

RESEARCH

Open Access



# Integrative multi-omics analysis reveals the contribution of neoVTX genes to venom diversity of *Synanceia verrucosa*

Zhiwei Zhang<sup>1†</sup>, Qian Li<sup>1,2†</sup>, Hao Li<sup>1</sup>, Shichao Wei<sup>1</sup>, Wen Yu<sup>1</sup>, Zhaojie Peng<sup>1</sup>, Fuwen Wei<sup>1,3,4</sup> and Wenliang Zhou<sup>1\*</sup>

## Abstract

**Background** Animal venom systems are considered as valuable model for investigating the molecular mechanisms underlying phenotypic evolution. Stonefish are the most venomous and dangerous fish because of severe human envenomation and occasionally fatalities, whereas the genomic background of their venom has not been fully explored compared with that in other venomous animals.

**Results** In this study, we followed modern venom pipelines to decode the *Synanceia verrucosa* venom components. A catalog of 478 toxin genes was annotated based on our assembled chromosome-level genome. Integrative analysis of the high-quality genome, the transcriptome of the venom gland, and the proteome of crude venom revealed mechanisms underlying the venom complexity in *S. verrucosa*. Six tandem-duplicated neoVTX subunit genes were identified as the major source for the neoVTX protein production. Further isoform sequencing revealed massive alternative splicing events with a total of 411 isoforms demonstrated by the six genes, which further contributed to the venom diversity. We then characterized 12 dominantly expressed toxin genes in the venom gland, and 11 of which were evidenced to produce the venom protein components, with the neoVTX proteins as the most abundant. Other major venom proteins included a presumed CRVP, Kuntiz-type serine protease inhibitor, calglandulin protein, and hyaluronidase. Besides, a few of highly abundant non-toxin proteins were also characterized and they were hypothesized to function in housekeeping or hemostasis maintaining roles in the venom gland. Notably, gastrotropin like non-toxin proteins were the second highest abundant proteins in the venom, which have not been reported in other venomous animals and contribute to the unique venom properties of *S. verrucosa*.

**Conclusions** The results identified the major venom composition of *S. verrucosa*, and highlighted the contribution of neoVTX genes to the diversity of venom composition through tandem-duplication and alternative splicing. The diverse neoVTX proteins in the venom as lethal particles are important for understanding the adaptive evolution of *S. verrucosa*. Further functional studies are encouraged to exploit the venom components of *S. verrucosa* for pharmaceutical innovation.

**Keywords** Multi-omics, Venom diversity, *Synanceia verrucosa*, Tandem duplication, Alternative splicing

<sup>†</sup>Zhiwei Zhang and Qian Li contributed equally to this work.

\*Correspondence:  
Wenliang Zhou  
zhouwl@gmlab.ac.cn

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Background

Venom is one of the most successful convergent evolutionary novelties across all major animal lineages for defense, predation, and competition [1]. Animal venom systems are considered as valuable models for investigating the molecular mechanisms underlying phenotypic evolution, and are widely used for exploring and engineering biologically active molecules for pharmaceutical innovation [2].

Modern venomomics involves genetics, evolution, ecology as well as translational studies, facilitating our understanding of the venomous animals including scorpions, snakes and spiders [3, 4]. Most of the animal venoms are highly complexed mixtures of molecules with modifications at the genomic, transcriptomic, and protein level. Recently, the emerging chromosomal level genomes have greatly facilitated venom toxin screening and gene identification in snake [5–7], spider [8, 9], and other venomous animals [10–13]. Large-scale comparative genomic analyses provide evolutionary views of toxin genes, which include co-option of pre-existing nontoxic genes, massive tandem duplication and domain loss [14–17]. Despite the varying venom composition and apparatus spanning venomous metazoans, a comparative analysis of 20 venom gland transcriptomes from wasps, spiders, fish and snakes revealed that a conserved genetic toolkit was recruited to evolve the ability to secrete toxins [18]. On the other hand, full-length transcriptomic analysis revealed that more than half of the venom genes exhibited multiple isoforms within the parasitoid wasp venom gland, suggesting that alternative splicing highly contributes to venom diversity and evolution [19]. In the future, integrative genomic, transcriptomic and proteomic approaches will contribute to novel toxin discovery and the elucidation of venom evolution.

Approximately 2500 fish species are venomous across the world, whereas most of their venom are left as under investigation and valuable unexplored resource [20]. The stonefish (Scorpaeniformes, Synanceiidae, *Synanceia*) are recognized as the most venomous and dangerous fish because of severe human envenomation and occasional fatalities [21, 22]. There are five species of stonefish, including *S. verrucosa*, *S. horrida*, *S. alula*, *S. nana*, and *S. platyrhyncha*. Along with the development of venomomic approaches, attempts have been devoted to decode the stonefish venom composition and functional activities [23]. The inceptive research in stonefish venom was predominantly reported in *S. horrida*, with the first fish-derived protein toxin characterized as stonustoxin (SNTX), which has inspired the venom research in the teleost [24, 25]. Venomous fish usually use their venom as defensive weapon against native predator by causing excruciating pain. Structure analysis showed that SNTX subunits belongs to the perforin superfamily of

pore-forming immune effectors, which accounts for its hemolytic activity [26].

As the most widely distributed stonefish species throughout the world, *S. verrucosa* demonstrated great potential in understanding venom pathology and toxin gene evolution in fish [27, 28]. The lethal components were purified in the venom in the early 1990s, including a tetramer protein VTX and a dimer neoVTX. Homologous to SNTX, VTX/neoVTX demonstrated similar hemolytic, hypotensive, and cardiotoxic activities [29–31]. Recently, a pathophysiological study showed that the cardiotoxicity of the *S. verrucosa* venom is in part due to binding to the voltage-gated  $\text{Ca}^{2+}$  channel, indicating a multifunctional activity of VTX/neoVTX as fish cytotoxin [32, 33]. Besides, two widely abundant components in the animal venom, hyaluronidase [34] and lectin [35], have also been characterized in *S. verrucosa* venom. Studies to date have been limited to single particle-level purification and molecular cloning, leaving most of the venom components remained uncharacterized, which limits our understanding of how venom exerts its poisoning action and how the different components contribute to the physiological effects of envenomation.

Two criteria have been widely adopted to identify major functional venom components in animals: (1) genes should exhibit highly reliable expression in the venom gland; (2) highly expressed genes should be further characterized in the venom proteome. In this study, we assembled the chromosomal level genome of *S. verrucosa* with a set of toxin genes preliminarily annotated. The major venom protein components were then deciphered and characterized through a comprehensive analysis of genomic, transcriptomic, and proteomic data. Notably, we revealed a mechanism of tandem-duplicated neoVTX genes that underwent massive alternative splicing, which underlines the venom diversity and evolution in *S. verrucosa*.

## Methods

### Animals

Adult *S. verrucosa* fish were purchased from a local fishery market and maintained in the laboratory aquarium system for at least 2 weeks before sampling. The fish were anesthetized using MS222 (General reagent, China) before dissection. Muscle, skin, and venom glands were dissected and rinsed with autoclaved artificial seawater, followed by flash-freezing in liquid nitrogen. The samples were then stored at  $-80^{\circ}\text{C}$ . All the experiments were conducted in accordance with guidelines approved by the Animal Ethical and Welfare Committee of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou).

### Genomic DNA extraction and sequencing

Genomic DNA was extracted from the muscle using the DNeasy Blood and Tissue Kit (Qiagen, USA). The integrity, purity, and concentration of the extracted DNA were analyzed by gel electrophoresis and NanoDrop™ (Thermo Fisher Scientific, USA). The DNA concentration was further quantified using a Qubit® 4.0 Fluorometer (Invitrogen, USA). A total of 15 µg DNA was then fragmented for HiFi sequencing library construction and the library was then sequenced using the PacBio Sequel II series sequencer (Pacific Biosciences, CA, USA). Whole genome sequencing (WGS) was included to facilitate genome assembly. Briefly, a library with a 300-bp inserted size was prepared using genomic DNA isolated from muscle following the protocol of the MGIEasy FS DNA Library Prep Kit (MGI Tech, China) and sequenced on the DNBSEQ-T7 platform (MGI Tech, China).

For Hi-C sequencing, the fresh muscle tissue was fixed with 2% formaldehyde to enable cross-linking between genomic DNA and chromatin. The crosslinked DNA was then extracted, and a library with a 150-bp inserted size was constructed. Sequencing was performed using the DNBSEQ T7 platform (MGI Tech, China).

### RNA extraction and sequencing

Total RNA was isolated from the muscle, skin, and venom glands ( $n=7$ ) using TRIzol reagent (Thermo Fisher Scientific, USA). The concentration of the RNA was analyzed by NanoDrop™ and a Bioanalyzer 2000 (Agilent, USA) was used to assess integrity. For next-generation sequencing (NGS), the cDNA libraries were constructed following the manufacturer's instructions and sequenced using the MGI DNBSEQ-T7 sequencing platform. Venom gland full-length transcriptome sequencing was performed using the ISO-Seq PacBio platform (Pacific Biosciences, CA, USA).

### Genome assembly and annotation

To generate a chromosomal level genome of *S. verrucosa*, a high-quality contig genome assembly was generated using Hifiasm v0.16.1 [36], followed by redundant sequence removal using Purge\_dups v1.2.5 [37]. The assembled contigs were then aligned to the Hi-C sequencing data of *S. verrucosa* with default parameters using chromap [38], followed by genome assembly with default parameters using the yaha pipeline [39]. Finally, the completeness of the genome assembly was assessed using BUSCO v5.4.7 with genome mode and referring to the database of actinopterygii\_odb10 [40].

