. 2001. Sequence diversity of *Vipera lebetina* snake venom gland *serine proteinase* homologs—result of alternative-splicing or genome alteration. *Gene* 263:199–203.

Sunagar K, Jackson TN, Undheim EA, Ali SA, Antunes A, Fry BG. 2013. Three-fingered RAVERs: rapid accumulation of variations in exposed residues of snake venom toxins. *Toxins* 5:2172–2208.

Takasaki C, Itoh Y, Onda H, Fujino M. 1992. Cloning and sequence analysis of a snake, *Atractaspis engaddensis* gene encoding *sarafotoxin* S6C. *Biochem Biophys Res Commun.* 189:1527–1533.

Tamiya T, Fujimi TJ. 2006. Molecular evolution of toxin genes in Elapidae snakes. *Mol Divers.* 10:529–543.

Terrat Y, Ducancel F. 2013. Are there unequivocal criteria to label a given protein as a toxin? Permissive versus conservative annotation processes. *Genome Biol.* 14:406.

Undheim EA, Jones A, Clauser KR, Holland JW, Pineda SS, King GF, Fry BG. 2014. Clawing through evolution: toxin diversification and convergence in the ancient lineage Chilopoda (Centipedes). *Mol Biol Evol.*

Vaiyapuri S, Wagstaff SC, Harrison RA, Gibbins JM, Hutchinson EG. 2011. Evolutionary analysis of novel serine proteases in the venom gland transcriptome of *Bitis gabonica rhinoceros*. *PLoS One* 6:e21532.

Vidal N. 2002. Colubroid systematics: evidence for an early appearance of the venom apparatus followed by extensive evolutionary tinkering. *Toxin Rev.* 21:21–41.

Vonk FJ, Casewell NR, Henkel CV, 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.* 110:20651–20656.

Whittington CM, Papenfuss AT, Kuchel PW, Belov K. 2008. Expression patterns of platypus *defensin* and related venom genes across a range of tissue types reveal the possibility of broader functions for *OvDLPs* than previously suspected. *Toxicon* 52:559–565.

Whittington CM, Papenfuss AT, Locke DP, et al. 2010. Novel venom gene discovery in the platypus. *Genome Biol.* 11:R95.