Article

Not peer-reviewed version

De Novo Venom Gland Transcriptome Assembly and Characterization for *Calloselasma rhodostoma* (Kuhl, 1824), the Malayan Pit Viper from Malaysia: Unravelling Toxin Gene Diversity in a Medically Important Basal Crotaline

Choo Hock Tan *, Kae Yi Tan, Tzu Shan Ng, Nget Hong Tan, Ho Phin Chong

Posted Date: 7 April 2023

doi: 10.20944/preprints202304.0125.v1

Keywords: kistomin; rhodostoxin; rhodostomin; rhodocytin; rhodocetin; ancrod

Article

De Novo Venom Gland Transcriptome Assembly and Characterization for *Calloselasma rhodostoma* (Kuhl, 1824), the Malayan Pit Viper from Malaysia: Unravelling Toxin Gene Diversity in a Medically Important Basal Crotaline

Choo Hock Tan [1,*], Kae Yi Tan [2], Tzu Shan Ng [2], Nget Hong Tan [2] and Ho Phin Chong [1]

[1] Department of Pharmacology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia; tanch@um.edu.my (CHT); schp_182@hotmail.com (HPC)

[2] Department of Molecular Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia; kytan_kae@um.edu.my (KYT); tzushang@gmail.com (TSN); tanngethong@yahoo.com.sg (NHT)

* Correspondence: tanchoohock@gmail.com; tanch@um.edu.my

Abstract: In Southeast Asia, the Malayan Pit Viper (*Calloselasma rhodostoma*) is a venomous snake species of medical importance and bioprospecting potential. To unveil the diversity of its toxin genes, this study assembled and analyzed the *de novo* venom gland transcriptome of *C. rhodostoma* from Malaysia. The expression of toxin genes dominates the gland transcriptome by 53.78% of total transcript abundance (based on overall FPKM), in which 92 non-redundant transcripts belonging to 16 toxin families were identified. Snake venom metalloproteinase (SVMP, PI>PII>PIII) is the most dominant family (37.84% of all toxin FPKM), followed by phospholipase A_2 (29.02%), bradykinin/angiotensinogen-converting enzyme inhibitor-C-type natriuretic peptide (16.30%), C-type lectin (CTL, 10.01%), snake venom serine protease (SVSP, 2.81%), L-amino acid oxidase (2.25%) and others (1.78%). The expressions of SVMP, CTL and SVSP correlate with hemorrhagic, anti-platelet and coagulopathic effects in envenoming. The SVMP metalloproteinase domains encode hemorrhagins (kistomin and rhodostoxin), while disintegrin (rhodostomin from P-II) is platelet-inhibitory. CTL gene homologues uncovered include rhodocytin (platelet aggregators) and rhodocetin (platelet inhibitor), which contribute to thrombocytopenia and platelet dysfunction. The major SVSP is a thrombin-like enzyme (ancrod homolog) responsible for defibrination in consumptive coagulopathy. The findings provide insight into the venom complexity of *C. rhodostoma* and the pathophysiology of envenoming.

Keywords: kistomin; rhodostoxin; rhodostomin; rhodocytin; rhodocetin; ancrod

Key Contribution: The study uncovers 16 toxin families in which numerous genes with novel sequences were identified for the Malayan Pit Viper, providing value information on the gene structures and functions of toxins for this unique species. Furthermore, the complete sequencing of toxins enriches snake venom databases and supports future studies in venom evolution, snakebite treatment, antivenom design and drug discovery.

1. Introduction

Venomous snakes from Elapidae and Viperidae are medically important snakes responsible for most snake envenoming cases [1]. Between the two clades, the elapids are a relatively young group (less than 40 Ma old) [2], while viperids first radiated in the Old World more than 40–50 million years ago, followed by invasion and diversification in the New World [3]. To date, viperids comprise more than 300 species grouped under the three subfamilies of Azemiopinae (Feas vipers), Viperinae (true vipers) and Crotalinae (pit vipers). The crotalines (pit vipers) are arguably the most speciose,

extensively distributed in all continents except for Antarctica, Australia, Hawaii, Madagascar, and various other isolated islands. Among the Old World crotalines, a basal clade of morphologically similar terrestrial pit vipers which consists of the genus *Calloselasma*, *Deinagkistrodon*, *Garthius*, *Hypnale* and *Tropidolaemus* was recognized [3]. The snakes of *Calloselasma*, *Deinagkistrodon* and *Hypnale* are particularly well known for causing severe snakebite envenoming in regions where they are prevalent, i.e., Southeast Asia, East Asia and South Asia, respectively. These species are designated by the World Health Organization (WHO) under Category 1 of venomous snakes, which implies the highest medical importance due to the snake's wide distribution and association with high mortality and morbidity [4].

In Southeast Asia, the genus *Calloselasma* is monotypic and represented by only one species, *Calloselasma rhodostoma*, whose type locality is Java [5]. Commonly known as the Malayan Pit Viper, this crotaline is found in the peninsula of Malaysia (northern west region), most parts of Indochina (Thailand, Laos, Cambodia and Vietnam), and in certain islands of Indonesia (Java, Karimunjawa, and Kangean) [6,7]. There have also been anecdotal reports of this species occurring in Nepal, Myanmar, middle and southern Peninsular Malaysia, northern Sumatra, and Borneo Island, especially in plantation areas. Presumably, these are non-native, invasive populations. On average, an adult snake is about 500–700 cm long, with females being slightly longer than males. The body of the Malayan Pit Viper is stout, dorsally reddish, grayish, or pale brown, flanked by two alternating series of large, dark brown, black-edged triangular blotches (see graphical abstract). The species is unique among many Asiatic viperids with its large crown scales and smooth dorsal scales. It has a reputation for being bad-tempered, quick to strike when disturbed, but also tends to stay in the same spot even hours after biting a person. This sedentary habit has earned it the local Malay name of "*ular kapak bodoh*", which is literally translated as *stupid viper*.

Clinically, envenoming caused by *C. rhodostoma* is highly toxic and fatal. *C. rhodostoma* envenoming results in local tissue damage and systemic hemotoxicity characterized by defibrination coagulopathy and thrombocytopenia [8–10]. The hemotoxic envenoming manifests as petechiae, epistaxis, hematuria, hemoptysis and consumptive coagulopathy, which may lead to complications such as intracranial hemorrhage with deadly consequences [4,11]. Earlier, a number of proteomic studies provided information on *C. rhodostoma* venom composition at varying depths [12–16]. A more thorough multi-step, quantitative proteomic study unveiled at least 96 distinct proteins belonging to 11 families in *C. rhodostoma* venom of Malayan origin [17]. The venom consists of mainly snake venom metalloproteinases (SVMP, 41.17% of total venom proteins), comprising the P-I (kistomin, 20.4%) and P-II (rhodostoxin, 19.8%) subclasses of SVMP. Other toxins identified were C-type lectins (snaclec, 26.3%), snake venom serine protease (SVSP, 14.9%), L-amino acid oxidase (7.0%), phospholipase A_2 (4.4%), cysteine-rich secretory protein (2.5%), and some minor proteins totaling 2.6%. In brief, the pathophysiology of *C. rhodostoma* envenoming is described as a constellation of hemotoxic activities caused by the multiple toxins in the venom: SVSP ancrod causes venom-induced consumptive coagulopathy, aggravated by thrombocytopenia (following platelet aggregation) caused by snaclec rhodocytin, while SVMP P-II rhodostoxin causes hemorrhage in various tissues, exacerbated by P-I kistomin and snaclec rhodocetin that inhibit platelet plug formation. Local inflammation and tissue necrosis at the bite site are likely caused by cytotoxic PLA_2 and LAAO as well as hemorrhagin SVMP in the venom [18].

The proteomic findings of *C. rhodostoma* venom will be useful information for antivenom design, and for bioprospecting of toxins in drug discovery [19]. However, a comprehensive gene profile reflecting the venom complexity of this species remains unavailable. We anticipate that a *de novo* assembly of its venom gland transcriptome will shed light on the diversity of venom genes expressed by *C. rhodostoma*. The findings will provide insights into the pathophysiology of *C. rhodostoma* envenoming, allowing practitioners to devise effective monitoring and treatment plan tailored to toxins in the venom. The uncovering of specific toxin gene sequences shall also contribute to the knowledge database of snake venom, thus enabling the comparison of toxin genes between species and facilitating the development of a pan-region antivenom, which aims to target specific toxins from various species. Hence, by applying the high throughput next-generation sequencing approach, this

study is the first to report the *de novo* transcriptome assembly, functional annotation, and expression profiling for the venom gland of *C. rhodostoma* from Malaysia. The diversity of toxin genes of this species was elucidated on the basis of its implication in snakebite envenoming and potential biomedical applications.

2. Results and Discussion

2.1. De novo transcriptome assembly, transcripts categorization and expression

De novo sequencing by Illumina HiSeq™ 2000 yielded a total of 46,636,384 clean reads, from which 146,916 contigs (N50 = 632) were assembled through Trinity (Q20 percentage = 97.17%) (Table 1). These contigs were connected resulting in 74,445 transcripts (N50=1636), validated in NCBI non-redundant (NR) protein database through BLASTx (e-value < 10^{-5}) (Supplementary File S1). The low-frequency transcripts (FPKM < 1) were filtered as non-expressing, and the remaining assemblies of 59,348 transcripts were further grouped into "unidentified", "non-toxin" and "toxin" categories accordingly. The "toxin" group made up 53.78% of the total transcript expression (FPKM), comprising 97 putative or classical toxin transcripts. The "non-toxin" and "unidentified" groups constituted 27.67% (20,962 transcripts) and 18.55% (38,289 transcripts), respectively, of the overall transcriptome (Table 1; Figure 1). The total expression level of toxin transcripts in *C. rhodostoma* venom gland (~54%) was high and comparable to pit vipers reported previously, for instance, *Bothrops atrox* (~59%) [20], *Deinagkistrodon acutus* (~40%) [21], and *Gloydius intermedius* (~61%) [22]. The high expressions of toxin transcripts were accompanied by a small subset of genes (97 out of 59,348), giving rise to a high redundancy of toxin genes (4549.90 FPKM/transcript). In contrast, the groups of "non-toxin" (10.39 FPKM/transcript) and "unidentified" (3.81 FPKM/transcript) transcripts have remarkably much lower redundancy in their gene expression. The same has been observed in a number of snake venom gland transcriptomics, which reported thousand-folds of higher expression of toxins genes in comparison to the non-toxin and unidentified genes [23–27]. The high expression of toxin genes is in line with the venom gland's function in synthesizing secretory proteins into the venom, while the non-toxin genes are mainly house-keeping genes for basic cellular functions.

Table 1. Overview of the NGS output statistics of *Calloselasma rhodostoma* venom gland transcriptomic analysis.

Parameter	Value
Total raw reads	47,365,132
Total clean reads	46,636,384
Total clean nucleotides (nt)	4,663,638,400
Q20 percentage	97.17%
N percentage	0.00%
GC percentage	44.52%
Contigs created	146,916
Total length (nt)	47,039,053
Mean length (nt)	320
N50	632
Unigenes/transcripts assembled	74,445
Total length (nt)	52,612,410
Mean length (nt)	707
N50	1636
Unigene/transcripts assembled (FPKM>1)	59,348
Unidentified	38,289
Non-toxin	20,962
Toxin	97

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2023 doi:10.20944/preprints202304.0125.v1

2.2. Overview of toxin gene expression in C. rhodostoma venom gland transcriptome

The 97 transcripts were subsequently classified into 16 protein families of snake venom toxins. After identifying redundant sequences, the number of unique toxin transcripts was further reduced to 92 (Table 2). Despite the very high overall expression of toxin genes in the transcriptome, the number of these genes are relatively small, and they are distributed with repetitions in a restricted set of protein family. In contrast to the generally perceived thought that snake venom contains myriad proteins, the transcriptomic profile and previous proteomic findings appear to be consistent with the dominant view of toxin recruitment through gene duplication and neofunctionalization, with consequent streamlining of venom functional complexity. Of these, 30 transcripts are identified as full-length protein-encoding transcripts (defined here as >90% coverage to the protein-coding region of the annotated protein sequences) (Table 3). Based on the FPKM as a quantitative parameter to gauge gene expression, the snake venom metalloproteinase (SVMP) was found to be the dominating toxin family in *C. rhodostoma* venom gland transcriptome, charting 37.84% of all toxin transcriptions. This is followed by phospholipase A_2 (PLA_2, 29.02%), bradykinin-potentiating peptide/angiotensin-converting enzyme inhibitor-C-type natriuretic peptide (BPP/ACEI-CNP), 16.30%) and snake venom C-type lectin (snaclec, 10.01%). Toxins from these four families, which constitute 93.16% of overall toxin FPKM, are most likely the key toxin components of *C. rhodostoma* venom (Figure 1), and responsible for most of the toxic activities in snakebite envenoming. The rest of toxin families identified from the venom gland transcriptome of *C. rhodostoma* are snake venom serine proteinase (SVSP, 2.81%), L-amino acid oxidase (LAAO, 2.25%), cysteine-rich secretory venom protein (CRiSP, 0.9%), 5'nucleotidase (5'NUC, 0.28%), phospholipase B (PLB, 0.25%), nucleobindin (NLB, 0.19%), nerve growth factor (NGF, 0.07%), vascular endothelial growth factor (VEGF, 0.05%), three-finger toxin (3FTX, 0.02%), aminopeptidase (APP, 0.01%), phosphodiesterase (PDE, 0.01%), and Kunitz-type serine proteinase inhibitor (KSPI, < 0.01%) (Figure 1).

Table 2. Overview of the families and subtypes of toxin genes in the venom gland transcriptome of *Calloselasma rhodostoma.*

Protein family/Protein subtype	Accession no. (Species)	Relative abundance % (subtype)
Snake venom metalloproteinase (SVMP)		**37.84 (22)**
P-I SVMP		**26.51 (4)**
Snake venom metalloproteinase kistomin	P0CB14 (*Calloselasma rhodostoma*)	26.29
Snake venom metalloproteinase BpirMP	P0DL29 (*Bothrops pirajai*)	0.17
Zinc metalloproteinase/disintegrin ussurin	Q7SZD9 (*Gloydius ussuriensis*)	0.03
Group I snake venom metalloproteinase	Q2UXQ3 (*Echis ocellatus*)	0.02
P-II SVMP		**7.94 (4)**
Zinc metalloproteinase/disintegrin	P30403 (*Calloselasma rhodostoma*)	7.89
Zinc metalloproteinase/disintegrin ussurin	Q7SZD9 (*Gloydius ussuriensis*)	0.04
Metalloprotease PIIa	V5IWE4 (*Trimeresurus gracilis*)	0.01
Zinc metalloproteinase-disintegrin VMP-II	J9Z332 (*Crotalus adamanteus*)	< 0.01
P-III SVMP		**3.39 (14)**
Zinc metalloproteinase-disintegrin-like halysase	Q8AWI5 (*Gloydius halys*)	3.32
Metalloprotease P-III	A0A077L6L9 (*Protobothrops elegans*)	0.03
Zinc metalloproteinase-disintegrin-like NaMP	A8QL59 (*Naja atra*)	0.01
Metalloprotease P-III	A0A077L6L9 (*Protobothrops elegans*)	0.01
Flavorase	G1UJB2 (*Protobothrops flavoviridis*)	0.01
Zinc metalloproteinase-disintegrin-like NaMP	A8QL59 (*Naja atra*)	0.01
Metalloprotease P-III 5	A0A077L7M5 (*Protobothrops flavoviridis*)	< 0.01
Metalloproteinase	A0A2Z4N9U9 (*Boiga irregularis*)	< 0.01
Zinc metalloproteinase-disintegrin-like NaMP	A8QL59 (*Naja atra*)	< 0.01