Repetitive sequences in the genome were screened using RepeatModeler v2.0 [41]. The *S. verrucosa* genome sequences were then mapped against *de novo* repeat libraries and the Repbase TE library v16.02 to identify transposable elements (TEs) using RepeatMasker v4.1.1

[42], and TE-related proteins were further examined using RepeatProteinMask v.4.0.6 [43]. Besides, Tandem Repeat Finder (TRF) v.4.07 was used to search for tandem repeat sequences [44].

Homology-based, transcriptome-based, and *de novo* prediction were used to annotate the genes in the genome, with the combined use of EvidenceModeler (EVM) v1.1.1 and PASA v2.4.0 [45]. For homology-based predictions, the protein sequences of *Danio rerio*, *Bufo gargarizans*, *Cyclopterus lumpus*, *Pseudoliparis swirei*, *Sebastes umbrosus*, *Gasterosteus aculeatus*, *Scatophagus argus*, *Thalassophryne amazonica*, *Siniperca chuatsi*, *Oncorhynchus mykiss*, *Oryzias latipes*, *Salmo salar*, *Oreochromis niloticus*, *Electrophorus electricus*, *Ostracion cubicus*, and *Acanthochromis polyacanthus* were obtained from the NCBI and aligned to the assembled *S. verrucosa* genome [46]. For transcriptome-based prediction, RNA-seq data were aligned to the genome with default parameters using Hisat v2.1.0 [47]. Subsequently, transcripts were reconstructed with default parameters using StringTie v2.0 [48]. For *de novo* gene predictions, Augustus v3.4.0, Genscan v3.1, and GlimmerHMM v3.0.1 were used to analyze the repeat-masked genome [49–51].

Functional annotation of the genes was performed using Blastp and Blastn ( $-e 1e-5$ ) to blast protein or nucleotide sequences against various databases including the NCBI nonredundant protein (NR), the NCBI nonredundant nucleotide sequence (NT), as well as the SwissProt databases. Gene Ontology (GO) terms were then retrieved and assigned to the *S. verrucosa* query sequences, and enzyme codes (EC) corresponding to GO terms were mapped to KEGG pathway annotations for further functional insights.

### Long-read RNA alignment and isoform identification

The raw reads obtained from the full-length transcriptome sequence were analyzed following the Iso-Seq3 pipeline (v3.1). Briefly, circular consensus sequences were created from the raw subreads, and the cDNA primers were removed using lima. Full-length reads were then filtered, clustered, and polished (Iso-Seq3, v3.1). The polished sequences were aligned to the *S. verrucosa* genome using pbmm2 (v1.13.0). Isoform identification was done using TAMA [52]. Redundant isoforms were removed by comparison with the raw data of the NGS transcriptome data.

An in-depth characterization of the isoforms and artifact removal was done using SQANTI3 [53]. Briefly, isoform characterization was performed using sqanti\_qc.py, and filtering was done using sqanti\_filter.py. The filtered data set was rerun through sqanti\_qc.py for the final confirmation. The SQANTI-filtered isoforms were used for the remaining analyses.

### Expression analysis of the genes in the venom gland

HISAT2 v2.1.0 was used to align the quality-controlled RNA-seq reads to the *S. verrucosa* genome. The aligned reads were assembled into transcripts, and expression was quantified using feature Counts [54]. Differential expression analysis was performed using DESeq2 v1.26.0 to identify differentially expressed genes in each tissue type [55]. Genes with an adjusted p-value < 0.05 and a  $\log_2(\text{fold change}) > 1$  were considered to be differentially expressed. Visualization of differentially expressed toxin genes across all samples was performed using the pheatmap package v1.0.12 in R. Transcript per million (TPM) values were used to assess the expression level among the venom glands, muscles, and skin of *S. verrucosa*.

### Toxin genes annotation

Because of the propensity of toxin gene families to be organized in extensive tandem arrays, the number of paralogs in particular gene families may be compromised, thereby impeding the accuracy of gene annotation [56]. Consequently, additional annotation steps were included in the present study. Combining the ToxProt-based annotation with the isoforms generated from the third-generation sequencing transcriptome of the venom gland, we performed empirical annotation using FGENESH (<http://www.softberry.com>) as well as manual annotation using IGV-GSaman (<https://gitee.com/CJchen/IGV-sRNA>). Finally, each re-annotated gene was searched against the ToxProt database to confirm its identity and further confirmed its homology through a phylogenetic analysis. We also downloaded HMM models from the Pfam database v35.0 [57], and then blasted the annotated toxin genes set using hmmsearch in HMMER v3.1b1 (<http://hmmer.org/>). Sequences with E-values below the default inclusion threshold (E-value <  $1e-05$ ) were retrieved and identified as toxin homologs.

### Proteomic analysis of the venom

Lyophilized venom (100  $\mu\text{g}$ ,  $n=4$ ) was dissolved in a triethylammonium bicarbonate buffer. Next, 10 mM of Tris (2-carboxyethyl) phosphine was added and the venom solution was incubated at 37 °C for 60 min to reduce the disulfide bonds in the proteins. Subsequently, 40 mM of iodoacetamide was added to block the sulfhydryl groups of the cysteine residues to prevent the reformation of disulfide bond reformation. Followed by the addition of pancreatic trypsin at an enzyme-to-protein ratio of 1:50, the venom was digested overnight at 37 °C. The resulting peptides were then desalted using HLB and quantified (Thermo Fisher Scientific). The samples were dissolved in sample loading buffer (2% acetonitrile, 0.1% formic acid) at a concentration of 0.25  $\mu\text{g}/\mu\text{L}$  for mass spectrometry analysis. The samples were separated by high-performance liquid chromatography and analyzed using

a timsTOF Pro2 mass spectrometer (Bruker, Germany) in data-dependent acquisition (DDA) mode. The ionization mode was set to positive, with an ion source voltage of 1.5 kV, and both MS and MS/MS data were acquired using time-of-flight analysis. The mass spectrometry scanning range was set to 100–1700 m/z. The data acquisition mode used parallel accumulation serial fragmentation (PASEF), in which 10 rounds of PASEF were performed after each full MS scan. Second-level spectra with charge states ranging from 0 to 5 were acquired and a dynamic exclusion time of 24 s was set for tandem mass spectrometry scans to avoid repeated ion selection. A local Scorpaeniformes venom protein database was established. This database included toxin gene set from *S. verrucosa* we annotated and the toxin protein of *S. horrida* in the Tox-Prot database. The venom proteomics data was searched against this local database using MaxQuant v2.4.2.0, and the results were filtered based on a parameter Peptide FDR  $\leq 0.01$ . Protein abundance was analyzed using intensity-based absolute quantification (iBAQ) method. iBAQ value is calculated by dividing the sum of peak intensities of all peptides matching to a specific protein by the number of theoretically observable peptides [58].

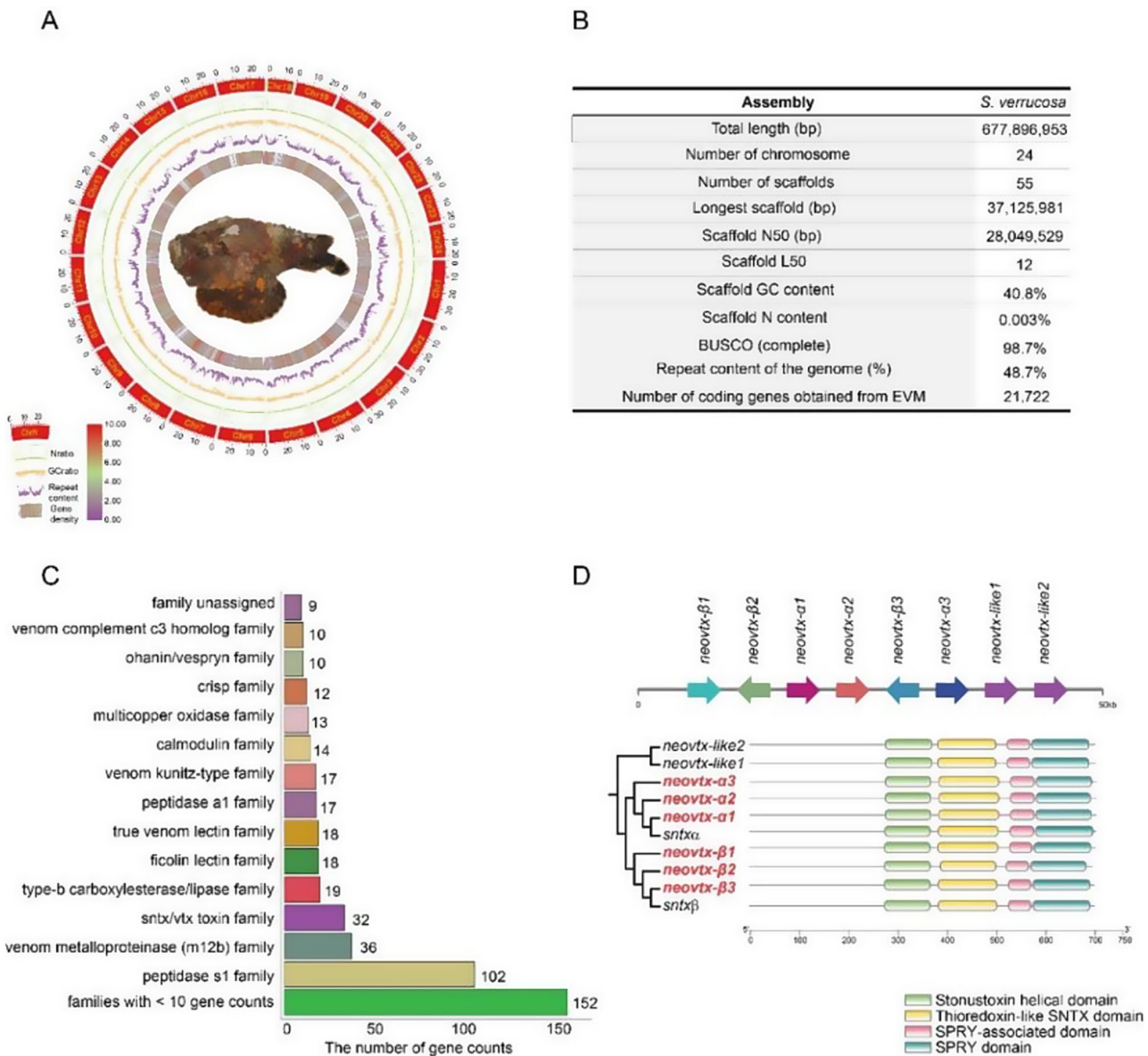
For native PAGE electrophoresis, the lyophilized venom was dissolved in a phosphate buffer solution (10 mg/mL) and centrifuged at 5,000 g for 15 min at room temperature to remove insoluble material and primarily cellular debris. The venom (50  $\mu\text{g}$ ) was then subjected to preparative native-PAGE (6%) gels after diluting with Tris-glycine buffer (0.05 M, pH 8.0). The gels were run at a constant voltage of 120 V at 4 °C. Gel strips were stained with Coomassie brilliant blue R-250 for 30 min and subsequently destained using water. The protein bands of interest were cut and sent for mass spectrometry analysis (Sangon Biotech, China).

