



PRIFYSGOL
BANGOR
UNIVERSITY

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

Hargreaves, A.D.; Swain, M.T.; Logan, D.W.; Mulley, J.F.

Toxicon

DOI:
[10.1016/j.toxicon.2014.10.004](https://doi.org/10.1016/j.toxicon.2014.10.004)

Published: 17/10/2014

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
Hargreaves, A. D., Swain, M. T., Logan, D. W., & Mulley, J. F. (2014). Testing the Toxicofera: comparative reptile transcriptomics casts doubt on the single, early evolution of the reptile venom system. *Toxicon*, 92, 140-156. <https://doi.org/10.1016/j.toxicon.2014.10.004>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1 **Title page**

2

3 **Title:**

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

6

7 **Authors:**

8 Adam D Hargreaves¹, Martin T Swain², Darren W Logan³ and John F Mulley^{1*}

9

10 **Affiliations:**

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

13

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

16

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

18

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

20

21

22 Author email addresses:

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

24 Adam Hargreaves – adamdhargreaves@gmail.com

25 Darren Logan - dl5@sanger.ac.uk

26 Martin Swain - mts11@aber.ac.uk

27 **Abstract**

28 The identification of apparently conserved gene complements in the venom and salivary
29 glands of a diverse set of reptiles led to the development of the Toxicofera hypothesis – the
30 single, early evolution of the venom system in reptiles. However, this hypothesis is based
31 largely on relatively small scale EST-based studies of only venom or salivary glands and
32 toxic effects have been assigned to only some putative Toxicoferan toxins in some species.
33 We set out to examine the distribution of these proposed venom toxin transcripts in order to
34 investigate to what extent conservation of gene complements may reflect a bias in previous
35 sampling efforts. Our quantitative transcriptomic analyses of venom and salivary glands and
36 other body tissues in five species of reptile, together with the use of available RNA-Seq
37 datasets for additional species, shows that the majority of genes used to support the
38 establishment and expansion of the Toxicofera are in fact expressed in multiple body tissues
39 and most likely represent general maintenance or “housekeeping” genes. The apparent
40 conservation of gene complements across the Toxicofera therefore reflects an artefact of
41 incomplete tissue sampling. We therefore conclude that venom has evolved multiple times in
42 reptiles.

43

44 **Keywords**

45 Snake venom, Toxicofera, Transcriptomics

46

47 **1. Introduction**

48 Snake venom is frequently cited as being highly complex or diverse (Li et al., 2005; Wagstaff
49 et al., 2006; Kini and Doley, 2010) and a large number of venom toxin genes and gene
50 families have been identified, predominantly from EST-based studies of gene expression
51 during the re-synthesis of venom in the venom glands following manually-induced emptying
52 (“milking”) of extracted venom (Pahari et al., 2007; Casewell et al., 2009; Siang et al., 2010;

53 Rokyta et al., 2011; Rokyta et al., 2012). It has been suggested that many of these gene
54 families originated via the duplication of a gene encoding a non-venom protein expressed
55 elsewhere in the body, followed by recruitment into the venom gland where natural selection
56 could act to increase toxicity. Subsequent additional duplications would then lead to a
57 diversification within gene families, often in a species-specific manner (Fry, 2005; Wong and
58 Belov, 2012; Casewell et al., 2013). However, since whole genome duplication is a rare event
59 in reptiles (Otto and Whitton, 2000), the hypothesis that novelty in venom originates via the
60 duplication of a “body” gene with subsequent recruitment into the venom gland requires that
61 gene duplication is a frequent event in the germline of venomous snakes. An additional
62 prerequisite is that the promoter and enhancer sequences that regulate venom gland-specific
63 expression are relatively simple and easy to evolve. It also suggests a high incidence of
64 neofunctionalisation rather than the more common process of subfunctionalisation (Lynch
65 and Force, 2000; Walsh, 2003; Lynch, 2007; Teshima and Innan, 2008). However, it has
66 recently been shown that in fact snake venom toxins are likely derived from pre-existing
67 salivary proteins that have been restricted to the venom gland rather than body proteins that
68 have been recruited (Hargreaves et al. 2014a).

69 The apparent widespread distribution of genes known to encode venom toxins in snakes in
70 the salivary glands of a diverse set of reptiles, (including both those that had previously been
71 suggested to have secondarily lost venom in favour of constriction or other predation
72 techniques, and those that had previously been considered to have never been venomous), led
73 to the development of the Toxicofera hypothesis – the single, early evolution of venom in
74 reptiles (Vidal and Hedges, 2005; Fry et al., 2006; Fry et al., 2009a; Fry et al., 2012a) (Figure
75 1). Analysis of a wide range of reptiles, including charismatic megafauna such as the
76 Komodo dragon, *Varanus komodoensis* (Fry et al., 2009b), has shown that the ancestral
77 Toxicoferan venom system comprises at least 16 genes, with additional gene families

78 subsequently recruited in different lineages (Fry et al., 2009a; Fry et al., 2012a; Fry et al.,
79 2013).

80 Although toxic effects have been putatively assigned to some Toxicoferan venom proteins in
81 certain species, the problem remains that their identification as venom components is based
82 largely on their expression in the venom gland during venom synthesis and their apparent
83 relatedness to other, known toxins in phylogenetic trees. It has long been known that all
84 tissues express a basic set of “housekeeping” or maintenance genes (Butte et al., 2002) and it
85 is therefore not surprising that similar genes might be found to be expressed in similar tissues
86 in different species of reptiles, and that these genes might group together in phylogenetic
87 trees. However, the identification of transcripts encoding putative venom toxins in other body
88 tissues would cast doubt on the classification of these Toxicoferan toxins as venom
89 components, as it is unlikely that the same gene could fulfil toxic and non-toxic (pleiotropic)
90 roles without evidence for alternative splicing to produce a toxic variant (as has been
91 suggested for *acetylcholinesterase* in the banded krait, *Bungarus fasciatus* (Vonk et al., 2011;
92 Casewell et al., 2013)) or increased expression levels in the venom gland (where toxicity
93 might be dosage dependent). In order to address some of these issues and to test the
94 robustness of the Toxicofera hypothesis, we have carried out a comparative transcriptomic
95 survey of the venom or salivary glands, skin and cloacal scent glands of five species of
96 reptile. Unlike the pancreas and other parts of the digestive system (Strydom, 1973; Kochva,
97 1987), these latter tissues (which include a secretory glandular tissue (the scent gland) and a
98 relatively inert, non-secretory tissue (skin)) have not previously been suggested to be the
99 source of duplicated venom toxin genes and we would therefore only expect to find
100 ubiquitous maintenance or “housekeeping” genes to be commonly expressed across these
101 tissues. We use the general term ‘salivary gland’ for simplicity, to encompass the oral glands
102 of the leopard gecko and rectal glands and Duvernoy’s gland of the royal python, corn snake
103 and rough green snake and do not imply any homology to mammalian salivary glands.

104 Our study species included the venomous painted saw-scaled viper (*Echis coloratus*); the
105 non-venomous corn snake (*Pantherophis guttatus*) and rough green snake (*Opheodrys*
106 *aestivus*) and a member of one of the more basal extant snake lineages, the royal python
107 (*Python regius*). As members of the Toxicofera *sensu* Fry et al. (Fry et al., 2013) we would
108 expect to find the basic Toxicoferan venom genes expressed in the venom or salivary glands
109 of all of these species. In addition, we generated corresponding data for the leopard gecko
110 (*Eublepharis macularius*), a member of one of the most basal lineages of squamate reptiles
111 that lies outside of the proposed Toxicofera clade (Figure 1). We have also taken advantage
112 of available transcriptomes or RNA-Seq data for corn snake vomeronasal organ
113 (Brykczynska et al., 2013) and brain (Tzika et al., 2011), garter snake (*Thamnophis elegans*)
114 liver (Schwartz and Bronikowski, 2013) and pooled tissues (brain, gonads, heart, kidney,
115 liver, spleen and blood of males and females (Schwartz et al., 2010)), Eastern diamondback
116 rattlesnake (*Crotalus adamanteus*) and eastern coral snake (*Micrurus fulvius*) venom glands
117 (Rokyta et al., 2011; Rokyta et al., 2012; Margres et al., 2013), king cobra (*Ophiophagus*
118 *hannah*) venom gland, accessory gland and pooled tissues (heart, lung, spleen, brain, testes,
119 gall bladder, pancreas, small intestine, kidney, liver, eye, tongue and stomach) (Vonk et al.,
120 2013), Burmese python (*Python molurus*) pooled liver and heart (Castoe et al., 2011), green
121 anole (*Anolis carolinensis*) pooled tissue (liver, tongue, gallbladder, spleen, heart, kidney and
122 lung), testis and ovary (Eckalbar et al., 2013) and bearded dragon (*Pogona vitticeps*), Nile
123 crocodile (*Crocodylus niloticus*) and chicken (*Gallus gallus*) brains (Tzika et al., 2011), as
124 well as whole genome sequences for the Burmese python and king cobra (Castoe et al., 2013;
125 Vonk et al., 2013).

126 Assembled transcriptomes were searched for genes previously suggested to be venom toxins
127 in *Echis coloratus* and related species (Wagstaff and Harrison, 2006; Casewell et al., 2009;
128 Wagstaff et al., 2009) as well as those that have been used to support the Toxicofera
129 hypothesis, namely *acetylcholinesterase*, *AVIT peptide* (Fry, 2005; Fry et al., 2009a; Vonk et

130 al., 2011; Fry et al., 2012a; Casewell et al., 2013), *complement c3/cobra venom factor*,
131 *epididymal secretory protein* (Alper and Balavitch, 1976; Fry et al., 2012a), *c-type lectins*
132 (Morita, 2005; Ogawa et al., 2005), *cysteine-rich secretory protein (crisp)* (Yamazaki et al.,
133 2003a; Yamazaki and Morita, 2004), *crotamine* (Rádis-Baptista et al., 2003; Oguiura et al.,
134 2005), *cystatin* (Ritonja et al., 1987; Richards et al., 2011), *dipeptidylpeptidase*, *lysosomal*
135 *acid lipase*, *renin aspartate protease* (Wagstaff and Harrison, 2006; Aird, 2008; Casewell et
136 al., 2009; Fry et al., 2012a), *hyaluronidase* (Tu and Hendon, 1983; Harrison et al., 2007),
137 *kallikrein* (Komori et al., 1988; Komori and Nikai, 1998), *kunitz* (Župunski et al., 2003), *l-*
138 *amino-acid oxidase* (Suhr and Kim, 1996; Du and Clemetson, 2002), *nerve growth factor*
139 (Angeletti, 1970; Kostiza and Meier, 1996), *phospholipase A₂* (Lynch, 2007), *phospholipase*
140 *b* (Bernheimer et al., 1987; Chatrath et al., 2011; Rokyta et al., 2011), *ribonuclease* (Aird,
141 2005), *serine protease* (Pirkle, 1998; Serrano and Maroun, 2005), *snake venom*
142 *metalloproteinase* (Bjarnason and Fox, 1994; Jia et al., 1996), *vascular endothelial growth*
143 *factor (vegf)* (Junqueira de Azevedo et al., 2001; Yamazaki et al., 2003b; Fry, 2005; Fry et
144 al., 2006), *veficolin* (OmPraba et al., 2010), *vespryn*, *waprin* (Torres et al., 2003; Pung et al.,
145 2006; Nair et al., 2007; Fry et al., 2012a) and *3-finger toxins* (Fry et al., 2003). Transcript
146 abundance estimation values were also calculated to allow the identification of any potential
147 occurrences of pleiotropy (a gene fulfilling a toxic and non-toxic role simultaneously) based
148 upon an elevated expression level in the venom or salivary gland compared to other body
149 tissues. All Transcript abundance values are given in FPKM (Fragments Per Kilobase of exon
150 per Million fragments mapped) and are mean values to account for variation between
151 individual samples (further details given in the methods section).

