Review

Snake Genome Sequencing: Results and Future Prospects

Harald M. I. Kerkkamp [1], R. Manjunatha Kini [2], Alexey S. Pospelov [3], Freek J. Vonk [4], Christiaan V. Henkel [1] and Michael K. Richardson [1,*]

[1] Institute of Biology, University of Leiden, Leiden 2300 RA, The Netherlands;
h.m.i.kerkkamp@biology.leidenuniv.nl (H.M.I.K.); henkel.c@hsleiden.nl (C.V.H.)

[2] Department of Biological Science, National University of Singapore, Singapore 117543, Singapore;
dbskinim@nus.edu.sg

[3] Department of Biosciences and Neuroscience Center, University of Helsinki, Helsinki 00014, Finland;
apospelo@mappi.helsinki.fi

[4] Naturalis Biodiversity Center, Darwinweg 2, Leiden 2333 CR, The Netherlands; freek.vonk@naturalis.nl

[*] Correspondence: m.k.richardson@biology.leidenuniv.nl

Abstract: Snake genome sequencing is in its infancy—very much behind the progress made in sequencing the genomes of humans, model organisms and pathogens relevant to biomedical research, and agricultural species. We provide here an overview of some of the snake genome projects in progress, and discuss the biological findings, with special emphasis on toxinology, from the small number of draft snake genomes already published. We discuss the future of snake genomics, pointing out that new sequencing technologies will help overcome the problem of repetitive sequences in assembling snake genomes. Genome sequences are also likely to be valuable in examining the clustering of toxin genes on the chromosomes, in designing recombinant antivenoms and in studying the epigenetic regulation of toxin gene expression.

Keywords: snake; genome; genomics; king cobra; reptile; Malayan pit viper

1. Introduction

The sequencing of animal genomes is uncovering a treasure trove of biological information. Genomes can be defined in various ways. Functional definitions based on concepts of information-encoding and transfer tend to ignore the role of extra-genomic (epigenetic) mechanisms in these processes [1]. Therefore, we shall simply assume the genome to comprise the nucleotide sequence of all nuclear and mitochondrial DNA of an organism. The genome may be sequenced in its entirety via whole genome sequencing [2,3]. It may be more practical for some research questions to sequence only the region of interest, using a 'targeted capture' approach [4]. Targeted approaches include the selective sequencing of bacterial artificial chromosome (BAC) libraries [5].

Genome sequencing has tended to focus on *Homo sapiens* and there are reportedly plans to sequence 2 million human genomes for biomedical research objectives including personalized medicine [6]. Further, the genomes of many animal species used as models in biomedical research, or reared in agriculture, have also been sequenced. The genomes of non-model species have received far less attention although there are plans to sequence many thousands of vertebrate genome in the near future [7].

1.1. Why Snakes Are Interesting

Snake genomics is a neglected topic, as can be seen by the relatively modest number of published genomes and projects in the pipeline (Table 1). Nonetheless, it is a topic that is attracting increasing

interest from a biologists in several sub-disciplines [8]. This interest in snake genomes stems from the medical importance of snakebite in many developing countries [9], the potential for finding novel drugs and other bioactive compounds in venoms [10] and, from the perspective of fundamental research, the extraordinary array of evolutionary novelties found in snakes [11,12].

Table 1. Snake genome projects published or in progress.

Trivial Name	Scientific Name	Family	Notes
Prong-snouted blind snake	*Anilios bituberculatus*	Typhlopidae	F.J. Vonk et al., in progress
Texas blind snake	*Rena dulcis*	Leptotyphlopidae	T.A. Castoe et al., in progress
Boa constrictor	*Boa constrictor*	Boidae	Ref. [13]; GenB: PRJNA210004
Boa constrictor	*Boa constrictor*	Boidae	Ref. [14]
Burmese python	*Python bivittatus*	Pythonidae	Published [2]; GenB: AEQU00000000
Garter snake	*Thamnophis sirtalis*	Colubridae	GenB: LFLD00000000
	Thamnophis elegans	Colubridae	Ref. [13]; GenB: PRJNA210004
Corn snake	*Pantherophis guttatus*	Colubridae	Ref. [15]; GenB: JTLQ01000000
Corn snake	*Pantherophis guttatus*	Colubridae	Targeted sequencing: 5′ hox genes [16]
King cobra	*Ophiophagus hannah*	Elapidae	Published [3]; GenB: AZIM00000000
Malayan pit viper	*Calloselasma rhodostoma*	Viperidae	F.J. Vonk et al., in progress
Five-pacer viper	*Deinagkistrodon acutus*	Viperidae	Ref. [17]
European adder	*Vipera berus berus*	Viperidae	Baylor College of Medicine, Human Genome Sequencing Center; GenB: JTGP00000000
Habu	*Protobothrops flavoviridis*	Viperidae	H. Shibata et al., in progress
Brown spotted pit viper	*Protobothrops mucrosquamatus*	Viperidae	A.S. Mikheyev et al., in progress; GenB: PRJDB4386
Prairie rattlesnake	*Crotalus viridis viridis*	Viperidae	T.A. Castoe et al., in progress
Western diamond-backed rattlesnake	*Crotalus atrox*	Viperidae	Ref. [5]
Timber rattlesnake	*Crotalus horridus*	Viperidae	GenB: LVCR00000000.1
Speckled rattlesnake	*Crotalus mitchellii pyrrhus*	Viperidae	Ref. [18]; GenB: JPMF01000000
Western Diamondback rattlesnake, Mojave rattlesnake and Eastern Diamondback rattlesnake	*Crotalus atrox, C. scutulatus, and C. adamanteus*	Viperidae	Targeted sequencing of bacterial artificial chromosome (BAC) clones containing phospholipase A_2 genes.
Pygmy rattlesnake	*Sistrurus miliarius*	Viperidae	Ref. [13]; GenB: PRJNA210004
Temple pit viper	*Tropidolaemus wagleri*	Viperidae	R.M. Kini et al., in progress

This list is not necessarily exhaustive. Abbreviation: GenB, GenBank accession number. Taxonomy according to the Pubmed Taxonomy database [19].

Snakes (Serpentes) are represented by around 3000 extant species [20]. They show a suite of adaptations common to many lineages of vertebrates that have independently evolved long, thin bodies. This suite includes limb reduction or loss, axial elongation, increase in vertebral count and asymmetry of paired viscera. Extant snakes have completely lost all traces of the forelimb and pectoral girdle. In most species there is also loss of the hindlimb and pelvic girdle [11]. Exceptions include the femoral and pelvic girdle remnants found on each side in Leptotyphlopidae (reviewed in Ref. [21]); and the single pelvic element on each side in Typhlopidae [22]. Pelvic vestiges are also present in Aniliidae, Cylindrophiidae and Anomochilidae; and in boas and pythons. There are both pelvic and femoral vestiges, the latter often tipped with a horny spur [22,23]. Compared to ancestral squamates, snakes show elongation of the primary axis with a high vertebral number and poor demarcation of the vertebral regions [24]. The left lung is reduced in size or absent [25].

Other adaptations include jaw modifications and metabolic adaptations associated with swallowing prey whole [2]; the presence of a venom delivery system [26], consisting of the venom glands and fangs; and heat-sensing "pit organs". Venom delivery systems are found in approximately 600 species in the Elapidae, Viperidae, Colubridae and Atractaspididae [27]. Heat-sensitive pit organs are represented by the loreal pit of Crotalinae [28], and the labial pits of some pythons and boas [29].

In the context of this special journal issue, the most relevant of these adaptations is venom, and the peptide or protein toxins that it contains. Venom can be defined as any glandular secretion produced by a metazoan that can be introduced into the tissues of another animal through a puncture (inflicted by the venomous animal for that purpose) and which incapacitates prey or deters attackers by virtue of its potent bioactivity [30]. Venom, and an associated venom delivery system (a gland connected to a puncturing device), has evolved independently in many animal clades [31,32].

