Snake venoms represent a danger to human health, but also a goldmine of bioactive proteins that can be harnessed for drug discovery purposes. The evolution of snakes and their venom has been studied for decades, particularly via traditional morphological and basic genetic methods alongside venom proteomics. However, while the field of genomics has matured rapidly over the past two decades due to the development of Next Generation Sequencing (NGS) technologies, snake genomics remains in its infancy. Here, we provide an overview of the state-of-the-art in snake genomics and discuss its potential implications for studying venom evolution and toxinology. Based on current knowledge, gene duplication and positive selection are key mechanisms in the neofunctionalization of snake venom proteins. This makes snake venoms important evolutionary drivers that explain the remarkable venom diversification and adaptative variation observed in these reptiles. Gene duplication and neofunctionalization have also generated a large number of repeat sequences in snake genomes that pose a significant challenge to DNA sequencing, resulting in the need for substantial computational resources and longer sequencing read length for high quality genome assembly. Fortunately, owing to constantly improving sequencing technologies and computational tools, we are now able to explore the molecular mechanisms of snake venom evolution in unprecedented detail. Such novel insights have the potential to impact the design and development of antivenoms and possibly other drugs, as well as provide new fundamental knowledge on snake biology and evolution.
Response to Reviewers:

We would like to thank the reviewers and the editor for a fast and very constructive review process. We have attempted to address the reviewer comments as good as possible, but please be critical about our responses/corrections, if you feel they or the manuscript could be further improved!

Reviewer #1: In this review, the authors provide an overview of recent applications of genomic analysis to study snake venom systems. This review covers an impressive span of topics related to genomics, snake venom, and the intersection of the two, and includes details from many relatively recent studies. There is minimal synthesis of the topics covered beyond echoing the major conclusions of the cited studies, and similar reviews have been published on this topic recently (Schield et al. 2021; Zancolli & Casewell, 2020).

I have several specific comments and requested edits that I feel should be addressed prior to the publication of this manuscript:

Response: References has been included to support the statement (lines 335-337).

Response: We have modified the sentence to defined the origin of the venom system as "prevalent hypothesis" rather than "generally accepted and we have included the suggested reference.

Section 3.4 – I would suggest that the authors add a statement at the end of the first paragraph noting that relatively few methods have been used to study selection in venom genes, and have done so in a small number of species. Additional studies are needed before broad conclusions can be made about selection on venom genes. In fact, some studies have found evidence for other modes of selection acting on venom regions (i.e., balancing selection; Rautsaw et al 2019).

Response: We have included and additional paragraph at the end of the section (lines 618-623) and referenced as suggested.

Figure 1 - the caption says the data is from snakedatabase.org, however the citation refers to reptile-database.org. Please correct the caption to be consistent with the citation (reptile-database.org).

Response: Corrected.

Table 1 - this table would be more informative if additional species were included. Specifically, only 2 of the 5 species in the table are venomous (Deinagkistrodon and Ophiophagus). As this review aims to be comprehensive, I think all currently available venomous snake data should be summarized here. Additions should include at least Naja naja (Suryamohan et al 2020), Crotalus viridis (Schield et al 2019), Crotalus tigris (Margres et al 2021), Thamnophis sirtalis (a rear-fanged venomous colubrid; Perry et al 2018), Protobothrops flavoviridis (Shibata et al 2018).

Response: We acknowledge the improvement upon implementation of the reviewer suggestion. The table has been updated with the suggested species and now, they are...
separated as non-venomous/venomous. A disclaimer has been included in the table description to highlight that the gene numbers for non-venomous species are homologue proteins found in other organs rather than in venom glands as listed for venomous species.

Tables 1 and 2 - indicate which species are and are not considered venomous. Response: Table 1 has been reformatted to differentiate venomous and non-venomous snakes, and table2 include a new column with this information

Line 231 - cite the NCBI accession for the T. elegans assembly Response: Reference included

Line 271 - typo, delete "0" from "0rattlesnake" Response: Corrected.

Reviewer #2: Review: GIGA-D-21-00410, The rise of genomics in snake venom research: recent advances and future perspectives

The authors present a general review of the state of genomics of snakes and their venoms, with specific focus on what we stand to learn about venom function and evolution from individual snake genomes and subsequent comparative genomics. The authors' approach is very general, covering several components. In some places the authors' include specific details from the several published genome studies in snakes. These details are welcome, and in some places further clarification is needed. The paper will be a nice primer for someone starting to think about the many promises and complexities of genomics and venom genomics in particular, and highlights some of the things we stand to learn from additional genomes that are sure to come in the future. I suggest several places for revisions that, with major revision of both wording and associated figures, can lead to a clearer and more impactful manuscript. Sincerely, Matt Holding.

Major:

Line 55-58, and Figure 1: These lines and the topology of the tree in the figure are somewhat misleading. First, saying snakes "evolved from lizards" belies the complexity here. Snakes are actually a very successful lineage of lizards. This leads to the more important point…calling snakes and lizards "two groups" may make sense colloquially, but not phylogenetically. Several recent studies have shown that Serpentes is the sister group to Anguimorph lizards, and that these two are therefore more closely related to one another than to other lizards. As such, from a perspective of monophyly, lizards and snakes do not form "two groups". Moreover, the topology in Figure 1 is therefore incorrect, as it shows Serpentes sharing a common ancestor with presumably all "Other squamates", when in fact Serpentes + Anguimorph lizards share a common ancestor, and that ancestral lizard shared a common ancestor with "Other squamates"

https://bmcecolevol.biomedcentral.com/articles/10.1186/1471-2148-13-93
https://royalsocietypublishing.org/doi/full/10.1098/rsbl.2012.0703

Response: Thank you for the suggested references. The text has been updated (lines 55-59) and the references has been included.

Lines 221-222: Authors should expand here as a lead-in to the downstream discussions of studying venom genes with whole genomes. You mention "multi-gene families", but the real need for high quality contiguous genomes comes from tandemly array "gene islands", which of course the case for all of the major venom gene families that you review in depth. It is therefore worth specific mention here and later that tandem duplication of similar sequence presents a particularly hard challenge for assembly software, and therefore long reads are needed for the best quality assessments of venom genes. Lines 388-393 contain the current mention of these
gene islands. They are hard too assemble without PacBIO HiFi or Minion long reads, and this should be mentioned extensively in a paper like this. One. Response: Good point, we agree with the Reviewer that long-read technology should be specifically mentioned here, as they are essential for the assembly of continuous genomes. We have therefore added two sentences in this paragraph that hopefully make the point clearer now, with a reference to Figure 2 where the long-read technology is also mentioned (lines 232-236).

Lines 346-347: A few lines to convince the reader that these differences aren't attributable to differences in the sequencing technologies (and therefore the ability to recover repetitive elements) would be appropriate here. Response: We have included a paragraph exemplifying that the sequencing approaches are not responsible for the registeres differences and references accordingly (lines 369-373)

Line 366-367: Where is the hypothesis coming from? There is no citation and it hasn't been previously mentioned. Why would microsatellites impact venom evolution? Are they enriched near venom genes? Otherwise this pure speculation needs to be removed. Response: We have removed the statement.

Lines 382 to 387: This is presented as if these are surprising instances of convergence. Rather, these are clearly homologous locations of these tandemly arrayed gene islands. Response: We have modified the text clarified the point and included a statement about the higher GC-content and faster recombination of macrocromosomes (lines 406-407 and 414-416).

Line 479-484: This is confusingly written. Which gene family(s) are you talking about? Name them rather than just saying "Toxin genes". The presentation of the idea in figure 3 doesn't help the confusion, but rather worsens it. MP, SP, CTL, etc., are large gene superfamilies, so of course they are all over the genome of vertebrates, and very few are venom. You need to undertake much clearer writing to get your point across and link the verbage here to the figure, or remove this content about whole genome duplications completely and just stick to the better understood components of venom evolution. If you think there is clear evidence for an importance of these more ancient WGD events in venom gene evolution, then it certainly merits a place in the paper. But if so, this needs much more detail to differentiate the idea of more recent common ancestors of the venom progenitor genes from the broader gene superfamilies (most of which are housekeeping genes) to which these venom genes belong. Response: We have detailed the toxin gene families studied in the P. flavoridis genome and modified the wording to point across figure 3.

Line 486-491: This is a major misrepresentation of these results. The scaffolds presented are tiny, on the order of only kilobases. Based on the more complete chromosome-scale assemblies done in Crotalus, it is clear that SVSP, SVMP, PLA2, and CTLP all occur in single gene clusters in pitvipers. You need to differentiate genomic scaffolds (a technical artifact of assembly quality) from gene clusters (a biological reality of genomic architecture). And then when doing so, it will always be best to default to the findings of those genomes that are done at chromosome scale, rather than in thousands of tiny scaffolds, until better genomes come out for the other species. See work from Sean Carroll's lab scaffolding individual gene families, or the Crotalus viridis or C. tigris genomes. Response: Thank you for pointing out. We acknowledge the misunderstanding in the terminology. We have modified the text to make it clear that we are referring to gene clusters rather than scaffolds (lines 498-503).

Minor:

Lines 117-120: Unclear why you would mention RADseq, as it is highly unlikely to hit a venom gene. Alternately, you could expand that it could help reveal potential population demographic trends that underly venom variation, as we did here: https://academic.oup.com/biolinnean/article/132/2/297/6042638?login=true
Response: We thank the Reviewer for a good point here. We have chosen to still make a mention about the use of RAD-seq technology, to exemplify to the community another RRS technique that could be useful to study venom evolution. We have modified the sentence as suggested by the Reviewer by adding the Holding et al. study as a direct example of the use of RAD-seq for this purpose while acknowledging that it is less suitable for the study of venom genes (lines 119-124).

Line 297: This also implies greater overall gene density in T. sirtalis (genes/Mb) and potentially differences is repetitive element content. The authors should list out these possible causes and then review which are supported by the literature, rather than just mentioning average intron length differences, which amount to a relatively small overall amount of any genome. Tim
Response: We acknowledge that this paragraph needed clarification. We have modified the text to include the considerations about the gene density and included references to support the statements (lines 311-319).

Line 322: There is little to no evidence that "environmental conditions" impact venom gene expression. Clarify this statement. What does matter is population of origin, age of snake, and time since last expulsion of venom.
Response: We have clarified the point as suggested and included supporting references (lines 343-344).

Line 336: "fluid genome" is a very imprecise term. I think you want to say "high degree of evolvability in structural features of the genome".
Response: We thank the Reviewer for the suggestion. We have chosen to remove the term and change it for the suggestion made by the Reviewer (lines 357-358).

Line 468: "chromatin-wise" is unclear wording. Reword.
Response: We have modified the text to emphasize that highly expressed toxin genes are usually on regions where chromatin is available and the methylation allows for higher transcription of the genes, when compared with non toxin genes (lines 491-492).

Line 470: "metavenom network" is also not a generally known, well-defined term. Define or reword to be more specific.
Response: We have reworded the sentence and defined “metavenom network” in the text (line 495).

Line 479-484: This is confusingly written. Which specific gene family(s) are you talking about? Name them rather than just saying "Toxin genes".
Response: We have modified the text and included a description of the specific toxin families (lines 504-513)

| Question | Response |
|----------|----------|
| Are you submitting this manuscript to a special series or article collection? | No |
| **Experimental design and statistics** | Yes |

Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our Minimum Standards Reporting Checklist. Information essential to interpreting the data presented should be made available in the figure legends.
| Question | Answer |
|----------|--------|
| Have you included all the information requested in your manuscript? | |
| Resources | Yes |
| A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite Research Resource Identifiers (RRIDs) for antibodies, model organisms and tools, where possible. | |
| Have you included the information requested as detailed in our Minimum Standards Reporting Checklist? | Yes |
| Availability of data and materials | Yes |
| All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in publicly available repositories (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript. | |
| Have you have met the above requirement as detailed in our Minimum Standards Reporting Checklist? | Yes |
The rise of genomics in snake venom research: recent
advances and future perspectives

Wei-qiao Rao1,2, Konstantinos Kalogeropoulos1, Morten E. Allentoft1,6, Shyam Gopalakrishnan1, Wei-
ing Zhao2, Christopher T. Workman1, Cecilie Knudsen1, Belén Jiménez-Mena3, Lorenzo Seneci1,
Mahsa Mousavi-Derazmahalleh3, Timothy P. Jenkins1, Esperanza Rivera-de-Torre1, Si-qi Liu2,6,
Andreas H. Laustsen1*

1Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby,
2Department of Mass Spectrometry, Beijing Genomics Institute-Research, Shenzhen, China
3Trace and Environmental DNA (TrEnD) Laboratory, School of Molecular and Life Sciences, Curtin
University, Perth, Australia
4Globe Institute, University of Copenhagen, Copenhagen, Denmark
5DTU Aqua, Technical University of Denmark, Silkeborg, Denmark

Address for correspondence:

Dr. Siqi Liu
Professor
Department of Mass Spectrometry,
Beijing Genomics Institute-Research
Shenzhen, China
*siqiliu@genomics.cn

Dr. Andreas H. Laustsen
Professor
Department of Biotechnology and Biomedicine
Technical University of Denmark
DK-2800 Kongens Lyngby, Denmark
*ahola@bio.dtu.dk
Abstract

Snake venoms represent a danger to human health, but also a goldmine of bioactive proteins that can be harnessed for drug discovery purposes. The evolution of snakes and their venom has been studied for decades, particularly via traditional morphological and basic genetic methods alongside venom proteomics. However, while the field of genomics has matured rapidly over the past two decades due to the development of Next Generation Sequencing (NGS) technologies, snake genomics remains in its infancy. Here, we provide an overview of the state-of-the-art in snake genomics and discuss its potential implications for studying venom evolution and toxinology. Based on current knowledge, gene duplication and positive selection are key mechanisms in the neofunctionalization of snake venom proteins. This makes snake venoms important evolutionary drivers that explain the remarkable venom diversification and adaptive variation observed in these reptiles. Gene duplication and neofunctionalization have also generated a large number of repeat sequences in snake genomes that pose a significant challenge to DNA sequencing, resulting in the need for substantial computational resources and longer sequencing read length for high quality genome assembly. Fortunately, owing to constantly improving sequencing technologies and computational tools, we are now able to explore the molecular mechanisms of snake venom evolution in unprecedented detail. Such novel insights have the potential to impact the design and development of antivenoms and possibly other drugs, as well as provide new fundamental knowledge on snake biology and evolution.