152 We find that many genes previously claimed to be venom toxins are in fact expressed in
153 multiple tissues (Figure 2) and that transcripts encoding these genes show no evidence of
154 consistently elevated expression level in venom or salivary glands compared to other tissues
155 (Supplemental tables S5-S9). Only two putative venom toxin genes (*l-amino acid oxidase b2*

156 and *PLA₂ IIA-c*) showed evidence of a venom gland-specific splice variant across our
157 multiple tissue data sets. We have also identified several cases of mistaken identity, where
158 non-orthologous genes have been used to claim conserved, ancestral expression and instances
159 of identical sequences being annotated as two distinct genes (see later sections). We propose
160 that the putative ancestral Toxicoferan venom toxin genes do not encode toxic venom
161 components in the majority of species and that the apparent venom gland-specificity of these
162 genes is a side-effect of incomplete tissue sampling. Our analyses show that neither increased
163 expression in the venom gland nor the production of venom-specific splice variants can be
164 used to support continued claims for the toxicity of these genes.

165

166 **2. Methods**

167 Experimental methods involving animals followed institutional and national guidelines and
168 were approved by the Bangor University Ethical Review Committee.

169 **2.1 RNA-Seq**

170 Total RNA was extracted from four venom glands taken from four individual specimens of
171 adult Saw-scaled vipers (*Echis coloratus*) at different time points following venom extraction
172 in order to capture the full diversity of venom genes (16, 24 and 48 hours post-milking).

173 Additionally, total RNA from two scent glands and two skin samples of this species and the
174 salivary, scent glands and skin of two adult corn snakes (*Pantherophis guttatus*), rough green
175 snakes (*Ophedryx aestivus*), royal pythons (*Python regius*) and leopard geckos (*Eublepharis*
176 *macularius*) was also extracted using the RNeasy mini kit (Qiagen) with on-column DNase
177 digestion. Only a single corn snake skin sample provided RNA of high enough quality for
178 sequencing. mRNA was prepared for sequencing using the TruSeq RNA sample preparation
179 kit (Illumina) with a selected fragment size of 200-500bp and sequenced using 100bp paired-
180 end reads on the Illumina HiSeq2000 or HiSeq2500 platform.

181 **2.2 Quality control, assembly and analysis**

182 The quality of all raw sequence data was assessed using FastQC (Andrews, 2010) and reads
183 for each tissue and species were pooled and assembled using Trinity (Grabherr et al., 2011)
184 (sequence and assembly metrics are provided in Supplemental tables S1-S3). Putative venom
185 toxin amino acid sequences were aligned using ClustalW (Larkin et al., 2007) and maximum
186 likelihood trees constructed using the Jones-Taylor-Thornton (JTT) model with 500
187 Bootstrap replicates. Transcript abundance estimation was carried out using RSEM (Li and
188 Dewey, 2011) as a downstream analysis of Trinity (version trinityrnaseq_r2012-04-27). Sets
189 of reads were mapped to species-specific reference transcriptome assemblies (Supplementary
190 table S4) to allow comparison between tissues on a per-species basis and all results values
191 shown are in FPKM (Fragments Per Kilobase of exon per Million fragments mapped).
192 Individual and mean FPKM values for each gene per tissue per species are given in
193 Supplementary tables S5-S9. All transcript abundance values given within the text are based
194 on the average transcript abundance per tissue per species to account for variation between
195 individual samples.
196 Transcriptome reads were deposited in the European Nucleotide Archive (ENA) database
197 under accession #ERP001222 and GenBank under the run accession numbers SRR1287707
198 and SRR1287715. Genes used to reconstruct phylogenies are deposited in GenBank under the
199 BioProject accession number PRJNA255316.

200

201

202 **3. Results**

203 **3.1 Genes unlikely to represent toxic components of the Toxicofera**

204 Based on our quantitative analysis of their expression pattern across multiple species, we
205 identify the following genes as unlikely to represent toxic venom components in the
206 Toxicofera clade (Vidal and Hedges, 2005). The identification of these genes as non-venom
207 is more parsimonious than alternative explanations such as the reverse recruitment of a

208 “venom” gene back to a “body” gene (Casewell et al., 2012), which requires a far greater
209 number of steps (duplication, recruitment, selection for increased toxicity, reverse
210 recruitment) to have occurred in each species, whereas a “body” protein remaining a “body”
211 protein is a zero-step process regardless of the number of species involved. The process of
212 reverse recruitment must also be considered doubtful given the rarity of gene duplication in
213 vertebrates (estimated to be between 1 gene per 100 to 1 gene per 1000 per million years
214 (Cotton and Page, 2005; Lynch and Conery, 2000; Lynch and Conery, 2003)).

215 3.1.1 Acetylcholinesterase

216 We find identical *acetylcholinesterase* (*ache*) transcripts in the *E. coloratus* venom gland and
217 scent gland (which we call transcript 1) and an additional splice variant expressed in skin and
218 scent gland (transcript 2). Whilst the previously known splice variants in banded krait
219 (*Bungarus fasciatus*) are differentiated by the inclusion of an alternative exon, analysis of the
220 *E. coloratus ache* genomic sequence (accession number KF114031) reveals that the shorter
221 transcript 2 instead comprises only the first exon of the *ache* gene, with a TAA stop codon
222 that overlaps the 5' GT dinucleotide splice site in intron 1. *ache* transcript 1 is expressed at a
223 low level in the venom gland (6.60 FPKM) and is found in multiple tissues in all study
224 species (Figure 2), as well as corn snake vomeronasal organ and garter snake liver. The
225 shorter transcript 2 is found most often in skin and scent glands (Figure 2, Supplementary
226 figure S1). The low expression level and diverse tissue distribution of transcripts of this gene
227 suggest that *acetylcholinesterase* does not represent a Toxicoferan venom toxin. Whilst some
228 ACHE activity has been recorded in the oral secretions of a number of colubrid snakes
229 (Mackessy, 2002), experiments with these secretions shows that several hours are needed to
230 achieve complete neuromuscular blockage. It should also be noted that the most frequently
231 cited sources for the generation of a toxic version of *ache* in banded krait via alternative
232 splicing include statements that *ache* “does not appear to contribute to the toxicity of the
233 venom” (Cousin et al., 1998), is “not toxic to mice, even at very high doses” (Cousin et al.,

234 1996a) and is “neither toxic by itself nor acting in a synergistic manner with the toxic
235 components of venom” (Cousin et al., 1996b).

236 3.1.2 AVIT

237 We find only a single transcript encoding an AVIT peptide in our dataset, in the scent gland
238 of the rough green snake (data not shown). The absence of this gene in all of our venom and
239 salivary gland datasets, as well as the venom glands of the king cobra, eastern coral snake and
240 Eastern diamondback rattlesnake, and the limited number of sequences available on Genbank
241 (one species of snake, *Dendroaspis polylepis* (accession number P25687) and two species of
242 lizard, *Varanus varius* and *Varanus komodoensis* (accession numbers AAZ75583 and
243 ABY89668 respectively)) despite extensive sampling, would suggest that it is unlikely to
244 represent a conserved Toxicoferan venom toxin.

245 3.1.3 Complement C3 (“cobra venom factor”)

246 We find identical transcripts encoding *complement c3* in all tissues in all species, with the
247 exception of royal python skin (Figures 2 and 3) and we find only a single *complement c3*
248 gene in the *E. coloratus* genome (data not shown). These findings, together with the
249 identification of transcripts encoding this gene in the liver, brain, vomeronasal organ and
250 tissue pools of various other reptile species (Figure 3) demonstrate that this gene does not
251 represent a Toxicoferan venom toxin. However, the grouping of additional *complement c3*
252 genes in the king cobra (*Ophiophagus hannah*) and monocled cobra (*Naja kaouthia*) in our
253 phylogenetic tree does support a duplication of this gene somewhere in the Elapid lineage.
254 One of these paralogs may therefore represent a venom toxin in at least some of these more
255 derived species and the slightly elevated expression level of this gene in the venom or
256 salivary gland of some of our study species suggests that *complement c3* has been exapted
257 (Gould and Vrba, 1982) to become a venom toxin in the Elapids. It seems likely that the
258 identification of the non-toxic paralog in other species (including veiled chameleon
259 (*Chamaeleo calypttratus*), spiny-tailed lizard (*Uromastyx aegyptia*) and Mitchell's water

260 monitor (*Varanus mitchelli*) has contributed to confusion about the distribution of this
261 “Cobra venom factor” (which should more rightly be called *complement c3b*), to the point
262 where genes in alligator (*Alligator sinensis*), turtles (*Pelodiscus sinensis*) and birds (*Columba*
263 *livia*) are now being annotated as venom factors (accession numbers XP 006023407-8, XP
264 006114685, XP 005513793, Figure 3).

265 3.1.4 Cystatin

266 We find two transcripts encoding cystatins expressed in the venom gland of *E. coloratus*
267 corresponding to *cystatin-e/m* and *f* (Supplementary figures S2 and S3). *cystatin-e/m* was
268 found to be expressed in all tissues from all species used in this study (Figure 2), as well as
269 corn snake vomeronasal organ and brain and garter snake liver and pooled tissues. The
270 transcript encoding *cystatin-f* (which has not previously been reported to be expressed in a
271 snake venom gland) is also expressed in the scent gland of *E. coloratus* and in the majority of
272 other tissues of our study species. We find no evidence for a monophyletic clade of
273 Toxicoforan cystatin-derived venom toxins and would agree with Richards et al. (Richards et
274 al., 2011) that low expression level and absence of *in vitro* toxicity represents a “strong case
275 for snake venom cystatins as essential housekeeping or regulatory proteins, rather than
276 specific prey-targeted toxins...” Indeed, it is unclear why cystatins should be considered to be
277 conserved venom toxins, since even the original discovery of cystatin in the venom of the
278 puff adder (*Bitis arietans*) states that there is “...no evidence that it is connected to the
279 toxicity of the venom” (Ritonja et al., 1987).

