Title page

Title:

Testing the Toxicofera: comparative reptile transcriptomics casts doubt on the single, early evolution of the reptile venom system

Authors:

Adam D Hargreaves[1], Martin T Swain[2], Darren W Logan[3] and John F Mulley[1]*

Affiliations:

1. School of Biological Sciences, Bangor University, Brambell Building, Deiniol Road, Bangor, Gwynedd, LL57 2UW, United Kingdom

2. Institute of Biological, Environmental & Rural Sciences, Aberystwyth University, Penglais, Aberystwyth, Ceredigion, SY23 3DA, United Kingdom

3. Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1HH, United Kingdom

*To whom correspondence should be addressed: j.mulley@bangor.ac.uk

Author email addresses:

John Mulley - j.mulley@bangor.ac.uk

Adam Hargreaves - bspa1b@bangor.ac.uk

Darren Logan - dl5@sanger.ac.uk

Martin Swain - mts11@aber.ac.uk

Abstract

Background

The identification of apparently conserved gene complements in the venom and salivary glands of a diverse set of reptiles led to the development of the Toxicofera hypothesis – the idea that there was a single, early evolution of the venom system in reptiles. However, this hypothesis is based largely on relatively small scale EST-based studies of only venom or salivary glands and toxic effects have been assigned to only some of these putative Toxcoferan toxins in some species. We set out to investigate the distribution of these putative venom toxin transcripts in order to investigate to what extent conservation of gene complements may reflect a bias in previous sampling efforts.

Results

We have carried out the first large-scale test of the Toxicofera hypothesis and found it lacking in a number of regards. Our quantitative transcriptomic analyses of venom and salivary glands and other body tissues in five species of reptile, together with the use of available RNA-Seq datasets for additional species shows that the majority of genes used to support the establishment and expansion of the Toxicofera are in fact expressed in multiple body tissues and most likely represent general maintenance or "housekeeping" genes. The apparent conservation of gene complements across the Toxicofera therefore reflects an artefact of incomplete tissue sampling. In other cases, the identification of a non-toxic paralog of a gene encoding a true venom toxin has led to confusion about the phylogenetic distribution of that venom component.

Conclusions

Venom has evolved multiple times in reptiles. In addition, the misunderstanding regarding what constitutes a toxic venom component, together with the misidentification of genes and the classification of identical or near-identical sequences as distinct genes has led to an

overestimation of the complexity of reptile venoms in general, and snake venom in particular, with implications for our understanding of (and development of treatments to counter) the molecules responsible for the physiological consequences of snakebite.

Keywords

Snake venom, Toxicofera, Transcriptomics,

Background

Snake venom is frequently cited as being highly complex or diverse [1-3] and a large number of venom toxin genes and gene families have been identified, predominantly from EST-based studies of gene expression during the re-synthesis of venom in the venom glands following manually-induced emptying ("milking") [4-8] and some proteomic studies of extracted venom. It has been suggested that many of these gene families have originated via the duplication of a gene encoding a non-venom protein expressed elsewhere in the body followed by recruitment into the venom gland where natural selection can act to increase toxicity, with subsequent additional duplications leading to a diversification within gene families, often in a species-specific manner [9-11]. However, since whole genome duplication is a rare event in reptiles [12], the hypothesis that novelty in venom originates via the duplication of a "body" gene with subsequent recruitment into the venom gland requires both that gene duplication is a frequent event in the germline of venomous snakes and that the promoter and enhancer sequences that regulate venom gland-specific expression are relatively simple and easy to evolve. It also suggests a high incidence of neofunctionalisation rather than the more common process of subfunctionalisation [13-16].

The apparent widespread distribution of genes known to encode venom toxins in snakes in the salivary glands of a diverse set of reptiles, including both those that had previously been suggested to have secondarily lost venom in favour of constriction or other predation

79 techniques and those that had previously been considered to have never been venomous led to
80 the development of the Toxicofera hypothesis – the single, early evolution of venom in
81 reptiles [17-19] (Figure 1). Analysis of a wide range of reptiles, including charismatic
82 megafauna such as the Komodo dragon, *Varanus komodoensis* [20], has shown that the basal
83 Toxicoferan venom system comprises at least 16 genes, with additional gene families
84 subsequently recruited in different lineages [18, 19, 21].
85 Although toxic effects have been putatively assigned to some Toxicoferan venom proteins in
86 some species, the problem remains that their identification as venom components is based
87 largely on their expression in the venom gland during venom synthesis and their apparent
88 relatedness to other, known toxins in phylogenetic trees. It has long been known that all
89 tissues express a basic set of "housekeeping" or maintenance genes [22] and it is therefore
90 not surprising that similar genes might be found to be expressed in similar tissues in different
91 species of reptiles and that these genes might group together in phylogenetic trees. However,
92 the identification of transcripts encoding putative venom toxins in other body tissues would
93 cast doubt on the classification of these Toxicoferan toxins as venom components, as it is
94 unlikely that the same gene could fulfil toxic and non-toxic roles without evidence for
95 alternative splicing to produce a toxic variant (as has been suggested for *acetylcholinesterase*
96 in the banded krait, *Bungarus fasciatus* [11, 23]) or increased expression levels in the venom
97 gland (where toxicity might be dosage dependent). In order to address some of these issues
98 and to test the robustness of the Toxicofera hypothesis we have carried out a comparative
99 transcriptomic survey of the venom or salivary glands, skin and cloacal scent glands of five
100 species of reptile. Unlike the pancreas and other parts of the digestive system [24, 25], these
101 latter tissues (which include a secretory glandular tissue (the scent gland) and a relatively
102 inert, non-secretory tissue (skin)) have not previously been suggested to be the source of
103 duplicated venom toxin genes and we would therefore only expect to find ubiquitous
104 maintenance or "housekeeping" genes to be commonly expressed across these tissues. Study

105 species included the venomous painted saw-scaled viper (*Echis coloratus*); the non-
106 venomous corn snake (*Pantherophis guttatus*) and rough green snake (*Opheodrys aestivus*)
107 and a member of one of the more basal extant snake lineages, the royal python (*Python*
108 *regius*). As members of the Toxicofera *sensu* Fry et al. [21] we would expect to find the basic
109 Toxicoferan venom genes expressed in the venom or salivary glands of all of these species. In
110 addition we generated corresponding data for the leopard gecko (*Eublepharis macularius*), a
111 member of one of the most basal lineages of squamate reptiles that lies outside of the
112 proposed Toxicofera clade (Figure 1). We have also taken advantage of available
113 transcriptomes or RNA-Seq data for corn snake vomeronasal organ [26] and brain [27], garter
114 snake (*Thamnophis elegans*) liver [28] and pooled tissues (brain, gonads, heart, kidney, liver,
115 spleen and blood of males and females [29]), eastern diamondback rattlesnake (*Crotalus*
116 *adamanteus*) and eastern coral snake (*Micrurus fulvius*) venom glands [7, 8, 30], king cobra
117 (*Ophiophagus hannah*) venom gland, accessory gland and pooled tissues (heart, lung, spleen,
118 brain, testes, gall bladder, pancreas, small intestine, kidney, liver, eye, tongue and stomach)
119 [31], Burmese python (*Python molurus*) pooled liver and heart [32], green anole (*Anolis*
120 *carolinensis*) pooled tissue (liver, tongue, gallbladder, spleen, heart, kidney and lung), testis
121 and ovary [33] and bearded dragon (*Pogona vitticeps*), Nile crocodile (*Crocodylus niloticus*)
122 and chicken (*Gallus gallus*) brains [27], as well as whole genome sequences for the Burmese
123 python and king cobra [31, 34].
124 Assembled transcriptomes were searched for genes previously suggested to be venom toxins
125 in *Echis coloratus* and related species [5, 35, 36] as well as those that have been used to
126 support the Toxicofera hypothesis, namely *acetylcholinesterase, AVIT peptide* [9, 11, 18, 19,
127 23], *complement c3/cobra venom factor, epididymal secretory protein* [19, 37], *c-type lectins*
128 [38, 39], *cysteine-rich secretory protein (crisp)* [40, 41], *crotamine* [42, 43], *cystatin* [44, 45],
129 *dipeptidylpeptidase, lysosomal acid lipase, renin aspartate protease* [5, 19, 35, 46],
130 *hyaluronidase* [47, 48], *kallikrein* [49, 50], *kunitz* [51], *l-amino-acid oxidase* [52, 53], *nerve*

growth factor [54, 55], *phospholipase* A_2 [15], *phospholipase b* [7, 56, 57], *ribonuclease* [58], *serine protease* [59, 60], *snake venom metalloproteinase* [61, 62], *vascular endothelial growth factor* (*vegf*) [9, 17, 63, 64], *veficolin* [65], *vespryn, waprin* [19, 66-68] and *3-finger toxins* [69].

We find that many genes previously claimed to be venom toxins are in fact expressed in multiple tissues (Figure 2) and that transcripts encoding these genes show no evidence of consistently elevated expression level in venom or salivary glands compared to other tissues (Supplemental tables S5-S9). Only two putative venom toxin genes (*l-amino acid oxidase b2* and *PLA*$_2$ *IIA-c*) showed evidence of a venom gland-specific splice variant across our multiple tissue data sets. We have also identified several cases of mistaken identity, where non-orthologous genes have been used to claim conserved, ancestral expression and instances of identical sequences being annotated as two distinct genes (see later sections). We propose that the putative ancestral Toxicoferan venom toxin genes do not encode toxic venom components in the majority of species and that the apparent venom gland-specificity of these genes is a side-effect of incomplete tissue sampling. Our analyses show that neither increased expression in the venom gland nor the production of venom-specific splice variants can be used to support continued claims for the toxicity of these genes.