## Results

### Chromosome-scale *de novo* assembly and toxin genes annotation

A high quality genome greatly facilitates the genetic decoding of the components of animal venom. In this study, a total of 19.80 Gb HiFi and 154.59 Gb Hi-C library sequencing data were generated respectively, resulting in an assembled genome size of 677.90 Mb for *S. verrucosa* (Fig. 1A). The assembled genome was further evaluated and the results indicated that a high-quality of *S. verrucosa* genome was obtained in this study (Fig. 1B and Supplementary Fig. 1).

The toxin gene identification was performed using a combined analysis with the assembled genome and venom gland transcriptomic data. A total of 478 potential toxin genes were preliminarily identified at the whole-genome level. Most of the genes were clustered into 56



**Fig. 1** Toxin genes in the genome of *S. verrucosa* **a**) Visualized circle plot of chromosomal level genome assembly of *S. verrucosa*, with N ratio, GC ratio, repeat content and gene density shown from outside to inside, corresponding to each pseudochromosome. **b**) Summary of the genome assembly evaluation. **c**) Protein families of all the annotated toxin genes. **d**) Tandem-duplicated SNTX/VTX genes clustered together with sequence divergence

gene families, with the peptidase S1 (102), venom metalloproteinase (36), the SNTX/VTX toxin (32), the type-B carboxylesterase/lipase (19), the ficolin lectin (18), the true venom lectin (18), peptidase A1 (17), and venom Kunitz-type (17) as the most abundant toxin gene families (Fig. 1C, Supplementary Data 1). Further GO annotation revealed that two-thirds (321 out of 476) of the genes exhibit explicit toxin activity (GO:0090729). The other 155 annotated genes were assigned to other functional categories, including calcium channel regulator activity (GO:0005246), killing of cells of another organism (GO:0031640), hyaluronoglucosaminidase activity

(GO:0004415), and serine-type endopeptidase inhibitor activity (GO:0004867) (Supplementary Fig. 2).

Toxin genes usually demonstrated remarkable tandem duplication events during evolution. In the present study, a total of 13 toxin genes were characterized as tandem repeated genes in *S. verrucosa* genome, including the SNTX/VTX genes, true venom lectin genes, peptidase S1 genes, glycosyl hydrolase 56 genes and etc. In particular, eight annotated SNTX/VTX genes were clustered on chromosome 14 (Fig. 1D). The front 6 SNTX/VTX genes showed high sequence homology (>85%) with *neovtxa* (GenBank: AB262392.1) and *neovtxb* (GenBank: AB262393.1) (Fig. 1D and Supplementary Fig. 3). The six

genes were arranged in the following order: *neovtx-b1*, *neovtx-b2*, *neovtx-a1*, *neovtx-a2*, *neovtx-b3*, *neovtx-a3*. Interestingly, the other two SNTX/VTX genes (*neovtx-like1* and *neovtx-like2*) clustered in this region shared a much lower sequence identity (<50%) and clustered into a separate branch of the phylogenetic tree, whereas an InterPro scan revealed that they all contained similar conserved domains, including a Stonustoxin helical domain (IPR048997), Thioredoxin\_11 domain (IPR040581), PRY domain (IPR006574) and SPRY domain (IPR003877) (Fig. 1D).

### Toxin genes expression in the venom gland

NGS of transcriptome was performed using venom gland ( $n=7$ ) to further characterize the toxin genes. A total of 14,380 genes were considered as actively expressed in the venom gland with a mean TPM>1. Further analysis revealed that 240 genes corresponded to the previous annotated toxin genes set (Supplementary Data 1).

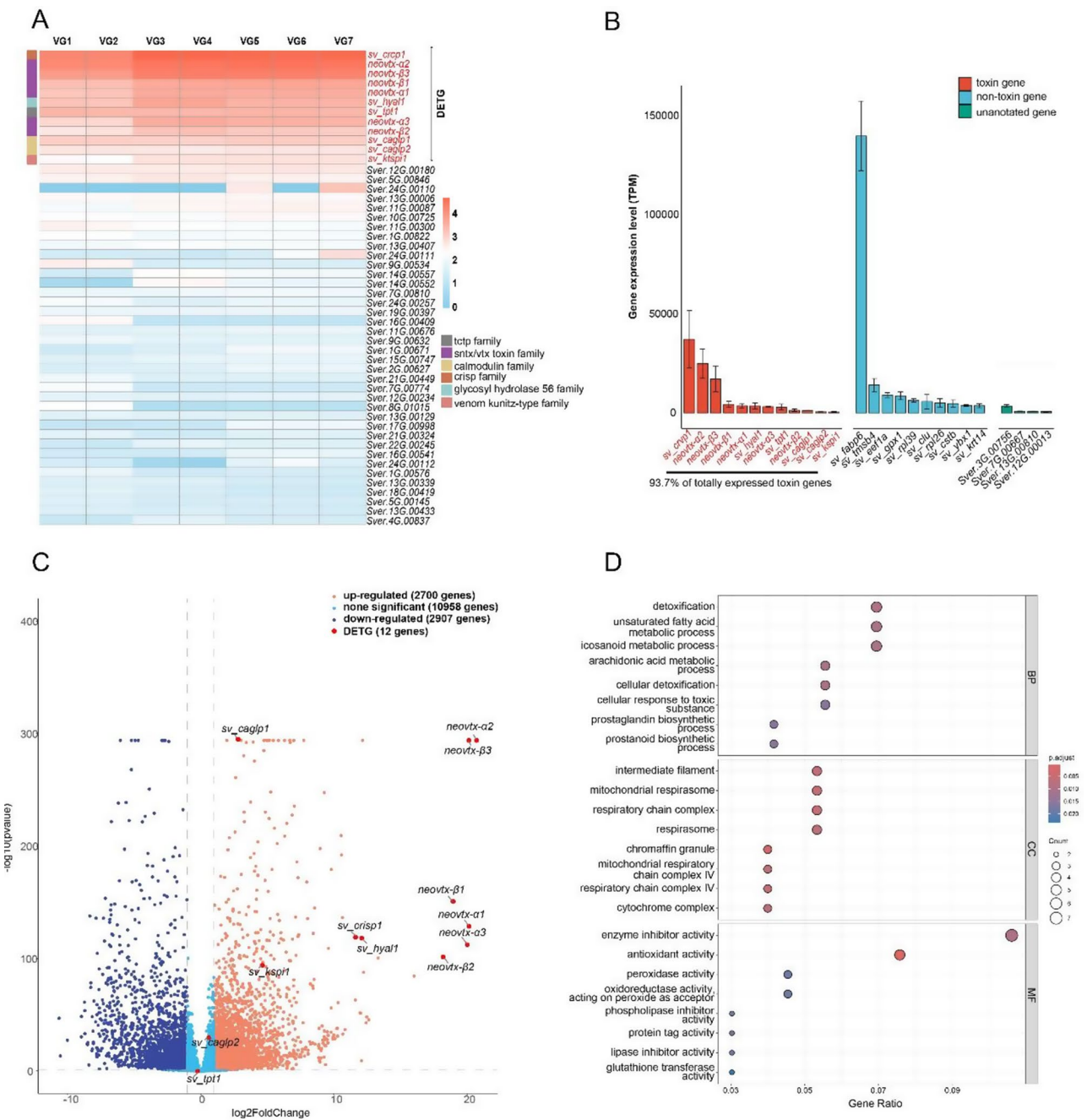
We characterized a number of dominantly expressed toxin genes (DETGs) with a TPM>500 in the venom gland. Twelve toxin genes were designated DETGs, including one cysteine rich venom protein (CRVP) gene (*sv\_crvp1*), the six tandem-repeated neoVTX subunit genes on chromosome 14, one hyaluronidase gene (*sv\_hyal1*), one translationally-controlled tumor protein gene (*sv\_tpt1*), two calglandulin protein genes (*sv\_caglp1* and *sv\_caglp2*), and one Kunitz-type serine protease inhibitor gene (*sv\_kspi1*) (Fig. 2A, B). The 12 DETGs accounted for 93.70% of the totally expressed toxin genes. Moreover, we also examined the highly expressed non-toxin genes in the venom gland, which may be pivotal to regulate the toxin gene expression and/or produce important non-toxic components in the venom. The top ten non-toxin genes expressed in the venom gland were shown in Fig. 2B, including *sv\_fabp6*, *sv\_tmsb4*, *sv\_eef1a*, *sv\_gpx1*, *sv\_rpl39*, *sv\_clu*, *sv\_rpl26*, *sv\_cstb*, *sv\_ybx1* and *sv\_krt14*. Interestingly, four unannotated genes with a TPM>500 were identified in the venom gland, suggesting the presence of unique genes in *S. verrucosa* not previously found in other venomous animals (Fig. 2B).

