UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE BIOCIÊNCIAS

Gabriel Gonzalez Sonoda

Composição e evolução do veneno de serpentes da tribo Hydropsini

Composition and evolution of venom in snakes from the tribe Hydropsini

SÃO PAULO
2024

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE BIOCIÊNCIAS

Gabriel Gonzalez Sonoda

Composição e evolução do veneno de serpentes da tribo Hydropsini

Composition and evolution of venom in snakes from the tribe Hydropsini

Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Mestre em Ciências, na Área de Genética e Biologia Evolutiva.

Orientador(a): Dr Inácio de Loiola Meirelles Junqueira de Azevedo

EXEMPLAR CORRIGIDO

SÃO PAULO
2024

Sonoda, Gabriel Gonzalez
Composição e evolução do veneno de serpentes da tribo Hydropsini
Número de páginas: 115

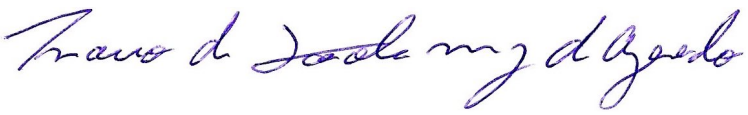
Dissertação (Mestrado) - Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva.

1. Toxinas 2. Lectinas do Tipo C 3.Helicops I. Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva.

Comissão Julgadora:

_____ _____

Prof(a). Dr(a). Prof(a). Dr(a).



_____

Prof(a). Dr.(a). Inácio de Loiola Meirelles Junqueira de Azevedo

Orientador(a)

Dedico este trabalho a todas as crianças que fazem perguntas complicadas sobre biologia. Que a alma de cientista e o pensamento crítico dirija estes pequeninos pesquisadores rumo aos seus sonhos.

“Mas o que salva a humanidade
É que não há quem cure a curiosidade”
Tom Zé, *Salva a Humanidade,* (2014)

“A cor do céu me compõe
Mar azul me dissolve
A equação me propõe
Computador me resolve”
- Os Mutantes , *2001 (Dois Mil E Um),* (1969)

Agradecimentos

Embora eu assine como único autor da dissertação, o trabalho apresentado teve a participação de muitos, pois a ciência se desenvolve de forma coletiva. Agradeço a todos que de alguma forma contribuíram para que a pesquisa pudesse ser realizada. Entre estes, ressalto meus agradecimentos:

Ao meu Orientador, Inácio de Loiola Meirelles Junqueira de Azevedo, por ter me aceitado em seu laboratório como aluno e ter oferecido uma orientação que explorou meus potenciais.

Ao Vinicius Carius de Souza, por toda ajuda na parte de Bioinformática, administração do servidor, desenvolvimento do algoritmo apresentado e modelagens 3D.

Ao Vincent Louis Viala, por todos ensinamentos e ajudas na parte de sequenciamento de transcriptomas por Nanopore. Da seleção de amostras até a montagem dos transcritos finais.

Ao Juan David Bayona Serrano e ao Pedro Gabriel Nacthigall, pelas ajudas na interpretação dos resultados de transcriptômica e evolução.

À equipe das coleções biológicas do Instituto Butantan, supervisionada pelo Felipe Gobbi Grazziotin, pela coleta das serpentes, preparação das amostras e organização da coleção. Ainda, à todos aqueles que se aventuraram em trabalhos de campo comigo buscando por *Helicops* (Weverton Azevedo, Julia Mayumi, Lucas Neves, Omar Entiauspe-Neto, Rafael Gusson e Eletra de Souza) e ajudaram com a manutenção dos animais coletados (Guilherme Oliveira e Larissa Silva).

Também agradeço ao Weverton Azevedo pela disponibilização das fotos usadas nesta dissertação.

À Solange Maria de Toledo Serrano e seus alunos e pósdocs: Jorge Moschem, Alison Chaves e Bianca Barros pela ajuda com os experimentos de SDS-PAGE, digestão *in gel* e proteomica.

À Luciana Freitas de Souza pela ajuda com a interpretação dos resultados de proteomica *shotgun.*

Às técnicas de Biologia Molecular Nancy da Rós e Mariana Morone por toda a ajuda com a parte de biomol, sobretudo na preparação das bibliotecas de Illumina e na localização dos reagentes.

À equipe administrativa do LETA, por toda ajuda com documentação e prestação de contas.

Aos demais pesquisadores do LETA, por terem me acompanhado em tantos experimentos.

Aos membros do Dojo Bioinfo do Instituto Butantan, pelas oportunidades de discussões acadêmicas na área de Bioinformática.

Ao comitê de acompanhamento, (Andrew Mason, Taran Grant e Sónia C.S. Andrade) pela atenção, críticas e sugestões, que foram fundamentais para o desenvolvimento da dissertação.
Aos meus antigos colegas, e hoje amigos, membros do LDG, por terem contribuído na minha formação como cientista no passado e até promoverem encontros bem humorados no presente.
Aos amigos que fiz e carrego fora da minha vida acadêmica. Obrigado por estarem ao meu lado nos momentos em que eu preciso. Pelos momentos de risada e papos absurdos.
À minha namorada, Larissa, por ter me acompanhado ao longo desses dois anos de Mestrado, me ouvindo e me ajudando sempre. Agradeço pela paciência com os surtos relativos ao (ou resultantes do) presente trabalho.
À minha família, principalmente pai, mãe e irmã, por todo apoio que tive durante o desenvolvimento deste Mestrado. Obrigado por acreditarem em mim e me incentivarem a buscar minhas ambições.
Ainda, agradeço à FAPESP (2021/10174-2 e 2016/50127-5) pelo financiamento da pesquisa. O fomento desta fundação é fundamental para o desenvolvimento da ciência, incluindo este Mestrado.

1. Introduction

Toxins are substances synthesized by organisms that cause deleterious effects to other organisms exposed to it. Animals with mechanisms to inoculate a mixture of said toxins are classified as venomous, and the mixtures are referred to as venom (WEXLER *et al.*, 2015). Research on venoms and toxins has contributed to many knowledge areas, such as the advances in Physiology stemming from the discovery of the Bradykinin/Kallikrein system (E SILVA *et al.*, 1949); in Pharmacognosy by the development of Captopril (KOCH-WESER *et al.*, 1982; MOHAMED ABD EL-AZIZ *et al.*, 2019); in Ecology, providing case studies of co-evolution (BRODIE; BRODIE, 1991; DAVIES; ARBUCKLE, 2019; PORAN *et al.*, 1987); in Evolutionary Biology, providing a framework to study molecular evolutionary mechanisms (DUDA; PALUMBI, 1999; LYNCH, 2007; ROKYTA *et al.*, 2011; SUNAGAR *et al.*, 2013; ZHANG *et al.*, 2015).

Venoms have evolved independently at least 101 times in metazoan phylogeny, usually acting in predation and defense (SCHENDEL *et al.*, 2019). Despite the diversity of venomous lineages, most studies focus on animals of medical interest, such as snakes, spiders and scorpions (VON REUMONT *et al.*, 2014). Snakebites have been notorious for their effects since the ancient Egypt, as the 2300 years old Brooklyn Papyrus depicts (GOLDING, 2020). It was no coincidence that the first modern study on metazoan venoms was with snakes (HAWGOOD, 1995). The anthropocentric interest in the pathology of snakebites and the urge to develop treatments led to numerous research on snake venom. However, the same motivations that made snakes one of the most studied taxa from a toxinological perspective also resulted in knowledge bias within the group, as most studies investigate the families often involved in accidents with humans: Atractaspididae, Viperidae and Elapidae (JUNQUEIRA-DE-AZEVEDO *et al.*, 2016). Although such bias is not detrimental to the development of ophidism treatments, it hinders the understanding on snake venoms, as these families roughly add up to 800 species, less than 25% of Caenophidia, the taxa containing most venomous snakes (FRY *et al.*, 2008).

Conversely, the family Dipsadidae, which comprises more than 800 species (UETZ *et al.*, 2021), many of which present venom glands, was far less studied from a

toxinological perspective. Dipsadidae snakes were formerly positioned within the Colubridae family, but currently, many authors consider it a family itself, along with the family Natricidae (GRAZZIOTIN *et al.*, 2012; ZAHER *et al.*, 2018, 2019). Recent studies group Dipsadidae, Natricidae and Colubridae within the superfamily Colubroidea, sister taxa to Elapoidea, which comprises the families Elapiadae, Atractaspiddidae and Lamprophiidae (ZAHER *et al.*, 2019) (Figure 1).

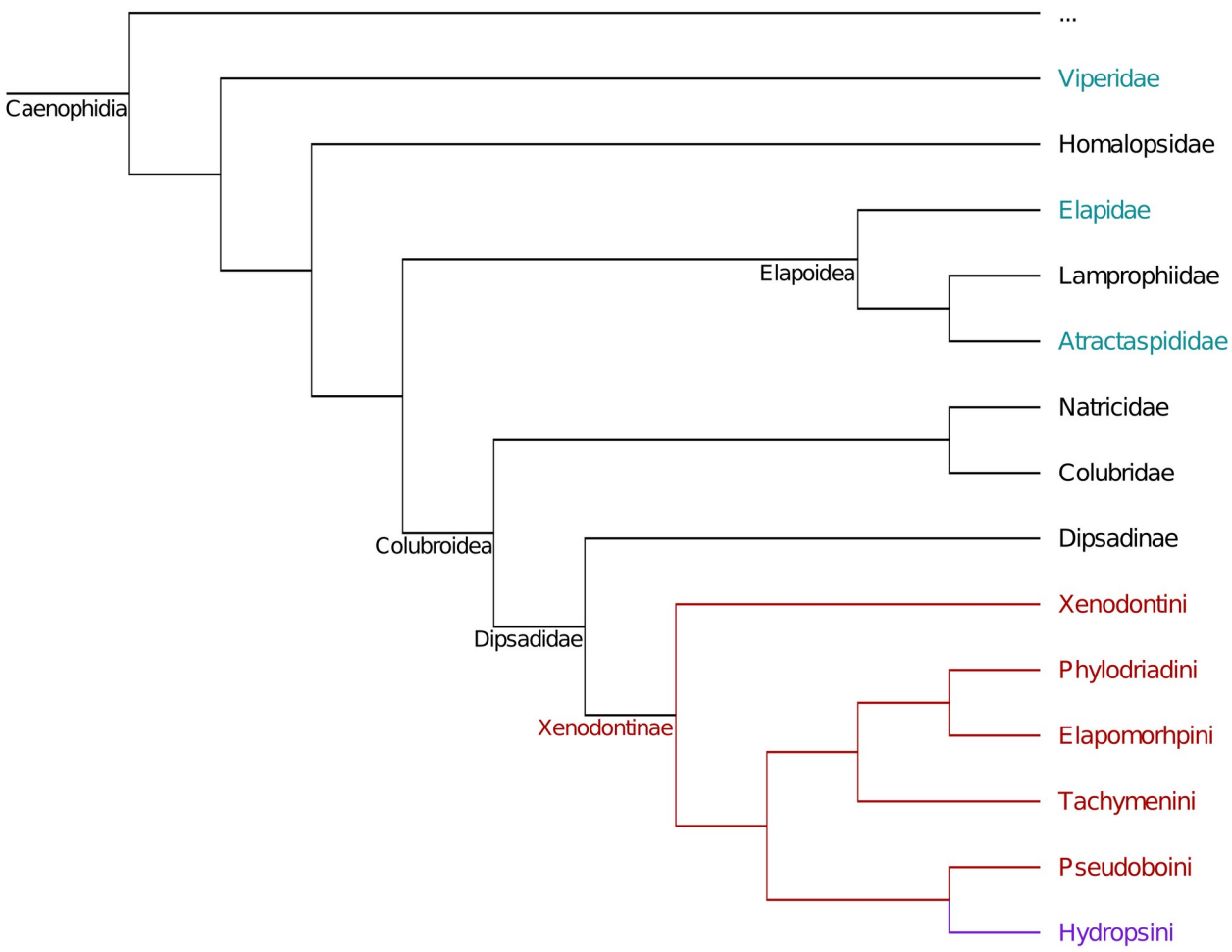


Figure 1: Simplified phylogeny of Caenophidia. In purple are the Hydropsini, the main focus of this work. Remaining Xenodontinae are represented in red. Taxa in Cyan are the three most studies families from the toxinological perspective.

Recently, thanks to technological advances, Omic studies allowed a better understanding of Dipsadidae venom (JUNQUEIRA-DE-AZEVEDO *et al.*, 2016). These studies showed that many toxin classes produced by Dipsadidae are homologous to those found in Viperidae and Elapidae (JUNQUEIRA-DE-AZEVEDO *et al.*, 2016), even though there is evidence of parallel recruitment of some toxin genes (BAYONA-SERRANO *et al.*, 2020;

JUNQUEIRA-DE-AZEVEDO *et al.*, 2016). Moreover, recent studies have shown expressive venom variations among the tribes of Dipsadidae, with each tribe presenting unique features.

Among these features are, for example, the unique venom phenotype of *Phalotris mertensi* (CAMPOS *et al.*, 2016); the convergent recruitment of different metalloproteases genes to play similar roles in Philodryadini, Conophini, Tachymenini and Xenodontini (BAYONA-SERRANO *et al.*, 2020; SCHRAMER *et al.*, 2022; TIOYAMA *et al.*, 2023); the rapid evolution of venom phenotypes, demonstrated by highly divergent venom profiles in closely related species within the Philodryadini tribe (TIOYAMA *et al.*, 2023). Additionally, studies on Dipsadidae venom have provided a more global perspective on the recruitment of toxin genes across snakes' phylogeny (BAYONA-SERRANO *et al.*, 2023). Surely, more studies are required to better understand the venom of Dipsadidae, as these few, but detailed studies only focus on a small portion of the diversity within the family.

Still, from these and a few other studies, it was possible to obtain in the recent years an overview of Dipsadidae venom composition. As in front-fanged snakes, Dipsadidae present high levels of Snake Venom Metaloproteases (SVMPs), Kunitz, C-type Lectins (CTL) and Cysteine Rich Secretory Proteins (CRISPs) (JUNQUEIRA-DE-AZEVEDO *et al.*, 2016). On the other hand, other venom components abundant in front-fanged snakes, such as three-finger toxins (3FTx), L-Amino Acid Oxidase (LAAO) and Snake Venom Serine Protease (SVSP) are not so common in Dipsadidae (JUNQUEIRA-DE-AZEVEDO *et al.*, 2016). Additionally, these studies strongly suggest high heterogeneity in the venom composition among the species. For instance, the Tachymenini *Mesotes strigatus* presents a venom rich in svMMP (CHING *et al.*, 2012), whereas the Elapomorphini *Phalotris mertensii* presents a venom uniquely rich in Kunitz. In some cases, such heterogeneity is within tribes, as in Philodryadini (TIOYAMA *et al.*, 2023). This suggests a potential undiscovered diversity that needs additional sampling of the Dipsadidae to be better understood.

In this context, a group worthy of toxinological investigations is the Hydropsini, a tribe of Dipsadidae divided into three genera: *Pseudoeryx*, *Hydrops*, and *Helicops*, the last one being the most diverse and geographically spread (Figure 1). These snakes present many adaptations to the freshwater habitat. For example, their eyes and nostrils are dorsally positioned (SCARTOZZONI, 2005), a possible adaptation that makes them less

conspicuous to non-aquatic predators, as it requires a smaller portion of the head to be emerged while breathing (GREENE, 1997).

To date, all Hydropsini species that had their diet investigated predominantly feed on fish (DE AGUIAR; DI-BERNARDO, 2004; DE CARVALHO TEIXEIRA *et al.*, 2017; SCARTOZZONI, 2009), the only tribe in the Xenodontinae sub-family to present such dietary habit. Adaptations to such a diet are also present in this tribe. Snakes of the *Helicops* genus present a wide head, possibly allowing for the ingestion of bulkier fishes (SCARTOZZONI, 2005). Moreover, curved teeth were found in the species *Helicops modestus* (OLIVEIRA *et al.*, 2016), a trait often associated with piscivory. The larger rear fangs of the species also present ridges, possibly aiding in the grip of slippery prey, as fishes (OLIVEIRA *et al.*, 2016). In these snakes, venom is thought to be delivered through a groove between the paired rear fangs on each side of the mouth (OLIVEIRA *et al.*, 2016).

As snakes' venom composition is tightly linked to their diet (DAVIES; ARBUCKLE, 2019; HEALY *et al.*, 2019; LYONS *et al.*, 2020) it is expected that a fish centered diet may have driven the evolution of venom in the tribe as well. *In vivo* experiments revealed neurotoxicity of this venom in both fishes (ALBOLEA *et al.*, 2000) and mice (ALBOLEA *et al.*, 2000; ESTRELLA *et al.*, 2009, 2011). No evidence of hemostatic disturbance was observed in mice (ESTRELLA *et al.*, 2011). Meanwhile, in humans, Hydropsini bites were reported to cause mild inflammation, increased clotting time (ALBOLEA *et al.*, 2000; VILLCA-CORANI *et al.*, 2021), and sometimes sweating, vomiting, headache and pain (SILVA *et al.*, 2019; VILLCA-CORANI *et al.*, 2021), with very weak evidence for neurological symptoms (VILLCA-CORANI *et al.*, 2021). The venom of *Helicops angulatus* was shown to contain a CRISP named Helicopsin, associated with neurotoxic effects (ESTRELLA *et al.*, 2011).

However, it was not until the time this project started that some knowledge about the toxin repertoire of the group became available. In 2022, a transcriptomic investigation involving some species of *Helicops* genus demonstrated that the CTL toxins are highly expressed in the venom glands of *Helicops leopardinus, Helicops* angulatus and *Helicops polylepis* (CERDA, 2023). In line with that, XIE et al., 2022 described CTL sequences with unique insertions in *Helicops leopardinus*.

This work aimed to understand the composition and evolution of the venom of Hydropsini snakes. To achieve this, proteo-transcriptomic approaches were used to describe the composition of these venoms, as well as the novelties present in their proteinaceous toxins at sequence level. Using transcriptomics, robust data on the expression of toxins in the venom gland of Hydropsini was obtained. Proteomic experiments were employed to validate these results by analyzing the venom of these snakes.

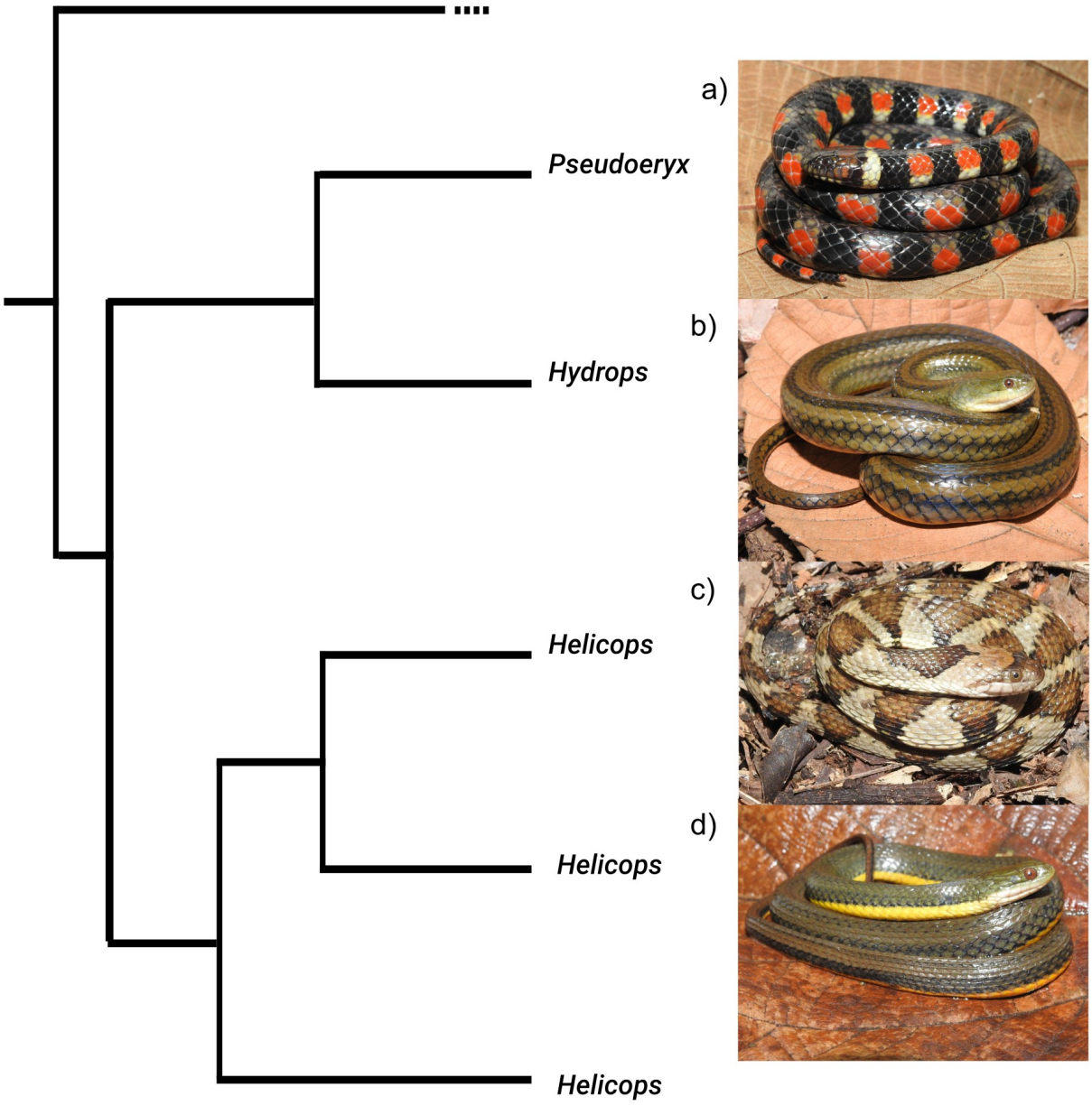


Figure 2: Simplified phylogeny of the Hydropsini tribe. a) Hydrops triangularis; b) *Helicops modestus*; c) *Helicops angulatus*; d) *Helicops carinicaudus.* Photos by Weverton Azevedo.

2. Objectives

General Objective:

To describe the venom composition of different Hydropsini species using different omics approaches and to understand the evolution of the most expressed toxins in snakes from this tribe. Furthermore, to investigate the association between the evolution of venom and feeding biology of this tribe.

Specific Objectives:

- To assemble Hydropsini venom gland transcriptomes in order to obtain highly supported putative toxin coding sequences.
- To elucidate the venom composition of the tribe Hydropsini, comparing its venom profile to closely related tribes.
- To investigate the molecular evolution of toxin sequences in Hydropsini, with a special focus on the novelties present in CTLs.
- To discuss possible drivers of the evolution of Hydropsini venom, including their diet.

3. Materials and Methods

3.1 Biological Sampling

Species used in this research were collected in nature from various localities in Brazil (ICMBio License number 57585) before and during the development of this work. From the three Hydropsini genera, *Helicops* and *Hydrops* were sampled (Table 1). The genus *Pseudoeryx* was not included in this study as it could not be found in the wild. Twenty-two individuals from seven different species belonging to the tribe Hydropsini were used for RNA-seq experiments of this study and eight individuals were included in proteomic experiments (Table 2). From the seven species sampled, six belonged to genus *Helicops*, which currently contains 20 species (UETZ *et al.*, 2021), and one belonged to the genus *Hydrops*, which contains 3 species((UETZ *et al.*, 2021)). In addition, 25 individuals from Tachymenini, Philodryadini and Pseudoboini, tribes phylogenetically close to Hydropsini according to recent studies (GRAZZIOTIN e*t al.,* 2012; ZAHER e*t al.,* 2009, 2018, 2019), were included as outgroups. Also, other five species were included for calibration of the phylogenetic tree (Table 2)

Venom was extracted following previous studies (MACKESSY *et al.*, 2006; MACKESSY; BAXTER, 2006) and stored at -80ºC. Four days after venom extraction, the animals were euthanized, and their venom glands were excised and stored in RNAlater (Invitrogen, USA). Venom extraction and euthanasia were in accordance with the Instituto Butantan ethics committee (CEUA No.4479020217).

In addition to the collected specimens, additional raw reads from *Philodryas nattereri* (PRJNA565388), an*d Psammophis schokari* (PRJNA668356) venom gland transcriptomes were downloaded from SRA database for use in the phylogenetic analyses. Table 2 summarizes the individuals used.

Table 1: Information of collected Hydropsini samples.

Code	Genus	Species	Sex	Age	Locality
SB1259	*Hydrops*	*triangularis*	M	Adult	Serrano do Maranhão
SB1946	*Hydrops*	*triangularis*	M	Juvenile	Mâncio Lima
SB1439	*Hydrops*	*triangularis*	N.D.	N.D.	Cruzeiro do Sul
SB0156	*Helicops*	*polylepis*	M	N.D.	Belterra
SB0155	*Helicops*	*polylepis*	M	N.D.	Belterra
SB0306	*Helicops*	*modestus*	M	N.D.	Alumínio
SB1347	*Helicops*	*modestus*	F	Adult	São Lourenço da Serra
SB1930	*Helicops*	*modestus*	F	Adult	Pinhalzinho
SB1164	*Helicops*	*leopardinus*	F	Adult	Corumbá
SB0544	*Helicops*	*leopardinus*	F	Adult	Corumbá
SB0124	*Helicops*	*infrataeniatus*	F	Adult	Bom Retiro
SB0125	*Helicops*	*infrataeniatus*	M	Adult	Bom Retiro
SB0162	*Helicops*	*infrataeniatus*	F	N.D.	Viamão
SB1845	*Helicops*	*infrataeniatus*	F	Adult	Álvaro de Carvalho
SB0141	*Helicops*	*carinicaudus*	M	Adult	Itajaí
SB0353	*Helicops*	*carinicaudus*	F	N.D.	São João da Barra
SB0354	*Helicops*	*carinicaudus*	M	N.D.	Biguaçu
SB0641	*Helicops*	*carinicaudus*	F	N.D.	Itajaí
SB0561	*Helicops*	*angulatus*	F	Adult	Sena Madureira
SB1256	*Helicops*	*angulatus*	F	Adult	Serrano do Maranhão
SB0795	*Helicops*	*angulatus*	F	Adult	Mâncio Lima
SB1945	*Helicops*	*angulatus*	F	Juvenile	Mâncio Lima

Table 2: Individuals used for each experiment. Note: For ONT Sequencing, X^1 = acquired at 260 bp/s; X^2 = acquired at 400 bp/s; S= shotgun; G=SDS-PAGE. Entries in gray correspond to outgroups. VGL: Left venom gland, VGR: Right venom gland.

Identifier	Species identifier	Illumina Sequencing	Phylogeny	Quantification	Toxin annotation	Proteome	ONT
SB1946VGR	*Hydrops triangularis_3*	X	Potentially contaminated	X	Potentially contaminated	S	
SB1439VGR	*Hydrops triangularis_2*	X	X	X	X		
SB1259VGR	*Hydrops triangularis_1*	X	X	X	X		X^1
SB156VGR	*Helicops polylepis_2*	X	X	X	X	S	
SB0155VGL	*Helicops polylepis_1*	X	X	X	X		
SB1930VGR	*Helicops modestus_3*	X	X	X	X		
SB1347VGR	*Helicops modestus_2*	X	X	X	X	S	X^2
SB0306VGL	*Helicops modestus_1*	X	X	Outlier	X		
SB544VGR	*Helicops leopardinus_2*	X	X	X	X		
SB1164VGR	*Helicops leopardinus_1*	X	X	X	X		
SB1845	*Helicops infrataeniatus_4*	X	X	X	X		
SB162VGR	*Helicops infrataeniatus_3*	X	X	X	X		X^1
SB0125VGR	*Helicops infrataeniatus_2*	X	X	X	X		
SB0124VGR	*Helicops infrataeniatus_1*	X	X	X	X	S	
SB0641	*Helicops carinicaudus_4*	X	X	X	X	S	
SB354VGR	*Helicops carinicaudus_3*	X	X	X	X		X^2
SB0353VGL	*Helicops carinicaudus_2*	X	X	X	X	SG	
SB0141VGR	*Helicops carinicaudus_1*	X	X	X	X		
SB1256	*Helicops angulatus_4*					G	
SB795VGR	*Helicops angulatus_3*	X	X	X	X	S	X^2
SB561VGR	*Helicops angulatus_2*	X	X	X	X	S	
SB1945VGR	*Helicops angulatus_1*	X	X	X	X		
SB0129VGL	*Tomodon dorsatus_3*	X	X	X	X		
SB0110VG	*Tomodon dorsatus_2*	X	X	X	X		
SB0109VG	*Tomodon dorsatus_1*	X	X	X	X		
SB0079VGL	*Mesotes strigatus_3*	X	X	X	X		
SB0029VGL	*Mesotes strigatus_2*	X	X	X	X		
SB0034VGL	*Mesotes strigatus_1*	X	X	X	X		
SB0215VGL	*Dryophylax aff. nattereri*	X	X				
SB0357VGL	*Dryophylax hypoconia_2*	X	X	X	X		
SB0033VGR	*Dryophylax hypoconia_1*	X	X	X	X		
SB0063VGR	*Dryophylax chaquensis*	X	X				
SB0404	*Ptychophis flavovirgatus*	X	X	X	X		
SB0403	*Ptychophis flavovirgatus*	X	X	X	X		
SB363VLG	*Pseudoboa nigra_3*	X	X	X	X		
SB0365VGL	*Pseudoboa nigra_2*	X	X	X	X		
SB0364VGL	*Pseudoboa nigra_1*	X	X	X	X		
SRR12802477	*Psammophis_schokari*	X	X				
SB0132VGL	*Philodryas olfersii_3*	X	X	X	X		
SB0026VGR	*Philodryas olfersii_2*	X	X	X	X		
SB0001VGL	*Philodryas olfersii_1*	X	X	X	X		
SRR11341145	*Philodryas nattereri*	X	X				
SB0238VGL	*Oxyrhopus guibei_3*	X	X	X	X		
SB0201VGL	*Oxyrhopus guibei_2*	X	X	X	X		
SB0188VGL	*Oxyrhopus guibei_1*	X	X	X	X		
SB0287	*Micrurus_corallinus*	X	X				
SB0065	*Mastigodryas_boddaerti*	X	X				
SB0292	*Leptoidera_anulata*	X	X				
SB0308VGR	*Gomesophis brasiliensis*	X	X				
SB0190VGL	*Erythrolamprus miliaris*	X	X				

3.2 RNA extraction and venom gland transcriptome sequencing

For RNA-Seq experiments, the excised venom glands were homogenized using Precellys 24 tissue homogenizer (Bertin Technologies SAS, France). TRIzol (Invitrogen, USA) method was used for RNA extraction based on the method of (CHOMZYNSKI; SACCHI, 1987). Sample purity was determined through the A260/280 ratio (NanoDrop, Thermo Fisher USA). RNA concentration was determined using fluorescence (Qubit3, Invitrogen, USA). RNA integrity number (RIN), measured with Bioanalyzer (Agilent, USA), was used as an indicator for RNA integrity.

For the short-reads sequencing, the extracted RNA was used to build Illumina libraries using TruSeq Stranded mRNA kit following manufacturer's recommendations (Illumina, USA). These were sequenced on a HiSeq 1500 (Illumina, USA) to obtain paired-end 150 base-pair reads.

To overcome limitations inherent in the small size of Illumina reads, long sequencing reads were obtained with Oxford Nanopore Technologies (ONT; Oxford, ENG) for five samples (Table 2), which were selected considering phylogenic position and RNA availability . For such, the extracted mRNA was retro-transcribed into single strand cDNA libraries with the SMARTer cDNA synthesis kits (Takara, USA), following the manufacturer's instructions. Double-strand synthesis and PCR amplification of the single strand cDNA libraries was done with the kit's 5' primer and Invitrogen's *taq* platinum (Invitrogen, USA). The number of PCR cycles was optimized following the kit's recommendations. The PCR product, purified with 1.5X AMPure beads (Beckman Coulter, USA), was used as input material to the Ligation Sequencing Kit V14 (Oxford, ENG) to obtain sequencing libraries, using the short fragment buffer and following the manufacturer's protocol. Libraries were loaded into a Mk1B MinION (Oxford, ENG) with a FLO-MIN114 (Oxford, ENG) flow cell. Data was acquired with the software MinKNOW. Due to changes in software version, data from the first two runs was acquired using the 260bp/s parameter and saved as fast5, while data from the remaining runs were acquired at 400bp/s and stored as pod5.

3.3 Illumina Reads Filtering and Transcriptome Assembly

Prior to trimming, each set of Illumina reads was scanned for cross-contamination using a custom Python script. Afterwards, TrimGalore! v0.5.0 (KRUEGER, 2015) was used to trim Illumina adapters and low-quality bases (phred >25). FASTQC v0.11.5 (ANDREWS, 2010) was used for quality check of trimmed reads. Trimmed paired-end reads were merged using PEAR v0.9.10 (ZHANG *et al.*, 2014). For assembling non-toxin transcripts used in quantification and phylogeny inference steps, only Trinity was used, as it assembles the highest diversity of BUSCO genes among the assemblers (HOLDING *et al.*, 2018). Downloaded samples used in phylogenetic inference were trimmed and assembled using the same methods as the others, taking their sequencing methodology into consideration. Toxin transcripts were assembled using Trinity v2.10.0 (GRABHERR *et al.*, 2011), Seqman Ngen (SWINDELL; PLASTERER, 1997), rnaSPAdes v3.13.0 (BUSHMANOVA *et al.*, 2019), Extender release version (ROKYTA *et al.*, 2012), and Bridger v2014-12-01 (CHANG *et al.*, 2015) in order to maximize the number of full length toxin transcripts recovered, as each of these assemblers tend to perform better at assembling different toxin families (HOLDING *et al.*, 2018).

Statistics of Trinity assemblies were accessed using Quast v5.0 (GUREVICH *et al.*, 2013). BUSCO V5.5.0 (WATERHOUSE *et al.*, 2018) was used examine the completeness of the assemblies by querying for sequences from a database of conserved orthologs (tetrapoda odb10).