Results

Based on our quantitative analysis of their expression pattern across multiple species, we identify the following genes as unlikely to represent toxic venom components in the Toxicofera. The identification of these genes as non-venom is more parsimonious than alternative explanations such as the reverse recruitment of a “venom” gene back to a “body” gene [70], which requires a far greater number of steps (duplication, recruitment, selection for increased toxicity, reverse recruitment) to have occurred in each species, whereas a “body” protein remaining a “body” protein is a zero-step process regardless of the number of species

involved. The process of reverse recruitment must also be considered doubtful given the rarity of gene duplication in vertebrates (estimated to be between 1 gene per 100 to 1 gene per 1000 per million years [71-73].

Acetylcholinesterase

We find identical *acetylcholinesterase* (*ache*) transcripts in the *E. coloratus* venom gland and scent gland (which we call transcript 1) and an additional splice variant expressed in skin and scent gland (transcript 2). Whilst the previously known splice variants in banded krait (*Bungarus fasciatus*) are differentiated by the inclusion of an alternative exon, analysis of the *E. coloratus ache* genomic sequence (accession number KF114031) reveals that the shorter transcript 2 instead comprises only the first exon of the *ache* gene, with a TAA stop codon that overlaps the 5' GT dinucleotide splice site in intron 1. *ache* transcript 1 is expressed at a low level in the venom gland (6.60 FPKM) and is found in multiple tissues in all study species (Figure 2), as well as corn snake vomeronasal organ and garter snake liver. The shorter transcript 2 is found most often in skin and scent glands (Figure 2, Supplementary figure S1). The low expression level and diverse tissue distribution of transcripts of this gene suggest that *acetylcholinesterase* does not represent a Toxicoferan venom toxin. It should also be noted that the most frequently cited sources for the generation of a toxic version of *ache* in banded krait via alternative splicing include statements that *ache* "does not appear to contribute to the toxicity of the venom" [74], is "not toxic to mice, even at very high doses" [75] and is "neither toxic by itself nor acting in a synergistic manner with the toxic components of venom" [76].

AVIT

We find only a single transcript encoding an AVIT peptide in our dataset, in the scent gland of the rough green snake (data not shown). The absence of this gene in all of our venom and salivary gland datasets, as well as the venom glands of the king cobra, eastern coral snake and Eastern diamondback rattlesnake and the limited number of sequences available on Genbank

(one species of snake, *Dendroaspis polylepis* (accession number P25687) and two species of lizard, *Varanus varius* and *Varanus komodoensis* (accession numbers AAZ75583 and ABY89668 respectively)) despite extensive sampling, would suggest that it is unlikely to represent a conserved Toxicoferan venom toxin.

Complement C3 ("cobra venom factor")

We find identical transcripts encoding *complement c3* in all tissues in all species, with the exception of royal python skin (Figures 2 and 3) and we find only a single *complement c3* gene in the *E. coloratus* genome (data not shown). These findings, together with the identification of transcripts encoding this gene in the liver, brain, vomeronasal organ and tissue pools of various other reptile species (Figure 3) demonstrate that this gene does not represent a Toxicoferan venom toxin. However, the grouping of additional *complement c3* genes in the king cobra (*Ophiophagus hannah*) and monocled cobra (*Naja kaouthia*) in our phylogenetic tree does support a duplication of this gene somewhere in the Elapid lineage. One of these paralogs may therefore represent a venom toxin in at least some of these more derived species and the slightly elevated expression level of this gene in the venom or salivary gland of some of our study species suggests that *complement c3* has been exapted [77] to become a venom toxin in the Elapids. It seems likely that the identification of the non-toxic paralogue in other species (including veiled chameleon (*Chamaeleo calyptratus*), spiny-tailed lizard (*Uromastyx aegyptia*) and Mitchell's water monitor (*Varanus mitchelli*)) has contributed to confusion about the distribution of this "Cobra venom factor" (which should more rightly be called *complement c3b*), to the point where genes in alligator (*Alligator sinensis*), turtles (*Pelodiscus sinensis*) and birds (*Columba livia*) are now being annotated as venom factors (accession numbers XP 006023407-8, XP 006114685, XP 005513793, Figure 3).

Cystatin

We find two transcripts encoding cystatins expressed in the venom gland of *E. coloratus* corresponding to *cystatin-e/m* and *f* (Supplementary figures S2 and S3). *cystatin-e/m* was found to be expressed in all tissues from all species used in this study (Figure 2), as well as corn snake vomeronasal organ and brain and garter snake liver and pooled tissues. The transcript encoding *cystatin-f* (which has not previously been reported to be expressed in a snake venom gland) is also expressed in the scent gland of *E. coloratus* and in the majority of other tissues of our study species. We find no evidence for a monophyletic clade of Toxicoferan cystatin-derived venom toxins and would agree with Richards et al. [45] that low expression level and absence of *in vitro* toxicity represents a "strong case for snake venom cystatins as essential housekeeping or regulatory proteins, rather than specific prey-targeted toxins…" Indeed, it is unclear why cystatins should be considered to be conserved venom toxins, since even from its earliest discovery in the venom of the puff adder (*Bitis arietans*) there has been "…no evidence that it is connected to the toxicity of the venom" [44].

Dipeptidyl peptidases

We find identical transcripts encoding *dipeptidyl peptidase 3* and *4* in all tissues in all species except the leopard gecko (Figures 2, 4a and 4b), and both of these have a low transcript abundance in the venom gland of *E. coloratus*. *dpp4* is expressed in garter snake liver and Anole testis and ovary and *dpp3* is also expressed in garter snake liver, king cobra pooled tissues and Bearded dragon brain (Figures 4a and b). It is therefore unlikely the either *dpp3* or *dpp4* represent venom toxins.

Epididymal secretory protein

We find one transcript encoding epididymal secretory protein (ESP) expressed in the venom gland of *Echis coloratus* corresponding to type E1. This transcript is also found to be expressed at similar levels in the scent gland and skin of this species and orthologous transcripts are expressed in all three tissues of all other species used in this study (Figure 2 and Supplementary figure S4a), suggesting that this is a ubiquitously expressed gene and not

a venom component. Previously described epididymal secretory protein sequences from varanids [78] and the colubrid *Cylindrophis ruffus* [21] do not represent *esp-e1* and their true orthology is currently unclear. However, our analysis of these and related sequences suggests that they are likely part of a reptile-specific expansion of esp-like genes and that the *Varanus* and *Cylindrophis* sequences do not encode the same gene (Supplementary figure S4b). Therefore there is not, nor was there ever, any evidence that epididymal secretory protein sequences represent venom components in the Toxicofera.

Ficolin ("veficolin")

We find one transcript encoding *ficolin* in the *E. coloratus* venom gland and identical transcripts in both scent gland and skin (Figure 2, Supplementary figure S5) and orthologus transcripts in all corn snake and leopard gecko tissues, as well as rough green snake salivary and scent glands and royal python salivary gland. Paralogous genes expressed in multiple tissues were also found in corn snake and rough green snake (Supplementary figure S5). These findings, together with additional data from available transcriptomes of pooled garter snake body tissues and bearded dragon and chicken brains show that *Ficolin* does not represent a Toxicoferan venom component.

Hyaluronidase

Hyaluronidase has been suggested to be a "venom spreading factor" to aid the dispersion of venom toxins throughout the body of envenomed prey, and as such it does not represent a venom toxin itself [79]. We do however find two hyaluronidase genes expressed in the venom gland of *E. coloratus*. The first appears to be venom gland specific (based on available data) and has two splice variants including a truncated variant similar to sequences previously characterised from *Echis carinatus sochureki* (accession number DQ840262) and *Echis pyramidum leakeyi* (accession number DQ840255) venom glands [48]. Although we cannot rule out hyaluronidase in playing an active (but non-toxic) role in *Echis* venom, it is worth commenting that hyaluronan has been suggested to have a role in wound healing and

the protection of the oral mucosa in human saliva [80]. The expression of hyaluronidases involved in hyaluronan metabolism in venom and/or salivary glands is therefore perhaps unsurprising.

Kallikrein

We find two Kallikrein-like sequences in *E. coloratus*, one of which is expressed in all three tissues in this species (at a low level in the venom gland) and a variety of other tissues in the other study species, and one of which is found only in scent gland and skin (Figure 2, Supplementary figure S6). These genes do not represent venom toxins in *E. coloratus* and appear to be most closely-related to a group of mammalian Kallikrein (KLK) genes containing *KLK1*, *11*, *14* and *15* and probably represent the outgroup to a mammalian-specific expansion of this gene family. The orthology of previously published Toxicoferan Kallikrein genes is currently unclear and the majority of these sequences can be found in our serine protease tree (see later section and Supplementary figure S19).

Kunitz

We find a number of transcripts encoding Kunitz-type protease inhibitors in our tissue data, with the majority of these encoding *kunitz1* and *kunitz2* genes (Figure 2 and Supplementary figure S7). The tissue distribution of these transcripts, together with the phylogenetic position of lizard and venomous snake sequences does not support a monophyletic clade of venom gland-specific Kunitz-type genes in the Toxicofera. The presence of protease inhibitors in reptile venom and salivary glands should perhaps not be too surprising and it again seems likely that the involvement of Kunitz-type inhibitors in venom toxicity in some advanced snake lineages (in this case mamba (*Dendroaspis sp.)* dendrotoxins and krait (*Bungarus multicinctus*) bungarotoxins [81, 82]) has led to confusion when non-toxic orthologs have been identified in other species.