A comparative analysis was further performed between the venom gland (VG) and its closely related anatomic tissues, including skin (SK) and muscle (MU). The results indicated that the VG exhibited distinct expression profiles, and a higher transcriptome similarity with the SK compared with MU (Supplementary Fig. 4). The results indicated that skin may serve as the potential evolutionary origin of VG in *S. verrucosa*. Differentially expressed genes (DEGs) in the VG compared with other tissues generally indicated their preferred contributing roles to the venom. A total of 2,700 genes showed significantly higher expression in the VG ( $\log_2$  Fold Change>1 and  $\text{padj}<0.05$ ) compared with the SK. By blasting with the

actively expressed 240 toxin genes, we obtained a total of 87 VG preferentially expressed toxin genes including ten of the 12 previously characterized DETGs. The six neoVTX subunit genes exhibited the highest differential expression level (Fig. 2C and Supplementary Data 2). The other two DETGs, the *sv\_tpt1* and *sv\_caglp2* (Sver.16G.00747), exhibited high expression levels in VG, as well as in SK, indicating their potential non-toxic roles outside the VG. We further examined the DEGs of the non-toxin genes set (with TPM>500 in VG) between the VG and SK. A total of 84 genes were characterized and predominantly showed the molecular function of enzyme inhibitor, antioxidant and peroxidase activities, and involved in the biological processes of detoxification, metabolism and biosynthesis (Fig. 2D and Supplementary Data 3). They may be partially involved in the resistance of *S. verrucosa* to its own venom.

### Proteomic evidence of the toxin genes in venom

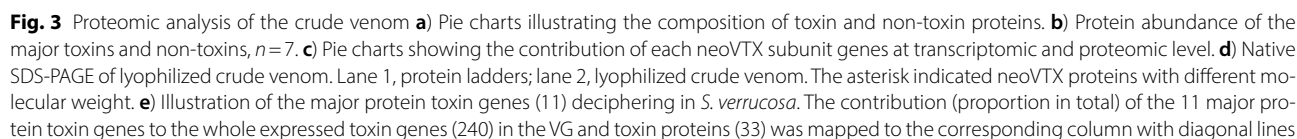
As the major components of the venom, the proteins encoded by the previously identified genes were further characterized by examining the venom proteome. A 4D-label free proteomic approach was employed to analyze the venom protein components of *S. verrucosa*. A total of 662 proteins were obtained, and 33 matched the previously annotated toxin genes (Supplementary Data 4). We characterized the major venom proteins into two categories: toxin proteins and non-toxin proteins. A quantitative analysis showed that the toxin proteins accounted for 43.19% of the total venom proteins (Fig. 3A). Next, we analyzed the proteomic data for the 12 DETGs. As shown in Fig. 3B, the most abundant toxin proteins were helothermine like protein (SvCRVP) from *sv\_crvp1* (22.35%). The neoVTX proteins accounted for approximately 17.21% of the total venom protein. Consistent with the VG transcriptomic data, almost all of the neoVTX proteins were derived from the six tandem-repeated neoVTX subunit genes. Of these, neoVTX a2 (50.55%) and neoVTX b3 (47.04%) accounted for most of the neoVTX proteins, and only a small proportion was contributed by the other four genes, indicating their dominant roles in protein production (Fig. 3C). Interestingly, native-PAGE analysis showed that two major protein bands with molecular weights of approximately 160 and 420 kDa, and both contained neoVTX proteins by mass spectrometry (Fig. 3D). Proteins from the other DETGs, including Kunitz-type serine protease inhibitors (SvKSPI) from *sv\_kspi1* (1.39%), the calglandulin-like protein (SvCAGL, 0.45%), and hyaluronidase (0.44%) were also evident in the venom proteome (Fig. 3B). The above 11 DETGs accounted for 96.91% of all the designated toxin proteins and served as the major toxin genes in *S. verrucosa* (Fig. 3E). The only DETGs with no proteomic



**Fig. 2** Transcriptomic analysis of the venom gland **a**) A heatmap of the highest 50 toxin genes transcription with hierarchical clustering in the venom gland among all the 7 samples. The color represented the expression level based on TPM value. **b**) Categorizes of the dominantly expressed toxin genes (DETGs) with TPM > 500, the highest 10 non-toxin genes, and 5 unannotated genes with TPM > 500. The value was shown as mean  $\pm$  SD,  $n = 7$ . **c**) Volcano plots showing the differentially expressed genes (DEGs) between the venom gland and the skin tissue, with log2 (fold change) thresholds of -1 and 1, and  $P < 0.05$ . **d**) GO enrichment bubble plot analysis of the DEGs from non-toxin genes set (with TPM > 500 in VG) between the VG and SK. The bubble size indicated the gene number, the color represented the adjusted p value. BP, biological process; CC, cellular components; MF, molecular function

evidence in the venom was *tpt1*, suggesting a major role in the VG rather than as venom components.

Other notable toxin proteins in the venom included homologs of putative endothelial lipase (Uniprot ID: J3RZ81), lactotransferrin (K9IMD0), snaclec coagulation factor IX/X binding protein subunit A (SLXA, Q9DG39), snaclec B1 (B4XT00), peroxiredoxin-4 (P0CV91), cobra venom factor (Q91132) and cystatin-2 (J3SE80), whose abundance were higher compared with the lowest expressed DETG (neoVTX-b2) in the VG. As for the non-toxin proteins, gastrotropin protein (25.13%) ranked the highest in abundance in the venom of *S. verrucose*.



The other highly expressed (abundance > 1%) non-toxin proteins included GPx1, PTGS1, cystatin-B, and TCAF1 (Fig. 3B).

#### Alternative splicing contributes to venom diversity

Alternative splicing (AS) of toxin genes was proposed as an important molecular mechanism underlying venom diversity and evolution. To further examine the complexity of the venom components at the transcriptional level, an integrative analysis pipeline of ISO-Seq and NGS and was performed with the same RNA samples from the VG (Fig. 4A).

A total of 379,860 ISO-Seq reads were obtained, with an N50 length of 2,408 bp. All the reads were then aligned to the above assembled genome and 374,821 of them matched the corresponding target genome block. A total of 27,600 unique isoforms were retained, corresponding to 10,945 gene loci. These unique isoforms were further blasted with the 476 annotated toxin genes, and a total of 825 cis-splicing isoforms were mapped to 191 of these genes (Fig. 4B, Supplementary Data 5). Further analysis revealed that the 12 DETGs accounted for 52.48% of the isoforms among all the toxin genes (Fig. 4C). The results indicated that AS predominantly occurred in the highly expressed toxin genes. The six neoVTX genes accounted for 411 AS isoforms, which was approximately half of the total number of toxin gene isoforms (Fig. 4C).

Mapping the isoforms to the six neoVTX genes revealed that *neovtx-a2* (154) and *neovtx-b3* (101) had the highest number of AS isoforms, followed by *neovtx-a3* (71), *neovtx-a1* (35), *neovtx-b1* (17), and *neovtx-b2* (4) (Fig. 4D). Interestingly, 29 trans-splicing isoforms were found among the six neoVTX genes, which utilized exons across these genes (Fig. 4D). The expression of these AS isoforms was further analyzed referring to the NGS data for quantitation analysis. A total of 27 isoforms accounted for about 90% expression of the six neoVTX genes predominantly associated with the *neovtx-a2*, *neovtx-b3* and *neovtx-a3* genes. In addition, the heterogeneous expression pattern of the AS isoforms among all the seven samples revealed the diversity of venom genetic regulation at an individual level (Supplementary Fig. 5). In contrast, all the trans-splicing isoforms exhibited much lower expression (Fig. 4E and Supplementary Data 5). These results indicated that the alternative splicing of the neoVTX genes contributes to the venom diversity.