Keywords: Snake genomics; DNA sequencing; venom; venom evolution; snakes; snake toxins
1. Background

Snakes (Squamata: Serpentes) represent a monophyletic lineage, comprising approximately 3,600 extant species found in all continents, except Antarctica [1,2]. From an evolutionary perspective, these reptiles stand out for their characteristic lack of limbs, elongated body shape, and exclusively carnivorous diet. Even before the advent of genetic approaches, conventional anatomical and morphology-based phylogenetic evidence unambiguously suggested that snakes are nested within evolved from lizards, with the Anguimorpha lineage (monitor lizards, glass lizards, beaded lizards, etc) as their closest relatives [3–5]. Together with amphisbaenians, snakes and all other lizards thus these two groups form the largest branch of terrestrial vertebrates, the squamate reptiles [3–5].

Snakes have many specialized adaptations compared to other reptile lineages. For example, the evolution of infrared sensing pits in pitvipers (Viperidae: Crotalinae), boas (Boidae), and pythons (Pythonidae), and of a venom apparatus in several snake families (Fig. 1), provide these animals with exceptional predatory capabilities despite the loss of limbs and the degradation of visual and auditory perception in many (but not all) species [6–8]. Moreover, severe jaw modifications and low metabolic rates enable snakes to swallow and digest large prey whole, further consolidating their position as formidable predators [9,10]. Thus, snakes are important model organisms for evolutionary studies, and have yielded insights into limb development [11–13], sex chromosome evolution [14], and venom evolution [15].

Fig. 1. Schematic diagram of snake evolution based on data from Reptile-database.org [16]. Snakes (Serpentes) are divided into the three main infraorders: Scolecophidia, Henophidia, and Alethinophidia, which combined encompass approximately 24 families (7 shown here). Families comprising venomous species have been marked with a skull and crossbones symbol. Colubridae constitutes the largest family of snakes, encompassing 52% of the approximately 3,566 snake species currently described. The total number of currently described venomous snake species (2,901), which predominantly fall within the families Homalopsidae, Lamprophiidae, Colubridae, Elapidae, and...
Viperidae. Only snake species that have undergone whole-genome sequencing and assembly are listed in this figure.

The development of next-generation sequencing (NGS) technologies in recent decades has allowed researchers to generate large genomic datasets and rendered the assembly and characterization of complete genomes a routine task. Despite the availability of NGS since the early 2000s, the utilization of these technologies to assemble and study complete snake genomes has been very limited, especially when compared to the amount of research that has been conducted in the fields of snake venom proteomics and transcriptomics [17]. It was not until 2013 that the first snake genomes based on high-throughput sequencing data were published for the Burmese python (*Python bivittatus*), the Red-tailed boa (*Boa constrictor constrictor*), and the King cobra (*Ophiophagus hannah*) [9,18,19]. Fortunately, snake genome research has eventually gained more attention, with 18 new genomes being released since 2013 and several more currently in progress [15,20–31]. These increased sequencing efforts have already revealed intriguing insights into the regulation and expression of venom-related genes. As an example, a large number of dormant toxin-encoding genes with unknown bioactivity were identified in the Okinawan habu (*Protobothrops flavoviridis*) [15]. Such discoveries could be of high scientific value and may improve our basic understanding of the interplay between protein function and evolution. Furthermore, as toxins from several animal lineages are known to possess different types of bioactivity, some of them could find utility in a variety of applications, from the development of novel therapeutics [32] to biopesticides [33] and molecular research tools [34]. With only 21 snake genomes publicly available to date, there is great unexplored scientific potential in sequencing and analyzing more snake genomes [17,35].

From a broader perspective, having access to a complete or nearly complete assembled genome provides an excellent basis for addressing a wide range of biological research questions. For example, genomic data can be used to predict protein-coding exons [36] (including exons in genes that
recently underwent pseudogenization), non-expressed genes, translated proteins, and microRNA (miRNA) genes [37]. Genomic data may also allow for the identification of toxin orthologs using comparative studies and homology searches [38]. Knowledge of homology is crucial for the reliability of functional annotation of genomes and can provide fundamental information on evolution and speciation processes [39,40]. Therefore, complete genomes are crucial to the field of proteomics as well, as the absence of reliable genome-derived protein libraries forces researchers to rely on homologous proteins from other organisms as a benchmark to compare newly characterized protein sequences against. This results in severely limited accuracy in identifying potentially homologous proteins, which consequently leads to overlooking and/or misrepresenting evolutionary patterns. This is especially relevant considering the likely widespread occurrence of alternative splicing in snake genomes, which gives rise to multiple mRNA products that in turn result in various isoform of a particular toxin [41–43]. Extensive post-genomic and post-translational modifications are also at play, leading to often remarkable discrepancies between genome, transcriptome, and proteome in terms of expression and sequence identity [44–46]. Along this line, comparative analysis of whole snake genomes could likely provide invaluable insight on the evolution and structure of the gene regulatory network responsible for the expression of venom genes in these animals (and arguably venomous amniotes in general) [47].

Several approaches are available to obtain reliable genomic data. Among them is reduced-representation sequencing (RRS), in which only a part of the genome is sequenced [48]. Conversely, restriction capture sequencing techniques allows for specific areas of interest (e.g., the exon part of the genome) to be targeted and sequenced, at a lower cost compared to WGS [49]. Although less suitable for studying venom genes, Restriction-Associated DNA sequencing (RAD-seq) uses restriction enzymes to obtain genome-wide sequencing data, which has recently been used e.g., to study population demographic trends that underly venom variation [50]. Nevertheless,
reliable detection of homologous genes across species and/or lineages can be hindered by the acquisition, loss, or pseudogenization of genes [40]. One way to overcome this challenge is to use whole-genome sequencing (WGS), which represents a more comprehensive resource for the detection of homologous genes as it provides the entire genotype of the target organism(s) [40]. WGS can also provide information on genomic variability of a species, and potentially discover and quantify the extent of selective (e.g., positive/purifying selection and hitchhiking effects) and neutral forces (e.g., genetic drift) driving venom evolution [51].

This review aims to provide a comprehensive summary of the current knowledge on snake genomics, with a particular focus on the current use and future potential of high-throughput DNA sequencing technologies in the field of snake toxinology. Moreover, we discuss how these technologies can be used to expand our current knowledge on snake venom evolution and toxin diversification.

2. Current status of snake venom research

2.1 Overview of snake toxin families

Studies have estimated that between 19,000 and 25,000 toxins are found in venoms from the Elapidae and Viperidae snake families, but only a few thousands have been characterized [52]. Nonetheless, this body of knowledge has proven sufficient for the systematic classification of snake venom toxins into 63 families, most of which are, however, only found in a small percentage of snake species and/or in negligible amounts within venom mixtures. [53]. The four families generally considered to be of highest relevance both from a clinical (human envenoming cases) and an ecological perspective (e.g. prey incapacitation) are the three-finger toxins (3FTxs), phospholipases A$_2$ (PLA$_2$s), snake venom metalloproteinases (SVMPs), and snake venom serine proteinases (SVSPs). Other widespread snake venom protein families include Cysteine-rich secretory proteins (CRISPs), L-amino acid oxidases...
LAAOs), and C-type lectin-like proteins (CTLPs) [53]. An overview of the main snake venom toxin families is provided in Table 1.
Table 1. Number of toxin-encoding genes for 22 toxin families in selected venomous and non-venomous reptile species. In venomous snake genomes, the numbers refer to the venom gland genome only. Non-venomous species lack venom glands, and the indicated numbers refer to homologous proteins expressed in other organs. *The green anole (Anolis carolinensis) was selected as outgroup taxon as it is a non-venomous, non-snake squamate with a complete genome sequence available.

| Venom protein family | Non-venomous | Venomous |
|----------------------|--------------|----------|
|                      | Anolis carolinensis* | Boa constrictor | Python bivittatus | Thamnophis sirtalis | Ophiophagus hannah | Naja naja | Deinagkistrodon acutus | Protobothrops flavoviridis | Crotalus viridis | Crotalus tigris | Bothrops jararaca |
| 5'-nucleotidases | 5Nase | 1 | 1 | 1 | - | 1 | 2 | 1 | 1 | 5 | - | 1 |
| Acetylcholinesterase | ACeH | 22 | 11 | 12 | - | 16 | 2 | 14 | - | 7 | - | - |
| Bovine pancreatic trypsin inhibitors | BPTI | 86 | 39 | 49 | - | 53 | 3 | 7 | - | 2 | 2 | - |
| Bradykinin-potentiating peptides and C-type natriuretic peptides | BNP | 1 | 3 | 1 | - | 6 | 3 | 2 | 1 | 1 | 1 | 1 |
| Cysteine-rich secretory proteins | CRISPs | 2 | 1 | 1 | 2 | 3 | 7 | 2 | 2 | 4 | 2 | 1 |
| C-type lectins and C-type lectin-like proteins | CTLPs | 5 | 7 | 6 | - | 13 | 2 | 22 | 10 | 6 | 5 | 6 |
| Disintegrins | Dvs | - | - | - | - | - | - | 3 | 2 | - | - | - |
| Factor V | - | 5 | 5 | 6 | - | 5 | - | 5 | - | 3 | - | - |
| Factor X | - | 9 | 11 | 11 | - | 11 | - | 11 | - | - | - | - |
| Hyaluronidases | HYAL | 5 | 6 | 6 | 1 | 6 | 3 | 6 | 1 | 4 | 1 | 1 |
| L-amino acid oxidases | LAAO | 4 | 5 | 6 | 2 | 3 | 3 | 4 | 1 | 4 | 2 | 2 |
| Nerve growth factors or neurotrophins | NGF | 5 | 5 | 5 | - | 5 | 3 | 4 | 1 | 2 | 1 | 1 |
| Phosphodiesterases | PDE | 6 | 6 | 5 | - | 5 | 1 | 5 | 1 | - | 1 |
| Phospholipases A<sub>2</sub> | PL<sub>A</sub> <sub>2</sub> | 1 | 1 | 1 | 1 | 4 | 8 | 1 | 9 | 5 | 3 | 1 |
| Phospholipases B | PLB | 1 | 1 | 1 | 1 | - | 4 | 1 | 1 | 1 | - | 1 |
| Snake venom metalloproteinases | SVM (PI) | - | 2 | - | - | - | - | 1 | - | - | - | - |
| SVM (PII) | - | 1 | - | - | - | - | 4 | 3 | - | 3 | 7 |
| SVM (PIII) | 1 | 1 | 2 | 7 | 4 | 8 | 5 | 6 | 11 | 2 | 20 |
| Snake venom serine proteinases | SVSP | 4 | 6 | 7 | 1 | 8 | 8 | 22 | 11 | 9 | 15 | 12 |
| Three-finger toxins | SFTx | - | - | - | - | 5 | 19 | - | 4 | 2 | 3 | - |
| Vascular endothelial growth factors | VEGF | 4 | 7 | 7 | - | 5 | 6 | 6 | 1 | 3 | 1 | 1 |
| Venom ficolins | Veficolins | 11 | 9 | 9 | - | 11 | - | 10 | - | 4 | 1 | - |
| Vespryn/ohanin-like proteins | - | 90 | 40 | 52 | - | 39 | 1 | 42 | 1 | 1 | - | - |
| Wapetins | - | 5 | 3 | 3 | - | 4 | - | 3 | - | 1 | 1 | - |
3FTxs belong to a superfamily of non-enzymatic proteins and are a major component in the venoms of most elapids, while they generally feature less prominently in viperid and colubrid venoms. These toxins have three β-stranded loops extending from a central core, contain four or five conserved disulfide bonds, and cause a wide range of pharmacological effects [54–56]. A prominent group of 3FTxs, α-neurotoxins, interfere with neuromuscular signal transmission of cholinergic neurons by binding to nicotinic acetylcholine receptors, causing flaccid paralysis [53, 55]. Other 3FTxs are toxic to cardiomyocytes and can lead to increased heart rate and ultimately cardiac arrest, while yet others function as calcium channel blockers or platelet aggregation inhibitors [54].

