



OPEN

## Identification of *Daboia siamensis* venom using integrated multi-omics data

Thammakorn Saethang<sup>1,10</sup>✉, Poorichaya Somparn<sup>2,3,10</sup>, Sunchai Payungporn<sup>4</sup>, Sira Sriswasdi<sup>5,6</sup>, Khin Than Yee<sup>7</sup>, Kenneth Hodge<sup>2</sup>, Mark A. Knepper<sup>8</sup>, Lawan Chanhome<sup>9</sup>, Orawan Khow<sup>9</sup>, Narongsak Chaibutr<sup>9</sup>, Visith Sitprijia<sup>9</sup> & Trairak Pisitkun<sup>2,8</sup>✉

Snakebite, classified by World Health Organization as a neglected tropical disease, causes more than 100,000 deaths and 2 million injuries per year. Currently, available antivenoms do not bind with strong specificity to target toxins, which means that severe complications can still occur despite treatment. Moreover, the cost of antivenom is expensive. Knowledge of venom compositions is fundamental for producing a specific antivenom that has high effectiveness, low side effects, and ease of manufacture. With advances in mass spectrometry techniques, venom proteomes can now be analyzed in great depth at high efficiency. However, these techniques require genomic and transcriptomic data for interpreting mass spectrometry data. This study aims to establish and incorporate genomics, transcriptomics, and proteomics data to study venomomics of a venomous snake, *Daboia siamensis*. Multiple proteins that have not been reported as venom components of this snake such as hyaluronidase-1, phospholipase B, and waprin were discovered. Thus, multi-omics data are advantageous for venomomics studies. These findings will be valuable not only for antivenom production but also for the development of novel therapeutics.

Snakebite is a crucial public health burden in tropical countries, with over one million bites reported annually according to the World Health Organization (WHO)<sup>1–3</sup>. Snakebite envenomation has been classified as a neglected health condition by WHO since it causes around 90,000–137,880 deaths and 1.8–2.7 million injuries per year and access to antivenoms is still obstructed in many developing countries<sup>4,5</sup>. Currently available commercial antivenoms are the intravenous administration of antivenom prepared from hyperimmune serum of horses, sheep or other large domesticated animals<sup>6</sup>. However, a mixture of antibodies contained in the antivenom do not bind with strong specificity to target toxins, which means that severe complications such as chronic disability and amputations can still occur despite treatment<sup>5–7</sup>. Even in cases of successful treatment, the cost of antivenom is expensive. There was a report showing that an antivenom could cost around 640 USD or even 1000 USD per treatment<sup>4,8</sup>. Furthermore, over the last two decades, several primary antivenom producers (such as Syntex, Behringwerke, and Sanofi Paliisteur) have ceased production of antivenoms. This has led to a current crisis in antivenom availability in various regions of the world, creating an important, life-threatening therapeutic gap<sup>9,10</sup>.

The knowledge of venom compositions is fundamental for producing a specific antivenom that has high effectiveness, low side effects, and ease of manufacture<sup>7,8</sup>. Snake venoms are complex mixtures of small molecules and peptides/proteins which have long evolved to help snakes in hunting prey and in protecting themselves from predators. Snake venom compositions vary among different snake species. Even within the same species, venom compositions vary according to age, geographic location, gender, diet, and seasons<sup>11</sup>. Mass spectrometry

<sup>1</sup>Department of Computer Science, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand. <sup>2</sup>Center of Excellence in Systems Biology, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand. <sup>3</sup>Translational Research in Inflammation and Immunology Research Unit (TRIRU), Department of Microbiology, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand. <sup>4</sup>Research Unit of Systems Microbiology, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand. <sup>5</sup>Center of Excellence in Computational Molecular Biology, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand. <sup>6</sup>Research Affairs, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand. <sup>7</sup>Chemical Toxicology Research Division, Department of Medical Research, Ministry of Health, Yangon 11191, Myanmar. <sup>8</sup>Epithelial Systems Biology Laboratory, NHLBI, National Institutes of Health, Bethesda, MD, USA. <sup>9</sup>Queen Saovabha Memorial Institute, The Thai Red Cross Society, Bangkok 10330, Thailand. <sup>10</sup>These authors contributed equally: Thammakorn Saethang and Poorichaya Somparn. ✉email: thammakorn.s@ku.th; pisitkut@nhlbi.nih.gov

techniques have been used to generate and analyze snake venom proteome data<sup>6,7,12,13</sup>. Compared to the aforementioned studies, recent advances in mass spectrometry mean that these proteomes can now be analyzed in great depth at high efficiency<sup>3,14</sup>. However, the modern approach requires genomic and transcriptomic data for identification of proteins, unveiling the contributions of gene content and regulation, and interpreting gene expression. To be clear, fully sequenced genomes are critical for interpreting mass spectrometry data<sup>15</sup>. The identified proteins from snake venom will be highly valuable for the production of antivenoms or new toxin inhibitors.

The increase in availability of snake genome sequencing data will facilitate not only the development of antivenom but also the discovery of new therapeutic candidates from snake venom. As of now, several venom-derived therapeutics have been approved by the Food and Drug Administration (FDA) such as Captopril, Tirofiban, Eptifibatide, Defibrase, Hemocoagulase, and Ximelagatran and multiple candidates are now in clinical trials<sup>16</sup>. Venom-derived drugs are being researched for application in the treatment of hypertension, acute coronary syndromes, perioperative bleeding, congestive heart failure, human immunodeficiency virus (HIV), multiple sclerosis, chronic pain treatment, blood loss during vascular trauma, non-compressible hemorrhage, and cancer<sup>3,6,7,17</sup>. In addition, snake venoms are also reported to possess potential therapeutic properties such as antitumor, antimicrobial, anticoagulating, and analgesic activities<sup>18</sup>. In the future, with the complete snake genome databases, it will be possible to identify novel therapeutic/bioactive molecules by scanning through snake genomes for protein coding regions and comparing these proteins against established pharmacophore databases<sup>17</sup>.

Apart from development of antivenoms and discovery of novel therapeutic agents, genome sequencing data is important for evolutionary research. Since the evolutionary origins of snake venom genes are still not fully understood, genome sequences can be used in evolutionary gene loss analysis, identification of pseudogenes, and phylogeny reconstruction<sup>17</sup>. For example, comparative genomics could identify gene loss of several visual pigment genes in some snakes<sup>19</sup>. This loss might relate to the fossorial behavior (digging and living underground) of ancestral snakes, thus causing reduction or loss of eyesight development<sup>20</sup>. Genome sequences are also important for the annotation of gene isoforms, especially toxin gene isoforms. Multiple isoforms of toxin genes have been found and hypothesized to provide broad spectrum toxicity against various prey species, or to serve as a toxin repertoire to prevent prey from developing resistance. In addition, multiple isoforms could relate to potentiation, since a complex composed of multiple toxins is more potent than a single toxin<sup>21</sup>. With the availability of snake genome sequencing data, evolutionary novelties can be discovered.

To this end, the primary objective of this study is to establish a multi-omics approach to identify venom composition of *Daboia siamensis* by incorporating genomics, transcriptomics, and proteomics data. The generated data will be highly valuable for antivenom production, discovery of novel therapeutics, and evolutionary studies. All novelties found from analyses of our initiated snake omics data will directly benefit individuals who are at risk of envenomation from *Daboia siamensis* which is endemic to Thailand and other parts of Southeast Asia as well as indirectly benefit relevant biomedical communities worldwide.

## Methods

**Snake tissue and venom acquisition and processing.** Specimens of Thai *Daboia siamensis* were retrieved from the snake farm of Queen Saovabha Memorial Institute (QSMI), the Thai Red Cross Society, Thailand. The experimental protocol was approved by the Ethic Committee of the Queen Saovabha Memorial Institute Animal Care and Use (Approval no. QSMI-ACUC-03-2016) in accordance with the guidelines of the National Research Council of Thailand. This study is reported in accordance with ARRIVE guidelines (<https://arriveguidelines.org>). Tissue samples were harvested from an adult female snake weighted 0.495 kg with a snout-to-vent length (SVL) of 96 cm and a total length (TOL) of 108.5 cm. Venom gland and crude venom were harvested from an adult male snake weighted 0.375 kg with SVL of 83 cm and TOL of 97 cm. All specimens were captured from Sakaew province, Thailand. Snakes were housed individually in a secure locked plastic cage, equipped with a hiding box and water bowl. They were fed mice once weekly until being experimented. Prior to venom glands and other organs removal, snakes were fully euthanized by a high dose of isoflurane inhalation and were monitored for at least 20 min to the point of loss of reflex. The euthanized snakes were disposed of according to the QSMI guideline.