3.4 Phylogeny inference

To properly understand the evolution of toxin genes and venom composition in the Hydropsini, a species phylogenetic tree was reconstructed. The inclusion of related species allowed the identification of evolutionary novelties exclusive to Hydropsini, while more distantly related species were included to allow fossil calibration.

To infer the tree topology, toxin depleted transcriptomes of the sampled individuals were used (see section 3.5.3). Coding Sequences (CDS) and amino-acid sequences were determined using TransDecoder V.5.5.0 (HAAS; PAPANICOLAOU, 2016). Redundant transcripts were removed from each transcriptome with cd-hit, with a threshold of 99%. The non-redundant translated amino-acid sequences were clustered into orthogroups, clusters containing paralogs and orthologs (EMMS; KELLY, 2015), using OrthoFinder V 2.5.4 (EMMS; KELLY, 2015). The set of curated toxins (see below) was added to the OrthoFinder analysis. Orthogroups containing sequences from this set were considered homologous to toxins and removed, as toxins may bias the phylogenetic inference due to their high substitution rate and complex duplication history (MEBS, 2001; XIE *et al.*, 2022).

Orthogroups containing less than 98 (twice the number of sampled individuals) and more than 39 (80% of the number of sampled individuals) amino-acid sequences were aligned using MAFFT (KATOH; STANDLEY, 2013). These 8691 protein alignments were converted into codon-alignments with Pal2Nal (SUYAMA *et al.*, 2006). The codon-alignments were trimmed using TrimAl (CAPELLA-GUTIÉRREZ *et al.*, 2009) and alignments with at least 300 nucleotides were selected, resulting in 8449 trimmed nucleotide alignments. Unrooted gene trees were estimated from these alignments using RaxML with 100 bootstraps and model GTRGAMMA (STAMATAKIS, 2014). Branches with support of at least 50 bootstraps, containing sequences from 39 individuals or more (80% of the samples), and presenting only orthologous and inparalogs genes, were selected using UphO (BALLESTEROS; HORMIGA, 2016). Here, inparalogs are defined as sequences from the same individual forming a monophyletic clade in the unrooted gene trees. To obtain clades corresponding to single copy orthologs and maximize the number of informative sites, only the longest inparalogs were kept.

The RCFV (a measure of compositional heterogeneity, (ZHONG *et al.*, 2011) was estimated with BaCoCa v1.1 (KÜCK; STRUCK, 2014). Alignments with a RCFV above a 5% confidence interval threshold, calculated with a custom R script, were removed, resulting in 896 alignments. These were used to infer final nucleotide gene trees with RaxML with model GTRGAMMA and 100 bootstraps. Nodes of the resulting 896 Maximum likelihood trees with bootstrap support below 10 were collapsed using nw_utils release version (https://github.com/nwutils/nwutils.github.io). The final gene trees were used as input by ASTRAL (ZHANG *et al.*, 2018) to estimate a species tree. Tree was rooted to the most recent common ancestor (MRCA) of *Micrurus corallinus* and *Psamophis schokari*.

For time calibrated evolutionary analysis, an ultrametric tree was made using the topology of the Species tree previously obtained. Since ASTRAL may underestimate the length of branches (SAYYARI; MIRARAB, 2016), adequate tree branches length were obtained with RAxML under the rapid hill-climbing mode using a concatenated matrix of the single copy orthogroups alignments, obtained with FASconCAT (KÜCK; MEUSEMANN, 2010). The topology was constrained to the tree obtained with ASTRAL.

A calibrated ultrametric tree was obtained using treePL and the maximum likelihood tree estimated with RAxML, following MAURIN (2020). Ages were attributed to nodes using the following fossil information: For the common ancestor of *Mastrigodryas boddaerti* and *Micrurus corallinus, Coluber carduci* was used, with ages ranging from 54 Milions year ago (Mya) to 30.9 Mya; for the common ancestor of *Leptoidera annulata* and *Philodryas olfersii*, *Paleoheterodon* was used, with ages ranging from 54 Mya to 12.5 Mya; For *Micrurus coralllinus* and *Psammophis schokari* common ancestor, Elapid Morphotype A was used with ages ranging from 54 Mya to 24.9 Mya. These calibrations were based on (HEAD *et al.*, 2016; ZAHER *et al.*, 2019).

3.5 Transcriptomes Annotation and Curation

3.5.1 Annotation

Annotation of toxin genes was only performed in species for which multiple individuals were available. For a more stringent annotation, transcripts obtained with different assemblers were annotated as toxins, putative toxins, or non-toxins using ToxCodAn Release version (NACHTIGALL *et al.*, 2021b). Transcripts annotated as non-toxins were submitted to a second and more permissive toxin identification pipeline. These were locally aligned to the ToxProt database (JUNGO *et al.*, 2012) using blastx from BLAST suit V 2.12.0+ (ALTSCHUL et al., 1990). Transcripts producing hits with e-value lower than 10e-5 were translated with CodAn V 1.2 (NACHTIGALL *et al.*, 2021a) using the vertebrate model. Sequences lacking a stop codon were filtered out. The remaining sequences were combined to the sequences annotated as toxins and putative toxins by ToxCodAn. To remove endophysiological paralogs, the combined set was queried to a database containing SwissProt (THE UNIPROT CONSORTIUM *et al.*, 2023) and ToxCodAn sequences using blastp (e-value <10e-5). The query was only considered a putative toxin if one of the five resulting hits with highest score matched sequences from ToxProt or ToxCodAn databases.

3.5.2 Curation

Sequences annotated as toxins or putative toxins were selected for a curation process. Such curation is necessary to remove mis-assembled and redundant sequences, resulting from the use of more than one assembler. A custom curation pipeline, including both manual and automated steps, was designed for this purpose (Figure 2). To select the best sequences, the pipeline sought to combine sequencing and phylogenetic evidence.

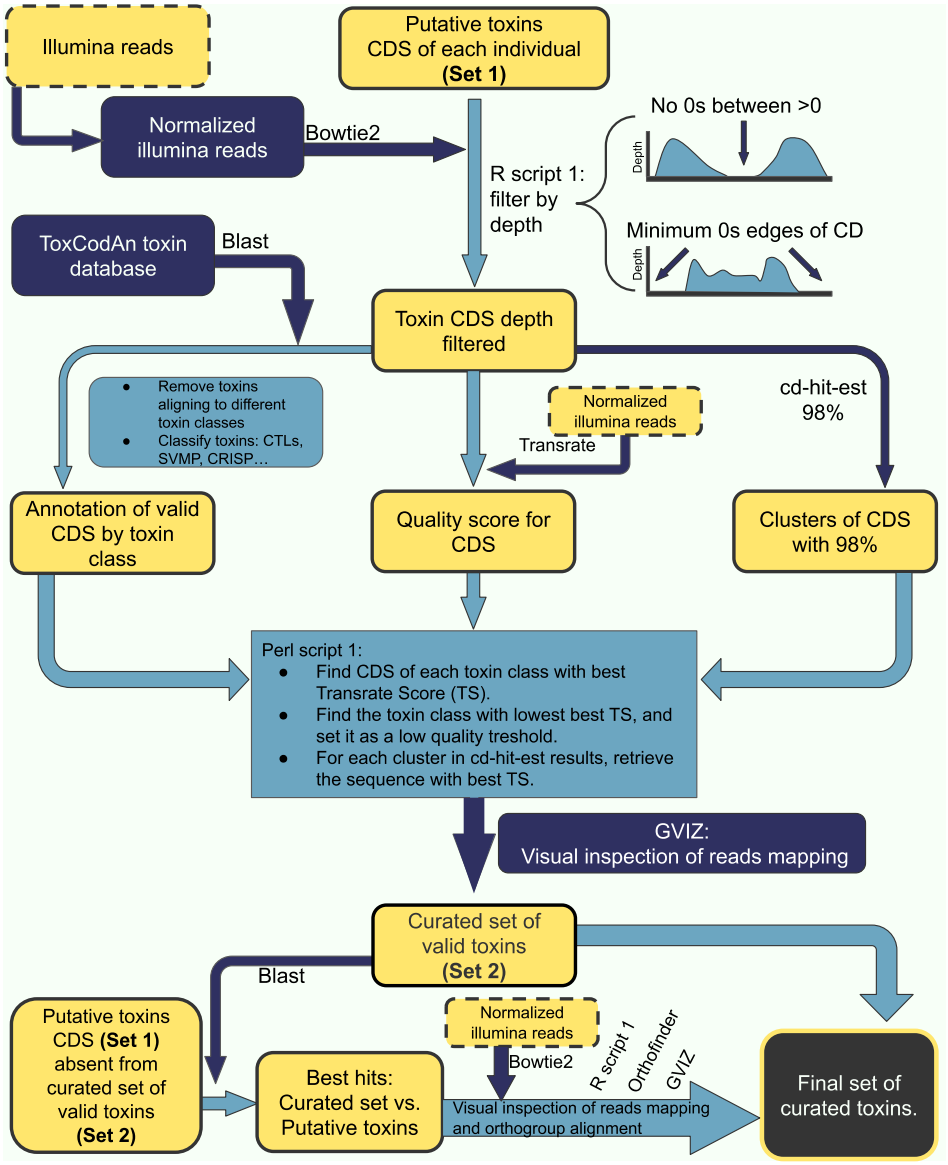


Figure 3: Curation pipeline employed for keeping supported toxin transcripts. Dark blue represent third-part scripts, while light blue represent in-house scripts. Dashed squares represent reads. As the Set 2 contained many partial transcripts, similar sequences were obtained from Set 1 and re-analyzed.

The automated part of the pipeline starts by normalizing reads from each individual using the insilico_read_normalization.pl, from Trinity Utils (GRABHERR *et al.*, 2011), reducing computational time in further steps. Then, these normalized reads were mapped to the toxin Coding Sequence (CDS) using bowtie2 (Version 2.4.5 with parameters --local and –very-sensitive) (LANGMEAD; SALZBERG, 2012). Using a custom Rscript (R CORE TEAM, 2022), CDS containing sites with no read mapped (depth = 0) were excluded if a) these sites were more than 90 bps away from the start or stop codon and b) these sites were between covered sites (R script 1; Figure 2). Next, the remaining CDS were aligned to the ToxCodAn (NACHTIGALL *et al.*, 2021b) toxin database with blastx. CDS aligning to more than one class of toxin were excluded, as these could be assembling artifacts. After that, Transrate (SMITH-UNNA *et al.*, 2016) and cd-hit (FU *et al.*, 2012) were used to remove redundancies and low quality transcripts, while keeping the CDS with highest Transrate Score (TS). The annotation obtained with ToxCodAn in the previous step were used to determine the minimum TS threshold to keep all identified toxin classes (Perl script 1; Figure 2).

For a manual inspection, the Jaccard index (defined as the ratio of the number of reads spanning a region in the CDS divided by the number of reads that only reach one of the regions' edges) of each transcript was obtained with inchworm_transcript_splitter.pl, from Trinity Utils. Additionally, OrthoFinder V 2.5.4 (EMMS; KELLY, 2015) was used to cluster the putative toxins in groups of homologous genes, which were aligned with MAFFT--auto (KATOH; STANDLEY, 2013). These alignments, the reads mapping and the Jaccard index were visually inspected using GVIZ (HAHNE; IVANEK, 2016) to remove low quality sequences. Briefly, if a given portion of a CDS was badly aligned to the remaining CDS in the orthogroup and the Jaccard index indicated low read support in that region, the CDS was discarded, otherwise, these were added to the final dataset.

Bias in the first mapping step was observed, possibly due to the redundancies in the bowtie2 database, leading to an abundance of partial toxins in the final set. To recover complete (or at least longer) sequences, the CDS from the final dataset were aligned to the initial dataset of putative toxins using blastn. Next, bowtie2 was used to map the normalized set of reads to the best hits resulting from the blastn alignments. The Rscript 1 was used to filter these alignments and, through a visual inspection, CDS supported by reads mapping were selected to the final dataset.

Sequences identical to those of other species, but with low read coverage were removed, as they likely correspond to contaminants. Additionally, sequences with truncated CDS were removed, as they likely correspond to assembly chimeras.

3.5.3 Species Reference Datasets

In order to obtain a representative, yet non-redundant, set of non-toxin transcripts, only the Trinity assembly was used, as it usually contains the highest diversity of non-toxin genes, (HOLDING *et al.*, 2018). Toxin coding contigs were identified through Blast searches (90% of identity and 95% of overlap) using the curated toxin dataset as query and removed from the assemblies. The resulting dataset corresponds to Trinity assembled venom gland transcriptomes depleted of toxins. ORFs from this dataset were predicted and translated using TransDecoder V5.5.0 (HAAS; PAPANICOLAOU, 2016) with its default parameters.

Reference sets of predicted proteomes for each species were obtained by joining the toxin depleted sequence datasets from all individuals in a single file and removing redundancies with cd-hit-est with a 95% threshold. Redundancies were also filtered out of the toxin sequences, with a 98% threshold. The toxin and non-toxins sets were merged and final redundancies were removed with a 98% threshold (Set 1). This final set was used for all analyses, except said otherwise.

3.6 ONT long-reads basecalling, reads pre-processing and transcript assembly.

Reads from the selected individuals (Table 2) were basecalled with Guppy V. 2.24-r1122 (ONT; Oxford, England) in SUP mode and the appropriate configurations for each case. During basecalling, a gross removal of adapters and read-splitting was performed while converting the binary fast5 (or pod5) files into fastq. Quality check of the basecalling was obtained with pycoQC v2.5.2 (LEGER; LEONARDI, 2019). A more sensitive removal of adapters and splitting of chimeric reads was obtained with pore-chopper V0.2.4 (https://github.com/rrwick/Porechop), setting the adapter identity threshold to 85% at both the middle and end of reads, while requiring a minimum of 60 bps to keep a split read. With the aid of a custom Perl script, reads were further split if one or more polyA (or polyT, if in the anti-sense), detected as consecutive series of 12 or more adenines (or thymidines), were present in the middle of the read.

With this final set of reads, transcriptomes were assembled using RNA-Bloom2 v1.0 (NIP *et al.*, 2023). The following flags were added to maximize the diversity of recovered transcripts (-p 0.95 -grad 0.8). Completeness of the final assembly was accessed with BUSCO.

To identify putative toxin transcripts, blastx was used to align the assembled transcripts to the same custom database used to annotate Illumina transcripts, containing the SwissProt and ToxCodAn set of toxins. Transcripts producing best alignments to proteins of ToxCodAn or ToxProt databases and covering at least 95% of the target were considered putative toxins. To account for mis-assemblies leading to frame shifts, coverage to the database was calculated using all alignments obtained between given query and target.

Validation of putative toxin transcripts was carried out by mapping the processed long-reads to the assembly using BWA-mem V.0.7.17-r1188 (LI, 2013). As no secondary alignments were allowed, disagreements between reads and reference should result from sequencing errors, absence of true transcript or badly assembled transcripts. Consistent

disagreements between reads and references, searched with bam-readcount (KHANNA *et al.*, 2022), were used as indicatives of badly assembled transcripts. A custom python script was used to classify the transcripts. As an error rate of at most 10% is expected, putative toxin transcripts were considered badly assembled and discarded if more than 90% of the mapped reads presented the same alternative base in any site. Further, transcripts with a proportion between 10% and 90% of reads supporting the same alternative alignment could result from badly assembled transcripts, absence of true transcripts or both, and were marked for reassembly. Transcripts with less than 10% of transcripts supporting the same alternative alignment were considered good.

Reassembly was carried out to obtain transcripts that could be validated by the sequencing reads (Figure 3). For such, a mapping of sequencing reads to a bait transcript was used to create groups of reads according to their original transcripts, which were then used to make a consensus. Firstly, the badly assembled transcripts, as defined above, were filtered out from the initial set of putative transcripts, then ambiguities were removed at an 80% threshold using cd-hit-est. Then the new set of transcripts was reclassified. These three steps were repeated until no additional bad transcript was found. Finally, the transcript with more aligned reads was selected as the bait. BWA-mem was used to align all reads to the bait. From this mapping, the ratio of the most common to the second most common base was computed for each site in the reference, using bam-readcount. If this ratio was above 10%, then such variation should indicate that reads from different transcripts mapped to the bait transcript (Figure 3). Sites with this ratio of divergence were saved for the next step. If more than 50 positions meet this criterion, only the ones with the highest ratio were used.

For all reads, the bases aligned to the saved positions were recorded as an index and the occurrence of each index in the dataset was counted (Figure 3). Given that reads coming from the same transcripts should share the same index, this information can be used to separate reads according to their transcript. For this clustering step, the distance between all pairs of indexes was calculated as the simple number of differences, resulting in a distance matrix. Using the python package ‘sklearn’ (HAO; HO, 2019), clusters were created using this matrix through the function AgglomerativeClustering (with the method ‘single’) and the distance threshold that maximize the Calinski-Harabasz score (CALINSKI; HARABASZ, 1974) of the clusters. In this point, each cluster represents the sets of bases that are expected to be found in reads originating from representative transcripts,

considering possibility of error. The most frequent index of each cluster was compared to the index of all reads. Reads with fewer differences to the cluster reference than the distance threshold between clusters (defined in the clustering process) are saved, aligned with MAFFT (--adjustdirectionaccurately --retree 2 --maxiterate 500) and used for for consensus building with ‘cons’ from the EMBOSS-suite v6.6.0.0(RICE *et al.*, 2000), generating the final transcripts. For the alignment with MAFFT, if more than 300 hundred reads were found, 300 reads are sampled for reducing processing time. For the consensus, the plurality parameter was set to 20%. The final set of transcripts was manually inspected and checked for redundancies. This step was repeated until no more sites with high discordance rate were identified.

This reassembly pipeline was wrapped in a python script, which was named TReCAO (Transcriptome Reassembly clustering algorithm for ONT).

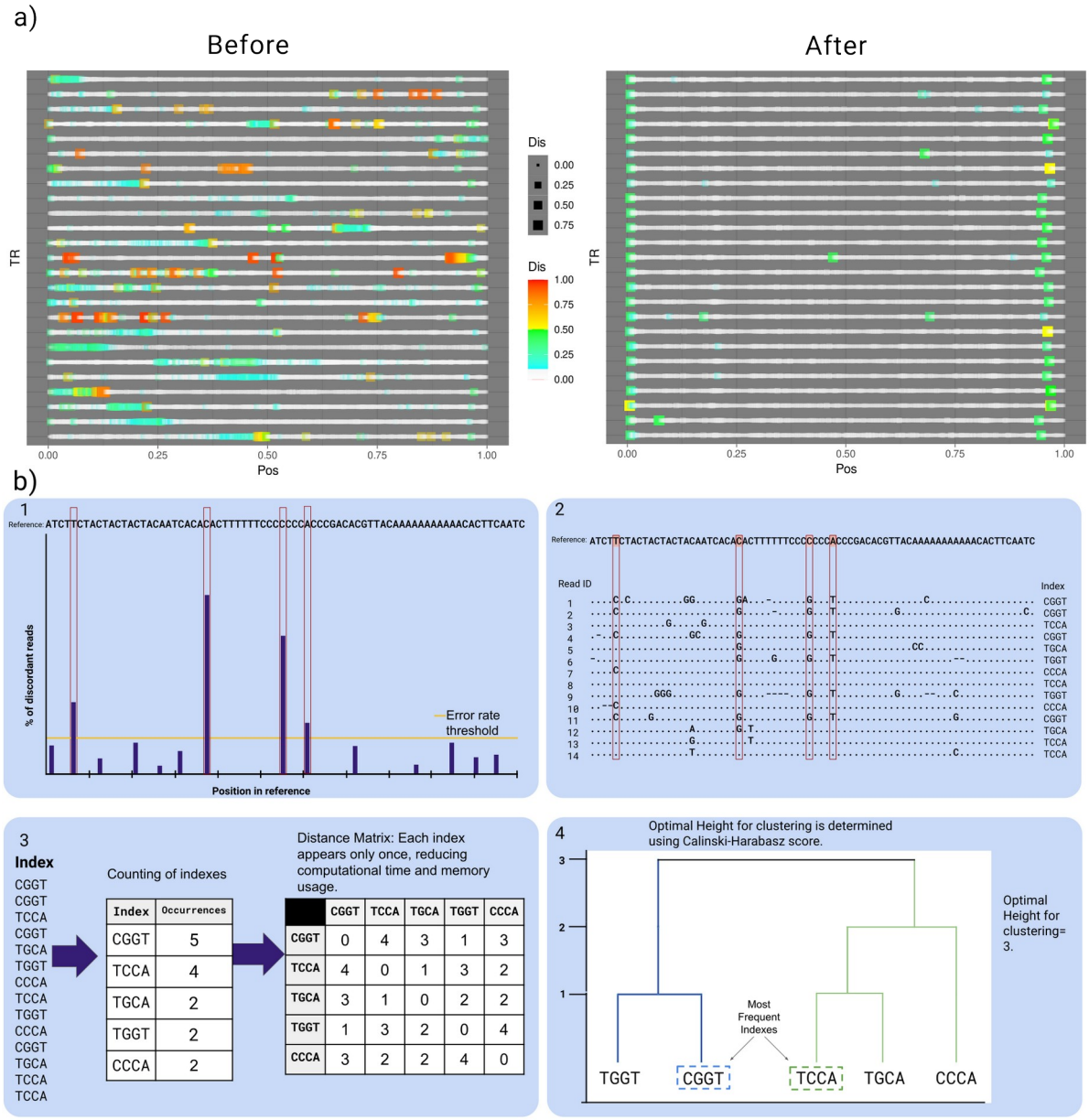


Figure 4: TReCAO pipeline. a) Transcript discordance before and after using the pipeline. Size and color of square indicate the the proportion of reads not supporting given base at each position (Discordance rate). TR: Transcripts; b) Pipeline used to cluster reads according to their original transcript: Positions in the transcript with high discordance are saved. Reads are indexed by the nucleotide present in the discordant sites. The frequency of each index is saved and used to calculate a distance matrix. Clusters of new reads are created based on the distance matrix.

3.7 Identification of Toxin Orthogroups

To compare the sequences and expression of toxins from different species, a homology based gene clustering methodology was used. OrthoFinder (inflation value = 3.0) was used to identify and cluster potentially homologous (*i.e.* paralogs and orthologs) sequences from different species into orthogroups, using as input the species reference proteome with toxins (Set 1). Orthogroups containing toxins sequences were annotated using the previously obtained best hits of their sequences.

These clusters allowed a proper comparison between the expression of snakes from different species. This approach was particularly convenient for the present framework, as the toxin families may present high heterogeneity in the number of transcripts across species.

To assign Orthogroups to the transcripts assembled using ONT long-reads, the transcripts were queried to Orthogroups containing toxin sequences using blastp. Next, the ONT transcripts were assigned to the Orthogroup containing the predicted protein that produced the best alignment.

3.8 Quantification of Toxin and non-toxin gene families transcript expression

A transcript expression analysis was conducted to estimate the venom composition of the studied snakes. Only species with more than one individual sampled were used. RSEM V1.3.1 (LI; DEWEY, 2011) pipeline was used to obtain values of Transcripts Per Million (TPM) through mapping of the Illumina reads to the corresponding species CDS, obtained previously (Set 1, section 3.5.4). For this analysis bowtie2 was chosen as aligner and bowtie2-mismatch-rate was set to 0.05 to reduce the rate of multi-aligned reads. After a first quantification, CDS without mapped reads were removed from the species data-set for a definitive quantification.

Venom glands from different individuals may present variation in the total number of reads corresponding to toxin transcripts. This may be caused by different amounts of non-glandular tissues retained during venom gland extraction or temporal differences in the venom synthesis cycle during fixation. In order to calculate mean values that only take into account the composition of the venom, toxin expression levels were normalized by the total TPM of toxins within each sample. The total TPM of toxins of each individual was obtained by summing the TPM of the CDS assigned to orthogroups containing at least one toxin CDS. The mean value of species relative expression for given CDS was calculated as

$$\overline{X} = \frac{\sum_{i=1}^{N} \frac{O_i}{T_i}}{N} \qquad \text{(Equation 1)}$$

in which O_i is the CDS expression in the individual i, T_i is the total toxin TPM in the individual i and N is the number of individuals sampled in the species.

Principal component analysis (PCA) of normalized expression values was performed with the package '*FactoMineR'* (LÊ *et al.*, 2008). For this analysis, the data was transformed using *center log-ratio* transformation to avoid biases arising from constant sum of the variables (MESSIAS, 2016).

For a more robust quantification of paralogs and isoforms, an expectation maximization approach was used to quantify the toxin expression using ONT long-reads, as these are long enough to cover whole CDS and should present a lower multi-mapping rate. For this analysis, CDS belonging to the CTLs orthogroup highly expressed in Hydropsini, (CTL0, CTL2, CTL12, CTL21; see results) were substituted with the CTLs assembled with ONT long-reads. For the quantification, ONT long-reads were mapped to the corresponding species reference CDS set using minimap2 2.17-r954-dirty; (LI, 2018) and the expression was calculated using Nanocount V1.1.0 (GLEESON *et al.*, 2022).

3.9 Venom proteomic experiments

Eight crude venom samples were subjected to shotgun proteomic assays. For such, the venom samples were lyophilized and resuspended in water. These were submitted to CEFAP-USP, where they were digested with trypsin and analyzed with High Performance Liquid Chromatography coupled to tandem Mass Spectrometry (HPLC-MS/MS), in a Orbitrap-Velos equipment. Data acquisition was performed as described in Freitas-de-Sousa et al., 2020. For each sample, two replicates were analyzed.

Obtained spectra were assigned to peptides using MaxQuant (TYANOVA *et al.*, 2016). The minimum peptide size was set to 7 and the False Discovery rate was set to 1%, using the backward database as decoy. Error tolerance was set to 20 ppm. For each sample, the respective species reference proteome (set 1) was used as database. With PERSEUS (TYANOVA; COX, 2018) proteomic ruler (WIŚNIEWSKI *et al.*, 2014), the concentration of each toxin in the samples was estimated from protein groups intensity (inferred by MaxQuant). These estimates of concentration were used for venom composition comparisons between species and individuals. Concentration of protein groups comprising sequences from different orthogroups was split among the found orthogroups. Given the lack of resolution provided by this method, comparisons were obtained by summing the concentration of each toxin class, instead of each orthogroup.

As a qualitative alternative to investigate the venom composition, 13% SDS-PAGEs of the extracted venom were performed with *Helicops angulatus* and *Helicops carinicaudus*. For each of these species, individual and pooled venom samples were analyzed under reduced and non-reduced condition. Bands were selected for subsequent sequencing with HPLC-MS/MS. I*n gel* digestion and peptide isolation were obtained following SHEVCHENKO et al., 2006, after which the peptides were re-disolved in 20ul of 0.1% formic acid. Next, 10ul of each sample was injected in a EASY-nLC (Thermo Scientific, Waltham, MA, USA) coupled to LTQ Orbitrap Velos (Thermo Scientific, Waltham, MA, USA). Separation was obtained with a linear gradient of solution A (formic acid 0.1%) solution B (acetonitrile 100%) (2%-24% in 8 minutes and 24%-80% in 12 minutes) under a constant flow of 300 nl/min. Percursor ion scan was obtained on Orbitrap

with a resolution of 30,000 within a scan range of 400 m/z to 1600 m/z. Percursor ions were fragmented by collision induced dissociation-CID with a normalized collision energy of 35. The resulting top 10 most intense ions within a window of 2 m/z were submitted to MS2 scan. Obtained spectra were identified using MaxQuant, with a maximum 5% FDR. For each band, intensity, spectral account and molecular weight were used for identification of the toxins in each band.

3.10 Evolutionary analyses of toxins

To describe and understand the evolution of toxins in Hydropsini, phylogenetic trees of the toxin genes were inferred. From the toxin classes determined in the annotation step the CTLs and Kunitz were selected, based on their considerable expression across the sampled Hydropsini transcriptomes as well as their presence in the proteome.

3.10.1 C-type Lectins

Orthogroups identified as CTLs and presenting transcriptional levels superior to 0.5% of the total toxin expression in any individual were aligned by codon with MAFFT and Geneious, along with 3 Elapidae, 4 Viperidae and 3 Dipsadidae sequences downloaded from EMBL database. The resulting alignment was trimmed with Trimal and used to estimate maximum likelihood tree using RAxML, with 1000 bootstrap, and setting the model to GTRGAMMA. The resulting tree was rooted to non-dimeric Viperidae CTLs. Trees were visualized and annotated with R (R CORE TEAM, 2022) package ‘ggtree’ (YU, 2020).

To explore the evolutionary novelties in Hydropsini CTLs, transcripts assembled with ONT long-reads and presenting high expression levels were compared to CTL sequences with three-dimensional information available. In order to find a good structural data model, these CTLs were aligned to a database containing PDB sequences using blastp and the most common best hit was selected. Next, using MAFFT, PDB sequence, the ONT assembled CTL transcripts and other Dipsadidade CTL sequences found in the same orthogroups as the ONT transcripts were aligned. The pairwise similarity between the PDB sequence and other alpha CTL sequences in the multiple alignment was used to identify regions in the reference that differ from Hydropsini, but not from other Dipsadidae. Additionally, a phylogeny of the CTLs from Hydropsini was obtained using RAxML, with 200 bootstraps and GTRGAMMA.

For a deeper investigation of potential structural novelties in the Hydropsini CTLs, AlphaFold2 (JUMPER *et al.*, 2021) was used to model the three-dimensional structure of these toxins through ColabFold (MIRDITA *et al.*, 2022). Modeling was carried out with the three CTL transcripts from *Helicops modestus* assembled with ONT long-reads. This species was chosen as it presented few Hydropsini CTL sequences and considerable CTL levels in the proteome. For each transcript, the predicted structure was relaxed with amber and visualized with pymol (SCHRÖDINGER, LLC, 2015).

3.10.2 Kunitz

CDS from Kunitz belonging to orthogroups with expression levels above 0.5% of the total venom expression were selected for the evolutionary analyses. As Kunitz have different organizations of domains in Dipsadidae (CAMPOS *et al.*, 2016), domains of toxins from this class were identified using HMMSCAN (MISTRY *et al.*, 2013) and the Pfam (FINN *et al.*, 2014) database. Codons of the Kunitz domains were extracted and used to create a gene tree using MAFFT and MrBayes (RONQUIST *et al.*, 2012) with the following parameters: rates=gamma, ngammacat=4, brlenspr=unconstrained:gammadir(1.0,0.1,1.0,1.0), aamodelpr=fixed(poisson), hapepr=exponential(10.0), ngen=1100000, samplefreq=200, nchains=4, temp=0.2, savebrlens=yes, starttree=random, seed=25357. Following, each transcript was annotated by its set of domains. With these annotations, expression levels for each different domain organization were calculated.

4. Results

4.1 Transcriptome Sequencing and Assembly

Venom gland transcriptomes of 43 individuals were sequenced on Illumina, of which 21 correspond to Hydropsini. The trimmed read sets, obtained with TrimGalore! showed a median read length ranging from 148 to 151 base pairs. The number of reads per individual ranged from 11.2 to 37.8 million reads (Table 3). Paired end reads were merged with an average success rate of approximately 86.7% (even though SB0287VGL presented an abnormally low rate of merging), resulting in longer reads for all sets of reads (Table 3, Figure 4).

Table 3: Sequencing statistics. Column ‘Number of reads after Trimgalore! (PE)’ reffers to the number of paired-endreads.