Metalloprotease P-III 5	A0A077L7M5 (*Protobothrops flavoviridis*)	< 0.01
Metalloproteinase	A0A2Z4N9U9 (*Boiga irregularis*)	< 0.01
Zinc metalloproteinase-disintegrin-like NaMP	A8QL59 (*Naja atra*)	< 0.01
Metalloprotease P-III 5	A0A077L7M5 (*Protobothrops flavoviridis*)	< 0.01
Zinc metalloproteinase-disintegrin-like NaMP	A8QL59 (*Naja atra*)	< 0.01
Phospholipase A_2 (PLA_2)		**29.02 (15)**
Phospholipase A2	A0A0H3U266 (*Calloselasma rhodostoma*)	16.47
Acidic phospholipase A2 S1E6-c	Q9PVE9 (*Calloselasma rhodostoma*)	5.19
K49 phospholipase A2	A8Y7N3 (*Deinagkistrodon acutus*)	3.62
Acidic phospholipase A2 Ts-A4	Q6H3C7 (*Trimeresurus stejnegeri*)	2.74
Phospholipase A2 homolog	P0DMT1 (*Echis pyramidum leakeyi*)	0.91
Phospholipase A2	A0A0H3U279 (*Ovophis makazayazaya*)	0.06
Phospholipase A2 group IIE	A0A2H4N3A5 (*Bothrops moojeni*)	0.02
Phospholipase A2, group IIE	A0A1J0R065 (*Crotalus atrox*)	0.01
Group 3 secretory phospholipase A2	A0A223PK36 (*Daboia russelii*)	< 0.01
Basic phospholipase A2 beta-bungarotoxin A4 chain	Q75S51 (*Bungarus candidus*)	< 0.01
Phospholipase A2 isoform 2	H8PG83 (*Parasuta nigriceps*)	< 0.01
Group 15 secretory phospholipase A2	A0A223PK35 (*Daboia russelii*)	< 0.01
Acidic phospholipase A2 homolog	P29601 (*Bungarus fasciatus*)	< 0.01
Acidic phospholipase A2	P00606 (*Bungarus multicinctus*)	< 0.01
Group 3 secretory phospholipase A2	A0A223PK36 (*Daboia russelii*)	< 0.01
Bradykinin-potentiating/Angiotensin-converting enzyme inhibitor/C-type natriuretic peptide (BPP/ACEI-CNP)		**16.30 (3)**
Angiotensin converting enzyme inhibitor and C-type natriuretic peptide	M5A7D0 (*Calloselasma rhodostoma*)	5.77
Angiotensin converting enzyme inhibitor and C-type natriuretic peptide	M5A7D0 (*Calloselasma rhodostoma*)	5.51
Angiotensin converting enzyme inhibitor and C-type natriuretic peptide	M5A7D0 (*Calloselasma rhodostoma*)	5.02
Snake C-type lectin (CTL)		**10.01 (7)**
Snaclec rhodocytin subunit beta	Q9I840 (*Calloselasma rhodostoma*)	4.33
C-type lectin	G8FML6 (*Calloselasma rhodostoma*)	3.26
Snaclec rhodocytin subunit alpha	Q9I841 (*Calloselasma rhodostoma*)	1.76
Snaclec rhodocetin subunit delta	D2YW40 (*Calloselasma rhodostoma*)	0.37
C-type lectin beta subunit	T2HPS7 (*Protobothrops flavoviridis*)	0.27
Lectoxin-Enh9	A7XQ58 (*Pseudoferania polylepis*)	0.01
C-type lectin 3	A0A346CLX6 (*Ahaetulla prasina*)	0.01
Snake venom serine proteinase (SVSP)		**2.81 (14)**
Thrombin-like enzyme ancrod	P26324 (*Calloselasma rhodostoma*)	1.92
Snake venom serine protease 3	O13058 (*Protobothrops flavoviridis*)	0.20
Snake venom serine protease ussurin	Q8UUJ2 (*Gloydius ussuriensis*)	0.19
Snake venom serine protease gussurobin	Q8UVX1 (*Gloydius ussuriensis*)	0.14
Venom thrombin-like enzyme	Q90Z47 (*Deinagkistrodon acutus*)	0.12
Thrombin-like enzyme	Q98TT5 (*Deinagkistrodon acutus*)	0.08
Thrombin-like enzyme stejnobin	Q8AY81 (*Trimeresurus stejnegeri*)	0.08
Snake venom serine protease 3	O13063 (*Trimeresurus gramineus*)	0.02
Venom plasminogen activator GPV-PA	P0DJF5 (*Trimeresurus albolabris*)	0.02
Thrombin-like enzyme ancrod-2	P47797 (*Calloselasma rhodostoma*)	0.02
Serine protease 3	A0A286S0D3 (*Gloydius intermedius*)	0.01
Thrombin-like enzyme kangshuanmei	P85109 (*Gloydius brevicaudus*)	0.01
Serine proteinase isoform 7	B0VXT9 (*Sistrurus catenatus edwardsii*)	< 0.01
Thrombin-like protein DAV-WY	B3V4Z6 (*Deinagkistrodon acutus*)	< 0.01

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2023 doi:10.20944/preprints202304.0125.v1

L-amino acid oxidase (LAAO)		**2.25 (1)**
L-amino-acid oxidase	P81382 (*Calloselasma rhodostoma*)	2.25
Cysteine-rich secretory protein (CRiSP)		**0.90 (2)**
Cysteine-rich secretory protein LCCL domain-containing 2	V8NV17 (*Ophiophagus hannah*)	0.90
Cysteine-rich seceretory protein Bc-CRPa	F2Q6G3 (*Bungarus candidus*)	< 0.01
5'Nucleotidase (5'NUC)		**0.28 (5)**
Snake venom 5'-nucleotidase	F8S0Z7 (*Crotalus adamanteus*)	0.27
5' nucleotidase	A6MFL8 (*Demansia vestigiata*)	< 0.01
5' nucleotidase	A6MFL8 (*Demansia vestigiata*)	< 0.01
5' nucleotidase	A6MFL8 (*Demansia vestigiata*)	< 0.01
5' nucleotidase 1	A0A346CLX4 (*Borikenophis portoricensis*)	< 0.01
Phospholipase B (PLB)		**0.25 (4)**
Phospholipase B-like	A0A2H4N395 (*Bothrops moojeni*)	0.25
Phospholipase B1, membrane-associated	V8NN21 (*Ophiophagus hannah*)	< 0.01
Phospholipase B-like	V8NLQ9 (*Ophiophagus hannah*)	< 0.01
Phospholipase B-like	V8NLQ9 (*Ophiophagus hannah*)	< 0.01
Nucleobindin (NLB)		**0.19 (1)**
Nucleobindin-1	V8P8E3 (*Ophiophagus hannah*)	0.19
Nerve growth factor		**0.07 (1)**
Nerve growth factor	B1Q3K2 (*Protobothrops flavoviridis*)	0.07
Snake venom vascular endothelial growth factor (VEGF)		**0.05 (1)**
Snake venom vascular endothelial growth factor toxin	P67862 (*Protobothrops flavoviridis*)	0.05
Three-finger toxin (3FTX)		**0.02 (9)**
Alpha-bungarotoxin isoform A31	P60615 (*Bungarus multicinctus*)	0.01
Neurotoxin-like protein pMD18-NTL1/2/4/5	Q7ZT13 (*Bungarus multicinctus*)	< 0.01
Muscarinic toxin BM14	Q8JFX7 (*Bungarus multicinctus*)	< 0.01
Kappa-3-bungarotoxin	P15817 (*Bungarus multicinctus*)	< 0.01
Gamma-bungarotoxin	Q9YGJ0 (*Bungarus multicinctus*)	< 0.01
Three finger toxin 1	A5X2W6 (*Sistrurus catenatus edwardsii*)	< 0.01
Short neurotoxin homolog NTL4	Q9YGI8 (*Bungarus multicinctus*)	< 0.01
Three finger toxin 2	A5X2W7 (*Sistrurus catenatus edwardsii*)	< 0.01
Putative three finger toxin	F5CPD4 (*Micrurus altirostris*)	< 0.01
Aminopeptidase A		**0.01 (1)**
Aminopeptidase	T2HQN1 (*Ovophis okinavensis*)	0.01
Phosphodiesterase (PDE)		**0.01 (5)**
Venom phosphodiesterase 1	J3SEZ3 (*Crotalus adamanteus*)	< 0.01
Venom phosphodiesterase 1	J3SEZ3 (*Crotalus adamanteus*)	< 0.01
Venom phosphodiesterase 2	J3SBP3 (*Crotalus adamanteus*)	< 0.01
Venom phosphodiesterase 1	J3SEZ3 (*Crotalus adamanteus*)	< 0.01
Venom phosphodiesterase 2	J3SBP3 (*Crotalus adamanteus*)	< 0.01
Kunitz-type serine proteinase inhibitor (KSPI)		**< 0.01 (1)**
Kunitz-type serine protease inhibitor homolog beta-bungarotoxin B2a chain	Q8AY45 (*Bungarus candidus*)	< 0.01

Table 3. Full-length toxin transcripts derived from the venom gland transcriptome of *Calloselasma rhodostoma.*

	Protein family/Protein ID	Annotated accession	Species	Amino acid chain	Mature chain of accession ID	Coverage (mature chain)	Coverage percentage (%)
Snake venom metalloproteinase (SVMP)							
Cr-SVMP01	Snake venom metalloproteinase kistomin	P0CB14	*Calloselasma rhodostoma*	417	417	1-417	100
Cr-SVMP05	Zinc metalloproteinase/disintegrin	P30403	*Calloselasma rhodostoma*	478	478	1-478	100
Phospholipase A_2 (PLA_2)							
Cr-PLA04	Acidic phospholipase A2 Ts-A4	Q6H3C7	*Trimeresurus stejnegeri*	139	139	1-139	100
Cr-PLA10	Basic phospholipase A2 beta-bungarotoxin A4 chain	Q75S51	*Bungarus candidus*	147	147	1-147	100
Cr-PLA11	Phospholipase A2 isoform 2	H8PG83	*Parasuta nigriceps*	136	136	1-136	100
Cr-PLA12	Group 15 secretory phospholipase A2	A0A223PK35	*Daboia russelii*	362	393	1-341	92
Snake C-type lectin (Snaclec)							
Cr-CTL01	Snaclec rhodocytin subunit beta	Q9I840	*Calloselasma rhodostoma*	146	146	1-146	100
Cr-CTL02	C-type lectin	G8FML6	*Agkistrodon piscivorus leucostoma*	157	158	1-158	99
Cr-CTL03	Snaclec rhodocytin subunit alpha	Q9I841	*Calloselasma rhodostoma*	136	136	1-136	100
Cr-CTL04	Snaclec rhodocetin subunit beta	P81398	*Calloselasma rhodostoma*	129	129	1-129	100
Cr-CTL05	Snaclec rhodocetin subunit delta	D2YW40	*Calloselasma rhodostoma*	150	150	1-150	100
Snake venom serine proteinase (SVSP)							
Cr-SSP01	Thrombin-like enzyme ancrod	P26324	*Calloselasma rhodostoma*	234	234	1-234	100
Cr-SSP03	Snake venom serine protease ussurin	Q8UUJ2	*Gloydius ussuriensis*	224	236	13-236	95

doi:10.20944/preprints202304.0125.v1

		L-amino acid oxidase (LAAO)					
Cr-LAO01	L-amino-acid oxidase	P81382	*Calloselasma rhodostoma*	516	516	1-516	100
		Cysteine-rich venom protein (CRiSP)					
Cr-CRP01	Cysteine-rich secretory protein LCCL domain-containing 2	V8NV17	*Ophiphagus Hannah*	495	472	1-472	100
		5′Nucleotidase (5′NUC)					
Cr-NUC01	Snake venom 5'-nucleotidase	F8S0Z7	*Crotalus adamanteus*	588	588	1-588	100
		Phospholipase B (PLB)					
Cr-PLB01	Phospholipase B-like	A0A2H4N395	*Bothrops moojeni*	553	558	1-553	99
Cr-PLB03	Phospholipase B-like	V8NLQ9	*Ophiophagus Hannah*	321	300	87-299	100
		Nucleobindin (NLB)					
Cr-NLB01	Nucleobindin-1	V8P8E3	*Ophiophagus Hannah*	452	397	22-372	100
		Nerve growth factor (NGF)					
Cr-NGF01	Nerve growth factor	B1Q3K2	*Protobothrops flavoviridis*	237	241	1-237	98
		Snake venom vascular endothelial growth factor (VEGF)					
Cr-VGF01	Snake venom vascular endothelial growth factor toxin	P67862	*Protobothrops flavoviridis*	145	146	1-145	99
		Three finger toxin (3FTX)					
Cr-FTX01	Alpha-bungarotoxin isoform A31	P60615	*Bungarus multicinctus*	95	95	1-95	100
Cr-FTX02	Neurotoxin-like protein pMD18-NTL1/2/4/5	Q7ZT13	*Bungarus multicinctus*	86	86	1-86	100
Cr-FTX03	Muscarinic toxin BM14	Q8JFX7	*Bungarus multicinctus*	97	103	7-103	94
		Aminopeptidase A					
Cr-APP01	Aminopeptidase	T2HQN1	*Ovophis okinavensis*	953	953	1-953	100
		Phosphodiesterase (PDE)					
Cr-PDE02	Venom phosphodiesterase 1	J3SEZ3	*Crotalus*	844	851	6-849	99

doi:10.20944/preprints202304.0125.v1

Cr-PDE03	Venom phosphodiesterase 2	J3SBP3	*Crotalus adamanteus*	808	810	1-808	99
Cr-PDE04	Venom phosphodiesterase 1	J3SEZ3	*Crotalus adamanteus*	849	851	1-849	99
Cr-PDE05	Venom phosphodiesterase 2	J3SBP3	*Crotalus adamanteus*	803	810	6-808	99
Kunitz-type serine proteinase inhibitor (KSPI)							
Cr-KUN01	Kunitz-type serine protease inhibitor homolog beta-bungarotoxin B2a chain	Q8AY45	*Bungarus candidus*	84	85	2-85	99

doi:10.20944/preprints202304.0125.v1

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2023 doi:10.20944/preprints202304.0125.v1

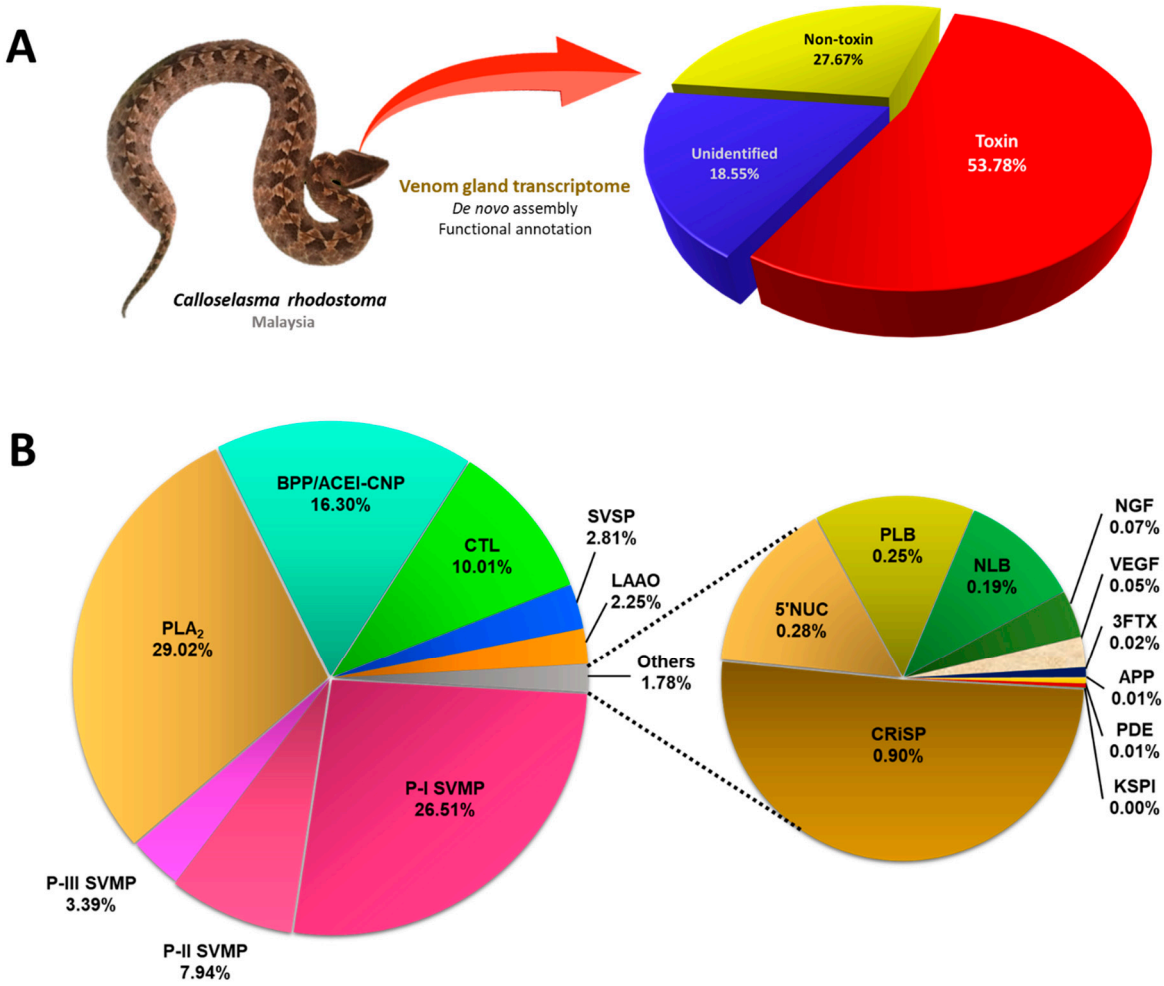

Figure 1. The transcriptomic profile of *Calloselasma rhodostoma* venom gland. (A) Classification of transcripts into toxins, non-toxins and unidentified genes derived from *de novo* transcriptome assembly of the venom gland of Malayan Pit Viper (inset). (B) Profiling of toxin transcripts into 16 families of genes coding for various venom proteins. Percentages indicate the relative abundance of transcripts based on FPKM. Abbreviations: 3FTXs, three-finger toxin; KSPI, Kunitz-type serine proteinase inhibitor; PLA_2, phospholipase A_2; SVMP, snake venom metalloproteinase; BPP/ACEI-CNP, bradykinin-potentiating peptide/angiotensin-converting enzyme inhibitor-C-type natriuretic peptide; NGF, nerve growth factor; CRiSP, cysteine-rich secretory protein; PLB, phospholipase-B; VEGF, vascular endothelial growth factor; CTL, snake C-type lectin; 5'NUC, 5'nucleotidase; APP, aminopeptidase A; PDE, phosphodiesterase; NLB, nucleobindin; LAAO, L-amino acid oxidase; SVSP, snake venom serine proteinase.