**Genomic DNA library preparation.** Whole genomic library preparation was carried out using the 10x Genomics linked-read technology. High Molecular Weight (HMW) genomic DNA (gDNA) is required for the preparation process. In this study, HMW gDNA was extracted from fresh *Daboia siamensis* liver tissue. First, 500 mg of tissue was harvested and subsequently snapped frozen in liquid nitrogen. Frozen tissue was stored at  $-80^{\circ}\text{C}$  before proceeding to gDNA library preparation by Novogene Co., Ltd. (China). Prior to the gDNA extraction process, the frozen tissue was kept in a tube and then thawed by placing a tube on ice. The tissue was cut into pieces ( $\sim 25$  mg per piece) and placed in 1.5 ml Kimble Kontes Pellet Pestle tubes. These tubes were maintained on ice. The nuclei isolation-digestion and DNA purification processes were achieved as described in the 10x Genomics Technical Note: Sample Preparation Demonstrated Protocol-Rev B (CG000072). The purified gDNA was stored at  $-20^{\circ}\text{C}$  and quantified using the Qubit dsDNA BR Kit as described in the 10x Genomics Technical Note: Chromium Genome Reagent Kits User Guide-Rev C (CG00022). The acceptable concentration of the diluted DNA is in the range of 0.8–1.2 ng/ $\mu\text{l}$ .

**Snake genome sequencing.** Purified gDNA was partitioned across Gel Bead-in-Emulsions (GEMs) using GemCode technology. The processes related to GEM generation and incubation, DNA barcoding, and genome library construction are described in the 10x Genomics Technical Note: Chromium Genome Reagent Kits User Guide-Rev C (CG00022). Illumina-ready sequencing libraries were generated using Chromium Genome Chip

and Chromium Controller. The resulting libraries were composed of P5 and P7 adaptors, Read 1, 16 bp 10x Barcode, gDNA insert, Read 2, and sample index ready for sequencing process. Read 1 sequence and the 10x Barcode were included during the GEM incubation whereas P5 and P7 primers, Read 2, and sample index were added during the library construction. Genome sequencing was carried out using the Illumina HiSeqX system to generate paired-end libraries with 150 bp read length. The final results of the sequencing process were outputted as FASTQ files. A conventional short-read sequencing was performed by using the Qiagen Blood and tissue DNeasy to isolate genomic DNA from liver tissues according to the manufacturer's description (Qiagen GmbH). The Paired-End Sequencing Sample Prep kit was used to prepare libraries from 5 µg of isolated gDNA according to the manufacturer's description (Illumina). Either a 200 bp band or a 500 bp band was cut from the gel (libraries PE200 and PE500, respectively). After amplification the resulting libraries were analyzed with an Agilent Bioanalyzer 2100 DNA 1000 series II chip according to the manufacturer's description (Agilent). Illumina HiSeq 2500 system will be used to perform short-read genome sequencing. Genomic libraries are paired-end sequenced with a read length of 150 nucleotides (150 bp). The results are genomic sequences in the form of FASTQ files. All-low quality reads (Q score < 30) were eliminated using Sickle<sup>22</sup>.

**De novo genome assembly.** The de novo assembly of the 10x Genomics linked-read sequencing was performed using Supernova version 2.1.1<sup>23</sup>. FASTQ files were used as input when executing the software, and parameters were configured as recommended by the software guidelines. The assembly process was done using a computer workstation with 32 CPU cores and 512 GB RAM. The assembly of conventional short reads was carried out using SOAPdenovo2 (version 2.04) executing on the National Institutes of Health (NIH) Biowulf cluster (<https://hpc.nih.gov/>).

**Venom gland transcriptome data generation.** Fresh tissue samples of the venom gland were harvested from a female *D. siamensis*. Illumina TruSeq RNA Sample Prep Kits were used for the RNA library preparation from these samples. The sequencing of the venom gland transcriptome was achieved using Illumina HiSeq 2500 system. Rcorrector<sup>24</sup> and Cutadapt<sup>25</sup> were used to correct sequencing errors and remove low-quality bases.

**Genome annotation.** Annotations for the assembled *Daboia siamensis* scaffolds were achieved using the automated genome annotation pipeline MAKER<sup>26</sup>. This software performs ab initio gene prediction, protein and transcript alignment, repeat identification, and 5' and 3' UTR inference. The final gene models along with quality control statistics are provided by MAKER. Inputs for MAKER included the assembled *Daboia siamensis* genome, repeats identified by RepeatModeler (<http://www.repeatmasker.org/RepeatModeler/>) and Repclass<sup>27</sup>, proteins corresponding to Serpentes suborder (taxon identifier 8570) downloaded from UniProt database<sup>28</sup>, and transcriptome data of the venom gland. Ab initio genes and coding DNA sequences (CDSs) were predicted by MAKER.

**Scaffold quality assessment.** The quality of scaffolds was assessed using BUSCO<sup>29</sup> to estimate the genome completeness score. In this study, the Vertebrate near-universal single-copy orthologs (SCOs) database was selected as a search database for quality assessment.

**Venom protein identification using in-gel and in-solution digestion techniques.** Venom proteins were separated by one-dimensional sodium dodecyl sulfate–polyacrylamide gel electrophoresis (1D SDS-PAGE). Briefly, 40 µg of crude venom was loaded onto bis–tris 12.5% acrylamide gels (Life Technologies). Gels were run at 200 V for 50 min, stained overnight with 0.1% Coomassie brilliant blue R-250, destained with water. Destained gels were scanned with the Odyssey infrared imaging system. Molecular weights were estimated based on protein standards. The gel was cut into 13 fractions from top to bottom. Each gel fraction was destained with 50% ACN/25 mM ammonium bicarbonate, reduced with 10 mM dithiothreitol (DTT) at 56 °C for 30 min, and alkylated with 55 mM iodoacetamide (IAA) at room temperature for 30 min in the dark. For trypsinization, 12.5 ng/µl trypsin solution was added to each sample and incubated for 60 min on ice. Excess trypsin was removed, an additional 50 µl of 25 mM ammonium bicarbonate was added to gel pieces, and samples were digested overnight at 37 °C. Following the overnight digestion, the samples were vortexed for 1 h at room temperature and sonicated in an ultrasonic bath for 3 min, and the tryptic peptides were extracted in 50 µl of 50% ACN/0.1% FA. The tryptic peptides were desalted with C18 StageTips. The samples were dried and stored at -80 °C for mass spectrometry analysis.

For in-solution digestion, 100 µg of crude venom was reconstituted in 8 M urea and added 10 mM DTT for reduction at 56 °C for 30 min. The sample was alkylated by adding 40 mM of IAA at room temperature for 30 min in the dark. Subsequently, 2 µg of trypsin was added and incubated overnight. On the following day, 10 µl of 100% FA was added, centrifuged at room temperature for 10 min, and then placed in a SpeedVac concentrator. The sample was reconstituted with 0.1% FA and fractionated by using Pierce high pH reversed-phase peptide fractionation kit (Thermo scientific). Peptides eluted from each fraction were dried and stored at -80 °C for mass spectrometry analysis.

**Nano-liquid chromatography-tandem mass spectrometry and database searches.** Peptides were separated by nano-liquid chromatography (EASY-nLC 1000, Thermo Fisher Scientific) coupled to a mass spectrometer (Q Exactive Plus Hybrid Quadrupole–Orbitrap, Thermo Fisher Scientific) through an EASY-Spray nanoelectrospray ion source (Thermo Fisher Scientific). The MS methods included a full MS scan at a resolution of 70,000 followed by 10 data dependent MS2 scans at a resolution of 17,500. The full MS scan range of

200–2000  $m/z$  was selected, and precursor ions with the charge states of +1 or greater than +8 were excluded. Normalized collision energy of HCD fragmentation was set at 28%. Raw LC–MS/MS files were searched using PEAKS Studio version 8.5 against UniProt proteins corresponding to suborder Serpentes and the CDS database predicted by MAKER. A target-decoy approach was used to limit a false discovery rate (FDR) of the identified peptides to less than 1%. Parent and fragment monoisotopic mass errors were set at 10 ppm. Carbamidomethylation of cysteine (C) was used as a fixed modification. Oxidation (M), acetylation (protein N-term), phosphorylation (STY), and deamidation (NQ) were set as variable modifications. A maximum of 1 missed cleavage was allowed.

**De novo sequencing-based MS/MS spectra search.** Additionally, raw LC–MS/MS files were searched by SMSNet<sup>30</sup> using the venom gland transcriptome, UniProt proteins corresponding to suborder Serpentes concatenated with the CDS database as the reference database. This software is implemented based on deep learning models incorporated with de novo sequencing processes for the MS/MS spectra interpretation. The SMSNet-M model which recognizes variable oxidation of methionine was used. Precursor mass tolerance was set at 30 ppm. The false discovery rate was set at 5% at amino acid level. Low confidence amino acid positions were transformed into mass tags. Partially identified peptides and peptides containing isoleucine/leucine were searched against the reference databases. For example, a partially identified peptides HG(196.1203)TSACK will be matched to all reference sequences that start with HG, end with TSACK, and contain a combination of amino acids whose total weight is 196.1203 in the middle. Only those with unique hits were retained as the final identified peptides. For example, since 196.1203 is roughly equal to the combined mass of a valine and a proline, if both HGVPTSACK and HGPVTSACK are present in the reference database, the partially identified peptide HG(196.1203)TSACK would be discarded because we could not distinguish between the two possibilities. All CDSs with at least one mapped peptide were compared with search results provided by PEAKS Studio.