1.2. What Genomes Can Tell Us

As has been pointed out [33], toxin evolution has been studied for many years using traditional sequencing and proteomics approaches; but new tools for genomics, transcriptomics and proteomics are greatly advancing the field. For biomedical research in general, there are many advantages in having genomic sequence data and we now summarize just a few of these advantages. A whole genome sequence allows, in principle, the prediction of all translated genes (the exome) by means of ab initio gene prediction algorithms [34] and homology searches using reference sequences [35]. Because of the paucity of genome sequences and the apparent frequent duplication of toxin-encoding genes, the use of transcriptome data from the same species, or one closely related, makes this task easier. The genes predicted may include genes for translated proteins as well as microRNAs (mRNAs) and other non-coding genes.

Predicted gene sequences can then be used in a host of applications and analyses, ranging from the design of probes for in situ hybridization [26], to searches for genes under positive or negative selection, as inferred by the d_N/d_S ratio [36]. The latter analyses have shown that multiple genes are under selection in snakes, or in clades within the snakes, including some venom toxin genes [3] and developmental genes possibly connected to development of the serpentiform body plan [2].

With genome sequences, evolutionary gene loss can be more confidently asserted than by looking at the transcriptome alone. Hypotheses about gene loss, or the degeneration of functional genes into pseudogenes, can be more easily tested because non-coding pseudogenes can be identified in the genome sequence on the basis of sequence homology or synteny [37]. Synteny refers to the location of loci on the same chromosome, or the order and orientation of neighboring genes, especially when compared across species [37].

Analysis of genome sequences shows that several visual pigment genes have been lost in snakes compared with other squamates [2]. This may be related to the putative fossorial (burrowing) lifestyle of an ancestral snake which might have had reduced eyes [38]. Genomics also reveals that some neurotoxin genes have been lost in the lineage leading to the Western and Eastern diamond-backed rattlesnakes (*Crotalus atrox* and *C. adamanteus*, respectively) [5].

Using genome sequences, it is possible to look for candidate regulatory regions; this in turn may allow genomic regulatory blocks to be identified [39]. In the context of toxinology, it will be especially interesting to examine whether duplicated toxin genes of the same toxin family are clustered [5] and functionally part of a common regulatory landscape—in a way analogous, perhaps, to the well-studied *hox* developmental genes [16]. Genome sequences allow the identification of structural variations, including inversions, insertions, deletions and tandem duplications and other large rearrangements [40]. It is also possible to look for transposable elements and other repetitive sequences [41].

Genomic data can be used in phylogeny reconstruction (which is one aspect of the discipline phylogenomics) although this endeavor is not without difficulties [37,42]. One such difficulty is that the evolution of nucleotide sequences effectively overwrites the ancestral sequence making homologies

(orthologues) more difficult to identify [37,42]. Horizontal gene transfer (between species), gene loss and genome duplications [43] can further obscure the phylogenetic relationships among species (reviewed in Ref. [37]).

1.3. Aims and Objectives of This Review

Our aim here is to review some of the biological results yielded, to date, by snake genomes; and to consider some of the research questions that may one day be solved by the analysis of snake genomes. We will focus mainly on the evolution of venom toxins, but discuss also some questions related to snake morphological and physiological adaptations that are being illuminated by genomics.

2. Status of Snake Genome Sequencing Projects

The sequencing of snake genomes is very much in its infancy. The first draft genomes of snakes to be published were those of the Boa constrictor (*Boa constrictor*) [13,14], Burmese python (*Python molurus bivittatus*) [2] and the king cobra (*Ophiophagus hannah*) [3], followed by a high coverage (238 ×) assembly of the first viper genome (*Crotalus mitchellii*) [18]. Some key data on the first two of these draft genomes are summarized in Table 2. The status of some other snake genome projects known to us, including studies based on targeted capture, is summarized in Table 1. As can be seen, the genome sizes of the Burmese python [2] and king cobra [3] are 1.44 and 1.36–1.59 Gbp, respectively. This is roughly half the size of the human genome and closer to the smaller genomes of some other sauropsids such as the chicken and the anole lizard (Table 2).

Table 2. Selected data from the Burmese python and king cobra draft genomes and comparison with genomes of other species.

Species	Coding Genes (k)	Genome Size (Gb)	Repeats (%)
Burmese python	25 [2]	1.44 [2]	31.8–59.4 [2]
King cobra	21.19 [3]	1.36–1.59 [3]	35.2–60.4 [2]
Chicken	20–23 * [44]	1.05 [44]	4.3–8.0 [45]; 9.4 [44]
Human	20.4 ¶; 19 [46]	3.54 ¶	>66–69 [47]
Anolis	18.5 †	1.70 †	30% ‡ [48]

* v. 85.4 in ensembl.org gives the number of coding genes in the chicken genome as 15,508; ¶ Human genome, build 38; ensembl.org; † GenBank Assembly ID GCA_000090745.1; ‡ Mobile elements.

3. Genome Data in the Reconstruction of Toxin Evolution

3.1. Overview of Possible Mechanisms of Toxin Evolution

Toxin evolution is reviewed in Ref. [32]. Waglerin toxins in Wagler's viper (*Tropidolaemus wagleri*) may well have arisen de novo since no orthologues have been found [49]. This is an exceptional case and in general, the evolution of genes de novo is thought to be comparatively rare. Thus, in the human genome, entirely new genes (i.e., those not found in other primates) are very few in number, and tend not to be expressed in the proteome, suggesting that they function as non-protein-coding genes [46]. In fact, the likelihood of a gene being expressed in the human proteome at all was found to be related to the age of evolutionary origin of that gene [46].

Cysteine-rich secretory protein (CRISP) and kallikrein toxins in Wagler's viper are suggested to have become toxic simply as a result of evolutionary changes in the coding sequence of existing salivary proteins [49]. Indeed, another study concluded that not just a few, but in fact most, snake venom toxins evolved from proteins expressed ancestrally in salivary glandular tissue [50]. In any case, it is clear most venom toxins share close sequence similarity, at least in their functional domains, with known, non-venom genes (physiological or body genes) [49].

Alternative splicing can result in both physiological and toxin isoforms being generated from the same gene in different tissues. This appears to be the case with acetylcholinesterase gene of *Bungarus fasciatus* [51].

3.2. Moonlighting: The Strange Case of Nerve Growth Factor

Nerve growth factor (NGF) is a component of venoms in many snakes. At first sight, it may seem to be nothing more than an innocuous neurotrophin apparently occurring in the venom for no good reason. However, NGF is an extremely potent inducer of mast cell degranulation; thus it is possible that it may produce increased local vascular permeability and toxin absorption; it may also produce or enhance anaphylaxis [52,53]. The possibility that venom nerve growth factor may contribute to the toxicity of venom is further suggested by the fact that, like other true venom toxins, it is under positive selection in at least some snakes [52,53]. Nerve growth factor may also play an ancillary (non-toxic) role while the venom is stored in the venom gland by inhibiting metalloprotease-mediated degradation [54]. Since a single isoform is present in *Bothrops jararaca* [55] it is possible that nerve growth factor may be 'moonlighting' as a venom component—that is, taking on functions in the venom additional to those of its function as a neurotropin (the concept of moonlighting is discussed in Ref. [56]). However, arguing against moonlighting is the fact that nerve growth factor is present in at least two copies in the king cobra genome [3] and in other cobras (reviewed in Ref. [50]; see also Table 3 in the current article).

Table 3. Number of copies (paralogues) of toxin genes in the king cobra (*Ophiophagus hannah*) genome; data from Ref. [3].

Venom Toxin or Toxin Family	Number of Paralogues
3FTx (three-finger toxin) *	21
PLA$_2$ (phospholipase A$_2$) *	12
Lectin *	11
Kunitz *	10
Waprins *	6
Cystatin	5
CRISP (cysteine-rich secretory protein)	3
CVF (cobra venom factor)	3
Kallikrein	3
SVMP (snake venom metalloproteinase)	3
LAAO (L-amino acid oxidase)	2
NGF (nerve growth factor)	2
NP (natriuretic peptide) *	2
Acetylcholinesterase	1
Hyaluronidase	1
PLB (phospholipase-B)	1
VEGF (vascular endothelial growth factors)	1
Vespryn	1

Key: (*) estimated number of paralogues; the current genome assembly is not sufficiently well-scaffolded to allow the number of paralogues to be determined with certainty.