PLA2s are found in the venoms of vipers, elapids, and certain rear-fanged species [57–60] and exert a wide variety of cytotoxic, myotoxic, cardiotoxic, and neurotoxic effects [57,58,60]. Of particular interest is a catalytically inactive, myotoxic category of PLA2s stemming from a single substitution of a highly conserved amino acid residue (Asp49 to Lys49/Asn49) [57]. Both non-catalytic and enzymatic PLA2s are able to form heterodimeric complexes with other PLA2s or other toxins in certain venoms, whereby their toxicity is greatly potentiated [58]. Most snake genomes contain multiple PLA2 genes, which likely originated from repeated gene duplication events [60,61]. These paralogs have diverse pharmacological activities, which were likely acquired through neofunctionalization (i.e. recruitment of a paralog to the venom gland following gene duplication and its subsequent evolution into a toxin-coding gene) [62,63]. Pseudogenization and deletion of PLA2 genes are also frequent in snakes, making this toxin family one of the most dynamic in terms of evolutionary history [28,39,64]. The annotation of more snake genomes, and the likely consequent discovery of more PLA2 genes, might provide an improved understanding of the evolution and the mechanisms of action of these proteins (including how the phenomenon of toxin synergism has
evolved), and potentially assist in the characterization of similar evolutionary processes for other enzymes.

Another major category of enzymes found in snake venoms are SVMPs [65,66]. These proteinases are enzymes that cleave peptide bonds in other proteins, which may result in the degradation or activation of the target [66]. Zinc-dependent SVMPs are often the major venom component in vipers [67], and these toxins hydrolyze extracellular matrix components, leading to rupture of capillaries and local and systemic bleeding [59]. Other clinical manifestation induced by SVMPs include edema, inflammation, myonecrosis, and reduced muscle regeneration [67]. Additionally, these enzymatic toxins can have anticoagulant, clotting factor-activating, or platelet-aggregating effects [68,69]. SVMPs are divided into three distinct classes depending on the domains present in the mature enzymes: P-I (metalloproteinase (M) domain only), P-II (M domain and disintegrin-like domain), and P-III (M domain, disintegrin-like domain, and cysteine-rich domain) [65]. Elucidation of snake genomes could help shed light on how these enzymes evolved from the ancestral P-III type via loss of domains [70–72] and postgenomic modifications, acquiring different functions and specificities in the process [65]. A better understanding of SVMP evolution via snake genomics could also provide insight into the evolutionary process that led to the diversification of SVMPs as a whole from the ancestral A Disintegrin and Metalloprotease (ADAM) family of metalloproteinases, which play significant roles in all stages of development and survival of higher-order organisms [73].

Finally, SVSPs are typically present in the venoms of vipers [74], but can also be found in elapid venoms [75]. SVSPs contain two six-stranded β-barrels and consist of approximately 245 amino acid residues. SVSPs also have a unique extended C-terminus that forms a disulfide bridge, which contributes to structural stability [76]. These toxins can induce blood coagulation through fibrin formation, Factor V activation, prothrombin activation, actin dissolution, or platelet aggregation;
conversely, they can also act as anticoagulants via fibrinolysis, fibrinolytic enzyme activation, or protein C activation [59,77–79]. This toxin family has received increased attention with recent genome studies on *P. flavoviridis* and *B. jararaca*, where the evolutionary pathway as well as the molecular regulation of SVSP expression was systematically investigated [15,29].

In summary, snake toxin families are numerous and their pharmacological actions are complex [80]. Knowledge on the toxicity and structure of different snake toxin families is essential to further our understanding of snake venom evolution, as well as to understand venoms as drug targets for antivenom development. Much knowledge has been gained from venom proteomics and transcriptomics, and new genomics technologies now allow for the investigation of the evolutionary relationships between toxins in different families in unprecedented detail.

### 2.2 State-of-the-art in snake genomics

With the rapid development of high-throughput sequencing technology, large-scale genomic projects have generated rich sequence information data of billions of base pairs and have paved the way for a new era in the field of phylogenetics, whereby the evolutionary history of organisms can be reconstructed from genomic data. The supermatrix method is the most well-known approach for analyzing concatenation of multiple gene sequences, and using genomic data sets with improved resolution can potentially mitigate phylogenetic problems previously caused by sampling errors [81]. However, since only 21 (approximately 0.6%) out of the ca. 3,600 existing snake species have undergone WGS so far [9,15,17,18,20–28,30,82–85], snake genomics will likely develop significantly in the coming years. A complete list of currently available snake genomes is provided below in Table 2.
Table 2. Whole-genome sequencing studies on snakes, published or in progress.

| Scientific classification | Genus       | Species              | Sequencing platform | DoC | GC% | N50 Size (kb) | N50 Size (kb) | Genomic Size (Gb) | Protein encoding genes identified | Venomous | INSDC ID                  | Ref.  |
|---------------------------|-------------|----------------------|---------------------|-----|-----|--------------|--------------|-------------------|-----------------------------------|-----------|--------------------------|-------|
| **Crotalidae**            | *Bothrops*  | *B. jararaca*        | Illumina; PacBio; BAC-SeqSc 20 PL 150 IL | 20 PL 150 IL | 36.6 | 139          | 15.74         | 1.3               | -                                  | Yes       | PRJNA691605               | [29]  |
|                           | *Crotalus*  | *C. viridis*         | Illumina; PaBio      | 100 | 36.6 | 139          | 15.74         | 1.3               | -                                  | Yes       | PDHV00000000.1            | [20,21]|
|                           |             | *C. tigris*          | Illumina; PaBio      | 33 PL 190 IL | 39.9/39.8 | 207.7/20    | 2.110         | 1.6               | 18,240                            | Yes       | VORL00000000            | [28]  |
|                           |             | *C. pyrrhus*         | Illumina             | 40  | 38.5 | 5.1          | 4.1           | 1.1               | -                                  | Yes       | JPMF00000000.1            | [26]  |
|                           |             | *C. horridus*        | Illumina             | 135 | 34.3 | 23.8         | 5.8           | 1.5               | -                                  | Yes       | LVCR00000000.1            | [82]  |
| **Colubridae**            |             | *Protobothrops*      | Illumina             | 96  | 38.5 | 467          | 3.8           | 1.4               | 20,540                            | Yes       | BFFQ00000000.1            | [15]  |
|                           |             | *P. flavoviridis*    | Illumina             | 86  | 40.6 | 424          | 22            | 1.6               | 20,122                            | Yes       | BCNE00000000.2            | [24]  |
|                           |             | *Sistrurus*          | Illumina; PacBio     | -   | 1,045 | 1,045        | 1,000         | 1.6               | -                                  | Yes       | PRJNA750087               | [31]  |
|                           |             | *Viper*              | Illumina             | 121 | 41.3 | 126.6        | 11.7          | 1.5               | -                                  | Yes       | JTGP00000000.1            | [25]  |
|                           | *Deinagkistrodon* | *D. acutus*       | Illumina             | 114 | 114  | 2120         | 22.42         | 1.4               | 21,194                            | Yes       | DQ343647.1               | [92]  |
| Class       | Genus       | Species       | Method     | Length (Kb) | GC%  | DOC (Mb) | N50 (Kb) | INSDC ID       |
|-------------|-------------|---------------|------------|-------------|------|----------|-----------|----------------|
| 14          | Colubridae  | Pantherophis  | *P. guttatus* | 13          | 38.3 | 1.4      | 4.3       | JTLQ00000000.1 |
|             |             | Thermophis    | *T. baileyi* | 185         | 43.6 | 1.8      | 2.414     | QLT00000000.8 |
|             |             | Thamnophis    | *T. sirtalis* | 72          | 41.8 | 1.4      | 516       | LFLD00000000.1 |
|             |             | Thamnophis    | *T. elegans* | 62          | 41.1 | 1.6      | 100.8     | PRJNA561996   |
| 13          | Elapidae    | Ophiophagus   | *O. hannah* | 28          | 40.6 | 1.6      | 226       | AZIM00000000.1 |
|             |             | Pseudonaja    | *P. textilis* | 73          | 40.1 | 1.6      | 14.68     | ULFR00000000.1 |
|             |             | Notechis      | *N. scutatus* | 71          | 40.2 | 1.6      | 5,997     | PRJEB27871    |
|             |             | Naja          | *N. naja*    | 250         | 40.4 | 1.79     | 223.3     | SOZL00000000.1 |
|             |             | Hydrophis     | *H. curtus*  | 120         | 37.2 | 1.62     | 1,346     | PRJNA597425   |
| 23          | Pythonidae  | Python        | *P. bivittatus* | 20          | 39.7 | 1.6      | 214       | AEQU00000000.2 |
| 13          | Boa         | *B. constrictor* |                | 125         | 1.6  | 1.6      | 214       |                |

228 GC% refers to the percentage of the Guanine (G) and Cytosine (C) bases in a genome, scaffold N50 is a measure of the assembly quality (see below), DoC is a measure of the depth of coverage (see below), and INSDC ID is the NCBI gene bank accession number of the respective genome. PB stands for PacBio and IL for Illumina.
Available snake genomes differ notably in their assembly and annotation qualities, which makes evaluation of genome quality an important factor in determining the suitability of a genome for addressing a given set of questions. For instance, while estimation of nucleotide composition and genomic repeat content can be achieved from a relatively fragmented genome assembly, high-quality genome assemblies are required for analyses of multi-gene families and regulatory elements [86]. The reason for this is that the majority of the known venom gene families form tandem-arrayed “gene islands” (significantly enriched in microchromosomes, see e.g. [13]), which generally represent a challenge for performing a continuous assembly. In order to achieve the best quality of assembly of venom genes, the use of long-read technology (e.g. PacBIO HiFi or MinIon) is therefore essential (Figure 2). Genome assembly quality is assessed using statistics that measure fragmentation of the genome assembly, such as total assembly length, total contig number, contig N50, and scaffold N50. The total length of the assembly represents the total length of all the contigs that are part of the de novo assembled genome. A high total assembly length usually indicates a high-quality genome assembly. The contig N50 expresses the contiguity of the assembled genome. For instance, a contig N50 of 10 kilo bases (kb) implies that 50% of the entire genome assembly is contained in contigs that are longer than 10kb. Thus, a high contig N50 value represents a high-quality assembly without too many gaps. Currently, the contig N50 values of most published snake genomes are <25 kb; exceptions include seven species with better assembly quality, namely Thamnophis elegans (Western terrestrial garter snake; 4,620 kb) [87], Crotalus tigris (Tiger rattlesnake; 2,110 kb) [32], Naja naja (Indian cobra; 304 kb) [31], Hydrophis curtus (Shaw’s sea snake; 183 kb) [30], Bothrops jararaca (Brazilian lancehead; 163.5 kb) [29], Pseudonaja textilis (Eastern brown snake; 51 kb) [84], and Notechis scutatus (Tiger snake; 32 kb) [83].

Another important parameter is the contig L50, which represents the minimum number of contigs required to cover 50% of the total assembly length. N50 and L50 values can be computed
both at the contig and scaffold level. The most complete published snake genomes to date are those of
N. naja and C. tigris, which were assembled by combining data obtained from long-read platforms
(PacBio and Nanopore) and short-read platforms (Illumina), as well as Chicago, Hi-C, and optical
mapping in the case of N. naja [31, Fig. 2]. The resulting assemblies have a scaffold N50 reaching a
staggering 207.72 Mb (C. tigris) and 223.35 Mb (N. naja) in length, which is roughly 2.5 times greater
than the previously assembled human reference genome (87 Mb) [31, 32, 88].

Fig. 2 Schematic representation of the next-generation sequencing (NGS) pipeline for
 genomic assembly. (1) Multiple companies have marketed sequencing platforms for genomic and
 transcriptomic studies, the most commonly used being Illumina (left), PacBio (middle), and Nanopore
 (right). (2) The three platforms differ in read length and accuracy of their generated sequences.
 Whilst Illumina sequencing generally yields short reads with low error rates, Nanopore sequences are
 substantially longer (up to 2 Mb), yet subject to frequent sequencing errors. Lastly, PacBio generates
 sequences with lengths and error rates in between the two other platforms. (3) After sequencing, reads
 are computationally processed and assembled into contigs, which in turn (4) serve as the building
 blocks for scaffolds. (5) The scaffolds are then aligned and annotated to produce the complete target
genome.

In addition to measures of genome contiguity – such as N50 scores – evaluating the
representation of genes in a genome assembly via tools such as Benchmarking Universal Single-Copy
Orthologues (BUSCO) provides great insight into genome assembly and annotation completeness [89].
A recent study using 611 published eukaryotic genomes showed that assemblies with high contig and
scaffold N50 scores were shown to have high BUSCO values as well. However, the study revealed that
assemblies with poor N50 scores may also (albeit rarely) show high BUSCO scores [90]. One example of this scenario in snakes is the case of the *P. flavoviridis* genome assembly where contig N50 was 3.8 kb, but percentages of complete and partial coverages for a set of 233 core vertebrate genes were 92.7% and 97.0%, respectively [18].