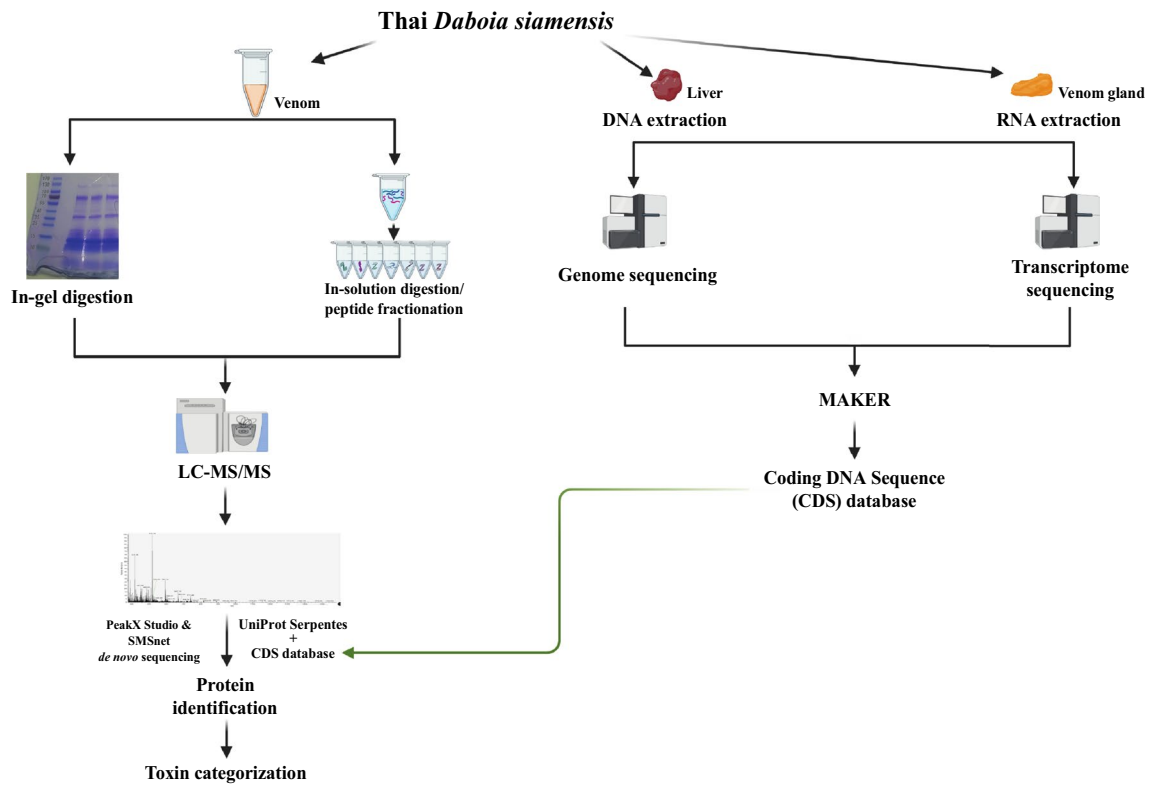
**Venomics studies using integrated genomics, transcriptomic, and proteomics data.** *Daboia siamensis* toxin genes, transcripts, and proteins were identified by (a) annotation process by MAKER and (b) BLAST search of predicted proteins against UniProt toxins corresponding to suborder Serpentes. Hereby, both entries in Swiss-Prot (manually annotated) and TrEMBL (automatically annotated) were used. For each entry of BLAST search, only the top hit protein with  $e\text{-value} \leq 1.00e-5$  and  $\%identities \geq 70$  was used for further analyses. Subsequently, identified toxins were clustered based on matching UniProt annotation into 2 major categories as described in the precedent research<sup>16</sup>: (1) toxins with enzymatic properties and (2) toxins with non-enzymatic properties. For the first group, there are 8 sub-categories: snake venom metalloproteinase (SVMP), snake venom serine protease (SVSP), phospholipase A<sub>2</sub> (PLA<sub>2</sub>), phospholipase B (PLB), L-amino acid oxidase (LAAO), Kunitz-type serine protease inhibitor (KSPI), hyaluronidases, and nucleases (RNase, DNase, and phosphodiesterase). For the second group, there are 11 sub-categories according to the well-known non-enzymatic protein families in snake venoms: cysteine-rich secretory proteins (CRISPs), snakecels, nerve growth factors (NGFs), natriuretic peptides, three-finger toxins (3FTs), sarafotoxins, cobra venom factors (CVFs), vascular endothelial growth factors (VEGFs), vesprins, waprins, and veficolins. Identified proteins were associated with their gene ontology (GO) terms using the “Retrieve/ID mapping” service provided by UniProt. By default, the GO term is not present in the mapping results. The option to show GO-related information must be selected first using the “edit column” button.

## Results and discussion

**Integrated multi-omics approach for comprehensive venomics profiling.** In order to comprehensively profile the venom of *Daboia siamensis*, we used integrated multi-omics data that incorporate genomics, transcriptomics, and proteomics layers (Fig. 1). Genome and transcriptome data from *Daboia siamensis* liver and venom gland, respectively, were first integrated to generate a more correct and complete coding DNA sequence (CDS) database. Subsequently, this CDS database was used for proteomic search of the venom sample analyzed via mass spectrometry. In addition to the database search strategy, a de novo search strategy was also applied to increase the protein identification in the venom sample. Finally, toxin categorization was performed on the identified proteins. The details and results of each step in this multi-omics approach were described as follows.

**Genome sequencing.** The sequencing process yielded 944.31 M paired-end reads. All raw reads were 150 bp in length with 69.38x coverage, exceeding the ideal coverage of 56x (Table 1). According to Supernova technical note<sup>31</sup>, higher coverage is occasionally advantageous in the assembly process. After trimming barcode sequences, the average length of reads was 138.5 bp, slightly lower than the ideal length at 140 bp. The estimated effective coverage of trimmed reads was 42.19x which is close to the ideal value. The percentage of Q30 bases in R2 reads and the median insert size were 76.81 and 361 bp, respectively. Both percentages were in the ideal ranges. Other measurements related to gDNA library are shown in the supplement Table S1.

**De novo assembly.** The assembly process generated a 1.67 Gb *Daboia siamensis* genome with a scaffold N50 of 1.10 Mb. The total number of scaffolds was 40.97 K. Among these assembled scaffolds, 9.23 K were longer than 10 Kb. The overall GC content was 38.89% (Table 2). Other measurements related to assembled scaffolds are shown in supplement Table S2. Sequences of genome scaffolds were deposited at The National Center for Biotechnology Information (NCBI) repository (BioProject: PRJNA794313). For the assembly quality assessment, genome completeness scores of 88% and 8% were estimated by BUSCO<sup>29</sup> based on complete and fragmented SCOs, respectively. The use of linked-read technology resulted in longer scaffolds compared to our study using



**Figure 1.** Workflow of our integrated multi-omics approach. The full-size figure of venom SDS-PAGE gel is shown in supplement Fig. S1.