Lysosomal acid lipase

We find two transcripts encoding Lysosomal acid lipase genes in the *E. coloratus* venom gland transcriptome, one of which (*lipa-a*) is also expressed in skin and scent gland in this species and all three tissues in our other study species. *lipa-a*, despite not being venom gland specific, is more highly expressed in the venom gland (3,337.33 FPKM) than in the scent gland (484.49 FPKM) and skin (22.79 FPKM) of *E. coloratus*, although there is no evidence of elevated expression in the salivary glands of our other study species. As this protein is involved in lysosomal lipid hydrolysis [83] and the venom gland is a highly active tissue, we suggest that this elevated expression is likely related to high cell turnover. Transcripts of *lipa-b* are found at a low level in the venom and scent glands of *E. coloratus* and the scent gland of royal python (Figure 2, Supplementary figure S8). Neither *lipa-a* or *lipa-b* therefore encode venom toxins.

Natriuretic peptide

We find only a single natriuretic peptide-like sequence in our dataset, in the skin of the royal python. The absence of this gene from the rest of our study species suggests that it is not a highly conserved Toxicoferan toxin.

Nerve growth factor

We find identical transcripts encoding *nerve growth factor* (*ngf*) in all three *E. coloratus* tissues. Transcripts encoding the orthologous gene are also found in the corn snake salivary gland and scent gland; rough green snake scent gland and skin; royal python skin and leopard gecko salivary gland, scent gland and skin (Figure 2 and Supplementary figure S9). *ngf* is expressed at a higher level in the venom gland (525.82 FPKM) than in the scent gland (0.18 FPKM) and skin (0.58 FPKM) of *E. coloratus*, but not at an elevated level in the salivary gland of other species, again hinting at the potential for exaptation of this gene. Based on these findings, together with the expression of this gene in garter snake pooled tissues, we suggest that *ngf* does not encode a Toxicoferan toxin. However, we do find evidence for the duplication of *ngf* in cobras (Supplementary figure S9) suggesting that it may represent a

311 venom toxin in at least some advanced snakes [84]. As with *complement c3*, it seems likely
312 that the identification of non-toxic orthologs in distantly-related species has led to the
313 conclusion that *ngf* is a widely-distributed venom toxin and confused its true evolutionary
314 history.

315 Phospholipase A_2 (PLA_2 Group IIE)

316 We find transcripts encoding Group IIE PLA_2 genes in the venom gland of *E. coloratus* and
317 the salivary glands of all other species (Figure 2, Supplementary figure S10). Although this
318 gene appears to be venom and salivary-gland-specific (based on available data), its presence
319 in all species (including the non-Toxicoferan leopard gecko) suggests that it does not
320 represent a toxic venom component.

321 Phospholipase B

322 We find a single transcript encoding *phospholipase b* expressed in all three *E. coloratus*
323 tissues (Figures 2 and 5) and transcripts encoding the orthologous gene are found in all other
324 tissues from all study species with the exception of rough green snake salivary gland. We also
325 find *plb* transcripts in corn snake vomeronasal organ, garter snake liver, Burmese python
326 pooled tissues (liver and heart) and bearded dragon brain (Figure 5). The two transcripts in
327 each of rough green snake and corn snake are likely alleles or the result of individual
328 variation, and actually represent a single *phospholipase b* gene from each of these species.
329 Transcript abundance analysis shows this gene to be expressed at a low level in all tissues
330 from all study species. Based on the phylogenetic and tissue distribution of this gene it is
331 unlikely to represent a Toxicoferan venom toxin.

332 Renin ("renin aspartate protease")

333 We find a number of transcripts encoding renin-like genes in the *E. coloratus* venom gland
334 (Figures 2 and 6), one of which (encoding the canonical *renin*) is also expressed in the scent
335 gland and is orthologous to a previously described sequence from the venom gland of the
336 ocellated carpet viper (*Echis ocellatus*, accession number CAJ55260). We also find that the

337 recently-published *Boa constrictor renin aspartate protease* (*rap*) gene (accession number
338 JX467165 [21]) is in fact a *cathepsin d* gene, transcripts of which are found in all three
339 tissues in all five of our study species. We suggest that this misidentification may be due to a
340 reliance on BLAST-based classification, most likely using a database restricted to squamate
341 or serpent sequences. It is highly unlikely that either *renin* or *cathepsin d* (or indeed any
342 renin-like aspartate proteases) constitute venom toxins in *E. coloratus* or *E. ocellatus*, nor do
343 they represent basal Toxicoferan toxins.

344 Ribonuclease

345 Ribonucleases have been suggested to have a role in the generation of free purines in snake
346 venoms [58] and the presence of these genes in the salivary glands of two species of lizard
347 (*Gerrhonotus infernalis* and *Celestus warreni*) and two colubrid snakes (*Liophis peocilogyrus*
348 and *Psammophis mossambicus*) has been used to support the Toxicofera [78, 85]. We did not
349 identify orthologous ribonuclease genes in any of our salivary or venom gland data, nor do
350 we find them in venom gland transcriptomes from the Eastern diamondback rattlesnake, king
351 cobra and eastern coral snake (although we have identified a wide variety of other
352 ribonuclease genes). The absence of these genes in seven Toxicoferans, coupled with the fact
353 that they were initially described from only 2 out of 11 species of snake [85] and 3 out of 18
354 species of lizard [78] would cast doubt on their status as conserved Toxicoferan toxins.

355 Three finger toxins (3ftx)

356 We find 2 transcripts encoding three finger toxin (3ftx)-like genes expressed in the *E.*
357 *coloratus* venom gland, one of which is expressed in all 3 tissues (*3ftx-a*) whilst the other is
358 expressed in the venom and scent glands (*3ftx-b*). Orthologous transcripts of *3ftx-a* are found
359 to be expressed in all three tissues of corn snake, rough green snake salivary gland and skin,
360 and royal python salivary gland. An ortholog of *3ftx-b* is expressed in rough green snake
361 scent gland. We also find a number of different putative *3ftx* genes in our other study species,
362 often expressed in multiple tissues (Figure 2, Supplementary figure S11). Based on the

phylogenetic and tissue distribution of both of these genes we suggest that they do not represent venom toxins in *E. coloratus*. As with other proposed Toxicoferan genes such as *complement c3* and *nerve growth factor*, it seems likely that *3ftx* genes are indeed venom components in some species, especially cobras and other elapids [31, 69], and that the identification of their non-venom orthologs in other species has led to much confusion regarding the phylogenetic distribution of these toxic variants.

Vespryn

We do not find *vespryn* transcripts in any *E. coloratus* tissues, although this gene is present in the genome of this species (accession number KF114032). We do however find transcripts encoding this gene in the salivary and scent glands of the corn snake, and skin and scent glands of the rough green snake, royal python and leopard gecko (Figure 2, Supplementary figure S12). We suggest that the tissue distribution of this gene in these species casts doubt on its role as a venom component in the Toxicofera.

Waprin

We find a number of "waprin"-like genes in our dataset, expressed in a diverse array of body tissues. Our phylogenetic analyses (Supplementary figure S13) show that previously characterised "waprin" genes [8, 66, 68, 86, 87] most likely represent *WAP four-disulfide core domain 2* (*wfdc2*) genes which have undergone a squamate-specific expansion and that there is no evidence for a venom gland-specific paralog. It is unlikely therefore that these genes represent a Toxicoferan venom toxin. Indeed, the inland taipan (*Oxyuranus microlepidotus*) "Omwaprin" has been shown to be "…non-toxic to Swiss albino mice at doses of up to 10 mg/kg when administered intraperitoneally" [68] and is more likely to have an antimicrobial function in the venom or salivary gland.

Implications for venom composition and complexity in *Echis coloratus*

The following genes show either a venom gland-specific expression or an elevated expression level in this tissue, but not both and as such we suggest that whilst they *may* represent venom toxins in *E. coloratus*, further analysis is needed in order to confirm this.

Vascular endothelial growth factor

We find four transcripts encoding vascular endothelial growth factor (VEGF) expressed in the venom gland of *E. coloratus*. These correspond to *vegf-a*, *vegf-b*, *vegf-c* and *vegf-f* and of these *vegf-a*, *b* and *c* are also expressed in the skin and scent gland of this species (Figure 2). Transcripts encoding orthologs of these genes are expressed in all three tissues of all other species used in this study (with the exception of the absence of *vegf-a* in corn snake skin). In accordance with previous studies [7] we find evidence of alternative splicing of *vegf-a* transcripts in all species although no variant appears to be tissue-specific. It is likely that a failure to properly recognise and classify alternatively spliced *vegf-a* transcripts (Aird et al. 2013) may have contributed to an overestimation of snake venom complexity. *vegf-d* was only found to be expressed in royal python salivary gland and scent gland and all three tissues from leopard gecko (Figure 2, Supplementary figure S14). The transcript encoding VEGF-F is found only in the venom gland of *E. coloratus* and, given the absence of any Elapid *vegf-f* sequences in public databases as well as absence of this transcript in the two species of colubrid in our study, we suggest that *vegf-f* is specific to vipers. Whilst *vegf-f* has a higher transcript abundance in *E. coloratus* venom gland (186.73 FPKM) than *vegf-a* (3.24 FPKM), *vegf-b* (1.28 FPKM) and *vegf-c* (1.54 FPKM), compared to other venom genes in this species (see next section) it has a considerably lower transcript abundance suggesting it represents at most a minor venom component in *E. coloratus*.

L-amino acid oxidase

We find transcripts encoding two *l-amino acid oxidase (laao)* genes in *E. coloratus*, one of which (*laao-b*) has two splice variants (Figure 2, Supplementary figure S15). *laao-a* transcripts are found in all three *E. coloratus* and leopard gecko tissues. *laao-b* is venom

gland-specific in *E. coloratus* (based on the available data) and transcripts of the orthologous gene are found in the scent glands of corn snake, rough green snake and royal python. The splice variant *laao-b2* may represent a venom toxin in *E. coloratus* based on its specific expression in the venom gland of this species and elevated expression level (628.84 FPKM).