The dominance of SVMP transcription (37.8% of all toxin FPKM) revealed in *C. rhodostoma* venom gland transcriptome is in agreement with the high abundance of SVMP (41.2% of total venom proteins) found in the proteome of Malaysian *C. rhodostoma* venom [13]. Nonetheless, the expression levels of other toxin transcripts appear to be at variance with the protein abundances of toxins found in its venom proteome. For example, while the proteomic study found relatively high abundances of snaclec (26.3%), snake venom serine proteases (SVSP, 14.9%), and L-amino acid oxidase (LAAO, 7%) in the venom, the FPKM, which indicate gene expression levels of the corresponding toxins, was only 10.1%, 2.6% and 2.25%, respectively. Statistical analysis with the Pearson correlation test found no significant correlation between gene expression (current transcriptomic study) and protein abundance (proteomic study, [13]) (correlation coefficient $r = 0.711$, p-value > 0.05). The discrepancy observed is not unusual as similar findings have been reported in a number of studies [24,25,28,29]. The causes of the no-correlation observed could be multifactorial and occur at different stages of protein synthesis. The process could be modulated by complex regulation of protein translation at

 doi:10.20944/preprints202304.0125.v1

different time points, varying synthesis rate and half-lives that varied among the transcripts of different toxins [28,29]. The transcriptomic profile, therefore, is a snap-shot picture of toxin gene expression at day 4 post-venom milking, while the proteome represented cumulation of various toxin proteins over an unspecified duration of venom production and storage in the glands. The dominance of SVMP in both data sets, nevertheless, suggests these toxins are actively and consistently expressed for important biological functions (discussed below).

2.3. Toxin gene diversity and implication on bioactivity of snake venom

The transcriptomic profile reveals substantial complexity of toxin genes expressed by *C. rhodostoma*. The diversity of the toxin genes, though complex, appears to conform to a small subset of predictable gene families that are characteristic of Viperidae snakes. The following section analyzed the sequence data of the toxins with correlation to bioactivities by respective gene families.

2.3.1. Snake venom metalloproteinase (SVMP)

While snake venom toxins are constrained to a restricted number of protein families, each protein family typically contains multiple proteoforms (isoforms), and exhibits a breadth of structural diversity across taxonomically distinct and geographically distant snake species [28]. Snake venom metalloproteinase (SVMP) is one such example, which is arguably the most structurally and functionally diverse snake toxin superfamily. Based on the domain structure and molecular weight, SVMPs are conventionally classified into three main subgroups or classes, i.e., P-I which are relatively small (20-30 kDa) consisting of only the metalloproteinase domain, P-II (30-60 kDa) characterized by the presence of disintegrin domain in addition to metalloproteinase, and P-III (>60 kDa) with the presence of a cysteine-rich domain, in addition to disintegrin-like and metalloproteinase domain [30,31]. In the present study, the SVMP family represents 34.87% of all toxin FPKM in the Malaysian *C. rhodostoma* venom gland transcriptome, where the subgroup P-I SVMP is the most abundant (26.51%), expressed at a higher level than the P-II SVMP (7.94%) and P-III SVMP (3.39%) (Figure 1). The dominance of P-I SVMP in venom gland transcriptome is consistent with the proteomic finding of *C. rhodostoma* venom (Malaysian origin) where the protein abundance of P-I SVMP was found to be 20.4% of total venom proteins. The P-II SVMP and P-III SVMP protein abundances, on the other hand, were reported to be 19.8% and 1%, respectively, in the venom proteome [13].

Among the P-I SVMP transcripts, Cr-SMP01, which is a homolog to kistomin (UniProt ID: P0CB14), is the most highly expressed toxin in the *C. rhodostoma* venom gland (Table 3). Similarly, this protein was found to be the most abundant SVMP form in the *C. rhodostoma* venom proteome [13]. Huang et al. first isolated and characterized kistomin, an anti-platelet protease composed of 227 amino acid residues and selectively cleaves human platelet glycoprotein GPIb [32]. Kistomin was further shown to affect glycoprotein Ib-von Willebrand factor interaction causing anti-thrombotic effect [33], and hydrolyze glycoprotein VI (GPVI, the transmembrane receptor on platelet), thereby inhibiting the interaction between platelet and collagen [33]. Clinically, *C. rhodostoma* envenoming invariably results in hemorrhagic effects [8–10], consistent with the high expression of this hemorrhagin protein. Our multiple sequence analysis shows Cr-SMP01 gene is nearly identical to kistomin, and it comprises the three domains of a signal peptide, a propeptide and the metalloproteinase (Figure 2). Its signal peptide and propeptide domains are also highly similar to two well characterized PI SVMP from *Deinagkistrodon acutus* (acutolysin, UniProt KB: Q9PW36) and *Crotalus atrox* (atrolysin, UniProt KB: Q90392), demonstrating a high degree of conservation of these non-protein coding regions. The metalloproteinase domain, on the contrary, is variable with multiple substitutions accumulated in these genes (Figure 2). These are coding genes for SVMP that are subject to the pressure of positive selection and the consequent accelerated evolution for adaption to various ecological niches occupied by the different species [34].

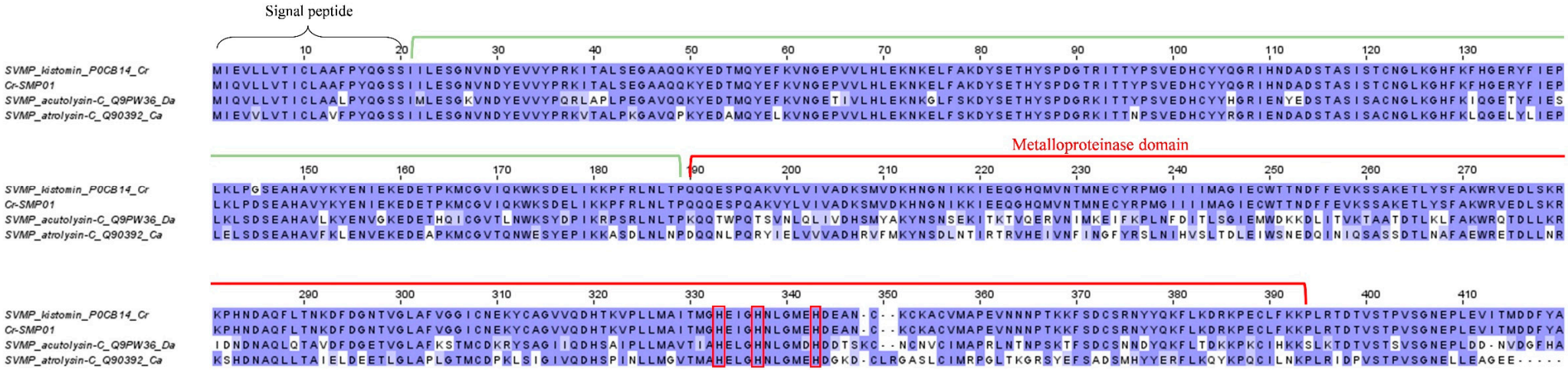

Figure 2. Multiple sequence alignment of Cr-SMP01 (putative snake venom metalloproteinase (SVMP) kistomin) identified in *Calloselasma rhodostoma* venom gland transcriptome in comparison to selected P-I SVMP retrieved from UniProt KB database. Black bracket: signal peptide domain; Green bracket: propeptide domain; Red bracket: metalloproteinase domain. Highlighted in red: metal binding site. Abbreviation: Cr – *Calloselasma rhodostoma*; Da – *Deinagkistrodon acutus*; Ca - *Crotalus atrox*.

doi:10.20944/preprints202304.0125.v1

 doi:10.20944/preprints202304.0125.v1

The major P-II SVMP gene, annotated as Cr-SMP05 in the *C. rhodostoma* venom gland has a high sequence identity to the protein zinc metalloproteinase/disintegrin (: P30403) (Table 3), which is more commonly just referred to as rhodostoxin or rhodostomin in vast publications [35–38]. The respective toxins, nonetheless, are distinct and each represents an individual cleavage product of the common P-II SVMP precursor (Figure 3). Au et al. [35] cloned the platelet aggregation inhibitor rhodostomin and showed its 68-amino acid sequence is in fact located at the carboxyl terminus of a much larger precursor polypeptide, which was later verified as part of a true hemorrhagin protein containing a metalloproteinase domain as in the SVMP of PI class [36,39]. As with the rhodotomin's precursor, the transcript Cr-SMP05 possesses the characteristic peroxisomal targeting sequence (Ser-His-Ala) at the C-terminus (Figure 3). Also, the signature motif of disintegrin, i.e., the RGD (Arg-Gly-Asp) motif is found (highlighted in green) in Cr-SMP05, rhodostomin (UniProt KB: P30403) and aculysin (UniProt KB: Q9WJ0, derived from another basal crotaline of the Old World, *D. acutus*), but intriguingly, not in atrolysin-E (UniProt KB: P34182, derived from the New World crotaline, *Crotalus atrox*). The RGD motif is deemed crucial for specific binding between disintegrin and integrin receptors [40], although a number of disintegrins have been found to have variable sequences not necessarily conforming to the RGD motif [37]. On the other hand, the metalloproteinase-containing domain of Cr-SMP05 resembles the sequence of rhodostoxin, the hemorrhagin derived from the same precursor gene of rhodostomin (Figure 3). Rhodostoxin is the first four-disulfide proteinase reported among all known venom metalloproteinases [36]. Two N-glycosylation sites have been identified (highlighted in orange) in the metalloproteinase domain of Cr-SMP05, consistent with that reported in rhodostoxin [36] (Figure 3).

Rhodostoxin is a major hemorrhagin in *C. rhodostoma* venom that exhibits strong proteolytic and potent hemorrhagic activity (minimum hemorrhage dose, MHD = 0.13 μg intradermally), although with low lethality (non-lethal even at an intravenous dose > 6 μg/g in rodents) [41]. Its target site of action and mechanism of action have not been investigated in depth, although based on its metalloproteinase domain one can deduce that it acts in a manner similar to other hemorrhagins by damaging the collagenous basement membrane and destabilizing the vascular integrity, leading to blood extravasation and hemorrhagic effect [42,43]. Rhodostomin, on the other hand, received more research attention over the years, presumably due to its pharmaceutical potential as an anti-platelet and anti-proliferative agent. Rhodostomin blocks binding of fibrinogen to fibrinogen receptor, the glycoprotein IIb/IIIa complex of an activated platelet [35]. It also blocks basic fibroblast growth factor (bFGF) and integrin $\alpha_v\beta_3$ receptor, the blockade of which extends its anti-angiogenesis property for tumor growth suppression [38].

Consistent with earlier cloning studies, the current transcriptomic finding in *C. rhodostoma* venom gland supports the view that hemorrhagin metalloproteinases (e.g., rhodostoxin) and anti-platelet disintegrins (e.g., rhodostomin) share common precursors [35,44]. Disintegrins including rhodostomin are typically isolated as individual low molecular weight proteins (MW 6–7 kDa) from the crude venoms [13,45,46], while stand-alone transcripts have not been identified in venom gland transcriptomes, implying they are generated by proteolysis of translated P-II SVMPs [47]. The yields of metalloproteinase and disintegrin cleaved thereof, obviously, do not exist at a balanced ratio of 1:1 in any proteomic study of snake venom [13,45,46]. The natural occurrence and abundance of disintegrins in any viperid venom are usually low relative to SVMP, and are virtually non-existent in elapid venoms [48]. The mechanism which determines the rate and extent of disintegrin release from the cleavage of P-II SVMP in a venom gland remains unresolved. Nonetheless, the derivation of the hemorrhagin and anti-platelet toxin from the same precursor gene, and their co-existence in an intact large protein suggest an important synergistic function in hemotoxic envenoming caused by the viperids. From the pathological point of view, the action of metalloproteinase (destruction of the basal membrane and vascular stability) coupled with that of disintegrin (anti-platelet plug formation) at the venom injection site would contribute to enhanced bleeding. In real envenoming, the local hemorrhagic activity would be exacerbated by systemic coagulopathy (induced by procoagulant and anticoagulant toxins) [9], thrombocytopenia and platelet dysfunction (caused by snaclecs) [49,50].

 doi:10.20944/preprints202304.0125.v1

Hypothetically, the stand-alone disintegrins could be absorbed from the bite wound, and thus contribute to the systemic effect on platelet.

The SVMP of P-III class have a higher molecular weight (typically > 60 kD), and are most intriguing in terms of structural complexity and function variety [51]. In addition to the typical "SVMP+disintegrin-like+cysteine rich" composition, some P-III SVMPs are found with an additional lectin-like domain that gives rise to unique substrate specificity, and hence, functionality. The present study uncovered a novel SVMP transcript, Cr-SMP09 from the venom gland transcriptome of *C. rhodostoma*, which is probably the first identified P-III SVMP gene in this species. The sequence of Cr-SMP05 shows the highest identity matched to halysase (UniProt KB: Q8AW15), an SVMP-disintegrin-like protein from *Gloydius halys* (Figure 4). The identification of this transcript as the main P-III SVMP of *C. rhodostoma* verifies previous proteomic findings in which peptide fragments were matched to the SVMP-disintegrin-like halysase instead of one belonging to Malayan Pit Viper [12,13]. The transcript identified, unfortunately, lack complete information for the signal peptide and propeptide, but it retains the full sequence of metalloproteinase, disintegrin-like and cysteine-rich domains (Figure 4). Comparing homologous P-III SVMP sequences among Cr-SMP09, halysase, acurhagin (UniProt KB: Q9W6M5, from *D. acutus*), and VAP1 (UniProt KB: Q9DGB9, from *C. atrox*), structural variation is high in acurhagin where multiple substitutions were found throughout the functional coding domains, suggesting its higher evolutionary events and possibly a more adaptive function for *D. acutus* (Figure 4). In *C. rhodostoma*, the expression of Cr-SMP09 is low (3.32% of all toxin FPKM), consistent with the low protein abundance of P-III SVMP in our proteomic study (~1% of total venom proteins). The low protein abundance has been a limitation to purification and characterization study of this novel protein.

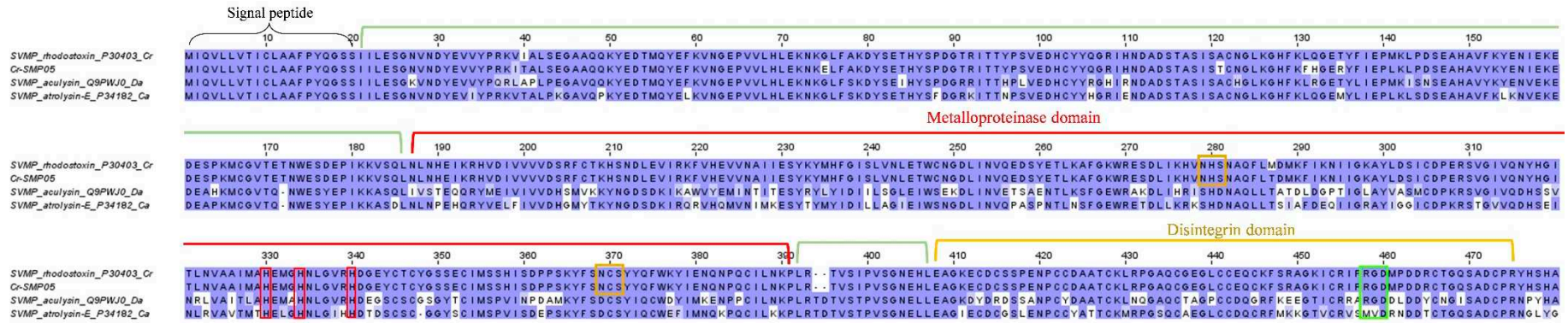

Figure 3. Multiple sequence alignment of Cr-SMP05 (putative snake venom metalloproteinase (SVMP) rhodostoxin) identified in *Calloselasma rhodostoma* venom gland transcriptome in comparison to selected P-II SVMP retrieved from UniProtKB database. Black bracket: signal peptide domain; Green bracket: propeptide domain; Red bracket: metalloproteinase domain; Orange bracket: disintegrin domain. Highlighted in red: metal binding site; highlighted in orange: N-glycosylation sites; highlighted in green: signature disintegrin motif. Abbreviation: Cr – *Calloselasma rhodostoma*; Da – *Deinagkistrodon acutus*; Ca – *Crotalus atrox*.