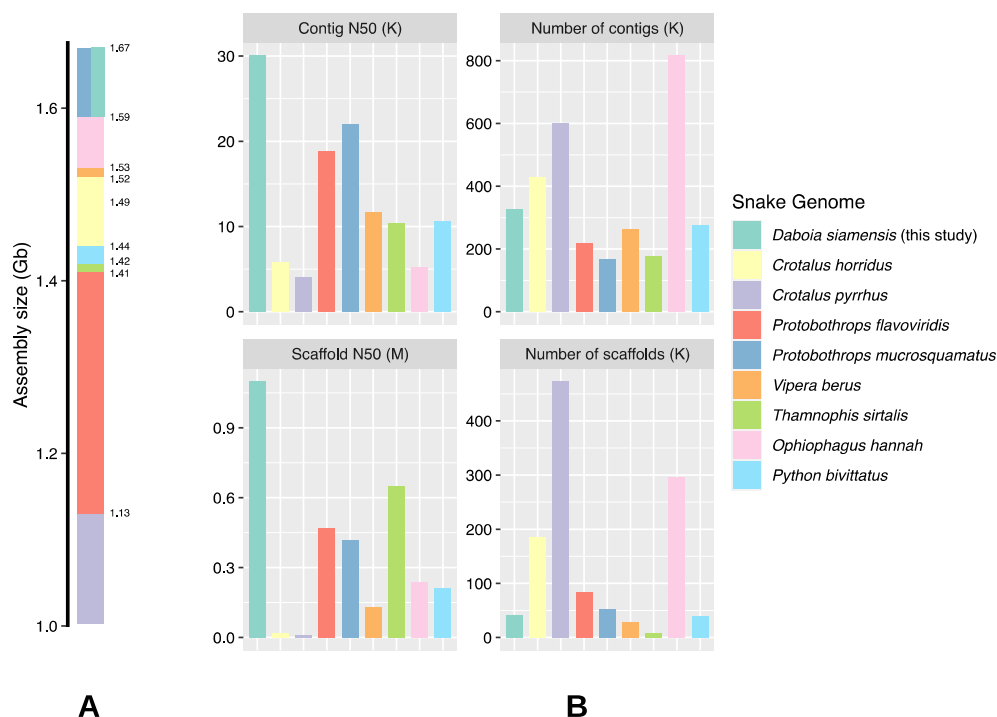
Measurement	Value from <i>Daboia siamensis</i> reads	Ideal value*
Number of reads	944.31 M	–
Read length	150 bp	–
Raw coverage	69.38x	~ 56x
Mean read length after trimming	138.5 bp	140 bp
Effective read coverage	42.19x	~ 42x
Fraction of Q30 bases in Read 2	76.81%	75–85%
Median insert size	361 bp	350–400 bp

**Table 1.** Statistics of raw sequenced reads used in the assembly process. \*The ideal statistics are defined by Supernova<sup>33</sup>.

Measurement	Value
Assembly size	1.67 Gb
Number of contigs	4.11 M
Number of scaffolds	40.97 K
Number of scaffolds with length ≥ 10 Kb	9.23 K
N50 contig size	30,074
N50 scaffold size	1.10 Mb
GC content of assembly	38.89%

**Table 2.** Statistics of assembled scaffolds for genomic data.

only a conventional short-read technology. The assembly of short reads was carried out using SOAPdenovo<sup>32</sup> which yielded 141 K scaffolds with N50 of 17 Kb, approximately 71 folds shorter than current scaffolds (Table S3). According to Fig. 2, the assembly sizes of published snake genomes range from 1.1 to 1.7 Gb. Contig and scaffold



**Figure 2.** Comparisons of snake genome assemblies. The overall assembly sizes (A) and contig and scaffold statistics (B) of *Daboia siamensis* (this study), *Crotalus horridus*<sup>38</sup>, *Crotalus pyrrhus*<sup>35</sup>, *Protobothrops flavoviridis*<sup>34</sup>, *Protobothrops mucrosquamatus*<sup>39</sup>, *Vipera berus* (GenBank assembly accession: GCA\_000800605.1), *Thamnophis sirtalis*<sup>40</sup>, *Ophiophagus Hannah*<sup>33</sup>, and *Python bivittatus*<sup>19</sup>.

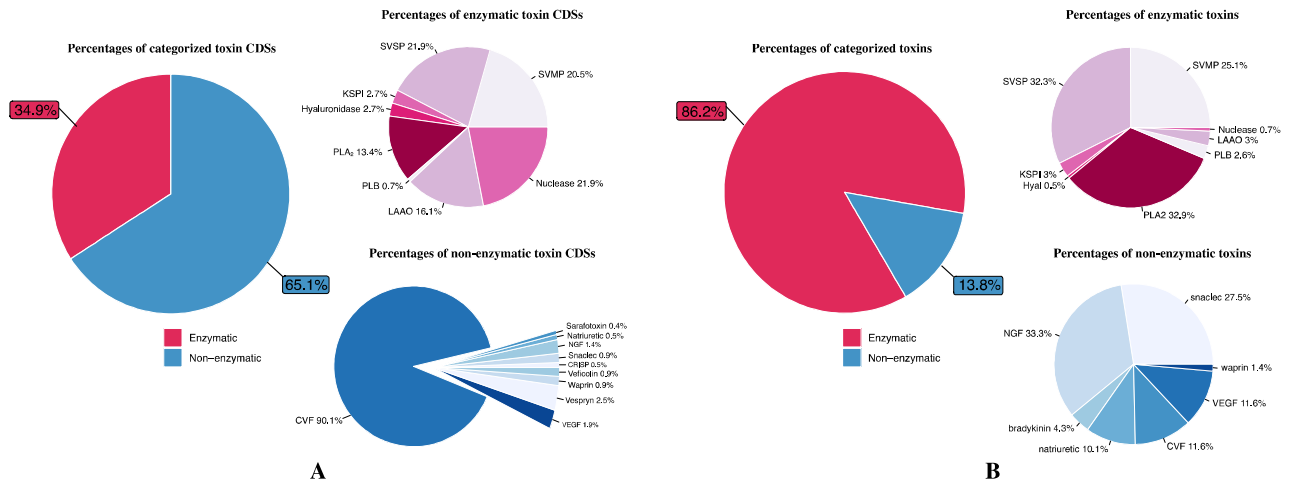
Region of	Number of predicted regions	Total region length (Mb)
Genes	19,691	87.77
Exons	72,598	14.85
Introns	52,826	73.18
CDS	19,772	14.04

**Table 3.** Number of genes, exons, introns, and mRNAs identified by MAKER.

N50s of *Daboia siamensis* assembly are longer than other snake genomes such as *Python bivittatus*<sup>19</sup>, *Ophiophagus Hannah*<sup>33</sup>, *Protobothrops flavoviridis*<sup>34</sup>, and *Crotalus pyrrhus*<sup>35</sup>. However, when comparing *Daboia siamensis*'s scaffold N50 with the recently published genomes of *Naja naja*<sup>36</sup> (scaffold N50 = 223.35 Mb) and *Pantherophis guttatus*<sup>37</sup> (scaffold N50 = 16.79 Mb), our scaffold N50 is shorter. These recent studies employed and ensembled scaffolds from multiple state-of-the-art long-read technologies such as 10x linked-read, PacBio, Chicago, and Oxford Nanopore altogether, thus scaffolds with exceptionally long N50 can be achieved.

**Annotation results.** In total, 19,691 genes were predicted from *Daboia siamensis* assembled genome by using the MAKER annotation pipeline (Table 3). Regions of genes cover ~5% of the assembled genome. From these regions, 19,772 CDSs were identified which cover ~0.8% of the assembled genome and ~15.99% of gene regions. These CDSs were compiled into the “CDS database” for the mass spectrometry search as described in the “Venom proteome” section below. Note that the transcriptome data used in the annotation process was generated from the *Daboia siamensis* venom gland. Therefore, many annotated CDSs are related to snake venoms. The detailed annotation report from MAKER is in supplement Table S4.

**Classification of venom CDS.** CDSs were categorized as described in “Methods”, “Venomics studies using integrated genomics, transcriptomic, and proteomics data” section (see supplement Table S5 for a full list of CDSs). From all 19,772 CDSs, 17,676 (~89.4%) can be annotated by using BLAST search and 915 CDSs can be categorized into two major toxin categories, enzymatic and non-enzymatic toxin groups (Fig. 3A). In the enzymatic group, there are 319 CDSs in total (34.9%). These CDSs were classified into 6 subgroups. The top-three common subgroups belong to SVSP, nuclease, and SVMMP followed by LAAO and PLA<sub>2</sub> subgroups. The least common subgroup is PLB which includes only 2 CDSs (2.8%). The non-enzymatic group includes 596 CDSs



**Figure 3.** Percentages of CDSs (A) and proteins (B) categorized in enzymatic and non-enzymatic groups.

(65.1%). These CDSs were divided into 7 subgroups. The most common subgroup is CVF which includes 534 CDSs (90.1%). Only 18 CDSs were classified as vespryn (2.5%), and the remaining 8 subgroups include 44 CDSs (7.4%).