Crotamine

We find a single *crotamine*-like transcript in *E. coloratus*, in the venom gland (Figure 2). Related genes are found in a variety of tissues in other study species (including the scent gland of the rough green snake, the salivary gland and skin of the leopard gecko, and in all three corn snake tissues), although the short length of these sequences precludes a definitive statement of orthology. This gene may represent a toxic venom component in *E. coloratus* based on its tissue distribution, but due to its low transcript abundance (10.95 FPKM) it is likely to play a minor role, if any.

The following genes are found only in the venom gland of *E. coloratus* and clearly show an elevated expression level (Figure 7). Whilst we classify these genes as encoding venom toxins in this species (Table 1) it should be noted that none of these genes support the monophyly of Toxicoferan venom toxins.

Cysteine-rich secretory proteins (CRISPs)

We find transcripts encoding two distinct CRISPs expressed in the *E. coloratus* venom gland, one of which is also found in skin and scent gland (Figure 2). Phylogenetic analysis of these genes (which we call *crisp-a* and *crisp-b*) reveals that they appear to have been created as a result of a gene duplication event earlier in the evolution of advanced snakes (Supplementary Figure S16). *crisp-a* transcripts are also found in all three corn snake tissues, as well as rough green snake skin and scent gland and royal python scent gland. *crisp-b* is also found in corn snake salivary gland (Figure 2 and Supplementary figure S16) and the phylogenetic and tissue distribution of this gene suggest that it does indeed represent a venom toxin, produced

via duplication of an ancestral *crisp* gene that was expressed in multiple tissues, including the salivary gland. The elevated transcript abundance of *crisp-b* (3,520.07 FPKM) in the venom gland of *E. coloratus* further supports its role as a venom toxin in this species (Figure 7). The phylogenetic and tissue distribution and low transcript abundance of *crisp-a* (0.61 FPKM in *E. coloratus* venom gland) shows that it is unlikely to be a venom toxin. We also find no evidence of a monophyletic clade of reptile venom toxins and therefore suggest that, contrary to earlier reports [20, 78], the CRISP genes of varanid and helodermatid lizards do not represent shared Toxicoferan venom toxins and, if they are indeed toxic venom components, have been recruited independently from those of the advanced snakes. Regardless of their status as venom toxins, it appears likely that the diversity of CRISP genes in varanid lizards in particular [17] has been overestimated as a result of the use of negligible levels of sequence variation to classify transcripts as representing distinct gene products (Supplementary figures S23 and S24).

C-type lectins

We find transcripts encoding 11 distinct C-type lectin genes in the *E. coloratus* venom gland, one of which (*ctl-a*) is also expressed in the scent gland of this species. The remaining 10 genes (*ctl-b* to *k*) are found only in the venom gland and form a clade with other viper C-type lectin genes (Figure 2, Supplementary figure S17). Of these, 6 are highly expressed in the venom gland (*ctl-b* to *d*, *ctl-f* to *g* and *ctl-j*) with a transcript abundance range of 3,706.21-24,122.41 FPKM (Figure 7). The remainder of these genes (*ctl-e*, *ctl-h* to *i* and *ctl-k*) show lower transcript abundance (0.80-1,475.88 FPKM), with two (*ctl-i* and *k*) being more lowly expressed than *ctl-a* (230.06 FPKM). A number of different C-type lectin genes are found in our other study species, often expressed in multiple tissues (Supplementary figure S17). We suggest therefore that the 6 venom-gland specific C-type lectin genes which are highly expressed are indeed venom toxins in *E. coloratus* and that these genes diversified via the duplication of an ancestral gene with a wide expression pattern, including in salivary/venom

glands. Based on their selective expression in the venom gland (from available data) the remaining four C-type lectin genes cannot be ruled out as putative toxins, although their lower transcript abundance suggests that they are likely to be minor components in *E. coloratus* venom. It should also be noted that a recent analysis of king cobra (*Ophiohagus hannah*) venom gland transcriptome and proteome suggested that "…lectins do not contribute to king cobra envenoming" [31].

Phospholipase A_2 (PLA_2 Group IIA)

We find five transcripts encoding Group IIA PLA_2 genes in *E. coloratus*, three of which are found only in the venom gland and two of which are found only in the scent gland (these latter two likely represent intra-individual variation in the same transcript) (Figure 2, Supplementary figure S18). The venom gland-specific transcript *PLA_2 IIA-c* is highly expressed (22,520.41 FPKM) and likely represents a venom toxin, and may also be a putative splice variant although further analysis is needed to confirm this. *PLA_2 IIA-d* and *IIA-e* show an elevated, but lower, expression level (1,677.15 FPKM and 434.67 FPKM respectively, Figure 7). Based on tissue and phylogenetic distribution we would propose that these three genes may represent putative venom toxins (Table 1).

Serine proteases

We find 6 transcripts encoding Serine proteases in *E. coloratus* (Figure 2, Supplementary figure S19) which (based on available data) are all venom gland specific. Four of these transcripts are highly expressed in the venom gland (*serine proteases a-c* and *e*; 3,076.01-7,687.03 FPKM) whilst two are expressed at a lower level (*serine proteases d* and *f*; 1,098.45 FPKM and 102.34 FPKM respectively, Figure 7). Based on these results we suggest *serine proteases a, b, c* and *e* represent venom toxins whilst *serine proteases d* and *f* may represent putative venom toxins (Table 1).

Snake venom metalloproteinases

We find 21 transcripts encoding snake venom metalloproteinases in *E. coloratus* and of these 14 are venom gland-specific, whilst another (*svmp-n*) is expressed in the venom gland and scent gland. Five remaining genes are expressed in the scent gland only whilst another is expressed in the skin (Figure 2, Supplementary figure S20). Of the 14 venom gland-specific SVMPs we find 4 to be highly expressed (5,552.84-15,118.41 FPKM, Figure 7). In the absence of additional data, we classify the 13 venom gland-specific *svmp* genes as venom toxins in this species (Table 1).

Discussion

Our transcriptomic analyses have revealed that all 16 of the basal venom toxin genes used to support the hypothesis of a single, early evolution of venom in reptiles (the Toxicofera hypothesis [17-21, 78]), as well as a number of other genes that have been proposed to encode venom toxins in multiple species are in fact expressed in multiple tissues, with no evidence for consistently higher expression in venom or salivary glands. Additionally, only two genes in our entire dataset of 74 genes in five species were found to encode possible venom gland-specific splice variants (*l-amino acid oxidase b2* and *PLA_2 IIA-c*). We therefore suggest that many of the proposed basal Toxicoferan genes most likely represent housekeeping or maintenance genes and that the identification of these genes as conserved venom toxins is a side-effect of incomplete tissue sampling. This lack of support for the Toxicofera hypothesis therefore prompts a return to the previously held view [88] that venom in different lineages of reptiles has evolved independently, once at the base of the advanced snakes, once in the helodermatid (gila monster and beaded lizard) lineage and, possibly, one other time in monitor lizards, although evidence for a venom system in this latter group [20, 78, 89] may need to be reinvestigated in light of our findings. The process of reverse recruitment [70], where a venom gene undergoes additional gene duplication events and is subsequently recruited from the venom gland back into a body tissue (which was proposed on

the basis of the placement of garter snake and Burmese python "physiological" genes within clades of "venom" genes) must also be re-evaluated in light of our findings.

Bites by venomous snakes are thought to be responsible for as many as 1,841,000 envenomings and 94,000 deaths annually, predominantly in the developing world [90, 91] and medical treatment of snakebite is reliant on the production of antivenoms containing antibodies, typically from sheep or horses, that will bind and neutralise toxic venom proteins [92]. Since these antivenoms are derived from the injection of crude venom into the host animal they are not targeted to the most pathogenic venom components and therefore also include antibodies to weakly- or non-pathogenic proteins, requiring the administration of large or multiple doses [11], increasing the risks of adverse reactions. A comprehensive understanding of snake venom composition is therefore vital for the development of the next generation of antivenoms [2, 11, 93] as it is important that research effort is not spread too thinly through the inclusion of non-toxic venom gland transcripts. Our results suggest that erroneous assumptions about the single origination and functional conservation of venom toxins across the Toxicofera has led to the complexity of snake venom being overestimated by previous authors, with the venom of the painted saw-scaled viper, *Echis coloratus* likely consisting of just 34 genes in 8 gene families (Table 1, based on venom gland-specific expression and a 'high' expression, as defined by presence in the top 25% of transcripts [94] in at least two of four venom gland samples), fewer than has been suggested for this and related species in previous EST or transcriptomic studies [5, 35]. However, it is noteworthy that the results of our analyses accord well with proteomic analyses of venom composition in snakes, which range from an almost identical complement of 35 toxins in 8 gene families for the related ocellated carpet viper, *Echis ocellatus* [36] to between 24-61 toxins in 6-14 families in a range of other species (Table 2). Far from being a "complex cocktail" [10, 11, 95, 96], snake venom may in fact represent a relatively simple mixture of toxic proteins

honed by natural selection for rapid prey immobilisation, with limited lineage-specific expansion in one or a few particular gene families.

In order to avoid continued overestimation of venom complexity, we propose that future transcriptome-based analyses of venom composition must include quantitative comparisons of multiple body tissues from multiple individuals and robust phylogenetic analysis that includes known paralogous members of gene families. We would also encourage the use of clearly explained, justifiable criteria for classifying highly similar sequences as new paralogs rather than alleles or the result of PCR or sequencing errors, as it seems likely that some available sequences from previous studies have been presented as distinct genes on the basis of extremely minor (or even non-existent) sequence variation (see Supplementary figures S21-S24 for examples of identical or nearly identical ribonuclease and CRISP sequences and Supplementary figures S25 and S26 for examples of the same sequence being annotated as two different genes). As a result, the diversity of "venom" composition in these species may have been inadvertently inflated.