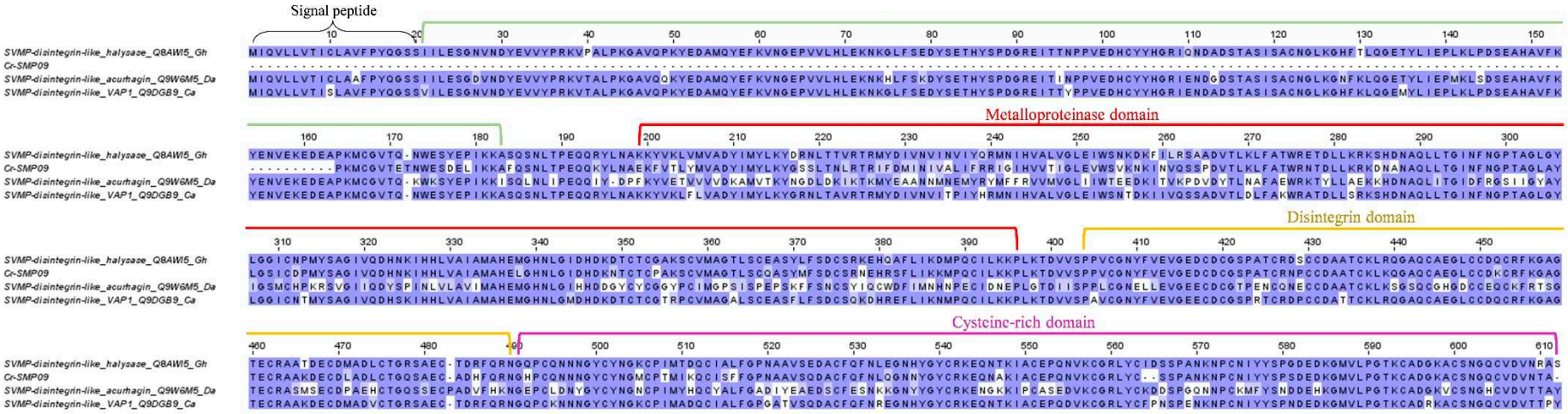

Figure 4. Multiple sequence alignment of Cr-SMP09 (putative P-III snake venom metalloproteinase (P-III SVMP)) identified in *Calloselasma rhodostoma* venom gland transcriptome in comparison to selected P-III SVMP retrieved from UniProtKB database. Black bracket: signal peptide domain; Blue bracket: propeptide domain; Red bracket: metalloproteinase domain; Green bracket: disintegrin domain; Purple bracket: cysteine-rich domain. Abbreviation: Gh – *Gloydius halys*; Da – *Deinagkistrodon acutus*; Ca – *Crotalus atrox*.

doi:10.20944/preprints202304.0125.v1

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2023 doi:10.20944/preprints202304.0125.v1

2.3.2. Phospholipase A_2 (PLA_2)

Snake venom PLA_2s ($svPLA_2$s) exhibit a wide array of pharmacological activities, including myotoxicity, nephrotoxicity, anti-coagulant effect and inflammation, owing to their structural versatility [52,53]. Used to be thought a ubiquitous group of enzymatic toxins in all snake venoms, $svPLA_2$ abundances are now known to vary remarkably from a high of >60% in Russell's Viper venoms [54,55] to virtually none in *Naja nivea*, *Naja annulifera* and *Naja senegalensis* (subgenus: *Uraeus*) [56–58]. Previous studies showed *C. rhodostoma* venom has low enzymatic activity [41,59], consistent with its proteome in which PLA_2 constitutes only 4.4% of the total venom proteins [13]. The apparently higher abundance of PLA_2 transcripts in the venom gland transcriptome is puzzling, suggesting complex regulation of the mRNA half-life and protein translation which cannot be simplified with a single snapshot of the transcriptomic profile. In fact, the disproportionate expression between PLA_2 transcripts and their proteins is not uncommon as reported in a number of venom gland transcriptomics for *Naja kaouthia* [23,60], *Naja sumatrana* [25,61], *Ophiophagus hannah* [24], *Hydrophis curtus* [62] and *Calliophis bivirgata* [26], when venom glands were harvested 3–4 days post-venom milking.

The present study identifies a number of acidic $svPLA_2$ transcripts in the *C. rhodostoma* venom gland transcriptome, notably Cr-PLA01, Cr-PLA02, Cr-PLA03 and Cr-PLA04 which form the main bulk of PLA_2 transcripts (Table 2). The mRNAs uncover the PLA_2 genes with higher expression (FPKM >10,000), and structurally, they belong to Group IIA PLA_2 as most viperid PLA_2 do (Figure 5). Cr-PLA01 and Cr-PLA02 are matched with the highest homology to previously deposited *C. rhodostoma* PLA_2 sequences of A0A0H3U266 (uncertain origin) and Q9PVE9 (Thailand, [63]), respectively. Cr-PLA03 and Cr-PLA04 are homologous to PLA_2s from *D. acutus* (ABY7N3, uncertain origin) and *Trimeresurus stejnegeri* (Q6H3C7, Taiwan, [64]), respectively. Variation is noted though among the matched PLA_2 pairs, in particular for Cr-PLA01 and Cr-PLA03. Cr-PLA01 (and the homologous A0A0H3U266) appears to have a higher substitution rate, exhibiting a considerably variable N-terminal sequence with marked non-synonymous substitution: --FNLWK---V-TG-EATKN-GM---N--PMKR-K (for the first 35 residues) (Figure 5). Cr-PLA03 has relatively less variability except for an important substitution at the 49th amino acid residue, in which the aspartic acid (D) is substituted for lysine (K), classifying it as a K49 PLA_2. The presence of aspartic acid at this position (D49) imparts catalytic activity critical for PLA_2 enzymatic reaction [65], and is highly conserved across lineages including *D. acutus* (Q1ZY03), *T. stejnegeri* (Q6H3C7), *T. puniceus* (Q2YHJ7) and *Gloydius brevicaudus* (A0A0H3U1W0) of the Old World as well as *Crotalus atrox* (P00624) and *Bothrops jararacussu* (Q90249) of the New World (Figure 5). The substitution of this amino acid (D) for lysine (K) in Cr-PLA04 is anticipated to result in a complete loss of phospholipase activity while gaining myotoxic activity, a toxic trait of neofunctionalization [66]. Clinically, systemic myotoxicity is rarely reported in *C. rhodostoma* envenoming, although the various PLA2 expressed may act in synergism with other toxins, contributing to inflammation, cytolytic and tissue destructive effects. Cr-PLA03, a novel K49 PLA_2 gene in *C. rhodostoma*, is highly homologous to the K49 variant found in *D. acutus* (A8Y7N3, Taiwan), and to a lesser degree, K49 PLA_2s from the New World pit vipers (e.g., Q90249, Q8UV27) (Figure 5) which are more clinically toxic and extensively studied. The finding supports the recruitment of K49 PLA_2 in the Asiatic basal crotalines and implies these are orthologous genes possibly with multiple ancestries.

Notwithstanding the sequence variability, these viperid PLA_2 sequences (Group II) show highly conserved 7 disulfide bonds and an extended C-terminal tail, with the absence of elapid or pancreatic loop which is apparent in elapid $svPLA_2$s which are structurally classified under Group IA and IB (Figure 5).

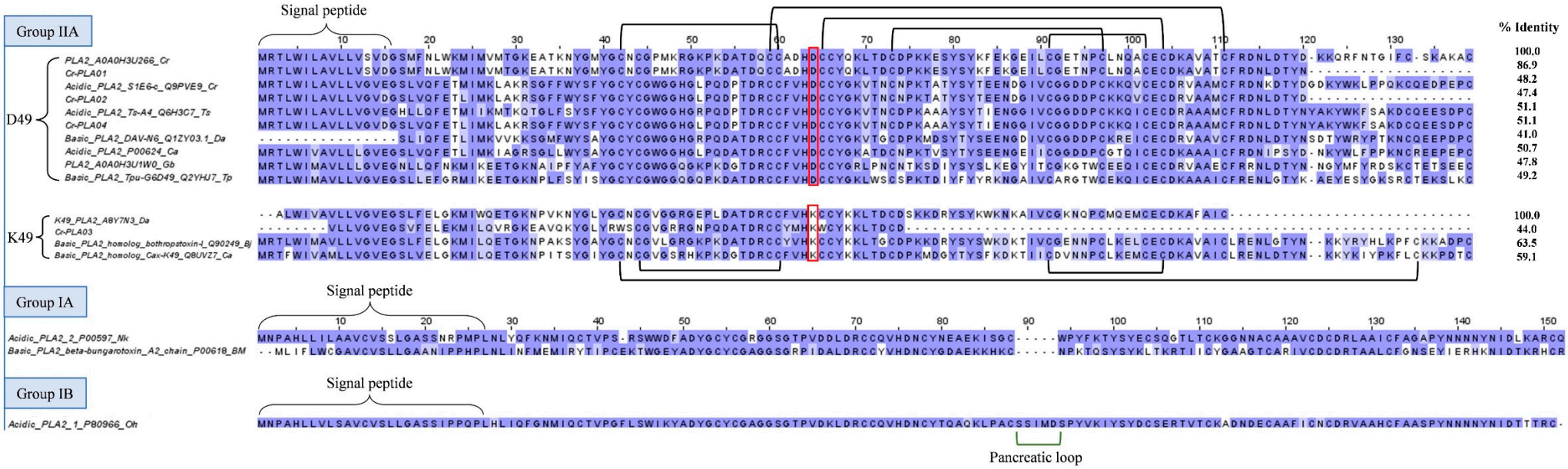

Figure 5. Multiple sequence alignment of PLA2s from *C. rhodostoma* venom gland transcriptome in comparison to PLA2s of related species, categorized according to Group (II, IA and IB) and subtype (D49, aspartic acid at 49th amino acid residue and K49, lysine at 49th amino acid residue). Black brackets: disulfide bond; highlighted in red box: D49 or K49 residue. Abbreviation: Cr – *Calloselasma rhodostoma*; Ts – *Trimeresurus stejnegeri*; Da – *Deinagkistrodon acutus*; Ca – *Crotalus atrox*; Gb: *Gloydius brevicaudus*; Tp: *Trimeresurus puniceus*; Bj – *Bothrops jararacussu*; Nk – *Naja kaouthia*; Bm – *Bungarus multicinctus*; Oh: *Ophiophagus hannah*.

doi:10.20944/preprints202304.0125.v1

 doi:10.20944/preprints202304.0125.v1

2.3.3. Bradykinin-potentiating peptide (BPP)/Angiotensinogen-converting enzyme inhibitor (ACEI) and Natriuretic peptide (NP)

Bradykinins (belong to the kallikrein-kinin system) and natriuretic peptides (NPs) are endogenous mammalian peptides involved in the homeostasis of body fluid and regulation of blood pressure [67]. Snake venoms would have evolved to target, mimic and disrupt these systems in the mammalian prey, subjecting it to a secondary effect of the venom, e.g., a significant drop in blood pressure that ultimately leads to subduing of the prey [68–70]. Bradykinin-potentiating peptides (BPPs) and C-type NPs (CNPs) are two well-characterized vasoactive blood pressure-modulating snake toxins found especially among pit vipers [69]. The occurrence of these peptides is mostly based on findings from cDNA sequencing, transcriptomics or genomics (e.g., *Bothrops jararaca*, *Trimeresurus flavoviridis*, *Trimeresurus gramineus* and *Agkistrodon halys blomhoffi* [71]; *Bothrops jararaca* [72]), while detection at the protein level has been scarce, presumably due to the peptides' low abundances and small molecular sizes that curtail detection (e.g., *Trimeresurus flavoviridis* [73]; Bitis gabonica rhinoceros [74]). The present study identified a substantial level of BPP/ACEI-CNP transcripts in the *C. rhodostoma* transcriptome (16.3% of all toxin FPKM), with homolog sequences matched to a multi-domain gene (UniProtKB: M5A7D0) coding for a BPP/ACEI-CNP precursor derived from a Thai specimen (Figure 6). Unfortunately, as of today no such protein has been isolated from *C. rhodostoma* venom, and no characterization has been done to locate the domains of BPP or ACEI and CNP in the gene.

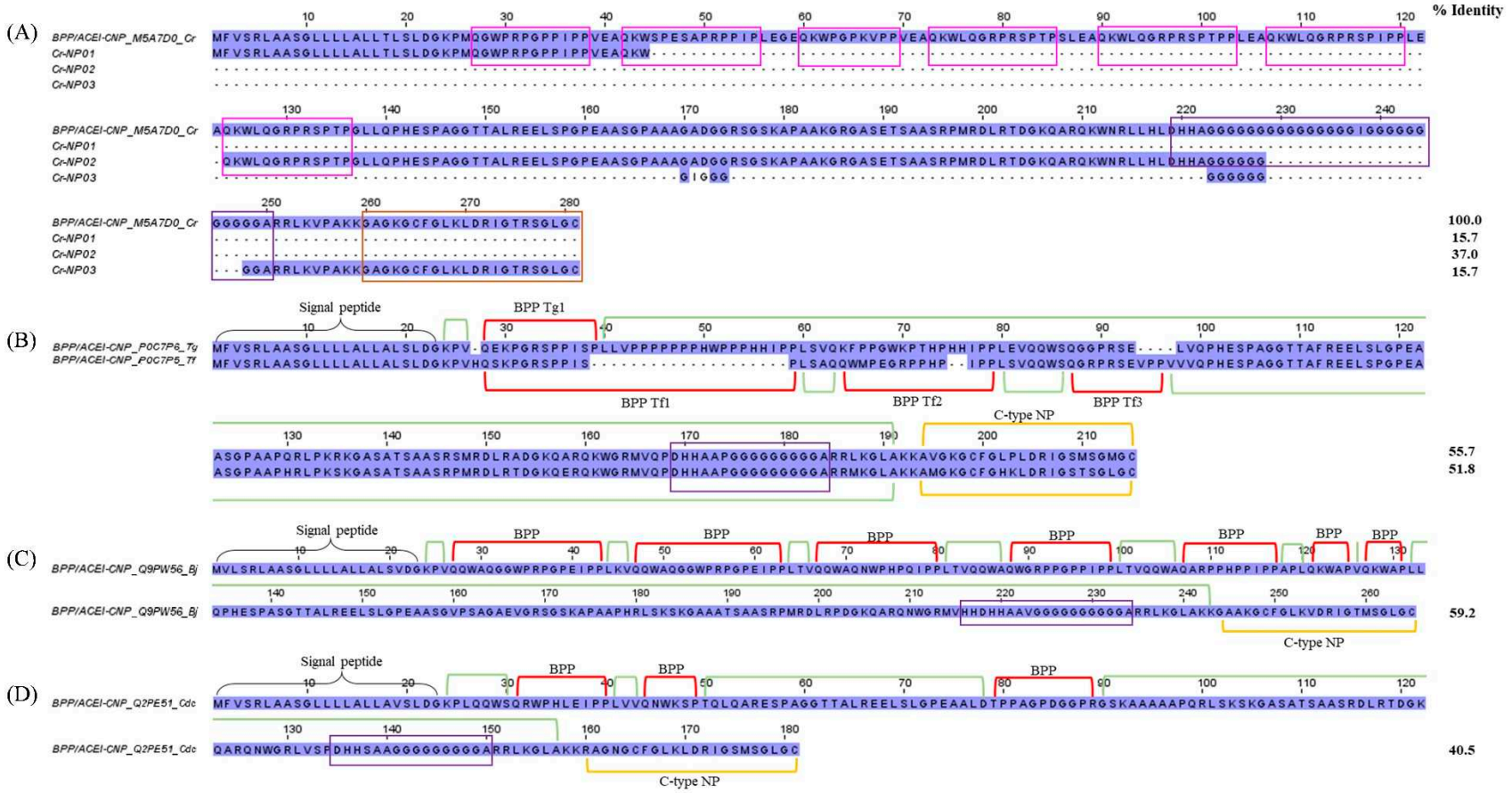

Figure 6. Multiple sequence alignment of bradykinin-potentiating peptides/angiotensin-converting enzyme inhibitors – C-type natriuretic peptide (BPP/ACEI-CNP) from *C. rhodostoma* venom gland transcriptome in comparison to BPP/ACEI-CNPs of related species. (A) CNP transcripts were aligned with BPP/ACEI-CNP of *C. rhodostoma* (UniProt KB: M5A7D0). (B) Alignments of BPP/ACEI-CNP from *Trimeresurus gramineus* (UniProt KB: P0C7P6) and *Trimeresurus flavoviridis* (UniProt KB: P0C7P5). (C) Alignment of BPP/ACEI-CNP from *Bothrops jararacussu* (UniProt KB: Q9PW56). (D) Alignment of BPP/ACEI-CNP from *Crotalus durissus collilineatus* (UniProt KB: Q2PE51). Highlighted in pink box: predicted bradykinin-potentiating peptides (BPP) domain of *C. rhodostoma*; highlighted in purple: spacer domain; highlighted in brown: probable C-type natriuretic peptide (NP) of *C. rhodostoma*. Black bracket: signal peptide domain; Green bracket: propeptide domain; Red bracket: BPP domain; Yellow bracket: C-type NP. Abbreviation: BPP Tg1 – Bradykinin-potentiating peptide *Trimeresurus gramineus* 1; BPP Tf1/2/3 – Bradykinin-potentiating peptide *Trimeresurus flavoviridis* 1/2/3; Cr – *Calloselasma rhodostoma*; Tg – *Trimeresurus gramineus*; Tf – *Trimeresurus flavoviridis*; Bj – *Bothrops jararacussu*; Cdc – *Crotalus durissus collilineatus*.

doi:10.20944/preprints202304.0125.v1

 doi:10.20944/preprints202304.0125.v1

To gain insight of its gene structure, we compare the transcript sequences with homologous BPP/ACEI-CNP genes retrieved from the database. Using the sequence from *Trimeresurus gramineus* (P0C7P6) as a classic template, this "composite" precursor gene has a signal sequence (a.a. residues 1-25) at the N-terminus, a pro-peptide and a modified residue (residues 24-26), followed by the BPP peptide coding region (residues 27-38), which is then spaced by another long pro-peptide or linker sequence (residues 39-186) with unidentified function. This is followed by another processing signal prior to the CNP sequence (residues 189-210) toward the C-terminus (Figure 6). It should be noted that in many other species, e.g., *Protobothrops flavoviridis* (P0C7P5), *Bothrops jararaca* (Q9PW56) and *Crotalus durissus collilineatus* (Q2PE51), more than one BPP coding regions (embedded variably within the long spacer prior to the signal sequence of CNP) have been identified in their respective orthologous genes (Figure 6). Our transcriptome finding revealed at least one BPP/ACEI and one CNP sequence that are unique to the *C. rhodostoma* of Malayan origin. By comparison to the other sequences, we managed to annotate the putative BPP- and CNP-coding regions in the transcripts and the precursor gene of *C. rhodostoma* (M5A7D0). Cr-NP01 has a propeptide identical to M5A7D0, and both share an identical BPP of 12 amino acid residues with an N-terminal sequence of QGWPRPGPPIPP. Classically, the BPPs have a pyroglutamic acid (Q) at the N-terminus and a notable high content of proline (P) residues [75], which gives to them some resistance to hydrolysis by aminopeptidases, carboxypeptidases and endopeptidases [76]. We further predicted a few additional BPP-coding domains in the precursor gene M5A7D0 based on the canonical BPP modular motif, and identified another BPP with the sequence QKWKQGRPRSPTP (13 amino acid residues) from the transcript Cr-NP02 (Figure 6). This was followed by a long spacer gene which carries an intriguingly high content of repeated glycine (G) residue within a well-conserved poly-His-poly-Gly region. The *C. rhodostoma* CNP sequence (GAGKGCFGLKLDRIGTRSGLGC) is discovered in the transcript Cr-NP03. Preceded by the conserved short signal peptide AKK, the CNP shares high homology with those of other pit vipers—consisting of 28 amino acid residues arranged in three highly conserved segments: GCFG, DRIG and SGLGC (C-terminus) (Figure 6). The first five residues of their N-termini undergo substitution but the overall structural constraint already established leaves little room for "improvisation". The natural occurrence of CNP is a redundancy to the BPPs, as both peptides ultimately promote the same biological effect, i.e., vasorelaxation induced by increased guanylate cyclase levels in the vascular smooth muscle cells [73]. While these vasodilating peptides can contribute to systemic hypotension, locally they may promote capillary permeability, thereby expediting the diffusion and spread of other toxins present in the venom. The exact pathophysiological role of BPP/ACEI and CNP in *C. rhodostoma* venom is as elusive as its gene origin and warrants further exploration.

