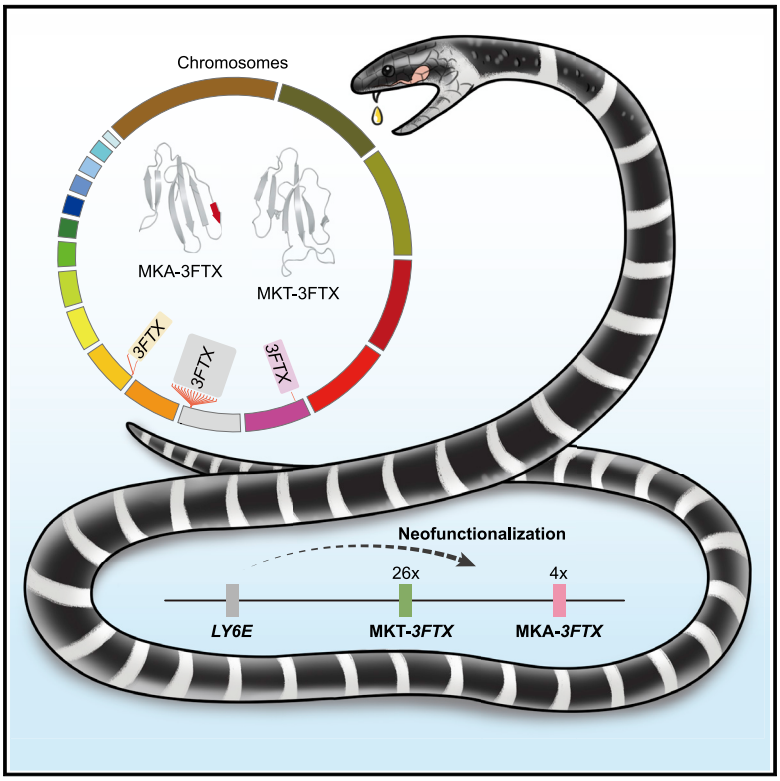
Cell Reports

Article

The structural and functional divergence of a neglected three-finger toxin subfamily in lethal elapids

Graphical abstract



Authors

Zhi-Yi Zhang, Yunyun Lv, Wei Wu, Chaochao Yan, Chen-Yang Tang, Changjun Peng, Jia-Tang Li

Correspondence

lijt@cib.ac.cn

In brief

Zhang et al. complete a high-quality reference genome of the lethal snake *Bungarus multicinctus*, through which the likely origin of three-finger toxins, as well as the structural and functional divergence of a neglected 3FTX subfamily, is revealed. They also provide a comprehensive toxin gene catalog valuable for evolutionary and medical research.

Highlights

- *De novo* chromosome-level assembly of a *Bungarus multicinctus* genome
- A comprehensive *B. multicinctus* toxin gene catalog for antivenom improvement
- Decoding the likely evolutionary origin of the three-finger toxin family
- A neglected three-finger toxin subfamily diverged in structure and function

Zhang et al., 2022, Cell Reports *40*, 111079
July 12, 2022
https://doi.org/10.1016/j.celrep.2022.111079

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

The structural and functional divergence of a neglected three-finger toxin subfamily in lethal elapids

Zhi-Yi Zhang,[1] Yunyun Lv,[1,2] Wei Wu,[1,3] Chaochao Yan,[1] Chen-Yang Tang,[1] Changjun Peng,[1,3] and Jia-Tang Li[1,3,4,5,*]
[1]CAS Key Laboratory of Mountain Ecological Restoration and Bioresource Utilization & Ecological Restoration and Biodiversity Conservation Key Laboratory of Sichuan Province, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, Sichuan 610041, China
[2]College of Life Science, Neijiang Normal University, Neijiang, Sichuan 641100, China
[3]University of Chinese Academy of Sciences, Beijing 101408, China
[4]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan 650223, China
[5]Lead contact
*Correspondence: lijt@cib.ac.cn
https://doi.org/10.1016/j.celrep.2022.111079

SUMMARY

Bungarus multicinctus is a widely distributed and medically important elapid snake that produces lethal neurotoxic venom. To study and enhance existing antivenom, we explore the complete repertoire of its toxin genes based on *de novo* chromosome-level assembly and multi-tissue transcriptome data. Comparative genomic analyses suggest that the three-finger toxin family (3FTX) may evolve through the neofunctionalization of flanking *LY6E*. A long-neglected 3FTX subfamily (i.e., MKA-3FTX) is also investigated. Only one MKA-*3FTX* gene, which evolves a different protein conformation, is under positive selection and actively transcribed in the venom gland, functioning as a major toxin effector together with MKT-3FTX subfamily homologs. Furthermore, this lethal snake may acquire self-resistance to its β-bungarotoxin via amino acid replacements on fast-evolving *KCNA2*. This study provides valuable resources for further evolutionary and structure-function studies of snake toxins, which are fundamental for the development of effective antivenoms and drug candidates.

INTRODUCTION

Snakes are broadly distributed worldwide and include more than 600 extant species with venom systems to help in prey capture and predator defense (Suryamohan et al., 2020). *Bungarus multicinctus* is a highly venomous elapid snake that feeds primarily on loaches, frogs, skinks, snakes (including conspecifics), and mice (Mao et al., 2010). It is widely distributed across southern Asia, including in India, Pakistan, Indonesia, Sri Lanka, Malaysia, Bangladesh, Vietnam, and China (Liang et al., 2021). In China, although bites are rare, its envenomation ranks first in lethality (Liang et al., 2021; Shan et al., 2016). Available antivenoms show good neutralizing efficacy but can exhibit batch-to-batch variation and elicit severe allergic reactions (Lin et al., 2020; Mao et al., 2017), which may be due to poorly defined antivenom components (Casewell et al., 2020; Suryamohan et al., 2020). Therefore, clarifying the full repertoire and composition of *B. multicinctus* toxin genes will facilitate the development of next-generation antivenom with well-defined composition (Casewell et al., 2020; Suryamohan et al., 2020).

Bungarus multicinctus venom contains postsynaptic three-finger toxins (3FTXs) as well as presynaptically acting β-bungarotoxins made up of type I phospholipase A_2 (PLA_2 I) (i.e., A chain) and snake venom Kunitz-type serine protease inhibitors (KUN) (i.e., B chain) (Jiang et al., 2011; Shan et al., 2016; Yin et al., 2020). The three-finger toxin family is one of the most abundant and diverse toxin families in elapid snakes. In total, 33 monophyletic 3FTX subgroups have been identified, 13 of which have been functionally characterized and named according to the activity of the representative members within (Fry et al., 2003). These subgroups include acn-esterase (i.e., acetylcholinesterase) inhibitor, antiplatelet toxin, L-type calcium blocker, type IA and IB cytotoxin, types I, II, and III α-neurotoxin, κ-neurotoxin, types A, B, and C muscarinic toxin, and synergistic toxin (both type A muscarinic and synergistic toxin subgroups are called aminergic toxin in the UniProt database). The remaining subgroups, which consist of members with unknown function, are termed orphan groups and numbered I to XX (Fry et al., 2003). Although the activities of some orphan groups have since been ascertained, the nomenclature has been preserved. *Bungarus multicinctus* 3FTXs can be phylogenetically categorized into type II α-neurotoxin (i.e., α-bungarotoxin), κ-neurotoxin (i.e., κ-bungarotoxin), and several orphan groups, which constitute the largest component of its venom (Jiang et al., 2011; Shan et al., 2016; Yin et al., 2020).

Studying the mechanisms that have generated the 3FTX family will not only help illustrate how venom systems evolved in Elapidae (Pahari et al., 2007) but will also help clarify their

structure-function relationships, predict the likely activity of toxins, reveal the neural transmission processes, and elucidate the pathogenesis of neuromuscular diseases (Fry et al., 2003; Kini and Doley, 2010). Such efforts require comprehensive knowledge of the corresponding sequences and their evolutionary patterns (Fry et al., 2003, 2013; Fry, 2005). However, the evolutionary relationships among 3FTX subgroups are difficult to resolve due to their short length and extreme variation caused by rapid evolution and diversification (Dashevsky et al., 2021; Sunagar et al., 2013). Therefore, no conclusion has been made as to which subgroup is more ancient (Dashevsky et al., 2021), and discussion of the relatedness among subgroups has been limited. For example, Fry et al. (2003) identified the close relationship between type A muscarinic and synergistic toxins through phylogenetic analysis, and Fujimi et al. (2003) suggested that type II α-neurotoxin/κ-neurotoxin may give rise to type I α-neurotoxin through insertion of "A" in the second intron. In addition, current methods used to identify 3FTXs (i.e., proteome and transcriptomic analysis) tend to be biased toward homologs highly expressed in the venom gland. Thus, the incomplete sampling of 3FTX sequences has obscured the evolutionary trajectory of this toxin family.

Current knowledge of *B. multicinctus* venom is based solely on venom proteomics and/or venom gland transcriptomes, which have provided valuable information on the complex composition of venom but have not clarified the genomic organization of the toxin gene families (Brahma et al., 2015; Lomonte and Calvete, 2017). High-quality reference genomes and transcriptome data are necessary to comprehensively identify toxin genes, and thus design safe and effective antivenoms and drug candidates (Abd El-Aziz et al., 2019; Bordon et al., 2020; Casewell et al., 2020; Suryamohan et al., 2020). Complete genomic information of toxin family members will also facilitate their evolutionary study (Dowell et al., 2016). To date, however, few chromosome-scale genome assemblies of snakes have been published. Thus, the scarcity of genetic resources of *B. multicinctus* has hindered functional and evolutionary investigations of its toxins.

Here, we provide the chromosome-level reference genome of *B. multicinctus*. Based on the assembly, the comprehensive *B. multicinctus* toxin gene catalog was acquired, showing considerable diversification within the 3FTX family. Moreover, comparative genomic analyses with seven other snake species revealed a potential mechanism exploited by *B. multicinctus* to evolve self-resistance to its own toxin by altering the toxin target. Additionally, a co-expression network involved in toxin synthesis and secretion was revealed through genomic and multi-tissue transcriptomic data analyses.

RESULTS

Chromosome-scale *de novo* assembly and annotation

To generate a high-quality reference genome (Figure 1A), we produced 98.5-fold genomic coverage of PacBio Sequel II reads (subread N50 length 33.74 kb), 83.9-fold short paired-end reads, and 128.5-fold Hi-C data from a *B. multicinctus* male (Table S1). PacBio reads were used for initial assembly, which was then polished with short paired-end reads to generate an assembly containing 609 contigs with an N50 of 14.45 Mb and total length of 1.61 Gb (Table S1). Contig assembly contiguity was improved using Hi-C (Figure 1B), resulting in 831 scaffolds with an N50 of 135.38 Mb (Table S1). In total, 1.57 Gb of assembled genome sequences (97.5%) were clustered onto 18 pseudo-chromosomes, ranging from 12.57 Mb to 276.03 Mb in length (Table S1). The average GC content was 40.0%, similar to that of other snakes (Margres et al., 2021). Analysis of genome assembly completeness identified 94.6% complete vertebrate benchmarking universal single-copy orthologs (BUSCO) genes (Table S1; Simão et al., 2015). Regarding BUSCO and N50 scores, the assembly was relatively complete compared with recently published snake genome assemblies (Margres et al., 2021; Suryamohan et al., 2020). In total, 19,004 protein-coding genes were annotated, 96.4% of which were annotated in public databases. Repetitive sequences accounted for 45.79% of the genome, consisting of 0.70% tandem repeats and 41.33% transposable elements (TEs). In the *B. multicinctus*, *Thamnophis elegans*, and *Naja naja* genomes, the TEs with smaller Kimura distances to their corresponding consensus sequences had higher genome coverage than the TEs with larger Kimura distances, indicating relatively recent burst of TEs. Long interspersed nuclear elements were obviously expanded in *B. multicinctus*, while long terminal repeat-retrotransposons were expanded in *N. naja* (Figure 1C).

Expansion of toxin gene families in *B. multicinctus*

Cluster analysis of gene families was performed for *B. multicinctus* and seven other snake species (i.e., *Python bivittatus*, *Crotalus viridis*, *Deinagkistrodon acutus*, *N. naja*, *Hydrophis curtus*, *Pantherophis guttatus*, and *T. elegans*). In total, 881 genes were identified as *B. multicinctus* specific and were significantly (adjusted $p < 0.05$) enriched in Gene Ontology (GO) terms related to toxins, including toxin activity (GO:0090729), phospholipase A_2 activity (GO:0004623), serine-type endopeptidase inhibitor activity (GO:0004867), and L-amino-acid oxidase activity (GO:0001716) (Table S2). The *B. multicinctus*-specific gene set contained seven type I phospholipase A_2 (*PLA_2 I*) and nine *KUN* copies, which encode the A-chain and B-chain subunits of the *Bungarus*-specific β-bungarotoxin, respectively (Table S3; Rowan, 2001). Furthermore, 116 gene families exhibited significant expansion in the *B. multicinctus* genome (Figure 2A). Functional enrichment analysis showed that these genes were significantly enriched (adjusted $p < 0.05$) in 251 GO terms, including gene functions related to toxin activity (GO:0090729), host cell postsynaptic membrane (GO:0035792), serine-type endopeptidase inhibitor activity (GO:0004867), and pathogenesis (GO:0009405) (Figure 2B and Table S2). The expanded toxin genes included five complement C3 homologs, 27 *3FTX* genes, and 19 *KUN* genes (Table S3).

Genomic landscape of toxin genes

A total of 118 toxin genes from 17 families were identified in *B. multicinctus*, including 30 *3FTX*, nine *PLA_2 I*, and 31 *KUN* genes (Figure 2C and Table S3). Most toxin families were organized on macrochromosomes, in accordance with the genome of *N. naja* but in contrast to the genomes of crotalids, where toxin genes are mainly distributed on microchromosomes (Figure 2C; Schield et al., 2019). Among the 30 *3FTXs*, 15 were clustered