Furthermore, much can be learned about the quality of a genome from its reported depth of coverage (DoC). A DoC of 10X implies that each position in the genome has been read on average 10 times from independent sequencing reads. High DoC values imply that each position (i.e. each nucleotide) can be determined with greater confidence. Consequently, the 21 snake genomes published to date or in progress can be categorized into two groups: (1) a high DoC group (>50X) comprising *B. jararaca* (Brazilian lancehead) [29], *Crotalus viridis* (Prairie rattlesnake) [20,21], *C. horridus* (Timber rattlesnake) [82], *P. flavoviridis* [15], *P. muraquenatus* (Brown-spotted pitviper) [24], *Vipera berus* (European adder) [25], *Thamnophis baileyi* (Tibetan hot-spring snake) [23], *Thamnophis sirtalis* (Common garter snake) [85], *T. elegans* (Western terrestrial garter snake), *P. textilis* [84], *C. tigris* (Tiger rattlesnake) [28], *N. scutatus* (Tiger snake) [91], *D. acutus* (ive-pacer viper) [92], *B. constrictor* (Red-tailed boa) [19], *H. curtus* (Shaw’s sea snake) [30], and *N. naja* [27]; and (2) a low DoC group (13–40X), which includes *Sistrurus catenatus* (Eastern massasauga rattlesnake), *C. pyrrhus* (Southwestern speckled rattlesnake) [26], *Pantherophis guttatus* (corn snake) [22], *O. hannah* [18], and *P. bivittatus* [9]. Unsurprisingly, the earliest published snake genomes are characterized by lower DoCs, whereas the more recently sequenced genomes benefitted from technological advancement and thus generally obtained better coverages. The best example of this is the *N. naja* genome, which reached a DoC of 250X [27] – by far the highest DoC reported for a snake genome to date (Table 2). This high DoC enabled the discovery of 43 new toxin-encoding genes, some of which are likely to be unique to *N. naja* [27].
Genome size (the total amount of DNA contained within one copy of a single complete genome [93]), number of genes, and guanine-cytosine content (GC-content, the percentage of the two nitrogenous bases in DNA [94]) vary from species to species and therefore may help elucidate phylogenetic relationships and molecular events (e.g., gene/genome duplication, pseudogenization, gene loss) in the evolution of species. Genome size can vary greatly and is typically correlated with organism size and complexity as well as with genome repeat content [93]. The reported genome sizes of snakes range from 1.3 Gb to 1.8 Gb, except for *C. pyrrhus* (1.1 Gb) and *B. jararaca* (2.1 Gb) (Table 2). This is consistent with previous findings that squamate reptiles and birds generally have smaller genomes than mammals (1-3 Gb for squamates vs 1-2 Gb for birds vs 2-6 Gb for mammals) (Table 3) [22].

Somewhat counterintuitively, genome size is not necessarily correlated with the number of genes in the genome. For example, although the *H. sapiens* genome (2.90 Gb) is roughly two times larger than the *T. sirtalis* genome (1.42 Gb), the number of genes is similar between the two (20,186 genes for *T. sirtalis* compared to 21,407 genes for *H. sapiens*) [95]. This implies higher average gene density (genes/Mbp) in *T. sirtalis* than in *H. sapiens*, which and is likely rooted in the larger percentage of repeat elements (REs) in the human genome compared to that of *T. sirtalis* (~70% and 37.12%, respectively) [96–98]. Thus, a considerably larger portion of the *H. sapiens* genome is not composed of protein-coding regions compared to than in the genome of *T. sirtalis*, which may compensate for the difference between their genome sizes. This also explains the greater gene density found in *T. sirtalis* compared to *H. sapiens*. Furthermore, even though the average gene length of *T. sirtalis* (13,384 bp) is significantly smaller than that of *H. sapiens* (23,247 bp), exon length is comparable between the two (280.12 bp vs 249.22 bp, respectively) [95].

Unlike their genome sizes and gene lengths, the genomic GC-contents for mammals, birds, and squamates are similar (~40%) (Table 3), and the GC-contents of reported snake genomes
range from 34.3% to 43.6% (Table 3). Interspecies variation in GC-content is thought to be caused by selective variation, mutation bias, and biased DNA repair-related recombination [94]. High GC-content might also be an indication of biased sequencing results [99]. It is advisable to obtain information regarding both genome size and GC-content prior to de novo assembly of a genome, as these key genomic features can guide the choice of the most appropriate assembly strategy.

Table 3. Selected genomic features compared across several vertebrate lineages [21].

| Tetrapod taxon            | Genome size (Gb) | GC-content | Transposable elements content |
|----------------------------|------------------|------------|-------------------------------|
|                            |                  |            | Range       | Mean Value |
| Mammals                    | 2.2-6.0          | ~40.9%     | 33.4%-56.4% | 44.5%      |
| Birds                      | 1.2-2.1          | ~40.2%     | 4.6%-10.1%  | 7.8%       |
| Colubroidea                | 1.5-3.0          | 39.3-47.8% | 33.0%-56.3% | 46.2%      |
| Non-colubroid snakes       | 1.7-2.1          | 38.8-43.4% | 28.7%-48.7% | 38.7%      |
| Scincioidea (skinks)       | 1.3-2.6          | 43.2-46.1% | 34.3%-44.0% | 37.6%      |

3. Understanding snake venom evolution through snake genomes

3.1 Genetic research on snake toxins

Phylogenetics is the cornerstone of our understanding of evolutionary relationships at all taxonomic levels and provides a historical basis for testing and inferring ecological and evolutionary processes [100–103]. In the past few decades, snake venom and its evolutionary origins have received considerable attention [46,104–106]. Although there is uncertainty and controversy about the origin of the venom system in squamate reptiles [29,104,107–109] a prevalent hypothesis is that, the core snake venom system evolved in the common ancestor of snakes and lizards [104].
Venom is a polygenic trait that has evolved many times in the tree of life, and it serves in both prey capture and defence against predators [105,110]. Unlike many polygenic traits [111,112], venom has a relatively direct pathway from transcription of toxin genes to translation into toxin proteins, which are then stored for use in the venom gland [46,113]. Thus, by combining venom-gland transcriptomics and venom proteomics, we can accurately map the progression from genotype to phenotype in this adaptive trait [105]. Although transcriptome data will vary depending on the population of origin, age of snake, sex and time since the last expulsion of venom [114,115] the snake was subjected to at the time of collection and/or sampling, as well as on the characteristics of the underlying genotype, transcriptomes represent a sample of the spatiotemporally-expressed genome and can be used as an entry into genome divergence analysis. Genome divergence analysis takes advantage of whole genome and/or transcriptome data to reconstruct phylogenies that chart the relationships among snakes, thus representing a precious resource for studies of snake venom evolution.

3.2 Structural characteristics of the toxin genes in snake genomes

More than 10,000 species of squamate reptiles have evolved over the last 200 million years, making this clade a major component of the vertebrate lineage [116]. The number of protein-coding genes is remarkably constant across vertebrates (including snakes), but vertebrate genomes differ considerably in size, structure, and composition [21]. An important genomic feature in this regard are transposable elements (TEs), which are self-replicating DNA sequences with the ability to insert themselves in new positions in the genome, thereby altering genome structure and gene regulation [117,118]. Having a high abundance of transposable elements could lead to a very high degree of evolvability in structural features of the genome where pseudogenization and gene duplication may occur more frequently, thus creating opportunities for neofunctionalization. As such, it is perhaps hardly surprising that TEs are consistently involved in the evolution of snake venom [17,18].
Preliminary research indicates that one of the main differences across snake genomes is the abundance and diversity of TEs, which ranges between 33.0%–56.3% in Colubroidea to 28.7%–56.3% in non-colubroid snakes [20,27,92,95]. For comparison, other reptiles, such as members of the order Scincioidea, have a lower variation in their number of transposable elements (34.3%–44.0%) (Table 3) [21,27,92]. Both abundance and diversity of TEs in snake genomes are exemplified by the genomes of *D. acutus* and *B. jararaca*. The former is made up of 13.84% long interspersed elements (LINEs, e.g. CR1, L1, and L2), 7.96% DNA transposons (e.g. hAT and TcMar elements), and 2.59% retrotransposons (e.g. Gypsy and DIRS elements) [92], whereas the latter comprises 14.6% LINEs with L2/CR1/Rex as the most abundant (8.8% of whole genome). The observed differences in the repeat content cannot be attributed only to varying sequencing technologies, as shown by the comparison of genome assembly qualities between snakes. For instance, while *B. constrictor* has a higher scaffold N50 (4505.2 kb) and less total gap length (55688.38 kb) compared to *D. acutus* (N50 2122.2 kb; gap length 82553.36 kb), the latter shows a higher total TE content (47.47 vs 39.59%) [119]. The genomes of *D. acutus* and *O. hannah* have a fairly low-divergence level (<10%) of CR1 and hAT elements from the inferred ancestral consensus sequences, while snakes belonging to more basal-branching clades (e.g. *B. constrictor* and *P. bivittatus*) have more than 20% divergence level [92]. Conversely, CR1 and hAT content is over three times higher in *D. acutus* and *O. hannah* than in *B. constrictor* and *P. bivittatus*, but the latter two species have undergone independent expansion of L2 repeat contents [92].

Another study that highlights genomic differences in TE content in snakes showed that repeat element abundance in the genomes of *D. acutus*, *T. sirtalis*, and *O. hannah* (all part of the Colubroidea clade) were characterized by a higher CR1-like and DNA transposon content compared to the genome of *P. bivittatus* [95]. Overall, repeat elements in the genomes of venomous snakes are generally more active, diverse, and dynamic compared to those of non-venomous species, indicating that different types of
transposable elements may have played multiple important roles in functional regulation of snake genes throughout evolution.

Another TE category that has attracted research attention are microsatellites (short-repeated DNA sequences). Microsatellites are so ubiquitous in certain snake species that a snake genome holds the record for containing the highest microsatellite content in any known eukaryote [21]. Bolstering this claim, a study of 11 viper species found an unprecedented average microsatellite content of 16,214 bp/Mbp [21]. In comparison, the average microsatellite density of four non-venomous snakes was roughly 55% of that amount, i.e. 8,953 bp/Mbp [21]. The same study found that the average genome density of Simple Sequence Repeat (SSR) loci (448-896 loci/Mbp) was roughly twice as large in venomous snake microsatellites as in non-venomous snakes homologs [21]. The study further found that the AATAG loci (which tend to be immediately adjacent to CR1-L3 LINEs in colubroid genomes) in venomous colubroids were increased 75-fold compared to other squamate reptiles and 71-fold compared to non-colubroid snakes [21]. Based on the significant expression of SSRs and LINE-SSR hybrid element content in venomous snakes compared to non-venomous snakes, the study also concluded that SSRs and LINE-SSR hybrid elements may have played key roles in the evolution of snake venom [21]. The dynamics and extent of the influence of SSRs and LINE-SSR on venom evolution therefore represent an intriguing venue for further research.

However, microsatellite content alone cannot explain the course of venom evolution. Indeed, another important factor is the chromosomal location of venom genes. What is known about snake chromosomes is largely based on cytogenetic experimental studies, which have revealed that the majority of snakes have 18 chromosomes (eight macrochromosomes and ten microchromosomes) [20]. It has been observed that a high proportion of venom genes are located on microchromosomes [15,21], revealing a consistent pattern of homologous chromosomal location for multiple venom gene families arranged in tandem-arrayed gene clusters. For example, 37% of all venom genes in the *C. viridis*
genome and ca. 57% (27/47 genes) of all annotated venom-related genes in the *P. flavoviridis* genome are located on microchromosomes (Fig. 3) [21,22]. This is the case for *C. tigris* as well, with all genes belonging to the major toxin family in the venom of this species (PLA$_2$s) located on microchromosome 7 [33]. Phylogenetic analysis of the three most abundant and well-characterized toxin families in *C. viridis* venom (SVMPs, SVSPs, and PLA$_2$s, all located on microchromosomes) revealed that each toxin gene family represents a distinct set of duplicated genes derived from a single ancestral homolog that produced a monophyletic cluster of venomous paraphyletic lineages [21]. Notably, microchromosomes have higher GC-content and faster recombination rates than macrochromosomes [21], as evident in the *C. viridis* genome [20]. Therefore, it appears that microchromosomes are generally enriched with venom genes, which together with their high recombination rate could explain the huge radiation and rapid evolution of venom-related genes [15]. Nonetheless, it should be noted that a substantial percentage of toxin-coding genes are found on macrochromosomes as well. This is evident in *N. naja*, where as many as 16 toxin gene families are located on macrochromosomes [32]. WGS of other venomous snake species will be essential to investigate how and to what extent chromosomal location of genes influences venom evolution.

Interestingly, the chromosome structure of *C. viridis* is comparable to that of *N. naja*. In fact, chromosome 4 of *N. naja* shares syntenic regions with *C. viridis* chromosomes 3 and 5, and chromosomes 5 and 6 of *N. naja* are syntenic with chromosome 5 of *C. viridis* [27]. This might indicate the occurrence of fusion and fission events, respectively [27]. The *N. naja* genome has also been compared to that of *O. hannah* (another elapid, and thus more closely related to *N. naja* than *C. viridis*), where 139 venom gland toxin genes from the *N. naja* genome were cross-referenced with genes in the *O. hannah* genome to find orthologs [27]. The results showed that 96 of the *N. naja* genes had counterparts in the *O. hannah* genome, while 43 did not [27]. Although some of these 43 genes may...
be unique to *N. naja*, others may simply not have been annotated in the *O. hannah* genome, possibly due to the high fragmentation of its assembly (which relied on short-reads) [14].

In the future, widespread access to different types of sequencing platforms providing researchers with both short and long reads, complementary tools for genome analysis (Hi-C and CHiCAGO), and higher quality sequence data will likely enable researchers to study snake genomes in greater detail. In turn, this will help elucidate differences and similarities between snake genomes and allow for more fine-grained studies of the structural characteristics of snake venom genes.