2.3.4. Snake C-type lectins

Snake C-type lectins are categorized into snake lectins and snake C-type lectin-like proteins (snaclecs). While snake lectins are classic sugar-binding proteins, snaclecs are usually unable to recognize carbohydrates as they lack the Ca^{2+} ion binding loop involved in sugar binding [77,78]. Snake lectins are 26–28 kDa homodimers whereas snaclecs are more structurally complex, consisting of loop-swapping heterodimers formed by homologous α- and β-subunits, each with a molecular mass ranging from 13 to 18 kDa [79]. Snaclecs are functionally more relevant to toxicity than lectins, in particular for viperid snakes such as *C. rhodostoma.* Snaclecs, in the major forms of rhodocytin and rhodocetin, have been found to constitute up to 26% of its total venom proteome [13]. In the current study, the overall expression of snaclecs in *C. rhodostoma* venom gland is 10.01% of all toxin total FPKM (Figure 1). Two complete sequences of snaclecs (Cr-CTL03 and Cr-CTL01) matching identically to the rhodocytin subunit α (UniProt KB: Q9I840) and rhodocytin subunit ß (UniProt KB: Q9I841), respectively, were identified from the venom gland transcriptome (Figure 7). Rhodocytin, also known as aggretin, is a 29 kDa heterodimer snaclec that interacts with the CLEC-2 (C-type lectin 2) receptor on platelet, inducing platelet aggregation [80,81]. The subunits Cr-CTL03 and Cr-CTL01 of *C. rhodostoma* show a lesser degree of sequence identities to the corresponding subunits of snaclec from *D. acutus*. Multiple substitutions contributing to the heterogeneity were observed between the

 doi:10.20944/preprints202304.0125.v1

two species. Six cysteine residues and three disulfide links are nonetheless well-conserved in each of the monomeric subunits (Figure 7).

The other two transcripts of lower expressions, Cr-CTL04 and Cr-CTL05, mark the expression of rhodocetin, a snaclec that binds to GPIb and inhibits VWF-dependent platelet aggregation [78,82]. Rhodocetin in fact is an $\alpha_2\beta_1$ integrin-specific antagonist, and a hetero-tetrameric molecule composed of alpha-beta and gamma-delta subunits arranged orthogonally in a cruciform pattern [83]. The sequence of Cr-CTL04 varies from the rhodocytin beta subunit Cr-CTL01 by acquiring an additional segment of LDLVI while losing some residues in comparison to rhodocytin (Figure 7). Cr-CTL05 is identical to the delta subunit of rhodocetin (Figure 7), while rhodocytin as a dimer has no comparable delta subunit. All subunits of rhodocetin and rhodocytin, nonetheless, shared conserved cysteine residues and disulfide bonds, which are responsible for the overall well-conserved and robust structure of snaclecs. Apparently, both snaclecs appear to share the same target receptor (platelet) but rhodocetin interacts with $\alpha_2\beta_1$ integrin (major collagen receptor on platelet) in an RGD-independent manner, hampering platelet activation [84]. Simply put, while rhodocytin is a platelet aggregator, rhodocetin inhibits platelet aggregation. This is again a redundancy of venom function—although the actions on platelet seem to be opposing, the resultant effect is thrombocytopenia with dysfunctional platelets as part of the pathophysiology of *C. rhodostoma* envenoming.

In addition, the *C. rhodostoma* venom gland transcriptomics reveals a novel transcript of snake C-type lectin (Cr-CTL02), which is homologous to venom lectins from *Agkistrodon piscivorus leucostoma* and *Crotalus atrox* (Figure 7). The role of galactose-biding snake lectins in *C. rhodostoma* envenoming is unclear, although experimentally, lectins from many New World species were found to exhibit hemagglutinating, inflammatory and platelet aggregating activities [85].

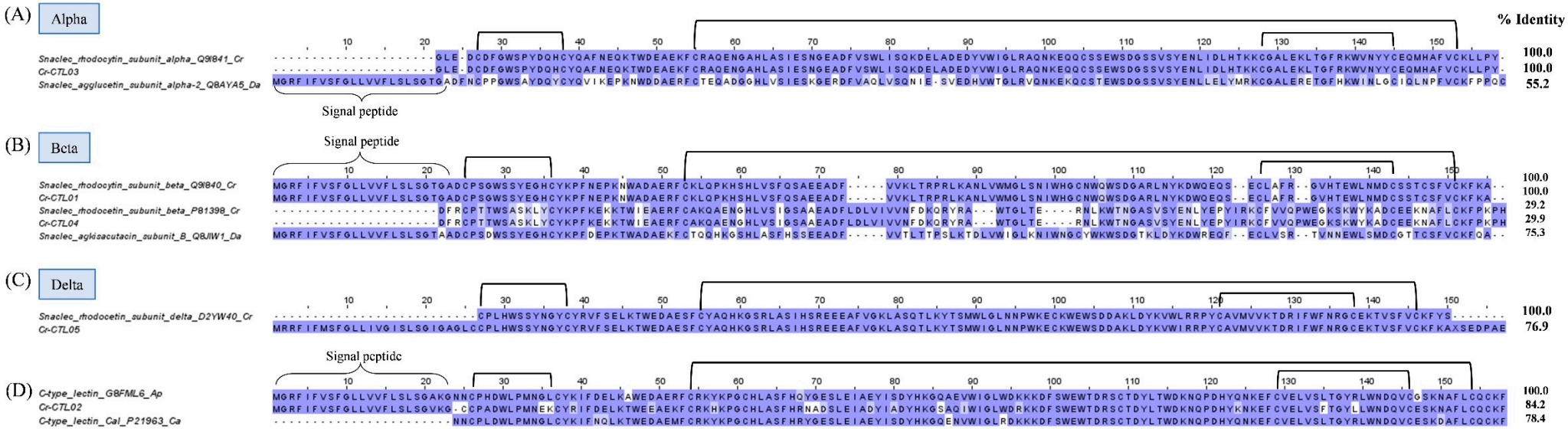

Figure 7. Multiple sequence alignment of Cr-CTL01–04 (Putative snake C-type lectin like protein (snaclec) identified in *Calloselasma rhodostoma* venom gland transcriptome in comparison to selected snaclec retrieved from UniProtKB database. Snaclecs were categorized according to alpha (A), beta (B), delta (C), and single (D) domain. Black brackets depict disulfide bonds. Abbreviation: Cr – *Calloselasma rhodostoma*; Da – Deinagkistrodon acutus; Ap – Agkistrodon piscivorus; Ca – *Crotalus atrox*.

doi:10.20944/preprints202304.0125.v1

 doi:10.20944/preprints202304.0125.v1

2.3.5. Snake venom serine proteinase (SVSP)

Snake venom serine proteinases (SVSPs) catalyze a broad range of reactions targeting the coagulation cascade, kallikrein-kinin and fibrinolytic systems, complement system, endothelial cells and platelets, all of which eventually augment the hemotoxic effect of viperid envenoming [86]. The SVSP abundance in *C. rhodostoma* venom proteome is well established to be around 15-25% of total venom proteins [12,13,15], while in the current transcriptomic study, the SVSP transcript level is disproportionately low (2.8% of all toxin FPKM) (Figure 1). The dominant transcript identified (Cr-SSP01) correlates with the major SVSP expressed in the venom proteome, i.e., ancrod (or arvin) [13], which belongs to the thrombin-like SVSP (TL-SVSP) subgroup. The sequence of Cr-SSP01 matches identically to ancrod of *C. rhodostoma* (unknown locality) reported in 1992 [87] but has a low degree of sequence identities when comparing with TL-SVSP from *D. acutus* (acutobin, 59.7%), *Gloydius brevicaudus* (kangshuanmei, 62.7%), *Trimeresurus (Viperidovipera) stejnegeri* (stejnobin, 63.1%), and *Crotalus durissus terrificus* (gyroxin, 57.1%) (Figure 8). Despite the variation, these TL-SVSP of various pit viper genera shares the conserved catalytic triad of serine (Ser195), histidine (His57) and aspartate (Asp102) (residue numbering according to the chymotrypsinogen system) (Figure 8) [88]. As reported previously, ancrod was heavily glycosylated [89], consistent with our observation of five N-linked glycosylation sites revealed in Cr-SSP01 (Figure 8). Stejnobin too has given N-link glycosylation site, while acutobinand kangshuanmei have four, and gyroxin has only one. Glycosylation plays an important role for the maintenance of homeostasis within the gland lumen by improving protein solubility, protection from proteolytic attack, quality control of protein folding, prolonging protein's plasmatic half-life, target site recognition, and modulation of immunogenicity [90–93].

TL-SVSPs are functionally similar to thrombin in some ways but are also dissimilar in many aspects (hence, "thrombin-like enzyme" is a slight misnomer). Ancrod represents a classical TL-SVSP that has been extensively studied and even clinically trialed as an anti-coagulant, though with a somewhat discouraging outcome [94,95]. Consumptive coagulopathy with defibrination is the major fatal hemotoxic effect of *C. rhodostoma* envenoming, and this is produced by the cleavage of fibrinogen α chain, releasing fibrinopeptides A, AP and AY, leaving fibrins to form tenuous, readily dissolved thrombi or micro-clots [49]. Unlike thrombin, most TL-SVSPs do not activate factor XIII to stabilize the thrombus. The repeated dissolution of clots in the background of continuous degradation of fibrinogen eventually leads to venom-induced consumptive coagulopathy and the consequent massive bleeding, hypovolemic shock, hypo-perfusion and multi-organ failure [96].

2.3.6. L-amino acid oxidase (LAAO)

L-Amino acid Oxidases (LAAO) are flavoenzymes that catalyze the stereo-specific oxidative deamination of L-amino acids to produce α-ketoacids, ammonia and hydrogen peroxide (H_2O_2), the last of which is believed to account for the diverse toxicity of the enzyme, including hemorrhagic, hemolytic, anti-microbial, cytotoxic and inflammatory activities [97,98]. The present study reveals a low level of LAAO transcripts in the *C. rhodostoma* venom gland transcriptome (2.25% of all toxin total FPKM), represented by one mRNA with full sequence (Cr-LAO01). The sequence of Cr-LAO01 is identical to the reported *C. rhodostoma* L-amino acid oxidase (UniProt KB: P81382) as shown in Figure xx covering the three flavine adenine dinucleotide (FAD) binding sites, and three substrate binding sites [99] (Figure 9). The abundance of LAAO in *C. rhodostoma* venom proteome is approximately 7% [13]. The pathological role of this toxin in *C. rhodostoma* envenoming is unclear but could be related to local inflammatory responses as it stimulates neutrophil activation and production of inflammatory mediators [100].

Comparison of Cr-LAO01 with LAAOs from different snake genera reveals well-conserved sequences among the viperids (vipers and pit vipers), while the elapid LAAOs appear to be relatively more variable (sequence identity well below 80%) (Figure 9). In general, LAAO abundances are much lower and sometimes undetected in the venoms of elapid snakes, with the exception of King Cobra (*Ophiophagus hannah*) whose venom contains LAAO >10% of total venom proteins [101].

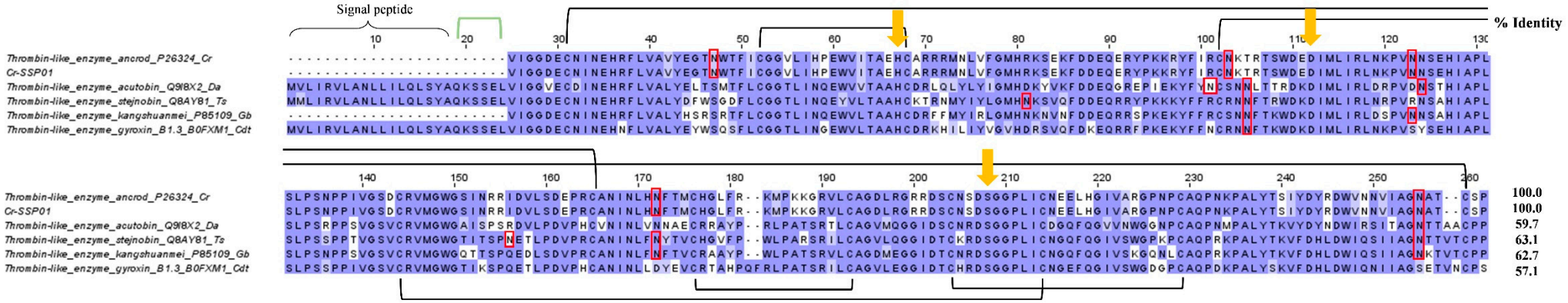

Figure 8. Sequence alignment of Cr-SSP01 (Putative snake venom serine proteinase (SVSP); thrombin-like enzyme) identified in *Calloselasma rhodostoma* venom gland transcriptome in comparison to selected SVSP from UniProtKB database. Black bracket: disulfide bond; Green bracket: propeptide. Highlighted in red: N-linked glycosylation site. Yellow arrows depicted conserved catalytic triad of histidine (His57), aspartate (Asp102) and serine (Ser195) with numbering based on chymotrypsinogen system. Abbreviation: Cr – *Calloselasma rhodostoma*, Da – *Deinagkistrodon acutus*; Ts – *Trimeresurus stejnegeri*; Gb – *Gloydius brevicaudus*; Cdt – *Crotalus durissus terrificus*.

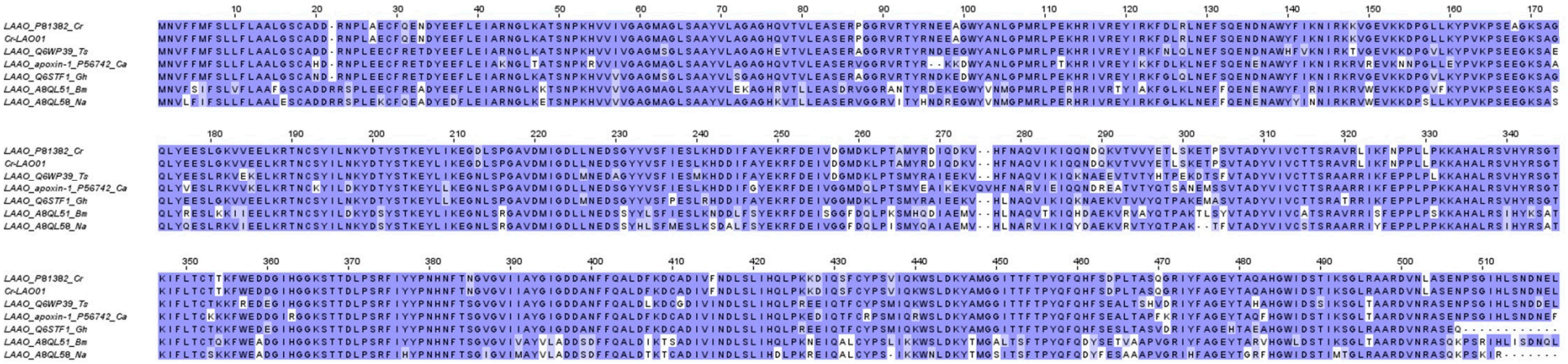

Figure 9. Sequence alignment of Cr-LAO01 (Putative L-Amino acid Oxidases (LAAO) identified in *Calloselasma rhodostoma* venom gland transcriptome in comparison to selected LAAO from UniProt KB database. Abbreviation: Cr – *Calloselasma rhodostoma*; Ts – *Trimeresurus stejnegeri*; Ca – *Crotalus atrox*; Gh – *Gloydius halys*; Bm – *Bungarus multicinctus*; Na – *Naja atra*.

doi:10.20944/preprints202304.0125.v1

 doi:10.20944/preprints202304.0125.v1

2.4. Low-abundance toxin transcripts

2.4.1. Toxins detected in both venom gland transcriptome and venom proteome

In *C. rhodostoma* venom proteome, a small amount of cysteine-rich secretory protein (CRiSP) was detected (1% of total venom proteins) [13], comparable to the transcript level of CRiSP (0.90% of all toxin total FPKM) found in this study (Figure 1). A full sequence of novel CRiSP (Cr-CRP01) was uncovered for this species. Probably due to limited information on CRiSP for closely related viperids in the database, Cr-CRP01 was matched with the highest homology to CRiSP LCCL domain (UniProt KB: VBN17) from the elapid *Ophiophagus hannah* (Supplementary File 2). Being cysteine-rich as the name implies, these toxins have a highly conserved pattern of 16 cysteine residues. The toxicity is unclear but they can cause blockage of calcium channels to inhibit smooth muscle contraction [102].