Sample	Species_ID	Number of reads after Trimgalore! (PE)	Merged (PEAR)	Reads merged (%)
SB1945VGR	Helicops angulatus_1	21,301,744	19,764,242	92.78
SB561VGR	Helicops angulatus_2	17,896,420	15,555,204	86.92
SB795VGR	Helicops angulatus_3	25,143,335	21,996,098	87.48
SB0141VGR	*Helicops carinicaudus_1*	34,999,584	30,741,894	87.84
SB0353VGL	*Helicops carinicaudus_2*	11,317,744	9,592,138	84.75
SB354VGR	Helicops carinicaudus_3	18,458,474	16,277,274	88.18
SB0641VGR	Helicops carinicaudus_4	21,266,330	19,182,987	90.2
SB162VGR	Helicops infrataeniatus_3	20,108,845	18,281,815	90.91
SB1845VGR	Helicops infrataeniatus_4	26,440,918	23,098,704	87.36
SB0124VGR	*Helicops infratateniatus_1*	30,702,252	25,585,949	83.34
SB0125VGR	*Helicops infratateniatus_2*	38,688,980	33,574,205	86.78
SB1164VGR	Helicops leopardinus_1	23,019,368	20,720,691	90.01
SB544VGR	Helicops leopardinus_2	22,190,368	19,890,551	89.64
SB0306VGL	*Helicops modestus_1*	11,208,419	10,039,382	89.57
SB1347VGR	Helicops modestus_2	20,767,076	18,626,457	89.69
SB1930VGR	Helicops modestus_3	19,289,576	17,701,610	91.77
SB0155VGL	*Helicops polylepis_1*	11,611,711	10,026,670	86.35
SB156VGR	Helicops polylepis_2	23,335,791	20,982,862	89.92
SB1439VGR	Hydrops triangularis_2	20,130,410	18,655,016	92.67
SB1946VGR	Hydrops triangularis_3	29,122,962	26,792,430	92
SB1259VGR	Hydrops triangularis)1	23,283,547	21,614,371	92.83
SB0292VGL	*Leptodeira annulata*	22,928,937	18,851,588	82.22
SB0065VGL	Mastigodryas boddaerti	14,981,346	12,998,956	86.77
SB0287VGL	*Micrurus corallinus*	18,760,041	9,841,360	52.46
SB190VGL	*Erythrolamprus miliaris*	19,459,237	16,289,530	83.71
SB0188VGL	*Oxyrhopus guibei_1*	15,164,421	12,447,479	82.08
SB0201VGL	*Oxyrhopus guibei_2*	20,795,979	18,647,239	89.67
SB0238VGL	*Oxyrhopus guibei_3*	21,983,515	15,943,888	72.53
SB0001VGL	*Philodryas olfersii_1*	2,432,779	2,403,169	98.78
SB0026VGR	*Philodryas olfersii_2*	3,438,718	3,413,445	99.27
SB0132VGL	*Philodryas olfersii_3*	14,719,790	13,401,460	91.04
SB0364VGL	*Pseudoboa_nigra_1*	22,250,892	17,798,270	79.99
SB0365VGL	*Pseudoboa_nigra_2*	38,255,675	30,278,264	79.15
SB363VLG	*Pseudoboa_nigra_3*	16,454,487	13,360,215	81.19
SB0403VGL	*Ptychophis flavovirgatus_1*	18,714,067	15,821,697	84.54
SB308VGR	*Gomesophis brasiliensis*	18,501,403	15,426,124	83.38
SB0404VGL	*Ptychophis flavovirgatus_2*	24,716,618	21,197,232	85.76
SB0033VGR	*Dryophylax hypoconia_1*	37,405,066	31,492,261	84.19
SB0357VGL	*Dryophylax hypoconia_2*	15,951,423	13,463,256	84.40
SB0063VGL	*Dryophylax chaquensis*	14,137,747	11,500,239	81.34
SB0215VGL	*Dryophylax aff nattereri*	25,716,665	21,137,043	82.19
SB0029VGL	*Mesotes strigatus_1*	13,645,237	11,048,617	80.97
SB0034VGL	*Mesotes strigatus_2*	17,174,532	14,331,883	83.45
SB0079VGL	*Mesotes strigatus_3*	18,153,316	14,607,215	80.47
SB0109VG	*Tomodon dorsatus_1*	7,237,200	7,049,950	97.41
SB0110VG	*Tomodon dorsatus_2*	5,570,918	5,482,769	98.42
SB0129VGL	*Tomodon dorsatus_3*	16,047,231	12,984,049	80.91

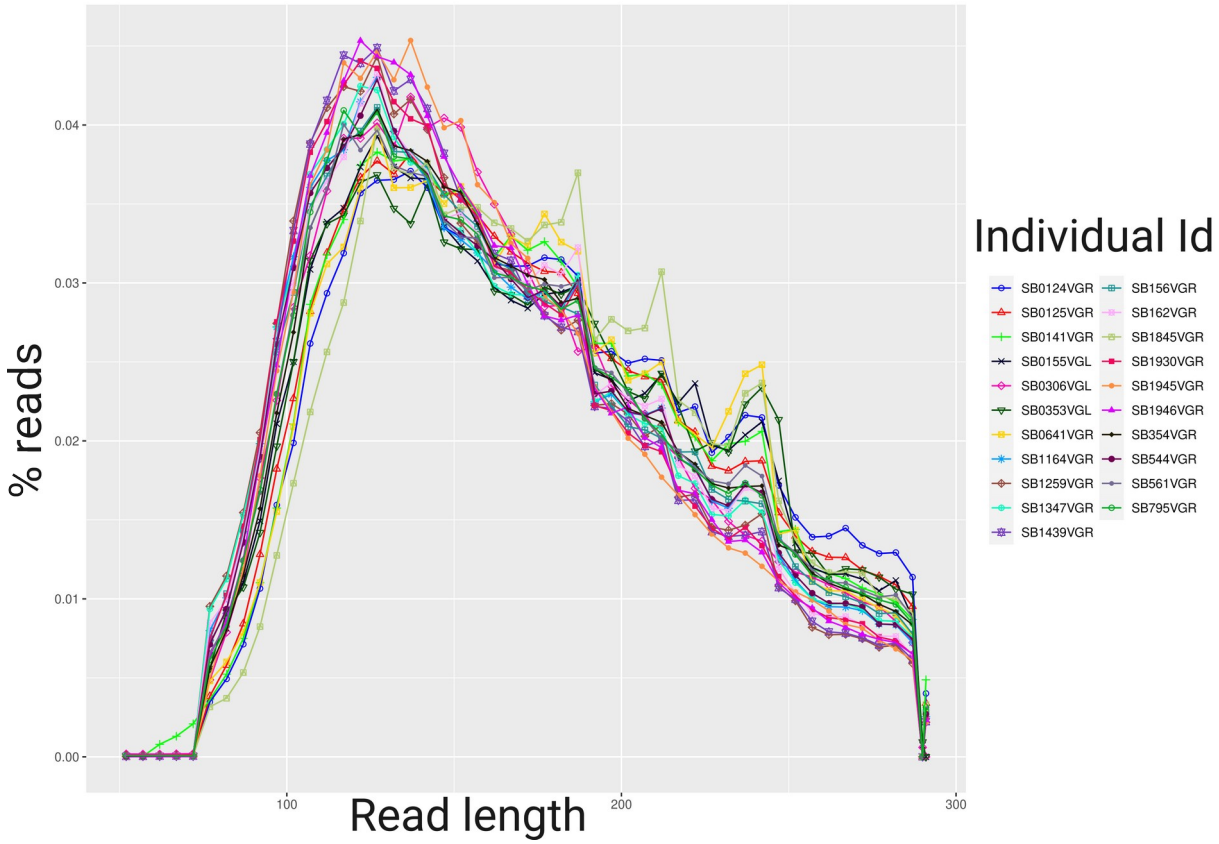


Figure 5: Size distribution of reads after merging with Pear only for sequenced Hydropsini. Each color represents an individual. After merging, reads with size greater than 150 were obtained.

With the merged sets of reads, transcriptomes were obtained using a diversity of assemblers to improve the chances of obtaining full-length transcripts of toxins. For the Trinity assemblies, used as non-toxin reference set, BUSCO identified a minimum of 12.8% of BUSCOs from the tetrapoda odb10 database in the assemblies (Figure 5). Nevertheless, the average percentage of complete BUSCOs between Hydropsini and the outgroup was similar (approximately 44% and 42%, respectively).

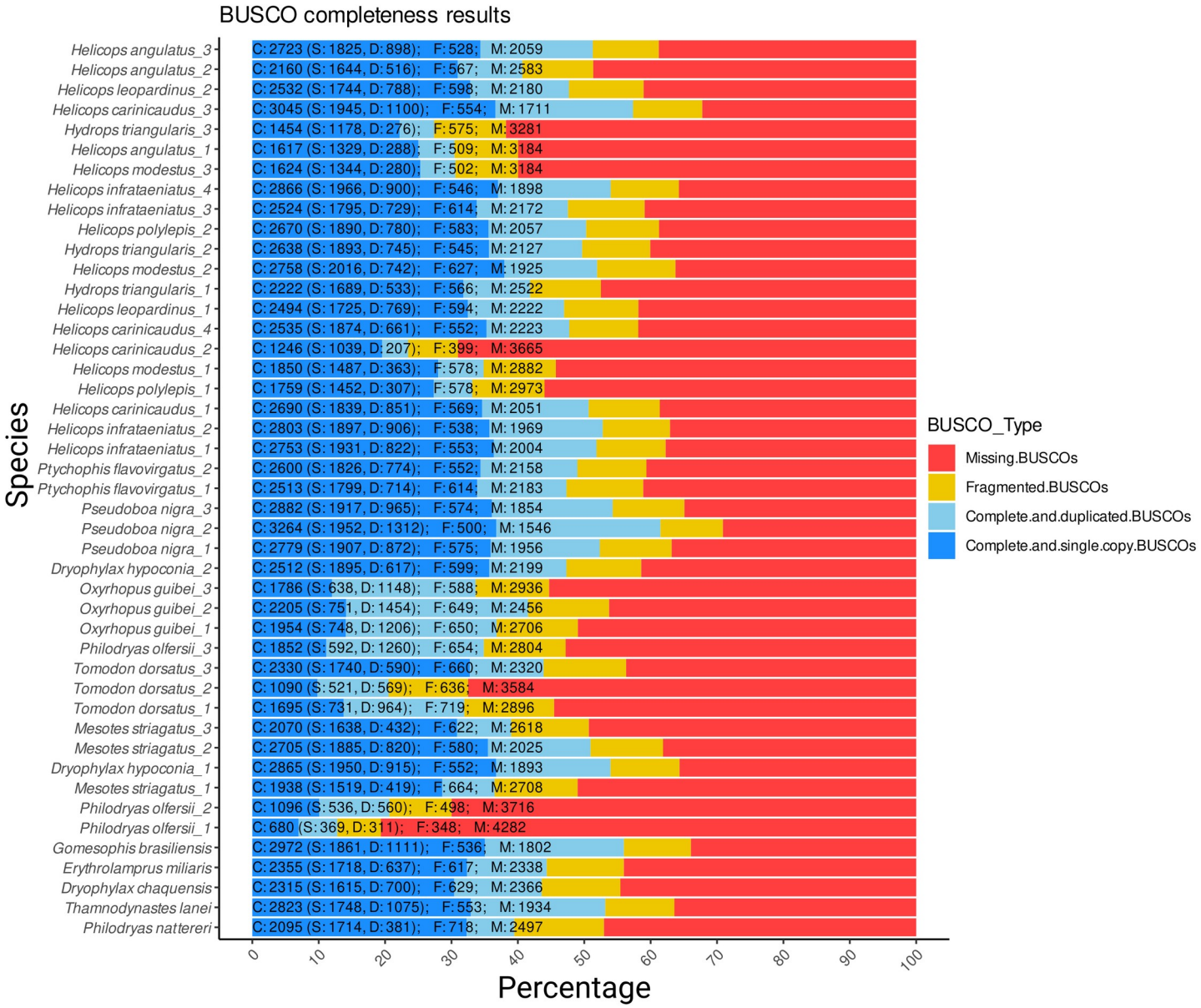


Figure 6: BUSCO analyses results from Illumina venom gland transcriptomes. Completeness lower than 100% is expected, as the snakes venom glands don't express all genes from BUSCO database. The consistency in completeness across samples is enough to conclude that the samples are comparable.

4.2 Phylogeny

The final set used to infer the species tree topology consisted in 896 gene trees resulting from alignments with an RCFV measure of at most 0.0309. The obtained tree (Figure 6) recovered the monophyly of Elapoidea, Colubroidea and Dipsadidae with support of at least 80% of the gene trees. Moreover, monophyly of the tribes used as outgroups (Philodryadini, Pseudoboini and Tachymenini) and of the focal group Hydropsini was also well supported. Although the phylogenetic relationship between most tribes was not well supported, Tachymenini was recovered as sister taxa to Hydropsini, supported by 62% of the gene trees, with a posterior probability of 1.

Inside Hydropsini clade, the genus *Helicops* was recovered as monophyletic with a posterior probability of 1. Two clades inside the genus were identified as highly supported, the first containing *Helicops angulatus* and *Helicops polylepis*, and a second one containing *Helicops leopardinus, Helicops infrataeniatus* and *Helicops modestus*.

To estimate branch lengths on the topology, a matrix of 733305 sites, built concatenating the alignments used to infer the gene trees was used. The species tree with appropriate branch lengths was used to obtain an ultrametric tree. The MRCA to Hydropsini dated to 21 Mya, while the MRCA of Tachymenini and Hydropsini dated to 27 Mya (Figure 7).

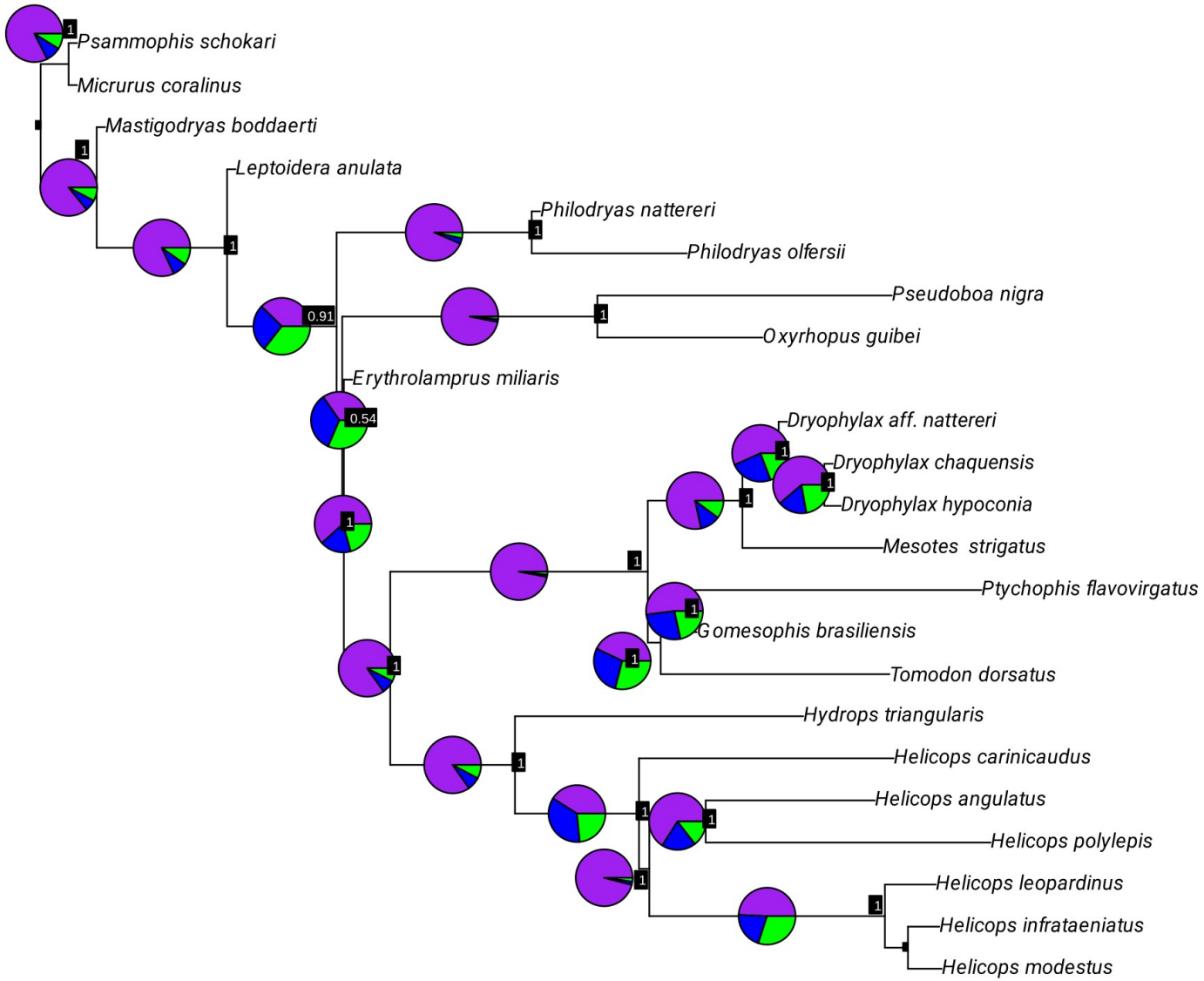


Figure 7: Astral tree topology obtained from 896 genes. Pie charts indicate quartet support for each node, in which purple corresponds to the shown quartets, while blue and green depicts alternative quartets support. Labels correspond to posterior probability.

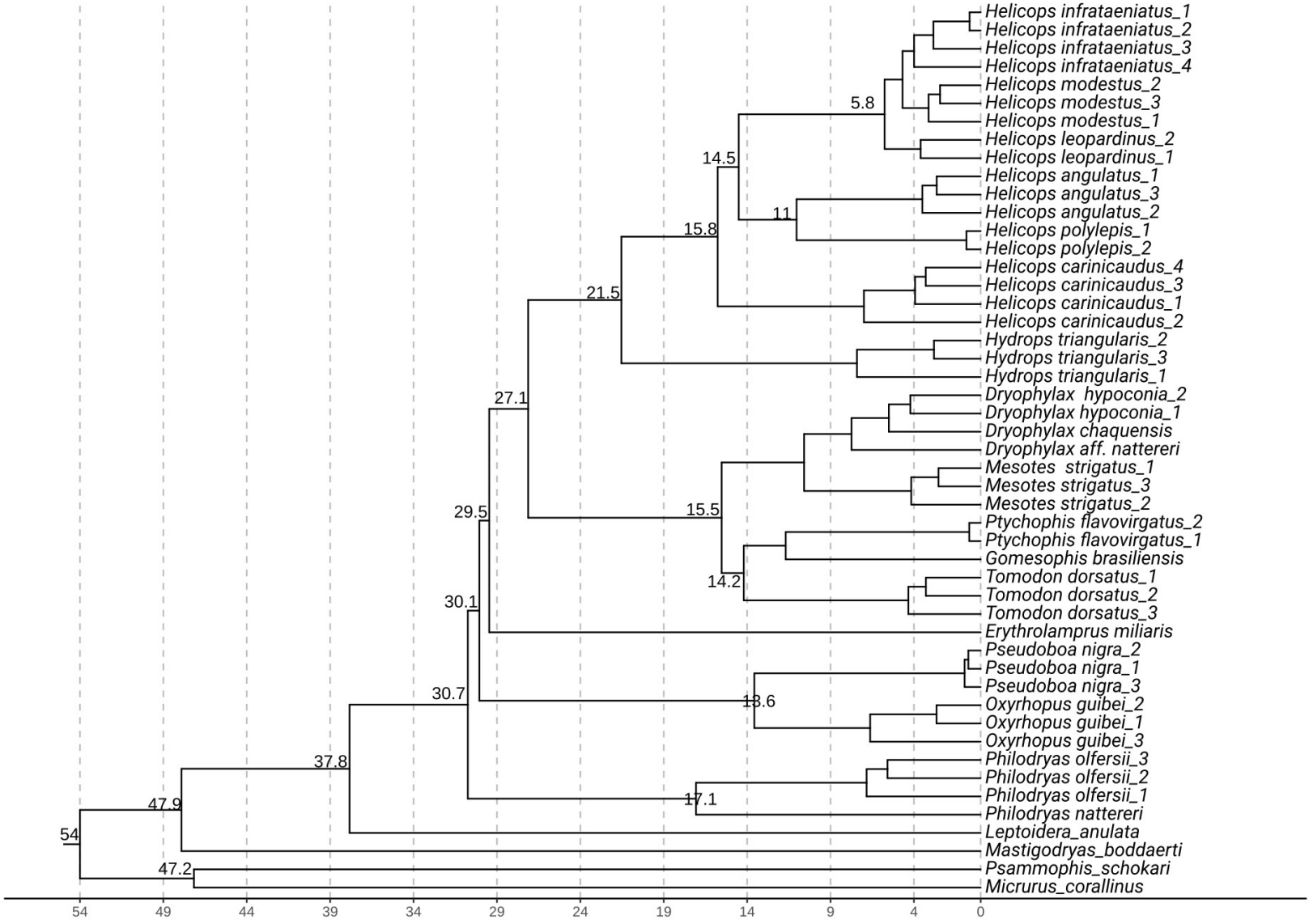


Figure 8: Ultrametric tree obtained using the topology in the Astral tree and branch lengths calculated by RAxML using a concatenated matrix of 896 genes. Numbers indicate Millons of years ago.

4.3 Transcriptome annotation

Using ToxCodAn and the custom pipeline, putative toxin transcripts from the venom glands of the sampled animals were annotated. The annotation process resulted in over 1000 contigs highly similar to toxins in certain individuals (Table 4). ToxCodAn found more putative toxins than the manual annotation pipeline (Table 4). From the initial set of transcripts highly similar to toxins, an average of 87% were removed during the validation step due to poor read mapping or redundancy. After validation and removal of redundancies, an average of 69 toxin transcripts were identified for each individual venom gland.

Table 4: Statistics of identified toxins for each sample. Entries in gray correspond to outgroups.

	Similarity search			Read coverage validation			
Species_ID	**Toxcodan**	**Manual**	**Total**	**Toxcodan**	**Manual**	**Total**	**Curation dropout rate**
Tomodon dorsatus_3	632	139	771	45	25	70	90.92%
Tomodon dorsatus_2	436	182	618	61	26	87	85.92%
Tomodon dorsatus_1	472	230	702	45	9	54	92.31%
Mesotes strigatus_3	669	219	888	43	21	64	92.79%
Mesotes strigatus_2	565	197	762	63	21	84	88.98%
Mesotes strigatus_1	617	136	753	66	23	89	88.18%
Dryophylax hypoconia_2	760	269	1029	32	15	47	95.43%
Dryophylax hypoconia_1	808	260	1068	70	24	94	91.20%
Ptychophis flavovirgatus_2	567	250	817	37	30	67	91.80%
Ptychophis flavovirgatus_1	654	280	934	40	24	64	93.15%
Pseudoboa nigra_3	571	276	847	52	8	60	92.92%
Pseudoboa nigra_2	808	342	1150	51	29	80	93.04%
Pseudoboa nigra_1	729	343	1072	47	21	68	93.66%
Philodryas olfersii_3	730	357	1087	64	38	102	90.62%
Philodryas olfersii_2	674	240	914	45	23	68	92.56%
Philodryas olfersii_1	776	354	1130	40	23	63	94.42%
Oxyrhopus guibei_3	899	532	1431	50	19	69	95.18%
Oxyrhopus guibei_2	718	469	1187	59	34	93	92.17%
Oxyrhopus guibei_1	688	458	1146	46	19	65	94.33%
Hydrops triangularis_3	705	427	1132	44	16	60	94.70%
Hydrops triangularis_2	706	434	1140	35	18	53	95.35%
Hydrops triangularis_1	602	418	1020	35	18	53	94.80%
Helicops polylepis_2	621	300	921	47	17	64	93.05%
Helicops polylepis_1	588	235	823	27	16	43	94.78%
Helicops modestus_3	488	215	703	55	20	75	89.33%
Helicops modestus_2	547	316	863	52	19	71	91.77%
Helicops modestus_1	198	94	292	54	18	72	75.34%
Helicops leopardinus_2	615	290	905	58	10	68	92.49%
Helicops leopardinus_1	584	269	853	50	30	80	90.62%
Helicops infrataeniatus_4	594	322	916	52	19	71	92.25%
Helicops infrataeniatus_3	642	353	995	59	25	84	91.56%
Helicops infrataeniatus_2	437	231	668	29	10	39	94.16%
Helicops infrataeniatus_1	717	301	1018	51	25	76	92.53%
Helicops carinicaudus_4	820	280	1100	55	41	96	91.27%
Helicops carinicaudus_3	634	272	906	52	23	75	91.72%
Helicops carinicaudus_2	714	327	1041	49	17	66	93.66%
Helicops carinicaudus_1	713	301	1014	41	12	53	94.77%
Helicops angulatus_3	586	284	870	50	15	65	92.53%
Helicops angulatus_2	577	279	856	20	1	21	97.55%
Helicops angulatus_1	422	171	593	59	28	87	85.33%

These validated transcripts were used to build a reference dataset for each species. After removing redundancies from the final toxin set, an average of 159 putative toxin transcripts were found for each species (Table 5). The number of transcripts attributed to toxin classes commonly found in snake venoms is represented in Figure 8.

Table 5: Number of putative toxin transcripts in each reference set of CDS used for downstream analyses.

Species	Toxin transcripts in reference file
Helicops angulatus	199
Helicops carinicaudus	293
Helicops infrataeniatus	272
Helicops leopardinus	167
Helicops modestus	163
Helicops polylepis	119
Hydrops triangularis	160
Oxyrhopus guibei	225
Philodryas olfersi	205
Pseudoboa nigra	216
Ptychophis flavovirgatus	113
Dryophilax hypoconia	181
Mesotes strigatus	198
Tomodon dorsatus	237

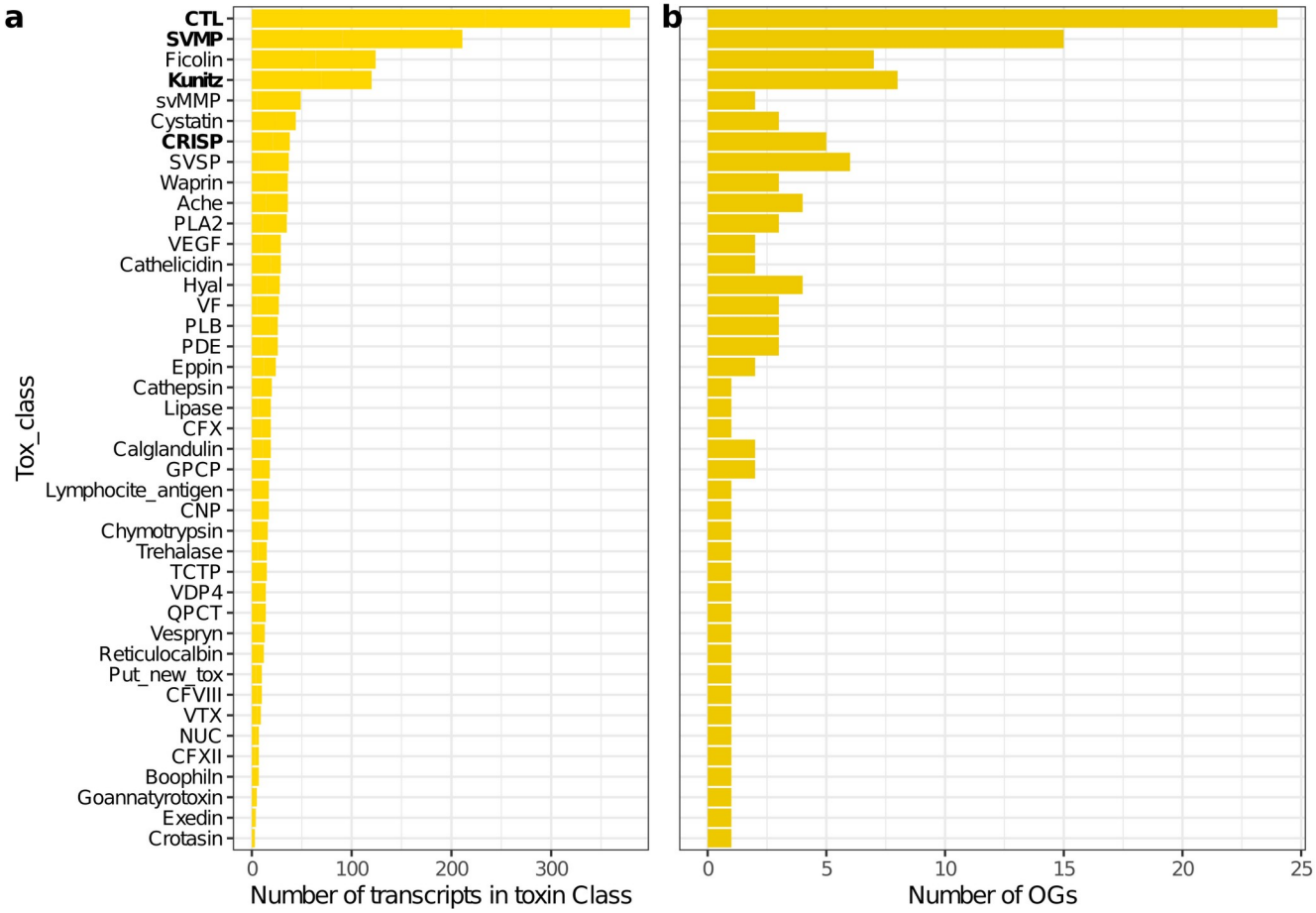


Figure 9: Protein classes identified in the reference toxin sets. **a**: Number of transcripts attributed to each protein class. **b:** Number of Orthogroups for each protein class. Toxins with significant expression in Hydropsini (see below) are in bold.

4.4 ONT long-reads assembly

Venom gland transcriptomes sequenced using Nanopore Flowcells yielded no less than 4 Gigabases and 7 million reads, except for *Helicops infrataeniatus_3*, which faced technical issues during the run (Table 6). During the basecalling process, an average of 30% of the reads were discarded due to low quality. After removing adapters and splitting reads, the final number of reads ranged from 90K to 12M (Table 6).

Table 6: ONT sequencing and trimming summary. Sequencing of SB1259 and SB795 was split in two different projects due to technical issues. Increase in number of reads after trimming and splitting results from split reads containing adapters or polyA tails.

Sequencing Run	Median read quality of passed reads (Phred score)	Passed Reads (Guppy)	Total Pass (Guppy)	% of Passing Reads (Guppy)	After trimming and splitting (Porechop + polyA_timmer)	% of reads After trimming and splitting (Porechop + polyA_timmer)
Hydrops triangularis_1 run1	14.7	8162644	11616015	87.41	12916828	111.2
Hydrops triangularis_1 run2	13.8	3453371				
Helicops modestus_2	13.2	5145571	5145571	66.43	5785231	112.43
Helicops carinicaudus_3	13.8	5689602	5689602	67.48	5837057	102.59
Helicops angulatus_3 run1	13.6	4677028	8460750	71.88	8938515	105.65
Helicops angulatus_3 run2	13	3783722				
Helicops infrataeniatus_3	12.3	96653	96653	54.88	90279	93.41

The number of transcripts assembled with RNA-bloom2 is shown in Table 7. The percentage of complete BUSCOs found was on average 58% of the total found in the Illumina assemblies (counting out the sample that encountered technical issues) (Table 8).

Table 7: Number of transcripts in each ONT sequenced transcriptome, assembled with RNA-Bloom2.

ID	Number of transcripts
Helicops infrataeniatus_3	1773
Hydrops triangularis_1	44996
Helicops modestus_2	75770
Helicops carinicaudus_3	104675
Helicops angulatus_3	115769

Table 8: Number of complete BUSCOs (single copy and duplicated) identified by Illumina and ONT.

Individual ID	Complete BUSCO Illumina	Complete BUSCO ONT	% (Illumina/ONT)
Helicops modestus_2	2758	1172	42.49%
Helicops carinicaudus_3	3045	2100	68.97%
Helicops angulatus_3	2723	1698	62.36%
Hydrops triangularis_1	2222	1282	57.70%
Helicops infrataeniatus_3	2524	71	2.81%

By aligning the Nanopore assemblies to a database containing toxins, CTL transcripts were identified (Figure 9). Most of these transcripts corresponded to poorly assembled or lowly covered transcripts, while the remaining were identified as dubious; none were classified as good. The absence of good transcripts suggests high disagreement rates between reads and the transcripts assembled by RNA-Bloom2 (Figure 9). These disagreements indicate that reads from different transcripts mapped to the same assembled contig, suggesting contigs with features from two different transcripts (chimeric transcripts), or absence of more appropriate contigs. Given these findings, reassembly of the transcripts was deemed necessary. With the developed algorithm, the final number of putative CTL transcripts ranged from 3 in *Helicops modestus* to 25 in *Hydrops triangularis* (Table 9; Figure 9).

Figure 10: Discordance rate for each site in the CTL transcripts identified in the sequenced individuals. Transcripts are sorted by mean discordance rate. For clearance, only the top 25 are shown for each assembler. Non-white squares indicate Discordance rate above 10%.

Table 9: Number of different transcripts identified as CTL in the Illumina dataset and after the reassembly of ONT reads.

Individual	Illumina	ONT
Helicops infrataeniatus_3	12	9
Hydrops triangularis_1	31	25
Helicops modestus_2	10	3
Helicops carinicaudus _3	24	15
Helicops angulatus_3	21	8

4.5 Orthogroups assignment

Orthogroup assignment for evolutionary and expression analyses was performed with OrthoFinder, using the reference translated transcriptomes from each species (Set 1). OrthoFinder identified 68,768 clusters of homologous sequences with at least 2 transcripts, 130 of which contained putative toxins. The mean percentage of transcripts identified as toxins in these orthogroups was 82%. Transcripts not previously identified as toxins could comprehend partial transcripts, mis-assembled transcripts, and toxins that escaped the depletion process.

These 130 toxin orthogroups contained toxins attributed to 42 different toxin classes. The toxin class with the most clusters was CTL, with 24 different orthogroups (Figure 8) and 379 transcripts. The same class also contained the orthogroup presenting most putative toxin transcripts (CTL0) with 107 transcripts (Figure 10). Notably, the sampled Hydropsini had many transcripts within this orthogroup, with the exception of *Helicops angulatus*, which presented a divergent and exclusive orthogroup (CTL12; Figure 10). Other toxin classes with numerous toxin transcripts, in descending order, were Snake Venom Metalloprotease (SVMP), Ficolin, Kunitz, and Snake venom Matrix Metalloprotease (svMMP) (Figure 8), the last being largely constituted by transcripts found in outgroups samples.

ONT assembled CTL transcripts were mostly assigned to the CTL0 orthogroup. Other orthogroups identified were CTL2 (in *Helicops carinicaudus)*, CTL12 (in *Helicops* angulatus), and CTL21 (in *Hydrops triangularis).* These identifications were based on alignments with at least 93% of identity, whereas alignments to transcripts from other orthogroups had at most 51% of identity.