280 3.1.5 Dipeptidyl peptidases

281 We find identical transcripts encoding *dipeptidyl peptidase 3* and *4* in all tissues in all species
282 except the leopard gecko (Figures 2, 4a and 4b), and both of these have a low transcript
283 abundance in the venom gland of *E. coloratus*. *dpp4* is expressed in garter snake liver and
284 Anole testis and ovary and *dpp3* is also expressed in garter snake liver, king cobra pooled

285 tissues and Bearded dragon brain (Figures 4a and b). It is therefore unlikely the either *dpp3* or
286 *dpp4* represent venom toxins.

287 3.1.6 Epididymal secretory protein

288 We find one transcript encoding epididymal secretory protein (ESP) expressed in the venom
289 gland of *Echis coloratus* (9.09 FPKM) corresponding to type E1. This transcript is also found
290 to be expressed at similar levels in the scent gland (13.71 FPKM) and skin (8.64 FPKM) of
291 this species and orthologous transcripts are expressed in all three tissues of all other species
292 used in this study (Figure 2 and Supplementary figure S4a), suggesting that this is a
293 ubiquitously expressed gene and not a venom component. Previously described epididymal
294 secretory protein sequences from varanids (Fry et al., 2010) and the colubrid *Cylindrophis*
295 *ruffus* (Fry et al., 2013) do not represent *esp-e1* and their true orthology is currently unclear.
296 However, our analysis of these and related sequences suggests that they are likely part of a
297 reptile-specific expansion of esp-like genes and that the *Varanus* and *Cylindrophis* sequences
298 do not encode the same gene (Supplementary figure S4b). Therefore there is not, nor was
299 there ever, any evidence that epididymal secretory protein sequences represent venom
300 components in the Toxicofera.

301 3.1.7 Ficolin (“veficolin”)

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

310 3.1.8 Hyaluronidase

311 Hyaluronidase has been suggested to be a “venom spreading factor” to aid the dispersion of
312 venom toxins throughout the body of envenomed prey, and as such it does not represent a
313 venom toxin itself (Kemparaju and Girish, 2006). We do however find two hyaluronidase
314 genes expressed in the venom gland of *E. coloratus*. The first appears to be venom gland
315 specific (based on available data) and has two splice variants including a truncated variant
316 similar to sequences previously characterised from *Echis carinatus sochureki* (accession
317 number DQ840262) and *Echis pyramidum leakeyi* (accession number DQ840255) venom
318 glands (Harrison et al., 2007). Although we cannot rule out hyaluronidase playing an active
319 (but non-toxic) role in *Echis* venom, it is worth commenting that hyaluronan has been
320 suggested to have a role in wound healing and the protection of the oral mucosa in human
321 saliva (Pogrel et al., 2003). The expression of hyaluronidases involved in hyaluronan
322 metabolism in venom and/or salivary glands is therefore perhaps unsurprising.

323 3.1.9 Kallikrein

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

333 3.1.10 Kunitz

334 We find a number of transcripts encoding Kunitz-type protease inhibitors in our tissue data,
335 with the majority of these encoding *kunitz1* and *kunitz2* genes (Figure 2 and Supplementary
336 figure S7). The tissue distribution of these transcripts, together with the phylogenetic position

337 of lizard and venomous snake sequences does not support a monophyletic clade of venom
338 gland-specific Kunitz-type genes in the Toxicofera. The presence of protease inhibitors in
339 reptile venom and salivary glands should perhaps not be too surprising and it again seems
340 likely that the involvement of Kunitz-type inhibitors in venom toxicity in some advanced
341 snake lineages (in this case mamba (*Dendroaspis spp.*) dendrotoxins and krait (*Bungarus*
342 *multicinctus*) bungarotoxins (Harvey, 2001; Kwong et al., 1995)) has led to confusion when
343 non-toxic orthologs have been identified in other species.

344 3.1.11 Lysosomal acid lipase

345 We find two transcripts encoding Lysosomal acid lipase genes in the *E. coloratus* venom
346 gland transcriptome, one of which (*lipa-a*) is also expressed in skin and scent gland in this
347 species and all three tissues in our other study species. *lipa-a*, despite not being venom gland
348 specific, is more highly expressed in the venom gland (3,337.33 FPKM) than in the scent
349 gland (484.49 FPKM) and skin (22.79 FPKM) of *E. coloratus*, although there is no evidence
350 of elevated expression in the salivary glands of our other study species. As this protein is
351 involved in lysosomal lipid hydrolysis (Warner et al., 1981) and the venom gland is a highly
352 active tissue, we suggest that this elevated expression is likely related to high cell turnover.
353 Transcripts of *lipa-b* are found at a low level in the venom and scent glands of *E. coloratus*
354 and the scent gland of royal python (Figure 2, Supplementary figure S8). Neither *lipa-a* or
355 *lipa-b* therefore encode venom toxins.

356 3.1.12 Natriuretic peptide

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

360 3.1.13 Nerve growth factor

361 We find identical transcripts encoding *nerve growth factor (ngf)* in all three *E. coloratus*
362 tissues. Transcripts encoding the orthologous gene are also found in the corn snake salivary

363 gland and scent gland; rough green snake scent gland and skin; royal python skin and leopard
364 gecko salivary gland, scent gland and skin (Figure 2 and Supplementary figure S9). *ngf* is
365 expressed at a higher level in the venom gland (525.82 FPKM) than in the scent gland (0.18
366 FPKM) and skin (0.58 FPKM) of *E. coloratus*, but not at an elevated level in the salivary
367 gland of other species, again hinting at the potential for exaptation of this gene. Based on
368 these findings, together with the expression of this gene in garter snake pooled tissues, we
369 suggest that *ngf* does not encode a Toxicoferan toxin. However, we do find evidence for the
370 duplication of *ngf* in cobras (Supplementary figure S9), suggesting that it may represent a
371 venom toxin in at least some advanced snakes (Sunagar et al., 2013). As with *complement c3*,
372 it seems likely that the identification of non-toxic orthologs in distantly-related species has
373 led to the conclusion that *ngf* is a widely-distributed venom toxin and confused its true
374 evolutionary history.

375 3.1.14 Phospholipase A₂ (PLA₂ Group IIE)

376 We find transcripts encoding Group IIE PLA₂ genes in the venom gland of *E. coloratus* and
377 the salivary glands of all other species (Figure 2, Supplementary figure S10). Although this
378 gene appears to be venom and salivary-gland-specific (based on available data), its presence
379 in all species (including the non-Toxicoferan leopard gecko) suggests that it does not
380 represent a toxic venom component.

381 3.1.15 Phospholipase B

382 We find a single transcript encoding *phospholipase b* expressed in all three *E. coloratus*
383 tissues (Figures 2 and 5) and transcripts encoding the orthologous gene are found in all other
384 tissues from all study species, with the exception of rough green snake salivary gland. We
385 also find *plb* transcripts in corn snake vomeronasal organ, garter snake liver, Burmese python
386 pooled tissues (liver and heart) and bearded dragon brain (Figure 5). The two transcripts in
387 the rough green snake and corn snake are likely alleles or the result of individual variation,
388 and actually represent a single *phospholipase b* gene from each of these species. Transcript

389 abundance analysis shows this gene to be expressed at a low level in all tissues from all study
390 species. Based on the phylogenetic and tissue distribution of this gene it is unlikely to
391 represent a Toxicoferan venom toxin.

392 3.1.16 Renin (“renin aspartate protease”)

393 We find a number of transcripts encoding renin-like genes in the *E. coloratus* venom gland
394 (Figures 2 and 6), one of which (encoding the canonical *renin*) is also expressed in the scent
395 gland and is orthologous to a previously described sequence from the venom gland of the
396 ocellated carpet viper (*Echis ocellatus*, accession number CAJ55260). We also find that the
397 recently-published *Boa constrictor renin aspartate protease (rap)* gene (accession number
398 JX467165 (Fry et al., 2013)) is in fact a *cathepsin d* gene, transcripts of which are found in all
399 three tissues in all five of our study species. We suggest that this misidentification may be
400 due to a reliance on BLAST-based classification, most likely using a database restricted to
401 squamate or serpent sequences. It is highly unlikely that either *renin* or *cathepsin d* (or indeed
402 any renin-like aspartate proteases) constitute venom toxins in *E. coloratus* or *E. ocellatus*, nor
403 do they represent basal Toxicoferan toxins.

404 3.1.17 Ribonuclease

405 Ribonucleases have been suggested to have a role in the generation of free purines in snake
406 venoms (Aird, 2005) and the presence of these genes in the salivary glands of two species of
407 lizard (*Gerrhonotus infernalis* and *Celestus warreni*) and two colubrid snakes (*Liophis*
408 *peocilogyrus* and *Psammophis mossambicus*) has been used to support the Toxicofera (Fry et
409 al., 2010; Fry et al., 2012b). We did not identify orthologous ribonuclease genes in any of our
410 salivary or venom gland data, nor do we find them in venom gland transcriptomes from the
411 Eastern diamondback rattlesnake, king cobra and eastern coral snake (although we have
412 identified a wide variety of other ribonuclease genes). The absence of these genes in seven
413 Toxicoferans, coupled with the fact that they were initially described from only 2 out of 11

414 species of snake (Fry et al., 2012b) and 3 out of 18 species of lizard (Fry et al., 2010) would
415 cast doubt on their status as conserved Toxicoferan toxins.

416 3.1.18 Three finger toxins (3ftx)

417 We find 2 transcripts encoding three finger toxin (3ftx)-like genes expressed in the *E.*
418 *coloratus* venom gland, one of which is expressed in all 3 tissues (*3ftx-a*) whilst the other is
419 expressed in the venom and scent glands (*3ftx-b*). Orthologous transcripts of *3ftx-a* are found
420 to be expressed in all three tissues of corn snake, rough green snake salivary gland and skin,
421 and royal python salivary gland. An ortholog of *3ftx-b* is expressed in rough green snake
422 scent gland. We also find a number of different putative *3ftx* genes in our other study species,
423 often expressed in multiple tissues (Figure 2, Supplementary figure S11). Based on the
424 phylogenetic and tissue distribution of both of these genes we suggest that they do not
425 represent venom toxins in *E. coloratus*. As with other proposed Toxicoferan genes such as
426 *complement c3* and *nerve growth factor*, it seems likely that *3ftx* genes are indeed venom
427 components in some species, especially cobras and other elapids (Vonk et al., 2013; Fry et
428 al., 2003), and that the identification of their non-venom orthologs in other species has led to
429 much confusion regarding the phylogenetic distribution of these toxic variants.