The study further unveils the gene expression of phosphodiesterases (PDE, a nuclease family) and 5'nucleotidases (5NT, a nucleotidase family) in the *C. rhodostoma* venom gland. The transcript levels of PDE and 5NT were found low at 0.01% and 0.28% of all toxin total FPKM, respectively. In the venom proteome of *C. rhodostoma*, these were known to be low-abundance proteins (> 1% of total venom proteins) [13], and their function is putatively associated with the dissemination of venom toxins in the prey (or snakebite victims) [103]. PDE, as an exonuclease, cleaves DNA in a 3',5'-direction, releasing 5'-mononucleotides that serve as nucleotide substrates for 5'-NTs, which in turn, liberate free nucleosides. The generation of purinic nucleosides (mainly adenosine) and other purine derivatives could lead to vasodilation and low blood pressure (facilitating prey immobilization), inflammation with enhanced toxin dissemination, and possible blockade of some neurotransmitters as in sympatholytic activity of adenosine [103]. Furthermore, the degradation of ADP and ATP by PDE may theoretically impairs platelet aggregation activity at the wound site. Despite the low expressions of these enzyme toxins, the present study successfully uncovered the full-length sequences of PDE and 5'-NT of *C. rhodostoma* (Supplementary File S2), which have not been characterized in depth. The sequences, representative of a basal pit viper from Asia, are matched to those of New World pit viper (*Crotalus* spp.), whose PDEs and 5'-NTs were more extensively studied and full sequences are readily available. Notwithstanding the distant phylogeographical relationship between *C. rhodostoma* and these New World pit vipers, a high degree of homology is found across the venom enzymes. There is very little or practically no important substitution observed in these orthologous genes of PDEs and 5'-NTs in spite of their high molecular weights (> 100 kDa). The finding suggests these proteins, though lowly expressed, are highly conserved with low evolution rates across different lineages. Presumably, these genes maintain basic or secondary function of venom but are not subject to intense selective pressures as with key toxins such as SVMP and SVSP, which are vital for predation. Minimum mutations involving these venom genes may help maintain existing beneficial phenotypes while promoting the rapid evolution of advantageous traits without interference from deleterious mutations.

In proteomic studies, phospholipase B (PLB), nerve growth factor (NGF) and aminopeptidase (APP) have been detected at very low abundances (< 1% of total venom proteins, respectively) in *C. rhodostoma* venom [13,16]. The present study identified the toxin genes of these minor proteins, and showed the transcriptions of these genes are indeed low: PLB, 0.25%; NGF, 0.07%; APP, 0.01%, of all toxin total FPKM (Table 2). PBL cleaves ester linkages of membrane glycerophospholipids at both positions *sn-1* and *sn-2* [104], and this may be the basic mechanism in which venom PLB exerts its strong hemolytic and cytotoxic activities [105–107]. The full sequence of *C. rhodostoma* venom PLB is matched with high homology (>90% sequence identity) to that of *Bothrops moojeni*, indicating well conservation of this protein in venom despite their distant phylogenetic relationship (Supplementary File S2).

The full sequence of a snake venom NGF gene was discovered in the *C. rhodsotoma* venom gland transcriptome. The sequence is near identical to that of *Protobothrops flavoviris* (UniProt KB: B1Q3K2) (Supplementary File S2). The pathophysiological role of NGF, a key member of the neurotrophin family in snakebite envenoming is vague, but this non-enzymatic toxin may have ancillary function

 doi:10.20944/preprints202304.0125.v1

in preserving the toxic cocktail of venom by inhibiting venom metalloproteinase-dependent proteolysis [108]. Sunagar et al. [109] demonstrated that venom-secreted NGFs have characteristics consistent with the typical accelerated molecular evolution of venom components driven by positive selection, suggesting their participation in prey envenomation.

One APP gene was also found in the study, identified based on its full sequence which is homologous to that of *Ovophis okinavensis* (Supplementary file S2). APP belongs to metalloproteases, but little is known regarding its function and how it varies from the more widely studied SVMPs. Earlier, the characterization of an APP isolated from the venom of *Bitis arientans* (an African pit viper) suggested this enzyme might cause alteration of blood pressure and impair brain function in snakebite victims [110].

2.4.2. Toxins detected exclusively in venom gland transcriptome

Previous studies have shown that not all toxin transcripts detected in snake venom gland transcriptomes are correspondingly translated and detected in the venom proteomes [24,28,29]. In the present study, we detected low-abundance transcripts of genes whose products have never been reported at the protein level, in any biochemical and proteomic study of *C. rhodostoma* venom [12–14,16]. These are genes putatively coding for nucleobindin (NCB), vascular endothelial growth factor (VEGF), three-finger toxin (3FTx), and Kunitz-type serine protease inhibitor (KSPI), each with a transcript level of 0.19%, 0.05%, 0.02%, and below 0.01%, respectively (Table 2).

The functions of these enigmatic genes are elusive (at least, in this species) but worth discussing for their potential application and bioprospecting use. Nucleobindin (NCB) is the precursor nesfatin-1, an endocrine factor associated with epilepsy in mammals [111]. In a proteopeptidome characterization of Bothrops jararaca venom, snake venom NCB was speculated to induce excitotoxicity and cause transient disorientation in prey, thus facilitating predation [111]. VEGF is found in some viperid venoms, notably from Russell's Viper (Daboia russelii and D. siamensis) [54,112], and its function has been linked to the increment of vascular permeability, the underlying mechanism of capillary leak syndrome [113,114] which is uncommon in C. rhodostoma envenoming. The transcriptomic finding of three-finger toxins (3FTx), which are canonical of and almost exclusive to elapid snakes, is puzzling but not totally unexpected, as several authors have also reported the detection of 3FTx transcripts in the venom glands of pit vipers and colubrid snakes [115,116]. Intriguingly, this is the first report of detection of 3FTx transcripts from a basal Old World pit viper, which is evolutionarily a more primitive species among the advanced venomous snakes. The presence of these genes across the various snake families suggests these genes have a much earlier recruitment that predates the divergence between viperids, elapids, and perhaps colubrids, with selection favoring the expansion and neofunctionalization of these genes in the more derived elapid snakes while silencing those in the viperids. The Kunitz-type serine protease inhibitors (KSPI), on the other hand, may serve to inhibit proteolysis by serine proteases in a venom. In some viperid venoms, KSPIs such as Ruvikunin and Rusvikunin-II purified from the native Rusvikunin complex of Pakistan Russell's Viper have been shown to exhibit anti-coagulant activity [117].

3. Conclusion

The present study unravels the complexity and diversity of venom genes in *C. rhodostoma* (Malayan Pit Viper), a basal pit viper species that is endemic to Southeast Asia. The *de novo* assembled venom gland transcriptome uncovers 16 toxin families in which numerous genes with novel sequences were identified. The major toxins determined, along with their differential expression levels are mostly reflective of the venom proteomic profile of *C. rhodostoma* notwithstanding incongruencies which could be resulted from several factors, including limited time points of tissue sampling (which unfortunately ignored the kinetics and time course of venom production), varying mRNA half-lives of toxins post-transcription, and complex regulations of protein translation in the venom gland and subsequent release as secretory proteins into the lumen. The findings derived from a snapshot of transcriptome profile as such, nonetheless, provide deep insights into the gene structures and functions of toxins from this unique species, improving our understanding of its

 doi:10.20944/preprints202304.0125.v1

relevance in the context of snakebite envenoming. Furthermore, the complete sequencing of toxins—many of which have not been fully characterized previously in this species, enriches snake venom databases and provides information to support further studies in venom evolution, snakebite treatment, antivenom design, and drug discovery.

5. Materials and Methods

5.1. Preparation of C. rhodostoma venom gland tissue

The adult *C. rhodostoma* specimen was collected from Kedah, a northern state in Peninsular Malaya. The venom was milked four days prior to venom gland tissue collection to promote transcription [118]. The venom glands were collected following euthanasia and sectioned into dimensions of 5 x 5 m. The sectioned tissue was immersed in RNAlater® solution (Ambion, TX, USA) at 4 °C overnight and stored at −80 °C until further use. The study was carried out in line with protocols approved by the Institutional Animal Use and Care Committee (IACUC) of the University of Malaya, Malaysia (Approval code: #2013-11-12/PHAR/R/TCH).

5.2. RNA extraction and mRNA purification

The venom gland tissue was homogenized in a 1 mL glass homogenizer with TRIzol solution (Invitrogen, Carlsbad, CA, USA). This was followed by the isolation using chloroform and treated with RNA-free DNAase I (Thermo Fisher Scientific, Waltham, MA, USA), to separate cellular debris and residual DNA. The isolated RNA was then purified via isopropyl alcohol ethanol precipitation. Polyadenylated mRNA was subsequently purified with oligo(dT) magnetic beads (Illumina TruSeq Stranded mRNA) (Illumina, San Diego, CA, USA) as per manufacturer's instructions. The quality of the purified total RNA was assessed using Agilent 2100 Bioanalyzer (RNA 6000 NanoKit) (Agilent Technologies, Waldbronn, Germany).

Enriched poly(A)+ mRNA isolated from the total venom gland RNA was used for cDNA construction. The isolated mRNA was fragmented into short fragments, which acted as templates for cDNA synthesis [119]. Random hexamer-primer (N6) was used to synthesis the first-stranded cDNA, followed by second-strand cDNA synthesis with the double-stranded cDNA as input materials, using second-strand buffers, dNTPs, RNase H and DNA polymerase I. From these cDNA, a paired-end library was synthesized using the Genomic Sample Prep kit (Illumina, San Diego, CA, USA), according to the manufacturer's instructions. The cDNA fragments generated were purified with QIAquick PCR extraction kit (Qiagen, Valencia, CA, USA) and dissolved in elution buffer for end repair and the addition of poly(A) to aid in the subsequent ligation of Illumina adaptors that contain a single thymine (T) base overhang at the 30 ends. Following the ligation, these cDNA fragments were amplified via polymerase chain reaction (PCR) electrophoresed on a 1.5–2% TAE (Tris base, acetic acid and EDTA) agarose gel. From the gel, suitable fragments (200–700 bp) were selected as templates for subsequent PCR amplification. Sequencing of the amplified samples library was achieved in a single lane on the Illumina HiSeq™ 2000 platform (Illumina, San Diego, CA, USA) with 100-base-pair, paired-end reads.

5.3. Filtration of raw sequenced reads

Sequenced data generated from Illumina HiSeq™ 2000 platform were transformed by base calling into sequence data, called the raw reads and stored in a FASTQ format. Prior to transcriptome assembly, raw reads were filtered to generate clean reads as part of the quality control process in the pre-analysis stage [120]. This involved the removal of (i) adaptors; (ii) reads with >5% of unknown nucleotides or (iii) low-quality reads with >20% of low-quality bases (determined as base quality < 10).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2023 doi:10.20944/preprints202304.0125.v1

5.4. De novo transcriptome assembly

The *de novo* transcriptome assembly was performed using a short-reads assembly program, Trinity (version 2.0.6) [121,122]. Three independent software modules, that is, Inchworm, Chrysalis and Butterfly, comprised the Trinity program were sequentially applied to process the large volumes of RNA-seq reads. In brief, this was based on the algorithm of *de Bruijn* graphs construction, which began by aligning k-mers (k = 25) and reads with a certain length of overlap were joined to form linear contigs. The reads were mapped back onto contigs and by referring to paired-end reads, contigs from the same transcript, as well as the distances between them were determined. The contigs were then partitioned into clusters, each of which carried a complete set of *de Bruijn* graphs (representing the transcriptional complexity at a given gene or locus). The graphs were independently processed to obtain full-length transcripts for alternatively spliced isoforms and to tease apart transcripts that corresponded to paralogous genes. The clean read Q20 percentage, a point of reference for quality control assessment was obtained as a benchmark for successful *de novo* assembly of the transcriptome.

5.5. Clustering and functional annotation of transcripts

The transcript sequences generated through Trinity were called Unigenes. Unigenes from the transcriptome assembly were further processed for sequence splicing and redundancy removal with TGI clustering tools (TGICL, version 2.1) to acquire non-redundant (NR) transcripts at the longest possible length [123]. The transcripts were then subjected to family clustering, which resulted in two classes of transcripts: (a) clusters, with a prefix CL and the cluster ID behind as contig; (b) singletons, whose ID was simply left with a prefix of Unigene. In each cluster, there were several transcripts with sequence similarities among them being >70%; while the singletons 'Unigenes' lack overlapping with other fragments at the given stringency. The value 70% was used to categorize the assembled sequences based on similarity; sequences similar to each other (may or may not be homologous as having > 90% similarity) were grouped under a cluster comprising various contigs.

Following this, transcript Unigenes were then aligned with BLASTx to protein database in priority order to NCBI non-redundant database (NR), with a cut-off value of E < 10-5. Proteins with the highest ranks in the BLASTx results were referred to determine the coding region sequences of Unigenes, followed by translation into amino acid sequences (using standard codon table). Hence, both nucleotide sequences (50 to 30) and amino acid sequences of the Unigene-coding regions were acquired. To remove redundancy from each cluster, the longest sequence in each cluster was chosen as the transcript, meanwhile, the length of scaffold was extended based on overlapping sequences using Phrap assembler (release 23.0) (http://www.phrap.org). The distributions of the length of contigs, scaffolds and Unigenes were calculated and the N50 length (assembly quality indicator) was set at N50 > 500 for assembly success.

5.6. Quantifying transcript abundance

Clean reads were aligned to Unigene using Bowtie2 [124]. The transcript abundances were calculated using RNA-seq with expectation maximization (RSEM) tool [125].

Fragments per kilobase of exon model per million reads mapped (FPKM) were used to determine the transcript abundance for the identified genes (Mortazavi et al., 2008). FPKM is the summation of normalized read counts based on gene length and the total number of mapped reads. The data was obtained using the RSEM tool in conjunction with Trinity based on a computational formula:

$$\text{FPKM of gene A} = \frac{10^6 \text{B}}{NC/1000}$$

FPKM is the expression of gene A; B is the number of fragments/reads which are aligned to gene A; N is the total number of fragments/reads that are aligned to all genes; C is the base number in the coding sequence of gene A.

5.7. Categorization of transcripts

The *de novo* assembled transcripts were subjected to BLASTx search to obtain the closest resembling sequences from the NR protein database for further classification based on functional annotations. The transcripts (Unigenes) were then sifted to remove those with an FPKM value of less than 1, followed by categorization into three groups: "toxins," "non-toxins" and "unidentified" " [25,27]. "Toxin" transcripts were recruited by toxin-related keyword searches against the annotated transcripts. "Non-toxin" and "unidentified" groups contain transcripts of cellular proteins or house-keeping genes and transcripts that could not be identified, respectively. The redundancy of gene expression was determined by dividing the total FPKM of each group by the total number of transcripts in the respective group of transcripts [27]. In the toxin group, the amino acid sequences were used to further validate the toxin identity through the BLASTp suite (Basic Local Alignment Search Tool-Protein) in the UniProt (Universal Protein Resource Knowledgebase) database platform. The transcripts were searched against the Serpentes database (taxid: 8570) and validated based on the lowest E-score value with the highest percentage of sequence similarity (updated as of 29 Jan 2023).

5.8. Sequence alignment and analysis

Multiple sequence alignment was conducted using Jalview software (version 2.10.5; University of Dundee with Dundee Resource for Protein Sequence Analysis and Structure Prediction, Scotland, United Kingdom) [126] and MUSCLE (Multiple Sequence Comparison by Log-Expectation, version 3.8.31) [127] program. Sequences of related species used in multiple sequence alignment were retrieved from UniProtKB depository (accessed date: 14 February 2023) (http://www.uniprot.org/). The selection was based on their relevance to the toxins in comparison to elucidate the similarity, variation and conserved regions of the sequences. Aligned sequences were coloured with BLOSUM62 colour scheme, where the colour intensity reflects the chance of amino acid substitution i.e., intense purple = low chance of amino acid substitution; white = high chance of amino acid substitution.

5.9. Supporting data

Sequencing data from the venom gland transcriptomics of *C. rhodostoma* was deposited in National Centre for Biotechnology Information (NCBI) Sequence Read Archive (https://submit.ncbi.nlm.nih.gov/subs/sra/) (submitted on 30 September 2019) under SRA accession: PRJNA549256.

5.10. Statistical analyses

The correlation between venom gland transcriptome (present study) and venom proteome data [13] of *C. rhodostoma* was analyzed. Transcripts that co-expressed in both sets of data were selected and assessed statistically (p-value < 0.05) with Pearson product-moment correlation coefficient test (IBM SPSS Version 23) to determine the correlativity of data.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: Conceptualization, C.H.T., K.Y.T. and N.H.T.; methodology, C.H.T., H.P.C. and T.S.N.; software, H.P.C. and T.S.N.; validation C.H.T., K.Y.T., H.P.C. and T.S.N.; formal analysis, C.H.T..; investigation, C.H.T., K.Y.T., ; resources, C.H.T, K.Y.T. and N.H.T.; data curation, C.H.T., K.Y.T.; writing—original draft preparation, C.H.T; writing—review and editing, C.H.T., T.S.N., K.Y.T., N.H.T. and H.P.C.; visualization, C.H.T. and H.P.C.; supervision, C.H.T. and K.Y.T.; project administration, C.H.T.; funding acquisition, C.H.T., K.Y.T. and N.H.T.. All authors have read and agreed to the published version of the manuscript.

Funding: The study was funded by the Ministry of Higher Education, Government of Malaysia (FRGS/1/2019/SKK08/UM/02/19), and Universiti Malaya, Malaysia (BKS003-2020).

Institutional Review Board Statement: The animal study protocol was approved by the Institutional Animal Use and Care Committee (IACUC) of the University of Malaya, Malaysia (Approval code: #2013-11-12/PHAR/R/TCH).