CellPress
OPEN ACCESS

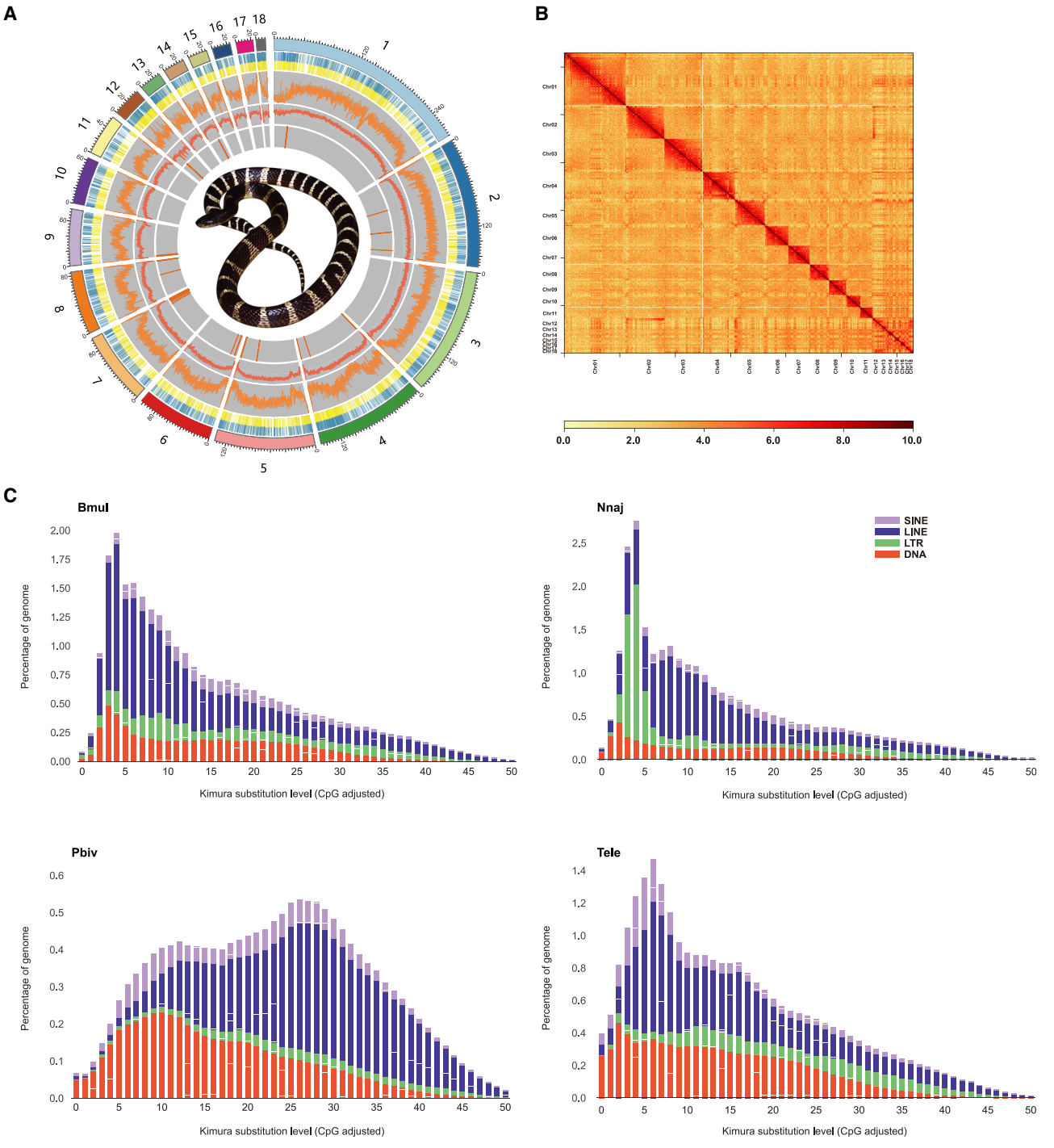
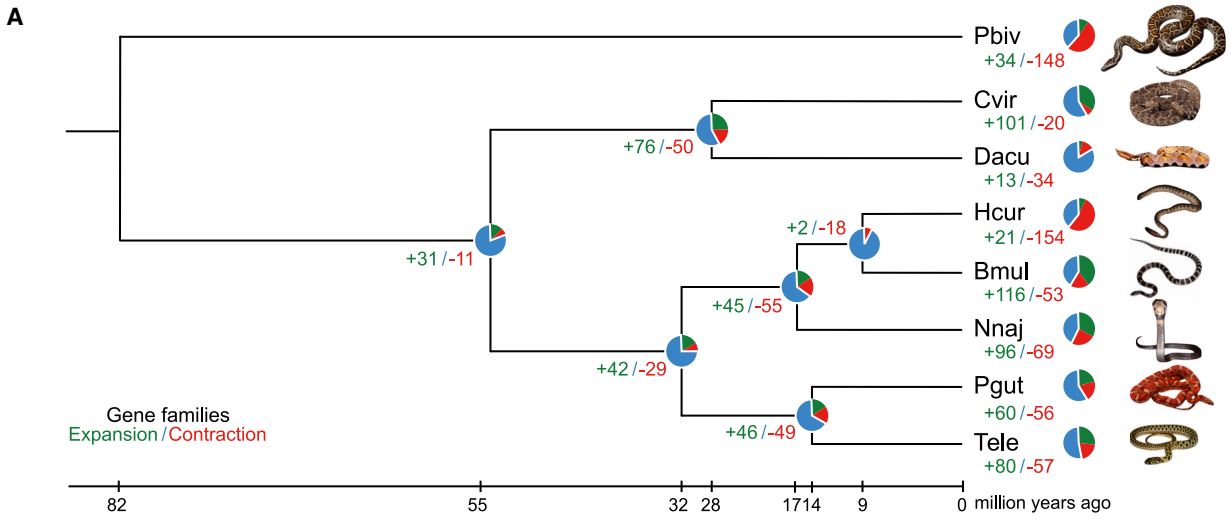


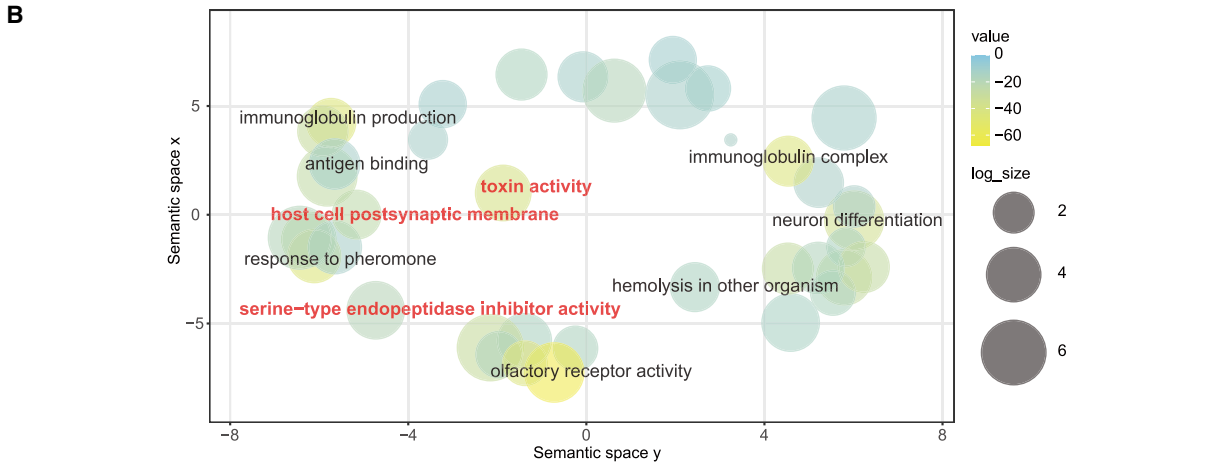
Figure 1. Genome assembly of *Bungarus multicinctus* compared with other snakes

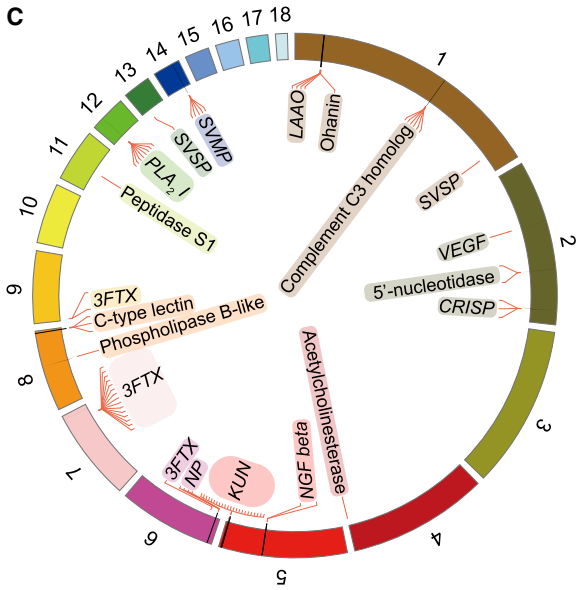
(A) Genomic features of *B. multicinctus* chromosomes. Circos plot of assembly showing sizes (each single scale represents 4 Mb) of 18 pseudo-chromosomes, gene density (percentage of genes per 100-kb window) of plus (yellow) and minus (blue) strand, GC content (minimum 30%, maximum 65%), transposable element (TE) occupancy (minimum 0%, maximum 100%), and distribution of toxin genes. Photo in circle shows *B. multicinctus*.

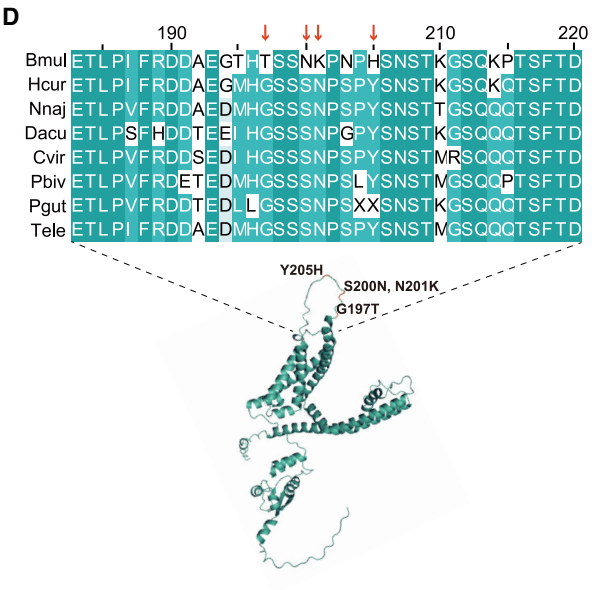
(B) Hi-C contact map showing genome-wide pairwise chromatin interactions of 18 chromosomes. Bar at bottom shows interaction strength.

(C) TE accumulation history in different snakes based on same analysis. x axis is CpG adjusted Kimura distance from consensus in TE library; y axis is percentage of TE occupancy in genome. Copies clustering on the left of the graph do not greatly diverge from the consensus sequence of the element and potentially correspond to recent copies, while sequences on the right might correspond to ancient copies. Bmul, *B. multicinctus*; Nnaj, *Naja naja*; Pbiv, *Python bivittatus*; Tele, *Thamnophis elegans*.









(legend on next page)

within a 13.2-Mb region on chromosome 7, one was located on chromosome 6, and two were located on chromosome 9, with the remaining homologs located on unassigned scaffolds (Figure 2C and Table S3). The *3FTX*s encode type II α-neurotoxin, κ-neurotoxin, aminergic toxin β-cardiotoxin subtype, and toxins from the orphan groups II, IV, V, VII, VIII, IX, XVII, and XIX (Fry et al., 2003; Jiang et al., 2011; Shan et al., 2016; Yin et al., 2020). There were also four *3FTX*s that could not be ascribed to any of the 33 previously defined subgroups (Table S3). Another large toxin gene family annotated in the genome consisted of 23 *KUN*s clustered on chromosome 5, including five encoding the B chain of β-bungarotoxin (Figure 2C and Table S3). *PLA_2 I*, which encodes a presynaptic-acting neurotoxin, was found on chromosome 12 with eight copies (Figure 2C and Table S3). Other toxin families, including 5′-nucleotidase, natriuretic peptide, cysteine-rich secreted protein, β-nerve growth factor (NGF-β), snake venom metalloproteinase (SVMP), and snake venom serine proteinase, were distributed on different chromosomes or unassigned scaffolds (Figure 2C and Table S3).

Positively selected genes and genes with high rates of molecular evolution

We identified 374 genes evolving under strong positive selection and 288 genes with high rates of molecular evolution (Table S2). Both gene sets contained *KCNA2*, which encodes potassium voltage-gated channel subfamily A member 2 (Kv1.2), a target of many animal toxins, including KUNs secreted by snakes (Harvey, 2001; Kwong et al., 1995; Mouhat et al., 2008). Of note, four conserved replacements among snakes, i.e., G197T, S200N, N201K, and Y205H, were found on Kv1.2 in *B. multicinctus* (Figures 2D and S1). These positively selected sites were located at the tip of the extracellular region of the voltage sensor domain (i.e., S1-S2 segment), which couples changes in transmembrane electrical potential with conformational changes that regulate channel opening or closing (Kim and Nimigean, 2016). *KCNA1*, another potassium channel subtype targeted by snake toxins (Harvey, 2001; Mouhat et al., 2008), was also identified as rapidly evolving, implying a higher rate of molecular evolution in *B. multicinctus*.

Divergence of *3FTX* gene family in snakes

The 3FTX family is one of the most diverse and important toxin families in snakes and constitutes the largest part of *B. multicinctus* venom (Figure S2; Jiang et al., 2011; Yin et al., 2020). Thus, we reconstructed the phylogeny of the 3FTX family and its possible ancestors, including LY6E, LYNX1, and SLURP2. The maximum-likelihood tree showed that LY6E was clustered with 3FTXs and indicated a reciprocally monophyletic relationship between two 3FTX subfamilies, which contained distinct conserved signal peptide sequences (Figures 3A, S3, and S4). According to the first three amino acids of the consensus signal peptide sequence of the corresponding alignment (Figure S4), the two subfamilies were defined as "MKA" and "MKT." More MKT-*3FTX* copies were found in the *B. multicinctus* than in the *N. naja* assembly (30 and 24, respectively). In comparison, MKA-*3FTX* subfamily showed considerable expansion in the *T. elegans* genome, with the largest number of MKA gene copies (Figures 3B and S5A).

Almost all venom 3FTXs belonged to the MKT subfamily, which showed remarkable expansion and diversity in the Elapidae (Figures 3A and S3). According to the phylogeny, the MKT-3FTXs could be further categorized into 32 subgroups exhibiting neurotoxicity, hemotoxicity, or cytotoxicity. Some subgroups are found in the venom of different genera. For example, type I and type II α-neurotoxins have been discovered in various genera across the phylogeny of elapids. In contrast, many subgroups were lineage specific. For instance, κ-neurotoxin and orphan groups V, VII, IX, and XVIII were *Bungarus* specific, and together with eight other subgroups, *Bungarus* exhibited the most diverse repertoire of MKT-3FTXs, followed by *Naja* (ten MKT-3FTX subgroups) (Figure S3). In comparison, the MKA subfamily was expanded in the non-toxic *T. elegans* (Figures 3B and S3), and only one MKA-3FTX homolog subtype (i.e., orphan group XIX) has been reported in the venom of elapid snakes (Abd El-Aziz et al., 2018; Kuhn et al., 2000; Rivera-Torres et al., 2016). Moreover, the syntenic relationship of *3FTX* clusters in different snakes was determined by retrieving conserved flanking *TOP1MT*, *ADGRB1* (*BAI1*), and Ly6/uPAR superfamily (which contains the *3FTX* family) members (Figures 3B and S5A). Generally, compared with the MKA-*3FTX* subfamily, the proportion of positively selected (i.e., Ka/Ks > 1, where Ka is number of non-synonymous substitutions per non-synonymous site and Ks is the number of synonymous substitutions per synonymous site) copies was higher in the MKT-*3FTX* subfamily (Figure S5B). In *B. multicinctus*, six MKT-*3FTX*s were under positive selection, while among the MKA-*3FTX* paralogs the only copy recruited into the venom