3.3. Gene Duplication

Gene duplication may be important in the evolution of venom toxins at two levels: (i) in the origin of the toxin gene from its ancestral counterpart and (ii) in subsequent expansion of the established toxin gene into a multigene family.

Some toxin genes may have undergone an initial duplication event, after which one copy came to be relatively highly expressed in the venom gland by some change in tissue-specific regulation [32]. The nascent toxin gene could then, in principle, undergo sequence evolution independently of its non-venom paralogue to evolve a new function. This process is called neo-functionalization [57,58]. One problem with neo-functionalization as a mechanism is that mutations are more likely to be deleterious than beneficial [59]. An alternative model of gene evolution after duplication is sub-functionalization. This phenomenon can account for the survival of both paralogues because, weakened in function by deleterious mutations, the two copies will need to be retained in the genome in order to make up, together, the full ancestral function by virtue of their complementary effects [60]. This has been called the duplication-degeneration-complementation (DDC) hypothesis [60].

There is no predictable pattern of duplication events in toxin evolution, as can be readily seen, for example, in the highly variable number of different toxin paralogues in the king cobra genome (Table 3). The number varies from one (hyaluronidase) to 21 (three-finger toxins). It is possible that some genes have undergone what we have referred to as 'hijacking' [3]; that is, sequence modification without duplication (Figure 1 in the current article). Comparative analysis of synteny (Figure 1) suggests that the ancestral PLBD1 gene may have evolved into the king cobra venom phospholipase-B (PLB), and that the HYALP1 gene similarly gave rise to venom-expressed hyaluronidase (HYAL).

An example of a gene that has undergone duplication is the ADAM gene which underwent duplication and subsequently these duplicates evolved into a venom-expressed snake venom metalloproteinase (SVMP) gene (Figure 1). Other toxins that have undergone multiple rounds of duplication to produce multigene families include phospholipase A_2 in rattlesnakes [5]. In that gene family, some paralogues subsequently disappeared from the genome in different lineages, possibly because of a change in prey type [5]. The origin of genes by duplication, and the subsequent loss of some paralogues in this way, is consistent with the birth-and-death model of the evolution of multigene families [61,62].
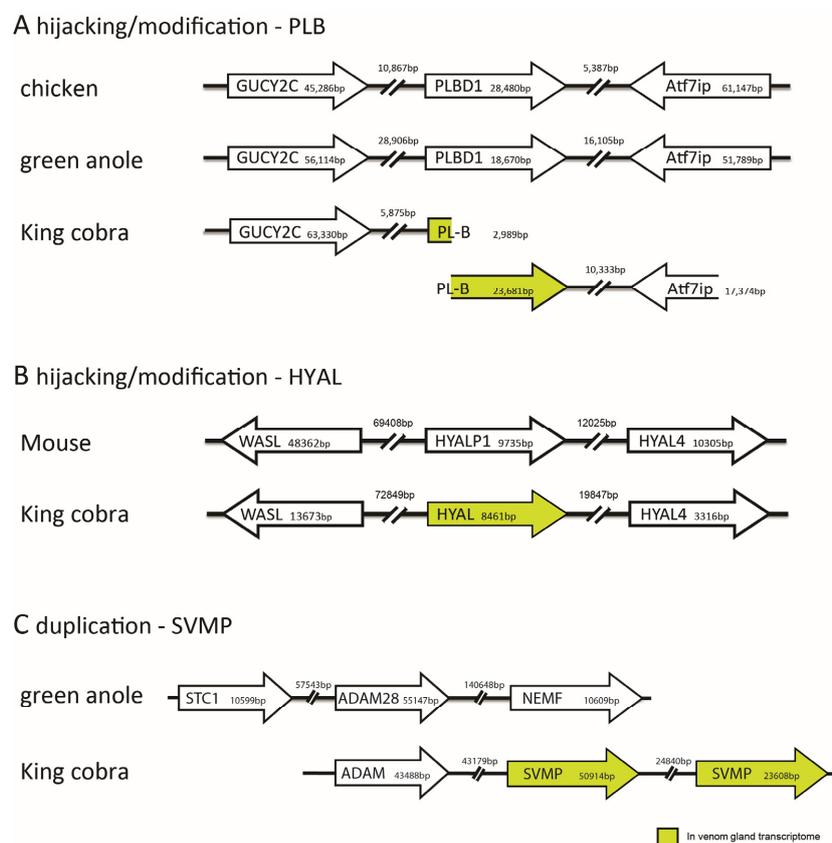

Figure 1. Syntenic comparisons of venom genes in the king cobra with other vertebrates revealing toxin recruitment by hijacking/modification and gene duplication. (**A**) Modification of PLBD1 gene found in the green anole lizard (*Anolis carolinensis*) and the chicken (*Gallus gallus*) results in the venom gland expressed phospholipase-B (PLB). Note that PLB is found split across two king cobra genome scaffolds; (**B**) Modification of HYALP1 gene found in the mouse (*Mus musculus*) results in the venom gland expressed hyaluronidase (HYAL); (**C**) Duplication of the non-venom gland expressed ADAM gene in the king cobra results in a venom gland expressed snake venom metalloproteinase (SVMP) gene. The ADAM gene in the green anole is flanked on both sides by non-SVMP genes, demonstrating the absence of gene duplication in this species. Note that subsequent downstream duplication of the SVMP gene in the king cobra results in multiple venom gland expressed SVMP isoforms. Based on Figure S5 from [3].

3.4. Possible Selective Advantage of Possessing Multigene Toxin Families

A preliminary analysis of the king cobra genome (Figure 2) suggests one possible selective advantage of duplication in the evolution of multigene toxin families. There is a tendency for paralogues that have undergone recent expansion to be more highly expressed in the venom gland transcriptome. More work is required to confirm this hypothesis although it is consistent, for example, with the relationship between amylase abundance and mRNA abundance in mice [63].
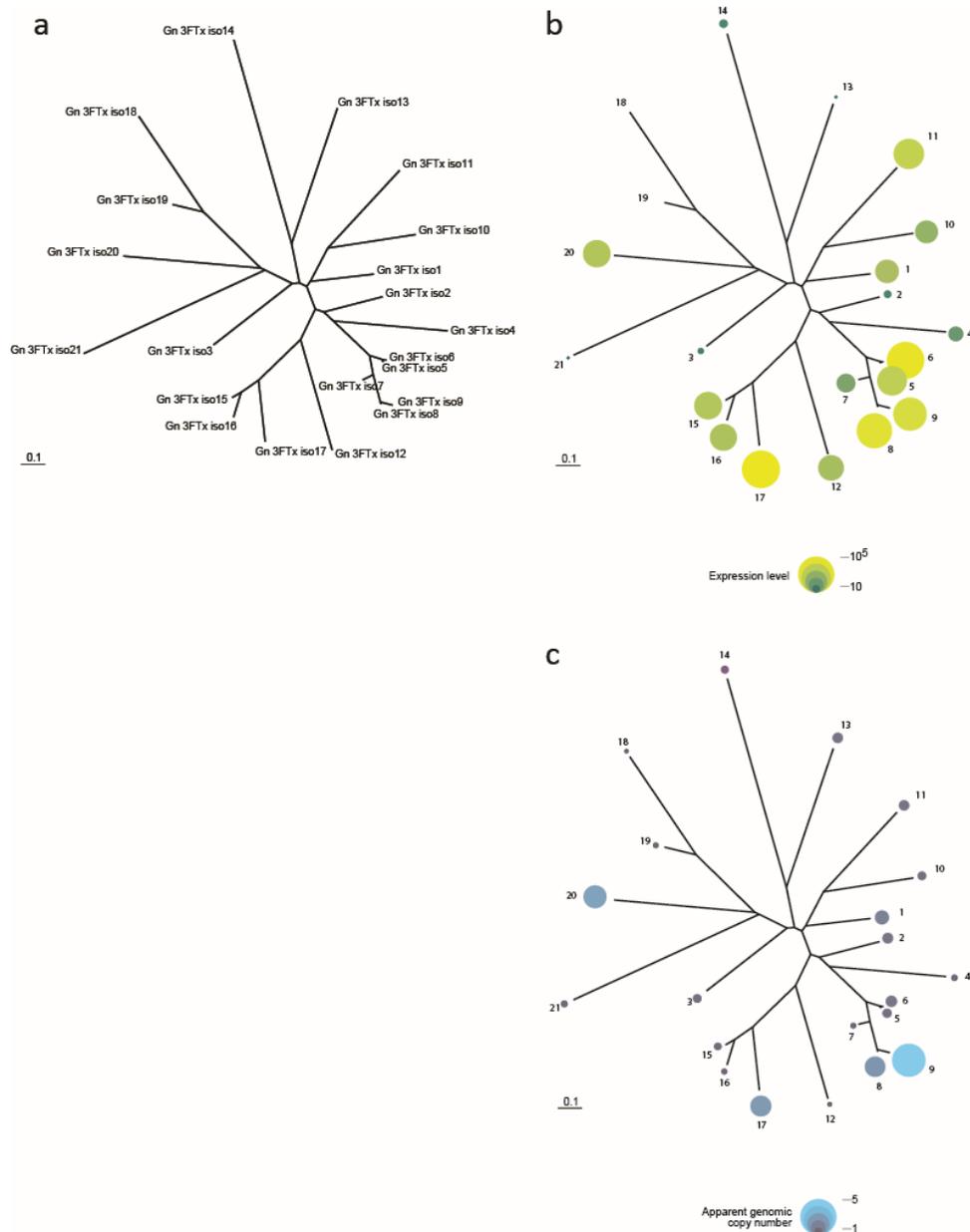