3.3 Molecular origin and regulation of snake venom genes

Snake venoms and their evolutionary origins have received substantial attention over the past decades, with more than 15,000 studies published on this topic [15]. Snake venoms have the dual functions of defense against predators and subduction of prey, with predation typically being the primary function [105]. This locks snakes and their prey in an evolutionary arms race, where the prey evolves biological strategies that make it resistant to toxins, and snakes are constantly pressured to optimize and adjust the composition of toxins in their venoms [105]. Indeed, dietary habits have often been indicated as a key driver of adaptive venom evolution in snakes, featuring among the main reasons behind inter- and intraspecific variation in venom composition [120].

Reports on trophic adaptations of snakes are plentiful. As an example, a study showing that venom variation in the Malayan pitviper (*Calloselasma rhodostoma*) throughout its range is significantly associated with the types of prey locally available [121]. This is also the case for the Mangrove catsnake (*Boiga dendrophila*), which was found to possess a 3FTx specific for birds and lizards (the bulk of this snake’s diet) but scarcely effective on mammals [122]. However, recent research reported that venom composition in the Mojave rattlesnake (*Crotalus scutulatus*) was associated with environmental factors (e.g. temperature, seasonality) rather than with diet [123].
suggests that a more complex scenario of factors could be affecting venom diversity than prey-related
drivers alone, as confirmed by the dynamics behind venom variation in the Northern Pacific rattlesnake
(Crotalus oreganus). In fact, the dichotomy in venom composition observed in this species is
consistently influenced not only by coevolution with its prey, but also by genetic distance and
elevation-based habitat gradients, in a pattern described as “phenotype matching” of venom
characteristics to multiple variables in the snake’s native ecosystem [124,125]. The genetic basis
underlying such complex adaptive processes could likely provide intriguing insight into the influence
of natural selection and phylogenetic relatedness on the evolution of a highly dynamic trait such as
snake venom. To this end, whole genome sequencing of snakes will likely be key to conclusively
determining the structural and evolutionary features of toxin genes and gene clusters. Analyzing such
patterns in a comparative framework would then enable researchers to identify similarities and
differences in adaptive drivers of venom evolution at all levels of snake taxonomy and phylogeny.

In recent years, venom evolution has been further explored through genome studies on
several species of venomous snakes [15,21,27,92]. One of these studies revealed that the venom gene
repertoire of D. acutus has a very different composition from those of O. hannah and the non-venomous
A. carolinensis (outgroup), B. constrictor, and P. bivittatus. These differences are exemplified both by
the absence of characteristic venom genes from the D. acutus genome relative to the O. hannah genome
and by the increased gene copy number of other venom gene families, including SVMPs, CTLPs, and
SVSPs (Table 1) [92]. Expression of most toxin-encoding genes shared by D. acutus and O. hannah
(especially older genes derived from the last common ancestor of these species) is limited to venom
glands or accessory glands [92]. Similarly, newer viper-specific toxin genes are expressed in the venom
and accessory glands of D. acutus, while equally recent elapid-specific toxin genes are expressed in
the venom and accessory glands of O. hannah [92]. Interestingly, genes closely related to the elapid-specific toxin genes expressed in the venom glands of O. hannah are expressed in the liver of D. acutus,
and genes related to viper-specific toxin genes expressed in the venom glands of *D. acutus* are
expressed in pooled organs from *O. hannah* [92].

These special expression patterns suggest that these venom genes may originate from
metabolic proteins that have undergone subfunctionalization (i.e., paralogs retaining only part of the
functional features of the original gene following duplication) or neofunctionalization, as well as that
changes in tissue-specific expression have occurred [17,92]. This is in accordance with previous
protein-based findings [126,127]. Similarly, analysis of the *O. hannah* genome demonstrated that the
regulatory components of the venomous secretion system may have evolved from the pancreas [18].
Several mechanisms likely contribute to the enhanced expression of toxin-coding genes in the venom
gland. At the chromosome level, methylation and chromatin accessibility were recently shown to play
a prominent role in gene regulation in *C. tigris*. In fact, methylation appears to be significantly more
prevalent in non-toxin and unexpressed toxin genes compared to expressed toxin counterparts in the
venom gland and pancreas of this species [33]. Furthermore, chromatin accessibility and methylation
levels is positively related with high the expression of toxin genes compared to, non-expressed
counterparts and non-toxin genes in *C. tigris*, further supporting a joint role for these two factors in
toxin gene expression [28]. Another important factor in regulation and expression of toxin genes is the
the gene regulatory network associated with them (recently termed "metavenom network"), which
comprises ~3000 genes that do not code for toxins but actively influence their expression and
postgenomic modifications (e.g. protein folding) in the venom gland as housekeeping genes [48].
Interestingly, this network presents highly conserved elements common to even distantly related
lineages such as snakes and venomous mammals; on the other hand, snakes (specifically *P. flavoviridis*
and *P. macrosquamatus*) also displayed several unique regulatory genes that were likely co-opted
together with neofunctionalized toxin genes absent in other lineages [48].
Gene duplication is thought to be one of the main mechanisms behind venom diversification [128]. The current consensus is that two rounds of whole-genome duplication (2R-WGD) occurred during the evolution of vertebrates [15, 129]. A study of the *P. flavoviridis* genome identified 18 families of venom-related genes, including both toxin and non-toxin gene copies. These include metalloproteinases (MP), serine proteases (SP), C-type lectin-like proteins (CTLP), phospholipases A2 (PLA2), three-finger toxins (3FTx), aminopeptidases (APaseN), cysteine-rich secretory proteins (CRISP), vespryns/SPa and ryanodine receptor domain proteins (Vesprrn), 5′-nucleotidases (5Nase), dipeptidyl peptidases (DDPase), hyaluronidases (Hyal), nerve growth factors or neurotrophins (NGF), vascular endothelial growth factors (VEGF), L-amino acid oxidases (LAAO), phosphodiesterases (PDE), phospholipases B (PLB), bradykinin-potentiating peptides and C-type natriuretic peptides (BNP), and glutaminyl peptide cyclotransferases (GPCase), [15]. The study suggested that 2R-WGD resulted in the creation of four paralogs from each of the 18 genes, which during the later evolution of venomous snakes, one of these four gene copies underwent neo- or subfunctionalization and evolved toxic properties, while the remaining three copies did not [15]. Both the toxin and non-toxin encoding genes subsequently underwent multiplication to different extents (Fig. 4A) [15], as is demonstrated by the multiple gene duplication events detected in the SVMP, SVSP, CTLP, PLA2, 3FTx, and CRISP gene families in *P. flavoviridis* and *N. naja* [15, 27]. However, this phenomenon was investigated to the greatest detail in rattlesnakes (*Crotalus* spp.), with comparative genomics between species revealing multiple duplication events in neurotoxic PLA2 genes as well as all SVMP classes. Chromosome mapping of the complete genomes of *C. viridis* and *C. tigris* provided further support for this scenario the occurrence of this phenomenon, highlighting similar duplication events for both gene families as well as SVSP genes (all of which are arranged in tandem-array single clusters) [18, 26].

Molecular phylogenetic analysis of *P. flavoviridis* shows that all toxin genes of a given gene family in this species are homologous to the same toxin gene families found in vipers and elapids,
such as *P. mucrosquamatus* (Brown-spotted pitviper) and *O. hannah* [15]. The notion that snake toxin genes massively expanded through gene duplication events and underwent sub- and/or neofunctionalization is also supported by other studies [18,27,109]. For example, the *N. naja* genome assembly contributes to our understanding of the origin of multiple unlinked venom gene clusters and provides new and conclusive evidence that each toxin family stems from a unique set of tandem duplicate genes [27].

**Fig 3.** Venom-related gene families in the *P. flavoviridis* genome. (A) Deduced evolutionary history of venom-related gene families through two rounds of whole-genome duplication (2R-WGD). An original set of 18 genes (shown in the top box) became 72 (four copies each). Then, a single copy of each family was likely co-opted to develop toxic functions, resulting in one snake venom (SV) copy (shown in a pale red box in the right column) and three non-venom (NV) paralogs (shown in the see-through box to the left). (B) Tandem duplications of SVMP genes. (C) Tandem duplications of SVSP genes. (D) Tandem duplications of CTLP genes. Based on Fig. 2 and Fig. S8 from [15].

While duplication either before or after gene recruitment to the venom gland is an established driving force of venom evolution in snakes, loss of genetic material has been no less pivotal in facilitating diversification of toxin families in certain venomous snake clades. For instance, the interplay between gene duplication and deletion (of entire genes as well as intragenic regions) is remarkable in rattlesnakes (*Crotalus* spp.). These pitvipers present signs of multiple independent losses of ancestral genes coding for SVMPs and neurotoxic PLAs – both of which had previously experienced a rampant expansion via repeated duplication episodes – across their phylogenetic tree [39,71]. Intriguingly, different genes underwent deletion among and even within species, such as observed in the Western diamondback rattlesnake (*C. atrox*), the Mojave rattlesnake (*C. scutulatus*),...
and the Southern Pacific rattlesnake (*C. helleri*) [39,64,71]. This resulted in great haplotype disparity and differential expression of toxin-encoding genes not only between species, but across conspecific individuals as well. Whole genome sequencing of *C. tigris* further corroborated this pattern, as this species is known for its remarkably simple venom composition largely based on neurotoxic PLA₂ isoforms [130]. However, the *C. tigris* genome revealed a deletion of three PLA₂ genes on microchromosome 7 and of ten SVMP genes on microchromosome 1 compared to homologous regions in *C. viridis*, indicating that even such a simple venom phenotype is the result of extensive genomic modifications over evolutionary time [28]. This pattern is not limited to rattlesnakes. For instance, the *Bothrops jararaca* genome also displays a great expansion of SVMP genes via duplication upon recruitment in the venom gland, followed by two deletions in the exon 14 region of PII-SVMP genes causing loss of the Cys-rich domain found in PIII-SVMPs [29]. This observation sheds further light on the genomic processes responsible for evolution and differentiation via domain loss in SVMPs, which has occurred in other viper lineages as well [46].

### 3.4. Adaptive and neutral evolution in snake venom

Determining and unraveling the driving factors behind the dynamic evolutionary processes in snake venom gene families has garnered the interest of scientists for decades – a quest that could only benefit from rising efforts in WGS of venomous snakes. Positive selection appears to be the force behind the evolution of genes involved in predator-prey arms races [131], and it seems to be pervasive across most toxin-related gene families in snakes. Positive selection leaves a well-defined pattern in the genome, with the accumulation of non-synonymous, amino-acid replacing nucleotide substitutions (denoted by KA), over synonymous substitutions (KS) in the gene [132]. In *P. flavoviridis*, the KA/KS ratios of the four main toxin gene families were consistently higher than 1 and/or higher than those
reported for non-venom genes (SVMPs: $1.047 \pm 0.438$, SVSPs: $1.253 \pm 0.090$, CTLPs: $0.871 \pm 0.071$, PLA$s: 1.093 \pm 0.062$) [15], suggesting positive selection behind the accelerated evolution of the major
toxin gene families in this species. Interestingly, *P. flavoviridis* also exhibited KA/KS > 1 in the 3FTx
and CRISP gene families, which therefore also displayed a tendency towards accelerated evolution
despite being present in far fewer copies [15]. Similarly, a high KA/KS ratio (2.034 $\pm$ 0.818) was
observed for the 3FTx gene family in the *N. naja* genome, again pointing towards rapid differentiation
and functional diversification for these genes [27]. Conversely, when KA/KS < 1 is indicative of either
neutral selection (random substitutions that confer neither evolutionary advantages nor disadvantages)
or purifying selection (i.e. removal of mutations that usually tend to be deleterious as they appear in
conserved areas of the gene). In the *P. flavoviridis* genome study, all non-dominant toxin gene families
had a KA/KS < 1 (Mean $\pm$ SE = 0.512 $\pm$ 0.018), indicating a more neutral nucleotide substitution and
the maintenance of similarity between gene copies [15]. On the other hand, when examining sequence
divergence using venom gland transcriptomes in sidewinder rattlesnakes (*Crotalus cerastes*),
examining sequence divergence using venom gland transcriptomes data showed evidence of stabilizing
selection being stabilized, which supports the maintenance of a generalist phenotype is favored
[133]. It must, however, be noted that despite various methods available for studying selection (see
[134]), relatively few have been applied in for the investigation of selection in venom and only in a
small number of species [15,27,133,135]. Therefore, additional studies are required before general
conclusions can be drawn.

New -omics tools and methods are rapidly advancing our knowledge of the mechanisms
behind venom evolution [136]. In particular, WGS has introduced advantages to snake venom research,
as WGS data can be used to identify structural variants, including inversions (Fig. 3A-B), insertions
(Fig. 3C), deletions, tandem repeats (Fig. 3A-C), transposable elements (TEs), and other repeat content
[21,137]. An increasing number of studies report venom variation at different levels, such as
ontogenetic, within-species, and between-species [46,132,138–141]. Once the reference genome of a species is available, population genomics can contribute to the identification of such intra- or interspecific variation. This further enhances the study of venom regulation, helping understand the evolution of complex regulatory networks [28]. Although it is generally acknowledged that positive selection appears to be the main driver behind venom evolution, genomic tools allow zooming in on specific venom-related genes to infer the role of neutral evolutionary processes, i.e. genetic drift or random changes in allele frequencies [142]. Genetic drift contributes to the accumulation of random neutral variation, which serves as the basis for natural selection to act upon in response to new evolutionary pressures [143]. Although most research to date has focused on the adaptive processes explaining venom evolution, recent studies have started assessing the role of such neutral forces in shaping venom characteristics. For example, genetic drift was identified as a prominent factor behind sequence divergence in venom genes in P. macrosquamatus, where dominant toxin-encoding genes displayed relaxed selective constraints for deleterious mutations despite statistically significant rates of positive selection [24]. Furthermore, it has been shown that variation in expression of the myotoxin, crotamine, in the Eastern diamondback rattlesnake (Crotalus adamanteus) and the South American rattlesnake (Crotalus durissus) is significantly more correlated with differences in number of duplication-derived gene copies between populations than with adaptive divergence in the sequences themselves [135,144].