Acknowledgments: Authors thank the University of Malaya, Malaysia for providing facilities for the research.

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

References

1. Gutiérrez, J.M.; Calvete, J.J.; Habib, A.G.; Harrison, R.A.; Williams, D.J.; Warrell, D.A. Snakebite envenoming. *Nature Reviews Disease Primers* **2017**, *3*, 17063.
2. Sanders, K.L.; Lee, M.S. Molecular evidence for a rapid late-Miocene radiation of Australasian venomous snakes (Elapidae, Colubroidea). *Molecular Phylogenetics and Evolution* **2008**, *46*, 1165-1173.
3. Alencar, L.R.V.; Martins, M.; Greene, H.W. Evolutionary History of Vipers. *eLS* **2018**, doi:doi:10.1002/9780470015902.a0027455
4. WHO, W.H.O. Guidelines for the management of snakebites, 2nd ed. *WHO Regional Office for South-East Asia*. **2016**.
5. Wüster, W.; Peppin, L.; Pook, C.E.; Walker, D.E. A nesting of vipers: phylogeny and historical biogeography of the Viperidae (Squamata: Serpentes). *Molecular Phylogenetics and Evolution* **2008**, *49*, 445-459.
6. Das, I. *Field guide to the reptiles of South-East Asia*; Bloomsbury Publishing: 2015.
7. Uetz, P.; Freed, P.; Aguilar, R.; Reyes, F.; Hošek, J. The Reptile Database, http://www.reptile-database.org, accessed [20 March 2022]. Available online: http://www.reptile-database.org (accessed on
8. Warrell, D.A. Tropical snakebite-clinical-studies in South East-Asia. *Toxicon* **1985**, *23*.
9. Reid, H.A.; Chan, K.; Thean, P. Prolonged coagulation defect (defibrination syndrome) in Malayan viper bite. *Lancet* **1963**, 621-626.
10. Reid, H.; Thean, P.; Chan, K.; Baharom, A. Clinical Effects of Bites by Malayan Viper (Ancistrodon rhodostomd). *Lancet* **1963**, 617-621.
11. Wongtongkam, N.; Wilde, H.; Sitthi-Amorn, C.; Ratanabanangkoon, K. A study of 225 Malayan pit viper bites in Thailand. *Military medicine* **2005**, *170*, 342-348.
12. Kunalan, S.; Othman, I.; Syed Hassan, S.; Hodgson, W. Proteomic Characterization of Two Medically Important Malaysian Snake Venoms, Calloselasma rhodostoma (Malayan Pit Viper) and Ophiophagus hannah (King Cobra). *Toxins* **2018**, *10*, 434.
13. Tang, E.L.H.; Tan, C.H.; Fung, S.Y.; Tan, N.H. Venomics of Calloselasma rhodostoma, the Malayan pit viper: A complex toxin arsenal unraveled. *Journal of Proteomics* **2016**, *148*, 44-56. https://doi.org/10.1016/j.jprot.2016.07.006.
14. Vejayan, J.; Khoon, T.L.; Ibrahim, H. Comparative analysis of the venom proteome of four important Malaysian snake species. *Journal of venomous animals and toxins including tropical diseases* **2014**, *20*, 6.
15. Ali, S.A.; Baumann, K.; Jackson, T.N.W.; Wood, K.; Mason, S.; Undheim, E.A.B.; Nouwens, A.; Koludarov, I.; Hendrikx, I.; Jones, A.; et al. Proteomic comparison of Hypnale hypnale (Hump-Nosed Pit-Viper) and Calloselasma rhodostoma (Malayan Pit-Viper) venoms. *Journal of Proteomics* **2013**, *91*, 338-343. https://doi.org/10.1016/j.jprot.2013.07.020.
16. Tang, E.L.H.; Tan, N.H.; Fung, S.Y.; Tan, C.H. Comparative proteomes, immunoreactivities and neutralization of procoagulant activities of Calloselasma rhodostoma (Malayan pit viper) venoms from four regions in Southeast Asia. *Toxicon* **2019**, *169*, 91-102.
17. Tang, E.L.; Tan, C.H.; Fung, S.Y.; Tan, N.H. Venomics of Calloselasma rhodostoma, the Malayan pit viper: A complex toxin arsenal unraveled. *J Proteomics* **2016**. https://doi.org/10.1016/j.jprot.2016.07.006.
18. Ismail, A.K. Snakebite and envenomation management in Malaysia. *Toxinology: Clinical Toxinology in Asia Pacific and Africa* **2015**, *2*, 71-102.
19. Tan, C.H. Snake Venomics: Fundamentals, Recent Updates, and a Look to the Next Decade. *Toxins* **2022**, *14*, 247.
20. Neiva, M.; Arraes, F.B.; de Souza, J.V.; Radis-Baptista, G.; da Silva, A.R.P.; Walter, M.E.M.; de Macedo Brigido, M.; Yamane, T.; Lopez-Lozano, J.L.; Astolfi-Filho, S. Transcriptome analysis of the Amazonian viper Bothrops atrox venom gland using expressed sequence tags (ESTs). *Toxicon* **2009**, *53*, 427-436.
21. Zhang, B.; Liu, Q.; Yin, W.; Zhang, X.; Huang, Y.; Luo, Y.; Qiu, P.; Su, X.; Yu, J.; Hu, S.; et al. Transcriptome analysis of Deinagkistrodon acutus venomous gland focusing on cellular structure and functional aspects using expressed sequence tags. *BMC genomics* **2006**, *7*, 152. https://doi.org/10.1186/1471-2164-7-152.
22. Yang, Z.-M.; Yang, Y.-E.; Chen, Y.; Cao, J.; Zhang, C.; Liu, L.-L.; Wang, Z.-Z.; Wang, X.-M.; Wang, Y.-M.; Tsai, I.-H. Transcriptome and proteome of the highly neurotoxic venom of Gloydius intermedius. *Toxicon* **2015**, *107*, 175-186. https://doi.org/10.1016/j.toxicon.2015.08.010.
23. Tan, K.Y.; Tan, C.H.; Chanhome, L.; Tan, N.H. Comparative venom gland transcriptomics of Naja kaouthia (monocled cobra) from Malaysia and Thailand: elucidating geographical venom variation and insights into sequence novelty. *PeerJ* **2017**, *5*, e3142. https://doi.org/10.7717/peerj.3142.

24. Tan, C.H.; Tan, K.Y.; Fung, S.Y.; Tan, N.H. Venom-gland transcriptome and venom proteome of the Malaysian king cobra (Ophiophagus hannah). *BMC genomics* **2015**, *16*, 687. https://doi.org/10.1186/s12864-015-1828-2.
25. Chong, H.P.; Tan, K.Y.; Tan, N.H.; Tan, C.H. Exploring the diversity and novelty of toxin genes in Naja sumatrana, the Equatorial spitting cobra from Malaysia through de novo venom-gland transcriptomics. *Toxins* **2019**, *11*, 104.
26. Palasuberniam, P.; Tan, K.Y.; Tan, C.H. De novo venom gland transcriptomics of Calliophis bivirgata flaviceps: uncovering the complexity of toxins from the Malayan blue coral snake. *Journal of Venomous Animals and Toxins including Tropical Diseases* **2021**, *27*.
27. Tan, C.H.; Tan, K.Y. De Novo venom-gland transcriptomics of spine-bellied sea snake (Hydrophis curtus) from Penang, Malaysia—next-generation sequencing, functional annotation and toxinological correlation. *Toxins* **2021**, *13*, 127.
28. Casewell, N.R.; Wagstaff, S.C.; Wüster, W.; Cook, D.A.N.; Bolton, F.M.S.; King, S.I.; Pla, D.; Sanz, L.; Calvete, J.J.; Harrison, R.A. Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. **2014**, *111*, 9205-9210. https://doi.org/10.1073/pnas.1405484111 %J Proceedings of the National Academy of Sciences.
29. Margres, M.J.; McGivern, J.J.; Wray, K.P.; Seavy, M.; Calvin, K.; Rokyta, D.R. Linking the transcriptome and proteome to characterize the venom of the eastern diamondback rattlesnake (Crotalus adamanteus). *Journal of Proteomics* **2014**, *96*, 145-158. https://doi.org/10.1016/j.jprot.2013.11.001.
30. Bjarnason, J.B.; Fox, J.W. Hemorrhagic metalloproteinases from snake venoms. *Pharmacology & Therapeutics* **1994**, *62*, 325-372. https://doi.org/10.1016/0163-7258(94)90049-3.
31. Takeda, S.; Takeya, H.; Iwanaga, S. Snake venom metalloproteinases: Structure, function and relevance to the mammalian ADAM/ADAMTS family proteins. *Biochim Biophys Acta Proteins Proteom* **2012**, *1824*, 164-176. https://doi.org/10.1016/j.bbapap.2011.04.009.
32. Huang, T.-F.; Chang, M.-C.; Teng, C.-M. Antiplatelet protease, kistomin, selectively cleaves human platelet glycoprotein Ib. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1993**, *1158*, 293-299. https://doi.org/10.1016/0304-4165(93)90028-7.
33. Hsu, C.; Wu, W.; Huang, T. A snake venom metalloproteinase, kistomin, cleaves platelet glycoprotein VI and impairs platelet functions. *Journal of ThrombosisHaemostasis* **2008**, *6*, 1578-1585.
34. Casewell, N.R.; Wagstaff, S.C.; Harrison, R.A.; Renjifo, C.; Wüster, W. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. *Molecular biology and evolution* **2011**, *28*, 2637-2649.
35. Au, L.-C.; Huang, Y.-B.; Huang, T.-F.; Teh, G.-W.; Lin, H.-H.; Choo, K.-B. A common precursor for a putative hemorrhagic protein and rhodostomin, a platelet aggregation inhibitor of the venom of Calloselasma rhodostoma: molecular cloning and sequence analysis. *Biochemical and biophysical research communications* **1991**, *181*, 585-593.
36. Chung, M.C.M.; Ponnudurai, G.; Kataoka, M.; Shimizu, S.; Tan, N.-H. Structural Studies of a Major Hemorrhagin (Rhodostoxin) from the Venom ofCalloselasma rhodostoma (Malayan Pit Viper). *Archives of biochemistry and biophysics* **1996**, *325*, 199-208.
37. Tan, N.-H.; Ponnudurai, G.; Chung, M.C.M. Proteolytic specificity of rhodostoxin, the major hemorrhagin of Calloselasma rhodostoma (Malayan pit viper) venom. *Toxicon* **1997**, *35*, 979-984.
38. Yeh, C.-H.; Peng, H.-C.; Yang, R.-S.; Huang, T.-F. Rhodostomin, a snake venom disintegrin, inhibits angiogenesis elicited by basic fibroblast growth factor and suppresses tumor growth by a selective $\alpha v\beta 3$ blockade of endothelial cells. *Molecular pharmacology* **2001**, *59*, 1333-1342.
39. Au, L.-C.; Chou, J.-S.; Chang, K.-J.; Teh, G.-W.; Lin, S.-B. Nucleotide sequence of a full-length cDNA encoding a common precursor of platelet aggregation inhibitor and hemorrhagic protein from Calloselasma rhodostoma venom. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* **1993**, *1173*, 243-245.
40. Fujii, Y.; Okuda, D.; Fujimoto, Z.; Horii, K.; Morita, T.; Mizuno, H. Crystal structure of trimestatin, a disintegrin containing a cell adhesion recognition motif RGD. *Journal of molecular biology* **2003**, *332*, 1115-1122.
41. Ponnudurai, G.; Chung, M.C.M.; Tan, N.-H. Isolation and characterization of a hemorrhagin from the venom of Calloselasma rhodostoma (Malayan pit viper). *Toxicon* **1993**, *31*, 997-1005.
42. Tan, N.-H.; Ponnudurai, G. The toxinology of Calloselasma rhodostoma (Malayan pit viper) venom. *Journal of Toxicology: Toxin Reviews* **1996**, *15*, 1-17.
43. Gutiérrez, J.M.; Escalante, T.; Rucavado, A.; Herrera, C.; Fox, J.W. A comprehensive view of the structural and functional alterations of extracellular matrix by snake venom metalloproteinases (SVMPs): novel perspectives on the pathophysiology of envenoming. *Toxins* **2016**, *8*, 304.
44. Chang, H.-H.; Chang, C.-P.; Chang, J.-C.; Dung, S.-Z.; Lo, S.J. Application of recombinant rhodostomin in studying cell adhesion. *Journal of Biomedical Science* **1997**, *4*, 235-243.

45. Tan, C.H.; Liew, J.L.; Navanesan, S.; Sim, K.S.; Tan, N.H.; Tan, K.Y. Cytotoxic and anticancer properties of the Malaysian mangrove pit viper (Trimeresurus purpureomaculatus) venom and its disintegrin (purpureomaculin). *Journal of Venomous Animals and Toxins including Tropical Diseases* **2020**, *26*.
46. Liew, J.L.; Tan, N.H.; Tan, C.H. Proteomics and preclinical antivenom neutralization of the mangrove pit viper (Trimeresurus purpureomaculatus, Malaysia) and white-lipped pit viper (Trimeresurus albolabris, Thailand) venoms. *Acta tropica* **2020**, *209*, 105528.
47. Moura-da-Silva, A.; Almeida, M.; Portes-Junior, J.; Nicolau, C.; Gomes-Neto, F.; Valente, R. Processing of snake venom metalloproteinases: generation of toxin diversity and enzyme inactivation. *Toxins* **2016**, *8*, 183.
48. Tasoulis, T.; Isbister, G.K. A review and database of snake venom proteomes. *Toxins* **2017**, *9*, 290.
49. Esnouf, M.P.; Tunnah, G.W. The isolation and properties of the thrombin-like activity from Ancistrodon rhodostoma Venom. *British Journal of Haematology* **1967**, *13*, 581-590.
50. Shin, Y.; Morita, T. Rhodocytin, a functional novel platelet agonist belonging to the heterodimeric C-type lectin family, induces platelet aggregation independently of glycoprotein Ib. *Biochemical and biophysical research communications* **1998**, *245*, 741-745.
51. Fox, J.W.; Serrano, S.M.T. Insights into and speculations about snake venom metalloproteinase (SVMP) synthesis, folding and disulfide bond formation and their contribution to venom complexity. *The FEBS journal* **2008**, *275*, 3016-3030.
52. Kini, R.M. Excitement ahead: structure, function and mechanism of snake venom phospholipase A2 enzymes. *Toxicon* **2003**, *42*, 827-840. https://doi.org/10.1016/j.toxicon.2003.11.002.
53. Ferraz, C.R.; Arrahman, A.; Xie, C.; Casewell, N.R.; Lewis, R.J.; Kool, J.; Cardoso, F.C. Multifunctional toxins in snake venoms and therapeutic implications: From pain to hemorrhage and necrosis. *Frontiers in ecology and evolution* **2019**, *7*, 218.
54. Faisal, T.; Tan, K.Y.; Tan, N.H.; Sim, S.M.; Gnanathasan, C.A.; Tan, C.H. Proteomics, toxicity and antivenom neutralization of Sri Lankan and Indian Russell's viper (Daboia russelii) venoms. *Journal of Venomous Animals and Toxins including Tropical Diseases* **2021**, *27*.
55. Pla, D.; Sanz, L.; Quesada-Bernat, S.; Villalta, M.; Baal, J.; Chowdhury, M.A.W.; León, G.; Gutiérrez, J.M.; Kuch, U.; Calvete, J.J. Phylovenomics of Daboia russelii across the Indian subcontinent. Bioactivities and comparative in vivo neutralization and in vitro third-generation antivenomics of antivenoms against venoms from India, Bangladesh and Sri Lanka. *Journal of proteomics* **2019**, *207*, 103443.
56. Tan, C.H.; Wong, K.Y.; Huang, L.-K.; Tan, K.Y.; Tan, N.H.; Wu, W.-G. Snake venomics and antivenomics of Cape Cobra (Naja nivea) from South Africa: insights into venom toxicity and cross-neutralization activity. *Toxins* **2022**, *14*, 860.
57. Tan, K.Y.; Wong, K.Y.; Tan, N.H.; Tan, C.H. Quantitative proteomics of Naja annulifera (sub-Saharan snouted cobra) venom and neutralization activities of two antivenoms in Africa. *International journal of biological macromolecules* **2020**, *158*, 605-616.
58. Wong, K.Y.; Tan, K.Y.; Tan, N.H.; Tan, C.H. A neurotoxic snake venom without phospholipase A2: proteomics and cross-neutralization of the venom from Senegalese Cobra, Naja senegalensis (Subgenus: Uraeus). *Toxins* **2021**, *13*, 60.
59. Tan, N.-H.; Kanthimathi, M.; Tan, C.-S. Enzymatic activities of Calloselasma rhodostoma (Malayan pit viper) venom. *Toxicon* **1986**, *24*, 626-630.
60. Tan, K.Y.; Tan, C.H.; Fung, S.Y.; Tan, N.H. Venomics, lethality and neutralization of Naja kaouthia (monocled cobra) venoms from three different geographical regions of Southeast Asia. *J Proteomics* **2015**, *120*, 105-125.
61. Yap, M.K.; Fung, S.Y.; Tan, K.Y.; Tan, N.H. Proteomic characterization of venom of the medically important Southeast Asian Naja sumatrana (Equatorial spitting cobra). *Acta Trop* **2014**, *133*, 15-25. https://doi.org/10.1016/j.actatropica.2014.01.014.
62. Tan, C.H.; Tan, K.Y.; Ng, T.S.; Sim, S.M.; Tan, N.H. Venom proteome of spine-bellied sea snake (*Hydrophis curtus)* from Penang, Malaysia: Toxicity correlation, immunoprofiling and cross-neutralization by sea snake antivenom. *Toxins (Basel)* **2019**, *11*, 3. https://doi.org/10.3390/toxins11010003.
63. Tsai, I.H.; Wang, Y.M.; Au, L.C.; Ko, T.P.; Chen, Y.H.; Chu, Y.F. Phospholipases A2 from Callosellasma rhodostoma venom gland: Cloning and sequencing of 10 of the cDNAs, three-dimensional modelling and chemical modification of the major isozyme. *European journal of biochemistry* **2000**, *267*, 6684-6691.
64. Tsai, I.-H.; Wang, Y.-M.; Chen, Y.-H.; Tsai, T.-S.; Tu, M.-C. Venom phospholipases A2 of bamboo viper (Trimeresurus stejnegeri): molecular characterization, geographic variations and evidence of multiple ancestries. *Biochemical Journal* **2004**, *377*, 215-223.
65. Van Den Bergh, C.J.; Slotboom, A.J.; Verheij, H.M.; De Haas, G.H. The role of aspartic acid-49 in the active site of phospholipase A2: A site-specific mutagenesis study of porcine pancreatic phospholipase A2 and the rationale of the enzymatic activity of [Iysine49] phospholipase A2 from Agkistrodon piscivorus piscivorus venom. *European journal of biochemistry* **1988**, *176*, 353-357.
66. Lomonte, B. Lys49 myotoxins, secreted phospholipase A2-like proteins of viperid venoms: A comprehensive review. *Toxicon* **2023**, 107024.