Figure 2. Preliminary analysis of three finger toxin isoforms in the king cobra genome. (**a**) Phylogeny showing isoform numbers; (**b**) Expression level (transcript abundance) in the venom gland; (**c**) Apparent copy number in genome. One hypothesis consistent with the figure is that the more recently expanded paralogues tend to be more highly expressed. The figure is an unpublished analysis by one of us (Christiaan Henkel) based on data in Ref. [3]. See Table 4 for corresponding genome sequencing and accession codes of the three finger toxin isoforms.

Other explanations for the evolution of multiple isoforms of the same toxin is that they might provide broad spectrum toxicity against a range of prey species. Presumably, this is more likely to be advantageous for generalists, than for specialists such as the king cobra. Multiple isoforms might also provide potentiation, so that the toxin complex is more potent than a single toxin. Potentiation is known, for example, in the cone snails [64]. The possession of multiple gene copies might make it more difficult for prey to evolve resistance.

Table 4. King cobra three finger toxin genome sequencing and accession codes. Isoforms correspond with the ones referred to in Figure 2.

3FTX Isoform	Nucleotide Sequence	Accession Code Genbank
Iso1	GATACACCTTGACATGTCTAACACATGAATCATTATTTTTTGAAACCACTGAGAC TTGTTCAGATGGGCAGAACCTATGCTATGCAAAATGGTTTGCAGTTTTTCCAGGTG	AZIM01011044.1
Iso2	GATACACCAGGATATGCCACAAATCTTCTTTTATCTCTGAGACTTGTCCAGATGG GCAGAACCTATGCTATTTAAAATCGTGGTGTGACATTTTTT	AZIM01016929.1
Iso3	GATACACCTTGACATGCATCACATCTGCTCGTAACTTTGAGACTTGTCCACCTGG GCAGAACCTATGCTTTTTAAAATCATGGTATGAAGCTTCAT	AZIM01214498.1
Iso4	TACAAAACCGGTGAACGTATTATTTCTGAGACTTGTCCCCCTGGGCAGGACCTAT GCTATATGAAGACTTGGTGTGACGTTTTTT	AZIM01146344.1
Iso5	GATACACCATGACATGTTACACACAGTACTCATTGTCTCCTCCAACCACTAAGAC TTGTCCAGATGGGCAGAACCTATGCTATAAAAGGTGATTTGCGTTTATTCCACATG	AZIM01015434.1
Iso6	GATACACCACGAAATGCTACGTAACACCTGATGCTACCTCTCAGACTTGTCCAG ATGGGGAGAACATATGCTATACAAAGTCTTGGTGTGACGGTTTTT	AZIM01133918.1
Iso7	GATACACCACGAAATGCTATGTAACACCTGATGCTACCTCTCAGACTTGTCCAGA TGGGGAGAACATATGCTATACAAAGTCTTGGTGTGACGTTTTTT	AZIM01229389.1
Iso8	GATACACCACGAAATGCTACATAACACCTGATGTGAAGTCTCAGACTTGTCCAG ATGGGGAGAACATATGCTATACAAAGACTTGGTGTGATGTTTGGT	AZIM01229389.1
Iso9	GATACACCACGAAATGCTACGTAACACCTGATGTTAAGTCTGAGACTTGTCCAG ATGGGCAGGACATATGCTATACAGAGACTTGGTGTGACGTTTGGT	AZIM01028336.1
Iso10	GATACACCACGAAATGCTACGTAACACCTGATGTTAAGTCTGAGACTTGTCCAG CTGGGCAGGACATATGCTATACAGAGACTTGGTGTGATGCTTGGT	AZIM01097792.1
Iso11	GACACACCAGGATATGTCTCACAGACTACTCAAAAGTTAGTGAAACCATTGAGA TTTGTCCAGATGGGCAGAACTTCTGCTTTAAAAAGTTTCCTAAGGGTATTCCATTTT	AZIM01006046.1
Iso12	GATACACCATGAAATGTCTCACAAAGTACTCCCGGGTTAGTGAAACCTCTCAGA CTTGTCACGTTTGGCAGAACCTATGTTTTAAAAAGTGGCAGAAGG	AZIM01011575.1
Iso13	GACACACCTTGATATGTGTCAAACAGTACACAATTTTTGGTGTAACCCCTGAGAT TTGCGCAGATGGGCAGAACCTATGCTATAAAACATGGCATATGGTGTATCCAGGTG	AZIM01011969.1
Iso14	GATACACCACGAAATGTTACAACCACCAGTCAACGACTCCTGAAACCACTGAAA TTTGTCCAGATTCAGGGTACTTTTGCTATAAAAGCTCTTGGATTGATGGACGTG	AZIM01034614.1
Iso15	GATACACCCTGATATGTCACCGAGTGCATGGACTTCAGACTTGTGAACCAGATG AGAAGTTTTGCTTTAGAAAGACGACAATGTTTTTTCCAAATC	AZIM01009352.1
Iso16	GATACACCAGGAAATGTCTCAACACACCGCTTCCTTTGATCTATANTTAAAATGA CTATTAAGAAGTTGCCATCTA	AZIM01009586.1
Iso17	NATACACCAGGATATGTTTAAAGCAAGAGCCATTTCAACCTGAAACCAGTACAA CTTGTCCAGATGGGGAAGATGCTTGCTATAGTACATTTTGGAGTGATAACC	AZIM01019523.1
Iso18	NATACACCAGGATATGTTTAAAGCAAGAGCCGTTTCAACCTGAAACCACTACAA CTTGTCCAGAAGGGGAGGATGCTTGCTATAATTTGTTTTGGAGTGATCACA	AZIM01052732.1
Iso19	GATACAGCTTGATATGTTTAACCAAGAGACGTATCGACCTGAAACCACTACAA CTTGTCCAGATGGGGAGGACACTTGCTATAGTACATTTTGGAATGATCACCATG	AZIM01009977.1
Iso20	CACAAACCAAGACATGTTACTCATGCACTGGAGCATTTTGTTCTAATCGTCAAAA ATGTTCGGGTGGGCAGGTCATATGCTTTAAAAGTTGGAAAAATACTCTTCTGATAT	AZIM01013260.1
Iso21	CACACACCCTGACATGTTACTCATGCAATGGATTATTATGTTCTGACCGTGAACA ATGTCCAGATGGGTAGGACATATGCTTTAAGAGATGGAATGATACTGATTGGTCAG	AZIM01013561.1
Iso22	GATACAGCTTGACATGTCTCAATTGCCCAGAACAGTATTGTAAAAGAATTCACA CTTGTCGAGATGGGGAGAACGTATGCTTTAAAAGGTTTTACGAGGGTAAACTATTAT	AZIM01071124.1
Iso23	GATACACTCTGTTGTGTTGCAAATGCAATCAAACGGTTTGTGATCTCAATTCGTAT TGTTCAGCAGGCAAGAACCAATGCTATATATTGCAGAATAATA	AZIM01008565.1

3.5. The Selective Expression of Toxin Genes, or Their Ancestral Orthologues, in the Venom Gland

Given that many toxins have arisen by duplication from genes whose ancestral function was something other than that of a venom toxin [5,49,55,65], how did one or more of the duplicates (paralogues) come to be selectively expressed in the venom gland?