**Fig 4.** Syntenic comparison of toxin gene clusters. Comparison showing the 3FTx, CRISP, and SVMP genes in N. naja, and C. viridis genomes. Orthologous gene pairs are indicated by the line linked across the corresponding genomic regions. Based on Fig. 4 and Extended Fig. 4 from [27].

The strength at which genetic drift acts on the genome is inversely proportional to effective population size (N_e, namely the number of reproductive individuals that actually produce
offspring) [143]. \( N_e \) greatly contributes to sequence variation, as the fate of a favourable mutation spreading is controlled by \( N_e \) and the strength of selection [145,146]. A prime example of this pattern in snake venom evolution is presented by the Eastern massasauga rattlesnake (\textit{Sistrurus catenatus}), a threatened species whose range consists of several scattered populations largely isolated from each other. Although the influence of genetic drift on venom evolution in this species is currently weak, it is likely to increase dramatically over time once the impact of drift is augmented due to the low \( N_e \) found in most populations [147]. Thus, complete genomes obtained through WGS together with cDNA libraries can expand our knowledge of the effects of selection on venom genes, with great potential to either corroborate or challenge the current positive selection-centered view of snake venom evolution.

4. Conclusions and perspectives

WGS is a revolutionary advance in genetic research that has only recently been applied to the fields of herpetology and toxinology. Nonetheless, sequencing of complete snake genomes has already shed light on the evolutionary history of toxin-encoding genes as well as their expression patterns in the venom gland. In the future, WGS may be harnessed to obtain a better understanding of the molecular mechanisms involved in snake evolution [6,104], find new bioactive molecules with potential clinical applications, and provide valuable information for antivenom development [35]. As only 21 complete snake genomes are currently available, there is ample opportunity for genomic research on the remaining thousands of snake species, including medically relevant venomous representatives. With the increasing power of sequencing technologies, the field of snake genomics is indeed likely to expand significantly in the years to come, with multiple complete genomes already in the process of being sequenced or published. However, this will not come without challenges, as the interplay of dietary and environmental factors that has fueled venom diversification via gene duplication, recruitment, and
neofunctionalization events makes it difficult to assemble whole venomous snake genomes. Another factor adding to the complexity of de novo genome assembly is the high content of repeat sequences in snake genomes. Some of these challenges might be adequately addressed by utilizing third-generation sequencing technology. As the costs and error rates of this and other approaches decrease, they are certain to be used more widely in snake genome research. In turn, the assembly of more venomous snake genomes will allow us to explore adaptation and venom evolution at all phylogenetic levels, bringing a new perspective to the study of snake genomes and venoms.

Data availability
Not applicable.

Competing interests
WR, WZ, and SL are employees at the BGI.

Funding
This research was supported by the Beijing Genomics Institute and the Technical University of Denmark. MEA is funded by the Independent Research Fund Denmark (7027-00147B). CK is funded by Innovation Fund Denmark (9065-00007B). TPJ is funded under Marie Sklodowska-Curie grant agreement no. 713683 (COFUNDfellowsDTU).

Authors' contributions
AHL and SL conceived the project. AHL, WR, KK, TPJ, CK, CTW, WZ, SG, LS, MMD, BJM and MEA structured the draft and provided final editing. AHL, KK, TPJ, and WR coordinated and drafted
the manuscript and implemented comments provided by all authors. All authors contributed critically

to the scientific content. All authors read and approved the final manuscript.

References

1. Burbrink FT, Pyron RA. The Taming of the Skew: Estimating Proper Confidence Intervals for Divergence Dates. Collins T, editor. Systematic Biology. 2008; doi: 10.1080/10635150802040605.

2. Wallach V, Williams KL, Boundy J. Snakes of the World: A Catalogue of Living and Extinct Species. 1st edition. Boca Raton: CRC Press;

3. Da Silva FO, Fabre A-C, Savriama Y, Ollonen J, Mahlow K, Herrel A, et al.. The ecological origins of snakes as revealed by skull evolution. Nat Commun. 2018; doi: 10.1038/s41467-017-02788-3.

4. Pyron RA, Burbrink FT, Wiens JJ. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC Evolutionary Biology. 2013; doi: 10.1186/1471-2148-13-93.

5. Wiens JJ, Hutter CR, Mulcahy DG, Noonan BP, Townsend TM, Sites JW, et al.. Resolving the phylogeny of lizards and snakes (Squamata) with extensive sampling of genes and species. Biol Lett. 2012; doi: 10.1098/rsbl.2012.0703.

6. Vonk FJ, Admiraal JF, Jackson K, Reshef R, de Bakker MAG, Vanderschoot K, et al.. Evolutionary origin and development of snake fangs. Nature. 2008; doi: 10.1038/nature07178.

7. Saviola AJ, Chiszar D, Busch C, Mackessy SP. Molecular basis for prey relocation in vipers. BMC Biology. 2013;

8. Gracheva EO, Ingolia NT, Kelly YM, Cordero-Morales JF, Hollopeter G, Chesler AT, et al.. Molecular basis of infrared detection by snakes. Nature. 2010; doi: 10.1038/nature08943.

9. Castoe TA, de Koning APJ, Hall KT, Card DC, Schield DR, Fujita MK, et al.. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. Proceedings of the National Academy of Sciences. 2013; doi: 10.1073/pnas.1314475110.

10. Greene HW, Fogden M, Fogden P. Snakes: The Evolution of Mystery in Nature. First edition. Oxford: University of California Press;

11. Cohn MJ, Tickle C. Developmental basis of limblessness and axial patterning in snakes. Nature. 1999; doi: 10.1038/20944.

12. Di-Poi N, Montoya-Burgos JJ, Miller H, Pourquié O, Milinkovitch MC, Duboule D. Changes in Hox genes’ structure and function during the evolution of the squamate body plan. Nature. 2010; doi: 10.1038/nature08789.

13. Guerreiro I, Nunes A, Woltering JM, Casaca A, Novoa A, Vinagre T, et al.. Role of a polymorphism in a Hox/Pax-responsive enhancer in the evolution of the vertebrate spine. Proceedings of the National Academy of Sciences. 2013; doi: 10.1073/pnas.1300592110.

14. Vicoso B, Emerson JJ, Zektser Y, Mahajan S, Bachtrog D. Comparative Sex Chromosome Genomics in Snakes: Differentiation, Evolutionary Strata, and Lack of Global Dosage Compensation. PLOS Biology. 2013; doi: 10.1371/journal.pbio.1001643.

15. Shibata H, Chijiwa T, Oda-Ueda N, Nakamura H, Yamaguchi K, Hattori S, et al.. The habu genome reveals accelerated evolution of venom protein genes. Scientific Reports. 2018; doi: 10.1038/s41598-018-28749-4.

16. Uetz, P., Freed, P, Hošek, J: The Reptile Database. http://www.reptile-database.org/ Accessed 2020 Jun 8.
17. Kerkkamp H, Kini R, Pospelov A, Vonk F, Henkel C, Richardson M. Snake Genome Sequencing: Results and Future Prospects. *Toxins*. 2016; doi: 10.3390/toxins8120360.

18. Vonk FJ, Casewell NR, Henkel CV, Heimberg AM, Jansen HJ, Mc Cleary RJR, et al.. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proceedings of the National Academy of Sciences*. 2013; doi: 10.1073/pnas.1314702110.

19. Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Barol I, et al.. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience*. 2013; doi: 10.1186/2047-217X-2-10.

20. Schield DR, Card DC, Hales NR, Perry BW, Pasquesi GM, Blackmon H, et al.. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proceedings of the National Academy of Sciences*. 2013; doi: 10.1073/pnas.1314702110.

21. Pasquesi GM, Adams RH, Card DC, Schield DR, Corbin AB, Perry BW, et al.. Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nat Commun*. 2018; doi: 10.1038/s41467-018-05279-1.

22. Ullate-Agote A, Milinkovitch MC, Tzika AC. The genome sequence of the corn snake (Pantherophis guttatus), a valuable resource for EvoDevo studies in squamates. *The International Journal of Developmental Biology*. 2014; doi: 10.1387/ijdb.150060at.

23. Li J-T, Gao Y-D, Xie L, Deng C, Shi P, Guan M-L, et al.. Comparative genomic investigation of high-elevation adaptation in ectothermic snakes. *Genome Res*. 2019; doi: 10.1101/gr.240952.118.

24. Aird SD, Arora J, Barua A, Qiu L, Terada K, Mikheyev AS. Population Genomic Analysis of a Pitviper Reveals Microevolutionary Forces Underlying Venom Chemistry. *Genome Biol Evol*. 2017; doi: 10.1093/gbe/evx199.

25. : Vipera berus berus isolate:VBER.BE-female (ID 170536) - BioProject - NCBI. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA170536 Accessed 2019 Nov 27.

26. Gilbert C, Meik JM, Dashevsky D, Card DC, Castoe TA, Schaack S. Endogenous hepadnaviruses, bornaviruses and circoviruses in snakes. *Proc Biol Sci*. 2014; doi: 10.1098/rspb.2014.1122.

27. Suryamohan K, Krishnankutty SP, Guillory J, Jevit M, Schröder MS, Wu M, et al.. The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. *Nature Genetics*. Nature Publishing Group; 2020; doi: 10.1038/s41588-019-0559-8.

28. Margres MJ, Rautsaw RM, Strickland JL, Mason AJ, Schramer TD, Hofmann EP, et al.. The Tiger Rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. *PNAS*. National Academy of Sciences; 2021; doi: 10.1073/pnas.2014634118.

29. Almeida DD, Viala VL, Nachitgall PG, Broe M, Gibbs HL, Serrano SM de T, et al.. Tracking the recruitment and evolution of snake toxins using the evolutionary context provided by the Bothrops jararaca genome. *PNAS*. National Academy of Sciences; 2021; doi: 10.1073/pnas.2015159118.

30. Peng C, Ren J-L, Deng C, Jiang D, Wang J, Qu J, et al.. The genome of Shaw’s sea snake (Hydrophis curtus) reveals secondary adaptation to its marine environment. *Mol Biol Evol*. doi: 10.1093/molbev/msaa043.

31. Ochoa A, Gibbs HL. Genomic signatures of inbreeding and mutation load in a threatened rattlesnake. *Molecular Ecology*. doi: 10.1111/mec.16147.

32. Vonk FJ, Jackson K, Doley R, Madaras F, Mirtschin PJ, 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; doi: 10.1002/bies.201000117.

33. Windley MJ, Herzig V, Dziemborowicz SA, Hardy MC, King GF, Nicholson GM. Spider-Venom Peptides as Bioinsecticides. *Toxins*. 2012; doi: 10.3390/toxins4030191.
36. Hucho F. Toxins as Tools in Neurochemistry. Angewandte Chemie International Edition in English. 1995; doi: 10.1002/anie.199500391.

35. Laustsen AH. Guiding recombinant antivenom development by omics technologies. New Biotechnology. 2018; doi: 10.1016/j.nbt.2017.05.005.

36. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 2004; doi: 10.1093/bioinformatics/bth315.

37. Liu X, Zheng Q, Vrettos N, Maragkakis M, Alexiou P, Gregory BD, et al.. A MicroRNA Precursor Surveillance System in Quality Control of MicroRNA Synthesis. Mol Cell. 2014; doi: 10.1016/j.molcel.2014.07.017.

38. Collins JE, White S, Searle SMJ, Stemple DL. Incorporating RNA-seq data into the Zebrafish Ensembl Gene Build. Genome Res. 2012; doi: 10.1101/gr.137901.112.

39. Dowell NL, Giorgianni MW, Kassner VA, Selegue JE, Sanchez EE, Carroll SB. The Deep Origin and Recent Loss of Venom Toxin Genes in Rattlesnakes. Current Biology. 2016; doi: 10.1016/j.cub.2016.07.038.

40. Tekaia F. Inferring Orthologs: Open Questions and Perspectives. Genomics Insights. 2016; doi: 10.4137/GEI.S37925.

41. Viala VL, Hildebrand D, Trusch M, Fucase TM, Sciani JM, Pimenta DC, et al.. Venomics of the Australian eastern brown snake (Pseudonaja textilis): Detection of new venom proteins and splicing variants. Toxicon. 2015; doi: 10.1016/j.toxicon.2015.06.005.

42. Ogawa T, Oda-Ueda N, Hisata K, Nakamura H, Chijiwa T, Hattori S, et al.. Alternative mRNA Splicing in Three Venom Families Underlying a Possible Production of Divergent Venom Proteins of the Habu Snake, Protobothrops flavoviridis. Toxins (Basel). 2019; doi: 10.3390/toxins1100581.