Additionally, we would encourage the adoption of a standard nomenclature for reptile genes, as the overly-complicated and confusing nomenclature used currently (Table 3) may also contribute to the perceived complexity of snake venom. We propose that such a nomenclature system should be based on the comprehensive standards developed for Anole lizards [97], for example:

- "Gene symbols for all…species should be written in lower case only and in italics, e.g., *gene2*."
- "Whenever criteria for orthology have been met… the gene symbol should be comparable to the human gene symbol, e.g., if the human gene symbol is *GENE2*, then the gene symbol would be *gene2*."

- "Duplication of the ortholog of a mammalian gene will be indicated by an "a" or "b" suffix, e.g., *gene2a* and *gene2b*. If the mammalian gene symbol already contains a suffix letter, then there would be a second letter added, e.g., *gene4aa* and *gene4ab*."

It seems likely that the application of our approach to other species (together with proteomic studies of extracted venom) will lead to a commensurate reduction in claimed venom diversity, with clear implications for the development of next generation antivenoms: since most true venom genes are members of a relatively small number of gene families, it is likely that a similarly small number of antibodies may be able to bind to and neutralise the toxic venom components, especially with the application of "string of beads" techniques [98] utilising fusions of short oligopeptide epitopes designed to maximise the cross-reactivity of the resulting antibodies [2].

Conclusions

We suggest that identification of the apparently conserved Toxicofera venom toxins in previous studies is most likely a side effect of incomplete tissue sampling, compounded by incorrect interpretation of phylogenetic trees and the use of BLAST-based gene identification methods. It should perhaps not be too surprising that homologous tissues in related species would show similar gene complements and the restriction of most previous studies to only the "venom" glands means that monophyletic clades of reptile sequences in phylogenetic trees have been taken to represent monophyletic clades of venom toxin genes. Whilst it is true that some of these genes do encode toxic proteins in some species (indeed, this was often the basis for their initial discovery) the discovery of orthologous genes in other species does not necessarily demonstrate shared toxicity. In short, toxicity in one does not equal toxicity in all.

Methods

590 Experimental methods involving animals followed institutional and national guidelines and
591 were approved by the Bangor Universty Ethical Review Committee.

592 *RNA-Seq*

593 Total RNA was extracted from four venom glands taken from four individual specimens of
594 adult Saw-scaled vipers (*Echis coloratus*) at different time points following venom extraction
595 in order to capture the full diversity of venom genes (16, 24 and 48 hours post-milking).
596 Additionally, total RNA from two scent glands and two skin samples of this species and the
597 salivary, scent glands and skin of two adult corn snakes (*Pantherophis guttatus*), rough green
598 snakes (*Opheodrys aestivus*), royal pythons (*Python regius*) and leopard geckos (*Eublepharis*
599 *macularius*) was also extracted using the RNeasy mini kit (Qiagen) with on-column DNase
600 digestion. Only a single corn snake skin sample provided RNA of high enough quality for
601 sequencing. mRNA was prepared for sequencing using the TruSeq RNA sample preparation
602 kit (Illumina) with a selected fragment size of 200-500bp and sequenced using 100bp paired-
603 end reads on the Illumina HiSeq2000 or HiSeq2500 platform.

604 *Quality control, assembly and analysis*

605 The quality of all raw sequence data was assessed using FastQC [99] and reads for each
606 tissue and species were pooled and assembled using Trinity [100] (sequence and assembly
607 metrics are provided in Supplemental tables S1-S3). Putative venom toxin amino acid
608 sequences were aligned using ClustalW [101] and maximum likelihood trees constructed
609 using the Jones-Taylor-Thornton (JTT) model with 500 Bootstrap replicates. Transcript
610 abundance estimation was carried out using RSEM [102] as a downstream analysis of Trinity
611 (version trinityrnaseq_r2012-04-27). Sets of reads were mapped to species-specific reference
612 transcriptome assemblies (Supplementary table S4) to allow comparison between tissues on a
613 per-species basis and all results values shown are in FPKM (Fragments Per Kilobase of exon
614 per Million fragments mapped). Individual and mean FPKM values for each gene per tissue
615 per species are given in Supplementary tables S5-S9. All transcript abundance values given

within the text are based on the average transcript abundance per tissue per species to account for variation between individual samples.

Transcriptome reads were deposited in the European Nucleotide Archive (ENA) database under accession #ERP001222 and GenBank under accession numbers XXX and XXX and genes used to reconstruct phylogenies are deposited in GenBank under accession numbers XXXX-XXXX

Competing interests

The authors declare that they have no competing interests

Authors' contributions

JFM, ADH and DWL designed the experiments and all authors carried out them out. JFM and ADH prepared the manuscript and this was seen, modified and approved by all authors.

Acknowledgements

The authors wish to thank A. Tweedale, R. Morgan, G. Cooke, A. Barlow, C. Wüster and M. Hegarty for technical assistance and N. Casewell, W. Wüster and A. Malhotra for useful discussions. We are grateful to the staff of High Performance Computing (HPC) Wales for enabling and supporting our access to their systems. This research was supported by a Royal Society Research Grant awarded to JFM (grant number RG100514) and Wellcome trust funding to DWL (grant number 098051). JFM and MTS are supported by the Biosciences, Environment and Agriculture Alliance (BEAA) between Bangor University and Aberystwyth University and ADH is funded by a Bangor University 125th Anniversary Studentship.

641 **References**

642

643 1. Li M, Fry BG, Kini RM: **Putting the brakes on snake venom evolution: the unique**
644 **molecular evolutionary patterns of *Aipysurus eydouxii* (Marbled sea snake)**
645 **phospholipase A_2 toxins.** *Mol Biol Evol* 2005, **22:**934-941.

646

647 2. Wagstaff SC, Laing GD, Theakston RDG, Papaspyridis C, Harrison RA: **Bioinformatics**
648 **and multiepitope DNA immunization to design rational snake antivenom.** *PLoS medicine*
649 2006, **3:**e184.

650

651 3. Kini RM, Doley R: **Structure, function and evolution of three-finger toxins: mini**
652 **proteins with multiple targets.** *Toxicon* 2010, **56:**855-867.

653

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

658

659 5. Casewell N, Harrison R, Wüster W, Wagstaff S: **Comparative venom gland**
660 **transcriptome surveys of the saw-scaled vipers (Viperidae: *Echis*) reveal substantial**
661 **intra-family gene diversity and novel venom transcripts.** *BMC Genomics* 2009, **10:**564.

662

663 6. Siang AS, Doley R, Vonk FJ, Kini RM: **Transcriptomic analysis of the venom gland of**
664 **the red-headed krait (*Bungarus flaviceps*) using expressed sequence tags.** *BMC Mol Biol*
665 2010, **11:**24.

666

667 7. Rokyta DR, Wray KP, Lemmon AR, Lemmon EM, Caudle SB: **A high-throughput**
668 **venom-gland transcriptome for the Eastern Diamondback Rattlesnake (*Crotalus***
669 ***adamanteus*) and evidence for pervasive positive selection across toxin classes.** *Toxicon*
670 2011, **57:**657-671.
671
672 8. Rokyta DR, Lemmon AR, Margres MJ, Aronow K: **The venom-gland transcriptome of**
673 **the eastern diamondback rattlesnake (*Crotalus adamanteus*).** *BMC Genomics* 2012,
674 **13:**312.
675
676 9. Fry BG: **From genome to "venome": molecular origin and evolution of the snake**
677 **venom proteome inferred from phylogenetic analysis of toxin sequences and related**
678 **body proteins.** *Genome Res* 2005, **15:**403-420.
679
680 10. Wong ES, Belov K: **Venom evolution through gene duplications.** *Gene* 2012, **496:**1-7.
681
682 11. Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG: **Complex cocktails: the**
683 **evolutionary novelty of venoms.** *Trends in ecology & evolution* 2013, **28:**219-229.
684
685 12. Otto SP, Whitton J: **Polyploid incidence and evolution.** *Annu Rev Genet* 2000, **34:**401-
686 437.
687
688 13. Lynch M, Force A: **The probability of duplicate gene preservation by**
689 **subfunctionalization.** *Genetics* 2000, **154:**459-473.
690
691 14. Walsh B: **Population-genetic models of the fates of duplicate genes.** *Genetica* 2003,
692 **118:**279-294.

15. Lynch VJ: **Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A_2 genes.** *BMC evolutionary biology* 2007, **7:**2.

16. Teshima KM, Innan H: **Neofunctionalization of duplicated genes under the pressure of gene conversion.** *Genetics* 2008, **178:**1385-1398.

17. Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SR, Kuruppu S, Fung K, Hedges SB, Richardson MK: **Early evolution of the venom system in lizards and snakes.** *Nature* 2006, **439:**584-588.

18. Fry BG, Vidal N, Van der Weerd L, Kochva E, Renjifo C: **Evolution and diversification of the Toxicofera reptile venom system.** *Journal of proteomics* 2009, **72:**127-136.

19. Fry BG, Casewell NR, Wüster W, Vidal N, Young B, Jackson TN: **The structural and functional diversification of the Toxicofera reptile venom system.** *Toxicon* 2012, **60:**434-448.

20. Fry BG, Wroe S, Teeuwisse W, van Osch MJ, Moreno K, Ingle J, McHenry C, Ferrara T, Clausen P, Scheib H: **A central role for venom in predation by *Varanus komodoensis* (Komodo Dragon) and the extinct giant *Varanus* (*Megalania*) *priscus*.** *Proceedings of the National Academy of Sciences* 2009, **106:**8969-8974.