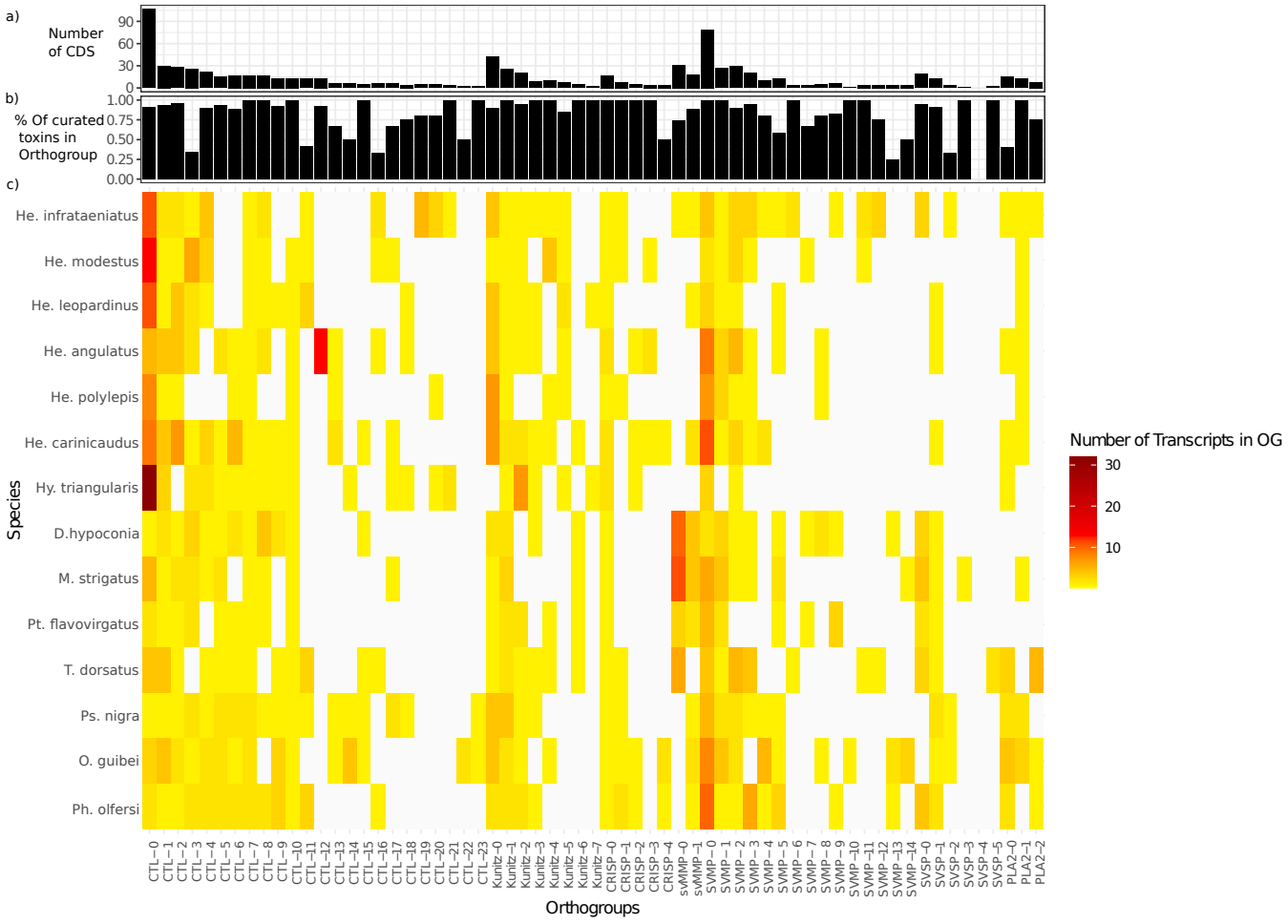


Figure 11: Orthogroup composition by species. a) Number of transcripts in Orthogroup, b) Percentage of transcripts in the orthogroup annotated as putative toxin. c) Number of transcripts per species. Only orthogroups for CTL, Kunitz, SvMMP, SVMP, SVSP and PLA2 are shown.

4.6 Toxin transcripts quantification

Transcript expression measures were obtained from mapping of reads to transcripts using an expectation maximization approach. A normalization was carried out to deal with individual differences in the amount of toxins in each transcriptome (Figure 11, 12). By summing the normalized expression of orthogroups containing the same toxin classes the mean venom profile of each species was calculated (Figure 13).

In the outgroups, the Philodryadini *Philodryas olfersii* presented a high expression of SVMP and CRISP (Figure 11, 13). The Pseudoboini *Oxyrhopus guibei* displayed a complex venom, with prevalence of CRISPs, CTLs and SVMPs, while the *Pseudoboa nigra* transcriptomes were also rich in Phospholipase-A 2 (PLA2) (Figure 11, 13). Species from the tribe Tachymenini presented a venom majorly composed of svMMP, in agreement to previous studies (BAYONA-SERRANO *et al.*, 2020; CHING *et al.*, 2012).

In Hydropsini transcriptome, the most expressed toxin transcripts were those encoding CTLs, comprising at least 83% of the total expression of toxins in *Helicops angulatus* and as much as 98% in *Hydrops triangularis* (Figure 13). Other toxins presenting a modest, yet relevant expression in Hydropsini were the Kunitz and the CRISPs (Figure 12,13). Additionally, SVMPs were found to compose a small portion of the transcriptome of *Helicops carinicaudus* (Figure 12,13).

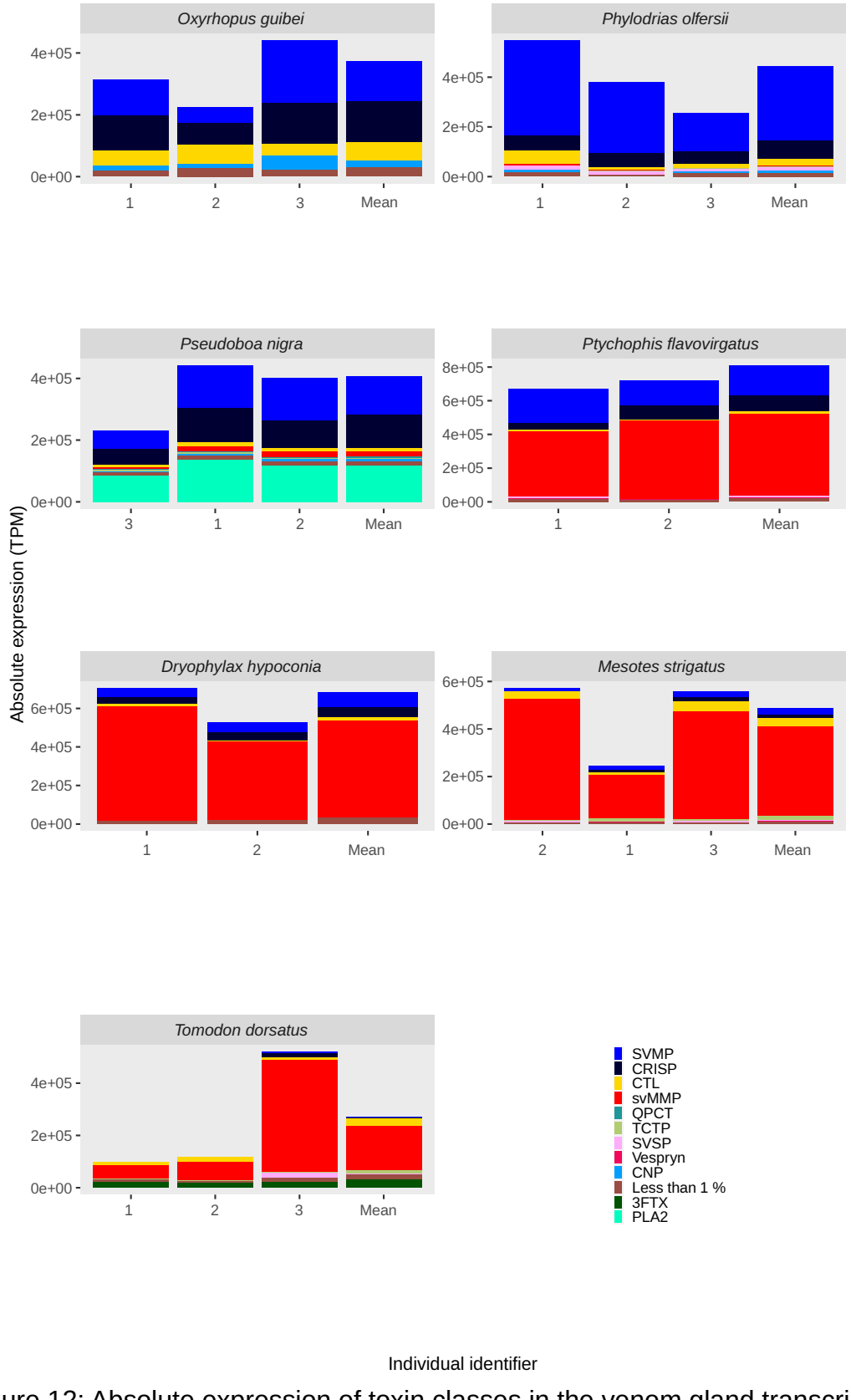


Figure 12: Absolute expression of toxin classes in the venom gland transcriptomes of outgroup individuals.

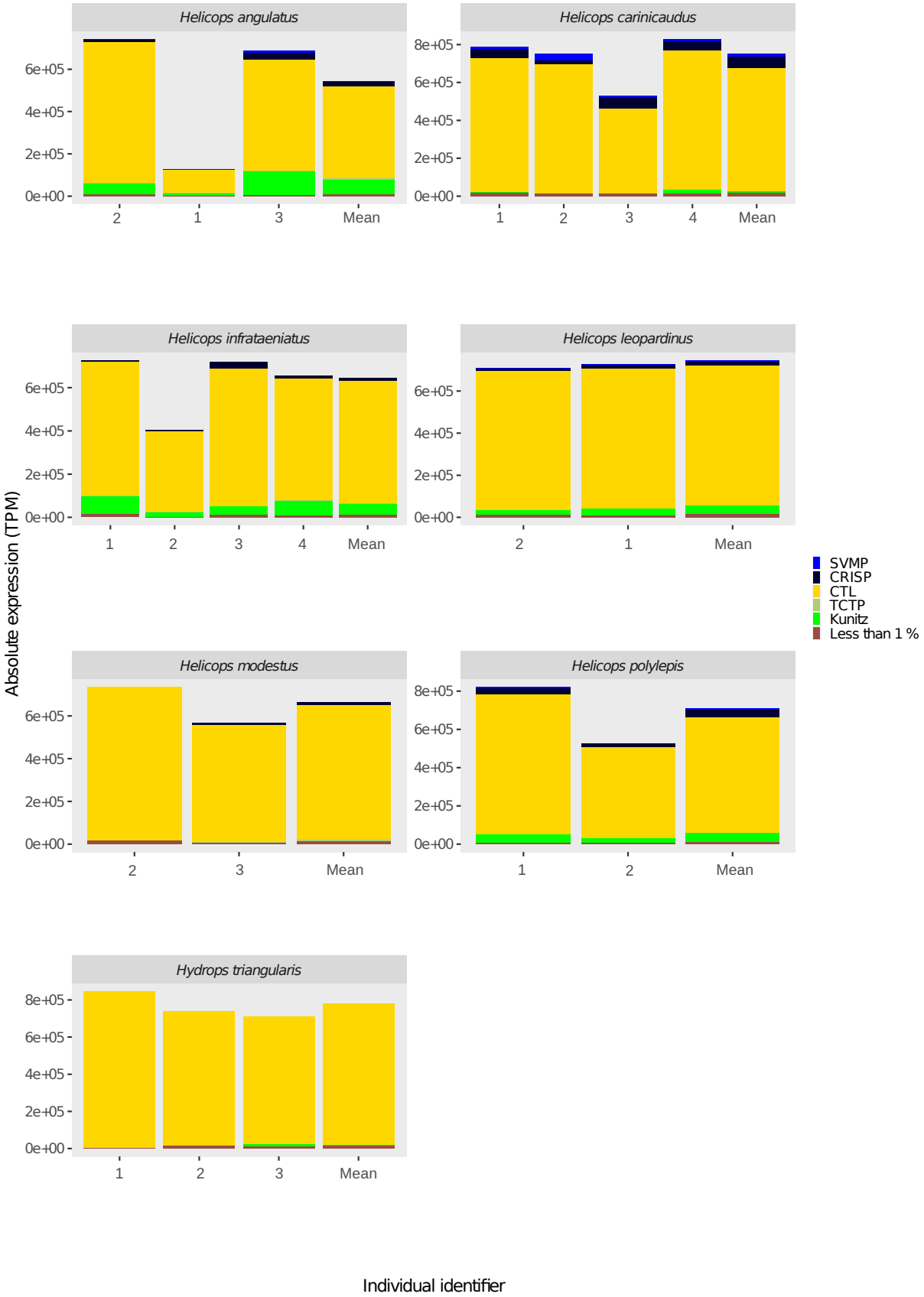


Figure 13: Absolute expression of toxin classes in the venom gland transcriptomes of Hydropsini individuals.

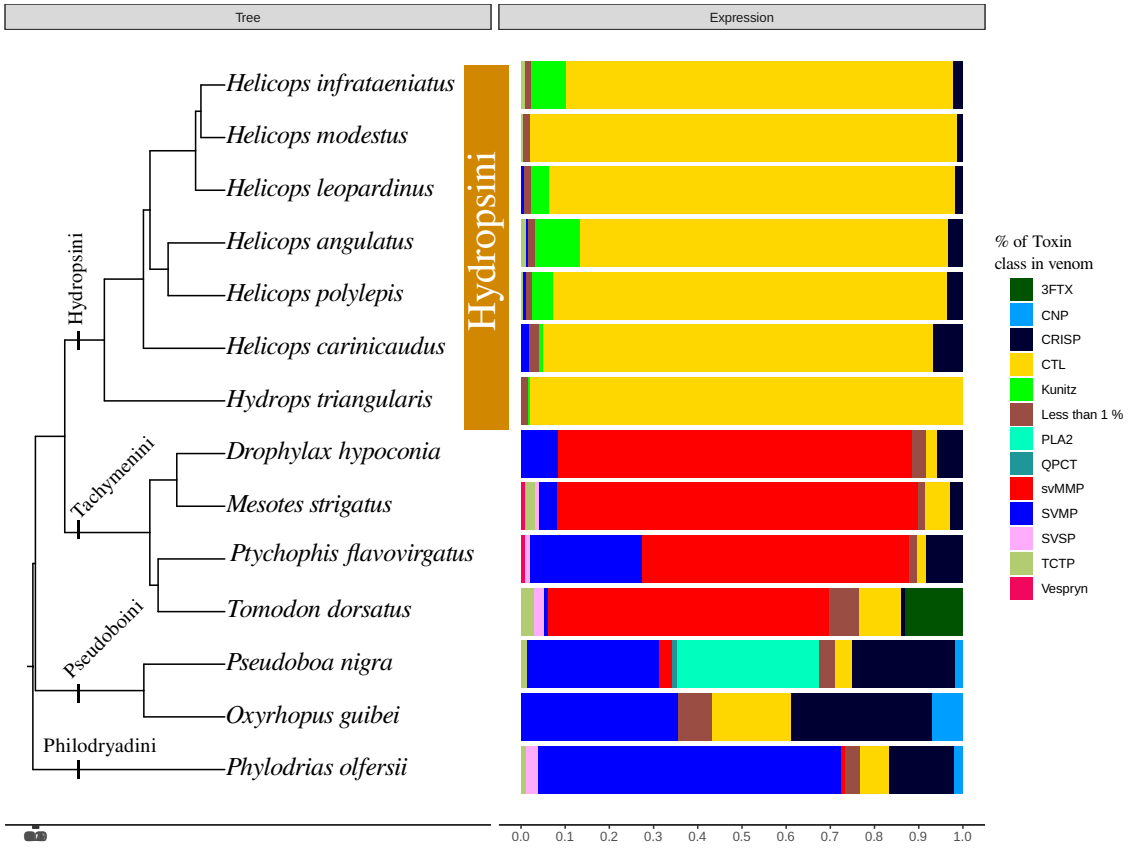


Figure 14: Left:Species tree for Hydropsini, Tachymenini, Pseudoboini and Philodryadini. Species not used in the quantification analyses were removed for clarity. Right: Average relative expression of toxin classes for each species. Less than 1%: Toxins with less than 1% of the total toxin expression.

More details on the patterns of venom expression were obtained by inspecting the expression of the toxin orthogroups (Figure 14). Three orthogroups retained most of the CTL expression in Philodryadini, Pseudoboini and Tachymenini (CTL0, CTL5 and CTL1). Also, one of these orthogroups (CTL0) contained highly expressed transcripts in most Hydropsini (Figure 14), with the exception of *Helicops angulatus* (presenting higher expression of CTL12 transcripts). Other CTL Orthogroups with considerable expression were the CTL2 (in *Helicops* carinicaudus) and CTL21 (in *Hydrops triangularis*).

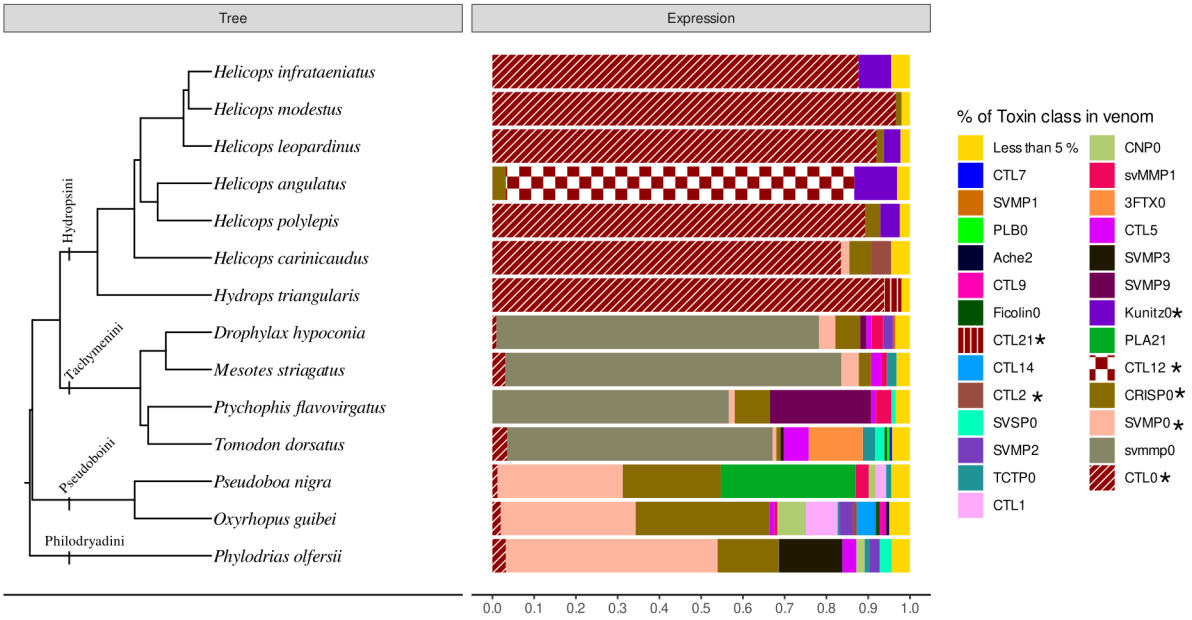


Figure 15: Left: Species tree for Hydropsini, Tachymenini, Pseudoboini and Philodryadini. Species not used in the quantification analyses were removed for clarity. Right: Average relative expression of toxin Orthogroups for each species. Less than 5%: Orthogroups summing 5% or less than the total toxin expression. Asterisks depict orthogroups with minimal expression within Hydropsini.

The use of ONT long-reads provided transcript quantification alternatives to Illumina. Although this data was restricted to the few samples sequenced by ONT, the long-reads offered a read-mapping quantification with a reasonable number of reads and reduced multi-mapping rate. Positive correlation between Illumina and ONT quantification of expression levels in toxin orthogroups was present in all sampled individuals (Figure 15), being strongly determined by the agreement in the high expression of the orthogroups CTL0, CTL12 and CTL21. Without CTL0, CTL12, and CTL 21 correlation was weaker (Figure 15). ONT-based quantification did not support the expression of a second CRISP orthogroup in *He. carinicaudus* and *He. modestus* (Figure 15).

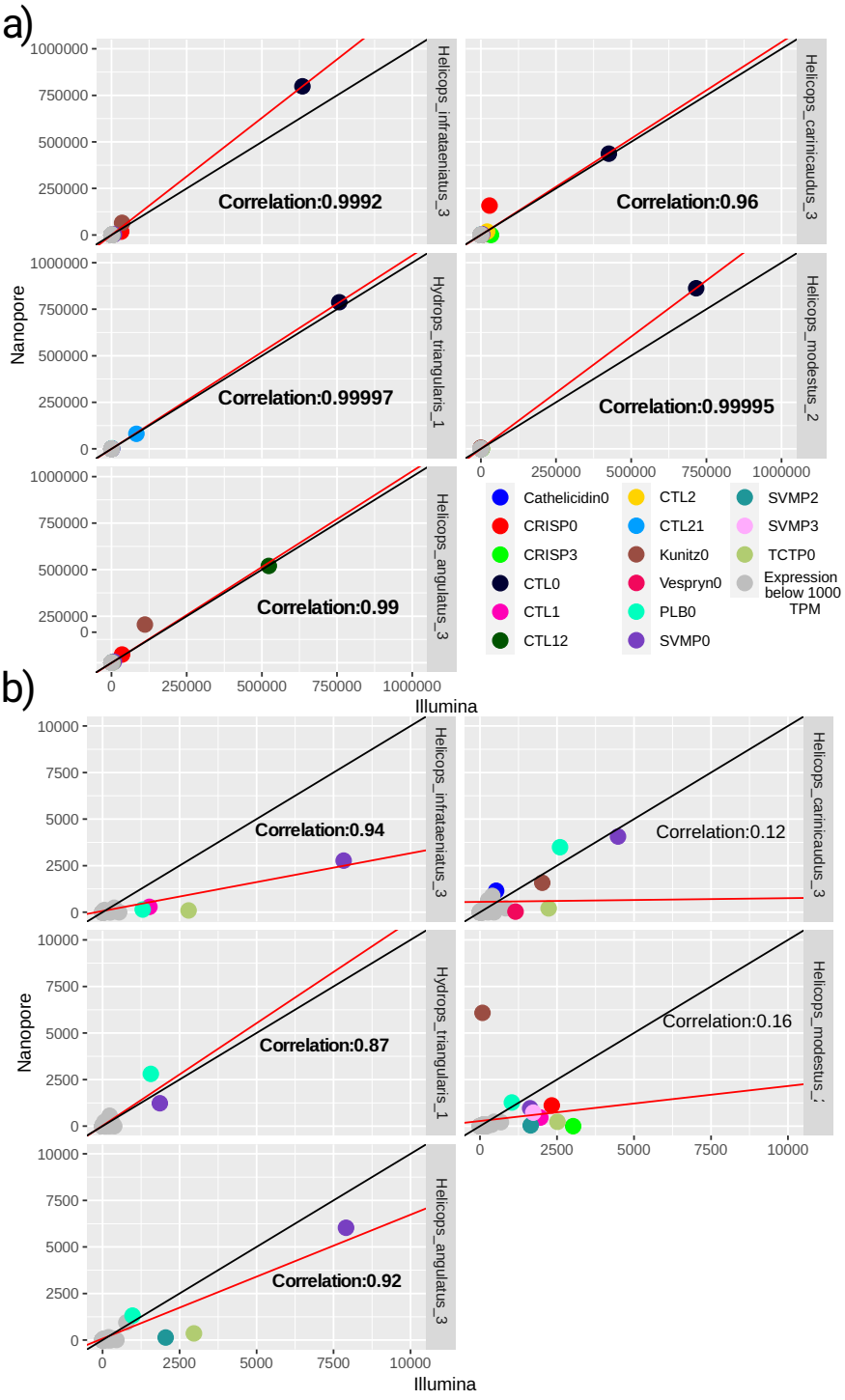


Figure 16: Correlation of expression between Illumina and ONT quantification analyses for toxin orthogroups. Values correspond to TPM. a) All toxins. b) Toxins with 10.000 or less TPM. Toxins with less than 1000 TPM in Illumina or in Nanopore are colored in gray. Black line is the 1:1 correlation, red line is the measured correlation. Bold: p-value <0.01.

The first two components of the PCA analysis of Hydropsini and outgroup individuals explained almost 49% of the variance in the data (Figure 16). The principal

component 1 (PC1) distinguished phenotypes rich in svMMP (Tachymenini), rich in CTL and Kunitz (Hydropsini) and an intermediate group (*Philodryas olfersii* and Pseudoboini) (Figure 16). Differences between Pseudoboini and *Philodryas* olfersii are well explained by the PC2 axis.

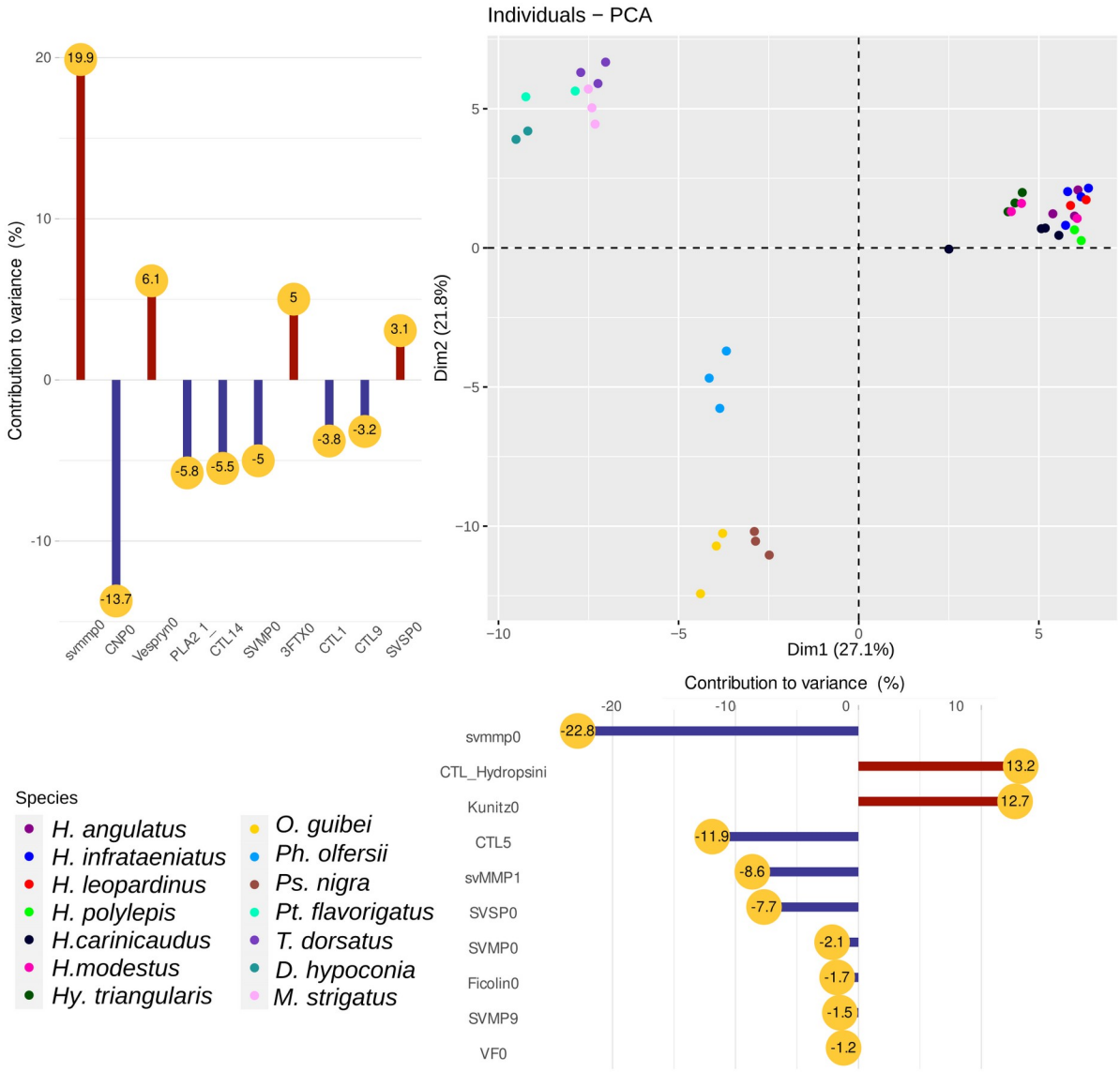


Figure 17: PCA of toxin expression. Lollipop plots represent the contribution of the top toxin to each PC. CTL Hydropsini corresponds to CTL0, CTL12 and CTL21, which are highly expressed in Hydropsini.

4.7 Proteomic experiments

Venom samples were subjected to HPLC-MS/MS shotgun analyses for a proteomic investigation of the venom. Using MaxQuant, at least 4564 MS1 and 34009 MS2 were acquired by the tandem mass spectrometer. The mean rate of peptides identification of MS2 was approximately 14%, yielding an average of 786 peptides belonging to an average of 170 protein groups, excluding decoys and human contaminants (Table 10). These protein groups consist in venom proteins and proteins from other tissues collected along with the venom. The sample with the lowest ratio of identified MS2 was SB1946, from *Hydrops triangularis* (less than 1% of identified MS/MS).

Table 10: Summary of shotgun proteomic identification results. Some identified proteins correspond to endogenous proteins associated with other tissues collected along with the venom.

Sample	MS	MS/MS	MS/MS identified	MS/MS identified [%]	Peptide sequences identified	Protein Sequences identified (excluding Contaminants and Decoys)
*Helicops infrataeniatus*_1	4874	40978	3756	8.6	445	86
*Helicops polylepis*_2	4684	44738	7519	16	1084	219
*Helicops carinicaudus*_2	4626	43889	8634	19	661	109
*Helicops carinicaudus*_4	4649	44637	8569	18	921	190
*Helicops angulatus*_2	4618	44527	7416	16	1067	245
*Helicops angulatus*_3	4564	44688	6956	15	711	164
*Helicops modestus*_2	4822	42073	8364	19	1320	286
*Hydrops triangularis*_3	5678	34009	130	0.34	79	64

MaxQuant was used to confirm the presence of toxin classes in the venom. In most samples, it was possible to identify toxin classes such as 3-Finger-Toxins, CTL, CRISP, SVMP, Kunitz. Phospholipase-B (PLB) and Glutaminyl-peptide cyclotransferase (QPCT), among others (Table 11-17), adding up to 18 different toxin classes across all samples.

Table 11: Toxins identified in the venom proteome of each sample. QPCT: Glutaminyl-peptide cyclotransferase, Hyal:Hyaluronidase, Ache:Acetylcholinesterase, CNP:C-type natriuretic peptide, VEGF:Vascular Endothelial Growth Factor, TCTP: Translationally controlled tumor protein, VF:Venom Factor.

Toxin_class	*Helicops infrataeniatus* 1	*Helicops polylepis* 2	*Helicops carinicaudus* 2	*Helicops angulatus* 2	*Helicops carinicaudus* 4	*Helicops angulatus* 3	*Helicops modestus* 2	*Hydrops triangularis* 3
CTL	X	X	X	X	X	X	X	X
SVMP	X	X	X	X	X	X	X	X
CRISP	X	X	X	X	X	X	X	
PLB	X	X	X	X	X	X	X	
QPCT	X	X	X	X	X	X	X	
Kunitz	X	X	X	X		X		X
Hyal	X		X			X	X	
Ache			X		X			
CNP		X					X	
VEGF	X						X	
Vespryn	X					X		
3FTX	X							
Cystatin		X						
SVSP				X				
TCTP							X	
VF			X					

Concentration estimations through proteomic ruler revealed different venom phenotypes among the species. Perseus's proteomic ruler suggests that venoms of *Hydrops triangularis* and *Helicops modestus* presented a high concentration of CTL, as in the transcriptome analyses. *Helicops infrataeniatus*, on the other hand, presented a high abundance of CRISPs. *Helicops angulatus* and *Helicops polylepis* presented a venom rich in Kunitz. One of the *Helicops carinicaudus* individual's venom was shown to have a venom majorly composed of SVMP (Figure 17).

Table 12: Protein groups containing toxins identified by maxquant in the HPLC-MS/MS analyses for the venom of *Helicops infrataeniatus*. These groups are formed by proteins containing shared peptides. Orthogroups column correspond to the orthogroups from all proteins within the group.

Sample	Orthogroups	ORF Sequence coverage [%]	ORF Sequence length	Intensity	MS/MS count
Helicops infrataeniatus_1	CTL0;	44.6	175	6.55E+08	90
Helicops infrataeniatus_1	SVMP3;	54	276	2.01E+07	24
Helicops infrataeniatus_1	PLB0;	52.6	557	7.25E+08	366
Helicops infrataeniatus_1	SVMP1;	3.2	621	6.02E+05	3
Helicops infrataeniatus_1	SVMP0;	39.3	506	1.11E+08	81
Helicops infrataeniatus_1	SVMP0;	37.5	614	5.50E+08	280
Helicops infrataeniatus_1	SVMP3;	25.3	431	6.24E+06	15
Helicops infrataeniatus_1	3FTX0;	17.5	80	1.01E+06	7
Helicops infrataeniatus_1	Vespryn0;	37.1	213	5.11E+07	43
Helicops infrataeniatus_1	VEGF0;	27.3	216	1.86E+07	6
Helicops infrataeniatus_1	CRISP0;CRISP1;	81.6	239	5.93E+09	536
Helicops infrataeniatus_1	Eppin0;	41.8	194	6.92E+07	25
Helicops infrataeniatus_1	QPCT0;	64.9	368	8.05E+08	216
Helicops infrataeniatus_1	Hyal0;	36.7	447	1.39E+08	146
Helicops infrataeniatus_1	SVMP0;	13.9	617	6.55E+06	3
Helicops infrataeniatus_1	CTL0;	24.6	175	6.58E+06	3
Helicops infrataeniatus_1	CTL0;	42.8	159	1.20E+08	34
Helicops infrataeniatus_1	Kunitz0;	11.9	84	4.22E+07	13

Table 13: Protein groups containing toxins identified the venom of *Helicops polylepis.* Legend goes as in table 12.

Sample	Orthogroups	ORF Sequence coverage [%]	ORF Sequence length	Intensity	MS/MS count
Helicops polylepis_2	CTL0;	64.4	174	5.27E+08	66
Helicops polylepis_2	CTL0;	61.8	178	2.24E+07	15
Helicops polylepis_2	SVMP0;	54.5	617	2.37E+09	250
Helicops polylepis_2	PLB0;	45	551	9.82E+08	291
Helicops polylepis_2	SVMP0;	30.3	555	2.86E+08	35
Helicops polylepis_2	CTL0;	60.9	179	9.98E+08	201
Helicops polylepis_2	SVMP3;	32.4	398	7.69E+08	206
Helicops polylepis_2	CTL0;	69.5	164	1.28E+08	18
Helicops polylepis_2	CTL6;	73.9	165	7.32E+08	132
Helicops polylepis_2	Cystatin0;	19.1	115	9.14E+06	1
Helicops polylepis_2	CTL0;	55.7	158	1.78E+08	24
Helicops polylepis_2	CTL0;	56	159	5.23E+08	51
Helicops polylepis_2	Kunitz0;	38.8	85	3.38E+09	266
Helicops polylepis_2	CRISP0;CRISP1;	65.3	239	1.44E+10	1195
Helicops polylepis_2	SVMP0;	35.1	555	2.21E+08	105
Helicops polylepis_2	SVMP1;	12.9	482	7.64E+06	12
Helicops polylepis_2	SVMP1;	6.8	621	3.37E+06	7
Helicops polylepis_2	CNP0;	14.3	63	6.94E+06	0
Helicops polylepis_2	Cystatin1;	8.8	136	7.78E+05	1
Helicops polylepis_2	QPCT0;	61.7	368	3.46E+08	175

Table 14: Protein groups containing toxins identified the venom of *Helicops carinicaudus.* Legend goes as in table 12.