430 3.1.19 Vespryn

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

437 3.1.20 Waprin

438 We find a number of “waprin”-like genes in our dataset, expressed in a diverse array of body
439 tissues. Our phylogenetic analyses (Supplementary figure S13) show that previously

440 characterised “waprin” genes (Torres et al., 2003; Fry et al., 2008; Rokyta et al., 2012; Aird
441 et al., 2013; Nair et al., 2007) most likely represent *WAP four-disulfide core domain 2*
442 (*wfdc2*) genes, which have undergone a squamate-specific expansion for which there is no
443 evidence for a venom gland-specific paralog. It is unlikely therefore that these genes
444 represent a Toxicoferan venom toxin. Indeed, the inland taipan (*Oxyuranus microlepidotus*)
445 “Omwaprin” has been shown to be “...non-toxic to Swiss albino mice at doses of up to 10
446 mg/kg when administered intraperitoneally” (Nair et al., 2007) and is more likely to have an
447 antimicrobial function in the venom or salivary gland.

448 **3.2 Putative venom toxins of *Echis coloratus***

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

452 3.2.1 Vascular endothelial growth factor

453 We find four transcripts encoding vascular endothelial growth factor (VEGF) expressed in the
454 venom gland of *E. coloratus*. These correspond to *vegfa*, *vegfb*, *vegfc* and *vegff* and of these,
455 *vegfa*, *b* and *c* are also expressed in the skin and scent gland of this species (Figure 2).
456 Transcripts encoding orthologs of these genes are expressed in all three tissues of all other
457 species used in this study (with the exception of the absence of *vegfa* in corn snake skin). In
458 accordance with previous studies (Rokyta et al., 2011), we find evidence of alternative splicing
459 of *vegfa* transcripts in all species although no variant appears to be tissue-specific. It is likely
460 that a failure to properly recognise and classify alternatively spliced *vegfa* transcripts (Aird et
461 al., 2013) may have contributed to an overestimation of snake venom complexity. *vegfd* was
462 only found to be expressed in royal python salivary gland and scent gland and all three tissues
463 from leopard gecko (Figure 2, Supplementary figure S14). The transcript encoding VEGF-F is
464 found only in the venom gland of *E. coloratus* and, given the absence of any Elapid *vegff*
465 sequences in public databases as well as absence of this transcript in the two species of colubrid

466 in our study, we suggest that *vegf-f* is specific to vipers. Whilst *vegf-f* has a higher transcript
467 abundance in *E. coloratus* venom gland (186.73 FPKM) than *vegf-a* (3.24 FPKM), *vegf-b* (1.28
468 FPKM) and *vegf-c* (1.54 FPKM), compared to other venom genes in this species (see next
469 section) it has a considerably lower transcript abundance suggesting it represents at most a
470 minor venom component in *E. coloratus*.

471 3.2.2 L-amino acid oxidase

472 We find transcripts encoding two *l-amino acid oxidase (laao)* genes in *E. coloratus*, one of
473 which (*laao-b*) has two splice variants (Figure 2, Supplementary figure S15). *laao-a*
474 transcripts are found in all three *E. coloratus* and leopard gecko tissues. *laao-b* is venom
475 gland-specific in *E. coloratus* (based on the available data) and transcripts of the orthologous
476 gene are found in the scent glands of corn snake, rough green snake and royal python. The
477 splice variant *laao-b2* may represent a venom toxin in *E. coloratus* based on its specific
478 expression in the venom gland of this species and elevated expression level (628.84 FPKM).

479 3.2.3 Crotonamine

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

487 **3.3 Proposed venom toxins in *Echis coloratus***

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

492 3.3.1 Cysteine-rich secretory proteins (CRISPs)

493 We find transcripts encoding two distinct CRISPs expressed in the *E. coloratus* venom gland,
494 one of which is also found in skin and scent gland (Figure 2). Phylogenetic analysis of these
495 genes (which we call *crisp-a* and *crisp-b*) reveals that they appear to have been created as a
496 result of a gene duplication event earlier in the evolution of advanced snakes (Supplementary
497 Figure S16). *crisp-a* transcripts are also found in all three corn snake tissues, as well as rough
498 green snake skin and scent gland and royal python scent gland. *crisp-b* is also found in corn
499 snake salivary gland (Figure 2 and Supplementary figure S16) and the phylogenetic and
500 tissue distribution of this gene suggest that it does indeed represent a venom toxin, produced
501 via duplication of an ancestral *crisp* gene that was expressed in multiple tissues, including the
502 salivary gland. The elevated transcript abundance of *crisp-b* (3,520.07 FPKM) in the venom
503 gland of *E. coloratus* further supports its role as a venom toxin in this species (Figure 7). The
504 phylogenetic and tissue distribution and low transcript abundance of *crisp-a* (0.61 FPKM in
505 *E. coloratus* venom gland) shows that it is unlikely to be a venom toxin. We also find no
506 evidence of a monophyletic clade of reptile venom toxins and therefore suggest that, contrary
507 to earlier reports (Fry et al., 2009b; Fry et al., 2010), the CRISP genes of varanid and
508 helodermatid lizards do not represent shared Toxicoferan venom toxins and, if they are
509 indeed toxic venom components, they have been recruited independently from those of the
510 advanced snakes. Regardless of their status as venom toxins, it appears likely that the
511 diversity of CRISP genes in varanid lizards in particular (Fry et al., 2006) has been
512 overestimated as a result of the use of negligible levels of sequence variation to classify
513 transcripts as representing distinct gene products (Supplementary figures S23 and S24).

514 3.3.2 C-type lectins

515 We find transcripts encoding 11 distinct C-type lectin genes in the *E. coloratus* venom gland,
516 one of which (*ctl-a*) is also expressed in the scent gland of this species. The remaining 10
517 genes (*ctl-b* to *k*) are found only in the venom gland and form a clade with other viper C-type

518 lectin genes (Figure 2, Supplementary figure S17). Of these, 6 are highly expressed in the
519 venom gland (*ctl-b* to *d*, *ctl-f* to *g* and *ctl-j*) with a transcript abundance range of 3,706.21-
520 24,122.41 FPKM (Figure 7). The remainder of these genes (*ctl-e*, *ctl-h* to *i* and *ctl-k*) show
521 lower transcript abundance (0.80-1,475.88 FPKM), with two (*ctl-i* and *k*) being more lowly
522 expressed than *ctl-a* (230.06 FPKM). A number of different C-type lectin genes are found in
523 our other study species, often expressed in multiple tissues (Supplementary figure S17). We
524 therefore suggest that the 6 venom-gland specific C-type lectin genes that are highly
525 expressed are indeed venom toxins in *E. coloratus* and that these genes diversified via the
526 duplication of an ancestral gene with a wide expression pattern, including in salivary/venom
527 glands. Based on their selective expression in the venom gland (from available data) the
528 remaining four C-type lectin genes cannot be ruled out as putative toxins, although their
529 lower transcript abundance suggests that they are likely to be minor components in *E.*
530 *coloratus* venom. It should also be noted that a recent analysis of king cobra (*Ophiophagus*
531 *hannah*) venom gland transcriptome and proteome suggested that “...lectins do not contribute
532 to king cobra envenoming” (Vonk et al., 2013).

533 3.3.3 Phospholipase A₂ (PLA₂ Group IIA)

534 We find five transcripts encoding Group IIA PLA₂ genes in *E. coloratus*, three of which are
535 found only in the venom gland and two of which are found only in the scent gland (these
536 latter two likely represent intra-individual variation in the same transcript) (Figure 2,
537 Supplementary figure S18). The venom gland-specific transcript *PLA₂ IIA-c* is highly
538 expressed (22,520.41 FPKM) and likely represents a venom toxin, and may also be a putative
539 splice variant although further analysis is needed to confirm this. *PLA₂ IIA-d* and *IIA-e* show
540 an elevated, but lower, expression level (1,677.15 FPKM and 434.67 FPKM respectively,
541 Figure 7). Based on tissue and phylogenetic distribution we would propose that these three
542 genes may represent putative venom toxins (Table 1).

543 3.3.4 Serine proteases

544 We find 6 transcripts encoding Serine proteases in *E. coloratus* (Figure 2, Supplementary
545 figure S19) which (based on available data) are all venom gland specific. Four of these
546 transcripts are highly expressed in the venom gland (*serine proteases a-c* and *e*; 3,076.01-
547 7,687.03 FPKM) whilst two are expressed at a lower level (*serine proteases d* and *f*; 1,098.45
548 FPKM and 102.34 FPKM respectively, Figure 7). Based on these results we suggest *serine*
549 *proteases a, b, c* and *e* represent venom toxins whilst *serine proteases d* and *f* may represent
550 putative venom toxins (Table 1).

551 3.3.5 Snake venom metalloproteinases

552 We find 21 transcripts encoding snake venom metalloproteinases in *E. coloratus* and of these
553 14 are venom gland-specific, whilst another (*svmp-n*) is expressed in the venom gland and
554 scent gland. Five remaining genes are expressed in the scent gland only whilst another is
555 expressed in the skin (Figure 2, Supplementary figure S20). Of the 14 venom gland-specific
556 SVMPs we find 4 to be highly expressed (5,552.84-15,118.41 FPKM, Figure 7). In the
557 absence of additional data, we classify the 13 venom gland-specific *svmp* genes as venom
558 toxins in this species (Table 1).

559 **4. Discussion**

560 Our transcriptomic analyses have revealed that all 16 of the basal venom toxin genes used to
561 support the hypothesis of a single, early evolution of venom in reptiles (the Toxicofera
562 hypothesis (Vidal and Hedges, 2005; Fry et al., 2006; Fry et al., 2009a; Fry et al., 2009b; Fry
563 et al., 2010; Fry et al., 2012a; Fry et al., 2013)), as well as a number of other genes that have
564 been proposed to encode venom toxins in multiple species are in fact expressed in multiple
565 tissues, with no evidence for consistently higher expression in venom or salivary glands.
566 Additionally, only two genes in our entire dataset of 74 genes in five species were found to
567 encode possible venom gland-specific splice variants (*l-amino acid oxidase b2* and *PLA₂ IIA-*
568 *c*). We therefore suggest that many of the proposed basal Toxicoferan genes most likely
569 represent housekeeping or maintenance genes and that the identification of these genes as

570 conserved venom toxins is a side-effect of incomplete tissue sampling. This lack of support
571 for the Toxicofera hypothesis therefore prompts a return to the previously held view
572 (Kardong et al., 2009) that venom in different lineages of reptiles has evolved independently,
573 once at the base of the advanced snakes, once in the helodermatid (gila monster and beaded
574 lizard) lineage and, possibly, one other time in monitor lizards, although evidence for a
575 venom system in this latter group (Fry et al., 2009b; Fry et al., 2010; Vikrant and Verma,
576 2013) may need to be reinvestigated in light of our findings. The process of reverse
577 recruitment (Casewell et al., 2012), where a venom gene undergoes additional gene
578 duplication events and is subsequently recruited from the venom gland back into a body
579 tissue (which was proposed on the basis of the placement of garter snake and Burmese
580 python “physiological” genes within clades of “venom” genes) must also be re-evaluated in
581 light of our findings.