**Venom proteome.** Proteins from *Daboia siamensis* venom gland were processed for mass spectrometry using two techniques: (1) SDS-PAGE followed by in-gel digestion and (2) in-solution digestion followed by high pH peptide fractionation to separate complex proteins and detect low abundance proteins. The generated MS/MS spectra were searched against the UniProt's Serpentes reference database (taxon identifier 8570) concatenated with the CDS database. The search results yielded 10,131 and 4367 unique peptides from in-gel and in-solution digestion techniques, respectively. The combined list of peptides contains 17,163 unique peptides corresponding to 1662 unique proteins (supplement Table S6). Subsequently, all identified proteins were clustered into two categories: enzymatic and non-enzymatic proteins as described in “Methods”, “Venomics studies using integrated genomics, transcriptomic, and proteomics data” section. The proteome is dominated by the enzymatic group (86.2%), whereas only 13.8% of proteins belong to the non-enzymatic group (Fig. 3B). The enzymatic group consists of the following protein families: PLA<sub>2</sub> 32.9%, SVSP 32.3%, SVMP 25.1%, LAO 3%, KSPI 3%, PLB 2.6%, hyaluronidase 0.5%, and nuclease 0.7%. PLA<sub>2</sub> family was found at the highest proportion among the enzymatic group, consistent with previous reports in Asian *Daboia siamensis* and Russell's viper<sup>41–43</sup>. It is a family of enzymes that cleave glycerophospholipids to dissociate fatty acids tail. The pharmacological effect of this enzyme is well known, such as coagulant, hemorrhagic, hemolytic, edema forming activities, and neurotoxicity. Whereas neurotoxic effect is a common feature of *Daboia russelii* venom in Sri Lanka and some parts of India, it is less common in Asian *Daboia siamensis*. SVSP and SVMP are commonly found in this venom and have hemotoxic activity. To contrast the hemolytic activity, Op B et al. (2021) compared the venom of *Daboia siamensis* from Thailand, Myanmar, and Taiwan. They found that Myanmar *Daboia siamensis* venom has the most hemotoxic effect<sup>44</sup>. In the present study, we can detect LAO toxin in the venom of *Daboia siamensis* from Thailand, similar to the reports from China and Myanmar<sup>42,45</sup>. In contrast, Changra TM (2020) reported that LAO was not detected in venoms from Thailand and Taiwan *Daboia siamensis* venom<sup>43</sup>. LAO does not directly cause hemorrhage, but it can cause an impairment of platelet aggregation via hydrogen peroxide production due to its catalyzing effects on the L-amino acid substrate. Interestingly, many proteins identified in this study have not been found in the published *Daboia siamensis* protein databases and reports, as shown in Table 4. For example, this study first discovered the enzyme hyaluronidase-1 (CDS ID 16116 with 62% sequence coverage) as a component of *Daboia siamensis* venom. The pairwise alignment showed that *Daboia siamensis* hyaluronidase-1 is 95.4% identical to Cerastes hyaluronidase-1 sequence (UniProt accession: A3QVN3, see Fig. S2). Hyaluronidase-1 enzyme degrades hyaluronic acid, a glycosaminoglycan commonly found in the extracellular matrix, to promote the local hemorrhagic effect. It is called a “spreading factor” because it promotes the spread of the toxin through the tissues and blood circulation of the prey. This enzyme could play an important role in destroying tissue at the bite site in the victim<sup>46</sup>. PLB, another rare toxin, was identified with 43% sequence coverage against CDS ID 14593. PLB is a protein in the phospholipase superfamilies. PLB toxin in *Pseudechis colletti* venom was reported to have hemolytic and cytotoxic activities in vitro in human erythrocytes<sup>47</sup>.

The non-enzymatic group consists of NGF 33.3%, snaclec 27.5%, VEGF 11.6%, CVF 11.6%, natriuretic peptides 10.1%, bradykinin 4.3%, and waprin 1.4% (Fig. 3B). Snaclec is found in Asian *Daboia siamensis* venom. It has been reported to possess platelet-modulating activity that enhances the activity of SVMP to induce hemorrhage<sup>48</sup>. Whereas VEGF, CVF, and NGF show potential in inducing hypotension and enhancing vascular permeability leading to bleeding complications<sup>49</sup>. We also found waprin toxin, a rare toxin that has not been reported in *Daboia siamensis*, but reported in another Russell's viper species, *Daboia russelii* (GenBank accession: ASU45069). We identified a single peptide with high confidence (with various post-translational modifications, see Table S6), and the CDS of this protein is highly conserved across venomous snake species (Fig. 4). This protein

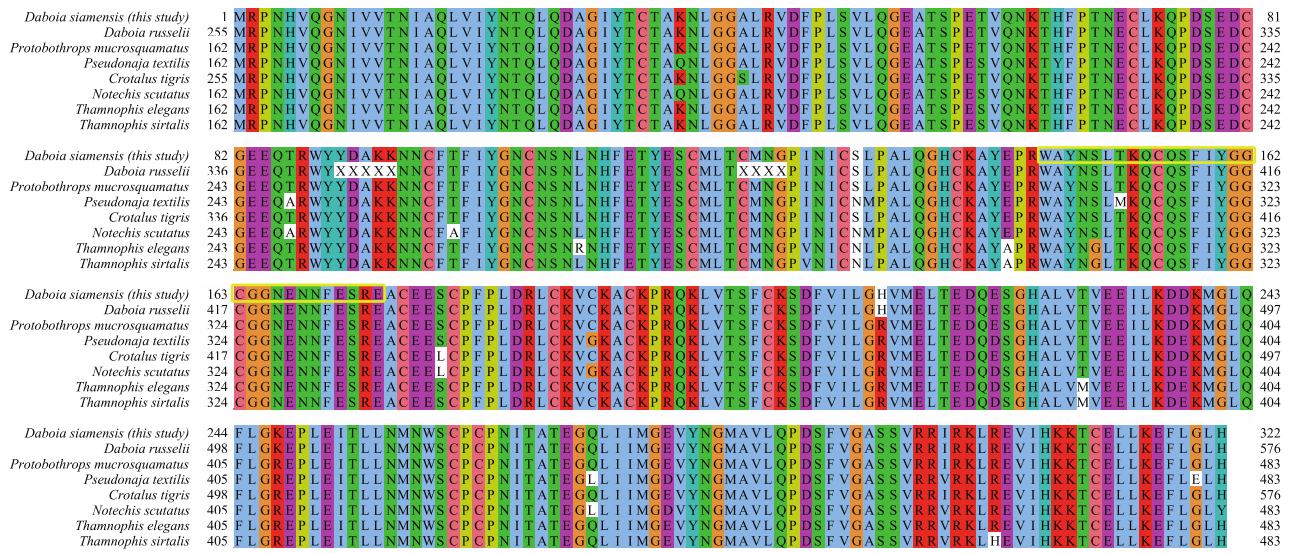
Venom category	Protein name	CDS ID	Source of specimen					Pathophysiological effects	
			Thailand (this study)	Thailand <sup>43</sup>	Indonesia <sup>43</sup>	Taiwan <sup>42</sup>	China <sup>42</sup>		Myanmar <sup>41</sup>
KSPI	Kunitz-type serine protease inhibitor C6	6337	✓	✓	✓	✓	✓	✓	Coagulopathy
PLA <sub>2</sub>	Acid PLA <sub>2</sub>	9491	✓	✓	✓	✓	✓	✓	Coagulopathy, neurotoxicity, edema
	Basic PLA <sub>2</sub>	11762	✓	✓	✓	✓	✓	✓	
	Phospholipase A2 inhibitor	2357	✓						
	Phospholipase A2 inhibitor subunit gamma B	3191	✓						
PLB	Phospholipase B	14593	✓						Hemotoxic activity
SVSP	Vaa serine proteinase homolog 1	8750	✓						Hemotoxic activity, coagulopathy
	Complement factor I	6095	✓						
	Peptidase S1 domain-containing protein	4112	✓						
	Factor V activator RVV-V gamma	18736	✓	✓	✓	✓		✓	
	Serine protease VLSP-1	370	✓					✓	
	Serine protease VLSP-3	18982	✓	✓	✓			✓	
	Beta-fibrinogenase-like	2680	✓	✓	✓			✓	
	Alpha-fibrinogenase-like	10579	✓	✓	✓	✓		✓	
SVMP	Coagulation factor X-activating enzyme heavy chain protein	2879	✓	✓	✓	✓		✓	Hemotoxic activity
	Snaclec coagulation factor X-activating enzyme light chain 1	4878	✓	✓	✓	✓		✓	
	Antihemorrhagic factor cHLP-B protein	18915	✓						
	Antihemorrhagic factor jMSF protein	3892	✓						
LAAO	L-amino-acid oxidase	14325	✓	✓	✓	-	✓	✓	Coagulopathy
	Amine oxidase	3374	✓						
Nuclease	Deoxyribonuclease II	8825	✓						Hemotoxic activity
	Angiogenin-2	437	✓						
Snaclec	C-type lectin-like protein subunit 2	4878	✓	✓	✓	✓	✓	✓	Hemorrhage
	C-type lectin snaclec-1	15452	✓						
CVF	Actin-depolymerizing factor	13591	✓						Bleeding, complications
NGF	Beta-nerve growth factor	4415	✓						
VEGF	Snake venom vascular endothelial growth factor toxin VR-1	9987	✓	✓	✓	✓	✓	✓	
Amino peptidase	Xaa-Pro aminopeptidase 2	6377	✓			✓	✓	✓	Hypertension
	Aminopeptidase	10654	✓					✓	
	Tripeptidyl aminopeptidase	6651	✓						
Cathepsin	Cathepsin B	1357	✓						Procoagulant
	Cathepsin H	6109	✓						
Waprin	WAP, follistatin/kazal, immunoglobulin, kunitz and netrin domain containing 2	16646	✓						Anti-microbial
Hyaluronidase	Hyaluronidase-1	16116	✓						Hemorrhage, inflammation
<b>Protein digestion methods for mass spectrometry</b>									
In-solution digestion			✓	✓	✓				
In-gel digestion			✓			✓	✓		
In-gel digestion (2D gel electrophoresis)								✓	