Sample	Orthogroups	ORF Sequence coverage [%]	ORF Sequence length	Intensity	MS/MS count
Helicops carinicaudus_2	CRISP0;CRISP3;	63.8	224	1.24E+08	37
Helicops carinicaudus_2	PLB0;	43.3	550	1.23E+08	144
Helicops carinicaudus_2	Eppin0;	20	195	5.13E+06	16
Helicops carinicaudus_2	CTL2;	28.8	163	2.02E+08	108
Helicops carinicaudus_2	QPCT0;	37.5	368	2.54E+07	65
Helicops carinicaudus_2	VF1;	25.7	323	6.63E+06	11
Helicops carinicaudus_2	Hyal0;	8.5	447	2.20E+06	5
Helicops carinicaudus_2	Ache0;	3	606	1.06E+06	5
Helicops carinicaudus_2	SVMP0;	58.1	613	3.97E+09	1418
Helicops carinicaudus_2	SVMP0;	30.4	611	2.90E+08	198
Helicops carinicaudus_2	SVMP0;	41.4	607	2.93E+08	214
Helicops carinicaudus_2	SVMP0;	52.1	612	2.02E+09	502
Helicops carinicaudus_2	CRISP0;	86.2	239	8.94E+08	545
Helicops carinicaudus_2	CTL0;	46.5	159	7.49E+07	39
Helicops carinicaudus_2	Kunitz2;	41	368	1.51E+07	42
Helicops carinicaudus_2	CTL5;	37.4	155	5.14E+07	44
Helicops carinicaudus_2	CTL0;	51.1	180	1.67E+08	153
Helicops carinicaudus_2	SVMP0;	29.8	614	1.48E+08	156
Helicops carinicaudus_2	SVMP0;	23.5	613	2.47E+07	24
Helicops carinicaudus_2	SVMP0;	39.2	617	3.63E+09	1202
Helicops carinicaudus_2	VF0;	20	575	3.32E+07	85
Helicops carinicaudus_2	Ache1;	6.4	562	4.12E+06	3
Helicops carinicaudus_4	CRISP0;CRISP3;	67.8	239	2.80E+09	662
Helicops carinicaudus_4	PLB0;	46.4	550	3.11E+08	238
Helicops carinicaudus_4	SVMP0; OG0000075 (Not_identified)	8.8	204	1.23E+06	2
Helicops carinicaudus_4	CTL2;	27	163	3.15E+08	104
Helicops carinicaudus_4	CTL0;	64.4	177	1.24E+09	416
Helicops carinicaudus_4	CTL2;	23.3	163	8.17E+07	39
Helicops carinicaudus_4	QPCT0;	35.9	368	4.35E+07	72
Helicops carinicaudus_4	VF1;	25.7	323	2.52E+07	13
Helicops carinicaudus_4	Hyal0;	2	447	1.03E+06	2
Helicops carinicaudus_4	Ache0;	17.8	606	1.37E+07	18
Helicops carinicaudus_4	CTL0;	57.3	164	2.61E+06	6
Helicops carinicaudus_4	SVMP0;	22.7	611	2.78E+08	131
Helicops carinicaudus_4	SVMP0;	18.8	612	7.27E+06	8
Helicops carinicaudus_4	CRISP0;	67.4	239	2.21E+08	87
Helicops carinicaudus_4	SVMP0;	24.9	614	1.60E+08	140
Helicops carinicaudus_4	SVMP0;	32.3	613	2.93E+07	36
Helicops carinicaudus_4	SVMP0;	32.3	610	7.77E+08	242
Helicops carinicaudus_4	SVMP0;	20.4	617	2.05E+07	31
Helicops carinicaudus_4	CTL13;	15.3	170	3.44E+06	8
Helicops carinicaudus_4	CTL0;	47.6	164	6.80E+07	53
Helicops carinicaudus_4	VF0;	35	575	2.12E+08	117
Helicops carinicaudus_4	CTL0;	57.9	164	7.23E+06	9
Helicops carinicaudus_4	Ache1;	18.7	562	1.97E+08	25
Helicops carinicaudus_4	SVMP1;	13.9	613	8.07E+06	17
Helicops carinicaudus_4	SVMP0;	17.4	610	1.15E+06	4
Helicops carinicaudus_4	SVMP0;	40.4	612	2.55E+08	153
Helicops carinicaudus_4	Kunitz0;	36.9	84	4.79E+08	101

Table 15: Protein groups containing toxins identified the venom of *Helicops angulatus.* Legend goes as in table 12.

Sample	Orthogroups	ORF Sequence coverage [%]	ORF Sequence length	Intensity	MS/MS count
Helicops angulatus_2	SVMP0;	27.1	613	1.05E+08	67
Helicops angulatus_2	SVMP0;	10.4	614	7.28E+06	11
Helicops angulatus_2	CTL7;	21.1	175	1.09E+07	9
Helicops angulatus_2	CTL0;	33.1	160	7.45E+06	3
Helicops angulatus_2	SVMP0;	26.6	617	2.73E+08	97
Helicops angulatus_2	CTL12;	54.2	179	8.75E+07	62
Helicops angulatus_2	CTL0;	18	161	6.20E+06	11
Helicops angulatus_2	CRISP0;	69.9	239	1.38E+09	426
Helicops angulatus_2	SVMP1;	3.1	621	1.37E+06	4
Helicops angulatus_2	SVSP1;	21.3	127	2.06E+06	4
Helicops angulatus_2	CRISP3;	46.4	224	1.34E+08	70
Helicops angulatus_2	CTL5;	20	155	9.43E+06	2
Helicops angulatus_2	Kunitz0;	39.6	96	5.53E+08	109
Helicops angulatus_2	CTL12;	30	180	5.35E+06	10
Helicops angulatus_2	CTL12;	26.1	180	1.59E+07	9
Helicops angulatus_2	QPCT0;GPCP1;	10.8	240	5.98E+05	2
Helicops angulatus_2	PLB0;PLB1;	7.3	551	2.42E+06	6
Helicops angulatus_2	SVMP0;	17	613	1.64E+07	21
Helicops angulatus_2	CTL1;	24.9	185	1.03E+07	16
Helicops angulatus_2	Kunitz0;	45.2	84	3.01E+08	102
Helicops angulatus_2	CRISP3;	30.7	212	1.97E+08	27
Helicops angulatus_2	SVMP0;SVMP1;	2.9	620	1.03E+07	13
Helicops angulatus_3	SVMP0;	34.7	613	1.27E+09	294
Helicops angulatus_3	SVMP0;	37.7	613	1.79E+09	344
Helicops angulatus_3	CTL12;	39.7	179	2.15E+07	17
Helicops angulatus_3	CTL0;	51.2	160	6.18E+07	31
Helicops angulatus_3	SVMP0;	35.3	617	2.71E+08	62
Helicops angulatus_3	CTL12;	35.8	179	8.76E+06	10
Helicops angulatus_3	CTL0;	36	161	1.99E+07	14
Helicops angulatus_3	Hyal0;	20.4	447	1.38E+07	33
Helicops angulatus_3	CRISP3;	25	224	0.00E+00	0
Helicops angulatus_3	QPCT0;	49.7	368	1.36E+08	163
Helicops angulatus_3	PLB0;	33.2	551	6.91E+07	99
Helicops angulatus_3	Vespryn0;	7.6	185	1.18E+06	2
Helicops angulatus_3	CTL12;	63.7	179	6.77E+07	97
Helicops angulatus_3	CRISP0;	61.9	239	1.22E+09	162
Helicops angulatus_3	SVMP0;	30.3	614	2.12E+08	179
Helicops angulatus_3	CRISP3;	69.3	212	8.05E+09	1234
Helicops angulatus_3	Kunitz0;	38.8	98	2.98E+09	368
Helicops angulatus_3	CTL5;	26.5	155	3.11E+07	34
Helicops angulatus_3	CTL12;	38.3	180	3.11E+07	41
Helicops angulatus_3	CTL12;	59.2	179	3.67E+08	132
Helicops angulatus_3	CTL12;	48.9	178	4.32E+07	28
Helicops angulatus_3	Kunitz0;	24.7	85	5.22E+06	5

Table 16: Protein groups containing toxins identified the venom of *Hydrops triangularis* Legend goes as in table 12.

Sample	Orthogroups	ORF Sequence coverage [%]	ORF Sequence length	Intensity	MS/MS count
Hydrops triangularis_3	CTL4;	13.9	158	1.13E+08	28
Hydrops triangularis_3	SVMP0;	1.5	613	1.82E+06	1
Hydrops triangularis_3	Kunitz2;	3.3	368	1.83E+06	5

Table 17: Protein groups containing toxins identified the venom of *Helicops modestus* Legend goes as in table 12.

Sample	Orthogroups	ORF Sequence coverage [%]	ORF Sequence length	Intensity	MS/MS count
Helicops_modestus_2	SVMP3;	68.2	428	1.56E+09	368
Helicops_modestus_2	CRISP0;CRISP1;	56.9	239	1.74E+09	298
Helicops_modestus_2	CTL0;	60.4	164	2.02E+09	50
Helicops_modestus_2	SVMP0;	41.9	614	3.65E+09	578
Helicops_modestus_2	CTL0;	59.7	159	5.11E+09	188
Helicops_modestus_2	CTL0;	67.1	164	6.10E+09	571
Helicops_modestus_2	Hyal0;	35.6	447	4.22E+08	197
Helicops_modestus_2	CRISP3;	61.6	224	8.46E+06	15
Helicops_modestus_2	CTL0;	63.4	175	2.79E+07	8
Helicops_modestus_2	QPCT0;	64.4	368	1.54E+09	354
Helicops_modestus_2	PLB0;	50.5	556	1.49E+09	400
Helicops_modestus_2	VEGF0;	34	141	5.62E+07	5
Helicops_modestus_2	TCTP0;	28.5	172	1.28E+07	12
Helicops_modestus_2	CTL0;	5.1	157	1.21E+07	10
Helicops_modestus_2	CNP0;	11.4	167	1.77E+06	5

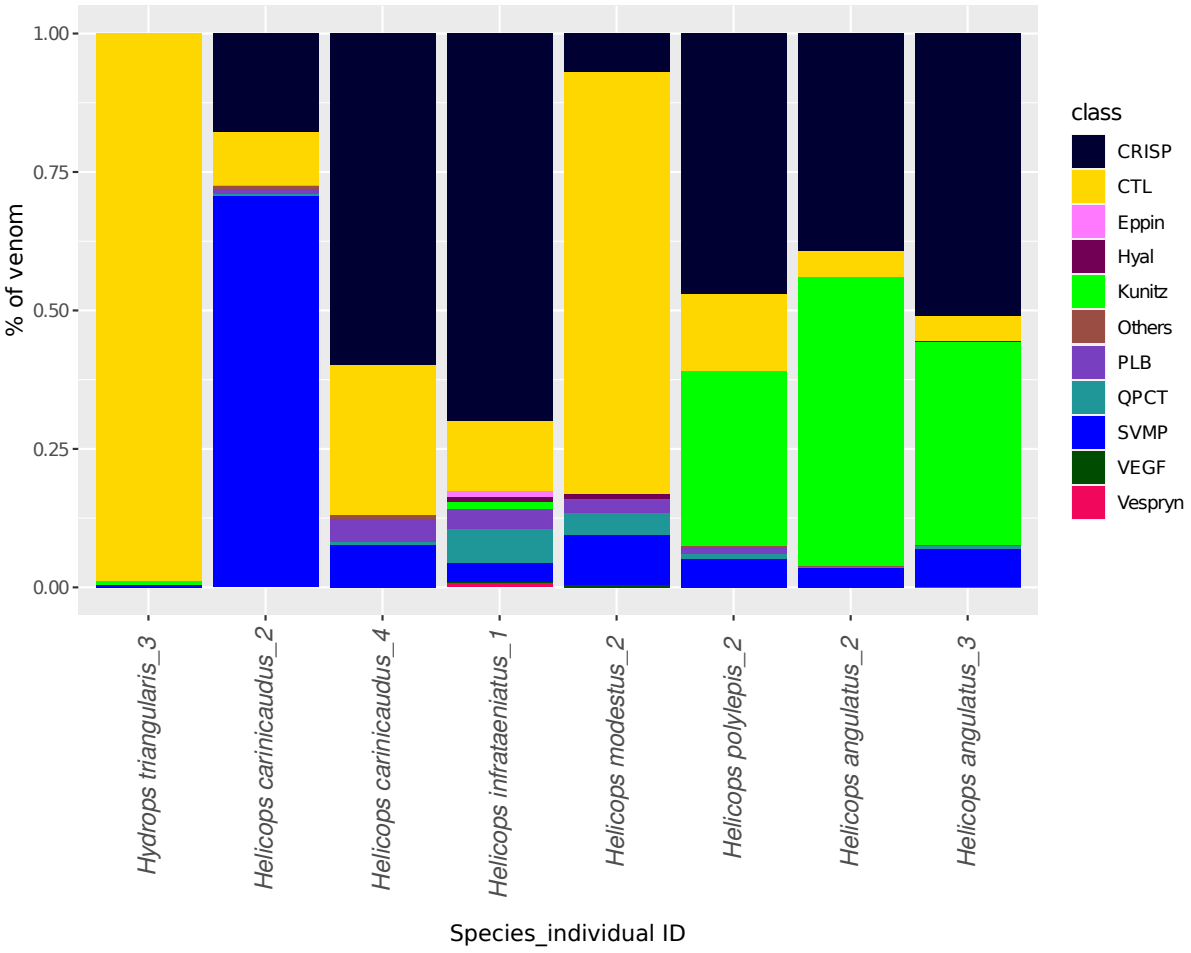


Figure 18: Percentage of each toxin in the venom as estimated by the Perseus proteomic ruler, based on intensity of peptides. Quantification is relative to the total abundance of toxins found in venom.

Venom SDS-PAGEs presented a strong band near 25 kDa (Figure 18, 19) which was identified as being a CRISP (Table 12, Table 13). Other identified toxins identified in the sampled bands were the Kunitz, CTL, SVMP and Hyalurinase (Table 12,13; Figures 19, 20). Many blood contaminant proteins were identified in the bands.

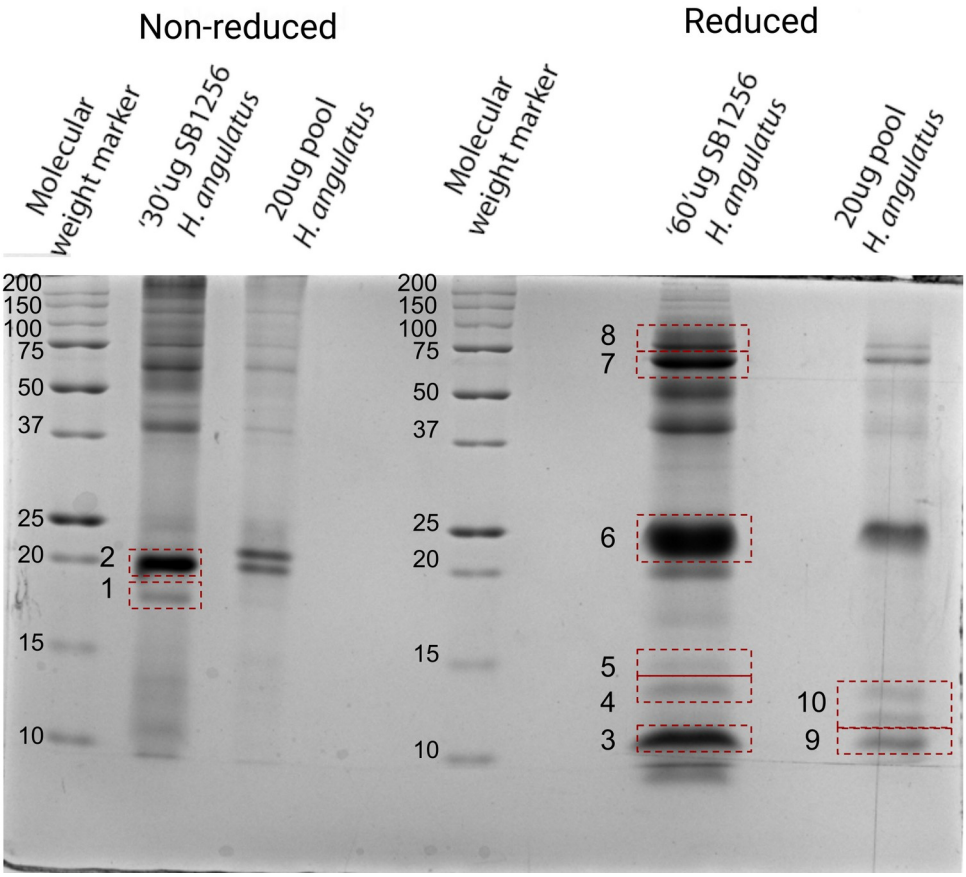


Figure 19: SDS-PAGE of venom from *Helicops angulatus*. Bands within the red boxes were submitted to HPLC-MS/MS. Respective identification is present in Table 12.

Table 18: Identification of proteins obtained in bands excised from SDS-PAGE of the venom from *Helicops angulatus*, shown in figure 18.

Band ID	Toxins identified	Main blood Contaminants Identified
1	CRISP, SVMP (autolysed)	NADH dehydrogenase
2	CRISP	Hemoglobin alpha, Hemoglobin-beta, Apolipoprotein A-I
3	Kunitz	Beta-Hemoglobin, Beta-lactoglobulin
4	None	Hemoglobin alpha, Hemoglobin-beta, Apolipoprotein A-I
5	None	Hemoglobin- Beta, Beta-lactohlobulin, Apolipoprotein A-I, Nudc-domain containing protein
6	CRISP	U5 small nuclear ribonucleoprotein 200 kDa helicase, Apolipoprotein A-I
7	CRISP, SVMP	Ovotransferrin
8	CRISP SVMP	Ovotransferrin
9	Kunitz	Beta-lactoglobulin, Nudc-domain containing protein, Beta-Hemoglobin
10	Kunitz	Beta-lactoglobulin, Hemoglobin Alpha,Nudc-domain containing protein, Beta-Hemoglobin, USP6

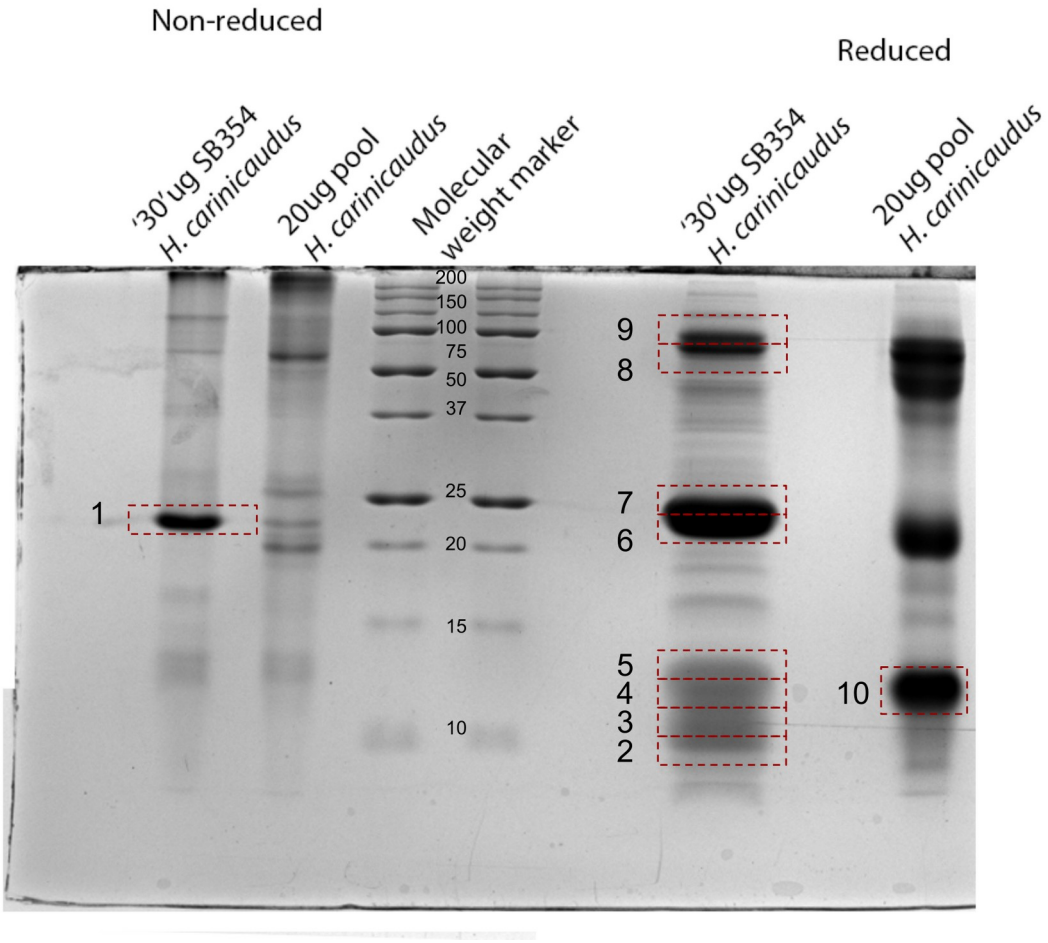


Figure 20: SDS-PAGE of venom from *Helicops carinicaudus*. Bands within the red boxes were submitted to HPLC-MS/MS. Respective identification is present in Table 13.

Table 19: Identification of proteins obtained in bands excised from SDS-PAGE of the venom from *Helicops carinicaudus* shown in figure 19.

Band ID	Toxins identified	Main blood Contaminants Identified
1	CRISP	Hemoglobin Beta, PLA2 inhibitor
2	CTL	Beta lactoglobulin, Hemoglobin-beta, Hemoglobin Alpha
3	CTL, Kunitz	glutaredoxin, Hemoglobin beta
4	CTL, Kunitz	Thioredoxin, Hemoglobin Alpha, Glutaredoxin, proteinS100-A11, Beta-lactoglobulin
5	CTL, Kunitz	Hemoglobin alpha, Hemoglobin-beta
6	CRISP	PLA2 inhibitor
7	CRISP, SVMP	Calbidin
8	SVMP, HYAL	Ovotransferrin, suldhydryl oxidase 1
9	SVMP	Ovotransferrin,hemopexin
10	CTL, Kunitz	Hemoglobin alpha, Hemoglobin-beta, Thierodoxin, Small serum protein 5-like i

4.8 Toxin alignments, Gene Trees and Protein features

4.8.1 The novelties of the highly expressed C-type Lectins

The CTL tree with expressed sequences was constructed with 314 sequences. After trimming, final alignments contained 373 nucleotide positions. The inferred tree presented monophyletic groups that closely reflect the inferred orthogroups (Figure 20). The phylogenetic relationship between these groups could not be precisely determined, as their nodes had a low support value. The analysis supports the monophyly of a clade containing sequences from EMBL classified as CTL-like, with alpha and beta subunits. Together, the orthogroups CTL0, CTL12 and CTL21 form a single monophyletic group containing the (1) CTLs highly expressed in Hydropsini, (2) CTL sequences from the outgroup and (3) the EMBL *C-type lectin like sub-unit A* from *Philodryas olfersii*. Within this clade, a monophyletic group containing only Hydropsini sequences was found (Figure 20). The CTL2, found mildly expressed in *Helicops carinicaudus,* forms a monophyletic group with sequences found in other species, including representatives from the outgroup. The expression of each orthogroup and the Phylogenetic tree can be found in Figure 21.

The sequences assembled using ONT long reads and belonging to the orthogroups CTL0, CTL12 and CTL21 were queried against the PDB database. From this search, 48 sequences had as best hit the *Snaclec agglucetin subunit alpha-1* (PDB entry: 6xfq_A) and 11 had as best hit *Snaclec rhodocytin subunit beta* (PDB entry: 3wwk_K). Comparisons of pairwise similarity of the Dipsadidae CTLs obtained in this work with the *Snaclec agglucetin subunit alpha-1* show that sequences from Hydropsini present a region of low similarity to this toxin, while outgroup sequences kept a constant quality throughout the alignment. The low similarity region was found to match the central loop of the *Snaclec agglucetin subunit alpha-1,* located between the beta strands 4 and 5 (Figure 22). The phylogeny of the CTLs exclusive to Hydropsini is depicted in Figure 23.

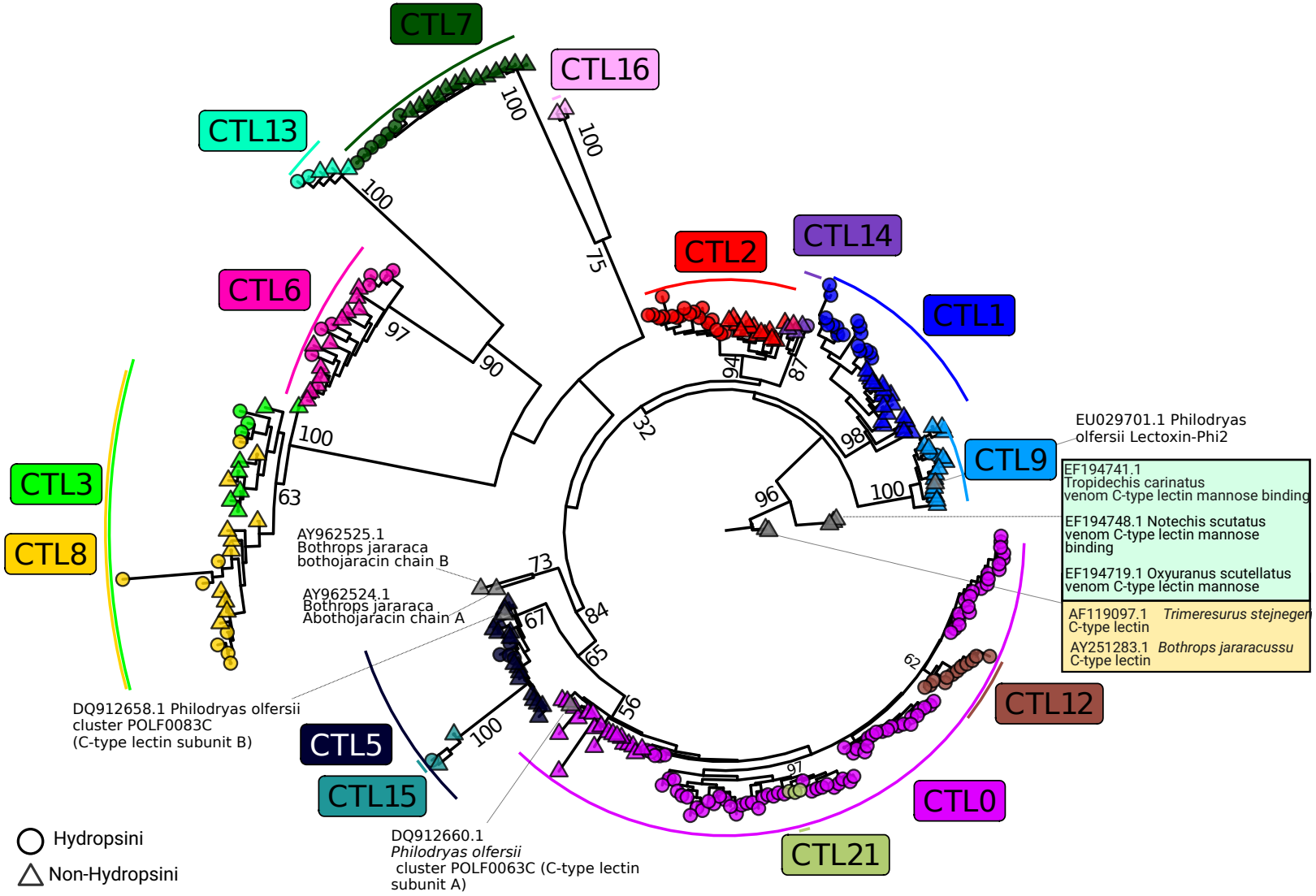


Figure 21: CTL Maximum likelihood gene tree. Circles indicate Hydropsini sequences, triangles correspond to outgroups. Gray triangles were downloaded from EMBL. Only Support values above 30 and outside orthogroups are shown.

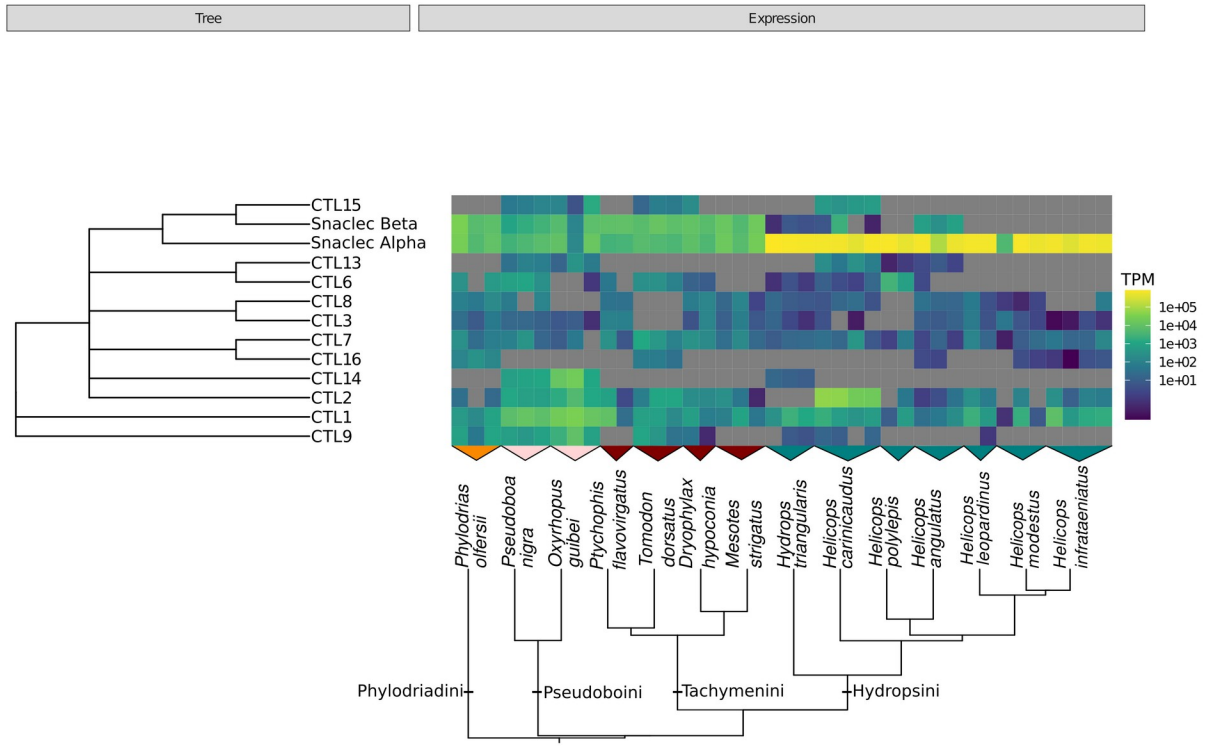


Figure 22: Expression for the different CTL orthogroups. The color scale is in logarithmic scale. CTL-alpha comprises CTL0, CTL12 and CTL21. CTL beta corresponds to CTL5.

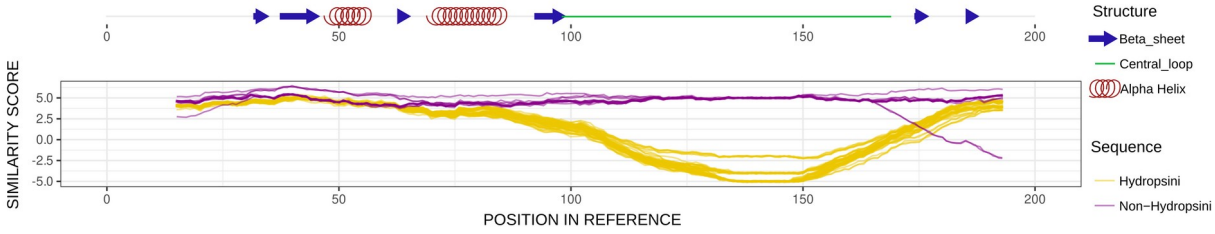


Figure 23: Pairwise similarity score of Dipsadidae CTL transcripts to *Snaclec agglucetin subunit alpha-1.* Non-Hydropsini: Sequences of non-Hydropsini Dipsadidae assembled with Illumina reads and belonging to the Orthogroup CTL0. Hydropsini: Sequences of Hydropsini assembled with ONT and belonging to CTL0.