582 Bites by venomous snakes are thought to be responsible for as many as 1,841,000
583 envenomings and 94,000 deaths annually (predominantly in the developing world
584 (Kasturiratne et al., 2008; Harrison et al., 2009)), and medical treatment of snakebite is reliant
585 on the production of antivenoms containing antibodies, typically from sheep or horses, that
586 will bind and neutralise toxic venom proteins (Chippaux and Goyffon, 1998). Since these
587 antivenoms are derived from the injection of crude venom into the host animal, they are not
588 targeted to the most pathogenic venom components and therefore also include antibodies to
589 weakly- or non-pathogenic proteins requiring the administration of large or multiple doses
590 (Casewell et al., 2013), increasing the risks of adverse reactions. A comprehensive
591 understanding of snake venom composition is therefore vital for the development of the next
592 generation of antivenoms (Harrison, 2004; Wagstaff et al., 2006; Casewell et al., 2013) as it
593 is important that research effort is not spread too thinly through the inclusion of non-toxic
594 venom gland transcripts. Our results suggest that erroneous assumptions about the single
595 origination and functional conservation of venom toxins across the Toxicofera has led to the

596 complexity of snake venom being overestimated by previous authors. We propose that the
597 venom of the painted saw-scaled viper, *Echis coloratus*, is likely to consist of just 34 genes in
598 8 gene families (Table 1, based on venom gland-specific expression and a ‘high’ expression,
599 as defined by presence in the top 25% of transcripts (Williford and Demuth, 2012) in at least
600 two of four venom gland samples), fewer than has been suggested for this and related species
601 in previous EST or transcriptomic studies (Wagstaff and Harrison, 2006; Casewell et al.,
602 2009).

603 It is noteworthy that the results of our analyses accord well with proteomic analyses of
604 venom composition in snakes, with an almost identical complement of 35 toxins in 8 gene
605 families known from the related ocellated carpet viper, *Echis ocellatus* (Wagstaff et al.,
606 2009), where SVMPs, CTLs and PLA₂s were found to be the most abundant proteins. Studies
607 of a range of other venomous snake species have identified a typical complement of between
608 24-61 toxins in 6-14 families (Table 2). Far from being a “complex cocktail” (Izidoro et al.,
609 2006; Calvete et al., 2007b; Wong and Belov, 2012; Casewell et al., 2013), snake venom may
610 in fact represent a relatively simple mixture of toxic proteins honed by natural selection for
611 rapid prey immobilisation, with limited lineage-specific expansion in one or a few particular
612 gene families.

613 In order to avoid continued overestimation of venom complexity, we propose that future
614 transcriptome-based analyses of venom composition must include quantitative comparisons
615 of multiple body tissues from multiple individuals and robust phylogenetic analysis that
616 includes known paralogous members of gene families. We would also encourage the use of
617 clearly explained, justifiable criteria for classifying highly similar sequences as new paralogs
618 rather than alleles or the result of PCR or sequencing errors, as it seems likely that some
619 available sequences from previous studies have been presented as distinct genes on the basis
620 of extremely minor (or even non-existent) sequence variation (see Supplementary figures
621 S21-S24 for examples of identical or nearly identical ribonuclease and CRISP sequences and

622 Supplementary figures S25 and S26 for examples of the same sequence being annotated as
623 two different genes). As a result, the diversity of “venom” composition in these species may
624 have been inadvertently inflated.

625 Additionally, we would encourage the adoption of a standard nomenclature for reptile genes,
626 as the overly-complicated and confusing nomenclature used currently (Table 3) may also
627 contribute to the perceived complexity of snake venom. We propose that such a nomenclature
628 system should be based on the comprehensive standards developed for anole lizards (Kusumi
629 et al., 2011; Hargreaves & Mulley, 2014b). It seems likely that the application of our
630 approach to other species (together with proteomic studies of extracted venom) will lead to a
631 commensurate reduction in claimed venom diversity, with clear implications for the
632 development of next generation antivenoms: since most true venom genes are members of a
633 relatively small number of gene families, it is likely that a similarly small number of
634 antibodies may be able to bind to and neutralise the toxic venom components, especially with
635 the application of “string of beads” techniques (Whitton et al., 1993) utilising fusions of short
636 oligopeptide epitopes designed to maximise the cross-reactivity of the resulting antibodies
637 (Wagstaff et al., 2006).

638

639 **5. Conclusions**

640 We suggest that identification of the apparently conserved Toxicofera venom toxins in
641 previous studies is most likely a side effect of incomplete tissue sampling, compounded by
642 incorrect interpretation of phylogenetic trees and the use of BLAST-based gene identification
643 methods. It should perhaps not be too surprising that homologous tissues in related species
644 would show similar gene complements and the restriction of most previous studies to only the
645 “venom” glands means that monophyletic clades of reptile sequences in phylogenetic trees
646 have been taken to represent monophyletic clades of venom toxin genes. Whilst it is true that
647 some of these genes do encode toxic proteins in some species (indeed, this was often the

648 basis for their initial discovery) the discovery of orthologous genes in other species does not
649 necessarily demonstrate shared toxicity. In short, toxicity in one does not equal toxicity in all.

650

651 **Acknowledgements**

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

662

663

664 **References**

665 Aird, S.D., 2008. Snake venom dipeptidyl peptidase IV: taxonomic distribution and
666 quantitative variation. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* 150, 222-228.

667

668 Aird, S.D., 2005. Taxonomic distribution and quantitative analysis of free purine and
669 pyrimidine nucleosides in snake venoms. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.*
670 140, 109-126.

671

672 Aird, S.D., Watanabe, Y., Villar-Briones, A., Roy, M.C., Terada, K., Mikheyev, A.S., 2013.
673 Quantitative high-throughput profiling of snake venom gland transcriptomes and proteomes
674 (*Ovophis okinavensis* and *Protobothrops flavoviridis*). *BMC Genomics* 14, 790.

675

676 Alper, C., Balavitch, D., 1976. Cobra venom factor: evidence for its being altered cobra C3
677 (the third component of complement). *Science* 191, 1275-1276.

678

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

681

682 Angeletti, R.H., 1970. Nerve growth factor from cobra venom. *Proc. Natl. Acad. Sci. U. S. A.*
683 65, 668-674.

684

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

687

688 Bjarnason, J.B., Fox, J.W., 1994. Hemorrhagic metalloproteinases from snake venoms.
689 *Pharmacol. Ther.* 62, 325-372.

690

691 Brykczynska, U., Tzika, A.C., Rodriguez, I., Milinkovitch, M.C., 2013. Contrasted evolution
692 of the vomeronasal receptor repertoires in mammals and squamate reptiles. *Genome Biol.*
693 *Evol.* 5, 389-401.

694

695 Butte, A.J., Dzau, V.J., Glueck, S.B., 2002. Further defining housekeeping, or "maintenance,"
696 genes: Focus on "A compendium of gene expression in normal human tissues". *Physiological*
697 *genomics* 7, 95-95.

698

699 Calvete, J.J., Escolano, J., Sanz, L., 2007a. Snake venomomics of *Bitis* species reveals large
700 intragenus venom toxin composition variation: application to taxonomy of congeneric taxa. *J.*
701 *Proteome Res.* 6, 2732-2745.

702

703 Calvete, J.J., Fasoli, E., Sanz, L., Boschetti, E., Righetti, P.G., 2009. Exploring the venom
704 proteome of the Western diamondback rattlesnake, *Crotalus atrox*, via snake venomomics and
705 combinatorial peptide ligand library approaches. *J. Proteome Res.* 8, 3055-3067.

706

707 Calvete, J.J., Juárez, P., Sanz, L., 2007b. Snake venomomics. Strategy and applications. *J. Mass*
708 *Spectrom.* 42, 1405-1414.

709

710 Calvete, J.J., Marcinkiewicz, C., Sanz, L., 2006. Snake venomomics of *Bitis gabonica gabonica*.
711 Protein family composition, subunit organization of venom toxins, and characterization of
712 dimeric disintegrins bitisgabonin-1 and bitisgabonin-2. *J. Proteome Res.* 6, 326-336.

713

714 Casewell, N.R., Huttley, G.A., Wüster, W., 2012. Dynamic evolution of venom proteins in
715 squamate reptiles. *Nat. Commun.* 3, 1066.

716

717 Casewell, N.R., Wüster, W., Vonk, F.J., Harrison, R.A., Fry, B.G., 2013. Complex cocktails:
718 the evolutionary novelty of venoms. *Trends Ecol. Evol.* 28, 219-229.

719

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

723

724 Castoe, T.A., de Koning, A.P., Hall, K.T., Card, D.C., Schield, D.R., Fujita, M.K., Ruggiero,
725 R.P., Degner, J.F., Daza, J.M., Gu, W., Reyes-Velasco, J., Shaney, K.J., Castoe, J.M., Fox,
726 S.E., Poole, A.W., Polanco, D., Dobry, J., Vandewege, M.W., Li, Q., Schott, R.K., Kapusta,
727 A., Minx, P., Feschotte, C., Uetz, P., Ray, D.A., Hoffmann, F.G., Bogden, R., Smith, E.N.,
728 Chang, B.S., Vonk, F.J., Casewell, N.R., Henkel, C.V., Richardson, M.K., Mackessy, S.P.,
729 Bronikowski, A.M., Yandell, M., Warren, W.C., Secor, S.M., Pollock, D.D., 2013. The
730 Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc.*
731 *Natl. Acad. Sci. U. S. A.* 110, 20645-20650.

732

733 Castoe, T.A., Fox, S.E., Jason de Koning, A., Poole, A.W., Daza, J.M., Smith, E.N., Mockler,
734 T.C., Secor, S.M., Pollock, D.D., 2011. A multi-organ transcriptome resource for the
735 Burmese Python (*Python molurus bivittatus*). *BMC Res. Notes* 4, 310-0500-4-310.