#### Discussion

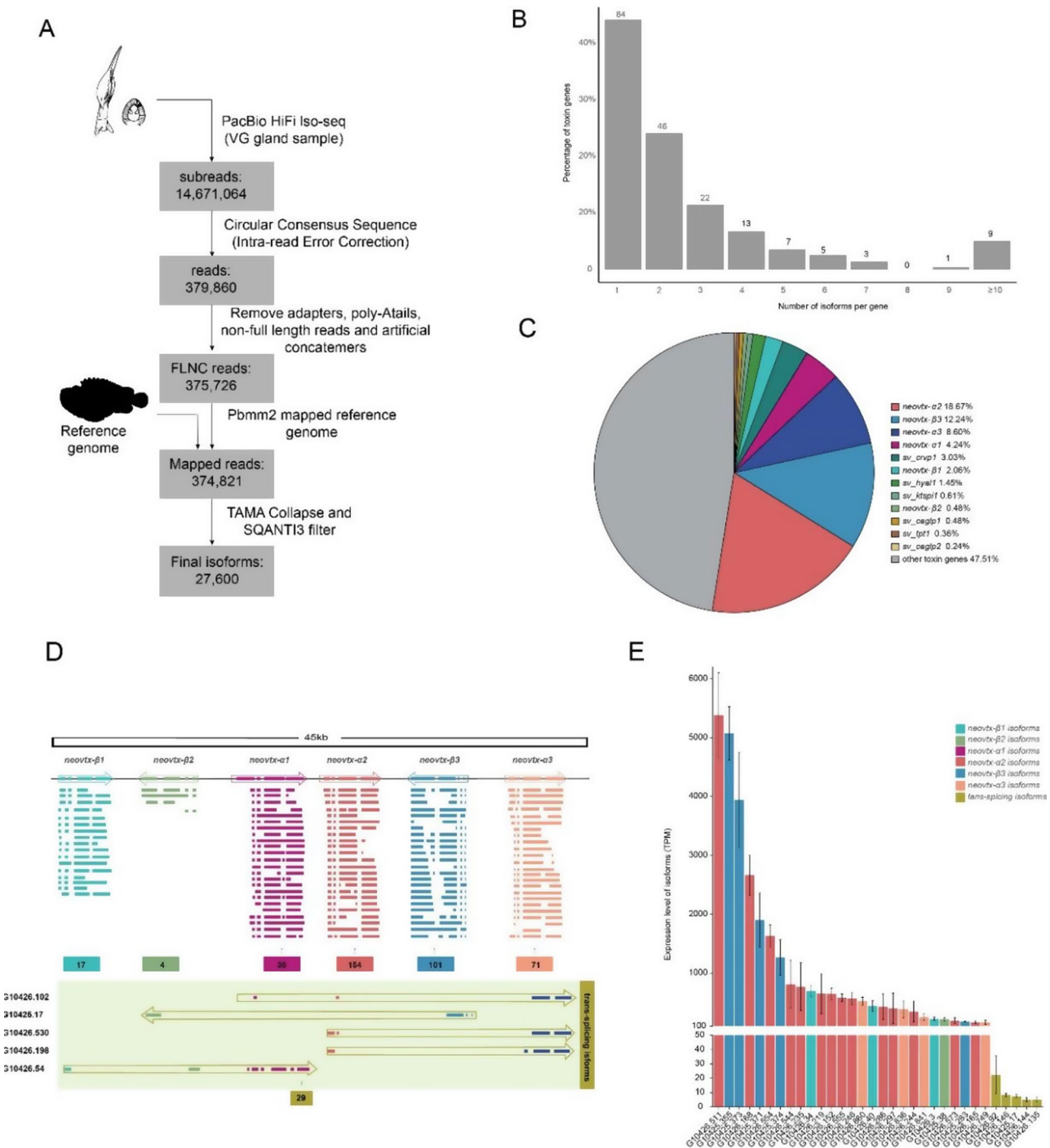
High-throughput sequencing and rapid development of bioinformatic analyses have greatly facilitated the modern venom research. We examined the venom components of *S. verrucosa* through an integrative analysis of genomic, transcriptomic and proteomic data. The chromosome-level genome provides a fundamental genetic

resource for venom components characterization and in-depth biomedical studies. We first assembled a high-quality *S. verrucosa* genome, which enabled us to preliminarily annotate 476 toxin genes. Further transcriptomic analysis of the VG characterized 12 DETGs as the predominantly expressed toxic genes and 11 of these were validated by analyzing the venom proteome.

#### The unique venom composition of *S. verrucosa* underlies its evolution adaptation

A total of 467 toxin genes were annotated in the *S. verrucosa* genome, with the peptidase S1 family, venom metalloproteinase, and SNTX/VTX toxin families as the most abundant (Fig. 1C). Peptidase S1 family genes encode serine proteases, which are responsible for coordinating various physiological functions, including digestion, the immune response, blood coagulation, and reproduction. Snake venom serine protease (SVSP) was characterized as one of the four dominant protein families, together with phospholipase A<sub>2</sub> (PLA<sub>2</sub>), snake venom metalloprotease (SVMP), and three-finger toxins (3FXs) [59]. In the *S. verrucosa* genome, peptidase S1 family genes (102) were listed as the most abundant toxin genes in *S. verrucosa* genome (Fig. 1C). Interestingly, only two of them showed expression levels with TPM > 100, and the peptidase S1 genes as a whole accounted for only a small proportion of the VG transcripts. This was similar with the cases in scorpion, spider and centipede [60–62]. Correspondingly, only around 0.13% of peptidase S1 proteins were found in the crude venom (Supplementary Data 3). A total of 36 venom metalloproteinase family genes were annotated in the *S. verrucosa* genome. All of these showed low-level expression with TPM < 50 and no metalloproteinase protein was identified in the crude venom. Similarly, six PLA<sub>2</sub> genes (TPM < 1) were barely detected in the VG transcriptome. Instead, neoVTX proteins from the six tandem-repeated genes were the most abundant toxin proteins in the venom (Fig. 3B). Although we characterized 12 DETGs in the VG transcripts, only three toxin proteins accounted for more than 1% in the crude venom, including a pair of neoVTX subunits, a CRVP like protein, and a Kuntiz-type serine protease inhibitor. The results indicated that the venom is relatively less complex compared with other venomous animals, which is consistent with a similar study in *S. horrida* [25]. The last but not the least, we found 4 unannotated genes with high expression (TPM > 500) in the VG but no proteins in the venom (Fig. 2B), which suggests their potential roles as unique genes involved in the VG function of *S. verrucosa*. Taken together, the above results suggest that the unique venom composition underlined the venom phenotype evolution of *S. verrucosa*.

Protein toxins offer invaluable source for the development of active drugs. Pore-forming proteins have



**Fig. 4** Massive alternative isoforms from the *neovtx* genes in the venom gland **a)** The work flow for sequencing and analysis for full length transcriptome construction. **b)** Distribution of the number of alternative splicing isoforms per annotated toxin gene. **c)** Percent of the AS isoforms from each DETGs in the total isoforms (856) mapped to the annotated toxin genes (476). **d)** The 6 tandem-repeated *neovtx* genes demonstrated the most isoforms. The isoforms were mapped to each gene with the exon usage shown. 5 trans-splicing isoforms were listed at the lower panel, showing a single isoform embraced exons from different genes. **e)** Expression level of 27 cis-splicing and 5 trans-splicing isoforms from the 6 *neovtx* genes, the value was illustrated as mean  $\pm$  SD,  $n = 7$

shown great potential in tumor therapy due to its ability to deliver antibodies into the cancer cells [63]. SNTX/neoVTX toxins demonstrate high pore-forming activities, and the purified SNTX proteins have been proved to inhibit MCF-7 breast cancer cell growth by triggering apoptosis [64]. SNTX/neoVTX toxins also show ability to bind to voltage gated  $\text{Ca}^{2+}$  channel mimotopes to exert cardiotoxicity. Molecular modified SNTX/neoVTX toxins toward a voltage gated  $\text{Ca}^{2+}$  channel blocker could also be used in cardiovascular therapy [65].

#### **Integrative analysis reveals a contribution of neoVTX genes to the venom diversity of *S. verrucosa***

The major lethal molecules of *S. verrucosa* venom have been purified and molecular characterized as early as the 1990s. The VTX and neoVTX were purified by two different research groups and characterized as different proteins and remained unclarified thus far. The VTX was first characterized as a tetrameric protein consisting of VTXa and VTXb subunits [30, 31], whereas the neoVTX was a dimeric protein consisting of neoVTXa and neoVTXb subunits [66]. They were considered different proteins primarily based on the difference in their deduced amino acid sequences. neoVTXb exhibited 90% identity compared with the reported VTXb, lower than that with the SNTXb from another species, *S. horrida*. Further analysis found that VTXb contains an additional nucleotide at the 3' end of the CDS, leading to a frame-shift, which contributes to the major difference between them. It was suggested that the geographical variation of the fish in the two studies contributed to major loci difference between the two toxin genes [66]. In our assembled genome, the discriminative loci in the three tandem-repeated beta subunit genes were all consistent with that of the neoVTXb gene (GenBank: AB262393.1) (Supplementary Fig. 6). Interestingly, we identified two protein bands with different molecular weights close to the reported dimeric and tetrameric forms of the neoVTX proteins by native PAGE electrophoresis. The two proteins were further confirmed by mass spectrometry sequencing (Fig. 3D). The results indicate that neoVTX proteins are the major lethal toxins in the fish strain that we used and they may be present in either a dimeric or tetrameric form.

Gene duplication plays an important role in the evolution and diversification of toxin genes [67]. The major toxins in the snake venom, including metalloproteinases, phospholipases  $\text{A}_2$ , and three-finger toxins, all experienced gene duplication events during evolution [15, 16]. A total of 32 SNTX/VTX family genes were annotated in our assembled *S. verrucosa* genome. The integrative analysis with the VG transcriptome and venom proteome revealed that most of the neoVTX proteins in the venom were derived from the six tandem-repeated neoVTX

subunit genes, predominantly a pair of a2 and b3 genes (Fig. 3C). Two other SNTX-like genes were also located adjacent to this region. They all contained four conserved domains, but with much lower sequence identity and were clustered into a separate branch of the phylogenetic tree (Fig. 1D). In addition, the transcriptomic data indicated that these two SNTX-like genes were minimally expressed in the VG, and no protein was detected in the venom proteome. The results indicate that they may share the same origin as the neoVTX genes but show no functional roles in *S. verrucosa*. It is hypothesized that SNTX/neoVTX genes were generated by gene duplication throughout the evolutionary history of Percomorpha fish [26]. The identification of gene tandem duplication extremely relies on high quality genome. There are only two published chromosomal level genomes of Scorpaeniformes, *Synanceia verrucosa* and *Sebastes schlegelii* up to now, making it difficult to perform evolutionary analysis. Recently published work has provided solid evidence that a clustered SNTX genes could only be identified in the two venomous Scorpaeniformes species, namely *S. verrucosa* and *S. schlegelii*, leading to the generation of authentic toxic SNTX/neoVTX genes by neofunctionalization [28]. The understanding of evolution pathway of the SNTX/neoVTX and their tandem duplication events requires additional genomic resource, especially of the closely related species.