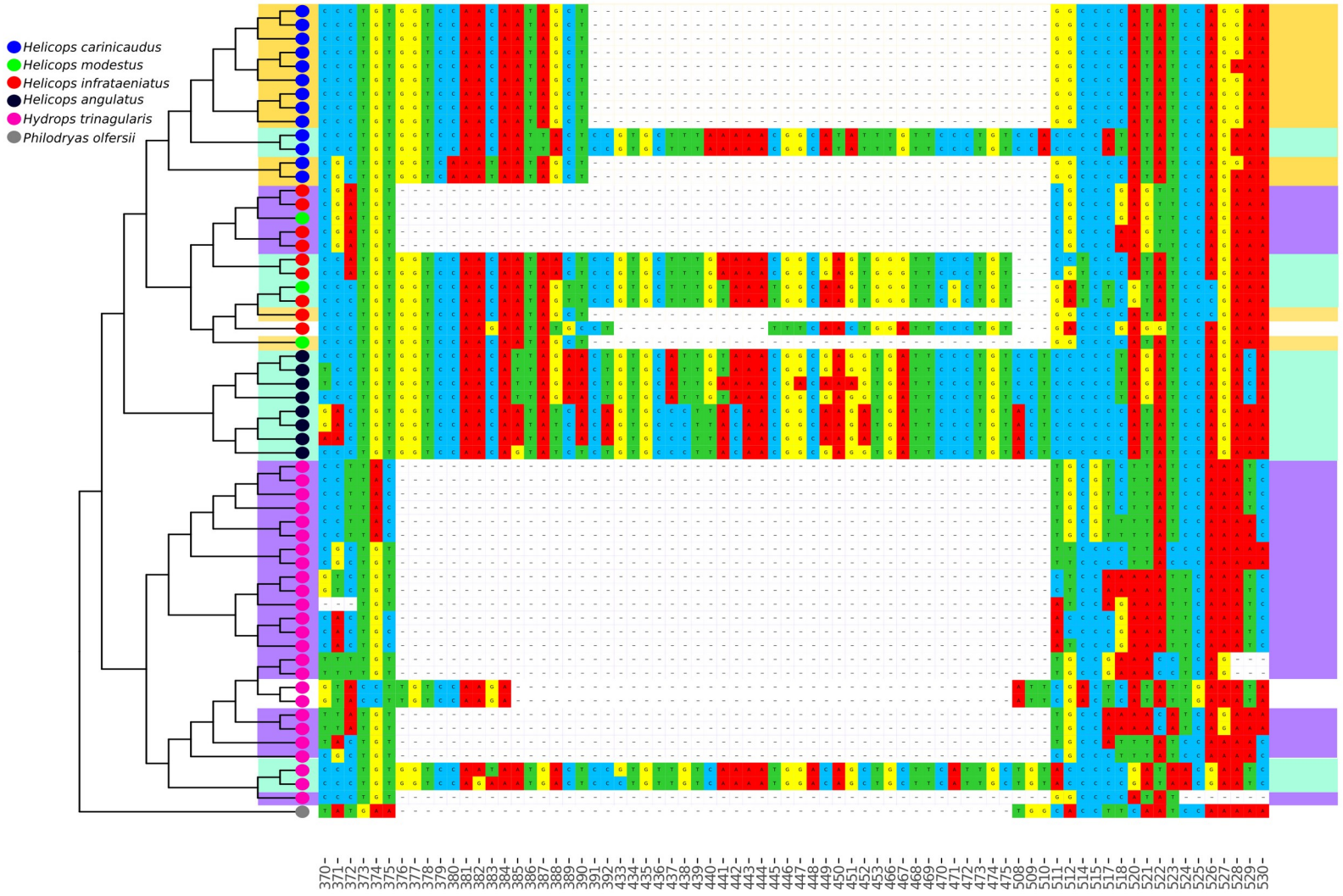


Figure 24: Left: Maximum likelihood cladogram of the entire sequences from CTL exclusive from Hydropsini. Circles indicate the species. Right: Alignment of the central loop of the CTL. Sequences with similar loops had their background painted with the same color.

Three-dimensional structures of the three CTL proteins predicted from ONT assembled transcripts were modeled to prospect the conformational changes resulting from the substitutions and insertions within the central loop. For all models, a similarity of the hydrophobic core to the reference was observed, with the main differences residing in the central loop. The loop of the constructed models presented either alpha-helix or beta sheets structures, in addition to disulfide bonds. None of these features were observed in the structure of canonical CTL alpha subunit (Figure 24).

In Hydropsini, CTL alpha-subunit transcripts with central loops similar to other Dipsadidae could only be retrieved in the Illumina assemblies of two species: *Helicops angulatus* and *Helicops carinicaudus*. Still, the summed TPM expression of those transcripts was at least 500 times lower than the CTLs presenting the insertions (Table 14).

Table 20: Summed TPM values for CTL-like alpha subunit transcripts presenting and not presenting the insertion in the loop. The insertion in the loop was identified by the absence of the ‘QC’ motif in the alignment. Data only shown for Hydropsini species presenting CTL-like alpha subunit with and without the novelty in the loop.

Individual ID	Total TPM - Insertion in loop	Total TPM - no Insertion in loop
Helicops angulatus_2	741,965.56	1,079.27
Helicops angulatus_1	126,733.96	56.33
Helicops angulatus_3	686,545.90	223.77
Helicops carinicaudus_1	786,428.92	700.33
Helicops carinicaudus_2	750,888.85	1,446.71
Helicops carinicaudus_3	526,847.50	541.68
Helicops carinicaudus_4	828,424.11	999.24

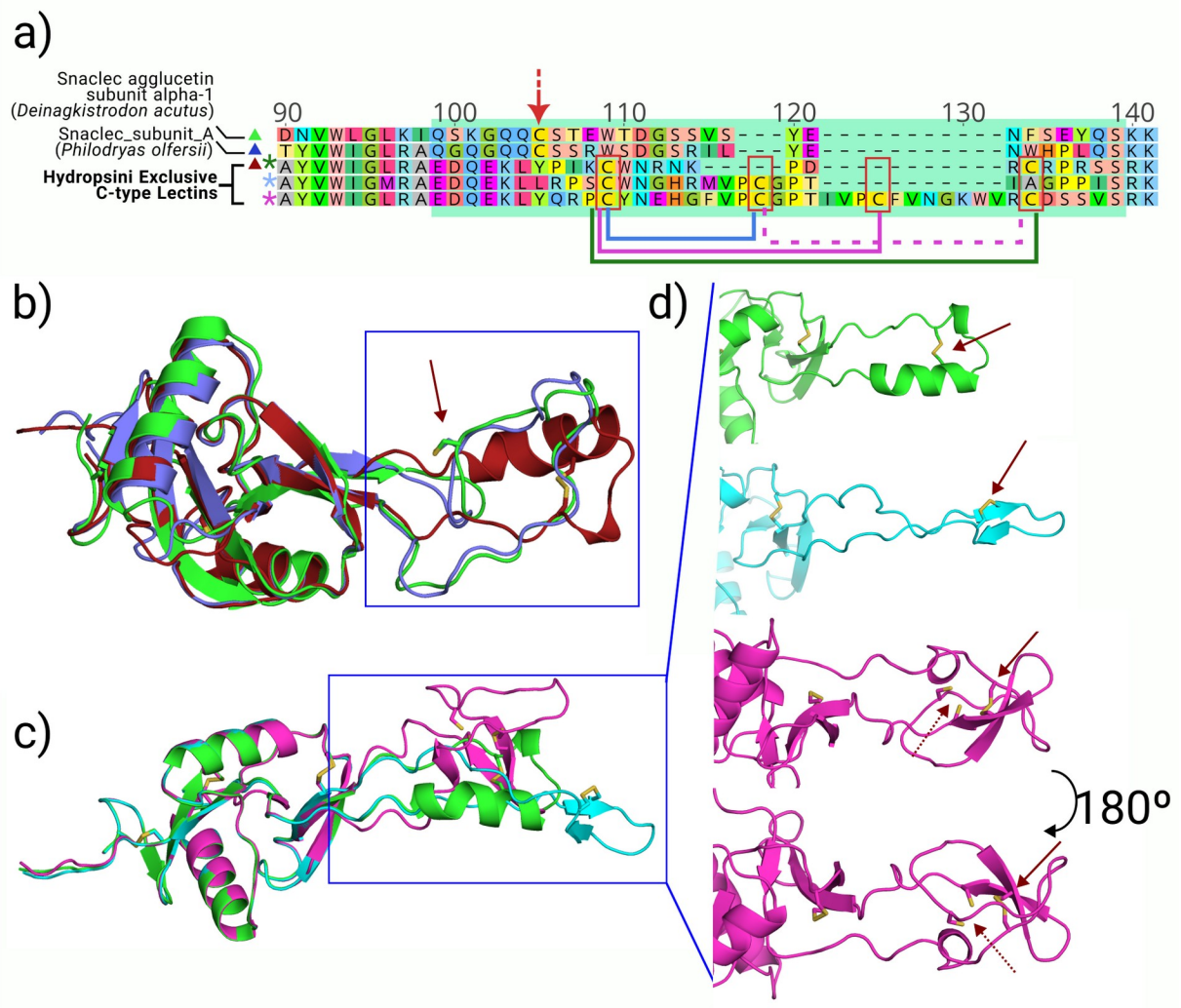


Figure 25: Prediction of three-dimensional structure of Hydropsini CTL. a) Amino-acid sequence alignment of the central loop of the CTL-like and the PDB model. The colors on the left correspond to the generated models. Predicted cysteine bonds for each CTL are indicated below, with matching colors. Green box: central loop. Red arrow: typical QC motif of alpha subunits. b) Alignment of structure of *Snaclec agglucetin subunit alpha-1* (green), *Philodryas olfersii* CTL0 protein (blue) and *Helicops modestus* CTL0 protein (red). Red arrow: Free cysteine that bonds to the beta subunit. Blue box: Central loop. c) Alignment of three *Helicops modestus* CTL0 protein from the same individual. Blue box: Central loop. d) Augmented view of the central loop of the sequences in b). Red arrows: Disulfide bonds. The dotted red arrow in the pink structure indicates two Cysteines physically close to each other.

4.8.2 The Kunitz expressed in Hydropsini present a single domain

Domain identification through HMM revealed 15 different domain configurations in proteins annotated as Kunitz. Among these, five were detected to compose more than 0.1% of all toxin transcripts according to the Illumina quantification: Single Kunitz (Kunitz_BPTI), Double Kunitz (Kunitz_BPTI; Kunitz_BPTI), Ku-Wap fusion (Kunitz_BPTI; WAP), Papilin like (Kunitz_BPTI; WAP; Kunitz_BPTI), and a complex domain configuration presenting three Kunitz and three WAP domains (Kunitz_BPTI; WAP;WAP;WAP; Kunitz_BPTI; Kunitz_BPTI) (Figure 24). Alignment and phylogenetic inference of the isolated Kunitz domains indicates a monophyletic Double Kunitz, with clades matching the Kunitz position in the protein. A similar pattern was also observed for the complex Kunitz-like protein: clades correspond to the position of the Kunitz domain in the protein. However, the Orthogroup containing these proteins was not monophyletic. There was no clear evidence for monophyly of the Kunitz domain in the single Kunitz.

In *Helicops*, Single Kunitz was the most expressed type of Kunitz. For the *Hydrops* genus, individuals presented different profiles of Kunitz expression: one individual barely expressed Kunitz, a second one expressed mostly Single-Kunitz transcripts while the third presented a more diverse Kunitz profile (Figure 25). Quantification by ONT long-reads also supported the expression of the Single Kunitz architecture in *Helicops,* suggesting a much smaller expression of different architectures in the genus. In *Hydrops triangularis,* long-reads quantification suggested prevalence of the ku-wa-wa-wa-ku-ku domain configuration, though the summed expression levels of Kunitz in the individual were less than 0.05% of the whole venom. Kunitz expression in outgroups was found to be much smaller and mostly composed of double Kunitz (Figure 26).

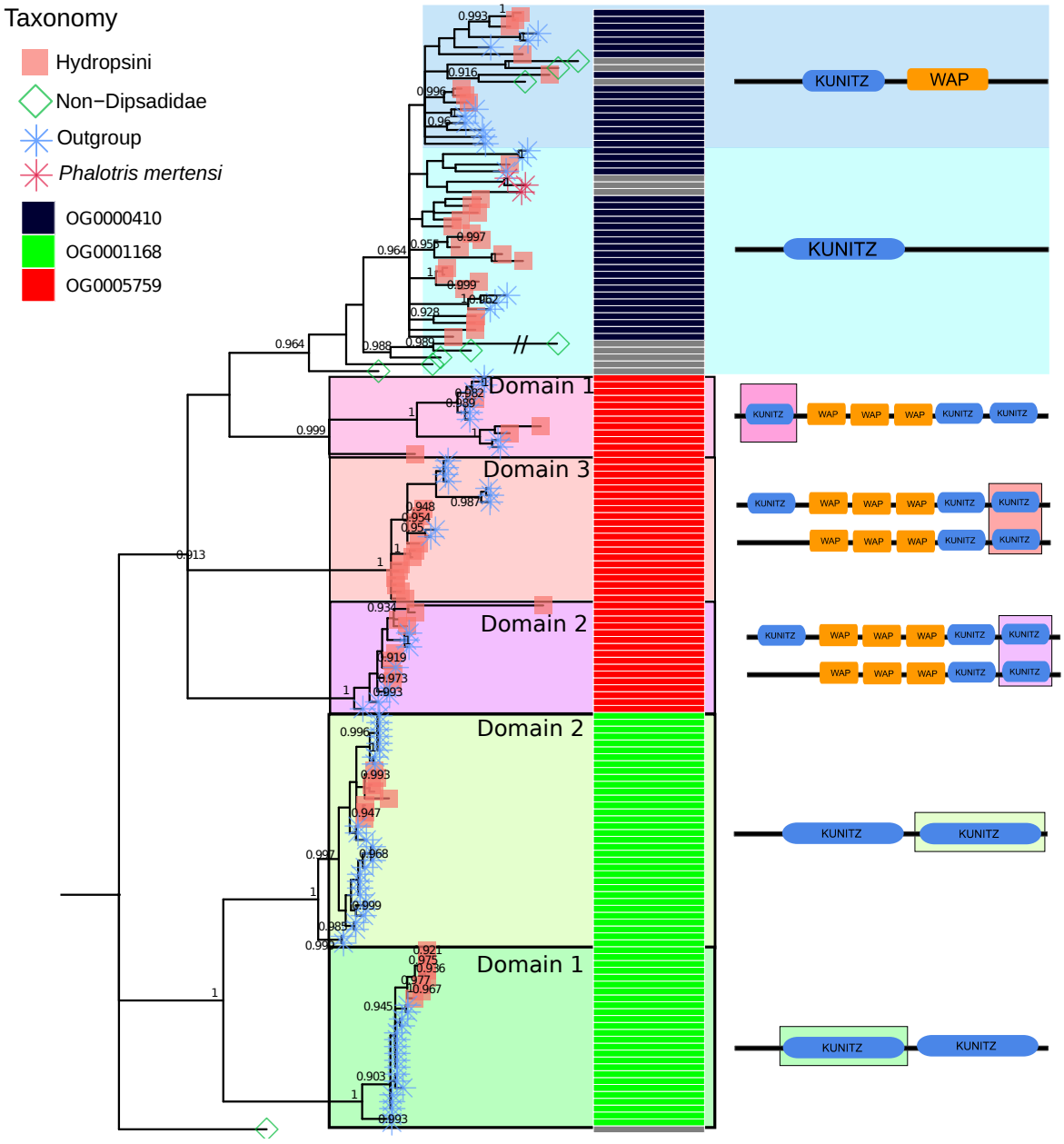


Figure 26: Left: Bayesian tree of Kunitz domains extracted from the CDS. Center: Corresponding orthogroup for each sequence. OG0000410: Kunitz0; OG0001168: Kunitz 1; OG0005759: Kunitz 2. Right: Domain configuration of the protein. Kunitz position in the protein is highlighted when needed.

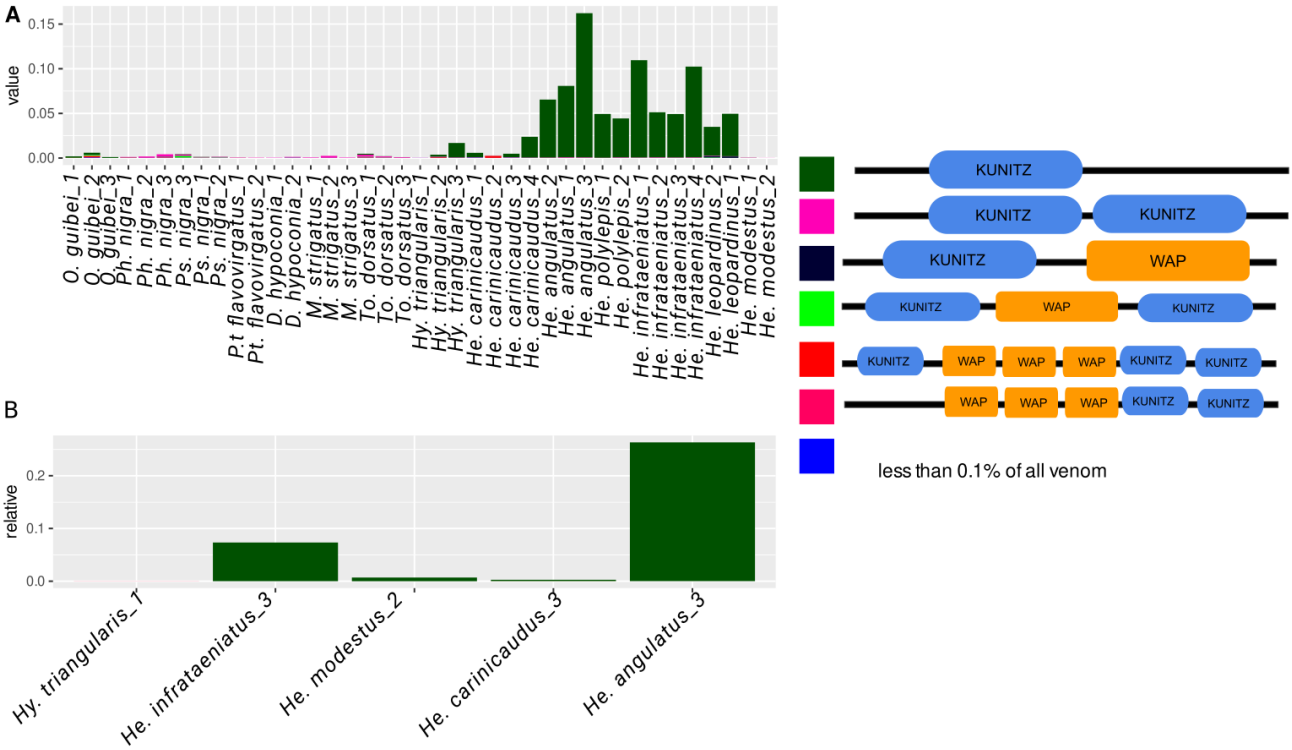


Figure 27: Kunitz expression relative to the total expression of venom. A: Illumina quantification; B: ONT quantification.

5. Discussion

5.1 Phylogenetics

The present work aimed to deeply investigate the venom of the Hydropsini. This tribe is unique among the Dipsadidae, presenting aquatic habits and an enigmatic venom, scarcely studied so far. Furthermore, unraveling the evolution and composition of the venom of Hydropsini may provide key insights into understanding of the venom of Dipsadidae snakes. For such, determining its phylogenetic position and relationship to other Dipsadidae species is a necessary step.

With this in sight, a phylogenetic tree including species from the Xenodontinae tribe was estimated using an algorithm that accounts for discordance in the gene tree (ZHANG *et al.*, 2018). As in (ZAHER *et al.*, 2018), the species tree strongly suggests Tachymenini as sister taxon to Hydropsini (Figure 6). Alternative topologies in literature suggest that, among the sampled tribes, Pseudoboini is the closest to Hydropsini (GRAZZIOTIN *et al.*, 2012; ZAHER *et al.*, 2019). The found topology of the tribe Hydropsini is similar to those found in literature, except for the position of *Helicops carinicaudus*, which was found as sister taxa to the remaining *Helicops* species in the present tree (MORAES-DA-SILVA *et al.*, 2019, 2021).

To obtain a fossil calibrated ultrametric tree, species of distant outgroups had to be sampled, as there are no fossil records of the Xenodontinae sub-family. Although node ages within Xenodontinae seem to be older in the present tree when compared to other calibrated trees (TREVINE *et al.*, 2022; ZAHER *et al.*, 2019), the radiation age of Hydropsini (21,5 Mya; Figure 7) was comparable to biogeographic estimations (CARVALHO, 2022).

5.2 Assemblying toxin transcripts with short and long reads

A short-read transcriptomic approach was used to assemble transcripts coding for toxins. The use of more than one assembler is necessary to obtain a complete set of toxin transcripts (HOLDING *et al.*, 2018), but it also increases the number of badly assembled sequences and redundancies. The high number of paralogs in toxin gene families further hinders appropriate assembly of toxin transcripts with short reads (HOLDING *et al.*, 2018; MACRANDER *et al.*, 2015). In line with that, a substantial proportion of putative toxin was removed due to evidence for misassembles.

In the final reference transcriptomes, CTL was the toxin class with the the most transcripts in Hydropsini, followed by SVMP and Kunitz. These results contrast with previous studies that suggested SVMP as the most diverse transcript in *Helicops* (CERDA *et al.*, 2022).

To rule out the possibility that the distinctive Hydropsini CTL resulted from chimeric transcripts, ONT long-reads were obtained. The CTL transcripts assembled by RNA-bloom2 were poorly supported by the sequencing reads (Figure 9). It is possible that expression of highly similar CTL transcripts may confound RNA-bloom2 strobomer-based digital normalization, which is efficient with dealing with large indels (typical of alternative splicing) (NIP *et al.*, 2023), but not with small (but recurrent) substitutions typical of paralogy. Moreover, a high false discovery rate has been recorded in spike-in cDNA control samples sequenced with ONT and assembled with RNA-bloom2 (NIP *et al.*, 2023). To reassemble highly supported transcripts, an algorithm was designed to cluster reads by transcript, from which consensus sequences were created. The final sequences were better supported by the reads, when compared to those assembled by RNA-bloom2 (Figure 9). Despite its longer running and hands-on time, the pipeline may be a viable approach for refining reference-free assemblies with only long-reads. Sharp differences in the expression levels of similar transcripts may result in misinterpretation of less abundant transcripts as persistent errors, potentially leading to their underrepresentation. Hence, the final set of CTL may be enriched with highly expressed transcripts.

This is, to the knowledge of the author, the first study to investigate the venom gland transcriptome of a Dipsadidae species using long-reads, and these results show how such approach complements the assembly transcripts from multi-gene families.

5.3 Composition of the Venom from Hydropsini

With the curated set of transcripts in hand, quantification of the transcriptomes was obtained by read mapping. Quantification by short reads and long reads agrees that CTL is the most expressed toxin class in Hydropsini venom gland transcriptome (Figure 13), reaching levels never described before in other groups of snakes, to the best of the author's knowledge. High expression of CTL has been described in *Helicops angulatus*, *Helicops polylepis* and *Helicops leopardinus* (CERDA *et al.*, 2022) during the development of this work, and the present study shows that this feature is spread throughout the Hydropsini tribe, including all sampled species of *Helicops* and *Hydrops.* Considering the sampling in this dissertation and other studies of Xenodontinae venom glands (CAMPOS *et al.*, 2016; CHING *et al.*, 2006; SCHRAMER *et al.*, 2022; TIOYAMA *et al.*, 2023), increase in CTL expression seems to have evolved in the common ancestor of all Hydropsini (Figure 13).

When compared to the outgroups, the single domain Kunitz was also found to be highly expressed (Figure 13, 26). This appears exclusively in *Helicops*, being more pronounced in *Helicops angulatus, Helicops polylepis, Helicops infrataeniatus* and *Helicops leopardinus* (Figure 13, 14, 26). High expression of single Kunitz toxins has been recorded in the Dipsadidae *Phalotris mertensi* (CAMPOS *et al.*, 2016), while other previously investigated Dipsadidae presented low levels of single Kunitz expression *(CHING et al.*, 2006; SCHRAMER *et al.*, 2022; TIOYAMA *et al.*, 2023).

Although rough, estimates based on proteomic shotgun provide insights on the venom composition. Here, these estimates differed significantly from the quantification obtained by transcriptomic methodologies. Although CTL was estimated as highly abundant in the venom of *Hydrops triangularis* and *Helicops modestus*, SVMPs, Kunitz and CRISPs were estimated to be in higher concentrations than CTL in other samples (Figure 17). Previous studies have described that the correlation between transcriptome and proteome quantifications may be poor (CASEWELL *et al.*, 2014; DURBAN *et al.*, 2011; TAN *et al.*, 2015a). On the other hand, several studies have found positive correlation between these two approaches (AIRD *et al.*, 2013; MODAHL *et al.*, 2018; ROKYTA *et al.*, 2015).

While it is possible to estimate the venom composition using shotgun proteomics, SDS-PAGE may provide a complementary option to understand the venom composition in a more qualitative way. Bands corresponding to CTLs were identified in the SDS-PAGE of *Helicops carinicaudus* venom (Figure 19, Table 13), but not in the SDS-PAGE of *Helicops angulatus'* venom (Figure 18, Table 12), even though CTL was found to be highly abundant in both species' transcriptome (Figure 13). For both, the strongest band was identified as CRISP (Figure 18, 19, Table 12, 13), most probably corresponding to the neurotoxin *Helicopsin* (ESTRELLA *et al.*, 2011). Even though estimations of protein concentration based on SDS-PAGE visualization may be imprecise, high CRISP abundance was also observed in shotgun analyses, indicatinga critical role for this toxin in *Helicops*, as opposed to what the transcriptomes suggested.

The agreement between shotgun proteomic experiments and SDS-PAGE suggests that differences between proteomes and transcriptomes in this group arise from biological mechanisms, rather than artifacts related to the shotgun quantification methodology. Natural post-transcriptional mechanisms, such as the proteolysis of toxins, could also impact the correlation between transcriptome and proteome (CASEWELL *et al.*, 2014). These would result in weaker bands in SDS-PAGE and a decrease in the number of ionizable peptides in shotgun analyses. If CTL proteins are more susceptible to degradation, then it is possible that their protein levels are maintained by an increased transcript expression. However, an unknown bias of the MS/MS based proteomic analysis, such as a possible low ionization potential of the CTLs cannot be ruled out.

Another possibility is that different toxins families are asynchronously expressed after venom milking in Hydropsini. Since all venom glands were fixed four days after venom extraction, this could explain the homogeneity of the venom gland profiles across individuals from the tribe. On the other hand, as the milked individuals were obtained from nature, determining the precise timing of their last feeding event is challenging, resulting in venom samples from various replenishment stages. In snakes, there is weak evidence for asynchronous expression of toxin gene families after venom extraction. In *Bitis arietans*, although differences were observed in toxin expression profile over time, different toxin families appear to be synthesized in parallel, while the venom composition during the synthesis resembles that of mature venom (CURRIER *et al.*, 2012). For *Vipera palestinae*, asynchronous synthesis of toxin gene families was considered the cause of differential activity of the venom in different stages of venom resynthesis (BROWN *et al.*, 1975;

ORON *et al.*, 1978). Meanwhile, for *Bungarus multicinctus*, the venom gland expression profile was constant throughout the synthesis cycle (YIN *et al.*, 2020).

The study of Hydropsini venom presents many challenges. The yield of venom is very low for each animal, and pilocarpine does not increase secretion volume, different to what is observed in other snakes (HILL; MACKESSY, 1997; MORAIS-ZANI *et al.*, 2018; ROSENBERG, 1992; ROSENBERG *et al.*, 1985). Moreover, the viscosity of the venom hinders its collection. Furthermore, their mouths are very sensitive and capillaries used in venom collection often damage their gums, leading to blood contamination of the venom samples (Pers. Obs.; Pers. Comm.). Blood contamination is especially problematic for CTL identification in the SDS-PAGE, as this toxin and Hemoglobin present similar molecular weight. In fact, both were identified in the same band in reduced SDS-PAGE (Figure 19, Table 13)

Nevertheless, the bias in shotgun quantification would be expected to be similar over the samples, but heterogeneity was observed. This indicates that the venom may be more heterogeneous among the species than the transcriptomic approach suggested. Regardless of what is causing this incongruence, two different sequencing methodologies agreed on the high expression of CTL in the venom glands (Figure 15). This indicates an important role of this toxin family for these snakes, as it demands high consumption of energy for the RNA synthesis. Unnecessary energy spending would be expected to be deleterious and removed by natural selection.

5.4 Evolution of toxins in Hydropsini

Many of the identified CTL transcripts have an insertion containing cysteines, as those recently described in *Helicops leopardinus* (XIE *et al.*, 2022). These sequences should have originated in the common ancestor of Hydropsini, as they are exclusive to and spread throughout the tribe. The final set of CTL sequences assembled with long-reads indicates that individuals express more than one CTL sequence (Figure 8, 23). Moreover, this supports that these CTL are a novelty in the venom of Hydropsini, and not assembling artifacts.

In snakes, CTLs are grouped in the Ca^{2+} dependent sugar binding proteins, and those which cannot bind to sugars, named CTL-like or Snaclecs (EBLE, 2019). Studies on Viperidae venom show that CTL-like proteins present a quaternary structure, forming heterodimers of alpha and beta subunits (CLEMETSON *et al.*, 2005; DOHNÁLEK; SKÁLOVÁ, 2022; EBLE, 2019; XIE *et al.*, 2022). In the alpha subunit, the motif ‘QC’ is conserved and contains the cysteine that forms the disulfide bond that stabilizes the heterodimer (EBLE, 2019; XIE *et al.*, 2022). The homology designation and maximum likelihood analyses show that the CTL-like highly expressed in Hydropsini are similar to the alpha subunit of Dipsadidae CTL-like sequences (Figure 20), in agreement to previous studies (XIE *et al.*, 2022). Thus, they possibly result from modifications of alpha-subunit CTL-like toxins from Dipsadidae (Figure 23). As in other studies (CERDA, 2023), the presented data indicates that these novel CTL are the ones being highly expressed in Hydropsini (Table 14), and hence should be subject to more analyses.

In the outgroups, transcript expression of alpha and beta subunit is similar (Figure 21), indicating a stoichiometric proportion of both subunits of the dimer in the venom (CHING *et al.*, 2006; JUNQUEIRA-DE-AZEVEDO *et al.*, 2006). However, in Hydropsini CTL-like, a sharp increase in the expression of the modified alpha subunit is observed, that is not accompanied by high expression levels for the beta subunits (Figure 21). Moreover, the diagnostic motif of the alpha subunit ‘QC’ is absent from the sequences highly expressed in Hydropsini (Table 14). In fact, CTL transcripts with such motif were only found in the Illumina assemblies of two Hydropsini species (Table 14) and in none of

the ONT assemblies. In addition to the loss of the motif, structural analyses indicate that in some sequences there is no free Cysteine to bond to the beta subunit, as the Cysteines present in the insertion likely form disulfide bonds with themselves. These results suggest that the Hydropsini CTL-like sequences don’t form the canonical alpha-beta dimers. In line with that, in the non-reduced SDS-PAGE of *Helicops carinicaudus*, a faint band is visible in the range of 14Kda, expected molecular weight of monomeric CTL (Figure 19).

The structural analyses coupled to the alignments to a reference protein with resolved three-dimensional structure suggest that the novel insertion of Hydropsini CTL-like is located in the central loop. The same region also presents sequence and structural diversity within Hydropsini (Figure 22, 23, 24). Such diversity likely resulted from gene duplications and subsequent divergence of gene copies, common features in the evolution of toxins from snakes (BAYONA-SERRANO *et al.*, 2020; FRY *et al.*, 2003; JUAREZ *et al.*, 2008; LYNCH, 2007), and other animals (DUDA; PALUMBI, 1999; SONODA *et al.*, 2023; SUNAGAR *et al.*, 2013).

Interestingly, the phylogeny of the CTLs from Hydropsini suggests a complex history of intra-specific duplications and convergent indels in the central loop (Figure 23). A hypothesis that could explain this pattern is that the phylogenetic signal of ancestral duplication events was confounded by inter-loci recombination between the gene copies. This would cause the paralogs within each species to be more similar to themselves than to their relative orthologs in a different species, resulting in species specific clades, as observed. Though genomic experiments are necessary to test it, this hypothesis provides a more parsimonious explanation to the observed phylogeny than the idea of multiple expansions of these CTL in each species. To the best of the author’s knowledge, no studies tested such hypothesis in toxin genes.

In true C-type Lectins, the central loop acts by binding to sugars (EBLE, 2019). In CTL-like, the same loop is important for dimerization and recognition of the molecular target (EBLE, 2019; FRY, 2015; LU *et al.*, 2005). The role of the new and diverse central loops in Hydropsini envenomation is enigmatic. A possibility is that they constitute adaptations for subduing fishes, which is the main item in the tribe’s diet. In Dipsadidae, CTL-like toxins are thought to act like the Viperidae CTL-like, disrupting hemostasis by targeting platelet aggregation factors. Fishes present nucleate thrombocytes instead of platelets, as other non-mammalian vertebrates (LEVIN, 2019), but the agglutination mechanisms of these cells are similar to platelets (JAGADEESWARAN *et al.*, 1999;

KHANDEKAR *et al.*, 2012; LANG *et al.*, 2010; LEVIN, 2019), although there are considerable sequence divergences between fish and mammalian platelet/thrombocyte adhesion proteins (LANG *et al.*, 2010). Additionally, fish have evolved coagulation mechanisms to prevent blood loss through their gills, such as potent trypsin that activate thrombocytes (KIM *et al.*, 2009). Still, the association between the particularities of the coagulation cascade in bony fishes and the rapid diversification observed in the Hydropsini CTL remains unclear and requires functional investigations.

As the central loop is responsible for recognizing the molecular target, it is also possible that these novelties resulted in a change of function. In the fish eating coral-snake *Micrurus surinamenses*, CTLs with cardiotoxic have been described (RINCON-FILHO *et al.*, 2020). Also, CTL-like proteins from *Echis multisquamatus* were found to inhibit nicotinic acetylcholine receptors (KRYUKOVA *et al.*, 2020). These studies show evidence of how versatile these molecules can be.

While most *in vivo* experiments on Hydropsini venom revealed a neurotoxic action (ALBOLEA *et al.*, 2000; ESTRELLA *et al.*, 2009, 2011), this action was attributed to the CRISP *Helicopsin (ESTRELLA et al., 2011)* and not to C-type lectins. On the other hand, Hydropsini snakebites in humans presented coagulopathies as main symptoms, although headache and nausea were also observed (ALBOLEA *et al.*, 1999; SILVA *et al.*, 2019; VILLCA-CORANI *et al.*, 2021). Interestingly the most severe hemostatic symptoms were registered in patients bitten by *Hydrops triangularis* and *Helicops modestus*, species presenting a CTL rich venom proteome (Figure 17), suggesting that these toxins may interfere in hemostasis (ALBOLEA *et al.*, 1999; SILVA *et al.*, 2019; VILLCA-CORANI *et al.*, 2021).