3.5.1. Recruitment and Neo-Functionalisation Hypothesis

One scenario for the selective expression of toxin genes in the venom gland is as follows [49]. One of the copies of the ancestral gene underwent a change in tissue-specific regulation so as to become expressed de novo in the venom gland [66]. This paralogue then underwent evolution of its coding sequence so as to become more effective as a venom toxin [3]. Such adaptive changes in

the coding sequence of the newly-recruited gene represent "neo-functionalization"—the evolution of a function not related to the ancestral function (reviewed in Ref. [57,58]). Changes in the coding sequence may be accompanied by additional changes in the regulation of toxin gene transcription, as well as in translation and post-translational modification of the protein [67]. Finally, there is evidence that a toxin gene may ultimately undergo a further change in tissue-specific regulation and revert to being expressed in a tissue or organ other than the venom gland [55,68]. While this hypothesis has been disputed [69], recent work comparing toxin expression in multiple different tissues of *B. jararaca* provided additional evidence for reverse recruitment [55].

3.5.2. Restriction and Sub-Functionalisation Hypothesis

The hypothesis of duplication and neo-functionalisation outlined above has been questioned [50] on the grounds that gene duplication in vertebrate genomes is an extremely rare event, and that persuasive examples of neo-functionalisation have rarely been described in any context. Furthermore, since new transcriptional regulatory relations have to be established for a gene to become highly expressed in the venom gland, the whole scenario is argued to be improbable [50].

An alternative hypothesis is that the ancestral gene was expressed in a wide range of tissues, including the venom gland; it then underwent duplication, with one paralogue becoming restricted in expression to the venom gland and losing expression in the other tissues [50]. Thus, while the recruitment and neo-functionalisation hypothesis is critically dependent on the acquisition of new regulatory regions (for the novel expression of a paralogue in the venom gland), the recruitment and sub-functionalisation depends on the loss of regulatory regions (that ancestrally drove expression in tissues other than the venom gland).

3.5.3. Testing the Recruitment and Restriction Hypotheses

It may well prove to be difficult or impossible to test these hypotheses using comparative transcriptomic data only. An essential pre-requisite will be the availability of multiple snake genomes that provide appropriate taxon sampling, together with the identification of regulatory regions that control the tissue-specific expression of toxin genes and their ancestral paralogues. Putative regulatory sequences will also have to be tested functionally. Progress is being made in this area, as we shall now discuss.

3.6. Mechanisms of Transcriptional Regulation That Might Have Led to Selective Expression of Toxin Genes in the Venom-Gland

3.6.1. Non-Coding RNA Genes

RPTLN are long, non-coding RNA genes that may have been involved in the evolution of snake venom metalloproteinases (SVMPs) from a disintegrin and metalloproteinase (ADAM) gene. According to one hypothesis [66], RPTLN was under the control of a venom gland promotor and its signal sequence became fused with the extracellular domain of the one copy of the ancestrally physiological ADAM gene (the latter having previously undergone tandem duplication). After thus being activated in the venom gland, the ADAM gene evolved into an SVMP [66]. The authors note that their hypothesis can be tested as soon as genome builds for the relevant snake species are available (see also this issue: see Ref. [70] for more information).

3.6.2. Transposable Elements

Another intriguing possibility is that *CR1 LINE* transposable elements, which are much more abundant in advanced than in basal snakes, may have played a role in toxin gene recruitment [2]. *CR1 LINEs* are abundant in the genome of the copperhead (*Agkistrodon contortrix*)—much more abundant than they are in the Burmese python (*Python molurus bivittatus*) genome [41]. We discuss transposable elements and other repetitive sequences in more detail below.

3.6.3. VERSE

It has previously been shown that the gene sequences of TroD (venom prothrombin activator gene) and TrFX (blood coagulation factor X gene) are highly similar, except for promoter and intron 1 regions, indicating that TroD probably evolved by duplication of its plasma counterpart [22]. The insertion, in the promoter of TroD, of a VERSE sequence (VEnom Recruitment/Switch Element) accounts for elevated, but not tissue-specific, expression [23].

3.6.4. AG-Rich Motifs

More recently, it was found that AG-rich motifs, in the first intron, silence gene expression in non-venom gland tissues [71]. These AG-rich motifs are promotor-independent silencers, and such cis-elements are also found in some snake toxin genes, but not in housekeeping genes. Several polycomb group proteins (transcription factors) were identified to bind these motifs to regulate expression. Genome sequences will help in identifying regulatory elements that control tissue-specific expression of toxin genes in venom glands as well as expression of cognate genes in respective tissues.

3.7. Evolution of Toxin Resistance in Snakes as Studied with Genomic Data

Genome sequences have cast light on the resistance by snakes to the toxins of their prey. Thus, the Eastern hog-nosed snake (*Heterodon platirhinos*) is resistant to the tetrodotoxin of *Notophthalmus viridescens* a newt on which it preys [72], and the garter snake *Thamnophis sirtalis* likewise shows resistance to the neurotoxic tetrodotoxin of newts in the genus *Taricha* [73]. Tetrodotoxin resistance in *Thamnophis* is due to modification of the amino acid sequence of tetrodotoxin targets: sodium ion channels on skeletal muscles (Na$_v$1.4) and peripheral neurons (Na$_v$1.6 and Na$_v$1.7; reviewed in Refs. [73,74]). Analysis of snake genomic data and partial sequences, suggests that tetrodotoxin resistance in *Thamnophis* arose stepwise over a long period of evolutionary time, with the ancient modifications of the sodium channels in nerves providing the necessary conditions for evolution of resistance in skeletal muscle sodium-channels [73].

4. Transposable Elements and Other Repetitive Sequences in Snake Genomes

Repetitive elements (repeats) are DNA sequences present in many copies in a genome; they can be classified into tandem repeats and transposable elements [75]. They are relatively abundant in snake genomes, especially the genomes of advanced snakes.

Studies of snake genomes have shown how transposable elements have accumulated in the Hox complex of developmental genes. The Hox complex consists of transcription factor genes that have important roles in regulating embryonic pattern formation. Di-Poï and colleagues selectively sequenced the 5′ regions of the Hox clusters of different species [6]. In the squamates studied, including the corn snake (*Pantherophis guttatus*) the clusters had become expanded in size due the accumulation of numerous transposable elements. These transposons include retrotransposons and DNA transposons, and occur mainly in the introns and intergenic regions [6]. The availability of genomic sequences is also helping to uncover possible incidences of horizontal gene transfer. Thus, the long interspersed element (LINE) non-LTR retrotransposon BovB is suggested to have been transferred, by tics, from squamates to bovids [28].

5. Future Prospects in Snake Genomics

In the future, it may be possible to scan snake genomes for bioactive molecules; this could be done, for example, by comparison with a pharmacophore database. Drug discovery from venoms has already delivered drugs such as Prialt, Integrilin, Captopril and Byetta, and multiple candidates are now progressing in clinical trials [76,77]. Furthermore, peptides derived from venoms are valid pharmacological tools to study diseases [78]. For example, the study of the snake toxin α-bungarotoxin led to characterisation of the nicotinic acetylcholine receptor (nAChR) and a new understanding

of the disease myasthenia gravis [79]. Genome sequences also provide the prospect of generating recombinant antivenoms [80]. Additionally, methylome sequencing will allow us to investigate the role of epigenetics in regulating toxin gene expression [81].