736

737 Chatrath, S.T., Chapeaurouge, A., Lin, Q., Lim, T.K., Dunstan, N., Mirtschin, P., Kumar,
738 P.P., Kini, R.M., 2011. Identification of novel proteins from the venom of a cryptic snake
739 *Drysdalia coronoides* by a combined transcriptomics and proteomics approach. *J. Proteome*
740 *Res.* 10, 739-750.

741

742 Chippaux, J., Goyffon, M., 1998. Venoms, antivenoms and immunotherapy. *Toxicon* 36,
743 823-846.
744

745 Cotton, J.A., Page, R.D., 2005. Rates and patterns of gene duplication and loss in the human
746 genome. *Proc. Biol. Sci.* 272, 277-283.
747

748 Cousin, X., Créminon, C., Grassi, J., Méflah, K., Cornu, G., Saliou, B., Bon, S., Massoulié,
749 J., Bon, C., 1996a. Acetylcholinesterase from *Bungarus* venom: a monomeric species. *FEBS*
750 *Lett.* 387, 196-200.
751

752 Cousin, X., Bon, S., Duval, N., Massoulie, J., Bon, C., 1996b. Cloning and expression of
753 acetylcholinesterase from *Bungarus fasciatus* venom. A new type of COOH-terminal
754 domain; involvement of a positively charged residue in the peripheral site. *J. Biol. Chem.*
755 271, 15099-15108.
756

757 Cousin, X., Bon, S., Massoulie, J., Bon, C., 1998. Identification of a novel type of
758 alternatively spliced exon from the acetylcholinesterase gene of *Bungarus fasciatus*.
759 Molecular forms of acetylcholinesterase in the snake liver and muscle. *J. Biol. Chem.* 273,
760 9812-9820.
761

762 Junqueira de Azevedo, I.L.M., Farsky, S.H.P., Oliveira, M.L.S., Ho, P.L., 2001. Molecular
763 Cloning and Expression of a Functional Snake Venom Vascular Endothelium Growth Factor
764 (VEGF) from the *Bothrops insularis* Pit Viper: a new member of the VEGF family of
765 proteins. *J. Biol. Chem.* 276, 39836-39842.
766

767 Du, X., Clemetson, K.J., 2002. Snake venom L-amino acid oxidases. *Toxicon* 40, 659-665.

768

769 Eckalbar, W.L., Hutchins, E.D., Markov, G.J., Allen, A.N., Corneveaux, J.J., Lindblad-Toh,
770 K., Di Palma, F., Alfoldi, J., Huentelman, M.J., Kusumi, K., 2013. Genome reannotation of
771 the lizard *Anolis carolinensis* based on 14 adult and embryonic deep transcriptomes. BMC
772 Genomics 14, 49-2164-14-49.

773

774 Fahmi, L., Makran, B., Pla, D., Sanz, L., Oukkache, N., Lkhider, M., Harrison, R.A., Ghalim,
775 N., Calvete, J.J., 2012. Venomics and antivenomics profiles of North African *Cerastes*
776 *cerastes* and *C. vipera* populations reveals a potentially important therapeutic weakness.
777 Journal of proteomics 75, 2442-2453.

778

779 Fry, B.G., Wüster, W., Kini, R.M., Brusica, V., Khan, A., Venkataraman, D., Rooney, A.,
780 2003. Molecular evolution and phylogeny of elapid snake venom three-finger toxins. J. Mol.
781 Evol. 57, 110-129.

782

783 Fry, B.G., Casewell, N.R., Wüster, W., Vidal, N., Young, B., Jackson, T.N., 2012a. The
784 structural and functional diversification of the Toxicofera reptile venom system. Toxicon 60,
785 434-448.

786

787 Fry, B.G., Scheib, H., Junqueira de Azevedo, Inacio de LM, Silva, D.A., Casewell, N.R.,
788 2012b. Novel transcripts in the maxillary venom glands of advanced snakes. Toxicon 59,
789 696-708.

790

791 Fry, B.G., Undheim, E.A., Ali, S.A., Debono, J., Scheib, H., Ruder, T., Jackson, T.N.,
792 Morgenstern, D., Cadwallader, L., Whitehead, D., 2013. Squeezers and leaf-cutters:

793 differential diversification and degeneration of the venom system in toxicoferan reptiles. Mol.
794 Cell. Proteomics 12, 1881-1899.

795

796 Fry, B.G., Vidal, N., Norman, J.A., Vonk, F.J., Scheib, H., Ramjan, S.R., Kuruppu, S., Fung,
797 K., Hedges, S.B., Richardson, M.K., 2006. Early evolution of the venom system in lizards
798 and snakes. Nature 439, 584-588.

799

800 Fry, B.G., Vidal, N., Van der Weerd, L., Kochva, E., Renjifo, C., 2009a. Evolution and
801 diversification of the Toxicofera reptile venom system. Journal of proteomics 72, 127-136.

802

803 Fry, B.G., Wroe, S., Teeuwisse, W., van Osch, M.J., Moreno, K., Ingle, J., McHenry, C.,
804 Ferrara, T., Clausen, P., Scheib, H., 2009b. A central role for venom in predation by *Varanus*
805 *komodoensis* (Komodo dragon) and the extinct giant *Varanus (Megalania) priscus*. Proc.
806 Natl. Acad. Sci. U. S. A. 106, 8969-8974.

807

808 Fry, B.G., 2005. From genome to "venome": molecular origin and evolution of the snake
809 venom proteome inferred from phylogenetic analysis of toxin sequences and related body
810 proteins. Genome Res. 15, 403-420.

811

812 Fry, B.G., Scheib, H., van der Weerd, L., Young, B., McNaughtan, J., Ramjan, S.F., Vidal,
813 N., Poelmann, R.E., Norman, J.A., 2008. Evolution of an arsenal: structural and functional
814 diversification of the venom system in the advanced snakes (Caenophidia). Mol. Cell.
815 Proteomics 7, 215-246.

816

817 Fry, B.G., Winter, K., Norman, J.A., Roelants, K., Nabuurs, R.J., van Osch, M.J., Teeuwisse,
818 W.M., van der Weerd, L., McNaughtan, J.E., Kwok, H.F., Scheib, H., Greisman, L., Kochva,

819 E., Miller, L.J., Gao, F., Karas, J., Scanlon, D., Lin, F., Kuruppu, S., Shaw, C., Wong, L.,
820 Hodgson, W.C., 2010. Functional and structural diversification of the Anguimorpha lizard
821 venom system. *Mol. Cell. Proteomics* 9, 2369-2390.

822

823 Georgieva, D., Risch, M., Kardas, A., Buck, F., von Bergen, M., Betzel, C., 2008.
824 Comparative analysis of the venom proteomes of *Vipera ammodytes ammodytes* and *Vipera*
825 *ammodytes meridionalis*. *J. Proteome Res.* 7, 866-886.

826

827 Gould, S.J., Vrba, E.S., 1982. Exaptation-a missing term in the science of form. *Paleobiology*
828 8, 4-15.

829

830 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis,
831 X., Fan, L., Raychowdhury, R., Zeng, Q., 2011. Full-length transcriptome assembly from
832 RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644-652.

833

834 Hargreaves, A.D., Swain, M.T., Hegarty, M.J., Logan, D.W., Mulley, J.F. (2014a) Restriction
835 and recruitment – gene duplication and the origin and evolution of snake venom toxins
836 *Genome Biology & Evolution* 6, 2088-2095

837

838 Hargreaves, A.D. Mulley, J.F. (2014b) A plea for standardized nomenclature of snake venom
839 toxins. *Toxicon* (in press) doi: 10.1016/j.toxicon.2014.08.070.

840

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

843

844 Harrison, R.A., Hargreaves, A., Wagstaff, S.C., Faragher, B., Lalloo, D.G., 2009. Snake
845 envenoming: a disease of poverty. PLoS Negl. Trop. Dis. 3, e569.
846

847 Harrison, R.A., Ibison, F., Wilbraham, D., Wagstaff, S.C., 2007. Identification of cDNAs
848 encoding viper venom hyaluronidases: cross-generic sequence conservation of full-length and
849 unusually short variant transcripts. Gene 392, 22-33.
850

851 Harvey, A.L., 2001. Twenty years of dendrotoxins. Toxicon 39, 15-26.
852

853 Izidoro, L.F.M., Ribeiro, M.C., Souza, G.R., Sant'Ana, C.D., Hamaguchi, A., Homs-
854 Brandeburgo, M.I., Goulart, L.R., Belebony, R.O., Nomizo, A., Sampaio, S.V., 2006.
855 Biochemical and functional characterization of an l-amino acid oxidase isolated from
856 *Bothrops pirajai* snake venom. Bioorg. Med. Chem. 14, 7034-7043.
857

858 Jia, L., Shimokawa, K., Bjarnason, J.B., Fox, J.W., 1996. Snake venom metalloproteinases:
859 structure, function and relationship to the ADAMs family of proteins. Toxicon 34, 1269-
860 1276.
861

862 Kardong, K., Weinstein, S., Smith, T., 2009. Reptile venom glands: form, function, and
863 future. In: Mackessy, S (Ed.), Handbook of venoms and toxins of reptiles, CRC Press, Boca
864 Raton, Florida, pp. 65-91.
865

866 Kasturiratne, A., Wickremasinghe, A.R., de Silva, N., Gunawardena, N.K., Pathmeswaran,
867 A., Premaratna, R., Savioli, L., Lalloo, D.G., de Silva, H.J., 2008. The global burden of
868 snakebite: a literature analysis and modelling based on regional estimates of envenoming and
869 deaths. PLoS medicine 5, e218.

870

871 Kemparaju, K., Girish, K., 2006. Snake venom hyaluronidase: a therapeutic target. *Cell*
872 *Biochem. Funct.* 24, 7-12.

873

874 Kini, R.M., Doley, R., 2010. Structure, function and evolution of three-finger toxins: mini
875 proteins with multiple targets. *Toxicon* 56, 855-867.

876

877 Kochva, E., 1987. The origin of snakes and evolution of the venom apparatus. *Toxicon* 25,
878 65-106.

879

880 Komori, Y., Nikai, T., 1998. Chemistry and Biochemistry of Kallikrein-Like Enzymes from
881 Snake Venoms. *Toxin Reviews* 17, 261-277.

882

883 Komori, Y., Nikai, T., Sugihara, H., 1988. Biochemical and physiological studies on a
884 kallikrein-like enzyme from the venom of *Crotalus viridis viridis* (Prairie rattlesnake).
885 *Biochim. Biophys. Acta* 967, 92-102.