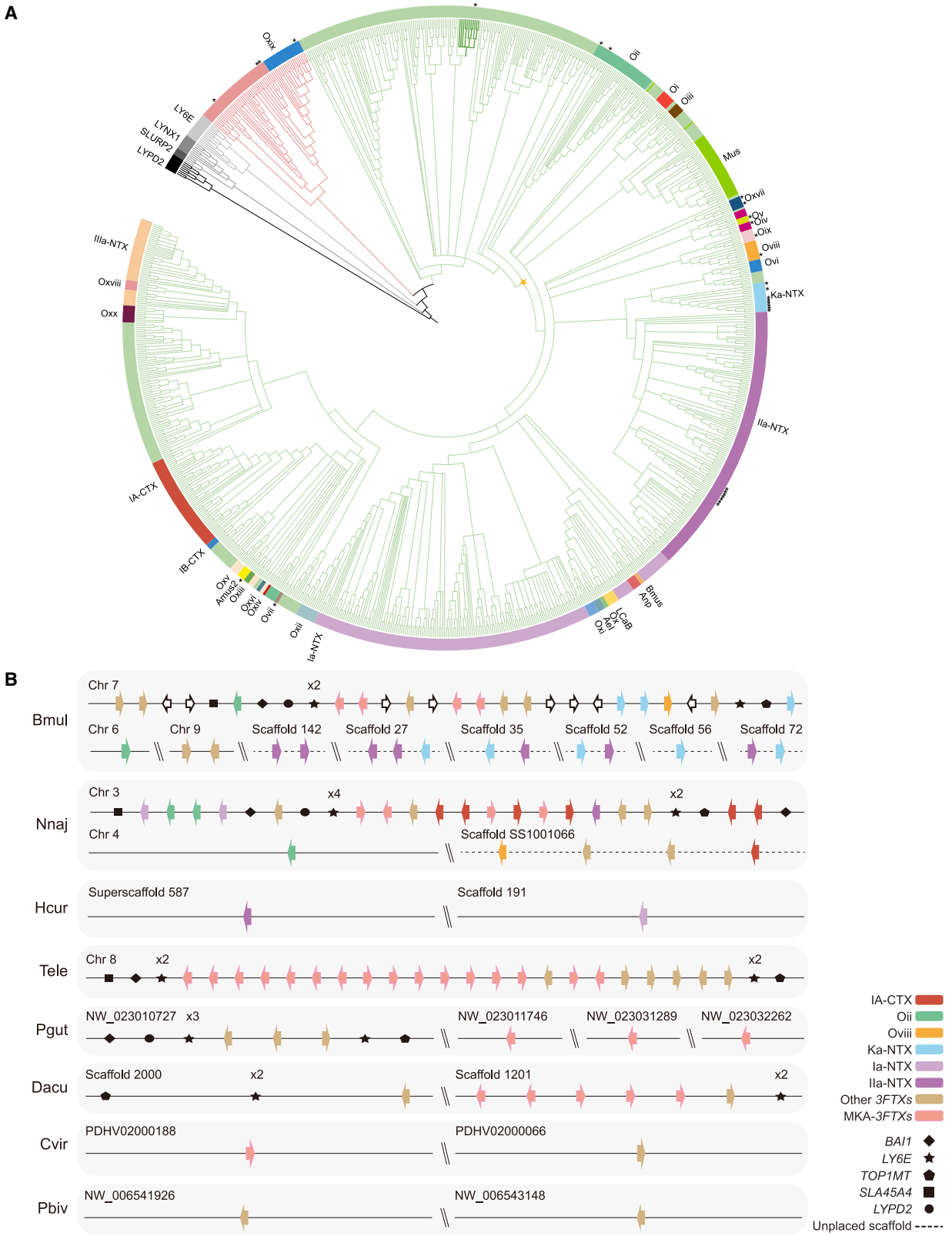
Figure 2. Comparative genomic analyses across eight snake species and genomic landscape of *B. multicinctus* toxin genes

(A) Phylogenetic tree with timescale indicating divergence times. Green and red numbers indicate number of significantly expanded (+) and contracted (−) gene families. Pie charts indicate proportions of expanded (green), contracted (red), and other (blue) gene families of each node. Timescale bar at the bottom shows the divergence time. Corresponding pictures of studied snakes are presented on the right.

(B) Enriched GO terms of expanded gene families. Bubble size represents number of annotations for a certain GO term in $\log_{10}$ scale, and color indicates adjusted enrichment p value in negative $\log_{10}$ scale. Significant terms with semantic similarity smaller than default threshold (0.15) are not presented, and term ID related to snake venom toxins is highlighted in red.

(C) Distribution of *B. multicinctus* toxin gene families on 18 chromosomes. Each red line represents a toxin gene copy. *LAAO*, L-amino acid oxidase; *SVMP*, snake venom metalloproteinase; *SVSP*, snake venom serine protease; *VEGF*, vascular endothelial growth factor; *NGF*, nerve growth factor; *CRISP*, cysteine-rich secreted protein; *NP*, natriuretic peptide; *3FTX*, three-finger toxin; *PLA_2 I*, type I phospholipase A_2; *KUN*, snake venom Kunitz-type serine protease inhibitors.

(D) Sequence alignment showing conserved replacements (indicated by arrows) on *B. multicinctus* KCNA2, voltage-gated channel subfamily A member 2 (Kv1.2). Alignment only exhibits S1-S2 extracellular region of the channel. AlphaFold2 v2.0.0 (Jumper et al., 2021) predicted tertiary structure of *B. multicinctus* Kv1.2 at the bottom shows the mutated sites (highlighted in orange) of the S1-S2 segment. Pbiv, *Python bivittatus*; Cvir, *Crotalus viridis*; Dacu, *Deinagkistrodon acutus*; Nnaj, *Naja naja*; Hcur, *Hydrophis curtus*; Bmul, *B. multicinctus*; Pgut, *Pantherophis guttatus*; Tele, *Thamnophis elegans*.



(legend on next page)

(i.e., the orphan group XIX homolog) experienced positive selection (Table S3).

Structural divergence of 3FTXs in *B. multicinctus*

The protein conformations of 3FTX homologs in *B. multicinctus* were generally similar, but differences among subfamilies were evident based on the TM-score (Figure S6A), an indicator for assessing topological similarity in protein structure. The 3FTXs consisted of three β-strand-containing loops (I, II, and III) protruding from a small core crosslinked by four conserved disulfide bonds, but with variation in loop length, number of β strands, and disulfide bonding patterns (Figures 4 and S6B).

In the MKT subfamily, 15 homologs had an extra disulfide bond in loop II, which formed a short helix at the tip of the loop characteristic of long-chain neurotoxins (Figures 4A and S6B). These long-chain neurotoxins were actively transcribed in the venom gland of *B. multicinctus* (Table S3), and based on phylogenetic inference, seven were identified as type II α-neurotoxin homologs (Figures 3A and S3), which are known to target muscular and neuronal nicotinic acetylcholine receptors (nAChRs) as well as GABA(A) channels (Hannan et al., 2015; Servent et al., 1997). Sequence similarity among the seven homologs was high, with only an A52V difference observed, which did not appear to affect higher structures (Figures 4A and S6B). The remaining eight long-chain toxins were κ-neurotoxin homologs, which selectively inhibit α-3-containing and several α-4-containing nAChRs (Chiappinelli et al., 1996), and were only found in the venom of *Bungarus*. Several amino acid replacements were observed in these κ-neurotoxins (Figure 4A), but the positions crucial for receptor binding were conserved (i.e., R55) or did not involve substitutions that can alter potency (i.e., position 57) (Figure 4A; Chiappinelli et al., 1996). We also identified seven 3FTXs with two extra cysteines in the first loop (Figure 4A). These types of 3FTXs contain a fifth disulfide bond in the first loop, which is an ancestral character lost in the long-chain and short-chain 3FTXs (Fry et al., 2003) and referred to as "non-conventional" in some studies (Kini and Doley, 2010; Nirthanan et al., 2003). The orphan groups II, IV, V, and XVII targeting nAChRs or mAChRs were non-conventional 3FTXs (Figures 4A and S6B; Chang et al., 2002; Chung et al., 2002; Mordvintsev et al., 2009; Nirthanan et al., 2002). Of note, the MKT paralog near the base of the MKT branch was also a non-conventional 3FTX but differed from the others in conformation (Figures 4A,S3, and S6B) and was transcribed at a lower level than the other non-conventional MKT-*3FTX*s (except for an orphan group II homolog that was not expressed in the venom gland) (Table S3). The remaining MKT paralogs were short-chain toxins with only four disulfide bridges (Figures 4A and S6B). Among these short-chain toxins, only the aminergic toxin (β-cardiotoxin homolog) was not expressed (Table S3).

Regarding the MKA subfamily, there were three short-chain and one non-conventional homolog (Figures 4A and S6B). The structure of the non-conventional homolog was similar to that of its counterparts (i.e., the orphan group XIX 3FTXs) found in the venom of *Dendroaspis angusticeps* (UniProtKB: C0HJT4), *Dendroaspis jamesoni* (UniProtKB: P25682), *Walterinnesia aegyptia* (UniProtKB: C0HKZ8), and *Bungarus candidus* (UniProtKB: P81782). Various functions of the orphan group XIX 3FTXs have been revealed but none have shown potent toxicity (Abd El-Aziz et al., 2018; Kuhn et al., 2000; Rivera-Torres et al., 2016). The non-conventional homolog contained an additional β strand in the third loop compared with the other three paralogs (Figures 4B and S6B). This may be due to the amino acids at positions 67 and 68 creating stronger hydrogen bonds, thus forming a stable β sheet (Figure 4).

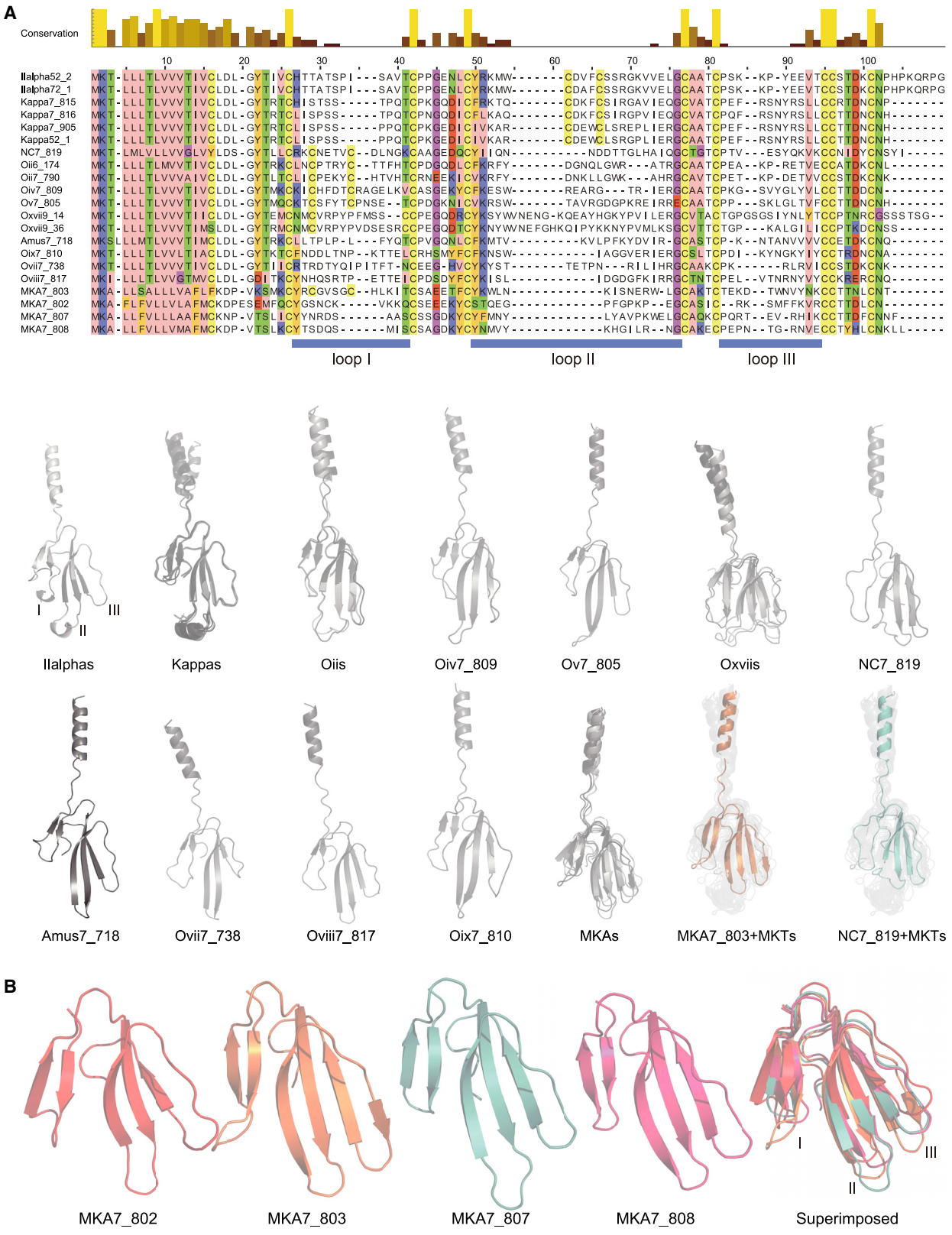
Different expression patterns between *3FTX* subfamilies

To explore the expression patterns of venom genes, we sequenced transcriptomes of multiple tissues, e.g., venom gland, kidney, heart, and liver. Weighted gene co-expression network analysis (WGCNA) was performed, and 1,550 genes showed high correlation with venom gland traits and were assigned into the corresponding brown module (Figure S7A). Functional enrichment analysis showed that genes in the venom gland module were significantly enriched in terms related to toxins as well as protein folding and cytoplasmic transportation (e.g., protein disulfide isomerase activity and ER-to-Golgi vesicle-mediated transport, respectively) (Table S4). Within the *3FTX* gene family, two subfamilies showed distinct expression profiles (Figure 5). Most MKT-*3FTX* homologs (except one orphan group II homolog and the aminergic toxin β-cardiotoxin subtype) were highly expressed in the venom gland (Figure 5 and Table S3) and 23 MKT-*3FTXs* were categorized into the co-expression network that is significantly and most closely associated with the venom gland according to WGCNA results. Most MKA-*3FTX* copies were lowly transcribed in the venom gland (Figure 5 and Table S3), with only one MKA copy

Figure 3. Evolution and diversification of 3FTXs

(A) Maximum-likelihood tree, with LYPD2 (black) as the outgroup, indicating the clustering of LY6E with 3FTX. The tree also shows a reciprocally monophyletic relationship of MKA-3FTX and MKT-3FTX subfamilies (pink and green branches, respectively). Darker green branches represent clusters of elapid MKT-3FTXs grouped with non-elapids homologs, and the node with a star sign indicates Elapidae-specific clusters. Colors of outermost rim represent different 3FTX subgroups, illustrating much more subdivisions in MKT subfamily relative to MKA subfamily. Names of the 3FTX subgroups are distributed around the phylogeny. Ia-NTX, type I α-neurotoxin; IIa-NTX, type II α-neurotoxin; IIIa-NTX, type III α-neurotoxin; IA-CTX, type IA cytotoxin; IB-CTX, type IB cytotoxin; Mus, representing type A (i.e., aminergic) and C muscarinic toxin; Amus2, aminergic toxin β-cardiotoxin subtype; Bmus, type B muscarinic toxin; Ka-NTX κ-neurotoxin; Oi to Oxx, orphan groups I to XX; AnP, antiplatelet toxin; AeI, Acn-esterase inhibitor; LCaB, L-type calcium blocker. The locations of *B. multicinctus* genes on the tree are indicated by asterisks (*). Nodes with <50 support are collapsed (support values are provided in Figure S3).

(B) Syntenic regions of the studied snakes containing *3FTX*s and conserved flanking anchor genes. Pink arrows represent MKA homologs, and other color arrows represent different subgroups of MKT-*3FTX*s. Subgroups, including aminergic toxin β-cardiotoxin subtype, type C muscarinic toxin, and orphan groups I, IV, V, VII, IX, XV, and XVII, are referred as "Other *3FTX*s." Detailed positions of each subgroup are provided in Figure S5A. Empty arrows are pseudo-*3FTX* genes that lack complete exons. Conserved flanking genes are represented in different shapes with copy number indicated at the top. Bmul, *B. multicinctus*; Nnaj, *Naja naja*; Hcur, *Hydrophis curtus*; Tele, *Thamnophis elegans*; Pgut, *Pantherophis guttatus*; Cvir, *Crotalus viridis*; Dacu, *Deinagkistrodon acutus*; Pbiv, *Python bivittatus*.