Another step forward in snake genome sequencing will be new sequencing techniques. Next- or second-generation sequencing, so-called because of the advance in Sanger sequencing, produces short reads with low error rates and high throughput [82]. Examples of next-generation platforms are Illumina and Roche 454. The newly-emerging third-generation sequencing platforms are able to provide reads many kB long [83,84]. These platforms include the so-called "PacBio" system—single molecule, real-time (SMRT) sequencing—from Pacific Biosciences; and MinION™ from Oxford Nanopore Technologies [82]. MinION™ has a higher error rate than PacBio [82] and both have higher error rates than second-generation sequencing. For a review of different sequencing platforms and their properties, see Ref. [84].

Given the relatively high percentage of repetitive sequences in advanced snake genomes, a hybrid approach is very promising, and has indeed proved useful for the tackling the same problem in the human genome [84]. This approach involves using the long reads of third-generation sequencing to bridge the gaps due to repeats; and combining them with reads from second-generation sequencing to ameliorate the problem of errors in the long reads [84]. We have used this approach to assemble a draft genome of the Malayan pit viper (unpublished data). However, as the error rate of long-read sequencing improves, the hybrid approach will likely lose favor.

Other interesting developments in sequencing include optical mapping [85], useful for examining the large-scale organization of genomic features, especially around large repetitive clusters; and single-cell RNA-seq [86]. The latter would be very useful in investigating the regulation of toxin production in the venom gland, as it provides a transcriptomic profile per individual cell (and thereby an overview of the different cell types in a gland).

Author Contributions: All authors contributed ideas, comments and references to this article. H.M.I.K. and M.K.R. took the lead in writing and structuring the article. H.M.I.K. and C.V.H. made the figures.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Goldman, A.D.; Landweber, L.F. What is a genome? *PLoS Genet.* **2016**, *12*. [CrossRef] [PubMed]
2. Castoe, T.A.; de Koning, A.P.; Hall, K.T.; Card, D.C.; Schield, D.R.; Fujita, M.K.; Ruggiero, R.P.; Degner, J.F.; Daza, J.M.; Gu, W.; et al. The burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20645–20650. [CrossRef] [PubMed]
3. Vonk, F.J.; Casewell, N.R.; Henkel, C.V.; Heimberg, A.M.; Jansen, H.J.; McCleary, R.J.; Kerkkamp, H.M.; Vos, R.A.; Guerreiro, I.; Calvete, J.J.; et al. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20651–20656. [CrossRef] [PubMed]
4. Jones, M.R.; Good, J.M. Targeted capture in evolutionary and ecological genomics. *Mol. Ecol.* **2016**, *25*, 185–202. [CrossRef] [PubMed]
5. Dowell, N.L.; Giorgianni, M.W.; Kassner, V.A.; Selegue, J.E.; Sanchez, E.E.; Carroll, S.B. The deep origin and recent loss of venom toxin genes in rattlesnakes. *Curr Biol* **2016**, *26*, 2434–2445. [CrossRef] [PubMed]
6. Ledford, H. Astrazeneca launches project to sequence 2 million genomes. *Nature* **2016**, *532*. [CrossRef] [PubMed]
7. Koepfli, K.P.; Paten, B.; O'Brien, S.J. The genome 10K project: A way forward. *Annu. Rev. Anim. Biosci.* **2015**, *3*, 57–111. [CrossRef] [PubMed]

8. Schield, D.R.; Card, D.C.; Reyes-Velasco, J.; Andrew, A.L.; Modahl, C.A.; Mackessy, S.M.; Pollock, D.D.; Castoe, T.A. A role for genomics in rattlesnake research—Current knowledge and future potential. In *Rattlesnakes of Arizona*; Schuett, G.W., Porras, L.W., Reiserer, R.S., Eds.; Eco Books: Rodeo, NM, USA, in press.

9. World Health Organization. *Rabies and Envenomings: A Neglected Public Health Issue: Report of a Consultative Meeting, World Health Organization, Geneva, 10 January 2007*; World Health Organization: Geneva, Switzerland, 2007.

10. Vonk, F.J.; Jackson, K.; Doley, R.; Madaras, F.; Mirtschin, P.J.; Vidal, N. Snake venom: From fieldwork to the clinic: Recent insights into snake biology, together with new technology allowing high-throughput screening of venom, bring new hope for drug discovery. *Bioessays* **2011**, *33*, 269–279. [CrossRef] [PubMed]

11. Coates, M.; Ruta, M. Nice snake, shame about the legs. *Trends Ecol. Evol.* **2000**, *15*, 503–507. [CrossRef]

12. Greene, H.W. *Snakes: The Evolution of Mystery in Nature*; University of California Press: Berkeley, CA, USA, 1997; p. 351.

13. Vicoso, B.; Emerson, J.J.; Zektser, Y.; Mahajan, S.; Bachtrog, D. Comparative sex chromosome genomics in snakes: Differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Biol.* **2013**, *11*. [CrossRef] [PubMed]

14. Bradnam, K.R.; Fass, J.N.; Alexandrov, A.; Baranay, P.; Bechner, M.; Birol, I.; Boisvert, S.; Chapman, J.A.; Chapuis, G.; Chikhi, R.; et al. Assemblathon 2: Evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience* **2013**, *2*. [CrossRef] [PubMed]

15. Ullate-Agote, A.; Milinkovitch, M.C.; Tzika, A.C. The genome sequence of the corn snake (*Pantherophis guttatus*), a valuable resource for evodevo studies in squamates. *Int. J. Dev. Biol.* **2014**, *58*, 881–888. [CrossRef] [PubMed]

16. Di-Poi, N.; Montoya-Burgos, J.I.; Miller, H.; Pourquie, O.; Milinkovitch, M.C.; Duboule, D. Changes in hox genes' structure and function during the evolution of the squamate body plan. *Nature* **2010**, *464*, 99–103. [CrossRef] [PubMed]

17. Yin, W.; Wang, Z.; Li, Q.; Lian, J.; Zhou, Y.; Lu, B.; Jin, L.; Qiu, P.; Zhang, P.; Zhu, W.; et al. Evolution trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. *Nat. Commun.* **2016**, *7*. [CrossRef] [PubMed]

18. Gilbert, C.; Meik, J.M.; Dashevsky, D.; Card, D.C.; Castoe, T.A.; Schaack, S. Endogenous hepadnaviruses, bornaviruses and circoviruses in snakes. *Proc. R. Soc.* **2014**, *281*. [CrossRef] [PubMed]

19. Pubmed taxonomy database. Available online: https://www.Ncbi.Nlm.Nih.Gov/taxonomy (accessed on 30 November 2016).

20. Vidal, N.; Delmas, A.S.; David, P.; Cruaud, C.; Couloux, A.; Hedges, S.B. The phylogeny and classification of caenophidian snakes inferred from seven nuclear protein-coding genes. *C. R. Biol.* **2007**, *330*, 182–187. [CrossRef] [PubMed]

21. Pinto, R.R.; Martins, A.R.; Curcio, F.; Ramos, L.O. Osteology and cartilaginous elements of trilepida salgueiroi (amaral, 1954) (scolecophidia: Leptotyphlopidae). *Anat. Rec. (Hoboken)* **2015**, *298*, 1722–1747. [CrossRef] [PubMed]

22. Boulenger, G.A. *Catalogue of the Snakes in the British Museum (Natural History)*; British Museum (Natural History): London, UK, 1893; Volume 1, p. 448.