886

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

889

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

892

893 Kusumi, K., Kulathinal, R.J., Abzhanov, A., Boissinot, S., Crawford, N.G., Faircloth, B.C.,
894 Glenn, T.C., Janes, D.E., Losos, J.B., Menke, D.B., Poe, S., Sanger, T.J., Schneider, C.J.,

895 Stapley, J., Wade, J., Wilson-Rawls, J., 2011. Developing a community-based genetic
896 nomenclature for anole lizards. *BMC Genomics* 12, 554-2164-12-554.
897

898 Kwong, P.D., McDonald, N.Q., Sigler, P.B., Hendrickson, W.A., 1995. Structure of β -
899 bungarotoxin: potassium channel binding by Kunitz modules and targeted phospholipase
900 action. *Structure* 3, 1109-1119.
901

902 Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H.,
903 Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins,
904 D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947-2948.
905

906 Li, M., Fry, B.G., Kini, R.M., 2005. Putting the brakes on snake venom evolution: the unique
907 molecular evolutionary patterns of *Aipysurus eydouxii* (Marbled sea snake) phospholipase A₂
908 toxins. *Mol. Biol. Evol.* 22, 934-941.
909

910 Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data
911 with or without a reference genome. *BMC Bioinformatics* 12, 323-2105-12-323.
912

913 Lomonte, B., Escolano, J., Fernández, J., Sanz, L., Angulo, Y., Gutiérrez, J.M., Calvete, J.J.,
914 2008. Snake venomomics and antivenomics of the arboreal neotropical pitvipers *Bothriechis*
915 *lateralis* and *Bothriechis schlegelii*. *Journal of proteome research* 7, 2445-2457.
916

917 Lynch, M., Conery, J.S., 2003. The evolutionary demography of duplicate genes. *J. Struct.*
918 *Funct. Genomics* 3, 35-44.
919

920 Lynch, V.J., 2007. Inventing an arsenal: adaptive evolution and neofunctionalization of snake
921 venom phospholipase A₂ genes. *BMC evolutionary biology* 7, 2.
922

923 Lynch, M., Conery, J.S., 2000. The evolutionary fate and consequences of duplicate genes.
924 *Science* 290, 1151-1155.
925

926 Lynch, M., Force, A., 2000. The probability of duplicate gene preservation by
927 subfunctionalization. *Genetics* 154, 459-473.
928

929 Mackessy, S.P., 2002. Biochemistry and pharmacology of colubrid snake venoms. *Toxin*
930 *Rev.* 21, 43-83.
931

932 Margres, M.J., Aronow, K., Loyacano, J., Rokyta, D.R., 2013. The venom-gland
933 transcriptome of the Eastern coral snake (*Micrurus fulvius*) reveals high venom complexity in
934 the intragenomic evolution of venoms. *BMC Genomics* 14, 1-18.
935

936 Morita, T., 2005. Structures and functions of snake venom CLPs (C-type lectin-like proteins)
937 with anticoagulant-, procoagulant-, and platelet-modulating activities. *Toxicon* 45, 1099-
938 1114.
939

940 Nair, D., Fry, B., Alewood, P., Kumar, P., Kini, R., 2007. Antimicrobial activity of
941 omwaprin, a new member of the waprin family of snake venom proteins. *Biochem. J.* 402,
942 93-104.
943

944 Ogawa, T., Chijiwa, T., Oda-Ueda, N., Ohno, M., 2005. Molecular diversity and accelerated
945 evolution of C-type lectin-like proteins from snake venom. *Toxicon* 45, 1-14.

946

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

949

950 OmPraba, G., Chapeaurouge, A., Doley, R., Devi, K.R., Padmanaban, P., Venkatraman, C.,
951 Velmurugan, D., Lin, Q., Kini, R.M., 2010. Identification of a novel family of snake venom
952 proteins veficolins from *Cerberus rynchops* using a venom gland transcriptomics and
953 proteomics approach. *J. Proteome Res.* 9, 1882-1893.

954

955 Otto, S.P., Whitton, J., 2000. Polyploid incidence and evolution. *Annu. Rev. Genet.* 34, 401-
956 437.

957

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

961

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

964

965 Pogrel, M.A., Low, M.A., Stern, R., 2003. Hyaluronan (hyaluronic acid) and its regulation in
966 human saliva by hyaluronidase and its inhibitors. *J. Oral Sci.* 45, 85-92.

967

968 Pung, Y.F., Kumar, S.V., Rajagopalan, N., Fry, B.G., Kumar, P.P., Kini, R.M., 2006. Ohanin,
969 a novel protein from king cobra venom: its cDNA and genomic organization. *Gene* 371, 246-
970 256.

971

972 Rádis-Baptista, G., Kubo, T., Oguiura, N., Svartman, M., Almeida, T., Batistic, R.F.,
973 Oliveira, E.B., Vianna-Morgante, ÂM., Yamane, T., 2003. Structure and chromosomal
974 localization of the gene for crotamine, a toxin from the South American rattlesnake, *Crotalus*
975 *durissus terrificus*. *Toxicon* 42, 747-752.

976

977 Richards, R., St Pierre, L., Trabi, M., Johnson, L.A., de Jersey, J., Masci, P.P., Lavin, M.F.,
978 2011. Cloning and characterisation of novel cystatins from elapid snake venom glands.
979 *Biochimie* 93, 659-668.

980

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

983

984 Rokyta, D.R., Lemmon, A.R., Margres, M.J., Aronow, K., 2012. The venom-gland
985 transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC*
986 *Genomics* 13, 312.

987

988 Rokyta, D.R., Wray, K.P., Lemmon, A.R., Lemmon, E.M., Caudle, S.B., 2011. A high-
989 throughput venom-gland transcriptome for the Eastern Diamondback Rattlesnake (*Crotalus*
990 *adamanteus*) and evidence for pervasive positive selection across toxin classes. *Toxicon* 57,
991 657-671.

992

993 Sanz, L., Escolano, J., Ferretti, M., Biscoglio, M.J., Rivera, E., Crescenti, E.J., Angulo, Y.,
994 Lomonte, B., Gutiérrez, J.M., Calvete, J.J., 2008. Snake venomomics of the South and Central
995 American Bushmasters. Comparison of the toxin composition of *Lachesis muta* gathered
996 from proteomic versus transcriptomic analysis. *J. Proteomics* 71, 46-60.

997

998 Schwartz, T.S., Bronikowski, A.M., 2013. Dissecting molecular stress networks: identifying
999 nodes of divergence between life-history phenotypes. *Mol. Ecol.* 22, 739-756.
1000

1001 Schwartz, T.S., Tae, H., Yang, Y., Mockaitis, K., Van Hemert, J.L., Proulx, S.R., Choi, J.,
1002 Bronikowski, A.M., 2010. A garter snake transcriptome: pyrosequencing, de novo assembly,
1003 and sex-specific differences. *BMC Genomics* 11, 694.
1004

1005 Serrano, S.M., Maroun, R.C., 2005. Snake venom serine proteinases: sequence homology vs.
1006 substrate specificity, a paradox to be solved. *Toxicon* 45, 1115-1132.
1007

1008 Siang, A.S., Doley, R., Vonk, F.J., Kini, R.M., 2010. Transcriptomic analysis of the venom
1009 gland of the red-headed krait (*Bungarus flaviceps*) using expressed sequence tags. *BMC Mol.*
1010 *Biol.* 11, 24-2199-11-24.
1011

1012 Strydom, D., 1973. Snake venom toxins: The evolution of some of the toxins found in snake
1013 venoms. *Syst. Biol.* 22, 596-608.
1014

1015 Suhr, S., Kim, D., 1996. Identification of the snake venom substance that induces apoptosis.
1016 *Biochem. Biophys. Res. Commun.* 224, 134-139.
1017

1018 Sunagar, K., Fry, B.G., Jackson, T.N., Casewell, N.R., Undheim, E.A., Vidal, N., Ali, S.A.,
1019 King, G.F., Vasudevan, K., Vasconcelos, V., 2013. Molecular Evolution of Vertebrate
1020 Neurotrophins: Co-Option of the Highly Conserved Nerve Growth Factor Gene into the
1021 Advanced Snake Venom Arsenal. *PloS one* 8, e81827.
1022

1023 Teshima, K.M., Innan, H., 2008. Neofunctionalization of duplicated genes under the pressure
1024 of gene conversion. *Genetics* 178, 1385-1398.
1025

1026 Torres, A.M., Wong, H.Y., Desai, M., Moochhala, S., Kuchel, P.W., Kini, R.M., 2003.
1027 Identification of a Novel Family of Proteins in Snake Venoms. Purification and structural
1028 characterization of nawaprin from *Naja nigricollis* snake venom. *J. Biol. Chem.* 278, 40097-
1029 40104.
1030

1031 Tu, A.T., Hendon, R.R., 1983. Characterization of lizard venom hyaluronidase and evidence
1032 for its action as a spreading factor. *Comp. Biochem. Physiol. B. Comp. Biochem.* 76, 377-
1033 383.
1034

1035 Tzika, A.C., Helaers, R., Schramm, G., Milinkovitch, M.C., 2011. Reptilian-transcriptome
1036 v1. 0, a glimpse in the brain transcriptome of five divergent Sauropsida lineages and the
1037 phylogenetic position of turtles. *EvoDevo* 2, 1-18.
1038

1039 Vidal, N., Hedges, S.B., 2005. The phylogeny of squamate reptiles (lizards, snakes, and
1040 amphisbaenians) inferred from nine nuclear protein-coding genes. *Comptes rendus biologies*
1041 328, 1000-1008.
1042

1043 Vikrant, S., Verma, B.S., 2013. Monitor lizard bite-induced acute kidney injury—a case
1044 report. *Ren. Fail.* 36, 444-446.
1045

1046 Vonk, F.J., Jackson, K., Doley, R., Madaras, F., Mirtschin, P.J., Vidal, N., 2011. Snake
1047 venom: From fieldwork to the clinic. *Bioessays* 33, 269-279.
1048

1049 Vonk, F.J., Casewell, N.R., Henkel, C.V., Heimberg, A.M., Jansen, H.J., McCleary, R.J.,
1050 Kerckamp, H.M., Vos, R.A., Guerreiro, I., Calvete, J.J., Wuster, W., Woods, A.E., Logan,
1051 J.M., Harrison, R.A., Castoe, T.A., de Koning, A.P., Pollock, D.D., Yandell, M., Calderon,
1052 D., Renjifo, C., Currier, R.B., Salgado, D., Pla, D., Sanz, L., Hyder, A.S., Ribeiro, J.M.,
1053 Arntzen, J.W., van den Thillart, G.E., Boetzer, M., Pirovano, W., Dirks, R.P., Spaink, H.P.,
1054 Duboule, D., McGlenn, E., Kini, R.M., Richardson, M.K., 2013. The king cobra genome
1055 reveals dynamic gene evolution and adaptation in the snake venom system. Proc. Natl. Acad.
1056 Sci. U. S. A. 110, 20651-20656.