(legend on next page)

(the orphan group XIX homolog), containing an additional β strand in the third loop (Figure 4B), assigned to the venom gland module. In contrast, three other MKA gene copies showed relatively high expression in the kidney (Figure 5). Furthermore, *3FTX* genes and toxin genes, including *PLA$_2$ I*, *KUN*, and *SVMP*, were also assigned to the venom gland module with high expression (Table S3). Several were identified as *B. multicinctus*-specific genes, including those encoding the A chain and B chain of β-bungarotoxin. Gene significance (GS) and module membership (MM) were used to quantify the association of individual genes with the traits of interest and to assess the correlation between module eigengenes and gene expression profiles, respectively. In the venom gland module, 427 genes with GS > 0.8 and MM > 0.8 were identified as central elements associated with the venom gland (Figure S7B and Table S4). This gene set included 18 MKT-*3FTX*s as well as genes related to protein folding (e.g., *PDIA6*, *ERP44*, and *SDF2L1*) and cytoplasmic transportation (e.g., *TMED7* and *TMED10*) (Table S4; Aber et al., 2019; Anelli et al., 2003; Eletto et al., 2014; Gong et al., 2021; Zhang et al., 2020), implying key roles in venom gene synthesis. Co-expression subnetworks regulated by central elements with weights >0.5 were generated. Three toxin genes (i.e., one orphan group XVII *3FTX*, one *SVMP*, and one NGF-β, with UniProt annotations of 3NOH_BUNMU (UniProtKB: Q9PW19), VM3_BUNMU (UniProtKB: A8QL49), and NGFV_BUNMU (UniProtKB: P34128), respectively) and five protein-folding and processing-related genes (i.e., *SDF2L1*, *ERP44*, *PDIA6*, *TMED10*, and *TMED7*) were identified as hub genes (Figure S7C).

Expansion and diversification of *KUN* family in *B. multicinctus*

The maximum-likelihood tree divided the KUN family into two main clades, one of which consisted almost entirely of elapid KUNs (Figure S8A). The tree also identified a KUN cluster from snakes of Australia and New Guinea, and a clade dominated by various viperid snakes (Figure S8A). The KUNs in *B. multicinctus* showed substantial expansion but the paralogs were not in a single monophyly, and some copies (e.g., Bmul_scaffold5_G01543.t1 and Bmul_scaffold5.mRNA.2, 10, 21, and 25) were grouped with non-elapids or elapids of other genera (Figure S8A). The *KUN* paralogs encoding the B chain of β-bungarotoxin were likely to have evolved through *Bungarus*-specific expansion. In *B. multicinctus*, this KUN subtype could be divided into two main clusters, each of which was clustered with different canonical *Bungarus* KUNs (Figure S8A). In terms of expression, *B. multicinctus KUN*s in the *Bungarus*-specific clusters generally exhibited higher transcription in the venom gland, and the B-chain-encoding paralogs in the cluster with more duplicates were more actively transcribed (Figure S8B).

DISCUSSION

Bungarus multicinctus is a widely distributed elapid snake that produces potent neurotoxic venom with high lethality. Its toxins have been engineered to create investigational ligands in biomedical and biochemical research (Kini and Doley, 2010). However, comprehensive genetic resources are needed to facilitate our understanding of the evolution, expression, and function of these toxins, which is fundamental for developing effective antivenoms. Our *B. multicinctus* chromosome-scale assembly and associated transcriptome data provide a valuable resource for exploring the origin and evolution of venom genes, thereby enabling venom-driven drug discovery.

We identified 118 toxin genes in *B. multicinctus*, including *3FTX*, *PLA$_2$ I*, and *KUN* (Figure 2C and Table S3). Several toxin families exhibited considerable expansion and high expression in the venom gland (e.g., *3FTX*) (Figures 2C, 3B, and S2; Table S3), which may contribute to the high neurotoxicity of *B. multicinctus* venom. In many snakes, toxin families closely related to venom properties have undergone marked expansion (Jackson and Koludarov, 2020). For instance, divergence in the neurotoxicity and cytotoxicity of pitviper venom has been attributed to the initial expansion and lineage-specific deletion of paralogs in the PLA$_2$ II toxin family (Dowell et al., 2016).

Toxin families are thought to have originated from protein families fulfilling common physiological functions, either through neofunctionalization or gene co-option without duplication (Almeida et al., 2021; Jackson and Koludarov, 2020). 3FTXs belong to the Ly6/uPAR superfamily, which includes other three-finger domain-possessing peptides such as LY6E, LYNX1, and SLURP2. As these homologs can also bind with nAChRs targeted by snake toxins, they are thought to be the possible ancestors of 3FTXs (Fry, 2005; Puddifoot et al., 2015; Roy et al., 2010; Wu et al., 2015). Previous study shows that 3FTXs form a polytomy with LY6E, LYNX1, and SLURP2, thus preventing the assignment of these toxins to a specific ancestor (Fry, 2005). Nevertheless, based on our reconstructed phylogeny of 3FTX together with the detailed sequence information of syntenic blocks that contain *3FTX*s and their flanking genes (Figures 3 and S5), it is likely that *3FTX*s have been recruited into venom through the neofunctionalization of flanking *LY6E* and have diverged into two subfamilies (i.e., MKA-3FTX and MKT-3FTX) prior to the evolution of the common ancestor of elapid snakes. There are other examples of descendants being

Figure 4. Multiple sequence alignment and predicted tertiary structures of 3FTXs identified in the *B. multicinctus* genome

(A) Upper panel shows alignment of 3FTXs in *B. multicinctus* after removing identical sequences. Histogram at the top presents the conservation level (0 to 10) of each site. Blue bars indicate loop regions of 3FTXs. The 3FTX names consist of subgroup/subfamily name, chromosome/unplaced scaffold number, and gene ID, e.g., IIalpha52_2 means gene 2 on scaffold 52 is a type II α-neurotoxin. Conformation representations of the above protein sequences are shown in the lower panel. Superimposed structure models of the same 3FTX subgroup highlight the same conformation between type II α-neurotoxins. Structural alignment also exhibits differences in loop length and conformation between MKA and MKT homologs, as well as differences between uncharacterized non-conventional MKT-3FTX and other MKT paralogs.

(B) Structure models of the MKA subfamily members and structural alignment highlighting extra β-strand on loop III of the homolog recruited into venom. IIalpha, type II α-neurotoxin; Kappa, κ-neurotoxin; NC, uncharacterized non-conventional 3FTX; Oii, orphan group II; Oiv, orphan group IV; Ov, orphan group V; Oxvii, orphan group XVII; Amus, aminergic toxin β-cardiotoxin subtype; Oix, orphan group IX; Ovii, orphan group VII; Oviii, orphan group VIII; MKA, MKA-3FTX homologs.

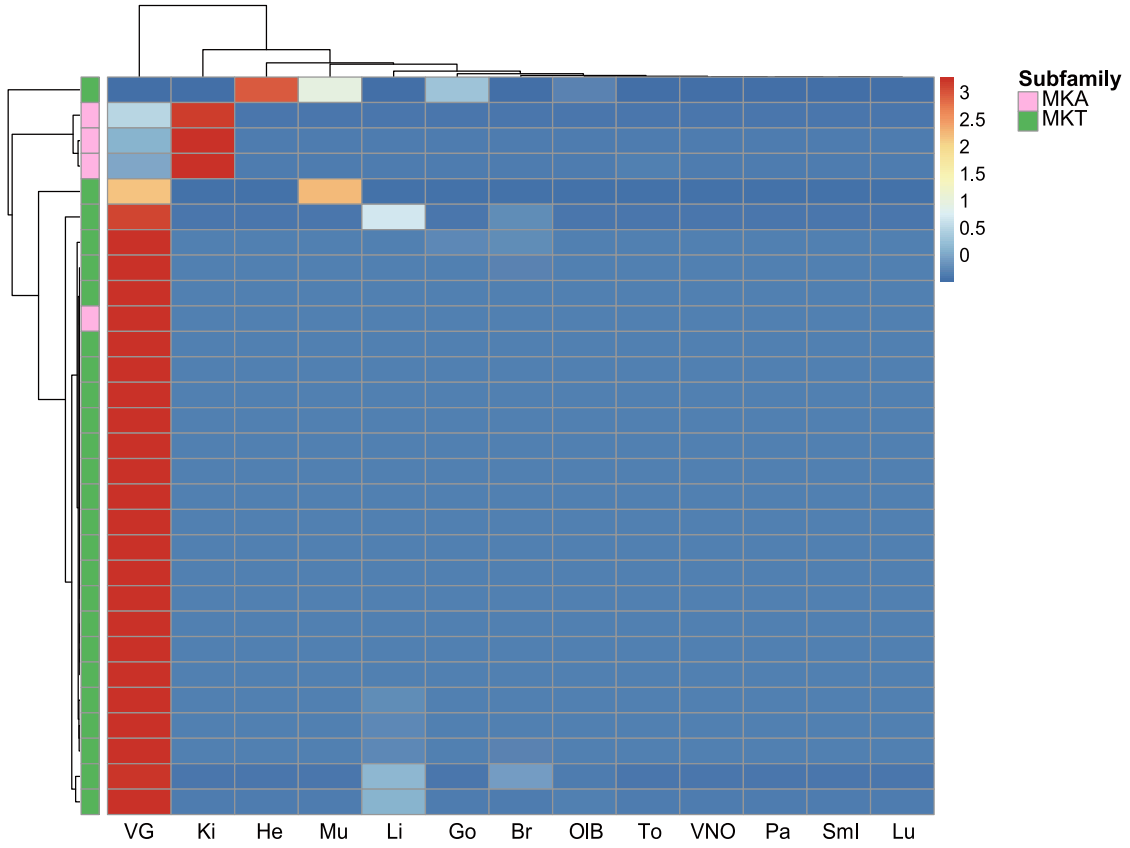


Figure 5. Expression patterns of *3FTX*s in *B. multicinctus* venom gland and other tissues

Each tissue has at least three biological replicates, and average transcripts per kilobase million values normalized by *Z* score are used to draw the heatmap. VG, venom gland; Ki, kidney; Br, brain; Li, liver; Go, gonads; OlB, olfactory bulb; To, tongue; VNO, vomeronasal organ; Pa, pancreas; SmI, small intestine; Lu, lung; He, heart; Mu, muscle.

recruited as toxins following duplication and diversification of their flanking ancestral homologs with similar functions. For instance, genes encoding SVMP, a major toxin family in viper venom, are flanked by the homologous non-toxic ADAM28 (Almeida et al., 2021; Giorgianni et al., 2020). Following initial duplication and transmembrane domain loss, ADAM28 gave rise to toxic SVMP paralogs through further diversification (Almeida et al., 2021; Giorgianni et al., 2020), a scenario that can also be applied to the emergence of 3FTXs from LY6E.

The MKT subfamily was substantially expanded in Elapidae and recruited into venom as the main effector toxins (Figures 3A and S3). The reconstructed phylogeny has revealed their considerable diversification and different evolutionary trajectories. For instance, the type I and type II α-neurotoxin perhaps emerged before the radiation of elapids (Fry et al., 2003), as they have been found in various genera across the elapid phylogeny (Figure S3). In contrast, type B muscarinic toxin, antiplatelet toxin, L-type calcium blocker, acn-esterase inhibitor, and orphan groups X and XI are distinct to *Dendroaspis* and nested within the type I α-neurotoxin cluster (also grouped with *Dendroaspis* type I α-neurotoxins), indicating that these subtypes emerged from type I α-neurotoxin following *Dendroaspis* divergence (Figure S3). Some 3FTX subtypes also likely experienced lineage-specific expansion. κ-Neurotoxins were to date found only in *Bungarus*, with many gene copies detected in the *B. multicinctus* genome. Similarly, type IA cytotoxins were only detected in *Naja* venom, and their coding genes have duplicated many times in the *N. naja* genome (Figures S3 and S5A). Moreover, several clades consisted solely of 3FTXs discovered in *Micrurus* and *Calliophis* (Dashevsky et al., 2021; Dashevsky and Fry, 2018). These 3FTXs were phylogenetically close to some conventionally recognized 3FTX subgroups but may be functionally different from them. For example, orphan group VI 3FTXs are short-chain toxins with only four disulfide bridges but, according to the sequence type, toxins in the closely related *Micrurus*-specific clade should have an extra disulfide bond in the first loop, indicating functional divergence (Figure S3). Considering the basal position of coral snakes in the species tree of Elapidae, further studies of evolutionary relatedness between 3FTXs of coral snakes and other elapids are essential for better understanding of how the functions of 3FTXs have evolved (Dashevsky et al., 2021; Dashevsky and Fry, 2018).

The MKT-*3FTX*s in *B. multicinctus* constituted the largest component of the venom gland transcriptome (Figure S2), with many copies showing high sequence variation (Figures 3B and

S5A). This is common in snakes, as families of toxins that are more abundantly expressed in venom glands typically undergo greater expansion and variation (Almeida et al., 2021). High diversification may be driven by functional pressure, which may fine-tune receptor interactions and prey specificity in different environments (Almeida et al., 2021). In North American pit vipers, the complexity of the major toxin family is strongly correlated with the phylogenetic diversity of prey (Holding et al., 2021). *Bungarus multicinctus* prey are taxonomically diverse and include loaches, frogs, skinks, snakes, and mice (Mao et al., 2010). Thus, 3FTXs may have evolved repeatedly to target the phylogenetic diversity of prey items, leading to the many subtypes of neurotoxic 3FTXs in *B. multicinctus*. We also identified a potential gene dosage effect, which may underpin MKT subfamily expansion in *B. multicinctus*. Specifically, the type II α-neurotoxin subgroup had the second highest number of copies with the same conformation but accounted for the largest proportion of 3FTXs in the venom (Figures 3B, 4A, and S2; Yin et al., 2020). Similarly, *Crotalus durissus* has up to 32 crotamine myotoxin paralogs with little sequence variation, and gene family copy number is positively correlated with crotamine concentration in venom, thereby exhibiting an obvious gene dosage effect (Oguiura et al., 2009).

Regarding the MKA subfamily, four copies of MKA-*3FTX*s were discovered in the elapid genomes (Figure 3B), and only one paralog (i.e., orphan group XIX homolog) was recruited into the venom, although its toxicity to prey has not yet been demonstrated (Abd El-Aziz et al., 2018; Kuhn et al., 2000; Rivera-Torres et al., 2016). Other MKA-*3FTX* paralogs in *B. multicinctus* were expressed in the kidney rather than in the venom gland, so this subfamily appears to play a different role outside of envenomation. However, in contrast, the subfamily exhibited marked expansion in the non-toxic *T. elegans*, although no studies have reported on 3FTXs in the saliva of this species (Hill and Mackessy, 2000; Perry et al., 2018), further obscuring the role of MKA homologs. Nevertheless, the differences in conformation and expression patterns between the orphan group XIX homolog and kidney-expressed MKA-*3FTX*s suggest functional divergence (Figures 4B and 5). The orphan group XIX homolog has an additional β strand on loop III, which contains K67 and D68. Through salt bridges formed by these two residues (Wang et al., 2012), this MKA-3FTX may form a homodimer in which the subunits are antiparallel to each other. 3FTXs such as κ-neurotoxin and haditoxin also form homodimers through the third loop where the monomers are oriented in the opposite direction, although the dimers are not stabilized by salt bridges (Kini and Doley, 2010). However, it is unknown whether the additional β strand is fundamental for the orphan group XIX homolog to acquire its role in venom. Identifying the functional sites of the MKA homologs and comparing them with those of other major toxins could help us to understand how 3FTXs obtain their toxicity.

KUN is another important family of toxins in *Bungarus* snake venom. This family is greatly expanded in *B. multicinctus*, and paralogs have undergone different evolutionary trajectories (Figure S8A). Several copies existed prior to Elapidae radiation (e.g., Bmul_scaffold5_G01543.t1, Bmul_scaffold5.mRNA.21 and 25), which are homologous to the waprin domain-fused KU-WAP gene, thought to be the ancestral *KUN* that evolved early in snakes (Župunski and Kordiš, 2016). Many other *KUN* copies have duplicated and diversified after *Bungarus* diverged from other elapid snakes, including *Bungarus*-specific β-bungarotoxin B-chain-encoding *KUN*s. In our study, the B-chain-encoding *KUN*s could be further divided into two clusters that evolved from different canonical *Bungarus KUN* paralogs. Similar to the type II α-neurotoxin, a gene dosage effect may underlie the expansion of *KUN*s in *B. multicinctus*. Specifically, copies that evolved through *Bungarus*-specific duplication generally exhibited higher transcription in the venom gland than the ancestral homologs, and copies in the larger B-chain-encoding *KUN* cluster were more actively transcribed (Figure S8B).