21. Fry BG, Undheim EA, Ali SA, Debono J, Scheib H, Ruder T, Jackson TN, Morgenstern D, Cadwallader L, Whitehead D: **Squeezers and leaf-cutters: differential diversification**

718 **and degeneration of the venom system in toxicoferan reptiles.** *Molecular & Cellular*
719 *Proteomics* 2013, **12:**1881-1899.
720
721 22. Butte AJ, Dzau VJ, Glueck SB: **Further defining housekeeping, or" maintenance,"**
722 **genes: Focus on" A compendium of gene expression in normal human tissues".**
723 *Physiological genomics* 2002, **7:**95-96.
724
725 23. Vonk FJ, Jackson K, Doley R, Madaras F, Mirtschin PJ, Vidal N: **Snake venom: From**
726 **fieldwork to the clinic.** *Bioessays* 2011, **33:**269-279.
727
728 24. Strydom D: **Snake venom toxins: The evolution of some of the toxins found in snake**
729 **venoms.** *Syst Biol* 1973, **22:**596-608.
730
731 25. Kochva E: **The origin of snakes and evolution of the venom apparatus.** *Toxicon* 1987,
732 **25:**65-106.
733
734 26. Brykczynska U, Tzika AC, Rodriguez I, Milinkovitch MC: **Contrasted evolution of the**
735 **vomeronasal receptor repertoires in mammals and squamate reptiles.** *Genome Biol Evol*
736 2013, **5:**389-401.
737 27. Tzika AC, Helaers R, Schramm G, Milinkovitch MC: **Reptilian-transcriptome v1. 0, a**
738 **glimpse in the brain transcriptome of five divergent Sauropsida lineages and the**
739 **phylogenetic position of turtles.** *EvoDevo* 2011, **2:**1-18.
740
741 28. Schwartz TS, Bronikowski AM: **Dissecting molecular stress networks: identifying**
742 **nodes of divergence between life-history phenotypes.** *Mol Ecol* 2013, **22:**739-756.
743

744 29. Schwartz TS, Tae H, Yang Y, Mockaitis K, Van Hemert JL, Proulx SR, Choi J,
745 Bronikowski AM: **A garter snake transcriptome: pyrosequencing, de novo assembly, and**
746 **sex-specific differences.** *BMC Genomics* 2010, **11:**694.
747
748 30. Margres MJ, Aronow K, Loyacano J, Rokyta DR: **The venom-gland transcriptome of**
749 **the eastern coral snake (*Micrurus fulvius*) reveals high venom complexity in the**
750 **intragenomic evolution of venoms.** *BMC Genomics* 2013, **14:**1-18.
751
752 31. Vonk FJ, Casewell NR, Henkel CV, Heimberg AM, Jansen HJ, McCleary RJ, Kerkkamp
753 HM, Vos RA, Guerreiro I, Calvete JJ, Wüster W, Woods AE, Logan JM, Harrison RA,
754 Castoe TA, de Koning AP, Pollock DD, Yandell M, Calderon D, Renjifo C, Currier RB,
755 Salgado D, Pla D, Sanz L, Hyder AS, Ribeiro JM, Arntzen JW, van den Thillart GE, Boetzer
756 M, Pirovano W, *et al.*: **The king cobra genome reveals dynamic gene evolution and**
757 **adaptation in the snake venom system.** *Proc Natl Acad Sci USA* 2013, **110:**20651-20656.
758
759 32. Castoe TA, Fox SE, Jason de Koning A, Poole AW, Daza JM, Smith EN, Mockler TC,
760 Secor SM, Pollock DD: **A multi-organ transcriptome resource for the Burmese Python**
761 **(*Python molurus bivittatus*).** *BMC Res Notes* 2011, **4**:310.
762
763 33. Eckalbar WL, Hutchins ED, Markov GJ, Allen AN, Corneveaux JJ, Lindblad-Toh K, Di
764 Palma F, Alfoldi J, Huentelman MJ, Kusumi K: **Genome reannotation of the lizard *Anolis***
765 ***carolinensis* based on 14 adult and embryonic deep transcriptomes.** *BMC Genomics*
766 2013, **14**:49.
767
768 34. Castoe TA, de Koning AP, Hall KT, Card DC, Schield DR, Fujita MK, Ruggiero RP,
769 Degner JF, Daza JM, Gu W, Reyes-Velasco J, Shaney KJ, Castoe JM, Fox SE, Poole AW,

770 Polanco D, Dobry J, Vandewege MW, Li Q, Schott RK, Kapusta A, Minx P, Feschotte C,
771 Uetz P, Ray DA, Hoffmann FG, Bogden R, Smith EN, Chang BS, Vonk FJ, *et al.*: **The**
772 **Burmese python genome reveals the molecular basis for extreme adaptation in snakes.**
773 *Proc Natl Acad Sci USA* 2013, **110:**20645-20650.
774
775 35. Wagstaff SC, Harrison RA: **Venom gland EST analysis of the saw-scaled viper, *Echis***
776 ***ocellatus*, reveals novel α_9 β_1 integrin-binding motifs in venom metalloproteinases and a**
777 **new group of putative toxins, renin-like aspartic proteases.** *Gene* 2006, **377:**21-32.
778
779 36. Wagstaff SC, Sanz L, Juárez P, Harrison RA, Calvete JJ: **Combined snake venomics**
780 **and venom gland transcriptomic analysis of the ocellated carpet viper, *Echis ocellatus*.**
781 *Journal of proteomics* 2009, **71:**609-623.
782
783 37. Alper C, Balavitch D: **Cobra venom factor: evidence for its being altered cobra C3**
784 **(the third component of complement).** *Science* 1976, **191:**1275-1276.
785 38. Morita T: **Structures and functions of snake venom CLPs (C-type lectin-like**
786 **proteins) with anticoagulant-, procoagulant-, and platelet-modulating activities.** *Toxicon*
787 2005, **45:**1099-1114.
788
789 39. Ogawa T, Chijiwa T, Oda-Ueda N, Ohno M: **Molecular diversity and accelerated**
790 **evolution of C-type lectin-like proteins from snake venom.** *Toxicon* 2005, **45:**1-14.
791
792 40. Yamazaki Y, Hyodo F, Morita T: **Wide distribution of cysteine-rich secretory proteins**
793 **in snake venoms: isolation and cloning of novel snake venom cysteine-rich secretory**
794 **proteins.** *Arch Biochem Biophys* 2003, **412:**133-141.
795

41. Yamazaki Y, Morita T: **Structure and function of snake venom cysteine-rich secretory proteins.** *Toxicon* 2004, **44:**227-231.

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

43. Oguiura N, Boni-Mitake M, Rádis-Baptista G: **New view on crotamine, a small basic polypeptide myotoxin from South American rattlesnake venom.** *Toxicon* 2005, **46:**363-370.

44. Ritonja A, Evans HJ, Machleidt W, Barrett AJ: **Amino acid sequence of a cystatin from venom of the African puff adder (*Bitis arietans*).** *Biochem J* 1987, **246:**799-802.

45. Richards R, St Pierre L, Trabi M, Johnson LA, de Jersey J, Masci PP, Lavin MF: **Cloning and characterisation of novel cystatins from elapid snake venom glands.** *Biochimie* 2011, **93:**659-668.

46. Aird SD: **Snake venom dipeptidyl peptidase IV: taxonomic distribution and quantitative variation.** *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 2008, **150:**222-228.

47. Tu AT, Hendon RR: **Characterization of lizard venom hyaluronidase and evidence for its action as a spreading factor.** *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* 1983, **76:**377-383.

48. Harrison RA, Ibison F, Wilbraham D, Wagstaff SC: **Identification of cDNAs encoding viper venom hyaluronidases: cross-generic sequence conservation of full-length and unusually short variant transcripts.** *Gene* 2007, **392:**22-33.

49. Komori Y, Nikai T, Sugihara H: **Biochemical and physiological studies on a kallikrein-like enzyme from the venom of *Crotalus viridis viridis* (Prairie rattlesnake).** *Biochimica et Biophysica Acta (BBA)-General Subjects* 1988, **967:**92-102.

50. Komori Y, Nikai T: **Chemistry and Biochemistry of Kallikrein-Like Enzyn from Snake Venoms.** *Toxin Reviews* 1998, **17:**261-277.

51. Župunski V, Kordiš D, Gubenšek F: **Adaptive evolution in the snake venom Kunitz/BPTI protein family.** *FEBS Lett* 2003, **547:**131-136.

52. Suhr S, Kim D: **Identification of the snake venom substance that induces apoptosis.** *Biochem Biophys Res Commun* 1996, **224:**134-139.

53. Du X, Clemetson KJ: **Snake venom L-amino acid oxidases.** *Toxicon* 2002, **40:**659-665.

54. Angeletti RH: **Nerve growth factor from cobra venom.** *Proc Natl Acad Sci USA* 1970, **65:**668-674.

55. Kostiza T, Meier J: **Nerve growth factors from snake venoms: chemical properties, mode of action and biological significance.** *Toxicon* 1996, **34:**787-806.

847 56. Bernheimer A, Linder R, Weinstein S, Kim K: **Isolation and characterization of a**
848 **phospholipase B from venom of collett's snake, *Pseudechis colletti*.** *Toxicon* 1987, **25:**547-
849 554.

850

851 57. Chatrath ST, Chapeaurouge A, Lin Q, Lim TK, Dunstan N, Mirtschin P, Kumar PP, Kini
852 RM: **Identification of novel proteins from the venom of a cryptic snake *Drysdalia***
853 ***coronoides* by a combined transcriptomics and proteomics approach.** *Journal of*
854 *Proteome Research* 2011, **10:**739-750.

855

856 58. Aird SD: **Taxonomic distribution and quantitative analysis of free purine and**
857 **pyrimidine nucleosides in snake venoms.** *Comparative Biochemistry and Physiology Part*
858 *B: Biochemistry and Molecular Biology* 2005, **140:**109-126.

859

860 59. Pirkle H: **Thrombin-like enzymes from snake venoms: an updated inventory.** *Thromb*
861 *Haemost* 1998, **79:**675-683.

862

863 60. Serrano SM, Maroun RC: **Snake venom serine proteinases: sequence homology vs.**
864 **substrate specificity, a paradox to be solved.** *Toxicon* 2005, **45:**1115-1132.