43. Siigur E, Aaspõllu A, Siigur J. Sequence diversity of Vipera lebetina snake venom gland serine proteinase homologs – result of alternative-splicing or genome alteration. Gene. 2001; doi: 10.1016/S0378-1119(00)00571-0.

44. Sunagar K, Khochare S, Senji Laxme RR, Attarde S, Dam P, Suranse V, et al.. A Wolf in Another Wolf's Clothing: Post-Genomic Regulation Dictates Venom Profiles of Medically-Important Cryptic Kraits in India. Toxins. Multidisciplinary Digital Publishing Institute; 2021; doi: 10.3390/toxins13010069.

45. Rokyta DR, Margres MJ, Calvin K. Post-transcriptional Mechanisms Contribute Little to Phenotypic Variation in Snake Venoms. G3 (Bethesda). 2015; doi: 10.1534/g3.115.020578.

46. Casewell NR, Wagstaff SC, Wuster W, Cook DAN, Bolton FMS, King SI, et al.. Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. Proceedings of the National Academy of Sciences. 2014; doi: 10.1073/pnas.1405484111.

47. Barua A, Mikheyev AS. An ancient, conserved gene regulatory network led to the rise of oral venom systems. Proc Natl Acad Sci U S A. 2021; doi: 10.1073/pnas.2021311118.

48. Elshure RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al.. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLOS ONE. Public Library of Science; 2011; doi: 10.1371/journal.pone.0019379.

49. Jones MR, Good JM. TARGETED CAPTURE IN EVOLUTIONARY AND ECOLOGICAL GENOMICS. Mol Ecol. 2016; doi: 10.1111/mec.13304.

50. Holding ML, Sovic MG, Colston TJ, Gibbs HL. The scales of coevolution: comparative phylogeography and genetic demography of a locally adapted venomous predator and its prey. Biological Journal of the Linnean Society. 2021; doi: 10.1093/biolinnean/blaa192.

51. Ellegren H. Genome sequencing and population genomics in non-model organisms. Trends in Ecology & Evolution. Elsevier; 2014; doi: 10.1016/j.tree.2013.09.008.
52. Laustsen A, Engmark M, Milbo C, Johannesen J, Lomonte B, Gutiérrez J, et al.. From Fangs to Pharmacology: The Future of Snakebite Envenoming Therapy. Current Pharmaceutical Design. 2016; doi: 10.2174/1381612822666160623073438.

53. Tasoulis T, Isbister GK. A Review and Database of Snake Venom Proteomes. Toxins. 2017; doi: 10.3390/toxins9090290.

54. Kini RM, Doley R. Structure, function and evolution of three-finger toxins: mini proteins with multiple targets. Toxicon. 2010; doi: 10.1016/j.toxicon.2010.07.010.

55. Fry BG, Wüster W, Kini RM, Brusic V, Khan A, Venkataraman D, et al.. Molecular Evolution and Phylogeny of Elapid Snake Venom Three-Finger Toxins. J Mol Evol. 2003; doi: 10.1007/s00239-003-2461-2.

56. Nirthanan S. Snake three-finger α-neurotoxins and nicotinic acetylcholine receptors: molecules, mechanisms and medicine. Biochemical Pharmacology. 2020; doi: 10.1016/j.bcp.2020.114168.

57. Gasanov SE, Dagda RK, Rael ED. Snake Venom Cytotoxins, Phospholipase A2s, and Zn2+-dependent Metalloproteinases: Mechanisms of Action and Pharmacological Relevance. J Clin Toxicol. 4:10001812014;

58. Lomonte B, Gutiérrez JM. Phospholipases A2 from viperidae snake venoms: how do they induce skeletal muscle damage? Acta Chim Slov. 58:647–592011;

59. Gutiérrez JM, Calvete JJ, Habib AG, Harrison RA, Williams DJ, Warrell DA. Snakebite envenoming. Nature Reviews Disease Primers. 2017; doi: 10.1038/nrdp.2017.63.

60. Tsai I-H. Snake Venom Phospholipase A2: Evolution and Diversity. In: Gopalakrishnakone P, Calvete JJ, editors. Venom Genomics and Proteomics. Dordrecht: Springer Netherlands; p. 291–306.

61. Kordiš D. Evolution of phospholipase A2 toxins in venomous animals. Acta Chim Slov. 58:638–462011;

62. Manjunatha Kini R. Excitement ahead: structure, function and mechanism of snake venom phospholipase A2 enzymes. Toxicon. 2003; doi: 10.1016/j.toxicon.2003.11.002.

63. Ohno M, Chijiwa T, Oda-Ueda N, Ogawa T, Hattori S. Molecular evolution of myotoxic phospholipases A2 from snake venom. Toxicon. 2003; doi: 10.1016/j.toxicon.2003.11.003.

64. Dowell NL, Giorgianni MW, Griffin S, Kassner VA, Selegue JE, Sanchez EE, et al.. Extremely Divergent Haplotypes in Two Toxin Gene Complexes Encode Alternative Venom Types within Rattlesnake Species. Current Biology. 2018; doi: 10.1016/j.cub.2018.02.031.

65. Markland FS, Swenson S. Snake venom metalloproteinases. Toxicon. 2013; doi: 10.1016/j.toxicon.2012.09.004.

66. Gutiérrez JM, Rucavado A. Snake venom metalloproteinases: their role in the pathogenesis of local tissue damage. Biochimie. 82:841–502000;

67. Gutiérrez JM, Escalante T, Rucavado A, Herrera C, Fox JW. A Comprehensive View of the Structural and Functional Alterations of Extracellular Matrix by Snake Venom Metalloproteinases (SVMPS): Novel Perspectives on the Pathophysiology of Envenoming. Toxins. 2016; doi: 10.3390/toxins8100304.

68. Sanchez EF, Flores-Ortiz RJ, Alvarenga VG, Eble JA. Direct Fibrinolytic Snake Venom Metalloproteinases Affecting Hemoestasis: Structural, Biochemical Features and Therapeutic Potential. Toxins (Basel). 2017; doi: 10.3390/toxins9120392.

69. Kini R, Koh C. Metalloproteinases Affecting Blood Coagulation, Fibrinolysis and Platelet Aggregation from Snake Venoms: Definition and Nomenclature of Interaction Sites. Toxins. 2016; doi: 10.3390/toxins8100284.

70. Casewell NR, Wagstaff SC, Harrison RA, Renjifo C, Wüster W. Domain Loss Facilitates Accelerated Evolution and Neofunctionalization of Duplicate Snake Venom Metalloproteinase Toxin Genes. Mol Biol Evol. Oxford Academic; 2011; doi: 10.1093/molbev/msr091.
71. Giorgianni MW, Dowell NL, Griffin S, Kassner VA, Selegue JE, Carroll SB. The origin and diversification of a novel protein family in venomous snakes. *Proc Natl Acad Sci USA*. 2020; doi: 10.1073/pnas.1920011117.

72. Brust A, Sunagar K, Undheim EAB, Vetter I, Yang DC, Casewell NR, et al. Differential Evolution and Neofunctionalization of Snake Venom Metalloprotease Domains *. Molecular & Cellular Proteomics*. Elsevier; 2013; doi: 10.1074/mcp.M112.023135.

73. Moura-da-Silva A, Almeida M, Fortes-Junior J, Nicolau C, Gomes-Neto F, Valente R. Processing of Snake Venom Metalloproteinases: Generation of Toxin Diversity and Enzyme Inactivation. *Toxins*. 2016; doi: 10.3390/toxins8060183.

74. Serrano SMT. The long road of research on snake venom serine proteinases. *Toxicon*. 2013; doi: 10.1016/j.toxicon.2012.09.003.

75. Kunalan S, Othman I, Syed Hassan S, Hodgson WC. Proteomic Characterization of Two Medically Important Malaysian Snake Venoms, Calloselasma rhodostoma (Malayan Pit Viper) and Ophiophagus hannah (King Cobra). *Toxins*. 2018; doi: 10.3390/toxins10110434.

76. Kang TS, Georgieva D, Genov N, Murakami MT, Sinha M, Kumar RP, et al. Enzymatic toxins from snake venom: structural characterization and mechanism of catalysis. *The FEBS Journal*. 2011; doi: 10.1111/j.1742-4658.2011.08115.x.

77. White J. Snake venoms and coagulopathy. *Toxicon*. 2005; doi: 10.1016/j.toxicon.2005.02.030.

78. Kini RM. The intriguing world of prothrombin activators from snake venom. *Toxicon*. 2005; doi: 10.1016/j.toxicon.2005.02.019.

79. Markland FS. Snake venoms and the hemostatic system. *Toxicon*. 1998; doi: 10.1016/S0041-0101(98)00126-3.

80. Meenakshisundaram R, Sweni S, Thirumalai-kolundusubramanian P. Hypothesis of snake and insect venoms against Human Immunodeficiency Virus: a review. *AIDS Res Ther*. 2009; doi: 10.1186/1742-6405-6-25.

81. Delsuc F, Brinkmann H, Philippe H. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet*. 2005; doi: 10.1038/nrg1603.

82. Hall JB, Cobb VA, Cahoon AB. The complete mitochondrial DNA sequence of *Crotalus horridus* (timber rattlesnake). *Mitochondrial DNA*. 2013; doi: 10.3109/19401736.2012.722999.

83. St Pierre L, Masci PP, Filipovitch I, Sorokina N, Marsh N, Miller DJ, et al. Comparative analysis of prothrombin activators from the venom of Australian elapids. *Mol Biol Evol*. 2005; doi: 10.1093/molbev/msi181.

84. Earl STH, Birrell GW, Wallis TP, St Pierre LD, Masci PP, de Jersey J, et al. Post-translational modification accounts for the presence of varied forms of nerve growth factor in Australian elapid snake venoms. *Toxicon*. 2006; doi: 10.1002/tox.200600263.

85. McGlothlin JW, Chuckalovcak JP, Janes DE, Edwards SV, Feldman CR, Brodie ED, et al. Parallel Evolution of Tetrodotoxin Resistance in Three Voltage-Gated Sodium Channel Genes in the Garter Snake Thamnophis sirtalis. *Mol Biol Evol*. 2014; doi: 10.1093/molbev/msu237.

86. Schield DR, Perry BW, Pasquesi GIM, Orton RW, Nikolakis ZL, Westfall AK, et al. Applications of Genomics and Related Technologies for Studying Reptile Venoms. *Handbook of Venoms and Toxins of Reptiles*. 2nd ed. CRC Press;

87. Thamnophis elegans isolate rThaEle1, whole genome shotgun sequencing project.

88. . Finishing the euchromatic sequence of the human genome. *Nature*. 2004; doi: 10.1038/nature03001.

89. Seppey M, Mani M, Zdobnov EM. BUSCO: Assessing Genome Assembly and Annotation Completeness. In: Kollmar M, editor. *Gene Prediction: Methods and Protocols*. New York, NY: Springer;
90. Jauhal AA, Newcomb RD. Assessing genome assembly quality prior to downstream analysis: N50 versus BUSCO. *Molecular Ecology Resources*. 2021; doi: 10.1111/1755-0998.13364.

91. Edwards R, Amos T, Tang J, Cawood B, Rispin S, Tuipulotu DE, et al.. Pseudodiploid pseudo-long-read whole genome sequencing and assembly of *Pseudonaja textilis* (eastern brown snake) and *Notechis scutatus* (mainland tiger snake). *F1000Research*. 2018; doi: 10.7490/f1000research.1115550.1.

92. Yin W, Wang Z, Li Q, Lian J, Zhou Y, Lu B, et al.. Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. *Nat Commun*. 2016; doi: 10.1038/ncomms13107.

93. Le TS, Yang F-J, Lo Y-H, Chang TC, Hsu J-C, Kao C-Y, et al.. Non-Mendelian assortment of homologous autosomes of different sizes in males is the ancestral state in the Caenorhabditis lineage. *Scientific Reports*. Nature Publishing Group; 2017; doi: 10.1038/s41598-017-13215-4.

94. Birdsell JA. Integrating Genomics, Bioinformatics, and Classical Genetics to Study the Effects of Recombination on Genome Evolution. *Mol Biol Evol*. 2002; doi: 10.1093/oxfordjournals.molbev.a004176.

95. Perry BW, Card DC, McGlothlin JW, Pasquesi GIM, Adams RH, Schield DR, et al.. Molecular Adaptations for Sensing and Securing Prey and Insight into Amniote Genome Diversity from the Garter Snake Genome. *Genome Biol Evol*. 2018; doi: 10.1093/gbe/evy157.

96. Perry BW, Card DC, McGlothlin JW, Pasquesi GIM, Adams RH, Schield DR, et al.. Molecular Adaptations for Sensing and Securing Prey and Insight into Amniote Genome Diversity from the Garter Snake Genome. *Genome Biol Evol*. 2018; doi: 10.1093/gbe/evy157.

97. de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PloS Genet*. 2011; doi: 10.1371/journal.pgen.1002384.

98. Liehr T. Repetitive Elements in Humans. *International Journal of Molecular Sciences*. Multidisciplinary Digital Publishing Institute; 2021; doi: 10.3390/ijms22042072.

99. Benjamini Y, Speed TP. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res*. 2012; doi: 10.1093/nar/gks001.

100. Figueroa A, McKevy AD, Grismer LL, Bell CD, Lailvaux SP. A Species-Level Phylogeny of Extant Snakes with Description of a New Colubrid Subfamily and Genus. *PLOS ONE*. 2016; doi: 10.1371/journal.pone.0161070.