23. Cohn, M.J.; Tickle, C. Developmental basis of limblessness and axial patterning in snakes. *Nature* **1999**, *399*, 474–479. [CrossRef] [PubMed]

24. Head, J.J.; Polly, P.D. Evolution of the snake body form reveals homoplasy in amniote hox gene function. *Nature* **2015**, *520*, 86–89. [CrossRef] [PubMed]

25. Van Soldt, B.J.; Metscher, B.D.; Poelmann, R.E.; Vervust, B.; Vonk, F.J.; Muller, G.B.; Richardson, M.K. Heterochrony and early left-right asymmetry in the development of the cardiorespiratory system of snakes. *PLoS ONE* **2015**, *10*. [CrossRef] [PubMed]

26. Vonk, F.J.; Admiraal, J.F.; Jackson, K.; Reshef, R.; de Bakker, M.A.; Vanderschoot, K.; van den Berge, I.; van Atten, M.; Burgerhout, E.; Beck, A.; et al. Evolutionary origin and development of snake fangs. *Nature* **2008**, *454*, 630–633. [CrossRef] [PubMed]

27. Jackson, K. Evolution of the venom conducting fang in snakes. *Integr. Comp. Biol.* **2002**, *42*, 1249.

28. Hofstadler Deiques, C. The development of the pit organ of bothrops jararaca and crotalus durissus terrificus (serpentes, viperidae): Support for the monophyly of the subfamily crotalinae. *Acta Zool.* **2002**, *83*, 175–182. [CrossRef]

29. Gracheva, E.O.; Ingolia, N.T.; Kelly, Y.M.; Cordero-Morales, J.F.; Hollopeter, G.; Chesler, A.T.; Sanchez, E.E.; Perez, J.C.; Weissman, J.S.; Julius, D. Molecular basis of infrared detection by snakes. *Nature* **2010**, *464*, 1006–1011. [CrossRef] [PubMed]

30. Weinstein, S.A. Snake venoms: A brief treatise on etymology, origins of terminology, and definitions. *Toxicon* **2015**, *103*, 188–195. [CrossRef] [PubMed]

31. Fry, B.G.; Roelants, K.; Champagne, D.E.; Scheib, H.; Tyndall, J.D.; King, G.F.; Nevalainen, T.J.; Norman, J.A.; Lewis, R.J.; Norton, R.S.; et al. The toxicogenomic multiverse: Convergent recruitment of proteins into animal venoms. *Annu. Rev. Genom. Hum. Genet.* **2009**, *10*, 483–511. [CrossRef] [PubMed]

32. Casewell, N.R.; Wuster, W.; Vonk, F.J.; Harrison, R.A.; Fry, B.G. Complex cocktails: The evolutionary novelty of venoms. *Trends Ecol. Evol.* **2013**, *28*, 219–229. [CrossRef] [PubMed]

33. Reyes-Velasco, J.; Card, D.C.; Andrew, A.L.; Shaney, K.J.; Adams, R.H.; Schield, D.R.; Casewell, N.R.; Mackessy, S.P.; Castoe, T.A. Expression of venom gene homologs in diverse python tissues suggests a new model for the evolution of snake venom. *Mol. Biol. Evol.* **2015**, *32*, 173–183. [CrossRef] [PubMed]

34. Majoros, W.H.; Pertea, M.; Salzberg, S.L. Tigrscan and glimmerhmm: Two open source ab initio eukaryotic gene-finders. *Bioinformatics* **2004**, *20*, 2878–2879. [CrossRef] [PubMed]

35. Collins, J.E.; White, S.; Searle, S.M.; Stemple, D.L. Incorporating RNA-seq data into the zebrafish ensembl genebuild. *Genome Res.* **2012**, *22*, 2067–2078. [CrossRef] [PubMed]

36. Spielman, S.J.; Wan, S.; Wilke, C.O. A comparison of one-rate and two-rate inference frameworks for site-specific dn/ds estimation. *Genetics* **2016**, *24*, 2499–2511. [CrossRef] [PubMed]

37. Tekaia, F. Inferring orthologs: Open questions and perspectives. *Genom. Insights* **2016**, *9*, 17–28. [CrossRef] [PubMed]

38. Simoes, B.F.; Sampaio, F.L.; Jared, C.; Antoniazzi, M.M.; Loew, E.R.; Bowmaker, J.K.; Rodriguez, A.; Hart, N.S.; Hunt, D.M.; Partridge, J.C.; et al. Visual system evolution and the nature of the ancestral snake. *J. Evol. Biol.* **2015**, *28*, 1309–1320. [CrossRef] [PubMed]

39. Irimia, M.; Maeso, I.; Roy, S.W.; Fraser, H.B. Ancient cis-regulatory constraints and the evolution of genome architecture. *Trends Genet.* **2013**, *29*, 521–528. [CrossRef] [PubMed]

40. Tattini, L.; D'Aurizio, R.; Magi, A. Detection of genomic structural variants from next-generation sequencing data. *Front. Bioeng. Biotechnol.* **2015**, *3*. [CrossRef] [PubMed]

41. Castoe, T.A.; Hall, K.T.; Mboulas, M.L.G.; Gu, W.; de Koning, A.P.; Fox, S.E.; Poole, A.W.; Vemulapalli, V.; Daza, J.M.; Mockler, T.; et al. Discovery of highly divergent repeat landscapes in snake genomes using high-throughput sequencing. *Genome Biol. Evol.* **2011**, *3*, 641–653. [CrossRef] [PubMed]

42. Telford, M.J.; Copley, R.R. Improving animal phylogenies with genomic data. *Trends Genet.* **2011**, *27*, 186–195. [CrossRef] [PubMed]

43. Taylor, J.S.; Van de Peer, Y.; Braasch, I.; Meyer, A. Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **2001**, *356*, 1661–1679. [CrossRef] [PubMed]

44. Hillier, L.W.; Miller, W.; Birney, E.; Warren, W.; Hardison, R.C.; Ponting, C.P.; Bork, P.; Burt, D.W.; Groenen, M.A.; Delany, M.E.; et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **2004**, *432*, 695–716. [CrossRef] [PubMed]

45. Wicker, T.; Robertson, J.S.; Schulze, S.R.; Feltus, F.A.; Magrini, V.; Morrison, J.A.; Mardis, E.R.; Wilson, R.K.; Peterson, D.G.; Paterson, A.H.; et al. The repetitive landscape of the chicken genome. *Genome Res.* **2005**, *15*, 126–136. [CrossRef] [PubMed]

46. Ezkurdia, I.; Juan, D.; Rodriguez, J.M.; Frankish, A.; Diekhans, M.; Harrow, J.; Vazquez, J.; Valencia, A.; Tress, M.L. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum. Mol. Genet.* **2014**, *23*, 5866–5878. [CrossRef] [PubMed]

47. De Koning, A.P.; Gu, W.; Castoe, T.A.; Batzer, M.A.; Pollock, D.D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **2011**, *7*. [CrossRef] [PubMed]

48. Alfoldi, J.; Di Palma, F.; Grabherr, M.; Williams, C.; Kong, L.; Mauceli, E.; Russell, P.; Lowe, C.B.; Glor, R.E.; Jaffe, J.D.; et al. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **2011**, *477*, 587–591. [CrossRef] [PubMed]

49.	Fry, B.G. From genome to "venome": Molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. *Genome Res.* **2005**, *15*, 403–420. [CrossRef] [PubMed]

50.	Hargreaves, A.D.; Swain, M.T.; Hegarty, M.J.; Logan, D.W.; Mulley, J.F. Restriction and recruitment-gene duplication and the origin and evolution of snake venom toxins. *Genome Biol. Evol.* **2014**, *6*, 2088–2095. [CrossRef] [PubMed]

51.	Cousin, X.; Bon, S.; Massoulie, J.; Bon, C. Identification of a novel type of alternatively spliced exon from the acetylcholinesterase gene of bungarus fasciatus. Molecular forms of acetylcholinesterase in the snake liver and muscle. *J. Biol. Chem.* **1998**, *273*, 9812–9820. [CrossRef] [PubMed]

52.	Sunagar, K.; Fry, B.G.; Jackson, T.N.; Casewell, N.R.; Undheim, E.A.; Vidal, N.; Ali, S.A.; King, G.F.; Vasudevan, K.; Vasconcelos, V.; et al. Molecular evolution of vertebrate neurotrophins: Co-option of the highly conserved nerve growth factor gene into the advanced snake venom arsenalf. *PLoS ONE* **2013**, *8*. [CrossRef]

53.	Kostiza, T.; Meier, J. Nerve growth factors from snake venoms: Chemical properties, mode of action and biological significance. *Toxicon* **1996**, *34*, 787–806. [CrossRef]

54.	Wijeyewickrema, L.C.; Gardiner, E.E.; Gladigau, E.L.; Berndt, M.C.; Andrews, R.K. Nerve growth factor inhibits metalloproteinase-disintegrins and blocks ectodomain shedding of platelet glycoprotein vi. *J. Biol. Chem.* **2010**, *285*, 11793–11799. [CrossRef] [PubMed]