1057

1058 Wagstaff, S.C., Harrison, R.A., 2006. Venom gland EST analysis of the saw-scaled viper,
1059 *Echis ocellatus*, reveals novel $\alpha_9 \beta_1$ integrin-binding motifs in venom metalloproteinases and
1060 a new group of putative toxins, renin-like aspartic proteases. Gene 377, 21-32.

1061

1062 Wagstaff, S.C., Laing, G.D., Theakston, R.D.G., Papaspyridis, C., Harrison, R.A., 2006.
1063 Bioinformatics and multiepitope DNA immunization to design rational snake antivenom.
1064 PLoS Med. 3, e184.

1065

1066 Wagstaff, S.C., Sanz, L., Juárez, P., Harrison, R.A., Calvete, J.J., 2009. Combined snake
1067 venomomics and venom gland transcriptomic analysis of the ocellated carpet viper, *Echis*
1068 *ocellatus*. J. Proteomics 71, 609-623.

1069

1070 Walsh, B., 2003. Population-genetic models of the fates of duplicate genes. Genetica 118,
1071 279-294.

1072

1073 Warner, T.G., Dambach, L.M., Shin, J.H., O'Brien, J.S., 1981. Purification of the lysosomal
1074 acid lipase from human liver and its role in lysosomal lipid hydrolysis. *J. Biol. Chem.* 256,
1075 2952-2957.
1076

1077 Whitton, J.L., Sheng, N., Oldstone, M.B., McKee, T.A., 1993. A "string-of-beads" vaccine,
1078 comprising linked minigenes, confers protection from lethal-dose virus challenge. *J. Virol.*
1079 67, 348-352.
1080

1081 Williford, A., Demuth, J.P., 2012. Gene expression levels are correlated with synonymous
1082 codon usage, amino acid composition, and gene architecture in the red flour beetle, *Tribolium*
1083 *castaneum*. *Mol. Biol. Evol.* 29, 3755-3766.
1084

1085 Wong, E.S., Belov, K., 2012. Venom evolution through gene duplications. *Gene* 496, 1-7.
1086

1087 Yamazaki, Y., Hyodo, F., Morita, T., 2003a. Wide distribution of cysteine-rich secretory
1088 proteins in snake venoms: isolation and cloning of novel snake venom cysteine-rich secretory
1089 proteins. *Arch. Biochem. Biophys.* 412, 133-141.
1090

1091 Yamazaki, Y., Morita, T., 2004. Structure and function of snake venom cysteine-rich
1092 secretory proteins. *Toxicon* 44, 227-231.
1093

1094 Yamazaki, Y., Takani, K., Atoda, H., Morita, T., 2003b. Snake venom vascular endothelial
1095 growth factors (VEGFs) exhibit potent activity through their specific recognition of KDR
1096 (VEGF receptor 2). *J. Biol. Chem.* 278, 51985-51988.
1097

- 1098 Župunski, V., Kordiš, D., Gubenšek, F., 2003. Adaptive evolution in the snake venom
1099 Kunitz/BPTI protein family. FEBS Lett. 547, 131-136.
- 1100
- 1101
- 1102
- 1103
- 1104
- 1105
- 1106
- 1107
- 1108
- 1109
- 1110
- 1111
- 1112
- 1113

1114 **Tables**

1115

1116 **Table 1.** Predicted venom composition of the painted saw-scaled viper, *Echis coloratus*

1117

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	34

1118

1119

1120

1121

1122

1123

1124

1125

1126

1127

1128

1129

1130 **Table 2.** Predicted numbers of venom toxins and venom toxin families from proteomic studies
 1131 of snake venom accord well with our transcriptome results.
 1132

Species	Number of toxins	Number of toxin families
<i>Bitis caudalis</i> (Calvete et al., 2007a)	30	8
<i>Bitis gabonica gabonica</i> (Calvete et al., 2006)	35	12
<i>Bitis gabonica rhinoceros</i> (Calvete et al., 2007a)	33	11
<i>Bitis nasicornis</i> (Calvete et al., 2007a)	28	9
<i>Bothriechis schlegelii</i> (Lomonte et al., 2008)	?	7
<i>Cerastes cerastes</i> (Fahmi et al., 2012)	25-30	6
<i>Crotalus atrox</i> (Calvete et al., 2009)	~24	~9
<i>Echis ocellatus</i> (Wagstaff et al., 2009)	35	8
<i>Lachesis muta</i> (Sanz et al., 2008)	24-26	8
<i>Naja kaouthia</i> (Kulkeaw et al., 2007)	61	12
<i>Ophiophagus hannah</i> (Vonk et al., 2013)	?	14

<i>Vipera ammodytes</i> (Georgieva et al., 2008)	38	9
--	----	---

1133

1134

1135

1136

1137

1138

1139

1140

1141

1142

1143

1144

1145

1146

1147 **Table 3.** Venom gene nomenclature. Lack of a formal set of nomenclatural rules for venom
 1148 toxins has led to an explosion of different gene names and may have contributed to the
 1149 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]

	<p>Serpentokallikrein [AAG27254]</p> <p>Salmobin [AAC61838]</p> <p>Batroxobin [AAA48553]</p> <p>Nikobin [CBW30778]</p> <p>Gloshedobin [POC5B4]</p> <p>Gussurobin [Q8UVX1]</p> <p>Pallabin [CAA04612]</p> <p>Pallase [AAC34898]</p>
Snake venom metalloproteinase (SVMP)	<p>Stejnihagin-B [ABA40759]</p> <p>Bothropasin [AAC61986]</p> <p>Atrase B [ADG02948]</p> <p>Mocarhagin 1 [AAM51550]</p> <p>Scutatease-1 [ABQ01138]</p> <p>Austrelease-1 [ABQ01134]</p>
Vascular endothelial growth factor (VEGF)	<p>Barietin [ACN22038]</p> <p>Cratrin [ACN22040]</p> <p>Apiscin [ACN22039]</p> <p>Vammin [ACN22045]</p>
Vespryn	<p>Ohanin [AAR07992]</p> <p>Thaicobrin [P82885]</p>
Waprin	<p>Nawaprin [P60589]</p> <p>Porwaprin [B5L5N2]</p> <p>Stewaprin [B5G6H3]</p> <p>Veswaprin [B5L5P5]</p> <p>Notewaprin [B5G6H5]</p>

	Carwaprin [B5L5P0]
--	--------------------

1150

1151

1152 **Figure legends**

1153

1154 **Figure 1. Relationships of key vertebrate lineages and the placement of species**

1155 **discussed in this paper.** A monophyletic clade of reptiles (which includes birds) is

1156 boxed and the Toxicofera (Fry et al., 2013) are shaded. Modified taxon names

1157 have been used for simplicity. Due to the lack of taxonomic resolution within the Colubridae,

1158 we have placed the term colubrids in inverted commas.

1159

1160 **Figure 2. Tissue distribution of proposed venom toxin transcripts.** The majority of

1161 transcripts proposed to encode Toxicoferan venom proteins are expressed in multiple body

1162 tissues. Transcript order follows descriptions in the main text and those transcripts found in

1163 the assembled transcriptomes but which are assigned transcript abundance of <1 FPKM are

1164 shaded orange. VG, venom gland; SAL, salivary gland; SCG, scent gland; SK, skin.

1165

1166 **Figure 3. Maximum likelihood tree of *complement c3* (“cobra venom factor”) sequences.**

1167 Whilst most sequences likely represent housekeeping or maintenance genes, a gene

1168 duplication event in the elapid lineage (marked with *) may have produced a venom-specific

1169 paralog. An additional duplication (marked with +) may have taken place in *Austrelaps*

1170 *superbus*, although both paralogs appear to be expressed in both liver and venom gland.

1171 Geographic separation in king cobras (*Ophiophagus hannah*) from Indonesia and China is

1172 reflected in observed sequence variation. Numbers above branches are Bootstrap values for

1173 500 replicates. Tissue distribution of transcripts is indicated using the following

1174 abbreviations: VG, venom gland; SK, skin; SCG, scent gland, AG, accessory gland; VMNO,

1175 vomeronasal organ and those genes found to be expressed in one or more body tissues are

1176 shaded blue.

1177

1178 **Figure 4. Maximum likelihood tree of dipeptidylpeptidase 3 (*dpp3*) and**
1179 ***dipeptidylpeptidase 4 (*dpp4*) sequences.*** Transcripts encoding *dpp3* and *dpp4* are found in a
1180 wide variety of body tissues, and likely represent housekeeping genes. Numbers above
1181 branches are Bootstrap values for 500 replicates. Tissue distribution of transcripts is indicated
1182 using the following abbreviations: VG, venom gland; SK, skin; SCG, scent gland, AG,
1183 accessory gland; VMNO, vomeronasal organ and those genes found to be expressed in one or
1184 more body tissues are shaded blue.

1185

1186 **Figure 5. Maximum likelihood tree of phospholipase b (*plb*) sequences.** Transcripts
1187 encoding *plb* are found in a wide variety of body tissues, and likely represent housekeeping
1188 genes. Numbers above branches are Bootstrap values for 500 replicates. Tissue distribution of
1189 transcripts is indicated using the following abbreviations: VG, venom gland; SK, skin; SCG,
1190 scent gland, AG, accessory gland; VMNO, vomeronasal organ and those genes found to be
1191 expressed in one or more body tissues are shaded blue.

1192

1193 **Figure 6. Maximum likelihood tree of renin-like sequences.** Renin-like genes are
1194 expressed in a diversity of body tissues. The recently published *Boa constrictor* “RAP-Boa-
1195 1” sequence is clearly a *cathepsin d* gene and is therefore not orthologous to the *Echis*
1196 *ocellatus* renin sequence as has been claimed (Fry et al., 2013). Numbers above branches are
1197 Bootstrap values for 500 replicates. Tissue distribution of transcripts is indicated using the
1198 following abbreviations: VG, venom gland; SK, skin; SCG, scent gland and those genes
1199 found to be expressed in one or more body tissues are shaded blue.

1200

1201 **Figure 7. Graph of transcript abundance values of proposed venom transcripts in the**
1202 ***Echis coloratus* venom gland.** The majority of Toxicoferan transcripts are expressed at
1203 extremely low level, with the most highly expressed genes falling into only four gene

1204 families (C-type lectins, Group IIA phospholipase A₂, serine proteases and snake venom
1205 metalloproteinases). FPKM = Fragments Per Kilobase of exon per Million fragments
1206 mapped.