Alternative splicing of venom genes is considered as a complementary mechanism for the generation of venom complexity across animal lineages [19, 68, 69]. We identified 856 isoforms expressed in the VG from 198 toxin genes, and 411 of them including 29 trans-splicing isoforms were mapped to the six tandem-repeated neoVTX subunit genes, which markedly increased the potential to produce diverse neoVTX proteins. Expression analysis of these AS isoforms revealed that a pair of neoVTX a2 and b3 genes transcribed the several highest expressed isoforms (Fig. 4E). Unfortunately, our proteomic sequencing approach failed to distinguish these proteins with various sizes and charges. A 2D gel electrophoresis of the crude venom from the closely related species *S. horrida* revealed 15 SNTXa isomers and 26 SNTXb isomers with varied length of ORE, partially supported the alternative splicing of neoVTX genes in this study. The results indicated that the AS of the SNTX/neoVTX genes were conserved mechanism in stonefish underlined venom diversity [25]. Interestingly, although we identified a total of 29 trans-splicing isoforms among the six tandem-duplicated neoVTX genes, they all showed much lower expression level (Fig. 4D and E), leaving their functional roles and biological significance unknown. The alternative splicing could contribute to the venom complexity, which is considered as an important adaptive strategy of *S. verrucosa* facing physiological changes and geographic

location, as well as against the native predators. Further work is required to investigate the biological significance of these neoVTX isomers.

#### The other major toxin proteins characterized in the venom of *S. verrucosa*

The most abundant toxin protein other than the neoVTX proteins was characterized as CRVP. Its homologue in *S. horrida* is defined as Golgi-associated plant pathogenesis related protein (*ShGAPR*) instead. Although *ShGAPR* showed a higher sequence similarity with the human *GAPR1*, it may act as a venom toxin similar to CVRPs [25]. The transcriptomic analysis of the *S. verrucosa* venom gland indicated that *sv\_crvp1* was found to be the highest expressed toxin gene in the VG and considered a DEG compared with the SK (Fig. 2B and C), which further supports its potential role as a toxin gene in *S. verrucosa*. The most homologous gene to *SvCRVP* annotated in ToxProt was helothermine, which was first isolated from the venom of the Mexican beaded lizard (*Heterodermus horridum horridum*) [70]. Lizard helothermine is a peptide toxin that blocks ryanodine receptors, which are responsible for  $\text{Ca}^{2+}$  release within skeletal, cardiac, and neuronal cells [71, 72]. Further studies are encouraged to illustrate the potential role of *SvCRVP* as the most abundant venom component besides the neoVTX proteins in *S. verrucosa*.

Venomous snakes use protease inhibitors in the venom as weapons to disrupt the homeostasis of the prey's physical and biochemical reactions, such as the blood coagulation and blood pressure regulation, which results in immobilization or death of the prey [73]. Protease inhibitors may protect peptide/protein toxins in the venom from degradation by proteases from the prey or its predators [74]. Kunitz-type serine protease inhibitors, C-type lectins and phospholipase  $\text{A}_2$  are the most extensively distributed protease inhibitors in the snake venom [75]. We annotated 17 Kunitz-type family genes in the *S. verrucosa* genome and only one of them (*sv\_kspi1/SvKSPI*) was characterized as a DETG in the VG (Figs. 2B and 3B) and accounted for 1.84% in total toxin proteins. Homologs of Snaclec B1 and SLXA belonging to C-type lectins were also identified in the crude venom (Fig. 3B), and together with *SvKSPI*, served as the major protease inhibitors in *S. verrucosa*.

Calglandulin was previously identified from the VG of *Bothrops insularis* and is characterized as a putative  $\text{Ca}^{2+}$  binding protein with four EF-hand motifs [76]. Using antiserum against recombinant calglandulin, it was only detectable in the VG of *Bothrops insularis*, but not in the crude venom or other tissues, indicating that calglandulin may be primarily involved in the cellular control mechanism of the secretion of toxins from the gland into the venom [76, 77]. Whereas, calglandulin was identified

as a major component in the crude venom of the fire ant *Solenopsis invicta* [78] and stingray [79], indicating its unknown functions beyond the toxins secretion. In *S. verrucosa* venom, a small proportion of calglandulin proteins (0.60%) were also evidenced, which indistinguishably corresponded to *sv\_caglp1* and *sv\_caglp2* due to their same deduced amino acid sequence (Supplementary Fig. 7). Notably, transcriptomic analysis revealed that only *sv\_caglp1* gene was characterized as DEGs between the VG and the skin, indicating a transcriptional regulatory mechanism driving *sv\_caglp1* as the major gene involved in toxins secretion (Fig. 2B, C).

Hyaluronidase is an enzyme commonly found in the venom of snake, scorpions, spiders and leeches [80]. Hyaluronidase activity is considered critical for the spreading of toxins and interrupting the integrity of the target's extracellular matrix by degrading hyaluronic acid, which is considered a therapeutic target against snake venom [81, 82]. The cDNA of *S. verrucosa* hyaluronidase has been previously cloned and characterized with a conserved active site consisting of one catalytic residue and four substrate positioning residues [34]. In the present study, we characterized the hyaluronidase gene as a DETG in the VG and further supported by the proteomic data (Figs. 2B and 3B).

Among all the 12 DETGs defined in the *S. verrucosa* VG, only the *sv\_tpt1* gene exhibited no proteomic evidence in the crude venom (Fig. 3B). The *Sv\_tpt1* gene encodes a translationally controlled tumor protein (TCTP), which is predominantly described as a venom toxins in spiders, that contributes to the allergic and inflammatory activity of the venom [83, 84]. There is few studies of the TCTP toxins in other venomous animals and its biological functions under investigation. The expression of the TCTP gene was reported in the VG from three scorpionfishes (Scorpaenidae.spp), indicating its wide distribution in the venomous teleost [85]. In the *S. verrucosa*, the *sv\_tpt1* gene showed comparable high expression levels in the VG and SK (Fig. 2C), along with its absence in the crude venom, which indicates other potential roles beyond as toxins in the *S. verrucosa*.

A group of toxin proteins beyond the 12 DETGs were also identified based on the abundance in the crude venom (Fig. 3B). A putative endothelial lipase first identified in the VG transcriptome of the eastern diamondback rattlesnake was suggested to exhibit phospholipase and triglyceride lipase activity [86]. Lactotransferrin protein is a major iron-binding protein usually found in exocrine fluids, including mucosal secretions and mammal breast milk. Lactotransferrin is annotated as a toxin because of its presence in the proteome of venom accessory glands in the vampire bat (*Desmodus rotundus*) and antimicrobials activity [87]. Similarly, peroxiredoxin-4 [88], Snaclec B1 [89], and cobra venom factor-like protein [90] are also

evidenced as venom proteins in other venomous animals. The presence of these proteins in the crude venom of *S. verrucosa* further supports their characterization as a toxin protein, whereas their underlying molecular mechanism need to be further investigated.

### A notable non-toxin protein in the venom

A set of non-toxin proteins are shown in Fig. 3B. Although some of these have not been reported to be present in the crude venom, they may have housekeeping or hemostasis maintaining roles in the VG. Of these, a gastrotropin-like protein was the second most enriched following the neoVTX proteins. It is encoded by the FABP6 gene, which also showed high expression levels in the VG (Figs. 2B and 3B). Gastrotropin/FABP6 is a fatty acid-binding protein and therapeutic target associated with immune infiltration [91]. Gastrotropin was first characterized as venom protein in *S. horrida*, but not in other venomous animals. It may contribute to the unique activities of the venom by maintaining venom hemostasis and protecting venom components from degradation [25]. We found that the extracted crude venom of *S. verrucosa* exhibits high activity at room temperature after hours; however, further studies are encouraged to investigate the biological significance of the gastrotropin-like protein in *S. verrucosa* venom.

Pharmaceutical innovation.

### Conclusions

In this study, the major venom components of *S. verrucosa* were thoroughly investigated by integrative multi-omics approaches in this work, for the first time in a venomous teleost. Six tandem-duplicated neoVTX subunit genes were identified in the genome, which contained massive AS isoforms and diverse neoVTX proteins. The diverse neoVTX proteins in the venom as lethal particles are important to understand the adaptive evolution of *S. verrucosa*, particularly when engaging nature predators. Further functional studies are needed to elucidate the mechanism of these alternative isoforms in venom. Additional biomedical studies are also encouraged to exploit the venom components of *S. verrucosa* for pharmaceutical innovation.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-11149-6>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

### Acknowledgements

Not applicable.

### Author contributions

Conceptualization, Z.Z. and W.Z.; Methodology, Q.L., H.L., W.Y., Z.P., and Z.Z.; Investigation, Z.Z., W.Z., and F.W.; Writing-Original Draft, Q.L. and Z.Z.; Writing-Review & Editing, Z.Z., S.W., and W.Z.; Funding Acquisition, Z.Z. and W.Z.; Resources, W.Z. and F.W.; Supervision, W.Z. and F.W.

### Funding

This work was supported by National Natural Science Foundation of China (32222014), Ministry of Science and Technology of China (2023YFF1304900, 2021YFF0502804), Science and Technology Department of Guangdong Province (2021QN02H103, 2024A1515013196), Guangdong Forestry Administration (SLYJ2023B4004), and the PI Project of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (GML2020GD0804, GML2022GD0804).

### Data availability

All the data generated and analyzed in this manuscript have now been deposited into the China National Center for Bioinformation with the accession number of PRJCA031620 (<https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA031620>).

### Declarations

#### Ethics approval and consent to participate

All the experiments were conducted in accordance with guidelines approved by the Animal Ethical and Welfare Committee of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou).