The β-bungarotoxin B-chain subunit guides the toxin to its site of action by binding to Kv1.2 channels (Kwong et al., 1995; Rowan, 2001). As cannibalism has been reported in *Bungarus* (Mao et al., 2010; Pandey et al., 2020), we speculate that *B. multicinctus* may have evolved resistance to β-bungarotoxins through mutations on Kv1.2, which may lower the affinity between the toxin and its targeting site through electrostatic repulsion. Indeed, there is evidence of self-toxin resistance through mutation on the potassium channel, as seen in scorpions on Kv1.3 (Zhang et al., 2016). Regarding Kv1.2 in *B. multicinctus*, mutations were located at the tip of the S1-S2 segment to which several spider toxins are known to bind (Chen and Chung, 2015). The Isoelectric Point Calculator (Kozlowski, 2016) estimated theoretical isoelectric point (p*I*) of the tip area increased from 5.91 to 10.04 due to the four mutations in the peptide sequence (Figures 2D and S1). As a result, Kv1.2 and the B-chain subunit (p*I* ∼ 8–10) may repel each other as they are both positively charged at neutral pH, suggesting that the β-bungarotoxin may not be guided to its targeting site and the snake may be less affected by the toxin if envenomed by conspecifics. Similarly, receptors targeted by the α-neurotoxin show charge reversal mutations in some snakes, whereby a negatively charged amino acid is replaced by a positively charged amino acid, which electrostatically repels the positively charged α-neurotoxins (Harris and Fry, 2021). However, the binding site of the B-chain subunit on Kv1.2 is still unclear, and further investigations are needed to confirm our hypothesis.

In *B. multicinctus*, *3FTX*, *PLA_2 I*, and *KUN* accounted for the largest proportion of toxin transcripts (Figure S2). A small proportion of other accessory venom proteins were also identified in the venom gland transcriptomes, including 5′-nucleotidase, phospholipase B-like, *VEGF* families, and so forth (Figure S2). This comprehensive catalog of actively transcribed toxin genes will provide a good reference for producing next-generation *B. multicinctus* antivenom (Table S3). In addition to the high expression of toxin genes, genes involved in protein folding and cytoplasmic transportation (i.e., *SDF2L1*, *ERP44*, *PDIA6*, *TMED7*, and *TMED10*) were also identified as hub genes in the venom gland co-expression network and may play crucial roles in toxin synthesis and secretion. ERP44 ensures correct disulfide formation during protein folding (Anelli et al., 2003), and PDIA6 mediates cellular response when unfolded proteins accumulate (Eletto et al., 2014). Thus, they may play a crucial role in maintaining normal cellular function during peak toxin synthesis.

Members of the TEMD family (e.g., TMED7 and TMED10) are important regulators of protein transport, regulating protein entry into vesicles and protein vesicular trafficking between membranes (Aber et al., 2019; Zhang et al., 2020). Therefore, cytoplasmic transportation of toxins may also be regulated by these peptides.

Limitations of the study

Although we identified actively transcribed toxin gene family members in the venom gland, we did not explore their expression via proteomic analysis. As post-transcriptional modification is common during toxin expression, some toxin genes may not be translated into proteins and thus do not function as toxins when envenomated. Therefore, our toxin gene expression profile can be different from the true cocktail of *B. multicinctus* venom. Moreover, the status of the venom glands used in our study differed, i.e., two individuals at 3 days post milking and two individuals at 14 days post milking. Thus, the current data are not suitable for time-series gene expression analysis, and more replications are necessary for such study in the future.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Genome sequencing
  - *De novo* genome assembly
  - RNA sequencing
  - Genome annotation
  - Toxin gene annotation
  - Evolutionary analyses
  - Evolutionary analysis of 3FTX family
  - Evolutionary analysis of KUN family
  - Differential expression analysis
  - Weighted gene co-expression network analysis
  - Three finger toxin structural modeling
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2022.111079.

ACKNOWLEDGMENTS

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (CAS) (XDB31000000); the National Natural Science Foundation of China (32100396; 32000296); the Key Research Program of Frontier Sciences, CAS (QYZDB-SSW-SMC058); the International Partnership Program of CAS (151751KYSB20190024); the CAS "Light of West China" Program (2018XBZG_JCTD_001); and the Sichuan Science and Technology Program (2021JDJQ0002). The authors would like to thank Jin-Long Ren, Kai Wang, Ke Jiang, and Yufan Wang for providing the photos of the snakes, and Zhongliang Peng, Dihao Wu, and Junjie Huang for providing assistance during sample processing.

AUTHOR CONTRIBUTIONS

J.-T.L. initiated and designed the study. Z.-Y.Z., Y.L., W.W., C.Y., C.-Y.T., and C.P. conducted this work. Z.-Y.Z. wrote the manuscript. J.-T.L., W.W., and Y.L. edited the manuscript. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 6, 2022
Revised: May 4, 2022
Accepted: June 20, 2022
Published: July 12, 2022

REFERENCES

Abd El-Aziz, T.M., Al Khoury, S., Jaquillard, L., Triquigneaux, M., Martinez, G., Bourgoin-Voillard, S., Seve, M., Arnoult, C., Beroud, R., and De Waard, M. (2018). Actiflagelin, a new sperm activator isolated from *Walterinnesia aegyptia* venom using phenotypic screening. J. Venom. Anim. Toxins *24*, 2. https://doi.org/10.1186/s40409-018-0140-4.

Abd El-Aziz, T.M., Soares, A.G., and Stockand, J.D. (2019). Snake venoms in drug discovery: valuable therapeutic tools for life saving. Toxins *11*, 564. https://doi.org/10.3390/toxins11100564.

Aber, R., Chan, W., Mugisha, S., and Jerome-Majewska, L.A. (2019). Transmembrane emp24 domain proteins in development and disease. Genet. Resour. *101*, e14. https://doi.org/10.1017/s0016672319000090.

Abrusán, G., Grundmann, N., DeMester, L., and Makalowski, W. (2009). TEclass—a tool for automated classification of unknown eukaryotic transposable elements. Bioinformatics *25*, 1329–1330. https://doi.org/10.1093/bioinformatics/btp084.

Almeida, D.D., Viala, V.L., Nachtigall, P.G., Broe, M., Gibbs, H.L., Serrano, S.M.d.T., Moura-da-Silva, A.M., Ho, P.L., Nishiyama-, M.Y., Jr., and Junqueira-de-Azevedo, I.L.M. (2021). Tracking the recruitment and evolution of snake toxins using the evolutionary context provided by the *Bothrops jararaca* genome. Proc. Natl. Acad. Sci. USA *118*, e2015159118. https://doi.org/10.1073/pnas.2015159118.

Anelli, T., Alessio, M., Bachi, A., Bergamelli, L., Bertoli, G., Camerini, S., Mezghrani, A., Ruffato, E., Simmen, T., and Sitia, R. (2003). Thiol-mediated protein retention in the endoplasmic reticulum: the role of ERp44. EMBO. J. *22*, 5015–5022. https://doi.org/10.1093/emboj/cdg491.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580. https://doi.org/10.1093/nar/27.2.573.

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., et al. (2021). The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. *49*, D344–D354. https://doi.org/10.1093/nar/gkaa977.

Bordon, K.d.C.F., Cologna, C.T., Fornari-Baldo, E.C., Pinheiro-Júnior, E.L., Cerni, F.A., Amorim, F.G., Anjolette, F.A.P., Cordeiro, F.A., Wiezel, G.A., Cardoso, I.A., et al. (2020). From animal poisons and venoms to medicines: achievements, challenges and perspectives in drug discovery. Front. Pharmacol. *11*, 1132. https://doi.org/10.3389/fphar.2020.01132.

Brahma, R.K., McCleary, R.J., Kini, R.M., and Doley, R. (2015). Venom gland transcriptomics for identifying, cataloging, and characterizing venom proteins in snakes. Toxicon *93*, 1–10. https://doi.org/10.1016/j.toxicon.2014.10.022.

Bryant, D., and Moulton, V. (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. *21*, 255–265. https://doi.org/10.1093/molbev/msh018.

Burge, S.W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E.P., Eddy, S.R., Gardner, P.P., and Bateman, A. (2013). Rfam 11.0: 10 years of RNA families. Nucleic Acids Res. *41*, D226–D232. https://doi.org/10.1093/nar/gks1005.

Casewell, N.R., Jackson, T.N.W., Laustsen, A.H., and Sunagar, K. (2020). Causes and consequences of snake venom variation. Trends. Pharmacol. Sci. *41*, 570–581. https://doi.org/10.1016/j.tips.2020.05.006.

Castoe, T.A., de Koning, A.P.J., Hall, K.T., Card, D.C., Schield, D.R., Fujita, M.K., Ruggiero, R.P., Degner, J.F., Daza, J.M., Gu, W., et al. (2013). The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. Proc. Natl. Acad. Sci. USA *110*, 20645–20650. https://doi.org/10.1073/pnas.1314475110.

Chan, P.P., Lin, B.Y., Mak, A.J., and Lowe, T.M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. *49*, 9077–9096. https://doi.org/10.1093/nar/gkab688.

Chang, L.S., Chung, C., Wu, B.N., and Yang, C.C. (2002). Characterization and gene organization of Taiwan banded krait (*Bungarus multicinctus*) gamma-bungarotoxin. J. Protein. Chem. *21*, 223–229. https://doi.org/10.1023/a:1019760401692.

Chen, R., and Chung, S.-H. (2015). Computational studies of venom peptides targeting potassium channels. Toxins *7*, 5194–5211. https://doi.org/10.3390/toxins7124877.

Chiappinelli, V.A., Weaver, W.R., McLane, K.E., ContiFine, B.M., Fiordalisi, J.J., and Grant, G.A. (1996). Binding of native kappa-neurotoxins and site-directed mutants to nicotinic acetylcholine receptors. Toxicon *34*, 1243–1256. https://doi.org/10.1016/s0041-0101(96)00110-9.

Chung, C., Wu, B.N., Yang, C.C., and Chang, L.S. (2002). Muscarinic toxin-like proteins from Taiwan banded krait (*Bungarus multicinctus*) venom: purification, characterization and gene organization. Biol. Chem. *383*, 1397–1406. https://doi.org/10.1515/bc.2002.158.

Dashevsky, D., and Fry, B.G. (2018). Ancient diversification of three-finger toxins in *Micrurus* coral snakes. J. Mol. Evol. *86*, 58–67. https://doi.org/10.1007/s00239-017-9825-5.

Dashevsky, D., Rokyta, D., Frank, N., Nouwens, A., and Fry, B.G. (2021). Electric blue: molecular evolution of three-finger toxins in the long-glanded coral snake species *Calliophis bivirgatus*. Toxins *13*, 124. https://doi.org/10.3390/toxins13020124.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics *22*, 1269–1271. https://doi.org/10.1093/bioinformatics/btl097.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dowell, N.L., Giorgianni, M.W., Kassner, V.A., Selegue, J.E., Sanchez, E.E., and Carroll, S.B. (2016). The deep origin and recent loss of venom toxin genes in rattlesnakes. Curr. Biol. *26*, 2434–2445. https://doi.org/10.1016/j.cub.2016.07.038.

Dress, A.W., Flamm, C., Fritzsch, G., Grünewald, S., Kruspe, M., Prohaska, S.J., and Stadler, P.F. (2008). Noisy: identification of problematic columns in multiple sequence alignments. Algorithms Mol. Biol. *3*, 7. https://doi.org/10.1186/1748-7188-3-7.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., and Aiden, E.L. (2017). *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science *356*, 92–95. https://doi.org/10.1126/science.aal3327.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016a). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. *3*, 99–101. https://doi.org/10.1016/j.cels.2015.07.012.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016b). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. *3*, 95–98. https://doi.org/10.1016/j.cels.2016.07.002.

Eletto, D., Eletto, D., Dersh, D., Gidalevitz, T., and Argon, Y. (2014). Protein disulfide isomerase A6 controls the decay of IRE1 alpha signaling via disulfide-dependent association. Mol. Cell. *53*, 562–576. https://doi.org/10.1016/j.molcel.2014.01.004.

Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome. Biol. *16*, 157. https://doi.org/10.1186/s13059-015-0721-2.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome. Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.

Fry, B.G. (2005). From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome. Res. *15*, 403–420. https://doi.org/10.1101/gr.3228405.

Fry, B.G., Undheim, E.A., Ali, S.A., Jackson, T.N., Debono, J., Scheib, H., Ruder, T., Morgenstern, D., Cadwallader, L., Whitehead, D., et al. (2013). Squeezers and leaf-cutters: differential diversification and degeneration of the venom system in toxicoferan reptiles. Mol. Cell. Proteomics. *12*, 1881–1899. https://doi.org/10.1074/mcp.m112.023143.

Fry, B.G., W⇅ster, W., Kini, R.M., Brusic, V., Khan, A., Venkataraman, D., and Rooney, A.P. (2003). Molecular evolution and phylogeny of elapid snake venom three-finger toxins. J. Mol. Evol. *57*, 110–129. https://doi.org/10.1007/s00239-003-2461-2.

Fujimi, T.J., Nakajyo, T., Nishimura, E., Ogura, E., Tsuchiya, T., and Tamiya, T. (2003). Molecular evolution and diversification of snake toxin genes, revealed by analysis of intron sequences. Gene *313*, 111–118. https://doi.org/10.1016/S0378-1119(03)00637-1.

Giorgianni, M.W., Dowell, N.L., Griffin, S., Kassner, V.A., Selegue, J.E., and Carroll, S.B. (2020). The origin and diversification of a novel protein family in venomous snakes. Proc. Natl. Acad. Sci. USA *117*, 10911–10920. https://doi.org/10.1073/pnas.1920011117.

Gong, W., Martin, T.A., Sanders, A.J., Jiang, A., Sun, P., and Jiang, W.G. (2021). Location, function and role of stromal cell-derived factors and possible implications in cancer (Review). Int. J. Mol. Med. *47*, 435–443. https://doi.org/10.3892/ijmm.2020.4811.

Grünewald, S., Forslund, K., Dress, A., and Moulton, V. (2007). QNet: an agglomerative method for the construction of phylogenetic networks from weighted quartets. Mol. Biol. Evol. *24*, 532–538. https://doi.org/10.1093/molbev/msl180.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. *31*, 5654–5666. https://doi.org/10.1093/nar/gkg770.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome. Biol. *9*, R7. https://doi.org/10.1186/gb-2008-9-1-r7.

Han, Y., and Wessler, S.R. (2010). MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res. *38*, e199. https://doi.org/10.1093/nar/gkq862.

Hannan, S., Mortensen, M., and Smart, T.G. (2015). Snake neurotoxin α-bungarotoxin is an antagonist at native GABAA receptors. Neuropharmacology *93*, 28–40. https://doi.org/10.1016/j.neuropharm.2015.01.001.

Harris, R.J., and Fry, B.G. (2021). Electrostatic resistance to alpha-neurotoxins conferred by charge reversal mutations in nicotinic acetylcholine receptors. Proc. Royal Soc. B. *288*, 20202703. https://doi.org/10.1098/rspb.2020.2703.

Harvey, A.L. (2001). Twenty years of dendrotoxins. Toxicon *39*, 15–26. https://doi.org/10.1016/s0041-0101(00)00162-8.

Hill, R.E., and Mackessy, S.P. (2000). Characterization of venom (Duvernoy's secretion) from twelve species of colubrid snakes and partial sequence of

four venom proteins. Toxicon *38*, 1663–1687. https://doi.org/10.1016/s0041-0101(00)00091-x.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–522. https://doi.org/10.1093/molbev/msx281.