865

866 61. Bjarnason JB, Fox JW: **Hemorrhagic metalloproteinases from snake venoms.**
867 *Pharmacol Ther* 1994, **62:**325-372.

868

869 62. Jia L, Shimokawa K, Bjarnason JB, Fox JW: **Snake venom metalloproteinases:**
870 **structure, function and relationship to the ADAMs family of proteins.** *Toxicon* 1996,
871 **34:**1269-1276.

872

873 63. Junqueira de Azevedo ILM, Farsky SHP, Oliveira MLS, Ho PL: **Molecular Cloning and**
874 **Expression of a Functional Snake Venom Vascular Endothelium Growth Factor**
875 **(VEGF) from the *Bothrops insularis* Pit Viper. a new member of the VEGF family of**
876 **proteins.** *J Biol Chem* 2001, **276:**39836-39842.
877
878 64. Yamazaki Y, Takani K, Atoda H, Morita T: **Snake venom vascular endothelial growth**
879 **factors (VEGFs) exhibit potent activity through their specific recognition of KDR**
880 **(VEGF receptor 2).** *J Biol Chem* 2003, **278:**51985-51988.
881
882 65. OmPraba G, Chapeaurouge A, Doley R, Devi KR, Padmanaban P, Venkatraman C,
883 Velmurugan D, Lin Q, Kini RM: **Identification of a novel family of snake venom proteins**
884 **veficolins from *Cerberus rynchops* using a venom gland transcriptomics and proteomics**
885 **approach.** *Journal of proteome research* 2010, **9:**1882-1893.
886
887 66. Torres AM, Wong HY, Desai M, Moochhala S, Kuchel PW, Kini RM: **Identification of**
888 **a Novel Family of Proteins in Snake Venoms. Purification and structural**
889 **characterization of nawaprin from *Naja nigricollis* snake venom.** *J Biol Chem* 2003,
890 **278:**40097-40104.
891
892 67. Pung YF, Kumar SV, Rajagopalan N, Fry BG, Kumar PP, Kini RM: **Ohanin, a novel**
893 **protein from king cobra venom: its cDNA and genomic organization.** *Gene* 2006,
894 **371:**246-256.
895
896 68. Nair D, Fry B, Alewood P, Kumar P, Kini R: **Antimicrobial activity of omwaprin, a**
897 **new member of the waprin family of snake venom proteins.** *Biochem J* 2007, **402**:93-104.
898

899 69. Fry BG, Wüster W, Kini RM, Brusic V, Khan A, Venkataraman D, Rooney A:
900 **Molecular evolution and phylogeny of elapid snake venom three-finger toxins.** *J Mol*
901 *Evol* 2003, **57:**110-129.
902
903 70. Casewell NR, Huttley GA, Wüster W: **Dynamic evolution of venom proteins in**
904 **squamate reptiles.** *Nature communications* 2012, **3**:1066.
905
906 71. Cotton JA, Page RD: **Rates and patterns of gene duplication and loss in the human**
907 **genome.** *Proc Biol Sci* 2005, **272:**277-283.
908
909 72. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes.**
910 *Science* 2000, **290:**1151-1155.
911
912 73. Lynch M, Conery JS: **The evolutionary demography of duplicate genes.** *Journal of*
913 *structural and functional genomics* 2003, **3:**35-44.
914
915 74. Cousin X, Bon S, Massoulie J, Bon C: **Identification of a novel type of alternatively**
916 **spliced exon from the acetylcholinesterase gene of *Bungarus fasciatus*. Molecular forms**
917 **of acetylcholinesterase in the snake liver and muscle.** *J Biol Chem* 1998, **273:**9812-9820.
918
919 75. Cousin X, Créminon C, Grassi J, Méflah K, Cornu G, Saliou B, Bon S, Massoulié J, Bon
920 C: **Acetylcholinesterase from *Bungarus* venom: a monomeric species.** *FEBS Lett* 1996,
921 **387:**196-200.
922
923 76. Cousin X, Bon S, Duval N, Massoulie J, Bon C: **Cloning and expression of**
924 **acetylcholinesterase from *Bungarus fasciatus* venom. A new type of cooh-terminal**

925 **domain; involvement of a positively charged residue in the peripheral site.** *J Biol Chem*
926 1996, **271:**15099-15108.
927
928 77. Gould SJ, Vrba ES: **Exaptation-a missing term in the science of form.** *Paleobiology*
929 1982, **8:**4-15.
930
931 78. Fry BG, Winter K, Norman JA, Roelants K, Nabuurs RJ, van Osch MJ, Teeuwisse WM,
932 van der Weerd L, McNaughtan JE, Kwok HF, Scheib H, Greisman L, Kochva E, Miller LJ,
933 Gao F, Karas J, Scanlon D, Lin F, Kuruppu S, Shaw C, Wong L, Hodgson WC: **Functional**
934 **and structural diversification of the Anguimorpha lizard venom system.** *Mol Cell*
935 *Proteomics* 2010, **9:**2369-2390.
936
937 79. Kemparaju K, Girish K: **Snake venom hyaluronidase: a therapeutic target.** *Cell*
938 *Biochem Funct* 2006, **24:**7-12.
939
940 80. Pogrel MA, Low MA, Stern R: **Hyaluronan (hyaluronic acid) and its regulation in**
941 **human saliva by hyaluronidase and its inhibitors.** *J Oral Sci* 2003, **45:**85-92.
942
943 81. Harvey AL: **Twenty years of dendrotoxins.** *Toxicon* 2001, **39:**15-26.
944
945 82. Kwong PD, McDonald NQ, Sigler PB, Hendrickson WA: **Structure of β2-**
946 **bungarotoxin: potassium channel binding by Kunitz modules and targeted**
947 **phospholipase action.** *Structure* 1995, **3:**1109-1119.
948

949 83. Warner TG, Dambach LM, Shin JH, O'Brien JS: **Purification of the lysosomal acid**
950 **lipase from human liver and its role in lysosomal lipid hydrolysis.** *J Biol Chem* 1981,
951 **256:**2952-2957.
952
953 84. Sunagar K, Fry BG, Jackson TN, Casewell NR, Undheim EA, Vidal N, Ali SA, King GF,
954 Vasudevan K, Vasconcelos V: **Molecular Evolution of Vertebrate Neurotrophins: Co-**
955 **Option of the Highly Conserved Nerve Growth Factor Gene into the Advanced Snake**
956 **Venom Arsenal.** *PloS one* 2013, **8:**e81827.
957
958 85. Fry BG, Scheib H, Junqueira de Azevedo, ILM, Silva DA, Casewell NR: **Novel**
959 **transcripts in the maxillary venom glands of advanced snakes.** *Toxicon* 2012, **59:**696-
960 708.
961
962 86. Fry BG, Scheib H, van der Weerd L, Young B, McNaughtan J, Ramjan SF, Vidal N,
963 Poelmann RE, Norman JA: **Evolution of an arsenal: structural and functional**
964 **diversification of the venom system in the advanced snakes (Caenophidia).** *Mol Cell*
965 *Proteomics* 2008, **7:**215-246.
966
967 87. Aird SD, Watanabe Y, Villar-Briones A, Roy MC, Terada K, Mikheyev AS:
968 **Quantitative high-throughput profiling of snake venom gland transcriptomes and**
969 **proteomes (*Ovophis okinavensis* and *Protobothrops flavoviridis*).** *BMC Genomics* 2013,
970 **14:**790.
971
972 88. Kardong K, Weinstein S, Smith T: **Reptile venom glands: form, function, and future.**
973 In *Handbook of venoms and toxins of reptiles*. Edited by Mackessy SP. Boca Raton, Florida:
974 CRC Press; 2009:65-91.

89. Vikrant S, Verma BS: **Monitor lizard bite-induced acute kidney injury–a case report.** *Renal failure* 2013, **36:**444-446.

90. Kasturiratne A, Wickremasinghe AR, de Silva N, Gunawardena NK, Pathmeswaran A, Premaratna R, Savioli L, Lalloo DG, de Silva HJ: **The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths.** *PLoS medicine* 2008, **5:**e218.

91. Harrison RA, Hargreaves A, Wagstaff SC, Faragher B, Lalloo DG: **Snake envenoming: a disease of poverty.** *PLoS neglected tropical diseases* 2009, **3:**e569.

92. Chippaux J, Goyffon M: **Venoms, antivenoms and immunotherapy.** *Toxicon* 1998, **36:**823-846.

93. Harrison R: **Development of venom toxin-specific antibodies by DNA immunisation: rationale and strategies to improve therapy of viper envenoming.** *Vaccine* 2004, **22:**1648-1655.

94. Williford A, Demuth JP: **Gene expression levels are correlated with synonymous codon usage, amino acid composition, and gene architecture in the red flour beetle, *Tribolium castaneum*.** *Mol Biol Evol* 2012, **29:**3755-3766.

95. Izidoro LFM, Ribeiro MC, Souza GR, Sant'Ana CD, Hamaguchi A, Homsi-Brandeburgo MI, Goulart LR, Beleboni RO, Nomizo A, Sampaio SV: **Biochemical and functional**

characterization of an l-amino acid oxidase isolated from *Bothrops pirajai* snake venom. *Bioorg Med Chem* 2006, **14:**7034-7043.

96. Calvete JJ, Juárez P, Sanz L: **Snake venomics. Strategy and applications.** *Journal of mass spectrometry* 2007, **42:**1405-1414.

97. Kusumi K, Kulathinal RJ, Abzhanov A, Boissinot S, Crawford NG, Faircloth BC, Glenn TC, Janes DE, Losos JB, Menke DB, Poe S, Sanger TJ, Schneider CJ, Stapley J, Wade J, Wilson-Rawls J: **Developing a community-based genetic nomenclature for anole lizards.** *BMC Genomics* 2011, **12:**554.