Another novelty present in some *Helicops* venom is the presence of a protein containing a single kunitz domain. This is supported by transcriptomic (Figure 14, 25 and 26) and proteomic evidence (Figure 17, 18 and 19). Snake venom Kunitz are small toxins homologous to bovine pancreatic trypsin inhibitor (DROCTOVÉ *et al.*, 2022; KUNITZ; NORTHROP, 1936). In Toxicofera, two-domain Kunitz is plesiomorphic, being the only Kunitz found in anguimorph and iguanian lizards (FRY, 2015), in which they likely play endophysiological roles. Meanwhile, in most Caenophidia venom glands, the single domain form is the most dominant (FRY, 2015). The Kunitz domain phylogeny suggests an ancestral duplication splitting the double Kunitz proteins from the rest (Figure 25). It also

suggests that, among the analyzed domain organizations, single Kunitz is evolutionarily closer to the complex ku-wa-wa-wa-ku-ku domain structure (Figure 25).

Few studies have investigated the roles that Kunitz may play on envenomation. The protease inhibition activity of these toxins, (CHANG *et al.*, 2001; CHENG *et al.*, 2005; SHAFQAT *et al.*, 1990; ZHOU *et al.*, 2004) possibly disturbs hemostasis (MASCI *et al.*, 2000). Non-inhibitory related activities, such as interaction with ion channels (HARVEY, 2001; SCHWEITZ *et al.*, 1994) and GPCR (DROCTOVÉ *et al.*, 2022) are also present in some Elapidae. Moreover, a Kunitz protein with both protease and ion channels inhibitors activities has also been recorded (YANG *et al.*, 2014).

The current result suggests an increase in expression of Kunitz toxins in some Hydropsini when compared to outgroups (Figure 14), especially for the single Kunitz domain (Figure 26). This feature appears to be more pronounced in *Helicops angulatus, Helicops polylepis, Helicops infrataeniatus* and *Helicops leopardinus.* Further, Kunitz were identified in considerable amounts in the shotgun venom of *Helicops angulatus* and *Helicops polylepis* (Figure 17) and found in the SDS-PAGE of *Helicops carinicaudus* and *Helicops angulatus* (Figure 18, 19).

Although lowly expressed, transcripts coding for single Kunitz sequences were identified in Dipsadidae from the outgroup. The evolutionary mechanisms that keep these genes being expressed are not clear, but this standing variability of the venom may be important for long-term evolutionary process, providing functional toxins which can be recruited on the face of new selective pressures.

5.5 The evolution of the venom profile of the Hydropsini driven by a fish centered diet

As snake venom evolution is tightly linked to their diet (DAVIES; ARBUCKLE, 2019; HEALY *et al.*, 2019; LYONS *et al.*, 2020), the differences in toxin expression between Hydropsini and the outgroups may have evolved along with the addition of fishes as the main item in the diet of the tribe. Most studies on fish eating snake venom focus on Elapidae species. Fish eating sea-dwelling Elapidae present venom rich in PLA and 3FTx (LOMONTE *et al.*, 2014; TAN *et al.*, 2015b, 2017). Similarly, *Micrurus surinamensis*, a semi-aquatic fresh water Elapidae that feeds on fishes (TAVARES-PINHEIRO *et al.*, 2021), presents a venom rich in 3FTx (SANZ *et al.*, 2019). In these snakes, the evolution of a neurotoxic and simplified (*i.e.* with low diversity of toxins) venom has been correlated to a necessity of quickly subduing fishes (TAN *et al.*, 2017), especially in an environment where bite and release behavior might not work (HEALY *et al.*, 2019). Nevertheless, neurotoxic venom is assumed to be the plesiomorphic condition in Elapidae; thus, the venom of these sea-snakes might result from simplifications of the ancestral venom. Accordingly, a sea-snake species that became specialized in feeding on fish eggs, which do not need subduction actions, presents frame shifts (LI *et al.*, 2005a) and mutations in conserved sites of these toxins (LI *et al.*, 2005b), reducing their toxicity.

In Dipsadidae, neurotoxic venom is not common, even in fish eating species. The semi-aquatic fish eating *Erythrolampus miliaris* (EISFELD *et al.*, 2021) presents a high abundance of svMMP in their venom gland transcriptome (BAYONA-SERRANO *et al.*, 2020), which is not assumed to be neurotoxic. In this study, the venom gland of *Ptychophis flavovirgatus,* which has been suggested to feed on fishes (SCARTOZZONI, 2005), was also shown to contain increased levels svMMP. Similarly, proteolytic activity was found in the venom of the semi-aquatic species *Hydrodynastes gigas* (HILL; MACKESSY, 2000), which also feeds on fish, although this species tends to be more generalist (LÓPEZ; GIRAUDO, 2004).

In Hydropsini, neither PLA2, 3FTx neurotoxins nor proteases were found highly expressed in the venom gland transcriptome. Instead, CTLs were identified as the most

expressed toxin family in the venom gland, suggesting hemostatic disruptor venom, as previously discussed. Nevertheless, the proteomic experiments also indicate CRISPs and Kunitz as important components of the venom from these species.

The supposed neurotoxicity of CRISPs in Hydropsini (ESTRELLA *et al.*, 2011) and their high abundance in the venom (Figure 17, 18 and 19) may be a case of convergent functional evolution to the fish eating Elapidae and their neurotoxic venom. However, CRISPs' abundance in the proteome or transcriptome does not clearly correlate with the amount of fish in the diet of the species in the tribe. *Helicops angulatus* diet is estimated to be 80% composed of fish, while in *Helicops polylepis* this proportion reaches 90% (DE CARVALHO TEIXEIRA *et al.*, 2017; SCARTOZZONI, 2009). Still, both transcriptome and proteome indicate that these two species express similar levels of CRISP (Figure 13, 17). The species *Helicops infrataeniatus* presents the most CRISP rich proteome (figure 17), while it has a smaller proportion of fish in diet than *Helicops angulatus* (DE AGUIAR; DI-BERNARDO, 2004; SCARTOZZONI, 2009). On the other hand, in the fish specialist *Hydrops triangularis (*SCARTOZZONI*,* 2009*),* CTL was identified as the main component in the venom proteome (Figure 17), with only a small fraction being attributed to Kunitz, SVMP or other toxins. A similar venom profile was found for *Helicops modestus*, in which only approximately 70% of the ingested items were fish.

Interestingly, snakes from the genus *Hydrops* present a diet rich in elongated fishes, while *Helicops* diet include a bigger proportion of non-elongated prey (SCARTOZZONI, 2009). Hence, it is possible that CRISP rich venoms have evolved as a response to the increase of consumption of non-elongated and more robust fishes in the ancestor of *Helicops* (SCARTOZZONI, 2009). In line with that, the cranium of *Helicops* species is more robust than *Hydrops triangularis* species, facilitating ingestion of bulkier fishes and other items (SCARTOZZONI, 2005, 2009). This could imply that different fishes impose different selective pressures over venom composition, suggesting that grouping Actinopterygii species, a highly diverse group, into a single alimentary class may be an oversimplification of data.

Regardless of what drove its composition, both literature and the presented results suggest that Hydropsini have a venom remarkably different from other previously investigated snakes, including many Dipsadidae. The high abundance of a novel type of CTL in their venom gland, as well as the presence of CRISPs and Kunitz in their venom makes the tribe unique. Though simple in composition, the venom of the Hydropsini

venom suggests a complex evolutionary history of the chemical arsenal of Dipsadidae snakes, especially in Xenodontinae.

6. Conclusions

1. High expression of CTL-like toxin transcripts is common to all Hydropsini snakes. Increase in expression of single Kunitz toxins appear to be exclusive to the *Helicops* genus.
2. Abundance of CTL-like transcripts is not clearly correlated with the venom proteome from these snakes. Instead, proteomic analyses suggest CRISP as an important venom component.
3. Discordance between transcriptome and proteome likely arises from biological causes.
4. Highly expressed CTL-like toxins are likely modified versions from the alpha subunit of CTL-like toxins. Current data support that these new CTL-like proteins do not form the canonical heterodimers of CTL-like toxins.
5. These new CTL sequences present a different central loop, providing a unique structural feature to a domain that is fundamental to protein-target interaction.
6. The role of these Hydropsini CTL remains to be clarified. Based on snakebite records, a possible hypothesis is that these CTL disrupt hemostasis.
7. Proteome analyses suggest abundance of CRISP in Hydropsini venom. These could be an evolutionary convergence to neurotoxic venom in fish feeding snakes.
8. The venom of the Hydropsini tribe further supports high heterogeneity between the tribes of Dipsadidae, indicating unique evolutionary paths for the venom of tribes in this family.

7. Resumo

Poucos estudos focaram no veneno dos Dipsadidae, ainda que esta família represente substancial porção da diversidade de serpentes. Ainda assim, publicações sugerem uma alta diversidade no arsenal químico da família, com cada tribo apresentando toxinas e venenos únicos. Esta dissertação teve como objetivo, investigar a composição e evolução do veneno das cobras d'água pertencentes a tribo Hydropsini, especialistas em peixes. A composição do seu veneno foi inferida usando uma combinação de abordagens ômicas, incluindo leituras-longas e proteômica para corroborar os resultados obtidos com transcriptomas. Para melhor compreender a evolução dos genes de toxinas e da composição do veneno dentro da tribo, venenos de Hydropsini foram comparados com os de tribos próximas, fornecendo uma perspectiva filogenética e evolutiva para o presente estudo. Os resultados obtidos indicaram uma alta expressão de transcritos de Lectinas do Tipo-C-like (CTL-like) nas glândulas de veneno nesta tribo. Em contrapartida, o proteoma mostrou a prevalência de outras toxinas, como Proteínas Secretadas Ricas em Cisteínas (CRISP) e Kunitz, enquanto não se observou uma abundância particularmente elevada de CTL-like. Ao investigar as CTL-like de Hydropsini montadas com leituras longas e curtas, foi observada uma inserção no *loop* central, perturbando uma região importante para a dimerização das CTL-like. Também, a falta de relação estequiométrica das subunidades alfa e beta de CTL-like (observada em outros Dipsadidae) indica uma ausência de dímeros canônicos de CTL-like em Hydropsini. Ainda, foi detectada uma considerável expressão de Kunitz de domínio único em Hydropsini, mas não no grupo externo. No geral, não se observou uma clara relação entre a composição do veneno de cada espécie de Hydropsini serpentes e a quantidade de peixe na sua dieta, mas a ingestão de peixes mais robustos pode ter pressionado a evolução de CRISPs neurotóxicas e de Kunitz em *Helicops*. Esta dissertação ajuda a entender o veneno de Dipsadidae, mostrando mais uma tribo com veneno peculiar na família.

8. Abstract

Few studies have focused on the venom of Dipsadidae, even though this family encompasses a substantial portion of the diversity of snakes. Yet, publications suggest a high diversity in the family's chemical arsenal, with each tribe presenting unique venoms and toxins. This dissertation aimed to investigate the composition and evolution of the venom from the water-snakes of the Hydropsini tribe, specialized in fish. The composition of the venom was assessed through combination of omics approaches, including the use of long-reads and proteomics to corroborate transcriptomic findings. To better understand the evolution of toxin genes and venom composition within the tribe, venoms of Hydropsini were compared to those of closely related tribes allowing a phylogenetic and evolutionary perspective to the present study. The obtained results indicated high expression of C-type Lectin-like (CTL-like) transcripts in the venom gland transcriptomes of Hydropsini. On the other hand, proteomic analyses showed prevalence of Cysteine-Rich Secretory Proteins (CRISPs) and Kunitz in the venom of Hydropsini, while low abundance was inferred for the CTL-like. Upon further investigation, CTL-like toxin from Hydropsini presented an insertion in the central-loop, disturbing a region important for dimerization in CTL-like. Moreover, lack of stoichiometric correlation between alpha and beta CTL-like subunits (observed in the remaining Dipsadidae) indicates absence of the canonical CTL-like dimers in Hydropsini. Additionally, considerable expression of single domain Kunitz proteins was detected in Hydropsini, but not in the outgroup. In general, for Hydropsini species, no clear correlation between their venom composition and the amount of fish in their diet was observed. Still, ingestion of bulkier prey may have driven the evolution of Kunitz and neurotoxic CRISPs in the genus *Helicops*. This dissertation helps in understanding the venom of Dipsadidae snakes, indicating yet another tribe with a peculiar venom in the family.

9. References

1. AIRD, S. D.; WATANABE, Y.; VILLAR-BRIONES, A.; ROY, M. C.; TERADA, K.; MIKHEYEV, A. S. Quantitative High-Throughput Profiling of Snake Venom Gland Transcriptomes and Proteomes (Ovophis Okinavensis and Protobothrops Flavoviridis). **BMC Genomics**, v. 14, n. 1, p. 790, 2013.
2. ALBOLEA, A. B. P.; SALOMÃO, M. D. G.; ALMEIDA-SANTOS, S. M.; JORDÃO, R. dos S. Epidemiologia de acidentes causados por serpentes não peçonhentas no estado de São Paulo, Brasil. **Ciência Biológicas e da Saúde**, v. 4, n. 5, p. 99–108, out. 1999.
3. ALBOLEA, A. B. P.; SALOMÃO, M. G.; JORDÃO, R. S.; ALMEIDA-SANTOS, S. M. Why Non-Poisonous Snakes Cause Accidents? **Toxicon**, v. 38, n. 4, p. 567–568, abr. 2000.
4. ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. **Journal of molecular biology**, v. 215, n. 3, p. 403–410, 1990.
5. ANDREWS, S. **FastQC: a quality control tool for high throughput sequence data**. [s.l.] Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom, 2010.
6. BALLESTEROS, J. A.; HORMIGA, G. A New Orthology Assessment Method for Phylogenomic Data: Unrooted Phylogenetic Orthology. **Molecular Biology and Evolution**, v. 33, n. 8, p. 2117–2134, ago. 2016.
7. BAYONA-SERRANO, J. D.; GRAZZIOTIN, F. G.; SALAZAR-VALENZUELA, D.; VALENTE, R. H.; NACHTIGALL, P. G.; COLOMBINI, M.; MOURA-DA-SILVA, A.; JUNQUEIRA-DE-AZEVEDO, I. L. M. Independent Recruitment of Different Types of Phospholipases A2 to the Venoms of Caenophidian Snakes: The Rise of PLA2-IIE within Pseudoboini (Dipsadidae). **Molecular Biology and Evolution**, v. 40, n. 7, p. msad147, 5 jul. 2023.
8. BAYONA-SERRANO, J. D.; VIALA, V. L.; RAUTSAW, R. M.; SCHRAMER, T. D.; BARROS-CARVALHO, G. A.; NISHIYAMA JR, M. Y.; FREITAS-DE-SOUSA, L. A.; MOURA-DA-SILVA, A. M.; PARKINSON, C. L.; GRAZZIOTIN, F. G. Replacement and parallel simplification of nonhomologous proteinases maintain venom phenotypes in rear-fanged snakes. **Molecular Biology and Evolution**, v. 37, n. 12, p. 3563–3575, 2020.
9. BRODIE, E. D.; BRODIE, E. D. Evolutionary Response of Predators to Dangerous Prey: Reduction of Toxicity of Newts and Resistance of Garter Snakes in Island Populations. **Evolution**, v. 45, n. 1, p. 221, fev. 1991.
10. BROWN, R.; BROWN, M.; BDOLAH, A.; KOCHVA, E. Accumulation of Some Secretory Enzymes in Venom Glands of Vipera Palaestinae. **American Journal of Physiology-Legacy Content**, v. 229, n. 6, p. 1675–1679, 1 dez. 1975.
11. BUSHMANOVA, E.; ANTIPOV, D.; LAPIDUS, A.; PRJIBELSKI, A. D. rnaSPAdes: A de Novo Transcriptome Assembler and Its Application to RNA-Seq Data. **GigaScience**, v. 8, n. 9, p. giz100, 1 set. 2019.
12. CALINSKI, T.; HARABASZ, J. A Dendrite Method for Cluster Analysis. **Communications in Statistics - Theory and Methods**, v. 3, n. 1, p. 1–27, 1974.
13. CAMPOS, P. F.; ANDRADE-SILVA, D.; ZELANIS, A.; PAES LEME, A. F.; ROCHA, M. M. T.; MENEZES, M. C.; SERRANO, S. M. T.; JUNQUEIRA-DE-AZEVEDO, I. D.

L. M. Trends in the Evolution of Snake Toxins Underscored by an Integrative Omics Approach to Profile the Venom of the Colubrid *Phalotris Mertensi*. **Genome Biology and Evolution**, v. 8, n. 8, p. 2266–2287, ago. 2016.

14. CAPELLA-GUTIÉRREZ, S.; SILLA-MARTÍNEZ, J. M.; GABALDÓN, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. **Bioinformatics**, v. 25, n. 15, p. 1972–1973, 2009.
15. CARVALHO, P. S. Systematics and biogeography of the Hydropsini tribe (Serpentes: Xenodontinae). 2022.
16. 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. **Proceedings of the National Academy of Sciences**, v. 111, n. 25, p. 9205–9210, 24 jun. 2014.
17. CERDA, P. **Toxic Forms Most Beautiful: The Evolutionary Dynamics of Rear-Fanged Snake Venoms**. 2023. 2023.
18. CERDA, P. A.; CROWE-RIDDELL, J. M.; GONÇALVES, D. J. P.; LARSON, D. A.; DUDA, T. F.; DAVIS RABOSKY, A. R. Divergent Specialization of Simple Venom Gene Profiles among Rear-Fanged Snake Genera (Helicops and Leptodeira, Dipsadinae, Colubridae). **Toxins**, v. 14, n. 7, p. 489, 15 jul. 2022.
19. CHANG, L.; CHUNG, C.; HUANG, H.-B.; LIN, S. Purification and Characterization of a Chymotrypsin Inhibitor from the Venom of Ophiophagus Hannah (King Cobra). **Biochemical and Biophysical Research Communications**, v. 283, n. 4, p. 862–867, maio 2001.
20. CHANG, Z.; LI, G.; LIU, J.; ZHANG, Y.; ASHBY, C.; LIU, D.; CRAMER, C. L.; HUANG, X. Bridger: A New Framework for de Novo Transcriptome Assembly Using RNA-Seq Data. **Genome Biology**, v. 16, n. 1, p. 30, dez. 2015.
21. CHENG, Y.-C.; YAN, F.-J.; CHANG, L.-S. Taiwan Cobra Chymotrypsin Inhibitor: Cloning, Functional Expression and Gene Organization. **Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics**, v. 1747, n. 2, p. 213–220, mar. 2005.
22. CHING, A. T. C.; PAES LEME, A. F.; ZELANIS, A.; ROCHA, M. M. T.; FURTADO, M. D. F. D.; SILVA, D. A.; TRUGILHO, M. R. O.; DA ROCHA, S. L. G.; PERALES, J.; HO, P. L.; SERRANO, S. M. T.; JUNQUEIRA-DE-AZEVEDO, I. L. M. Venomics Profiling of *Thamnodynastes Strigatus* Unveils Matrix Metalloproteinases and Other Novel Proteins Recruited to the Toxin Arsenal of Rear-Fanged Snakes. **Journal of Proteome Research**, v. 11, n. 2, p. 1152–1162, 3 fev. 2012.
23. CHING, A. T. C.; ROCHA, M. M. T.; PAES LEME, A. F.; PIMENTA, D. C.; DE FÁTIMA D. FURTADO, M.; SERRANO, S. M. T.; HO, P. L.; JUNQUEIRA-DE-AZEVEDO, I. L. M. Some Aspects of the Venom Proteome of the Colubridae Snake *Philodryas Olfersii* Revealed from a Duvernoy's (Venom) Gland Transcriptome. **FEBS Letters**, v. 580, n. 18, p. 4417–4422, 7 ago. 2006.
24. CHOMZYNSKI, P.; SACCHI, N. Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate–Phenol–Chloroform Extraction. **Analytical Biochemistry**, v. 162, n. 1, p. 156–159, abr. 1987.
25. CLEMETSON, K. J.; LU, Q.; CLEMETSON, J. M. Snake C-Type Lectin-Like Proteins and Platelet Receptors. **Pathophysiology of Haemostasis and Thrombosis**, v. 34, n. 4–5, p. 150–155, 2005.
26. CURRIER, R. B.; CALVETE, J. J.; SANZ, L.; HARRISON, R. A.; ROWLEY, P. D.; WAGSTAFF, S. C. Unusual Stability of Messenger RNA in Snake Venom Reveals Gene Expression Dynamics of Venom Replenishment. **PLoS ONE**, v. 7, n. 8, p. e41888, 7 ago. 2012.

27. DAVIES, E.-L.; ARBUCKLE, K. Coevolution of snake venom toxic activities and diet: Evidence that ecological generalism favours toxicological diversity. **Toxins**, v. 11, n. 12, p. 711, 2019.
28. DE AGUIAR, L. F. S.; DI-BERNARDO, M. Diet and Feeding Behavior of *Helicops Infrataeniatus* (Serpentes: Colubridae: Xenodontinae) in Southern Brazil. **Studies on Neotropical Fauna and Environment**, v. 39, n. 1, p. 7–14, abr. 2004.
29. DE CARVALHO TEIXEIRA, C.; DE ASSIS MONTAG, L. F.; DOS SANTOS-COSTA, M. C. Diet Composition and Foraging Habitat Use by Three Species of Water Snakes, *Helicops* Wagler, 1830, (Serpentes: Dipsadidae) in Eastern Brazilian Amazonia. **Journal of Herpetology**, v. 51, n. 2, p. 215–222, jun. 2017.
30. DOHNÁLEK, J.; SKÁLOVÁ, T. C-Type Lectin-(like) Fold – Protein-Protein Interaction Patterns and Utilization. **Biotechnology Advances**, v. 58, p. 107944, set. 2022.
31. DROCTOVÉ, L.; CIOLEK, J.; MENDRE, C.; CHORFA, A.; HUERTA, P.; CARVALHO, C.; GOUIN, C.; LANCIEN, M.; STANAJIC-PETROVIC, G.; BRACO, L.; BLANCHET, G.; UPERT, G.; DE PAUW, G.; BARBE, P.; KECK, M.; MOURIER, G.; MOUILLAC, B.; DENIS, S.; RODRÍGUEZ DE LA VEGA, R. C.; QUINTON, L.; GILLES, N. A New Kunitz-type Snake Toxin Family Associated with an Original Mode of Interaction with the Vasopressin 2 Receptor. **British Journal of Pharmacology**, v. 179, n. 13, p. 3470–3481, jul. 2022.
32. DUDA, T. F.; PALUMBI, S. R. Molecular Genetics of Ecological Diversification: Duplication and Rapid Evolution of Toxin Genes of the Venomous Gastropod *Conus*. **Proceedings of the National Academy of Sciences**, v. 96, n. 12, p. 6820–6823, 8 jun. 1999.
33. DURBAN, J.; JUÁREZ, P.; ANGULO, Y.; LOMONTE, B.; FLORES-DIAZ, M.; ALAPE-GIRÓN, A.; SASA, M.; SANZ, L.; GUTIÉRREZ, J. M.; DOPAZO, J.; CONESA, A.; CALVETE, J. J. Profiling the Venom Gland Transcriptomes of Costa Rican Snakes by 454 Pyrosequencing. **BMC Genomics**, v. 12, n. 1, p. 259, dez. 2011.
34. E SILVA, M. R.; BERALDO, W. T.; ROSENFELD, G. BRADYKININ, A HYPOTENSIVE AND SMOOTH MUSCLE STIMULATING FACTOR RELEASED FROM PLASMA GLOBULIN BY SNAKE VENOMS AND BY TRYPSIN. **American Journal of Physiology-Legacy Content**, v. 156, n. 2, p. 261–273, 1 fev. 1949.
35. EBLE, J. Structurally Robust and Functionally Highly Versatile—C-Type Lectin (-Related) Proteins in Snake Venoms. **Toxins**, v. 11, n. 3, p. 136, 1 mar. 2019.
36. EISFELD, A.; PIZZATTO, L.; VRCIBRADIC, D. Diet of the Semiaquatic Snake Erythrolamprus miliaris (Dipsadidae, Xenodontinae) in the Brazilian Atlantic Forest. **Journal of Herpetology**, v. 55, n. 4, 27 set. 2021. Disponível em: <https://bioone.org/journals/journal-of-herpetology/volume-55/issue-4/20-117/Diet-of-the-Semiaquatic-Snake-Erythrolamprus-miliaris-Dipsadidae-Xenodontinae-in/10.1670/20-117.full>. Acesso em: 27 nov. 2023.
37. EMMS, D. M.; KELLY, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. **Genome biology**, v. 16, n. 1, p. 1–14, 2015.
38. ESTRELLA, A.; RODRIGUEZ-TORRES, A.; SERNA, L.; NAVARRETE, L. F.; RODRIGUEZ-ACOSTA, A. Is the South American Water Snake Helicops Angulatus (Linnaeus, 1758) (Dipsadidae:Xenodontinae) Venomous?/?Es Venenosa La Serpiente de Agua Suramericana Helicops Angulatus (linnaeus, 1758) (Colubridae:Xenodontinae)? **Herpetotropicos: Tropical Amphibians & Reptiles**, v. 5, n. 2, p. 79+, 2009.

39. ESTRELLA, A.; SÁNCHEZ, E. E.; GALÁN, J. A.; TAO, W. A.; GUERRERO, B.; NAVARRETE, L. F.; RODRÍGUEZ-ACOSTA, A. Characterization of Toxins from the Broad-Banded Water Snake Helicops Angulatus (Linnaeus, 1758): Isolation of a Cysteine-Rich Secretory Protein, Helicopsin. **Archives of Toxicology**, v. 85, n. 4, p. 305–313, abr. 2011.
40. FINN, R. D.; BATEMAN, A.; CLEMENTS, J.; COGGILL, P.; EBERHARDT, R. Y.; EDDY, S. R.; HEGER, A.; HETHERINGTON, K.; HOLM, L.; MISTRY, J.; SONNHAMMER, E. L. L.; TATE, J.; PUNTA, M. Pfam: The Protein Families Database. **Nucleic Acids Research**, v. 42, n. D1, p. D222–D230, jan. 2014.
41. FREITAS-DE-SOUSA, L. A.; NACHTIGALL, P. G.; PORTES-JUNIOR, J. A.; HOLDING, M. L.; NYSTROM, G. S.; ELLSWORTH, S. A.; GUIMARÃES, N. C.; TIOYAMA, E.; ORTIZ, F.; SILVA, B. R.; KUNZ, T. S.; JUNQUEIRA-DE-AZEVEDO, I. L. M.; GRAZZIOTIN, F. G.; ROKYTA, D. R.; MOURA-DA-SILVA, A. M. Size Matters: An Evaluation of the Molecular Basis of Ontogenetic Modifications in the Composition of Bothrops Jararacussu Snake Venom. **Toxins**, v. 12, n. 12, p. 791, 11 dez. 2020.
42. FRY, B. **Venomous reptiles and their toxins: evolution, pathophysiology and biodiscovery**. [s.l.] Oxford University Press, 2015.
43. FRY, B. G.; SCHEIB, H.; VAN DER WEERD, L.; YOUNG, B.; MCNAUGHTAN, J.; RAMJAN, S. F. R.; VIDAL, N.; POELMANN, R. E.; NORMAN, J. A. Evolution of an Arsenal. **Molecular & Cellular Proteomics**, v. 7, n. 2, p. 215–246, fev. 2008.
44. FRY, B. G.; WÜSTER, W.; KINI, R. M.; BRUSIC, V.; KHAN, A.; VENKATARAMAN, D.; ROONEY, A. P. Molecular evolution and phylogeny of elapid snake venom three-finger toxins. **Journal of molecular evolution**, v. 57, p. 110–129, 2003.
45. FU, L.; NIU, B.; ZHU, Z.; WU, S.; LI, W. CD-HIT: accelerated for clustering the next-generation sequencing data. **Bioinformatics**, v. 28, n. 23, p. 3150–3152, 2012.
46. GLEESON, J.; LEGER, A.; PRAWER, Y. D. J.; LANE, T. A.; HARRISON, P. J.; HAERTY, W.; CLARK, M. B. Accurate Expression Quantification from Nanopore Direct RNA Sequencing with NanoCount. **Nucleic Acids Research**, v. 50, n. 4, p. e19–e19, 28 fev. 2022.
47. GOLDING, W. R. J. The Brooklyn Papyrus (47.218. 48 and 47.218. 85) and its Snakebite Treatments. **University of South Africa**, 2020.
48. GRABHERR, M. G.; HAAS, B. J.; YASSOUR, M.; LEVIN, J. Z.; THOMPSON, D. A.; AMIT, I.; ADICONIS, X.; FAN, L.; RAYCHOWDHURY, R.; ZENG, Q. Full-length transcriptome assembly from RNA-Seq data without a reference genome. **Nature biotechnology**, v. 29, n. 7, p. 644–652, 2011.
49. GRAZZIOTIN, F. G.; ZAHER, H.; MURPHY, R. W.; SCROCCHI, G.; BENAVIDES, M. A.; ZHANG, Y.-P.; BONATTO, S. L. Molecular Phylogeny of the New World Dipsadidae (Serpentes: Colubroidea): A Reappraisal. **Cladistics**, v. 28, n. 5, p. 437–459, out. 2012.
50. GREENE, H. W. **Snakes: the evolution of mystery in nature**. [s.l.] Univ of California Press, 1997.
51. GUREVICH, A.; SAVELIEV, V.; VYAHHI, N.; TESLER, G. QUAST: Quality Assessment Tool for Genome Assemblies. **Bioinformatics**, v. 29, n. 8, p. 1072–1075, 15 abr. 2013.
52. HAAS, B.; PAPANICOLAOU, A. TransDecoder (find coding regions within transcripts). **Google Scholar**, 2016.
53. HAHNE, F.; IVANEK, R. Visualizing Genomic Data Using Gviz and Bioconductor. *Em*: MATHÉ, E.; DAVIS, S. **Statistical Genomics**. Methods in Molecular Biology. New York, NY: Springer New York, 2016. p. 335–351.

54. HAO, J.; HO, T. K. Machine Learning Made Easy: A Review of *Scikit-Learn* Package in Python Programming Language. **Journal of Educational and Behavioral Statistics**, v. 44, n. 3, p. 348–361, jun. 2019.
55. HARVEY, A. L. Twenty Years of Dendrotoxins. **Toxicon**, v. 39, n. 1, p. 15–26, jan. 2001.
56. HAWGOOD, B. J. Abbé Felice Fontana (1730–1805): Founder of Modern Toxinology. **Toxicon**, v. 33, n. 5, p. 591–601, maio 1995.
57. HEAD, J. J.; MAHLOW, K.; MÜLLER, J. Fossil calibration dates for molecular phylogenetic analysis of snakes 2: Caenophidia, Colubroidea, Elapoidea, Colubridae. **Palaeontologia electronica**, v. 19, n. 2, 2016. Disponível em: <http://palaeo-electronica.org/content/fc-9>. Acesso em: 7 jul. 2023.
58. HEALY, K.; CARBONE, C.; JACKSON, A. L. Snake Venom Potency and Yield Are Associated with Prey-Evolution, Predator Metabolism and Habitat Structure. **Ecology Letters**, v. 22, n. 3, p. 527–537, mar. 2019.
59. HILL, R. E.; MACKESSY, S. P. Venom Yields from Several Species of Colubrid Snakes and Differential Effects of Ketamine. **Toxicon**, v. 35, n. 5, p. 671–678, maio 1997.
60. HILL, R. E.; MACKESSY, S. P. Characterization of Venom (Duvernoy's Secretion) from Twelve Species of Colubrid Snakes and Partial Sequence of Four Venom Proteins. **Toxicon**, v. 38, n. 12, p. 1663–1687, dez. 2000.
61. HOLDING, M.; MARGRES, M.; MASON, A.; PARKINSON, C.; ROKYTA, D. Evaluating the Performance of De Novo Assembly Methods for Venom-Gland Transcriptomics. **Toxins**, v. 10, n. 6, p. 249, 19 jun. 2018.
62. JAGADEESWARAN, P.; SHEEHAN, J. P.; CRAIG, F. E.; TROYER, D. Identification and Characterization of Zebrafish Thrombocytes. **British Journal of Haematology**, v. 107, n. 4, p. 731–738, dez. 1999.
63. JUAREZ, P.; COMAS, I.; GONZALEZ-CANDELAS, F.; CALVETE, J. J. Evolution of snake venom disintegrins by positive Darwinian selection. **Molecular biology and evolution**, v. 25, n. 11, p. 2391–2407, 2008.
64. JUMPER, J.; EVANS, R.; PRITZEL, A.; GREEN, T.; FIGURNOV, M.; RONNEBERGER, O.; TUNYASUVUNAKOOL, K.; BATES, R.; ŽÍDEK, A.; POTAPENKO, A.; BRIDGLAND, A.; MEYER, C.; KOHL, S. A. A.; BALLARD, A. J.; COWIE, A.; ROMERA-PAREDES, B.; NIKOLOV, S.; JAIN, R.; ADLER, J.; BACK, T.; PETERSEN, S.; REIMAN, D.; CLANCY, E.; ZIELINSKI, M.; STEINEGGER, M.; PACHOLSKA, M.; BERGHAMMER, T.; BODENSTEIN, S.; SILVER, D.; VINYALS, O.; SENIOR, A. W.; KAVUKCUOGLU, K.; KOHLI, P.; HASSABIS, D. Highly Accurate Protein Structure Prediction with AlphaFold. **Nature**, v. 596, n. 7873, p. 583–589, 26 ago. 2021.
65. JUNGO, F.; BOUGUELERET, L.; XENARIOS, I.; POUX, S. The UniProtKB/Swiss-Prot Tox-Prot Program: A Central Hub of Integrated Venom Protein Data. **Toxicon**, v. 60, n. 4, p. 551–557, set. 2012.
66. JUNQUEIRA-DE-AZEVEDO, I.; CAMPOS, P.; CHING, A.; MACKESSY, S. Colubrid Venom Composition: An -Omics Perspective. **Toxins**, v. 8, n. 8, p. 230, 23 jul. 2016.
67. JUNQUEIRA-DE-AZEVEDO, I. L. M.; CHING, A. T. C.; CARVALHO, E.; FARIA, F.; NISHIYAMA, M. Y.; HO, P. L.; DINIZ, M. R. V. *Lachesis Muta* (Viperidae) cDNAs Reveal Diverging Pit Viper Molecules and Scaffolds Typical of Cobra (Elapidae) Venoms: Implications for Snake Toxin Repertoire Evolution. **Genetics**, v. 173, n. 2, p. 877–889, 1 jun. 2006.