**Table 4.** Potential venom proteins identified from a different source of *Daboia siamensis* specimens.

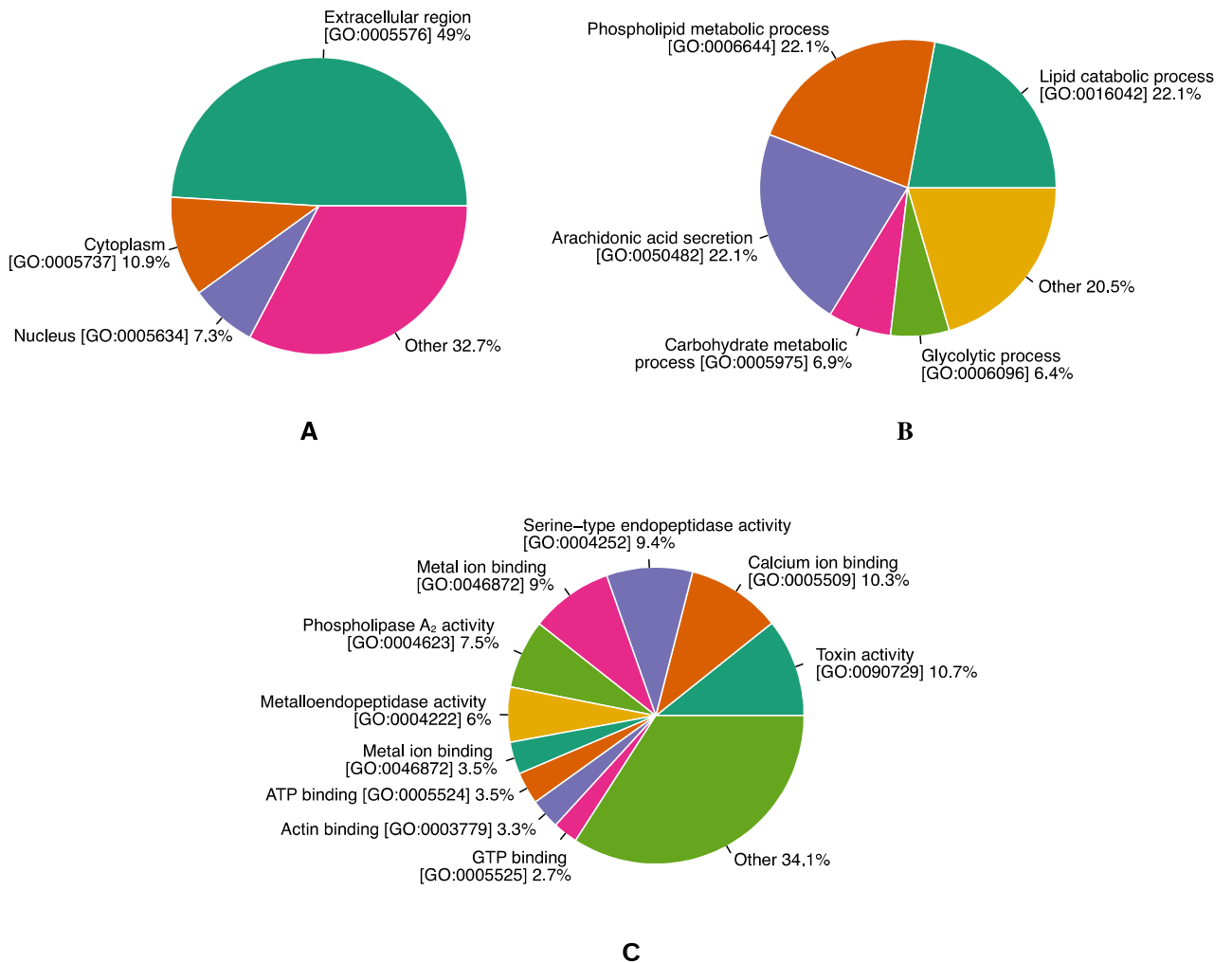
exhibits proteinase inhibitor activity with potent antimicrobial activity<sup>50</sup>. We summarize the pathophysiological effects of all toxins identified in this study in Table 4.

All proteins identified in the venom were assigned to functional group-based Gene Ontology (GO) terms using the “Retrieve/ID mapping” service provided by UniProt database as described in “Methods”, “Venomics studies using integrated genomics, transcriptomic, and proteomics data” section. Figure 5 shows GO annotation of putative *Daboia siamensis* venom proteome. Most of the identified venom proteins (49%) seem to function in extracellular regions since SVMP, SVSP, LAAO, PLA<sub>2</sub>, and KSPI are located in extracellular regions (Fig. 5A). The top-five GO molecular function terms were annotated as toxin activity, calcium ion binding, serine-type

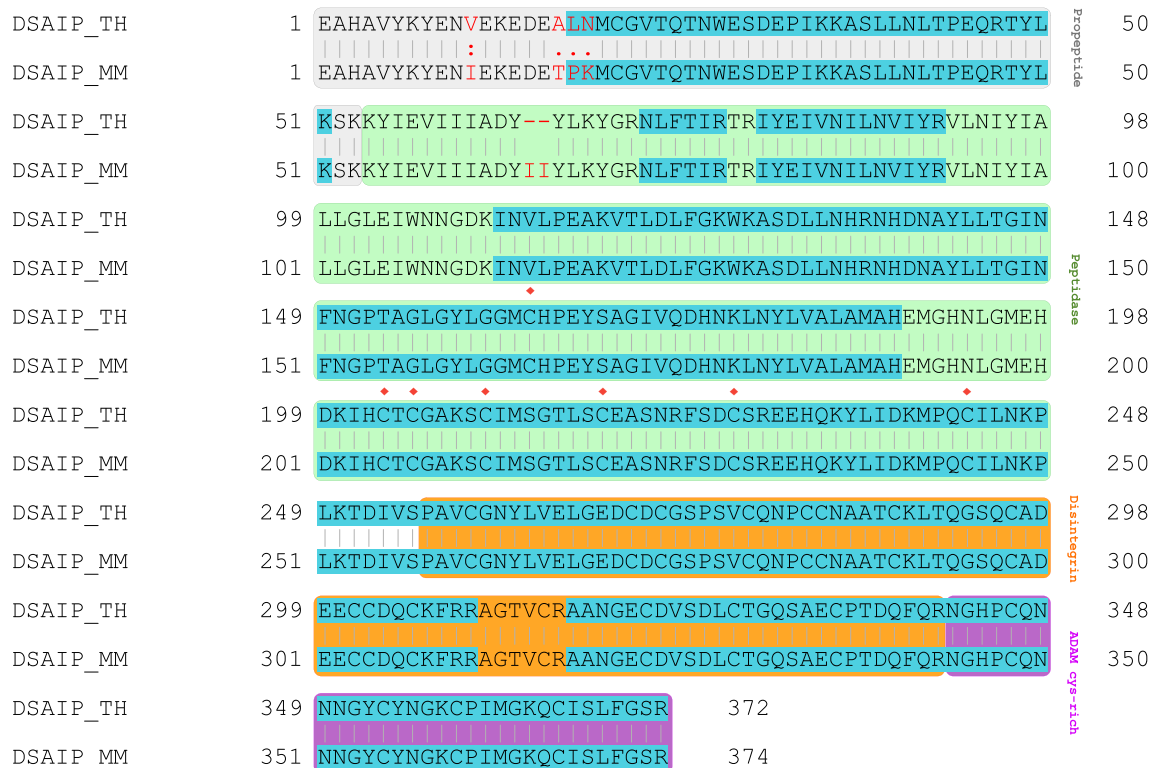




**Figure 4.** Multiple alignment of waprins protein from various venomous snakes. The NCBI reference sequence/ GenBank accession numbers for *Daboia russelii*, *Protobothrops mucrosquamatus*, *Pseudonaja textilis*, *Crotalus tigris*, *Notechis scutatus*, *Thamnophis elegans*, and *Thamnophis sirtalis* waprins are ASU45069, XP\_029139923, XP\_026557480, XP\_039199862, XP\_026523009, XP\_032094882, and XP\_013912228, respectively. The single peptide of *Daboia siamensis* waprins identified in this study is highlighted in the yellow box.



**Figure 5.** Percentages of functional group-based GO terms calculated from *Daboia siamensis* venom proteome according to cellular component terms (A), biological process terms (B), and molecular function terms (C).