67. De Bold, A.J.; Borenstein, H.B.; Veress, A.T.; Sonnenberg, H. A rapid and potent natriuretic response to intravenous injection of atrial myocardial extract in rats. *Life sciences* **1981**, *28*, 89-94.
68. Vink, S.; Jin, A.H.; Poth, K.J.; Head, G.A.; Alewood, P.F. Natriuretic peptide drug leads from snake venom. *Toxicon* **2012**, *59*, 434-445.
69. Pimenta, D.C.; Spencer, P.J. Bradykinin-potentiating and related peptides from reptile venoms. In *Handbook of Venoms and Toxins of Reptiles*; CRC Press: 2021; pp. 241-250.
70. Lameu, C.; Neiva, M.; Hayashi, M. Venom Bradykinin-related peptides (BRPs) and its multiple biological roles. *An Integrated View of the Molecular Recognition and Toxinology—From Analytical Procedures to Biomedical Applications* **2013**, 119-151.
71. Higuchi, S.; Murayama, N.; Saguchi, K.; Ohi, H.; Fujita, Y.; Camargo, A.C.; Ogawa, T.; Deshimaru, M.; Ohno, M. Bradykinin-potentiating peptides and C-type natriuretic peptides from snake venom. *Immunopharmacology* **1999**, *44*, 129-135.
72. Hayashi, M.A.; Camargo, A.C. The Bradykinin-potentiating peptides from venom gland and brain of Bothrops jararaca contain highly site specific inhibitors of the somatic angiotensin-converting enzyme. *Toxicon* **2005**, *45*, 1163-1170.
73. Michel, G.H.; Murayama, N.; Sada, T.; Nozaki, M.; Saguchi, K.; Ohi, H.; Fujita, Y.; Koike, H.; Higuchi, S. Two N-terminally truncated forms of C-type natriuretic peptide from habu snake venom. *Peptides* **2000**, *21*, 609-615.
74. Fucase, T.M.; Sciani, J.M.; Cavalcante, I.; Viala, V.L.; Chagas, B.B.; Pimenta, D.C.; Spencer, P.J. Isolation and biochemical characterization of bradykinin-potentiating peptides from Bitis gabonica rhinoceros. *Journal of Venomous Animals and Toxins including Tropical Diseases* **2018**, *23*.
75. Ianzer, D.; Konno, K.; Marques-Porto, R.; Portaro, F.C.V.; Stöcklin, R.; de Camargo, A.C.M.; Pimenta, D.C. Identification of five new bradykinin potentiating peptides (BPPs) from Bothrops jararaca crude venom by using electrospray ionization tandem mass spectrometry after a two-step liquid chromatography. *Peptides* **2004**, *25*, 1085-1092.
76. Cheung, H.; Cushman, D. Inhibition of homogeneous angiotensin-converting enzyme of rabbit lung by synthetic venom peptides of Bothrops jararaca. *Biochimica et Biophysica Acta (BBA)-Enzymology* **1973**, *293*, 451-463.
77. Morita, T. Structures and functions of snake venom CLPs (C-type lectin-like proteins) with anticoagulant-, procoagulant-, and platelet-modulating activities. *Toxicon* **2005**, *45*, 1099-1114.
78. Arlinghaus, F.T.; Eble, J.A. C-type lectin-like proteins from snake venoms. *Toxicon* **2012**, *60*, 512-519. https://doi.org/10.1016/j.toxicon.2012.03.001.
79. Clemetson, K.; Morita, T.; Kini, R.M. Classification and nomenclature of snake venom C-type lectins and related proteins. *Toxicon: official journal of the International Society on Toxinology* **2009**, *54*, 83.
80. Watson, A.A.; O'callaghan, C.A. Molecular analysis of the interaction of the snake venom rhodocytin with the platelet receptor CLEC-2. *Toxins* **2011**, *3*, 991-1003.
81. Suzuki-Inoue, K.; Fuller, G.L.; García, Á.; Eble, J.A.; Pöhlmann, S.; Inoue, O.; Gartner, T.K.; Hughan, S.C.; Pearce, A.C.; Laing, G.D. A novel Syk-dependent mechanism of platelet activation by the C-type lectin receptor CLEC-2. *Blood* **2006**, *107*, 542-549.
82. Wang, Y.M.; Liew, Y.F.; Chang, K.Y.; Tsai, I.H. Purification and characterization of the venom phospholipases A2 from Asian monotypic crotalinae snakes. *Journal of Natural Toxins* **1999**, *8*, 331-340.
83. Eble, J.A.; Niland, S.; Bracht, T.; Mormann, M.; Peter-Katalinic, J.; Pohlentz, G.; Stetefeld, J. The $\alpha 2\beta 1$ integrin-specific antagonist rhodocetin is a cruciform, heterotetrameric molecule. *The FASEB Journal* **2009**, *23*, 2917-2927.
84. Eble, J.A.; Beermann, B.; Hinz, H.-J.r.; Schmidt-Hederich, A. $\alpha 2\beta 1$ integrin is not recognized by rhodocytin but is the specific, high affinity target of rhodocetin, an RGD-independent disintegrin and potent inhibitor of cell adhesion to collagen. *Journal of Biological Chemistry* **2001**, *276*, 12274-12284.
85. Sartim, M.A.; Sampaio, S.V. Snake venom galactoside-binding lectins: a structural and functional overview. *Journal of Venomous Animals and Toxins including Tropical Diseases* **2015**, *21*, 1-11.
86. Swenson, S.D.; Stack, S.; Markl, F.S. Thrombin-Like Serine Proteinases in Reptile Venoms. In *Handbook of Venoms and Toxins of Reptiles*; CRC Press: 2021; pp. 351-362.
87. Burkhart, W.; Smith, G.F.; Su, J.-L.; Parikh, I.; LeVine, H. Amino acid sequence determination of Ancrod, the thrombin-like α-fibrinogenase from the venom of Akistrodon rhodostoma. *FEBS letters* **1992**, *297*, 297-301.
88. Rawlings, N.D.; Barrett, A.J. *Serine peptidases and their clans. In Handbook of Proteolytic Enzymes, 2nd edition, edited by N.D. Rawlings, A.J. Barrett, J. F. Wossner. San Diego, CA: Academic Press Ltd., pp. 1417–39.*; 2004.
89. Pfeiffer, G.; Linder, D.; Strube, K.; Geyer, R. Glycosylation of the thrombin-like serine protease ancrod fromAgkistrodon rhodostoma venom. Oligosaccharide substitution pattern at eachN-glycosylation site. *Glycoconjugate journal* **1993**, *10*, 240-246.
90. Roth, J.; Zuber, C.; Park, S.; Jang, I.; Lee, Y.; Kysela, K.G.; Le Fourn, V.; Santimaria, R.; Guhl, B.; Cho, J.W. Protein N-glycosylation, protein folding, and protein quality control. *Molecules and cells* **2010**, *30*, 497-506.

91. Lin, C.-W.; Chen, J.-M.; Wang, Y.-M.; Wu, S.-W.; Tsai, I.-H.; Khoo, K.-H. Terminal disialylated multiantennary complex-type N-glycans carried on acutobin define the glycosylation characteristics of the Deinagkistrodon acutus venom. *Glycobiology* **2011**, *21*, 530-542.
92. Wormald, M.R.; Dwek, R.A. Glycoproteins: glycan presentation and protein-fold stability. *Structure* **1999**, *7*, R155-R160.
93. Lannoo, N.; Van Damme, E.J. Review/N-glycans: The making of a varied toolbox. *Plant Science* **2015**, *239*, 67-83.
94. Hennerici, M.G.; Kay, R.; Bogousslavsky, J.; Lenzi, G.L.; Verstraete, M.; Orgogozo, J.M. Intravenous ancrod for acute ischaemic stroke in the European Stroke Treatment with Ancrod Trial: a randomised controlled trial. *The Lancet* **2006**, *368*, 1871-1878.
95. Sherman, D.G.; Atkinson, R.P.; Chippendale, T.; Levin, K.A.; Ng, K.; Futrell, N.; Hsu, C.Y.; Levy, D.E. Intravenous ancrod for treatment of acute ischemic stroke: the STAT study: a randomized controlled trial. *Jama* **2000**, *283*, 2395-2403.
96. Maduwage, K.; Isbister, G.K. Current treatment for venom-induced consumption coagulopathy resulting from snakebite. *PLoS neglected tropical diseases* **2014**, *8*, e3220.
97. Li, Z.-Y.; Yu, T.-F.; Lian, E.C.Y. Purification and characterization of l-amino acid oxidase from king cobra (Ophiophagus hannah) venom and its effects on human platelet aggregation. *Toxicon* **1994**, *32*, 1349-1358. https://doi.org/10.1016/0041-0101(94)90407-3.
98. Samel, M.; Tõnismägi, K.; Rönnholm, G.; Vija, H.; Siigur, J.; Kalkkinen, N.; Siigur, E. l-Amino acid oxidase from Naja naja oxiana venom. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **2008**, *149*, 572-580. https://doi.org/10.1016/j.cbpb.2007.11.008.
99. Moustafa, I.M.; Foster, S.; Lyubimov, A.Y.; Vrielink, A. Crystal structure of LAAO from Calloselasma rhodostoma with an L-phenylalanine substrate: insights into structure and mechanism. *Journal of molecular biology* **2006**, *364*, 991-1002.
100. Paloschi, M.V.; Boeno, C.N.; Lopes, J.A.; Rego, C.M.A.; Silva, M.D.S.; Santana, H.M.; Serrath, S.N.; Ikenohuchi, Y.J.; Farias, B.J.C.; Felipin, K.P. Reactive oxygen species-dependent-NLRP3 inflammasome activation in human neutrophils induced by l-amino acid oxidase derived from Calloselasma rhodostoma venom. *Life Sciences* **2022**, *308*, 120962.
101. Lee, M.L.; Tan, N.H.; Fung, S.Y.; Sekaran, S.D. Antibacterial action of a heat-stable form of L-amino acid oxidase isolated from king cobra (Ophiophagus hannah) venom. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* **2011**, *153*, 237-242.
102. Yamazaki, Y.; Morita, T. Structure and function of snake venom cysteine-rich secretory proteins. *Toxicon* **2004**, *44*, 227-231. https://doi.org/10.1016/j.toxicon.2004.05.023.
103. Dhananjaya, B.L.; D'souza, C.J.M. An overview on nucleases (DNase, RNase, and phosphodiesterase) in snake venoms. *Biochemistry (Moscow)* **2010**, *75*, 1-6. https://doi.org/10.1134/s0006297910010013.
104. Coronado, M.A.; da Silva Olivier, D.; Eberle, R.J.; do Amaral, M.S.; Arni, R.K. Modeling and molecular dynamics indicate that snake venom phospholipase B-like enzymes are Ntn-hydrolases. *Toxicon* **2018**.
105. Takasaki, C.; Tamiya, N. Isolation and properties of lysophospholipases from the venom of an Australian elapid snake, Pseudechis australis. *Biochemical Journal* **1982**, *203*, 269-276.
106. Bernheimer, A.; Weinstein, S.; Linder, R. Isoelectric analysis of some Australian elapid snake venoms with special reference to phospholipase B and hemolysis. *Toxicon* **1986**, *24*, 841-849.
107. Bernheimer, A.; Linder, R.; Weinstein, S.; Kim, K.-S. Isolation and characterization of a phospholipase B from venom of Collett's snake, Pseudechis colletti. *Toxicon* **1987**, *25*, 547-554.
108. Lavin, M.; Earl, S.; Birrel, G.; St Pierre, L.; Guddat, L.; de Jersey, J.; Masci, P. Snake venom nerve growth factors. In *Handbook of Venoms and Toxins of Reptiles*, SP, M., Ed.; Taylor and Francis Group, CRC Press: Boca Raton, 2009.
109. Sunagar, K.; Fry, B.G.; Jackson, T.N.; Casewell, N.R.; Undheim, E.A.; Vidal, N.; Ali, S.A.; King, G.F.; Vasudevan, K.; Vasconcelos, V. Molecular evolution of vertebrate neurotrophins: co-option of the highly conserved nerve growth factor gene into the advanced snake venom arsenalf. *PloS one* **2013**, *8*, e81827.
110. Vaiyapuri, S.; Wagstaff, S.C.; Watson, K.A.; Harrison, R.A.; Gibbins, J.M.; Hutchinson, E.G. Purification and functional characterisation of rhiminopeptidase A, a novel aminopeptidase from the venom of Bitis gabonica rhinoceros. *PLoS neglected tropical diseases* **2010**, *4*, e796.
111. Nicolau, C.A.; Carvalho, P.C.; Junqueira-de-Azevedo, I.L.M.; Teixeira-Ferreira, A.; Junqueira, M.; Perales, J.; Neves-Ferreira, A.G.C.; Valente, R.H. An in-depth snake venom proteopeptidome characterization: Benchmarking Bothrops jararaca. *Journal of proteomics* **2017**, *151*, 214-231.
112. 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. *Journal of Venomous Animals and Toxins including Tropical Diseases* **2020**, *26*.
113. Lingam, T.M.C.; Tan, K.Y.; Tan, C.H. Capillary leak syndrome induced by the venoms of Russell's Vipers (Daboia russelii and Daboia siamensis) from eight locales and neutralization of the differential toxicity by

three snake antivenoms. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* **2021**, *250*, 109186.
114. Rucavado, A.; Escalante, T.; Camacho, E.; Gutiérrez, J.M.; Fox, J.W. Systemic vascular leakage induced in mice by Russell's viper venom from Pakistan. *Scientific Reports* **2018**, *8*, 16088.
115. Pawlak, J.; Mackessy, S.P.; Fry, B.G.; Bhatia, M.; Mourier, G.; Fruchart-Gaillard, C.; Servent, D.; Ménez, R.; Stura, E.; Ménez, A. Denmotoxin, a three-finger toxin from the colubrid snake Boiga dendrophila (Mangrove Catsnake) with bird-specific activity. *Journal of Biological Chemistry* **2006**, *281*, 29030-29041.
116. Pahari, S.; Mackessy, S.P.; Kini, R.M. The venom gland transcriptome of the Desert Massasauga Rattlesnake (Sistrurus catenatus edwardsii): towards an understanding of venom composition among advanced snakes (Superfamily Colubroidea). *BMC molecular biology* **2007**, *8*, 115.
117. Mukherjee, A.K.; Mackessy, S.P.; Dutta, S. Characterization of a Kunitz-type protease inhibitor peptide (Rusvikunin) purified from Daboia russelii russelii venom. *International journal of biological macromolecules* **2014**, *67*, 154-162.
118. Rotenberg, D.; Bamberger, E.; Kochva, E. Studies on ribonucleic acid synthesis in the venom glands of Vipera palaestinae (Ophidia, Reptilia). *Biochemical Journal* **1971**, *121*, 609-612.
119. Wery, M.; Descrimes, M.; Thermes, C.; Gautheret, D.; Morillon, A. Zinc-mediated RNA fragmentation allows robust transcript reassembly upon whole transcriptome RNA-Seq. *Methods* **2013**, *63*, 25-31.
120. Conesa, A.; Madrigal, P.; Tarazona, S.; Gomez-Cabrero, D.; Cervera, A.; McPherson, A.; Szcześniak, M.W.; Gaffney, D.J.; Elo, L.L.; Zhang, X. A survey of best practices for RNA-seq data analysis. *Genome biology* **2016**, *17*, 1-19.
121. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **2011**, *29*, 644-652. https://doi.org/10.1038/nbt.1883.
122. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* **2013**, *8*, 1494-1512.
123. Pertea, G.; Huang, X.; Liang, F.; Antonescu, V.; Sultana, R.; Karamycheva, S.; Lee, Y.; White, J.; Cheung, F.; Parvizi, B. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics (Oxford, England)* **2003**, *19*, 651-652.
124. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **2012**, *9*, 357-359.
125. Li, B.; Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **2011**, *12*, 1-16.
126. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)* **2009**, *25*, 1189-1191. https://doi.org/10.1093/bioinformatics/btp033.
127. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **2004**, *32*, 1792-1797. https://doi.org/10.1093/nar/gkh340.