68. KATOH, K.; STANDLEY, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. **Molecular Biology and Evolution**, v. 30, n. 4, p. 772–780, 1 abr. 2013.
69. KHANDEKAR, G.; KIM, S.; JAGADEESWARAN, P. Zebrafish Thrombocytes: Functions and Origins. **Advances in Hematology**, v. 2012, p. 1–9, 2012.
70. KHANNA, A.; LARSON, D.; SRIVATSAN, S.; MOSIOR, M.; ABBOTT, T.; KIWALA, S.; LEY, T.; DUNCAVAGE, E.; WALTER, M.; WALKER, J.; GRIFFITH, O.; GRIFFITH, M.; MILLER, C. Bam-readcount - rapid generation of basepair-resolution sequence metrics. **Journal of Open Source Software**, v. 7, n. 69, p. 3722, 29 jan. 2022.
71. KIM, S.; CARRILLO, M.; KULKARNI, V.; JAGADEESWARAN, P. Evolution of Primary Hemostasis in Early Vertebrates. **PLoS ONE**, v. 4, n. 12, p. e8403, 23 dez. 2009.
72. KOCH-WESER, J.; VIDT, D. G.; BRAVO, E. L.; FOUAD, F. M. Captopril. **New England Journal of Medicine**, v. 306, n. 4, p. 214–219, 28 jan. 1982.
73. KRUEGER, F. Trim Galore!: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. **Babraham Institute**, 2015.
74. KRYUKOVA, E. V.; VULFIUS, C. A.; ZIGANSHIN, R. H.; ANDREEVA, T. V.; STARKOV, V. G.; TSETLIN, V. I.; UTKIN, Y. N. Snake C-type lectin-like proteins inhibit nicotinic acetylcholine receptors. **Journal of Venom Research**, v. 10, p. 23, 2020.
75. KÜCK, P.; MEUSEMANN, K. FASconCAT: Convenient Handling of Data Matrices. **Molecular Phylogenetics and Evolution**, v. 56, n. 3, p. 1115–1118, set. 2010.
76. KÜCK, P.; STRUCK, T. H. BaCoCa – A Heuristic Software Tool for the Parallel Assessment of Sequence Biases in Hundreds of Gene and Taxon Partitions. **Molecular Phylogenetics and Evolution**, v. 70, p. 94–98, jan. 2014.
77. KUNITZ, M.; NORTHROP, J. H. Isolation from Beef Pancreas of Crystalline Trypsinogen, Trypsin, a Trypsin Inhibitor, and an Inhibitor-Trypsin Compound. **Journal of General Physiology**, v. 19, n. 6, p. 991–1007, 20 jul. 1936.
78. LANG, M. R.; GIHR, G.; GAWAZ, M. P.; MÜLLER, I. I. Hemostasis in Danio Rerio: Is the Zebrafish a Useful Model for Platelet Research? **Journal of Thrombosis and Haemostasis**, v. 8, n. 6, p. 1159–1169, jun. 2010.
79. LANGMEAD, B.; SALZBERG, S. L. Fast Gapped-Read Alignment with Bowtie 2. **Nature Methods**, v. 9, n. 4, p. 357–359, abr. 2012.
80. LÊ, S.; JOSSE, J.; HUSSON, F. **FactoMineR** : An *R* Package for Multivariate Analysis. **Journal of Statistical Software**, v. 25, n. 1, 2008. Disponível em: <http://www.jstatsoft.org/v25/i01/>. Acesso em: 7 jul. 2023.
81. LEGER, A.; LEONARDI, T. pycoQC, interactive quality control for Oxford Nanopore Sequencing. **Journal of Open Source Software**, v. 4, n. 34, p. 1236, 28 fev. 2019.
82. LEVIN, J. The Evolution of Mammalian Platelets. *Em*: **Platelets**. [s.l.] Elsevier, 2019. p. 1–23.
83. LI, B.; DEWEY, C. N. RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. **BMC Bioinformatics**, v. 12, n. 1, p. 323, dez. 2011.
84. LI, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. Disponível em: <https://arxiv.org/abs/1303.3997>. Acesso em: 15 nov. 2023.
85. LI, H. Minimap2: Pairwise Alignment for Nucleotide Sequences. **Bioinformatics**, v. 34, n. 18, p. 3094–3100, 15 set. 2018.

86. LI, M.; FRY, B. G.; KINI, R. M. Eggs-Only Diet: Its Implications for the Toxin Profile Changes and Ecology of the Marbled Sea Snake (Aipysurus Eydouxii). **Journal of Molecular Evolution**, v. 60, n. 1, p. 81–89, jan. 2005a.
87. LI, M.; FRY, B. G.; KINI, R. M. Putting the Brakes on Snake Venom Evolution: The Unique Molecular Evolutionary Patterns of Aipysurus Eydouxii (Marbled Sea Snake) Phospholipase A2 Toxins. **Molecular Biology and Evolution**, v. 22, n. 4, p. 934–941, abr. 2005b.
88. LOMONTE, B.; PLA, D.; SASA, M.; TSAI, W.-C.; SOLÓRZANO, A.; UREÑA-DÍAZ, J. M.; FERNÁNDEZ-MONTES, M. L.; MORA-OBANDO, D.; SANZ, L.; GUTIÉRREZ, J. M.; CALVETE, J. J. Two Color Morphs of the Pelagic Yellow-Bellied Sea Snake, Pelamis Platura, from Different Locations of Costa Rica: Snake Venomics, Toxicity, and Neutralization by Antivenom. **Journal of Proteomics**, v. 103, p. 137–152, maio 2014.
89. LÓPEZ, M. S.; GIRAUDO, A. Diet of the large water snake Hydrodynastes gigas (Colubridae) from northeast Argentina. **Amphibia-Reptilia**, v. 25, n. 2, p. 178–184, 2004.
90. LU, Q.; NAVDAEV, A.; CLEMETSON, J. M.; CLEMETSON, K. J. Snake Venom C-Type Lectins Interacting with Platelet Receptors. Structure–Function Relationships and Effects on Haemostasis. **Toxicon**, v. 45, n. 8, p. 1089–1098, jun. 2005.
91. LYNCH, V. J. Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. **BMC Evolutionary Biology**, v. 7, n. 1, p. 2, 2007.
92. LYONS, K.; DUGON, M. M.; HEALY, K. Diet Breadth Mediates the Prey Specificity of Venom Potency in Snakes. **Toxins**, v. 12, n. 2, p. 74, 23 jan. 2020.
93. MACKESSY, S. P.; BAXTER, L. M. Bioweapons Synthesis and Storage: The Venom Gland of Front-Fanged Snakes. **Zoologischer Anzeiger - A Journal of Comparative Zoology**, v. 245, n. 3–4, p. 147–159, nov. 2006.
94. MACKESSY, S. P.; SIXBERRY, N. M.; HEYBORNE, W. H.; FRITTS, T. Venom of the Brown Treesnake, Boiga Irregularis: Ontogenetic Shifts and Taxa-Specific Toxicity. **Toxicon**, v. 47, n. 5, p. 537–548, abr. 2006.
95. MACRANDER, J.; BROE, M.; DALY, M. Multi-Copy Venom Genes Hidden in de Novo Transcriptome Assemblies, a Cautionary Tale with the Snakelocks Sea Anemone Anemonia Sulcata (Pennant, 1977). **Toxicon**, v. 108, p. 184–188, dez. 2015.
96. MASCI, P.; WHITAKER, A.; SPARROW, L.; DE JERSEY, J.; WINZOR, D.; WATTERS, D.; LAVIN, M.; GAFFNEY, P. Textilinins from Pseudonaja textilis textilis. Characterization of two plasmin inhibitors that reduce bleeding in an animal model. **Blood coagulation & fibrinolysis**, v. 11, n. 4, p. 385–393, 2000.
97. MAURIN, K. J. L. An empirical guide for producing a dated phylogeny with treePL in a maximum likelihood framework. 2020. Disponível em: <https://arxiv.org/abs/2008.07054>. Acesso em: 7 jul. 2023.
98. MEBS, D. Toxicity in Animals. Trends in Evolution? **Toxicon**, v. 39, n. 1, p. 87–96, jan. 2001.
99. MESSIAS, R. M. **Transformações em dados composicionais para a aplicação da análise de componentes principais**. 2016. Universidade de São Paulo, São Paulo, 2016. Disponível em: <http://www.teses.usp.br/teses/disponiveis/45/45133/tde-12072016-211056/>. Acesso em: 7 jul. 2023.

100. MIRDITA, M.; SCHÜTZE, K.; MORIWAKI, Y.; HEO, L.; OVCHINNIKOV, S.; STEINEGGER, M. ColabFold: Making Protein Folding Accessible to All. **Nature Methods**, v. 19, n. 6, p. 679–682, jun. 2022.
101. MISTRY, J.; FINN, R. D.; EDDY, S. R.; BATEMAN, A.; PUNTA, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. **Nucleic acids research**, v. 41, n. 12, p. e121–e121, 2013.
102. MODAHL, C. M.; FRIETZE, S.; MACKESSY, S. P. Transcriptome-Facilitated Proteomic Characterization of Rear-Fanged Snake Venoms Reveal Abundant Metalloproteinases with Enhanced Activity. **Journal of Proteomics**, v. 187, p. 223–234, set. 2018.
103. MOHAMED ABD EL-AZIZ; GARCIA SOARES; STOCKAND. Snake Venoms in Drug Discovery: Valuable Therapeutic Tools for Life Saving. **Toxins**, v. 11, n. 10, p. 564, 25 set. 2019.
104. MORAES-DA-SILVA, A.; AMARO, R. C.; NUNES, P. M. S.; RODRIGUES, M. T.; CURCIO, F. F. Long known, brand new, and possibly threatened: a new species of watersnake of the genus Helicops Wagler, 1828 (Serpentes; Xenodontinae) from the Tocantins-Araguaia River Basin, Brazil. **Zootaxa**, v. 4903, n. 2, 7 jan. 2021. Disponível em: <https://www.mapress.com/zt/article/view/zootaxa.4903.2.3>. Acesso em: 24 nov. 2023.
105. MORAES-DA-SILVA, A.; AMARO, R. C.; NUNES, P. M. S.; STRÜSSMANN, C.; TEIXEIRA, M. J.; ANDRADE, A. Jr.; SUDRÉ, V.; RECODER, R.; RODRIGUES, M. T.; CURCIO, F. F. Chance, luck and a fortunate finding: a new species of watersnake of the genus Helicops Wagler, 1828 (Serpentes: Xenodontinae), from the Brazilian Pantanal wetlands. **Zootaxa**, v. 4651, n. 3, 6 ago. 2019. Disponível em: <https://www.mapress.com/zt/article/view/zootaxa.4651.3.3>. Acesso em: 24 nov. 2023.
106. MORAIS-ZANI, K. D.; SERINO-SILVA, C.; GALIZIO, N. D. C.; TASIMA, L. J.; PAGOTTO, J. F.; ROCHA, M. M. T. D.; MARCELINO, J. R.; SANT'ANNA, S. S.; TASHIMA, A. K.; TANAKA-AZEVEDO, A. M.; GREGO, K. F. Does the Administration of Pilocarpine Prior to Venom Milking Influence the Composition of Micrurus Corallinus Venom? **Journal of Proteomics**, v. 174, p. 17–27, mar. 2018.
107. NACHTIGALL, P. G.; KASHIWABARA, A. Y.; DURHAM, A. M. CodAn: Predictive Models for Precise Identification of Coding Regions in Eukaryotic Transcripts. **Briefings in Bioinformatics**, v. 22, n. 3, p. bbaa045, 20 maio 2021a.
108. NACHTIGALL, P. G.; RAUTSAW, R. M.; ELLSWORTH, S. A.; MASON, A. J.; ROKYTA, D. R.; PARKINSON, C. L.; JUNQUEIRA-DE-AZEVEDO, I. L. M. ToxCodAn: A New Toxin Annotator and Guide to Venom Gland Transcriptomics. **Briefings in Bioinformatics**, v. 22, n. 5, p. bbab095, 2 set. 2021b.
109. NIP, K. M.; HAFEZQORANI, S.; GAGALOVA, K. K.; CHIU, R.; YANG, C.; WARREN, R. L.; BIROL, I. Reference-Free Assembly of Long-Read Transcriptome Sequencing Data with RNA-Bloom2. **Nature Communications**, v. 14, n. 1, p. 2940, 22 maio 2023.
110. OLIVEIRA, L. D.; SCARTOZZONI, R. R.; ALMEIDA-SANTOS, S. M. D.; JARED, C.; ANTONIAZZI, M. M.; SALOMÃO, M. D. G. Morphology of Duvernoy's Glands and Maxillary Teeth and a Possible Function of the Duvernoy's Gland Secretion in *Helicops Modestus* Günther, 1861 (Serpentes: Xenodontinae). **South American Journal of Herpetology**, v. 11, n. 1, p. 54–65, abr. 2016.
111. ORON, U.; KINAMON, S.; BDOLAH, A. Asynchrony in the Synthesis of Secretory Proteins in the Venom Gland of the Snake *Vipera Palaestinae*. **Biochemical Journal**, v. 174, n. 3, p. 733–739, 15 set. 1978.

112. PORAN, N. S.; COSS, R. G.; BENJAMINI, E. Resistance of California Ground Squirrels (Spermophilus Beecheyi) to the Venom of the Northern Pacific Rattlesnake (Crotalus Viridis Oreganus): A Study of Adaptive Variation. **Toxicon**, v. 25, n. 7, p. 767–777, jan. 1987.
113. R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing, 2022.
114. RICE, P.; LONGDEN, I.; BLEASBY, A. EMBOSS: The European Molecular Biology Open Software Suite. **Trends in Genetics**, v. 16, n. 6, p. 276–277, jun. 2000.
115. RINCON-FILHO, S.; NAVES-DE-SOUZA, D. L.; LOPES-DE-SOUZA, L.; SILVANO-DE-OLIVEIRA, J.; BONILLA FERREYRA, C.; COSTAL-OLIVEIRA, F.; GUERRA-DUARTE, C.; CHÁVEZ-OLÓRTEGUI, C. Micrurus Surinamensis Peruvian Snake Venom: Cytotoxic Activity and Purification of a C-Type Lectin Protein (Ms-CTL) Highly Toxic to Cardiomyoblast-Derived H9c2 Cells. **International Journal of Biological Macromolecules**, v. 164, p. 1908–1915, dez. 2020.
116. ROKYTA, D. R.; LEMMON, A. R.; MARGRES, M. J.; ARONOW, K. The Venom-Gland Transcriptome of the Eastern Diamondback Rattlesnake (Crotalus Adamanteus). **BMC Genomics**, v. 13, n. 1, p. 312, 2012.
117. ROKYTA, D. R.; MARGRES, M. J.; CALVIN, K. Post-Transcriptional Mechanisms Contribute Little to Phenotypic Variation in Snake Venoms. **G3 Genes|Genomes| Genetics**, v. 5, n. 11, p. 2375–2382, 1 nov. 2015.
118. ROKYTA, D. R.; WRAY, K. P.; LEMMON, A. R.; LEMMON, E. M.; CAUDLE, S. B. A High-Throughput Venom-Gland Transcriptome for the Eastern Diamondback Rattlesnake (Crotalus Adamanteus) and Evidence for Pervasive Positive Selection across Toxin Classes. **Toxicon**, v. 57, n. 5, p. 657–671, abr. 2011.
119. RONQUIST, F.; TESLENKO, M.; VAN DER MARK, P.; AYRES, D. L.; DARLING, A.; HÖHNA, S.; LARGET, B.; LIU, L.; SUCHARD, M. A.; HUELSENBECK, J. P. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. **Systematic Biology**, v. 61, n. 3, p. 539–542, 1 maio 2012.
120. ROSENBERG, H. I. An Improved Method for Collecting Secretion from Duvernoy's Gland of Colubrid Snakes. **Copeia**, v. 1992, n. 1, p. 244, 3 fev. 1992.
121. ROSENBERG, H. I.; BDOLAH, A.; KOCHVA, E. Lethal Factors and Enzymes in the Secretion from Duvernoy's Gland of Three Colubrid Snakes. **Journal of Experimental Zoology**, v. 233, n. 1, p. 5–14, jan. 1985.
122. SANZ, L.; DE FREITAS-LIMA, L. N.; QUESADA-BERNAT, S.; GRAÇA-DE-SOUZA, V. K.; SOARES, A. M.; CALDERÓN, L. D. A.; CALVETE, J. J.; CALDEIRA, C. A. S. Comparative Venomics of Brazilian Coral Snakes: Micrurus Frontalis, Micrurus Spixii Spixii, and Micrurus Surinamensis. **Toxicon**, v. 166, p. 39–45, ago. 2019.
123. SAYYARI, E.; MIRARAB, S. Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. **Molecular Biology and Evolution**, v. 33, n. 7, p. 1654–1668, jul. 2016.
124. SCARTOZZONI, R. R. Morfologia de serpentes aquáticas neotropicais: um estudo comparativo. **São Paulo: Universidade de São Paulo**, 2005.
125. SCARTOZZONI, R. R. **Estratégias reprodutivas e ecologia alimentar de serpentes aquáticas da tribo Hydropsini (Dipsadidae, Xenodontinae).** 2009. Universidade de São Paulo, 2009.
126. SCHENDEL; RASH; JENNER; UNDHEIM. The Diversity of Venom: The Importance of Behavior and Venom System Morphology in Understanding Its Ecology and Evolution. **Toxins**, v. 11, n. 11, p. 666, 14 nov. 2019.

127. SCHRAMER, T. D.; RAUTSAW, R. M.; BAYONA-SERRANO, J. D.; NYSTROM, G. S.; WEST, T. R.; ORTIZ-MEDINA, J. A.; SABIDO-ALPUCHE, B.; MENESES-MILLÁN, M.; BORJA, M.; JUNQUEIRA-DE-AZEVEDO, I. L. M.; ROKYTA, D. R.; PARKINSON, C. L. An Integrative View of the Toxic Potential of Conophis Lineatus (Dipsadidae: Xenodontinae), a Medically Relevant Rear-Fanged Snake. **Toxicon**, v. 205, p. 38–52, jan. 2022.
128. SCHRÖDINGER, LLC. **The JyMOL Molecular Graphics Development Component, Version 1.8**. nov. 2015. .
129. SCHWEITZ, H.; HEURTEAUX, C.; BOIS, P.; MOINIER, D.; ROMEY, G.; LAZDUNSKI, M. Calcicludine, a Venom Peptide of the Kunitz-Type Protease Inhibitor Family, Is a Potent Blocker of High-Threshold Ca2+ Channels with a High Affinity for L-Type Channels in Cerebellar Granule Neurons. **Proceedings of the National Academy of Sciences**, v. 91, n. 3, p. 878–882, fev. 1994.
130. SHAFQAT, J.; BEG, O. U.; YIN, S.; ZAIDI, Z. H.; JÖRNVALL, H. Primary Structure and Functional Properties of Cobra (*Naja Naja Naja*) Venom Kunitz-type Trypsin Inhibitor. **European Journal of Biochemistry**, v. 194, n. 2, p. 337–341, dez. 1990.
131. SHEVCHENKO, A.; TOMAS, H.; HAVLI, J.; OLSEN, J. V.; MANN, M. In-Gel Digestion for Mass Spectrometric Characterization of Proteins and Proteomes. **Nature Protocols**, v. 1, n. 6, p. 2856–2860, dez. 2006.
132. SILVA, A. M. D.; MENDES, V. K. D. G.; MONTEIRO, W. M.; BERNARDE, P. S. Non-venomous snakebites in the Western Brazilian Amazon. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 52, p. e20190120, 2019.
133. SMITH-UNNA, R.; BOURSNELL, C.; PATRO, R.; HIBBERD, J. M.; KELLY, S. TransRate: Reference-Free Quality Assessment of de Novo Transcriptome Assemblies. **Genome Research**, v. 26, n. 8, p. 1134–1144, ago. 2016.
134. SONODA, G. G.; TOBARUELA, E. D. C.; NORENBURG, J.; FABI, J. P.; ANDRADE, S. C. S. Venomous Noodles: The Evolution of Toxins in Nemertea through Positive Selection and Gene Duplication. **Toxins**, v. 15, n. 11, p. 650, 12 nov. 2023.
135. STAMATAKIS, A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. **Bioinformatics**, v. 30, n. 9, p. 1312–1313, 1 maio 2014.
136. SUNAGAR, K.; UNDHEIM, E.; CHAN, A.; KOLUDAROV, I.; MUÑOZ-GÓMEZ, S.; ANTUNES, A.; FRY, B. Evolution Stings: The Origin and Diversification of Scorpion Toxin Peptide Scaffolds. **Toxins**, v. 5, n. 12, p. 2456–2487, 13 dez. 2013.
137. SUYAMA, M.; TORRENTS, D.; BORK, P. PAL2NAL: Robust Conversion of Protein Sequence Alignments into the Corresponding Codon Alignments. **Nucleic Acids Research**, v. 34, n. Web Server, p. W609–W612, 1 jul. 2006.
138. SWINDELL, S. R.; PLASTERER, T. N. SEQMAN. *Em*: SWINDELL, S. R. **Sequence Data Analysis Guidebook**. Totowa, NJ: Humana Press, 1997. p. 75–89.
139. 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**, v. 16, n. 1, p. 687, dez. 2015a.
140. TAN, C. H.; TAN, K. Y.; LIM, S. E.; TAN, N. H. Venomics of the Beaked Sea Snake, Hydrophis Schistosus: A Minimalist Toxin Arsenal and Its Cross-Neutralization by Heterologous Antivenoms. **Journal of Proteomics**, v. 126, p. 121–130, ago. 2015b.
141. TAN, C. H.; WONG, K. Y.; TAN, K. Y.; TAN, N. H. Venom Proteome of the Yellow-Lipped Sea Krait, Laticauda Colubrina from Bali: Insights into Subvenomic Diversity,

Venom Antigenicity and Cross-Neutralization by Antivenom. **Journal of Proteomics**, v. 166, p. 48–58, ago. 2017.

142. TAVARES-PINHEIRO, R.; MELO, F. S.; DE FIGUEIREDO, V. A. M. B.; SANCHES, P. R.; ANAISSI, J. S. C.; SANTANA, M. M. S.; OLIVEIRA-SOUZA, A. E.; DOS SANTOS REIS, T.; REBELO, G. L.; MELO, F. T. V.; OTHERS. “In living color”: predation by the coralsnake Micrurus surinamensis (Cuvier, 1816)(Serpentes: Elapidae) on a knifefish in the eastern Amazon, Brazil. **Herpetology Notes**, v. 14, p. 755–758, 2021.
143. THE UNIPROT CONSORTIUM; BATEMAN, A.; MARTIN, M.-J.; ORCHARD, S.; MAGRANE, M.; AHMAD, S.; ALPI, E.; BOWLER-BARNETT, E. H.; BRITTO, R.; BYE-A-JEE, H.; CUKURA, A.; DENNY, P.; DOGAN, T.; EBENEZER, T.; FAN, J.; GARMIRI, P.; DA COSTA GONZALES, L. J.; HATTON-ELLIS, E.; HUSSEIN, A.; IGNATCHENKO, A.; INSANA, G.; ISHTIAQ, R.; JOSHI, V.; JYOTHI, D.; KANDASAAMY, S.; LOCK, A.; LUCIANI, A.; LUGARIC, M.; LUO, J.; LUSSI, Y.; MACDOUGALL, A.; MADEIRA, F.; MAHMOUDY, M.; MISHRA, A.; MOULANG, K.; NIGHTINGALE, A.; PUNDIR, S.; QI, G.; RAJ, S.; RAPOSO, P.; RICE, D. L.; SAIDI, R.; SANTOS, R.; SPERETTA, E.; STEPHENSON, J.; TOTOO, P.; TURNER, E.; TYAGI, N.; VASUDEV, P.; WARNER, K.; WATKINS, X.; ZARU, R.; ZELLNER, H.; BRIDGE, A. J.; AIMO, L.; ARGOUD-PUY, G.; AUCHINCLOSS, A. H.; AXELSEN, K. B.; BANSAL, P.; BARATIN, D.; BATISTA NETO, T. M.; BLATTER, M.-C.; BOLLEMAN, J. T.; BOUTET, E.; BREUZA, L.; GIL, B. C.; CASALS-CASAS, C.; ECHIOUKH, K. C.; COUDERT, E.; CUCHE, B.; DE CASTRO, E.; ESTREICHER, A.; FAMIGLIETTI, M. L.; FEUERMANN, M.; GASTEIGER, E.; GAUDET, P.; GEHANT, S.; GERRITSEN, V.; GOS, A.; GRUAZ, N.; HULO, C.; HYKA-NOUSPIKEL, N.; JUNGO, F.; KERHORNOU, A.; LE MERCIER, P.; LIEBERHERR, D.; MASSON, P.; MORGAT, A.; MUTHUKRISHNAN, V.; PAESANO, S.; PEDRUZZI, I.; PILBOUT, S.; POURCEL, L.; POUX, S.; POZZATO, M.; PRUESS, M.; REDASCHI, N.; RIVOIRE, C.; SIGRIST, C. J. A.; SONESSON, K.; SUNDARAM, S.; WU, C. H.; ARIGHI, C. N.; ARMINSKI, L.; CHEN, C.; CHEN, Y.; HUANG, H.; LAIHO, K.; MCGARVEY, P.; NATALE, D. A.; ROSS, K.; VINAYAKA, C. R.; WANG, Q.; WANG, Y.; ZHANG, J. UniProt: The Universal Protein Knowledgebase in 2023. **Nucleic Acids Research**, v. 51, n. D1, p. D523–D531, 6 jan. 2023.
144. TIOYAMA, E. C.; BAYONA-SERRANO, J. D.; PORTES-JUNIOR, J. A.; NACHTIGALL, P. G.; DE SOUZA, V. C.; BERALDO-NETO, E.; GRAZZIOTIN, F. G.; JUNQUEIRA-DE-AZEVEDO, I. L. M.; MOURA-DA-SILVA, A. M.; FREITAS-DE-SOUSA, L. A. The Venom Composition of the Snake Tribe Philodryadini: ‘Omic’ Techniques Reveal Intergeneric Variability among South American Racers. **Toxins**, v. 15, n. 7, p. 415, 27 jun. 2023.
145. TREVINE, V. C.; GRAZZIOTIN, F. G.; GIRAUDO, A.; SALLABERRY-PINCHEIRA, N.; VIANNA, J. A.; ZAHER, H. The Systematics of Tachymenini (Serpentes, Dipsadidae): An Updated Classification Based on Molecular and Morphological Evidence. **Zoologica Scripta**, v. 51, n. 6, p. 643–663, nov. 2022.
146. TYANOVA, S.; COX, J. Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. *Em*: VON STECHOW, L. **Cancer Systems Biology**. Methods in Molecular Biology. New York, NY: Springer New York, 2018. p. 133–148.
147. TYANOVA, S.; TEMU, T.; COX, J. The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics. **Nature Protocols**, v. 11, n. 12, p. 2301–2319, dez. 2016.

148. VILLCA-CORANI, H.; NIETO-ARIZA, B.; LEÓN, R.; ROCABADO, J. A.; CHIPPAUX, J.-P.; URRA, F. A. First Reports of Envenoming by South American Water Snakes Helicops Angulatus and Hydrops Triangularis from Bolivian Amazon: A One-Year Prospective Study of Non-Front-Fanged Colubroid Snakebites. **Toxicon**, v. 202, p. 53–59, out. 2021.
149. VON REUMONT, B.; CAMPBELL, L.; JENNER, R. Quo Vadis Venomics? A Roadmap to Neglected Venomous Invertebrates. **Toxins**, v. 6, n. 12, p. 3488–3551, 19 dez. 2014.
150. WATERHOUSE, R. M.; SEPPEY, M.; SIMÃO, F. A.; MANNI, M.; IOANNIDIS, P.; KLIOUTCHNIKOV, G.; KRIVENTSEVA, E. V.; ZDOBNOV, E. M. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. **Molecular Biology and Evolution**, v. 35, n. 3, p. 543–548, 1 mar. 2018.
151. WEXLER, P.; FONGER, G. C.; WHITE, J.; WEINSTEIN, S. Toxinology: taxonomy, interpretation, and information resources. **Science & Technology Libraries**, v. 34, n. 1, p. 67–90, 2015.
152. WIŚNIEWSKI, J. R.; HEIN, M. Y.; COX, J.; MANN, M. A “Proteomic Ruler” for Protein Copy Number and Concentration Estimation without Spike-in Standards. **Molecular & Cellular Proteomics**, v. 13, n. 12, p. 3497–3506, dez. 2014.
153. XIE, B.; DASHEVSKY, D.; ROKYTA, D.; GHEZELLOU, P.; FATHINIA, B.; SHI, Q.; RICHARDSON, M. K.; FRY, B. G. Dynamic Genetic Differentiation Drives the Widespread Structural and Functional Convergent Evolution of Snake Venom Proteinaceous Toxins. **BMC Biology**, v. 20, n. 1, p. 4, 7 jan. 2022.
154. YANG, W.; FENG, J.; WANG, B.; CAO, Z.; LI, W.; WU, Y.; CHEN, Z. BF9, the First Functionally Characterized Snake Toxin Peptide with Kunitz-Type Protease and Potassium Channel Inhibiting Properties: KUNITZ-TYPE PROTEASE AND POTASSIUM CHANNEL INHIBITOR FROM SNAKE. **Journal of Biochemical and Molecular Toxicology**, v. 28, n. 2, p. 76–83, fev. 2014.
155. YIN, X.; GUO, S.; GAO, J.; LUO, L.; LIAO, X.; LI, M.; SU, H.; HUANG, Z.; XU, J.; PEI, J.; CHEN, S. Kinetic Analysis of Effects of Temperature and Time on the Regulation of Venom Expression in Bungarus Multicinctus. **Scientific Reports**, v. 10, n. 1, p. 14142, 24 ago. 2020.
156. YU, G. Using Ggtree to Visualize Data on Tree-Like Structures. **Current Protocols in Bioinformatics**, v. 69, n. 1, p. e96, mar. 2020.
157. ZAHER, H.; GRAZZIOTIN, F. G.; CADLE, J. E.; MURPHY, R. W.; MOURA-LEITE, J. C. de; BONATTO, S. L. Molecular Phylogeny of Advanced Snakes (Serpentes, Caenophidia) with an Emphasis on South American Xenodontines: A Revised Classification and Descriptions of New Taxa. **Papéis Avulsos de Zoologia (São Paulo)**, v. 49, n. 11, p. 115–153, 2009.
158. ZAHER, H.; MURPHY, R. W.; ARREDONDO, J. C.; GRABOSKI, R.; MACHADO-FILHO, P. R.; MAHLOW, K.; MONTINGELLI, G. G.; QUADROS, A. B.; ORLOV, N. L.; WILKINSON, M.; ZHANG, Y.-P.; GRAZZIOTIN, F. G. Large-Scale Molecular Phylogeny, Morphology, Divergence-Time Estimation, and the Fossil Record of Advanced Caenophidian Snakes (Squamata: Serpentes). **PLOS ONE**, v. 14, n. 5, p. e0216148, 10 maio 2019.
159. ZAHER, H.; YÁNEZ-MUÑOZ, M. H.; RODRIGUES, M. T.; GRABOSKI, R.; MACHADO, F. A.; ALTAMIRANO-BENAVIDES, M.; BONATTO, S. L.; GRAZZIOTIN, F. G. Origin and Hidden Diversity within the Poorly Known Galápagos Snake Radiation (Serpentes: Dipsadidae). **Systematics and Biodiversity**, v. 16, n. 7, p. 614–642, 3 out. 2018.

160. ZHANG, C.; RABIEE, M.; SAYYARI, E.; MIRARAB, S. ASTRAL-III: Polynomial Time Species Tree Reconstruction from Partially Resolved Gene Trees. **BMC Bioinformatics**, v. 19, n. S6, p. 153, maio 2018.
161. ZHANG, J.; KOBERT, K.; FLOURI, T.; STAMATAKIS, A. PEAR: A Fast and Accurate Illumina Paired-End reAd mergeR. **Bioinformatics**, v. 30, n. 5, p. 614–620, 1 mar. 2014.
162. ZHANG, S.; GAO, B.; ZHU, S. Target-Driven Evolution of Scorpion Toxins. **Scientific Reports**, v. 5, n. 1, p. 14973, 7 out. 2015.
163. ZHONG, M.; HANSEN, B.; NESNIDAL, M.; GOLOMBEK, A.; HALANYCH, K. M.; STRUCK, T. H. Detecting the Symplesiomorphy Trap: A Multigene Phylogenetic Analysis of Terebelliform Annelids. **BMC Evolutionary Biology**, v. 11, n. 1, p. 369, dez. 2011.
164. ZHOU, X.-D.; JIN, Y.; LU, Q.-M.; LI, D.-S.; ZHU, S.-W.; WANG, W.-Y.; XIONG, Y.-L. Purification, Characterization and Primary Structure of a Chymotrypsin Inhibitor from Naja Atra Venom. **Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology**, v. 137, n. 2, p. 219–224, fev. 2004.

10. Biografia

Formação

Bacharel em Ciências Biológicas pela USP (2016-2020) - “Molecular Evolution and Diversity of Toxins in Nemertea”

Licenciatura em Ciências Biológicas pela USP (2016-2021)

Lista de Publicações

1. Sonoda, G.G.; Tobaruela, E.d.C.; Norenburg, J.; Fabi, J.P.; Andrade, S.C.S. (2023).Venomous Noodles: The Evolution of Toxins in Nemertea through Positive Selectionand Gene Duplication. Toxins, 15,(11), 650

2. Verdes, A., Taboada, S., Hamilton, B. R., Undheim, E. A., Sonoda, G. G., Andrade, S.C., ... & Riesgo, A. (2022). Evolution, expression patterns, and distribution of novel ribbon worm predatory and defensive toxins. Molecular Biology and Evolution, 39(5), msac096.

3. Santos, C. A., Sonoda, G. G., Cortez, T., Coutinho, L. L., & Andrade, S. C. (2021). Transcriptome expression of biomineralization genes in Littoraria flava gastropod in Brazilian rocky shore reveals evidence of local adaptation. Genome biology and evolution, 13(4), evab050.