98. Whitton JL, Sheng N, Oldstone MB, McKee TA: **A "string-of-beads" vaccine, comprising linked minigenes, confers protection from lethal-dose virus challenge.** *J Virol* 1993, **67:**348-352.

99. Andrews S: **FastQC: A quality control tool for high throughput sequence data.** 2010, Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

100. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q: **Full-length transcriptome assembly from RNA-Seq data without a reference genome.** *Nat Biotechnol* 2011, **29:**644-652.

101. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23:**2947-2948.

102. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *BMC Bioinformatics* 2011, **12:**323.

103. Calvete JJ, Escolano J, Sanz L: **Snake venomics of *Bitis* species reveals large intragenus venom toxin composition variation: application to taxonomy of congeneric taxa.** *Journal of proteome research* 2007, **6:**2732-2745.

104. Calvete JJ, Marcinkiewicz C, Sanz L: **Snake venomics of *Bitis gabonica gabonica*. Protein family composition, subunit organization of venom toxins, and characterization of dimeric disintegrins bitisgabonin-1 and bitisgabonin-2.** *Journal of proteome research* 2006, **6:**326-336.

105. Lomonte B, Escolano J, Fernández J, Sanz L, Angulo Y, Gutiérrez JM, Calvete JJ: **Snake venomics and antivenomics of the arboreal neotropical pitvipers *Bothriechis lateralis* and *Bothriechis schlegelii*.** *Journal of proteome research* 2008, **7:**2445-2457.

106. Fahmi L, Makran B, Pla D, Sanz L, Oukkache N, Lkhider M, Harrison RA, Ghalim N, Calvete JJ: **Venomics and antivenomics profiles of North African *Cerastes cerastes* and *C. vipera* populations reveals a potentially important therapeutic weakness.** *Journal of proteomics* 2012, **75:**2442-2453.

107. Calvete JJ, Fasoli E, Sanz L, Boschetti E, Righetti PG: **Exploring the venom proteome of the western diamondback rattlesnake, *Crotalus atrox*, via snake venomics and combinatorial peptide ligand library approaches.** *Journal of proteome research* 2009, **8:**3055-3067.

108. Sanz L, Escolano J, Ferretti M, Biscoglio MJ, Rivera E, Crescenti EJ, Angulo Y, Lomonte B, Gutiérrez JM, Calvete JJ: **Snake venomics of the South and Central American Bushmasters. Comparison of the toxin composition of *Lachesis muta* gathered from proteomic versus transcriptomic analysis.** *Journal of proteomics* 2008, **71:**46-60.

109. Kulkeaw K, Chaicumpa W, Sakolvaree Y, Tongtawe P, Tapchaisri P: **Proteome and immunome of the venom of the Thai cobra, *Naja kaouthia*.** *Toxicon* 2007, **49:**1026-1041.

110. Georgieva D, Risch M, Kardas A, Buck F, von Bergen M, Betzel C: **Comparative analysis of the venom proteomes of *Vipera ammodytes ammodytes* and *Vipera ammodytes meridionalis*.** *Journal of proteome research* 2008, **7:**866-886.

Tables

Table 1. Predicted venom composition of the painted saw-scaled viper, *Echis coloratus*

Gene family	Number of genes
SVMP	13
C-type lectin	8
Serine protease	6
PLA2	3
CRISP	1
L-amino acid oxidase	1
VEGF	1
Crotamine	1
Total 8	34

Table 2. Predicted numbers of venom toxins and venom toxin families from proteomic studies of snake venom accord well with our transcriptome results.

Species	Number of toxins	Number of toxin families
Bitis caudalis [103]	30	8
Bitis gabonica gabonica [104]	35	12
Bitis gabonica rhinoceros [103]	33	11
Bitis nasicornis [103]	28	9
Bothriechis schlegelii [105]	?	7
Cerastes cerastes [106]	25-30	6
Crotalus atrox [107]	~24	~9
Echis ocellatus [36]	35	8
Lachesis muta [108]	24-26	8
Naja kaouthia [109]	61	12
Ophiophagus hannah [31]	?	14
Vipera ammodytes [110]	38	9

Table 3. Venom gene nomenclature. Lack of a formal set of nomenclatural rules for venom toxins has led to an explosion of different gene names and may have contributed to the overestimation of reptile venom diversity.

Gene/gene family	Alternative name and accession number
3 Finger toxin (3Ftx)	Denmotoxin [Q06ZW0] Candoxin [AY142323]
CRISP	Piscivorin [AAO62994] Catrin [AAO62995] Ablomin [AAM45664] Tigrin [Q8JGT9] Kaouthin [ACH73167, ACH73168] Natrin-1 [Q7T1K6] CRVP [Q8UW25, Q8UW11] Pseudechetoxin [Q8AVA4] Pseudechin [Q8AVA3] Serotriflin [P0CB15] Latisemin [Q8JI38] Ophanin [AAO62996] Opharin [ACN93671] Bc-CRP [ACE73577, ACE73578]
Ficolin	Veficolin [ADK46899] Ryncolin [D8VNS7-9, D8VNT0]
Serine proteases	Acubin [CAB46431] Gyroxin [B0FXM3] Ussurase [AAL48222] Serpentokallikrein [AAG27254] Salmobin [AAC61838] Batroxobin [AAA48553] Nikobin [CBW30778] Gloshedobin [POC5B4] Gussurobin [Q8UVX1] Pallabin [CAA04612] Pallase [AAC34898]
Snake venom metalloproteinase (SVMP)	Stejnihagin-B [ABA40759] Bothropasin [AAC61986] Atrase B [ADG02948] Mocarhagin 1 [AAM51550] Scutatease-1 [ABQ01138] Austrelease-1 [ABQ01134]
Vascular endothelial growth factor (VEGF)	Barietin [ACN22038] Cratrin [ACN22040] Apiscin [ACN22039] Vammin [ACN22045]
Vespryn	Ohanin [AAR07992] Thaicobrin [P82885]
Waprin	Nawaprin [P60589] Porwaprin [B5L5N2] Stewaprin [B5G6H3] Veswaprin [B5L5P5] Notewaprin [B5G6H5] Carwaprin [B5L5P0]

Figure legends

Figure 1. Relationships of key vertebrate lineages and the placement of species discussed in this paper. A monophyletic clade of reptiles (which includes birds) is shaded green and the Toxicofera [21] are shaded red. Modified taxon names have been used for simplicity.

Figure 2. Tissue distribution of proposed venom toxin transcripts. The majority of transcripts proposed to encode Toxicoferan venom proteins are expressed in multiple body tissues. Transcripts found in the assembled transcriptomes but which are assigned transcript abundance of <1 FPKM are shaded orange. Eco, painted saw-scaled viper (*Echis coloratus*); Pgu, corn snake (*Pantherophis guttatus*); Oae, rough green snake (*Opheodrys aestivus*); Pre, royal python (*Python regius*); Ema, leopard gecko (*Eublepharis macularius*). VG, venom gland; SAL, salivary gland; SCG, scent gland; SK, skin.

Figure 3. Maximum likelihood tree of *complement c3* ("*cobra venom factor*") sequences. Whilst most sequences likely represent housekeeping or maintenance genes, a gene duplication event in the elapid lineage (marked with *) may have produced a venom-specific paralog. An additional duplication (marked with +) may have taken place in *Austrelaps superbus*, although both paralogs appear to be expressed in both liver and venom gland. Geographic separation in king cobras (*Ophiophagus hannah*) from Indonesia and China is reflected in observed sequence variation. Numbers above branches are Bootstrap values for 500 replicates. Tissue distribution of transcripts is indicated using the following abbreviations: VG, venom gland; SK, skin; SCG, scent gland, AG, accessory gland; VMNO, vomeronasal organ and those genes found to be expressed in one or more body tissues are shaded blue.

Figure 4. Maximum likelihood tree of *dipeptidylpeptidase 3* (*dpp3*) and *dipeptidylpeptidase 4* (*dpp4*) sequences. Transcripts encoding *dpp3* and *dpp4* are found in a wide variety of body tissues, and likely represent housekeeping genes. Numbers above branches are Bootstrap values for 500 replicates. Tissue distribution of transcripts is indicated using the following abbreviations: VG, venom gland; SK, skin; SCG, scent gland, AG, accessory gland; VMNO, vomeronasal organ and those genes found to be expressed in one or more body tissues are shaded blue.

Figure 5. Maximum likelihood tree of *phospholipase b* (*plb*) sequences. Transcripts encoding *plb* are found in a wide variety of body tissues, and likely represent housekeeping genes. Numbers above branches are Bootstrap values for 500 replicates. Tissue distribution of transcripts is indicated using the following abbreviations: VG, venom gland; SK, skin; SCG, scent gland, AG, accessory gland; VMNO, vomeronasal organ and those genes found to be expressed in one or more body tissues are shaded blue.

Figure 6. Maximum likelihood tree of renin-like sequences. Renin-like genes are expressed in a diversity of body tissues. The recently published *Boa constrictor* "RAP-Boa-1" sequence is clearly a *cathepsin d* gene and is therefore not orthologous to the *Echis ocellatus* renin sequence as has been claimed [21]. Numbers above branches are Bootstrap values for 500 replicates. Tissue distribution of transcripts is indicated using the following abbreviations: VG, venom gland; SK, skin; SCG, scent gland and those genes found to be expressed in one or more body tissues are shaded blue.

Figure 7. Graph of transcript abundance values of proposed venom transcripts in the *Echis coloratus* venom gland. The majority of Toxicoferan transcripts are expressed at

1173 extremely low level, with the most highly expressed genes falling into only four gene
1174 families (C-type lectins, Group IIA phospholipase A_2, serine proteases and snake venom
1175 metalloproteinases). FPKM = Fragments Per Kilobase of exon per Million fragments
1176 mapped.