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>Center for Evolution and Conservation Biology, Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China

<sup>2</sup>Key Laboratory of Conservation and Application in Biodiversity of South China, School of Life Sciences, Guangzhou University, Guangzhou, China

<sup>3</sup>Jiangxi Provincial Key Laboratory of Conservation Biology, College of Forestry, Jiangxi Agricultural University, Nanchang, China

<sup>4</sup>Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

Received: 11 October 2024 / Accepted: 11 December 2024

Published online: 18 December 2024

### References

1. Zancolli G, Casewell NR. Venom systems as models for studying the origin and regulation of evolutionary novelties. *Mol Biol Evol*. 2020;37(10):2777–90.
2. Bordon KCF, Cologna CT, Fornari-Baldo EC, Pinheiro-Júnior EL, Cerni FA, Amorim FG, Anjolette FAP, Cordeiro FA, Wiesel GA, Cardoso IA, et al. From animal poisons and venoms to medicines: achievements, challenges and perspectives in drug discovery. *Front Pharmacol*. 2020;11:1132.
3. von Reumont BM, Anderlüh G, Antunes A, Ayvazyan N, Beis D, Caliskan F, Crnković A, Damm M, Dutertre S, Ellgaard L et al. Modern venomomics-Current insights, novel methods, and future perspectives in biological and applied animal venom research. *GigaScience* 2022, 11.
4. Suranse V, Srikanthan A, Sunagar K. Animal Venoms: Origin, Diversity and Evolution. In: *eLS*. John Wiley & Sons, Ltd; 2018: 1–20.

5. Suryamohan K, Krishnankutty SP, Guillory J, Jevit M, Schröder MS, Wu M, Kuriakose B, Mathew OK, Perumal RC, Koludarov I, et al. The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. *Nat Genet.* 2020;52(1):106–17.
6. Ludington AJ, Hammond JM, Breen J, Deveson IW, Sanders KL. New chromosome-scale genomes provide insights into marine adaptations of sea snakes (*Hydrophis*: Elapidae). *BMC Biol* 2023, 21(1).
7. Margres MJ, Rautsaw RM, Strickland JL, Mason AJ, Schramer TD, Hofmann EP, Stiers E, Ellsworth SA, Nystrom GS, Hogan MP et al. The Tiger Rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. *Proc Natl Acad Sci U S A* 2021, 118(4).
8. Zhu B, Jin P, Hou Z, Li J, Wei S, Li S. Chromosomal-level genome of a sheet-web spider provides insight into the composition and evolution of venom. *Mol Ecol Resour.* 2022;22(6):2333–48.
9. You Y, Tang Y, Yin W, Liu X, Gao P, Zhang C, Tembrock LR, Zhao Y, Yang Z. From genome to proteome: Comprehensive identification of venom toxins from the Chinese funnel-web spider (Macrothelidae: *Macrothele yani*). *Int J Biol Macromol.* 2024;268(Pt 2).
10. Herráez-Pérez A, Pardos-Blas JR, Afonso CML, Tenorio MJ, Zardoya R. Chromosome-level genome of the venomous snail *Kalloconus canariensis*: a valuable model for venomomics and comparative genomics. *Gigascience.* 2022;12.
11. Peng C, Huang Y, Bian C, Li J, Liu J, Zhang K, You X, Lin Z, He Y, Chen J, et al. The first *Conus* genome assembly reveals a primary genetic central dogma of conopeptides in *C. betulinus*. *Cell Discov.* 2021;7(1):11.
12. Farhat S, Modica MV, Puillandre N. Whole genome duplication and gene evolution in the hyperdiverse venomous gastropods. *Mol Biol Evol.* 2023;40(8):msad171.
13. Wei S, Zhou W, Fan H, Zhang Z, Guo W, Peng Z, Wei F. Chromosome-level genome assembly of the yellow boxfish (*Ostracion cubicus*) provides insights into the evolution of bone plates and ostracitoxin secretion. *Front Mar Sci.* 2023;10:1170704.
14. Martinson EO, Mrinalini, Kelkar YD, Chang C-H, Werren JH. The evolution of venom by co-option of single-copy genes. *Curr Biol.* 2017;27(13):2007–13. e2008.
15. Zhang Z-Y, Lv Y, Wu W, Yan C, Tang C-Y, Peng C, Li J-T. The structural and functional divergence of a neglected three-finger toxin subfamily in lethal elapids. *Cell Rep.* 2022;40(2):111079.
16. Almeida DD, Viala VL, Nachtigall PG, Broe M, Gibbs HL, Serrano SMT, Mourada-Silva AM, Ho PL, Nishiyama-Jr MY, Junqueira-de-Azevedo ILM. Tracking the recruitment and evolution of snake toxins using the evolutionary context provided by the *Bothrops jararaca* genome. *Proc Natl Acad Sci U S A.* 2021;118(20):e2015159118.
17. Koludarov I, Senoner T, Jackson TNW, Dashevsky D, Heinzinger M, Aird SD, Rost B. Domain loss enabled evolution of novel functions in the snake three-finger toxin gene superfamily. *Nat Commun.* 2023;14(1):4861.
18. Zancolli G, Reijnders M, Waterhouse RM, Robinson-Rechavi M. Convergent evolution of venom gland transcriptomes across Metazoa. *Proc Natl Acad Sci U S A.* 2022;119(1):e2111392119.
19. Ye X, He C, Yang Y, Sun YH, Xiong S, Chan KC, Si Y, Xiao S, Zhao X, Lin H, et al. Comprehensive isoform-level analysis reveals the contribution of alternative isoforms to venom evolution and repertoire diversity. *Genome Res.* 2023;33(9):1554–67.
20. Harris RJ, Jenner RA. Evolutionary ecology of fish venom: Adaptations and consequences of evolving a venom system. *Toxins.* 2019;11(2):60.
21. Maillaud C, Hoang-Oppermann T, Hoang-Oppermann V, Rigot H, Girardot S, Nour M. Is stonefish *Synanceia verrucosa* envenomation potentially lethal? *Toxicon.* 2020;184:78–82.
22. Lopez CAJ, Magee CAJ, Belyea CCM, Gumboc L. Finger flexor tenosynovitis from Stonefish envenomation injury. *J Am Acad Orthop Surg Glob Res Rev.* 2019;3(5):e024.
23. Khoo HE. Bioactive proteins from stonefish venom. *Clin Exp Pharmacol Physiol.* 2002;29(9):802–6.
24. Poh CH, Yuen R, Khoo HE, Chung M, Gwee M, Gopalakrishnakone P. Purification and partial characterization of stonustoxin (lethal factor) from *Synanceia horrida* venom. *Comp Biochem Physiol B.* 1991;99(4):793–8.
25. Ziegman R, Undheim EAB, Baillie G, Jones A, Alewood PF. Investigation of the estuarine stonefish (*Synanceia horrida*) venom composition. *J Proteom.* 2019;201:12–26.
26. Ellisdon AM, Reboul CF, Panjikar S, Huynh K, Oellig CA, Winter KL, Dunstone MA, Hodgson WC, Seymour J, Deardorff PK, et al. Stonefish toxin defines an ancient branch of the perforin-like superfamily. *Proc Natl Acad Sci U S A.* 2015;112(50):15360–5.
27. Saggiomo SL, Firth C, Wilson DT, Seymour J, Miles JJ, Wong Y. The geographic distribution, venom components, pathology and treatments of stonefish (*Synanceia spp.*) venom. *Mar Drugs.* 2021;19(6):302.
28. Tang T, Huang Y, Peng C, Liao Y, Lv Y, Shi Q, Gao B. A chromosome-level genome assembly of the reef stonefish (*Synanceia verrucosa*) provides novel insights into Stonustoxin (*sntx*) Genes. *Mol Biol Evol.* 2023;40(10):msad215.
29. Garnier P, Sauviat M-p, Goudey-perriere F, Perriere C. Cardiotoxicity of verrucotoxin, a protein isolated from the venom of *Synanceia verrucosa*. *Toxicon.* 1997;35(1):47–55.
30. Garnier P, Goudey-Perrière F, Breton P, Dewulf C, Petek F, Perrière C. Enzymatic properties of the stonefish (*Synanceia verrucosa*) Bloch and Schneider, 1801) venom and purification of a lethal, hypotensive and cytolytic factor. *Toxicon.* 1995;33(2):143–55.
31. Garnier P, Ducancel F, Ogawa T, Boulain J-C, Goudey-Perrière F, Perrière C, Ménez A. Complete amino-acid sequence of the  $\beta$ -subunit of VTx from venom of the stonefish (*Synanceia verrucosa*) as identified from cDNA cloning experiments. *Biochim Biophys Acta.* 1997;1337(1):1–5.
32. Harris RJ, Youngman NJ, Chan W, Bosmans F, Cheney KL, Fry BG. Getting stoned: Characterisation of the coagulotoxic and neurotoxic effects of reef stonefish (*Synanceia verrucosa*) venom. *Toxicol Lett.* 2021;346:16–22.
33. Campos FV, Fiorotti HB, Coitinho JB, Figueiredo SG. Fish Cytolysins in All Their Complexity. *Toxins.* 2021;13(12):877.
34. Madokoro M, Ueda A, Kiriake A, Shiomi K. Properties and cDNA cloning of a hyaluronidase from the stonefish *Synanceia verrucosa* venom. *Toxicon.* 2011;58(4):285–92.
35. Kato K, Nakagawa H, Shinohara M, Ohura K. Purification of a novel lectin from the dorsal spines of the stonefish, *Synanceia verrucosa*. *J Osaka Dent Univ.* 2016;50(2):55–61.
36. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat Methods.* 2021;18(2):170–5.
37. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics.* 2020;36(9):2896–8.
38. Zhang H, Song L, Wang X, Cheng H, Wang C, Meyer CA, Liu T, Tang M, Aluru S, Yue F, et al. Fast alignment and preprocessing of chromatin profiles with Chromap. *Nat Commun.* 2021;12(1):6566.
39. Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* 2023, 39(1).
40. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol.* 2021;38(10):4647–54.
41. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A.* 2020;117(17):9451–7.
42. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA.* 2015;6:11.
43. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinf.* 2009;Chap. 4:Unit 4.10.
44. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999;27(2):573–80.
45. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using EvidenceModeler and the program to assemble spliced alignments. *Genome Biol.* 2008;9(1):R7.
46. Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, Bibi F, Yang Y, Wang J, Nie W, et al. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science.* 2019;364(6446):eaav6202.
47. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37(8):907–15.
48. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33(3):290–5.
49. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics.* 2008;24(5):637–44.
50. Burge CB, Karlin S. Finding the genes in genomic DNA. *Curr Opin Struct Biol.* 1998;8(3):346–54.
51. Majoros WH, Salzberg SL. An empirical analysis of training protocols for probabilistic gene finders. *BMC Bioinformatics.* 2004;5:206.