**Figure 6.** Pairwise sequence alignment of Thai (TH) and Myanmar (MM) DSAIP proteins. The grey, green, orange, and purple boxes denote regions of propeptide, peptidase, disintegrin, and ADAM cys-rich, respectively. Alphabets in red indicate differences between Thai and Myanmar DSAIP proteins. Blue-highlighted alphabets are regions covered by SMSNet search results. Red diamonds denote conserved cysteine residues.

endopeptidase activity, metal ion binding, and phospholipase A<sub>2</sub> activity. These terms are consistent with the activities of component proteins shown in the venom proteome (Fig. 3B).

**Venom proteins identified by de novo sequencing-based MS/MS spectra searching.** A total of 903 CDSs were identified by SMSNet (Table S7). Compared to the search results generated by PEAKS Studio, 175 proteins are overlapped (Table S8). Interestingly, 728 CDSs were solely identifiable by SMSNet. One of these CDSs is identified as *Daboia siamensis* apoptosis-inducing protease (DSAIP), a P-III class SVMP, first discovered in Burmese *Daboia siamensis*<sup>41</sup> (UniProt accession A0A2H4Z2X4) and also found in China<sup>42</sup> and Taiwan<sup>51</sup>. DSAIP protein is a homolog of *Vipera lebetina* apoptosis-inducing protease (VLAIP) which is also found in Russell's vipers endemic to Sri Lanka<sup>52</sup> and Pakistan<sup>53</sup> but not found in specimens from Western Indian<sup>54</sup> and Southern Indian<sup>55</sup>.

Peptide sequences identified from SMSNet also validate the difference between Thai and Burmese DSAIP proteins as illustrated in the pairwise sequence alignment (Fig. 6). Amino acids leucine and asparagine at positions 17 and 18 of Thai DSAIP are consistent with proteomics results (Table S7). This may indicate one of the divergences between Thai and Myanmar DSAIP proteins. All differences at positions 11, 17, 18, and 19 are located in the propeptide region. However, sequences in peptidase, disintegrin, and ADAM cysteine-rich regions are identical. The majority of P-III class SVMPs consists of six conserved cysteinyl residues in the peptidase region, but for DSAIP there are seven conserved cysteinyl residues. The seventh cysteinyl residue is involved in the apoptosis-inducing activity<sup>56</sup>. In this study, seven conserved cysteinyl residues were identified in the peptidase region (Fig. 6).

In addition, many proteins identified by SMSNet have not been reported as a component of *Daboia siamensis* venom (Table S9). Similar to the use of PEAK Studio, a number of proteins were detectable when using CDSs as a reference database but undetectable when using the conventionally UniProt's Serpentes reference database. These findings support the use of integrated multi-omics data as an improved approach for discoveries in venom studies.

## Conclusion

This study incorporates genomics, transcriptomics, and proteomics data to study venomomics of *Daboia siamensis*. Using integrated multi-omics data, we found numerous proteins that have not been reported as *Daboia siamensis* venom components such as hyaluronidase-1, phospholipase B, and wiprin. Our findings indicate that the use of integrated multi-omics data is advantageous for venomomics studies. The integrated multi-omics data along with our discoveries will be valuable not only for antivenom production but also for the development of novel therapeutics.

## Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD030613. Genome data have been deposited at The National Center for Biotechnology Information (NCBI) repository (BioProject: PRJNA794313).

Received: 10 February 2022; Accepted: 22 July 2022

Published online: 30 July 2022

## References

- Ahmed, S. M. *et al.* Emergency treatment of a snake bite: Pearls from literature. *J. Emerg. Trauma Shock* **1**, 97–105. <https://doi.org/10.4103/0974-2700.43190> (2008).
- Chippaux, J. P. Snake-bites: Appraisal of the global situation. *Bull. World Health Organ.* **76**, 515–524 (1998).
- Jimenez, R., Ikonopoulou, M. P., Lopez, J. A. & Miles, J. J. Immune drug discovery from venoms. *Toxicon* **141**, 18–24. <https://doi.org/10.1016/j.toxicon.2017.11.006> (2018).
- Jenkins, T. P. & Laustsen, A. H. Cost of manufacturing for recombinant snakebite antivenoms. *Front. Bioeng. Biotechnol.* **8**, 703. <https://doi.org/10.3389/fbioe.2020.00703> (2020).
- Silva, A. & Isbister, G. K. Current research into snake antivenoms, their mechanisms of action and applications. *Biochem. Soc. Trans.* **48**, 537–546. <https://doi.org/10.1042/Bst20190739> (2020).
- Valente, R. H., Nicolau, C. A., Perales, J. & Neves-Ferreira, A. G. D. C. Snake venom proteoepitomics: What lies behind the curtain. In *Venom Genomics and Proteomics* (ed. Juan J. Calvete), 333–365 (Springer Dordrecht Heidelberg New York London, 2016). <https://doi.org/10.1007/978-94-007-6649-56-1>.
- Risch, M. *et al.* Snake venomomics of the Siamese Russell's viper (*Daboia russelli siamensis*)—Relation to pharmacological activities. *J. Proteom.* **72**, 256–269. <https://doi.org/10.1016/j.jpro.2009.01.006> (2009).
- Laustsen, A. H., Johansen, K. H., Engmark, M. & Andersen, M. R. Recombinant snakebite antivenoms: A cost-competitive solution to a neglected tropical disease?. *PLoS Negl. Trop. Dis.* **11**, e0005361. <https://doi.org/10.1371/journal.pntd.0005361> (2017).
- Gutierrez, J. M. *et al.* Snakebite envenoming. *Nat. Rev. Dis. Primers* **3**, 17063. <https://doi.org/10.1038/nrdp.2017.63> (2017).
- Theakston, R. D. & Warrell, D. A. Crisis in snake antivenom supply for Africa. *Lancet* **356**, 2104 (2000).
- Chippaux, J. P., Williams, V. & White, J. Snake venom variability: Methods of study, results and interpretation. *Toxicon* **29**, 1279–1303 (1991).
- Kamiguti, A. S., Theakston, R. D., Sherman, N. & Fox, J. W. Mass spectrophotometric evidence for P-III/P-IV metalloproteinases in the venom of the Boomslang (*Dispholidus typus*). *Toxicon* **38**, 1613–1620 (2000).
- Quinton, L. *et al.* Characterization of toxins within crude venoms by combined use of Fourier transform mass spectrometry and cloning. *Anal. Chem.* **77**, 6630–6639. <https://doi.org/10.1021/ac050575k> (2005).
- Liu, C. C. *et al.* Snake venom proteome of *Protobothrops mucrosquamatus* in Taiwan: Delaying venom-induced lethality in a rodent model by inhibition of phospholipase A2 activity with varespladib. *J. Proteom.* **234**, 104084. <https://doi.org/10.1016/j.jpro.2020.104084> (2021).
- Dowell, N. L. *et al.* 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> (2016).
- Mohamed Abd El-Aziz, T., Garcia Soares, A. & Stockand, J. D. Snake venoms in drug discovery: Valuable therapeutic tools for life saving. *Toxins (Basel)* <https://doi.org/10.3390/toxins11100564> (2019).
- Kerckamp, H. M. *et al.* Snake genome sequencing: Results and future prospects. *Toxins (Basel)* <https://doi.org/10.3390/toxins8120360> (2016).
- Chan, Y. S. *et al.* Snake venom toxins: Toxicity and medicinal applications. *Appl. Microbiol. Biotechnol.* **100**, 6165–6181. <https://doi.org/10.1007/s00253-016-7610-9> (2016).
- Castoe, T. A. *et al.* The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20645–20650. <https://doi.org/10.1073/pnas.1314475110> (2013).
- Simoes, B. F. *et al.* Visual system evolution and the nature of the ancestral snake. *J. Evol. Biol.* **28**, 1309–1320. <https://doi.org/10.1111/jeb.12663> (2015).
- Olivera, B. M., Seger, J., Horvath, M. P. & Fedosov, A. E. Prey-capture strategies of fish-hunting cone snails: Behavior, neurobiology and evolution. *Brain Behav. Evol.* **86**, 58–74. <https://doi.org/10.1159/000438449> (2015).
- Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files (*Version 1.33*) (2011).
- Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767. <https://doi.org/10.1101/gr.214874.116> (2017).
- Song, L. & Florea, L. Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. *Gigascience* **4**, 48. <https://doi.org/10.1186/s13742-015-0089-y> (2015).
- Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
- Holt, C. & Yandell, M. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* **12**, 491. <https://doi.org/10.1186/1471-2105-12-491> (2011).
- Feschotte, C., Keswani, U., Ranganathan, N., Guibotsy, M. L. & Levine, D. Exploring repetitive DNA landscapes using REPEATCLASS, a tool that automates the classification of transposable elements in eukaryotic genomes. *Genome Biol. Evol.* **1**, 205–220. <https://doi.org/10.1093/gbe/evp023> (2009).
- The UniProt, C. UniProt: The universal protein knowledgebase. *Nucl. Acids Res.* <https://doi.org/10.1093/nar/gky092> (2018).
- Seppy, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227–245. [https://doi.org/10.1007/978-1-4939-9173-0\\_14](https://doi.org/10.1007/978-1-4939-9173-0_14) (2019).
- Karunratanakul, K., Tang, H. Y., Speicher, D. W., Chuangsuwanich, E. & Sriswasdi, S. Uncovering thousands of new peptides with sequence-mask-search hybrid de novo peptide sequencing framework. *Mol. Cell. Proteom.* **18**, 2478–2491. <https://doi.org/10.1074/mcp.TIR119.001656> (2019).
- 10XGenomics. *Achieving Success with De Novo Assembly*. <https://support.10xgenomics.com/de-novo-assembly/guidance/doc/achieving-success-with-de-novo-assembly> (2018).
- Luo, R. *et al.* SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18. <https://doi.org/10.1186/2047-217X-1-18> (2012).
- Vonk, F. J. *et al.* The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20651–20656. <https://doi.org/10.1073/pnas.1314702110> (2013).
- Shibata, H. *et al.* The habu genome reveals accelerated evolution of venom protein genes. *Sci. Rep.* **8**, 11300. <https://doi.org/10.1038/s41598-018-28749-4> (2018).
- Gilbert, C. *et al.* Endogenous hepadnaviruses, bornaviruses and circoviruses in snakes. *Proc. Biol. Sci.* **281**, 20141122. <https://doi.org/10.1098/rspb.2014.1122> (2014).
- Suryamohan, K. *et al.* 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> (2020).