Holding, M.L., Strickland, J.L., Rautsaw, R.M., Hofmann, E.P., Mason, A.J., Hogan, M.P., Nystrom, G.S., Ellsworth, S.A., Colston, T.J., Borja, M., et al. (2021). Phylogenetically diverse diets favor more complex venoms in North American pitvipers. Proc. Natl. Acad. Sci. USA *118*, e2015579118. https://doi.org/10.1073/pnas.2015579118.

Hu, J., Fan, J., Sun, Z., and Liu, S. (2020). NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics *36*, 2253–2255. https://doi.org/10.1093/bioinformatics/btz891.

Jackson, T.N.W., and Koludarov, I. (2020). How the toxin got its toxicity. Front. Pharmacol. *11*, 574925. https://doi.org/10.3389/fphar.2020.574925.

Jiang, Y., Li, Y., Lee, W., Xu, X., Zhang, Y., Zhao, R., Zhang, Y., and Wang, W. (2011). Venom gland transcriptomes of two elapid snakes (*Bungarus multicinctus* and *Naja atra*) and evolution of toxin genes. BMC Genom. *12*, 1. https://doi.org/10.1186/1471-2164-12-1.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

Jungo, F., Bougueleret, L., Xenarios, I., and Poux, S. (2012). The UniProtKB/Swiss-Prot Tox-Prot program: a central hub of integrated venom protein data. Toxicon *60*, 551–557. https://doi.org/10.1016/j.toxicon.2012.03.010.

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., et al. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome. Res. *24*, 1384–1395. https://doi.org/10.1101/gr.170720.113.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589. https://doi.org/10.1038/nmeth.4285.

Katoh, K., Rozewicki, J., and Yamada, K.D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. *20*, 1160–1166. https://doi.org/10.1093/bib/bbx108.

Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S.O., and Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. BMC. Bioinform. *19*, 189. https://doi.org/10.1186/s12859-018-2203-5.

Keilwagen, J., Wenk, M., Erickson, J.L., Schattat, M.H., Grau, J., and Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. Nucleic Acids Res. *44*, e89. https://doi.org/10.1093/nar/gkw092.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4.

Kim, D.M., and Nimigean, C.M. (2016). Voltage-gated potassium channels: a structural examination of selectivity and gating. CSH Perspect. Biol. *8*, a029231. https://doi.org/10.1101/cshperspect.a029231.

Kini, R.M., and Doley, R. (2010). Structure, function and evolution of three-finger toxins: mini proteins with multiple targets. Toxicon *56*, 855–867. https://doi.org/10.1016/j.toxicon.2010.07.010.

Kozlowski, L.P. (2016). IPC–isoelectric point calculator. Biol. Direct *11*, 55. https://doi.org/10.1186/s13062-016-0159-9.

Kuhn, P., Deacon, A.M., Comoso, S., Rajaseger, G., Kini, R.M., Usón, I., and Kolatkar, P.R. (2000). The atomic resolution structure of bucandin, a novel toxin isolated from the Malayan krait, determined by direct methods. Acta Crystallogr. Sect. D Biol. Crystallogr. *56*, 1401–1407. https://doi.org/10.1107/s0907444900011501.

Kwong, P.D., McDonald, N.Q., Sigler, P.B., and Hendrickson, W.A. (1995). Structure of beta 2-bungarotoxin: potassium channel binding by Kunitz modules and targeted phospholipase action. Structure *3*, 1109–1119. https://doi.org/10.1016/s0969-2126(01)00246-5.

Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. *35*, 3100–3108. https://doi.org/10.1093/nar/gkm160.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC. Bioinform. *9*, 559. https://doi.org/10.1186/1471-2105-9-559.

Li, W., Jaroszewski, L., and Godzik, A. (2001). Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics *17*, 282–283. https://doi.org/10.1093/bioinformatics/17.3.282.

Li, W., Jaroszewski, L., and Godzik, A. (2002). Tolerating some redundancy significantly speeds up clustering of large protein databases. Bioinformatics *18*, 77–82. https://doi.org/10.1093/bioinformatics/18.1.77.

Liang, Q., Huynh, T.M., Ng, Y.Z., Isbister, G.K., and Hodgson, W.C. (2021). In vitro neurotoxicity of Chinese krait (*Bungarus multicinctus*) venom and neutralization by antivenoms. Toxins *13*, 49. https://doi.org/10.3390/toxins13010049.

Lin, B., Zhang, J.-R., Lu, H.-J., Zhao, L., Chen, J., Zhang, H.-F., Wei, X.-S., Zhang, L.-Y., Wu, X.-B., and Lee, W.-H. (2020). Immunoreactivity and neutralization study of Chinese *Bungarus multicinctus* antivenin and lab-prepared anti-bungarotoxin antisera towards purified bungarotoxins and snake venoms. PLoS Neglect. Trop. D. *14*, e0008873. https://doi.org/10.1371/journal.pntd.0008873.

Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., Li, Z., Chen, Y., Mu, D., and Fan, W. (2013). Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. Preprint at arXiv. https://doi.org/10.48550/arXiv.1308.2012.

Lomonte, B., and Calvete, J.J. (2017). Strategies in 'snake venomics' aiming at an integrative view of compositional, functional, and immunological characteristics of venoms. J. Venom. Anim. Toxins *23*, 26. https://doi.org/10.1186/s40409-017-0117-8.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome. Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

Mao, J.-J., Norval, G., Hsu, C.-L., Chen, W.-H., and Ger, R. (2010). Observations and comments on the diet of the many-banded krait (*Bungarus multicinctus multicinctus*) in Taiwan. Reptil. Amphib. *17*, 73–76.

Mao, Y.-C., Liu, P.-Y., Chiang, L.-C., Liao, S.-C., Su, H.-Y., Hsieh, S.-Y., and Yang, C.-C. (2017). *Bungarus multicinctus multicinctus* snakebite in Taiwan. Am. J. Trop. Med. Hyg. *96*, 1497–1504. https://doi.org/10.4269/ajtmh.17-0005.

Margres, M.J., Rautsaw, R.M., Strickland, J.L., Mason, A.J., Schramer, T.D., Hofmann, E.P., Stiers, E., Ellsworth, S.A., Nystrom, G.S., Hogan, M.P., et al. (2021). The tiger rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. Proc. Natl. Acad. Sci. USA *118*, e2014634118. https://doi.org/10.1073/pnas.2014634118.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: the protein families database in 2021. Nucleic Acids Res. *49*, D412–D419. https://doi.org/10.1093/nar/gkaa913.

Mordvintsev, D.Y., Polyak, Y.L., Rodionov, D.I., Jakubik, J., Dolezal, V., Karlsson, E., Tsetlin, V.I., and Utkin, Y.N. (2009). Weak toxin WTX from *Naja kaouthia* cobra venom interacts with both nicotinic and muscarinic acetylcholine receptors. FEBS J. *276*, 5065–5075. https://doi.org/10.1111/j.1742-4658.2009.07203.x.

Mouhat, S., Andreotti, N., Jouirou, B., and Sabatier, J.-M. (2008). Animal toxins acting on voltage-gated potassium channels. Curr. Pharm. Des. *14*, 2503–2518. https://doi.org/10.2174/138161208785777441.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics *29*, 2933–2935. https://doi.org/10.1093/bioinformatics/btt509.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274. https://doi.org/10.1093/molbev/msu300.

Nirthanan, S., Charpantier, E., Gopalakrishnakone, P., Gwee, M.C., Khoo, H.E., Cheah, L.S., Bertrand, D., and Kini, R.M. (2002). Candoxin, a novel toxin from *Bungarus candidus*, is a reversible antagonist of muscle ($\alpha\beta\gamma\delta$) but a poorly reversible antagonist of neuronal α7 nicotinic acetylcholine receptors. J. Biol. Chem. *277*, 17811–17820. https://doi.org/10.1074/jbc.M111152200.

Nirthanan, S., Gopalakrishnakone, P., Gwee, M., Khoo, H., and Kini, R.M. (2003). Non-conventional toxins from Elapid venoms. Toxicon *41*, 397–407. https://doi.org/10.1016/s0041-0101(02)00388-4.

Oguiura, N., Collares, M.A., Furtado, M.F.D., Ferrarezzi, H., and Suzuki, H. (2009). Intraspecific variation of the crotamine and crotasin genes in *Crotalus durissus* rattlesnakes. Gene *446*, 35–40. https://doi.org/10.1016/j.gene.2009.05.015.

Pahari, S., Bickford, D., Fry, B.G., and Kini, R.M. (2007). Expression pattern of three-finger toxin and phospholipase A_2 genes in the venom glands of two sea snakes, *Lapemis curtus* and *Acalyptophis peronii*: comparison of evolution of these toxins in land snakes, sea kraits and sea snakes. BMC. Evol. Biol. *7*, 175. https://doi.org/10.1186/1471-2148-7-175.

Pandey, D.P., Bhattarai, P., and Piya, R.C. (2020). Food spectrum of common kraits (*Bungarus caeruleus*): an implication for snakebite prevention and snake conservation. J. Herpetol. *54*, 87. https://doi.org/10.1670/18-054.

Pasquesi, G.I.M., Adams, R.H., Card, D.C., Schield, D.R., Corbin, A.B., Perry, B.W., Reyes-Velasco, J., Ruggiero, R.P., Vandewege, M.W., Shortt, J.A., and Castoe, T.A. (2018). Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. Nat. Commun. *9*, 2774. https://doi.org/10.1038/s41467-018-05279-1.

Peng, C., Ren, J.-L., Deng, C., Jiang, D., Wang, J., Qu, J., Chang, J., Yan, C., Jiang, K., Murphy, R.W., et al. (2020). The genome of Shaw's sea snake (*Hydrophis curtus*) reveals secondary adaptation to its marine environment. Mol. Biol. Evol. *37*, 1744–1760. https://doi.org/10.1093/molbev/msaa043.

Perry, B.W., Card, D.C., McGlothlin, J.W., Pasquesi, G.I.M., Adams, R.H., Schield, D.R., Hales, N.R., Corbin, A.B., Demuth, J.P., Hoffmann, F.G., et al. (2018). Molecular adaptations for sensing and securing prey and insight into amniote genome diversity from the garter snake genome. Genome. Biol. Evol. *10*, 2110–2129. https://doi.org/10.1093/gbe/evy157.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295. https://doi.org/10.1038/nbt.3122.

Puddifoot, C.A., Wu, M., Sung, R.-J., and Joiner, W.J. (2015). Ly6h regulates trafficking of alpha7 nicotinic acetylcholine receptors and nicotine-induced potentiation of glutamatergic signaling. J. Neurosci. *35*, 3420–3430. https://doi.org/10.1523/jneurosci.3630-14.2015.

Rivera-Torres, I.O., Jin, T.B., Cadene, M., Chait, B.T., and Poget, S.F. (2016). Discovery and characterisation of a novel toxin from *Dendroaspis angusticeps*, named Tx7335, that activates the potassium channel KcsA. Sci. Rep. *6*, 23904. https://doi.org/10.1038/srep23904.

Rowan, E.G. (2001). What does beta-bungarotoxin do at the neuromuscular junction? Toxicon *39*, 107–118. https://doi.org/10.1016/s0041-0101(00)00159-8.

Roy, A., Zhou, X., Chong, M.Z., D'Hoedt, D., Foo, C.S., Rajagopalan, N., Nirthanan, S., Bertrand, D., Sivaraman, J., and Kini, R.M. (2010). Structural and functional characterization of a novel homodimeric three-finger neurotoxin from the venom of *Ophiophagus hannah* (king cobra). J. Biol. Chem. *285*, 8302–8315. https://doi.org/10.1074/jbc.M109.074161.

Schield, D.R., Card, D.C., Hales, N.R., Perry, B.W., Pasquesi, G.M., Blackmon, H., Adams, R.H., Corbin, A.B., Smith, C.F., Ramesh, B., et al. (2019). The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. Genome. Res. *29*, 590–601. https://doi.org/10.1101/gr.240952.118.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome. Biol. *16*, 259. https://doi.org/10.1186/s13059-015-0831-x.

Servent, D., WincklerDietrich, V., Hu, H.Y., Kessler, P., Drevet, P., Bertrand, D., and Ménez, A. (1997). Only snake curaremimetic toxins with a fifth disulfide bond have high affinity for the neuronal alpha 7 nicotinic receptor. J. Biol. Chem. *272*, 24279–24286. https://doi.org/10.1074/jbc.272.39.24279.

Shan, L.-L., Gao, J.-F., Zhang, Y.-X., Shen, S.-S., He, Y., Wang, J., Ma, X.-M., and Ji, X. (2016). Proteomic characterization and comparison of venoms from two elapid snakes (*Bungarus multicinctus* and *Naja atra*) from China. J. Proteomics. *138*, 83–94. https://doi.org/10.1016/j.jprot.2016.02.028.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome. Res. *13*, 2498–2504. https://doi.org/10.1101/gr.1239303.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Singh, L. (1972). Evolution of karyotypes in snakes. Chromosoma *38*, 185–236. https://doi.org/10.1007/bf00326193.

Solovyev, V. (2007). Statistical approaches in eukaryotic gene prediction. In Handbook of Statistical Genetics, D.J. Balding, M. Bishop, and C. Cannings, eds. (Wiley-Interscience), pp. 97–159.

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics *24*, 637–644. https://doi.org/10.1093/bioinformatics/btn013.

Supek, F., Bošnjak, M., Skunca, N., and Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One *6*, e21800. https://doi.org/10.1371/journal.pone.0021800.

Sunagar, K., Jackson, T.N.W., Undheim, E.A.B., Ali, S.A., Antunes, A., and Fry, B.G. (2013). Three-fingered RAVERs: rapid accumulation of variations in exposed residues of snake venom toxins. Toxins *5*, 2172–2208. https://doi.org/10.3390/toxins5112172.

Suryamohan, K., Krishnankutty, S.P., Guillory, J., Jevit, M., Schröder, M.S., Wu, M., Kuriakose, B., Mathew, O.K., Perumal, R.C., Koludarov, I., et al. (2020). The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. Nat. Genet. *52*, 106–117. https://doi.org/10.1038/s41588-019-0559-8.

Ullate-Agote, A., Burgelin, I., Debry, A., Langrez, C., Montange, F., Peraldi, R., Daraspe, J., Kaessmann, H., Milinkovitch, M.C., and Tzika, A.C. (2020). Genome mapping of a LYST mutation in corn snakes indicates that vertebrate chromatophore vesicles are lysosome-related organelles. Proc. Natl. Acad. Sci. USA *117*, 26307–26317. https://doi.org/10.1073/pnas.2003724117.

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Dev. Reprod. Biol. *8*, 77–80. https://doi.org/10.1016/s1672-0229(10)60008-3.

Wang, W., Rasmussen, T., Harding, A.J., Booth, N.A., Booth, I.R., and Naismith, J.H. (2012). Salt bridges regulate both dimer formation and monomeric flexibility in HdeB and may have a role in periplasmic chaperone function. J. Mol. Biol. *415*, 538–546. https://doi.org/10.1016/j.jmb.2011.11.026.

Wang, X., and Wang, L. (2016). GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. Front. Plant Sci. *7*, 1350. https://doi.org/10.3389/fpls.2016.01350.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033.

Wu, M., Puddifoot, C.A., Taylor, P., and Joiner, W.J. (2015). Mechanisms of inhibition and potentiation of alpha α4β2 nicotinic acetylcholine receptors by members of the Ly6 protein family. J. Biol. Chem. *290*, 24509–24518. https://doi.org/10.1074/jbc.M115.647248.

Xu, J., and Zhang, Y. (2010). How significant is a protein structure similarity with TM-score=0.5? Bioinformatics *26*, 889–895. https://doi.org/10.1093/bioinformatics/btq066.

Yang, Z.H. (2007). Paml 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088.