101. Harvey PH, Harvey PP, Harvey R in BD of ZPH, Pagel MD, Pagel MD. The Comparative Method in Evolutionary Biology. Oxford University Press;

102. Whelan S, Liò P, Goldman N. Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends Genet*. 2001; doi: 10.1016/s0168-9525(01)02272-7.

103. Phylogenetic Methods Come of Age: Testing Hypotheses in an Evolutionary Context | Science. https://science.sciencemag.org/content/276/5310/227 Accessed 2021 Mar 21.

104. Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SFR, et al.. Early evolution of the venom system in lizards and snakes. *Nature*. 2006; doi: 10.1038/nature04328.

105. Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG. Complex cocktails: the evolutionary novelty of venoms. *Trends in Ecology & Evolution*. 2013; doi: 10.1016/j.tree.2012.10.020.

106. Fry BG, Scheib H, Weerd L van der, Young B, McNaughton J, Ramjan SFR, et al.. Evolution of an Arsenal: Structural and Functional Diversification of the Venom System in the Advanced Snakes (Caenophidia). *Molecular & Cellular Proteomics*. 2008; doi: 10.1074/mcp.M700094-MCP200.

107. Hargreaves AD, Swain MT, Hegarty MJ, Logan DW, Mulley JF. Restriction and Recruitment—Gene Duplication and the Origin and Evolution of Snake Venom Toxins. *Genome Biology and Evolution*. 2014; doi: 10.1109/gbe/evu166.
108. Reyes-Velasco J, Card DC, Andrew AL, Shaney KJ, Adams RH, Schield DR, et al.. Expression of venom gene homologs in diverse python tissues suggests a new model for the evolution of snake venom. *Mol Biol Evol.* 2015; doi: 10.1093/molbev/msu294.

109. Casewell NR, Huttley GA, Wüster W. Dynamic evolution of venom proteins in squamate reptiles. *Nature Communications,* Nature Publishing Group; 2012; doi: 10.1038/ncomms2065.

110. Gibbs HL, Mackessy SP. Functional basis of a molecular adaptation: Prey-specific toxic effects of venom from Sistrurus rattlesnakes. *Toxicon,* 2009; doi: 10.1016/j.toxicon.2009.01.034.

111. Comparative studies of gene expression and the evolution of gene regulation | Nature Reviews Genetics. https://www.nature.com/articles/nrg3229 Accessed 2021 Mar 21.

112. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, et al.. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature,* 2004; doi: 10.1038/nature02415.

113. Margres MJ, McGivern JJ, Rokyta DR. Linking the transcriptome and proteome to characterize the venom of the eastern diamondback rattlesnake (Crotalus adamanteus). *J Proteomics.* 2014; doi: 10.1016/j.jprot.2013.11.001.

114. Schwartz TS, Tae H, Yang Y, Mockaitis K, Van Hemert JL, Proulx SR, et al.. A garter snake transcriptome: pyrosequencing, de novo assembly, and sex-specific differences. *BMC Genomics.* 2010; doi: 10.1186/1471-2164-11-694.

115. Hofmann EP, Rautsaw RM, Strickland JL, Holding ML, Hogan MP, Mason AJ, et al.. Comparative venom-gland transcriptomics and venom proteomics of four Sidewinder Rattlesnake (Crotalus cerastes) lineages reveal little differential expression despite individual variation. *Sci Rep.* Nature Publishing Group; 2018; doi: 10.1038/s41598-018-33943-5.

116. Zheng Y, Wiens JJ. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution.* 2016; doi: 10.1016/j.ympev.2015.10.009.

117. Feiner N. Accumulation of transposable elements in Hox gene clusters during adaptive radiation of Anolis lizards. *Proceedings of the Royal Society B: Biological Sciences.* 2016; doi: 10.1098/rspb.2016.1555.

118. Platt RN, Vandewege MW, Ray DA. Mammalian transposable elements and their impacts on genome evolution. *Chromosome Res.* 2018; doi: 10.1007/s10577-017-9570-z.

119. Yin W, Wang Z, Li Q, Lian J, Zhou Y, Lu B, et al.. Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. *Nat Commun.* 2016; doi: 10.1038/ncomms13107.

120. Barlow A, Pook CE, Harrison RA, Wüster W. Coevolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution. *Proceedings of the Royal Society B: Biological Sciences.* 2009; doi: 10.1098/rspb.2009.0048.

121. Daltry JC, Wüster W, Thorpe RS. Diet and snake venom evolution. *Nature.* 1996; doi: 10.1038/379573a0.

122. Pawlak J, Mackessy SP, Fry BG, Bhatia M, Mouriir G, Fruchart-Gaillard C, et al.. Denmotoxin, a Three-finger Toxin from the Colubrid Snake Boiga dendrophila (Mangrove Catsnake) with Bird-specific Activity. *J Biol Chem.* American Society for Biochemistry and Molecular Biology; 2006; doi: 10.1074/jbc.M605850200.

123. Zancolli G, Calvete JJ, Cardwell MD, Greene HW, Hayes WK, Hegarty MJ; et al.. When one phenotype is not enough: divergent evolutionary trajectories govern venom variation in a widespread rattlesnake species. *Proceedings of the Royal Society B.* The Royal Society; 2019; doi: 10.1098/rspb.2018.2735.
124. Holding ML, Biardi JE, Gibbs HL. Coevolution of venom function and venom resistance in a rattlesnake predator and its squirrel prey. *Proceedings of the Royal Society B: Biological Sciences*. Royal Society; 2016; doi: 10.1098/rspb.2015.2841.

125. Holding ML, Margres MJ, Rokyta DR, Gibbs HL. Local prey community composition and genetic distance predict venom divergence among populations of the northern Pacific rattlesnake (*Crotalus oreganus*). *Journal of Evolutionary Biology*. 2018; doi: 10.1111/jeb.13347.

126. Fry BG. 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 Research*. 2005; doi: 10.1101/gr.3228405.

127. Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, et al.. The Toxicogenomic Multiverse: Convergent Recruitment of Proteins Into Animal Venoms. *Annual Review of Genomics and Human Genetics*. 2009; doi: 10.1146/annurev.genom.9.081307.164356.

128. Conant GC, Wolfe KH. Turning a hobby into a job: How duplicated genes find new functions. *Nature Reviews Genetics*. 2008; doi: 10.1038/nrg2482.

129. Holland PW, Garcia-Fernández J, Williams NA, Sidow A. Gene duplications and the origins of vertebrate development. *Dev Suppl.* :125–33 1994;

130. Calvete JJ, Pérez A, Lomonte B, Sánchez EE, Sanz L. Snake Venomics of *Crotalus tigris*: The Minimalist Toxin Arsenal of the Deadliest Nearctic Rattlesnake Venom. Evolutionary Clues for Generating a Pan-Specific Antivenom against Crotalid Type II Venoms. *J Proteome Res*. American Chemical Society; 2012; doi: 10.1021/pr201021d.

131. Charlesworth B, Charlesworth D. Elements of evolutionary genetics, 1st edition | macmillan learning for instructors.

132. Casewell NR, Jackson TNW, Laustsen AH, Sunagar K. Causes and Consequences of Snake Venom Variation. *Trends in Pharmacological Sciences*. 2020; doi: 10.1016/j.theips.2020.05.006.

133. Rautsaw RM, Hofmann EP, Margres MJ, Holding ML, Strickland JL, Mason AJ, et al.. Intraspecific sequence and gene expression variation contribute little to venom diversity in side-winder rattlesnakes (*Crotalus cerastes*). *Proceedings of the Royal Society B: Biological Sciences*. Royal Society; 2019; doi: 10.1098/rspb.2019.0810.

134. Kosiol C, Anisimova M. Selection Acting on Genomes. In: Anisimova M, editor. *Evolutionary Genomics: Statistical and Computational Methods*. New York, NY: Springer;

135. Margres MJ, Bigelow AT, Lemmon EM, Lemmon AR, Rokyta DR. Selection To Increase Expression of Venom Gene Homologs in Diverse Python Tissues Suggests a New Model for the Evolution of Snake Venom. *Molecular Biology and Evolution*. 2015; doi: 10.1093/molbev/msu294.

136. Massey DJ, Calvete JJ, Sánchez EE, Sanz L, Richards K, Curtis R, et al.. Venom variability and envenoming severity outcomes of the *Crotalus scutulatus scutulatus* (Mojave rattlesnake) from Southern Arizona. *Journal of Proteomics*. 2012; doi: 10.1016/j.jprot.2012.02.035.

137. Pia D, Sanz L, Quesada-Bernat S, Villalta M, Baal J, Chowdhury MAW, et al.. Phylovenomics of *Daboia russelli* across the Indian subcontinent. Bioactivities and comparative in vivo neutralization and in vitro third-generation antivenomics of antivenoms against venoms from India, Bangladesh and Sri Lanka. *Journal of Proteomics*. 2019; doi: 10.1016/j.jprot.2019.103443.
simus, C. tzabcan, and C. culminatus Revealed miRNA-modulated Ontogenetic Shifts. J Proteome Res. American Chemical Society; 2017; doi: 10.1021/acs.jproteome.7b00414.

141. Laxme RRS, Khochare S, Souza HF de, Ahuja B, Suranse V, Martin G, et al.. Beyond the ‘big four’: Venom profiling of the medically important yet neglected Indian snakes reveals disturbing antivenom deficiencies. PLOS Neglected Tropical Diseases. Public Library of Science; 2019; doi: 10.1371/journal.pntd.0007899.

142. Kimura M. Evolutionary Rate at the Molecular Level. Nature. 1968; doi: 10.1038/217624a0.

143. Wright S. Evolution in Mendelian Populations. Genetics. 16:97–1591931;

144. Oguiura N, Collares MA, Furtado MFD, Ferrarezzi H, Suzuki H. Intraspecific variation of the crotamine and crotasin genes in Crotalus durissus rattlesnakes. Gene. 2009; doi: 10.1016/j.gene.2009.05.015.

145. Charlesworth B. Effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 2009; doi: 10.1038/nrg2526.

146. Ludington AJ, Sanders KL. Demographic analyses of marine and terrestrial snakes (Elapidae) using whole genome sequences. Molecular Ecology. 2021; doi: https://doi.org/10.1111/mec.15726.

147. Ochoa A, Broe M, Moriarty Lemmon E, Lemmon AR, Rokyta DR, Gibbs HL. Drift, selection and adaptive variation in small populations of a threatened rattlesnake. Molecular Ecology. 2020; doi: 10.1111/mec.15517.
Figure 2

1. Sample preparation
   - Sample
   - Extract DNA

2. Sequencing preparation
   - PCR amplification
   - Sequencing amplification
   - Library construction (<1000bp)
     - Amplification
     - Bridge PCR
     - Emulsion PCR
     - DNA nanoballs
     - Library construction (>30,000bp)

3. Sequencing
   - First generation
     - Chain termination
   - Second generation
     - Sequencing by synthesis
     - Sequencing by ligation
     - Combinatorial probe anchor synthesis
   - Third generation
     - Single-molecule real-time sequencing
     - Nanopore sequencing

4. Data analysis
   - Raw data
   - Clean data
   - Assembly
   - Mapping
   - De novo
   - Genome annotation
Figure 3

A. Evolution mechanism of venom-related gene families

Ancestral NV genes
MP (5)  
SP (3)  
CTLP (3)  
MP (3)

3FTX (3)  
APase (3)  
CRISp (3)

Vespryn (3)  
5Nase (3)  
DDPase (3)  
Hyal (3)  
NGF (3)  
VEGF (3)  
LAOO (3)  
PDE (3)  
PLB (3)  
BNP (3)  
GPCase (3)

Present NV genes
MP (57)  
SP (34)  
CTLP (40)  
MP (31)

3FTX (2)  
APase (10)  
CRISp (2)

Vespryn (11)  
5Nase (10)  
DDPase (7)  
Hyal (5)  
NGF (3)  
VEGF (2)  
LAOO (2)  
PDE (2)  
PLB (4)  
BNP (1)  
GPCase (1)

Gain of toxin function

2R-WGD

Multiplication
Present SV genes
MP (11)  
SP (11)  
CTLP (10)  
MP (9)

3FTX (4)  
APase (2)  
CRISp (2)

Vespryn (1)  
5Nase (1)  
DDPase (1)  
Hyal (1)  
NGF (1)  
VEGF (1)  
LAOO (1)  
PDE (1)  
PLB (1)  
BNP (1)  
GPCase (1)

Accelerated evolution

Moderate multiplication

Ancestral SV genes
MP (3)  
SP (3)  
CTLP (3)  
MP (3)

3FTX (3)  
APase (3)  
CRISp (3)

Vespryn (3)  
5Nase (3)  
DDPase (3)  
Hyal (3)  
NGF (3)  
VEGF (3)  
LAOO (3)  
PDE (3)  
PLB (3)  
BNP (3)  
GPCase (3)

Multiplication

No multiplication

B. MP

{scaffold}_2862(70kb)  
{scaffold}_14911(51kb)  
{scaffold}_4106(46kb)  
{scaffold}_6789(84kb)

{scaffold}_7597(57kb)

D. CTLP

{scaffold}_3168(96kb)  
{scaffold}_6789(84kb)