52. Kuo RI, Cheng Y, Zhang R, Brown JWS, Smith J, Archibald AL, Burt DW. Illuminating the dark side of the human transcriptome with long read transcript sequencing. *BMC Genomics*. 2020;21(1):751.
53. Tardaguila M, de la Fuente L, Marti C, Pereira C, Pardo-Palacios FJ, Del Risco H, Ferrell M, Mellado M, Macchietto M, Verheggen K, et al. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Res*. 2018;28(3):396–411.
54. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–30.
55. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
56. Margres MJ, Rautsaw RM, Strickland JL, Mason AJ, Schramer TD, Hofmann EP, Stiers E, Ellsworth SA, Nystrom GS, Hogan MP, et al. The Tiger Rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. *Proc Natl Acad Sci U S A*. 2021;118(4):e2014634118.
57. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar Gustavo A, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. The protein families database in 2021. *Nucleic Acids Res*. 2020;49(D1):D412–9.
58. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. Global quantification of mammalian gene expression control. *Nature*. 2011;473(7347):337–42.
59. Tasoulis T, Isbister GK. A current perspective on snake venom composition and constituent protein families. *Arch Toxicol*. 2023;97(1):133–53.
60. Morgenstern D, Rohde BH, King GF, Tai T, Sher D, Zlotkin E. The tale of a resting gland: transcriptome of a replete venom gland from the scorpion *Hottentotta judaicus*. *Toxicon*. 2011;57(5):695–703.
61. Almeida DD, Scortecchi KC, Kobashi LS, Agnez-Lima LF, Medeiros SR, Silva-Junior AA, Junqueira-de-Azevedo Ide L, Fernandes-Pedrosa Mde F: Profiling the resting venom gland of the scorpion *Tityus stigmurus* through a transcriptomic survey. *BMC Genomics*. 2012;13:362.
62. Liu ZC, Zhang R, Zhao F, Chen ZM, Liu HW, Wang YJ, Jiang P, Zhang Y, Wu Y, Ding JP, et al. Venomic and transcriptomic analysis of centipede *Scolopendra subspinipes dehaani*. *J Proteome Res*. 2012;11(12):6197–212.
63. Benton JT, Bayly-Jones C. Challenges and approaches to studying pore-forming proteins. *Biochem Soc Trans*. 2021;49(6):2749–65.
64. Wijanarko A, Firdianna A, Ginting MJ, Lischer K, Sahlan M, Hermansyah H. Isolation and anticancer activity assay of stonefish (*Synanceia horrida*) bioactive stonustoxin. *AIP Conference Proceedings* 2022, 2537(1).
65. Angus JA, Wright CE, Xi Q. Targeting voltage-gated calcium channels in cardiovascular therapy. *Lancet*. 2000;356(9238):1287–9.
66. Ueda A, Suzuki M, Honma T, Nagai H, Nagashima Y, Shiomi K. Purification, properties and cDNA cloning of neoverrucotoxin (neoVTX), a hemolytic lethal factor from the stonefish *Synanceia verrucosa* venom. *Biochim Biophys Acta*. 2006;1760(11):1713–22.
67. Wong ESW, Belov K. Venom evolution through gene duplications. *Gene*. 2012;496(1):1–7.
68. Haney RA, Matte T, Forsyth FS, Garb JE. Alternative transcription at venom genes and its role as a complementary mechanism for the generation of venom complexity in the common house spider. *Front Ecol Evol*. 2019;7:85.
69. Ogawa T, Oda-Ueda N, Hisata K, Nakamura H, Chijiwa T, Hattori S, Isomoto A, Yugeta H, Yamasaki S, Fukumaki Y, et al. Alternative mRNA splicing in three venom families underlying a possible production of divergent venom proteins of the habu snake, *Protobothrops flavoviridis*. *Toxins (Basel)*. 2019;11(10):581.
70. Mochca-Morales J, Martin BM, Possani LD. Isolation and characterization of Helothermine, a novel toxin from *Heloderma horridum* (Mexican beaded lizard) venom. *Toxicon*. 1990;28(3):299–309.
71. Van Petegem F. Ryanodine receptors: structure and function. *J Biol Chem*. 2012;287(38):31624–32.
72. Morrisette J, Krätzschnmar J, Haendler B, el-Hayek R, Mochca-Morales J, Martin BM, Patel JR, Moss RL, Schleuning WD, Coronado R, et al. Primary structure and properties of helothermine, a peptide toxin that blocks ryanodine receptors. *Biophys J*. 1995;68(6):2280–8.
73. Inagaki H. Snake Venom Protease Inhibitors: Enhanced Identification, Expanding Biological Function, and Promising Future. In: *Snake Venoms*. Edited by Inagaki H, Vogel C-W, Mukherjee AK, Rahmy TR, Gopalakrishnakone P. Dordrecht: Springer Netherlands; 2017: 161–186.
74. Ma H, Xiao-Peng T, Yang S-L, Lu Q-M, Lai R. Protease inhibitor in scorpion (*Mesobuthus eupeus*) venom prolongs the biological activities of the crude venom. *Chin J Nat Med*. 2016;14(8):607–14.
75. Thakur R, Mukherjee AK. Pathophysiological significance and therapeutic applications of snake venom protease inhibitors. *Toxicon*. 2017;131:37–47.
76. Junqueira-de-Azevedo ILM, Pertinhez T, Spisni A, Carreño FR, Farah CS, Ho PL. Cloning and expression of calglandulin, a new EF-hand protein from the venom glands of *Bothrops insularis* snake in *E. coli*. *Biochim Biophys Acta*. 2003;1648(1):90–8.
77. Pierre LS, Woods R, Earl S, Masci PP, Lavin MF. Identification and analysis of venom gland-specific genes from the coastal taipan (*Oxyuranus scutellatus*) and related species. *Cell Mol Life Sci*. 2005;62(22):2679–93.
78. Cai L, Yang F, Wang Y, Yang J, Zhu Y, Ma X, Höfer J, Wang Y, Ma Y, Xiao L. A combined protein toxin screening based on the transcriptome and proteome of *Solenopsis invicta*. *Proteome Sci*. 2022;20(1):15.
79. Kirchhoff KN, Billion A, Voolstra CR, Kremb S, Wilke T, Vilcinskis A. Stingray venom proteins: Mechanisms of action revealed using a novel network pharmacology approach. *Mar Drugs*. 2021;20(1):27.
80. Bala E, Hazarika R, Singh P, Yasir M, Shrivastava R. A biological overview of Hyaluronidase: A venom enzyme and its inhibition with plants materials. *Materials Today* 2018, 5(2, Part 1):6406–6412.
81. Silva de França F, Tambourgi DV. Hyaluronan breakdown by snake venom hyaluronidases: From toxins delivery to immunopathology. *Front Immunol*. 2023;14:1125899.
82. Bittenbinder MA, van Thiel J, Cardoso FC, Casewell NR, Gutiérrez J-M, Kool J, Vonk FJ. Tissue damaging toxins in snake venoms: mechanisms of action, pathophysiology and treatment strategies. *Commun Biol*. 2024;7(1):358.
83. Boia-Ferreira M, Moreno KG, Basilio ABC, da Silva LP, Vuitika L, Soley B, Wille ACM, Donatti L, Barbaro KC, Chaim OM, et al. TCTP from *Loxosceles Intermedia* (Brown Spider) venom contributes to the allergic and inflammatory response of cutaneous loxoscelism. *Cells*. 2019;8(12):1489.
84. Senff-Ribeiro A. Translationally Controlled Tumor Protein (TCTP/HRF) in Animal Venoms. In: *TCTP/tpt1 - Remodeling Signaling from Stem Cell to Disease*. Edited by Telerman A, Amson R. Cham: Springer International Publishing; 2017: 193–200.
85. Xie B, Yu H, Kerkkamp H, Wang M, Richardson M, Shi Q. Comparative transcriptome analyses of venom glands from three scorpionfishes. *Genomics*. 2019;111(3):231–41.
86. Rokyta DR, Lemmon AR, Margres MJ, Aronow K. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics*. 2012;13(1):312.
87. Francischetti IMB, Assumpção TCF, Ma D, Li Y, Vicente EC, Uieda W, Ribeiro JMC. The Vampirome: Transcriptome and proteome analysis of the principal and accessory submaxillary glands of the vampire bat *Desmodus rotundus*, a vector of human rabies. *J Proteom*. 2013;82:288–319.
88. Calvete JJ, Fasoli E, Sanz L, Boschetti E, Righetti PG. Exploring the venom proteome of the western diamondback rattlesnake, *Crotalus atrox*, via snake venomomics and combinatorial peptide ligand library approaches. *J Proteome Res*. 2009;8(6):3055–67.
89. Jebali J, Bazaa A, Sarray S, Benhaj K, Karboul A, El Ayeb M, Marrakchi N, Gargouri A. C-type lectin protein isoforms of *Macrovipera lebetina*: cDNA cloning and genetic diversity. *Toxicon*. 2009;53(2):228–37.
90. Fritzinger DC, Bredehorst R, Vogel CW. Molecular cloning and derived primary structure of cobra venom factor. *Proc Natl Acad Sci U S A*. 1994;91(26):12775–9.
91. McKillop IH, Girardi CA, Thompson KJ. Role of fatty acid binding proteins (FABPs) in cancer development and progression. *Cell Signal*. 2019;62:109336.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.