Ye, C., Hill, C.M., Wu, S., Ruan, J., and Ma, Z.S. (2016). DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Sci. Rep. *6*, 31900. https://doi.org/10.1038/srep31900.

Yin, W., Wang, Z.-J., Li, Q.-Y., Lian, J.-M., Zhou, Y., Lu, B.-Z., Jin, L.-J., Qiu, P.-X., Zhang, P., Zhu, W.-B., et al. (2016). Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. Nat. Commun. *7*, 13107. https://doi.org/10.1038/ncomms13107.

Yin, X., Guo, S., Gao, J., Luo, L., Liao, X., Li, M., Su, H., Huang, Z., Xu, J., Pei, J., and Chen, S. (2020). Kinetic analysis of effects of temperature and time on the regulation of venom expression in *Bungarus multicinctus*. Sci. Rep. *10*, 14142. https://doi.org/10.1038/s41598-020-70565-2.

Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC. Bioinformatics. *19*, 153. https://doi.org/10.1186/s12859-018-2129-y.

Zhang, C., Xie, Z., Li, X., Chen, J., Feng, J., Lang, Y., Yang, W., Li, W., Chen, Z., Yao, J., et al. (2016). Molecular basis for the toxin insensitivity of scorpion voltage-gated potassium channel MmKv1. Biochem. J. *473*, 1257–1266. https://doi.org/10.1042/bcj20160178.

Zhang, M., Liu, L., Lin, X., Wang, Y., Li, Y., Guo, Q., Li, S., Sun, Y., Tao, X., Zhang, D., et al. (2020). A translocation pathway for vesicle-mediated unconventional protein secretion. Cell *181*, 637–652.e15. https://doi.org/10.1016/j.cell.2020.03.031.

Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. *33*, 2302–2309. https://doi.org/10.1093/nar/gki524.

Župunski, V., and Kordiš, D. (2016). Strong and widespread action of site-specific positive selection in the snake venom Kunitz/BPTI protein family. Sci. Rep. *6*, 37054. https://doi.org/10.1038/srep37054.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Python bivittatus reference genome	Castoe et al., 2013	NCBI: GCF_000186305.1
Crotalus viridis reference genome	Pasquesi et al., 2018	NCBI: GCA_003400415.2
Deinagkistrodon acutus reference genome	Yin et al., 2016	http://gigadb.org/dataset/100196
Naja Naja reference genome	Suryamohan et al., 2020	NCBI: GCA_009733165.1
Hydrophis curtus reference genome	Peng et al., 2020	https://doi.org/10.6084/m9.figshare.11391606.v5
Pantherophis guttatus reference genome	Ullate-Agote et al., 2020	NCBI: GCF_001185365.1
Thamnophis elegans reference genome	Vertebrate Genomes Project	NCBI: GCF_009769535.1
Bungarus multicinctus reference genome	This paper	CNGBdb: CNA0045869
PacBio reads data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509256
Illumina short-reads data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509254
Illumina short-reads data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509255
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509258
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509259
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509260
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509261
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509262
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509263
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509264
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509265
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509266
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509267
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509268
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509269
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509270
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509271
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509272
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509273
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509274
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509275
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509276
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509277
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509278
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509279
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509280
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509281
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509282
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509283
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509284
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509285
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509286

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509287
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509288
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509289
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509290
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509291
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509292
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509293
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509294
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509295
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509296
RNA-seq data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509297
Hi-C data of *Bungarus multicinctus*	This paper	CNGBdb: CNR0509257
Software and algorithms		
Genomic Charactor Estimator	Liu et al., 2013	https://github.com/fanagislab/GCE
Platanus v1.2.4	Kajitani et al., 2014	http://platanus.bio.titech.ac.jp/platanus-assembler/platanus-1-2-4
DGB2OLC	Ye et al., 2016	https://github.com/yechengxi/DBG2OLC
NextPolish v1.4	Hu et al., 2020	https://github.com/Nextomics/NextPolish
HiC-Pro v2.2	Servant et al., 2015	https://github.com/nservant/HiC-Pro
Juicer v1.5	Durand et al., 2016b	https://github.com/aidenlab/juicer
3D-DNA v180922	Dudchenko et al., 2017	https://github.com/aidenlab/3d-dna
Juicebox	Durand et al., 2016a	https://github.com/aidenlab/Juicebox
BUSCO v3.1.0	Simão et al., 2015	https://busco.ezlab.org/
GMATA v2.2	Wang and Wang, 2016	https://sourceforge.net/projects/gmata/?source=navbar
Tandem Repeats Finder v4.07b	Benson, 1999	https://tandem.bu.edu/trf/trf.html
MITE-hunter	Han and Wessler, 2010	https://github.com/jburnette/MITE-Hunter
RepeatModeler v1.0.11		http://www.repeatmasker.org
RepeatMasker v1.331		http://www.repeatmasker.org
STAR v2.7.3a	Dobin et al., 2013	https://github.com/alexdobin/STAR
GeMoMa v1.6.1	Keilwagen et al., 2016, 2018	http://www.jstacs.de/index.php/GeMoMa
StringTie v1.3.4d	Pertea et al., 2015	https://ccb.jhu.edu/software/stringtie/
PASA v2.3.3	Haas et al., 2003, 2008	https://github.com/PASApipeline/PASApipeline/releases
Augustus v3.3.1	Stanke et al., 2008	https://github.com/Gaius-Augustus/Augustus
EVidenceModeler v1.1.1	Haas et al., 2008	https://evidencemodeler.github.io/
TransposonPSI v1.0.0	NA	http://transposonpsi.sourceforge.net/
InterProScan v5.32–71.0	Blum et al., 2021; Jones et al., 2014	https://github.com/ebi-pf-team/interproscan
tRNAscan-SE v2.0	Chan et al., 2021	http://lowelab.ucsc.edu/tRNAscan-SE/
Infernal cmscan v1.1.2	Nawrocki and Eddy, 2013	http://eddylab.org/infernal/
HMMER v3.1b1	NA	http://hmmer.org/
Jalview v2.11.1	Waterhouse et al., 2009	https://www.jalview.org/
MolQuest v2.4.5	Solovyev, 2007	http://www.molquest.com
OrthoFinder v2.2.7	Emms and Kelly, 2015, 2019	https://github.com/davidemms/OrthoFinder
IQ-TREE v1.6.574	Nguyen et al., 2015	http://www.iqtree.org/
PAML 4.7	Yang, 2007	http://abacus.gene.ucl.ac.uk/software/paml.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CAFE v3.1	De Bie et al., 2006	https://hahnlab.github.io/CAFE/src_docs/html/index.html
REViGO	Supek et al., 2011	http://revigo.irb.hr/
ParaAT v2.0	NA	https://github.com/jdebarry/paraat/tree/master/ParaAT2.0
KaKs_Calculator 2.0	Wang et al., 2010	https://github.com/kullrich/kakscalculator2
HISAT2 v2.1.0	Kim et al., 2019	http://daehwankimlab.github.io/hisat2/
DESeq2 v1.26.0	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
WGCNA package	Langfelder and Horvath, 2008	https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/
Cytoscape v3.7.2	Shannon et al., 2003	https://cytoscape.org/cy3.html
AlphaFold2 v2.0.0	Jumper et al., 2021	https://github.com/deepmind/alphafold
TM-align	Zhang and Skolnick, 2005	https://zhanggroup.org/TM-align/
TEclass	Abrusán et al., 2009	https://www.compgen.uni-muenster.de/tools/teclass/index.hbi?
ASTRAL v5.6.3	Zhang et al., 2018	https://github.com/smirarab/ASTRAL
CD-HIT v4.6	Li et al., 2001, 2002	http://weizhong-lab.ucsd.edu/cd-hit/
Noisy v1.5.12	Bryant and Moulton, 2004; Dress et al., 2008; Grünewald et al., 2007	http://www.bioinf.uni-leipzig.de/Software/noisy/
Isoelectric Point Calculator	Kozlowski, 2016	http://isoelectric.org/calculate.php
RNAmmer v1.2	Lagesen et al., 2007	https://services.healthtech.dtu.dk/service.php?RNAmmer-1.2
Other		
UniProtKB/Swiss-Prot Tox-Prot program	Jungo et al., 2012	http://www.uniprot.org/program/Toxins
Original 3FTXs alignment	This paper	https://doi.org/10.6084/m9.figshare.19698985
The in-house perl script used for enrichment analysis	This paper	https://doi.org/10.6084/m9.figshare.19881574
Genome annotation of *Bungarus multicinctus*	This paper	https://doi.org/10.6084/m9.figshare.19915051

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jia-Tang Li (lijt@cib.ac.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The genome assembly of *B. multicinctus*, as well as all PacBio reads, Illumina short-reads and Hi-C data have been deposited at the China National GeneBank DataBase (CNGBdb) under project accession number CNP0002662, and are publicly available as of the date of publication. Accession numbers of the data files are listed in the key resources table. The genome annotation file has been deposited at figshare and is publicly available as of the date of publication. The DOI is listed in the key resources table.
- The original alignment used for studying the phylogeny of 3FTXs and all original code have been deposited at figshare and are publicly available. The DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

A total of nine adult individuals of *B. multicinctus* - two males from Fujian (BM000 and BM001), as well as two females and five males from Zhejiang (BM002-BM008) - were collected and processed in accordance with the approval of Animal Experiment Ethics Committee in Chengdu Institution of Biology, Chinese Academy of Sciences. A summary of tissues obtained from each animal is shown in Table S5.

METHOD DETAILS

Genome sequencing

High-molecular weight DNA were extracted from the liver of a male animal (BM000). DNA quantity, purity and integrity were assessed by Qubit fluorometer, Nanodrop spectrophotometer and pulse-field gel electrophoresis respectively.

Three types of whole genome sequencing data were generated: Illumina short-reads, PacBio long-reads and Hi-C. For short-reads sequencing, libraries with 350 insert size were prepared as per the manufactures' instructions and sequenced on an Illumina NovaSeq 6000 platform using 2 × 150 bp paired-end sequencing. Approximately 141.23 Gb of raw data were generated (Table S1). After filtering out adapters, low-quality bases and reads, 135.04 Gb of clean data were retained (Table S1). The K-mer frequencies were calculated for paired-end short reads with GCE (genomic charactor estimator) (Liu et al., 2013) using K = 17. Based on the K-mer distribution the estimated genome size was ~1.62 Gb.

For long-reads sequencing, libraries with 20 kb insert size were prepared following the manufactures' protocol and sequenced using PacBio Sequel II platform. A total of 6,872,388 subreads and 158.62 Gb data were generated (Table S1), giving an overall coverage of ~98x based on the estimated genome size. The mean and N50 lengths of the subreads were 23.1 kb and 33.7 kb respectively (Table S1).

One Hi-C libraries was prepared using venom gland tissue from a male animal (BM001). Briefly, the tissue was cross-linked with formaldehyde and quenched with glycine. The cross-linked tissue was subsequently lysed to extract the nuclei. Fixed chromatin was then extracted and digested. The cohesive ends were filled in by adding biotinylated nucleotides, and free blunt ends were ligated. After ligation, the cross-linking was reversed and the DNA purified to remove proteins. Purified DNA was then sheared to a length of ~400 bp and point ligation junctions were pulled down. Afterward, the Hi-C library for BGI sequencing was prepared according to manufacturers' instructions. The final library was sequenced on the BGI MGISEQ-2000 platform with PE150 mode to generate 689 million 150 bp paired-end reads, which provided ~128x (206.87 Gb) physical coverage of the genome.

De novo genome assembly

Firstly, reads from Illumina sequencing were assembled into contigs using Platanus v1.2.4 with the following optimized parameters: -k 29 -d 0.3 -t 16 -m 300 (Kajitani et al., 2014). The contigs were then aligned against the PacBio reads with DGB2OLC to generate consensus contigs (Ye et al., 2016). Finally, to improve the accuracy of the assembly, the resulting consensus was subjected to three iterations of polishing using NextPolish v1.4 (Hu et al., 2020).

The chromosome-scale assembly was achieved using the Hi-C data. Read pairs generated from the Hi-C library were firstly mapped onto the polished assembly, and the valid read pairs accepted by HiC-Pro v2.2 were retained for further analysis (Servant et al., 2015). With the Juicer v1.5 + 3D-DNA v180922 pipeline (Dudchenko et al., 2017; Durand et al., 2016b), valid reads were mapped to the assembly again and the alignment information was used for scaffolding the contigs. The assembly error occurred during the Hi-C scaffolding was further corrected in visual by using Juicebox (Durand et al., 2016a). Finally, we successfully clustered 1290 contigs into 831 scaffolds, with the N50 lengths increased from approximately 14 to 135 Mb (Table S1), and the number of scaffolds longer than 10 Mb presumed to be at the chromosome level became close to the actual number of chromosomes (Table S1; Singh 1972).

Genome assembly completeness was assessed by using gene sets of benchmarking universal single-copy orthologs (BUSCO v3.1.0) with genome mode and lineage data from vertebrata (Simão et al., 2015).

RNA sequencing

Total RNA was extracted from the venom gland, brain, vomeronasal organ, olfactory bulb, tongue, heart, muscle, lung, liver, pancreas, small intestine, kidney and gonads from different individuals (Table S5) by using a standard TRIzol method. RNA quantity, purity and integrity were assessed by Qubit 2.0 fluorometer, Nanodrop spectrophotometer and Agilent 2100 Bioanalyzer respectively. PolyA RNA-sequencing libraries were prepared and sequenced on a NovaSeq 6000 platform to generate ~6 Gb of raw RNA-seq data for each sample. After filtering out adapters and low-quality reads (number of N base >10% of the read length, number of bases with quality value smaller than five >50% of the read length), clean data were used for following analyses.

Genome annotation

Firstly, the simple repeat sequences (SSRs) and all tandem repeat elements in the whole genome were annotated using the software GMATA v2.2 (Wang and Wang 2016) and Tandem Repeats Finder (TRF) v4.07b (Benson 1999) respectively. Transposable elements (TE) in the *B. multicinctus* genome were then identified using a combination of *ab initio* and homology-based methods. Briefly, an *ab initio* repeat library for the species was first predicted using MITE-hunter (Han and Wessler 2010) and RepeatModeler v1.0.11

(http://www.repeatmasker.org) with default parameters. The library was then aligned to TEclass Repbase (Abrusán et al., 2009) to identify the type of repeat families. Moreover, RepeatMasker v1.331 (http://www.repeatmasker.org) was applied to search for known and novel TEs by mapping sequences against the *ab initio* repeat library and Repbase TE library. Finally, overlapping results from the two methods were collated and retained. RepeatMasker v1.331 built-in scripts were used to estimate the divergence level between the individual TE copies versus their consensus sequences based on CpG adjusted Kimura distance.

Three approaches, including *ab initio* prediction, homology search, and transcriptome-based prediction, were used independently for gene prediction in a repeat-masked genome. In homology search, GeMoMa v1.6.1 (Keilwagen et al., 2016, 2018) was used to align the homologous peptides from related species to the assembly to get the gene structure information. For RNAseq-based gene prediction, filtered RNA-seq reads (including sequencing data from venom gland and other tissues) were aligned to the reference genome using STAR v2.7.3a (Dobin et al., 2013) with default settings. The transcripts were then assembled using StringTie v1.3.4d (Pertea et al., 2015) and open reading frames (ORFs) were predicted using PASA v2.3.3 (Haas et al., 2003, 2008). For the *de novo* prediction, RNAseq reads were firstly *de novo* assembled by StringTie v1.3.4d and analyzed with PASA v2.3.3 to produce a training set. The training set was then applied to train Augustus v3.3.1 (Stanke et al., 2008) iteratively for gene prediction. Finally, EVidenceModeler (EVM) v1.1.1 (Haas et al., 2008) was used to produce an integrated gene set. Predicted genes with TE were removed by TransposonPSI v1.0.0 (http://transposonpsi.sourceforge.net/) and the miscoded genes were further filtered. Untranslated regions (UTRs) and alternative splicing regions were determined using PASA v2.3.3 based on RNA-seq assemblies. The longest transcripts for each locus was retained, and regions outside of the ORFs were designated as UTRs.