37. Ullate-Agote, A. *et al.* Genome mapping of a LYST mutation in corn snakes indicates that vertebrate chromatophore vesicles are lysosome-related organelles. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 26307–26317. <https://doi.org/10.1073/pnas.2003724117> (2020).
38. Hall, J. B., Cobb, V. A. & Cahoon, A. B. The complete mitochondrial DNA sequence of *Crotalus horridus* (timber rattlesnake). *Mitochondrial DNA* **24**, 94–96. <https://doi.org/10.3109/19401736.2012.722999> (2013).
39. Aird, S. D. *et al.* Population genomic analysis of a pitviper reveals microevolutionary forces underlying venom chemistry. *Genome Biol. Evol.* **9**, 2640–2649. <https://doi.org/10.1093/gbe/evx199> (2017).
40. McGlothlin, J. W. *et al.* Parallel evolution of tetrodotoxin resistance in three voltage-gated sodium channel genes in the garter snake *Thamnophis sirtalis*. *Mol. Biol. Evol.* **31**, 2836–2846. <https://doi.org/10.1093/molbev/msu237> (2014).
41. Yee, K. T. *et al.* Analysis of snake venom metalloproteinases from Myanmar Russell's viper transcriptome. *Toxicon* **146**, 31–41. <https://doi.org/10.1016/j.toxicon.2018.03.005> (2018).
42. Tan, K. Y., Tan, N. H. & Tan, C. H. Venom proteomics and antivenom neutralization for the Chinese eastern Russell's viper, *Daboia siamensis* from Guangxi and Taiwan. *Sci. Rep.* **8**, 8545. <https://doi.org/10.1038/s41598-018-25955-y> (2018).
43. Lingam, T. M. C., Tan, K. Y. & Tan, C. H. Proteomics and antivenom immunoprofiling of Russell's viper (*Daboia siamensis*) venoms from Thailand and Indonesia. *J. Venom Anim. Toxins Incl. Trop. Dis.* **26**, e20190048. <https://doi.org/10.1590/1678-9199-JVATI-TD-2019-0048> (2020).
44. Op den Brouw, B. *et al.* A genus-wide bioactivity analysis of *Daboia* (Viperinae: Viperidae) viper venoms reveals widespread variation in haemotoxic properties. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms222413486> (2021).
45. Belt, P. J., Malhotra, A., Thorpe, R. S., Warrell, D. A. & Wuster, W. In *Symposia of the Zoological Society of London* 219–234 (The Society, 1960–1999).
46. Fox, J. W. A brief review of the scientific history of several lesser-known snake venom proteins: L-amino acid oxidases, hyaluronidases and phosphodiesterases. *Toxicon* **62**, 75–82. <https://doi.org/10.1016/j.toxicon.2012.09.009> (2013).
47. Bernheimer, A. W., Linder, R., Weinstein, S. A. & Kim, K. S. Isolation and characterization of a phospholipase B from venom of Collett's snake, *Pseudechis colletti*. *Toxicon* **25**, 547–554. [https://doi.org/10.1016/0041-0101\(87\)90290-x](https://doi.org/10.1016/0041-0101(87)90290-x) (1987).
48. Sajevic, T., Leonardi, A. & Krizaj, I. Haemostatically active proteins in snake venoms. *Toxicon* **57**, 627–645. <https://doi.org/10.1016/j.toxicon.2011.01.006> (2011).
49. Yamazaki, Y. *et al.* Snake venom vascular endothelial growth factors (VEGF-Fs) exclusively vary their structures and functions among species. *J. Biol. Chem.* **284**, 9885–9891. <https://doi.org/10.1074/jbc.M809071200> (2009).
50. Thompson, R. C. & Ohlsson, K. Isolation, properties, and complete amino acid sequence of human secretory leukocyte protease inhibitor, a potent inhibitor of leukocyte elastase. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 6692–6696. <https://doi.org/10.1073/pnas.83.18.6692> (1986).
51. Sanz, L. *et al.* Translational venomomics: Third-generation antivenomics of anti-Siamese Russell's viper, *Daboia siamensis*, antivenom manufactured in Taiwan CDC's vaccine center. *Trop. Med. Infect. Dis.* <https://doi.org/10.3390/tropicalmed3020066> (2018).
52. Tan, N. H. *et al.* Functional venomomics of the Sri Lankan Russell's viper (*Daboia russelii*) and its toxinological correlations. *J. Proteom.* **128**, 403–423. <https://doi.org/10.1016/j.jprot.2015.08.017> (2015).
53. Mukherjee, A. K., Kalita, B. & Mackessy, S. P. A proteomic analysis of Pakistan *Daboia russelii russelii* venom and assessment of potency of Indian polyvalent and monovalent antivenom. *J. Proteom.* **144**, 73–86. <https://doi.org/10.1016/j.jprot.2016.06.001> (2016).
54. Kalita, B., Patra, A. & Mukherjee, A. K. Unraveling the proteome composition and immuno-profiling of Western India Russell's viper venom for in-depth understanding of its pharmacological properties, clinical manifestations, and effective antivenom treatment. *J. Proteome Res.* **16**, 583–598. <https://doi.org/10.1021/acs.jproteome.6b00693> (2017).
55. Sharma, M., Das, D., Iyer, J. K., Kini, R. M. & Doley, R. Unveiling the complexities of *Daboia russelii* venom, a medically important snake of India, by tandem mass spectrometry. *Toxicon* **107**, 266–281. <https://doi.org/10.1016/j.toxicon.2015.06.027> (2015).
56. Olaoba, O. T., Karina Dos Santos, P., Selistre-de-Araujo, H. S. & Ferreira de Souza, D. H. Snake venom metalloproteinases (SVMPs): A structure-function update. *Toxicon X* **7**, 100052. <https://doi.org/10.1016/j.toxcx.2020.100052> (2020).

## Acknowledgements

This study was supported by National Research Council of Thailand (Contract No. DBG6180012), Thailand Science research and Innovation Fund Chulalongkorn University (CU\_FRB65\_hea(49)\_056\_30\_37) and (CU\_FRB65\_hea(44)\_051\_30\_32), and International SciKU Branding (ISB), Faculty of Science, Kasetsart University. We are thankful to all members of Center of Excellence in Systems Biology for valuable comments and advice.

## Author contributions

Conception and design of study: T.S. and T.P.; acquisition of data: P.S., S.P., L.C., V.S.; analysis and/or interpretation of data: T.S., P.S., K.T.Y., M.A.K., and S.S.; Drafting the manuscript: T.S., P.S., K.H., T.P., and K.T.Y.; Approval of the version of the manuscript to be published: T.S., P.S., S.P., S.S., K.T.Y., K.H., M.A.K., L.C., V.S., and T.P.

## Funding

Open Access funding provided by the National Institutes of Health (NIH). This research project was supported by National Research Council of Thailand [Contract No. DBG6180012], Ratchadapisek Sompoch Endowment Fund [RA61/103], Ratchadaphiseksomphot Endowment Fund [RA MF14/61], Thailand Science research and Innovation Fund Chulalongkorn University (CU\_FRB65\_hea(49)\_056\_30\_37) and (CU\_FRB65\_hea(44)\_051\_30\_32), and Chulalongkorn Academic Advancement into Its 2nd Century Project.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-17300-1>.

**Correspondence** and requests for materials should be addressed to T.S. or T.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

© The Author(s) 2022