55.	Junqueira-de-Azevedo, I.L.; Bastos, C.M.; Ho, P.L.; Luna, M.S.; Yamanouye, N.; Casewell, N.R. Venom-related transcripts from bothrops jararaca tissues provide novel molecular insights into the production and evolution of snake venom. *Mol. Biol. Evol.* **2015**, *32*, 754–766. [CrossRef] [PubMed]

56.	Jeffery, C.J. Protein species and moonlighting proteins: Very small changes in a protein's covalent structure can change its biochemical function. *J. Proteom.* **2016**, *134*, 19–24. [CrossRef] [PubMed]

57.	True, J.R.; Carroll, S.B. Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol.* **2002**, *18*, 53–80. [CrossRef] [PubMed]

58.	Taylor, J.S.; Raes, J. Duplication and divergence: The evolution of new genes and old ideas. *Annu. Rev. Genet.* **2004**, *38*, 615–643. [CrossRef] [PubMed]

59.	Loewe, L.; Hill, W.G. The population genetics of mutations: Good, bad and indifferent. *Philos. Trans. R. Soc. B Biol. Sci.* **2010**, *365*, 1153–1167. [CrossRef] [PubMed]

60.	Force, A.; Lynch, M.; Pickett, F.B.; Amores, A.; Yan, Y.L.; Postlethwait, J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **1999**, *151*, 1531–1545. [PubMed]

61.	Fry, B.G.; Wuster, W.; Kini, R.M.; Brusic, V.; Khan, A.; Venkataraman, D.; Rooney, A.P. Molecular evolution and phylogeny of elapid snake venom three-finger toxins. *J. Mol. Evol.* **2003**, *57*, 110–129. [CrossRef] [PubMed]

62.	Nei, M.; Rooney, A.P. Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* **2005**, *39*, 121–152. [CrossRef] [PubMed]

63.	Meisler, M.H.; Antonucci, T.K.; Treisman, L.O.; Gumucio, D.L.; Samuelson, L.C. Interstrain variation in amylase gene copy number and mRNA abundance in three mouse tissues. *Genetics* **1986**, *113*, 713–722. [PubMed]

64.	Olivera, B.M.; Seger, J.; Horvath, M.P.; Fedosov, A.E. Prey-capture strategies of fish-hunting cone snails: Behavior, neurobiology and evolution. *Brain Behav. Evol.* **2015**, *86*, 58–74. [CrossRef] [PubMed]

65.	Fry, B.G.; Vidal, N.; Norman, J.A.; Vonk, F.J.; Scheib, H.; Ramjan, S.F.; Kuruppu, S.; Fung, K.; Hedges, S.B.; Richardson, M.K.; et al. Early evolution of the venom system in lizards and snakes. *Nature* **2006**, *439*, 584–588. [CrossRef] [PubMed]

66.	Sanz-Soler, R.; Sanz, L.; Calvete, J.J. Distribution of rptln genes across reptilia: Hypothesized role for rptln in the evolution of svmps. *Integr. Comp. Biol.* **2016**. [CrossRef] [PubMed]

67.	Casewell, N.R.; Wagstaff, S.C.; Wuster, W.; Cook, D.A.; Bolton, F.M.; King, S.I.; Pla, D.; Sanz, L.; Calvete, J.J.; Harrison, R.A. Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 9205–9210. [CrossRef] [PubMed]

68.	Casewell, N.R.; Huttley, G.A.; Wuster, W. Dynamic evolution of venom proteins in squamate reptiles. *Nat. Commun.* **2012**, *3*. [CrossRef] [PubMed]

69. Hargreaves, A.D.; Swain, M.T.; Logan, D.W.; Mulley, J.F. Testing the toxicofera: Comparative transcriptomics casts doubt on the single, early evolution of the reptile venom system. *Toxicon* **2014**, *92*, 140–156. [CrossRef] [PubMed]

70. Sanz, L.; Calvete, J.J. Insights into the evolution of a snake venom multi-gene family from the genomic organization of echis ocellatus svmp genes. *Toxins (Basel)* **2016**, *8*. [CrossRef] [PubMed]

71. Han, S.X.; Kwong, S.; Ge, R.; Kolatkar, P.R.; Woods, A.E.; Blanchet, G.; Kini, R.M. Regulation of expression of venom toxins: Silencing of prothrombin activator trocarin d by ag-rich motifs. *FASEB J.* **2016**, *30*, 2411–2425. [CrossRef] [PubMed]

72. Feldman, C.R.; Durso, A.M.; Hanifin, C.T.; Pfrender, M.E.; Ducey, P.K.; Stokes, A.N.; Barnett, K.E.; Brodie, E.D., 3rd; Brodie, E.D., Jr. Is there more than one way to skin a newt? Convergent toxin resistance in snakes is not due to a common genetic mechanism. *Heredity (Edinb)* **2016**, *116*, 84–91. [CrossRef] [PubMed]

73. McGlothlin, J.W.; Kobiela, M.E.; Feldman, C.R.; Castoe, T.A.; Geffeney, S.L.; Hanifin, C.T.; Toledo, G.; Vonk, F.J.; Richardson, M.K.; Brodie, E.D.; et al. Historical contingency in a multigene family facilitates adaptive evolution of toxin resistance. *Curr. Biol.* **2016**, *26*, 1616–1621. [CrossRef] [PubMed]

74. Soong, T.W.; Venkatesh, B. Adaptive evolution of tetrodotoxin resistance in animals. *Trends Genet.* **2006**, *22*, 621–626. [CrossRef] [PubMed]

75. Padeken, J.; Zeller, P.; Gasser, S.M. Repeat DNA in genome organization and stability. *Curr. Opin. Genet. Dev.* **2015**, *31*, 12–19. [CrossRef] [PubMed]

76. Marcinkiewicz, C. Functional characteristic of snake venom disintegrins: Potential therapeutic implication. *Curr. Pharm. Des.* **2005**, *11*, 815–827. [CrossRef] [PubMed]

77. Laing, G.D.; Moura-da-Silva, A.M. Jararhagin and its multiple effects on hemostasis. *Toxicon* **2005**, *45*, 987–996. [CrossRef] [PubMed]

78. McCleary, R.J.; Kini, R.M. Non-enzymatic proteins from snake venoms: A gold mine of pharmacological tools and drug leads. *Toxicon* **2013**, *62*, 56–74. [CrossRef] [PubMed]

79. Kini, R.M.; Doley, R. Structure, function and evolution of three-finger toxins: Mini proteins with multiple targets. *Toxicon* **2010**, *56*, 855–867. [CrossRef] [PubMed]

80. Wagstaff, S.C.; Laing, G.D.; Theakston, R.D.; Papaspyridis, C.; Harrison, R.A. Bioinformatics and multiepitope DNA immunization to design rational snake antivenom. *PLoS Med.* **2006**, *3*. [CrossRef] [PubMed]

81. Bird, A. Perceptions of epigenetics. *Nature* **2007**, *447*, 396–398. [CrossRef] [PubMed]

82. Karlsson, E.; Larkeryd, A.; Sjodin, A.; Forsman, M.; Stenberg, P. Scaffolding of a bacterial genome using minion nanopore sequencing. *Sci. Rep.* **2015**, *5*. [CrossRef] [PubMed]

83. Lu, H.; Giordano, F.; Ning, Z. Oxford nanopore minion sequencing and genome assembly. *Genom. Proteom. Bioinform.* **2016**, *31*, 265–279. [CrossRef] [PubMed]

84. Xiao, W.; Wu, L.; Yavas, G.; Simonyan, V.; Ning, B.; Hong, H. Challenges, solutions, and quality metrics of personal genome assembly in advancing precision medicine. *Pharmaceutics* **2016**, *8*. [CrossRef] [PubMed]

85. Levy-Sakin, M.; Ebenstein, Y. Beyond sequencing: Optical mapping of DNA in the age of nanotechnology and nanoscopy. *Curr. Opin. Biotechnol.* **2013**, *24*, 690–698. [CrossRef] [PubMed]

86. Saliba, A.E.; Westermann, A.J.; Gorski, S.A.; Vogel, J. Single-cell RNA-seq: Advances and future challenges. *Nucleic Acids Res.* **2014**, *42*, 8845–8860. [CrossRef] [PubMed]