Public databases including SwissProt, NR, KEGG, KOG and Gene Ontology were used to acquire gene function and protein domain/motif information. In brief, the putative domains and GO terms of genes were identified using the InterProScan v5.32–71.0 (Blum et al., 2021; Jones et al., 2014) with default parameters. BLASTp v2.7.1 was used to align the EVM-integrated protein sequences against the other four database with an E value cutoff of 1×10^{-5}, and the hit with the lowest E value was retained. At last, results from the five database searches were integrated.

Non-coding RNAs including transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) were also annotated. tRNAs were predicted using tRNAscan-SE v2.0 (Chan et al., 2021) with eukaryote parameters, and other non-coding RNAs were detected by searching against Rfam database with Infernal cmscan v1.1.2 (Burge et al., 2013; Nawrocki and Eddy 2013). The rRNA subunits were predicted using RNAmmer v1.2 (Lagesen et al., 2007).

Toxin gene annotation

Because venom-gene families are known to occur in large tandem arrays which lowers the accuracy of *ab initio* gene prediction (Margres et al., 2021), additional annotation steps were performed for venom genes. We focused on three-finger toxin (3FTX), type I phospholipase A_2 (PLA_2 I) and snake venom Kunitz-type serine protease inhibitor (KUN) toxin families, since they can make up more than 90% of the *B. multicinctus* venom component (Jiang et al., 2011; Shan et al., 2016; Yin et al., 2020). Take 3FTX family as an example, we firstly downloaded HMM model of the toxin family (Toxin_TOLIP, PF00087) from Pfam database v35.0 (Mistry et al., 2021), and then searched against annotated proteins of the *B. multicinctus* using hmmsearch in HMMER v3.1b1 (http://hmmer.org/). Sequences with E-values below the default inclusion threshold were retrieved and identified as toxin homologs if the BLAST hit was to a known 3FTX sequence (E-value < 1e-05). The complete coding sequences (CDS) of these toxins were retrieved and aligned by Jalview v2.11.1 after translation (Waterhouse et al., 2009). Then we identified conserved CDS region (which encodes signal peptide of 3FTX) based on the alignment and built a nucleotide HMM model specifically for *3FTXs* using hmmbuild in HMMER v3.1b1. This HMM model was then used to query the genome by nhmmer in HMMER v3.1b1 with –max option, so that to locate the genomic fragment that may belong to a *3FTX*. We extracted the target fragment and its 4 kb flanking sequences (empirically, *3FTX* is shorter than 4 kb) and then queried the whole region by known 3FTX sequences using FGENESH+ in MolQuest v2.4.5 (Solovyev 2007; http://www.molquest.com). Each gene re-annotated in this way was again searched against NCBI to confirm its identity and further confirmed its homology through phylogenetic analysis. The above re-annotation procedure was repeated to identify *3FTXs* in other seven snake genomes (Table S5).

Other toxin genes were confirmed by combining the results of original gene functional annotation and differential expression analysis described below. Briefly, genes up-regulated significantly in venom gland or belong to GO term "toxin activity (GO:0090729)" were picked out and blast against NCBI to confirm their identity. Genes were identified as toxin homologs if the BLAST hit was to a known toxin sequence (E-value < 1e-05).

Evolutionary analyses

Phylogenetic analyses

OrthoFinder v2.2.7 (Emms and Kelly 2015, 2019) was applied to identify single-copy gene families from the genome assembly of *B. multicinctus* and seven other snakes (*P. bivittatus*, *C. viridis*, *D. acutus*, *N. Naja*, *H. curtus*, *P. guttatus* and *T. elegans*) downloaded from public database (Table S5). For phylogenetic tree construction, CDS alignments of each single-copy gene family were concatenated to construct a super-gene for each species, which were partitioned corresponding to the 1st, 2nd and 3rd codon site in the CDS. A maximum-likelihood tree was then constructed using IQ-TREE v1.6.574 (Nguyen et al., 2015) with parameters: -nt AUTO -st DNA -bb 2000 -alrt 2000. We also built 8,690 gene trees based on all single-copy genes using IQ-TREE with the parameters listed above. The resulting gene trees were processed by ASTRAL v5.6.3 to generate a species tree (Zhang et al., 2018). Species trees

generated by the two methods were compared, and the one with the best topology was chosen for estimating divergence times under a relaxed clock model using the MCMCTREE program in the PAML 4.7 package (Yang 2007).

Gene family analysis

When the number of genes within an orthogroup defined by OrthoFinder was more than two and all of them came from one species, these genes were considered as species-specific. Gene functions that were over-represented significantly (adjusted $p < 0.05$) among *B. multicinctus* specific genes were identified by conducting a hypergeometric test with the GO and KEGG annotations of the species' genes as the reference. The family expansion or contraction analysis was performed by CAFE v3.1 (De Bie et al., 2006) with the phylogenetic tree generated above as input. The expanded and contracted gene families on each branch of the tree were detected by comparing the cluster size of each branch with the maximum-likelihood cluster size of the ancestral node leading to that branch; a smaller ancestral node indicates gene family expansion, whereas a larger ancestral node indicates family contraction. The overall p value (family-wide p value in CAFE v3.1 based on Monte Carlo resampling) of each branch and node was then calculated, and the exact p values (Viterbi method in CAFE v3.1) of each significant overall p value (<0.01) were also calculated. The gene family with a p value < 0.01 was defined as significant expansion or contraction. Significantly (adjusted p value < 0.05) over-represented GO terms among expanded families were identified using in-house perl scripts (refer to "Data and Code Availability"). REViGO (Supek et al., 2011) was then used to cluster the top 50 overrepresented GO terms.

QEGs and PSGs analysis

Single copy gene families were picked out from the OrthoFinder v2.2.7 results described above for identifying positively selected genes (PSGs) and genes with high rates of molecular evolution (quickly evolving genes, QEGs). To identify QEGs, we firstly used the branch model of CODEML in PAML v4.7, with *B. multicinctus* being set as the foreground branch and the others as background branches. The null hypothesis was that the ω value of each branch was equal, and the alternative hypothesis was that the ω values on the foreground branch was not equal to these of the background branches. Then, a likelihood ratio test (LRT) was performed after correcting the p values using the FDR test with Bonferroni correction. Identified genes were considered QEGs when the ω value on the foreground branch was greater than that on the background branches, and the corrected p value < 0.05.

A branch-site model of CODEML in PAML v4.7 was used to find PSGs. The null hypothesis was that the ω value of each site on each branch was ≤ 1, whereas the alternative hypothesis was that the ω values of particular sites on the foreground branch were >1. A likelihood ratio test was then performed: the null distribution was a 50:50 mixture of $\chi 2$ distributions with 1 degree of freedom and a point mass at zero. The p values calculated based on this mixture distribution were further corrected for multiple testing by conducting an FDR test with a Bonferroni correction. According to the Bayes empirical Bayes inference, the PSGs should met the requirements of a corrected p value < 0.05 and contained at least one positively selected site with a posterior probability >0.95.

Evolutionary analysis of 3FTX family

As one of the most abundant and diversified toxin families in snakes, the phylogeny of 3FTXs were studied based on eight snakes' genomes. Firstly, the syntenic *3FTX* gene clusters were determined for each snake by locating the flanking conserved genes. The flanking *LY6E*s are members of Ly6/uPAR superfamily, to which 3FTX family also belongs (Puddifoot et al., 2015; Wu et al., 2015), we therefore examined the evolutionary relatedness between *3FTX*, *LY6E* and the other two possible *3FTX* ancestors, *LYNX1* and *SLURP2* (Fry, 2005). We used brown rat (*Rattus norvegicus*) LYNX1 (GenBank: NP_001124018.1) and SLURP2 (GenBank: NP_001124023.1) as the query to search the nr, nr/nt, est and TSA databases of birds, reptiles and amphibians using BLASTP/tBLASTn (E value <10), and further tBLASTn against (E value <10) the genomes of multiple taxa (16 amphibians, 29 reptiles and 20 avians; Table S5). Although some sequences of birds and reptiles found on NCBI were named LYNX1, they actually did not form a monophyletic group with the true (i.e mammalian) LYNX1 when we reconstructed their phylogeny with different Ly6/uPAR superfamily members using IQ-TREE (-bb 30,000 -bnni), which suggested that these genes were not LYNX1 orthologs (Figure S9). Since no real orthologs of mammalian LYNX1 and SLURP2 were found in birds, reptiles and amphibians, only LYNX1 and SLURP2 from mammals were used for following analysis.

3FTXs of different subgroups were downloaded from UniProtKB/Swiss-Prot Tox-Prot program (Jungo et al., 2012; http://www.uniprot.org/program/Toxins). From each subgroup, a representative 3FTX was chosen as the query to search the GenBank nr/nt, est and TSA databases using tBLASTn. For each BLAST search, top 500 significant (E value < 1e-05) matches were downloaded and translated. After all the target sequences were integrated, CD-HIT v4.6 (threshold 90% similarity) was used to identify and remove redundant sequences (Li et al., 2001, 2002). Together with sequences downloaded from UniProtKB/Swiss-Prot and GenBank, all annotated 3FTX peptides were aligned using MAFFT with L-INS-i mode and BLOSUM45 matrix (Katoh et al., 2019; Waterhouse et al., 2009). The alignment was trimmed by using Noisy v1.5.12 (Bryant and Moulton 2004; Dress et al., 2008; Grünewald et al., 2007), and a maximum likelihood tree was constructed using IQ-TREE v1.6.5 with the best substitution model determined by ModelFinder (Kalyaanamoorthy et al., 2017; Nguyen et al., 2015). Fifteen thousand ultrafast bootstrap (UFBoot) replicates with nearest neighbor interchange (NNI) optimization were performed to assess node support (Hoang et al., 2018). LYPD2, another member of Ly6/uPAR superfamily, was used as the outgroup.

In order to examine the selective pressure on *3FTXs* in studied species, we calculated Ka/Ks ratio for each gene. Firstly, ancestral sequences for each subfamily was reconstructed using IQ-TREE v2.0 (Nguyen et al., 2015), and homologs were then aligned against their corresponding ancestors using ParaAT v2.0 (https://github.com/jdebarry/paraat/tree/master/ParaAT2.0), from which the output was further processed by KaKs_Calculator 2.0 (Wang et al., 2010) to calculate Ka/Ks values.

Evolutionary analysis of KUN family

*KUN*s have expanded greatly with very high gene copy number in *B. multicinctus*, we thus investigated its evolution in snakes. Previous study showed that in some species a WAP domain can be present in *KUN* forming a KU-WAP protein (Župunski and Kordiš 2016). So we used a KU-WAP protein (UniProtKB: A5X2X1) from massasauga (*Sistrurus catenatus*) together with two typical KUNs (one is beta-bungarotoxin B-chain encoding gene) as queries to search GenBank nr, nr/nt, est and TSA databases using BLASTP/tBLASTn (E value < 1e-05). For each BLAST search, top 500 significant (E value < 1e-05) matches were downloaded and translated. KUNs on UniProtKB/Swiss-Prot were downloaded and integrated with the GenBank dataset, then CD-HIT v4.6 (threshold 90% similarity) was used to identify and remove redundant sequences. Together with downloaded sequences, all annotated KUNs were aligned using MAFFT with L-INS-i mode. The alignment was first manually checked to remove non-KUN homologs that also contain Kunitz domain (e.g. kunitz-type protease inhibitor 1, kunitz 1), and then trimmed by using Noisy v1.5.12. The processed alignment was submitted to IQ-TREE v1.6.5 to construct a maximum likelihood tree based on the best substitution model determined by ModelFinder. Fifteen thousand ultrafast bootstrap (UFBoot) replicates with nearest neighbor interchange (NNI) optimization were performed to assess node support. Kunitz 3, another KUN homolog, was used as the outgroup.

Differential expression analysis

HISAT2 v2.1.0 (Kim et al., 2019) was used to align the quality controlled RNAseq reads to the *B. multincintus* genome. Aligned reads were assembled into transcripts and expression was quantified by using StringTie v2.0.4. Differential expression analysis was conducted by using DESeq2 v1.26.0 (Love et al., 2014) to identify differentially expressed genes for each tissue type. Genes with an adjusted p value <0.05 and a log-fold change >2 were considered to express differentially. Differentially expressed *3FTXs* were visualized by using pheatmap package v1.0.12 in R across all samples with average TPM (transcripts per kilobase million) values normalized by *Z* score. $Log_2(n+1)$ transformed average TPM values were used to better illustrate the difference in expression among *KUN*s within the venom gland. Functional enrichment analysis was conducted for genes significantly up-regulated in venom gland by using an in-house perl scripts based on hypergeometric distribution approach (refer to "Data and Code Availability").

Weighted gene co-expression network analysis

WGCNA package (Langfelder and Horvath 2008) in R was used to conduct weighted gene co-expression analysis. Normalized counts (the raw counts divided by the size factors estimated by DESeq2) were used as input. After checking for excessive missing values and identification of outlier samples, a soft threshold was estimated (power = 4; Figures S7D and S7E). Based on the selected soft threshold, unsigned co-expression networks were constructed. To identify modules of highly connected genes, similar networks were merged with parameters "minModuleSize = 50, pamRespectsDendro = FALSE, mergeCutHeight = 0.25". Finally, modules closely related (cor >0.90, p < 0.05) to each tissue were selected and functional enrichment analysis was conducted for genes in the selected modules by using the above in-house perl scripts (refer to "Data and Code Availability"). Hubgenes in the co-expression network were identified using Cytoscape v3.7.2 (Shannon et al., 2003) based on the degree method.

Three finger toxin structural modeling

Three-dimensional structural models of 3FTXs were predicted with whole sequence using AlphaFold2 v2.0.0 in casp14 mode (Jumper et al., 2021). Pairwise TM-scores of the predicted tertiary structures were calculated with TM-align (Xu and Zhang 2010; Zhang and Skolnick 2005), based on which the intra-subfamily similarity were compared against that of inter-subfamily using independent two sample t test to examine if there is a variation in conformation between the subfamilies.

QUANTIFICATION AND STATISTICAL ANALYSIS

Identification of expanded gene families, as well as QEGs and PSGs were done with CAFE v3.1 (De Bie et al., 2006) and PAML v4.7 (Yang 2007) respectively. Each tool implements their own method to assess significance, as described in the "Method details" section. All enrichment analyses described in the "Method details" section were performed with an in-house perl scripts (refer to "Data and Code Availability"), which implements a hypergeometric test. p values for enrichment tests were corrected with the Benjamini-Hochberg method. Ka/Ks ratio and related p value for each *3FTX* was calculated with KaKs_Calculator 2.0 (Wang et al., 2010) as described in the "Method details" section. Differential expression analyses and WGCNA were done with DESeq2 v1.26.0 (Love et al., 2014) and WGCNA package (Langfelder and Horvath 2008) in R respectively. Each tool implements their own method to assess significance, as described in the "Method details" section. For structural similarity comparisons between 3FTXs, intra-subfamily structural similarity was compared against that of inter-subfamily using independent two sample t test (one-tailed) in R, as described in the "Method details" section. Shapiro-Wilk normality test and F-test were conducted in R to confirm that our data have met normal distribution and equal variance assumption of the t test respectively. Intra-subfamily structural similarity is defined as the sum of pairwise TM-scores between MKA-3FTXs divided by the number of pairwise comparisons between MKA-3FTXs. Inter-subfamily structural similarity is defined as the sum of pairwise TM-scores between MKA and MKT-3FTXs divided by the number of pairwise comparisons between the two subfamilies. p value <0.05 indicates that intra-subfamily similarity of MKA-3FTXs was significantly higher than inter-subfamily similarity between MKA- and MKT-3FTXs.