

Ministério da Saúde
Fundação Oswaldo Cruz
Centro de Pesquisas René Rachou
Programa de Pós-graduação em Ciências da Saúde

Transcritoma da glândula venenífera da serpente *Bothrops neuwiedi*: montagem,
anotação e identificação de proteínas de interesse farmacológico

por

Valéria Gonçalves de Alvarenga

Belo Horizonte

Abril / 2014

DISSERTAÇÃO MBCM-CPqRR V.G. ALVARENGA 2014

Ministério da Saúde
Fundação Oswaldo Cruz
Centro de Pesquisas René Rachou
Programa de Pós-graduação em Ciências da Saúde

Transcritoma da glândula venenífera da serpente *Bothrops neuwiedi*: montagem,
anotação e identificação de proteínas de interesse farmacológico

por

Valéria Gonçalves de Alvarenga

Dissertação apresentada com vistas à
obtenção do título de Mestre em Ciências na
área de concentração Biologia Celular e
Molecular.

Orientação: Dr. Guilherme Corrêa Oliveira

Coorientação: Dra. Maria Inácia E. Costa

Dr. Eladio O. Flores Sanchez

Belo Horizonte

Abril / 2014

Catálogo-na-fonte
Rede de Bibliotecas da FIOCRUZ
Biblioteca do CPqRR
Segemar Oliveira Magalhães CRB/6 1975

A473t Alvarenga, Valéria Gonçalves.
2014

Transcritoma da glândula venenífera da serpente *Bothrops neuwiedi*: montagem, anotação e identificação de proteínas de interesse farmacológico / Valéria Gonçalves de Alvarenga. – Belo Horizonte, 2014.

XVII, 56 f.: il.; 210 x 297mm.

Bibliografia: f.: 67- 73

Dissertação (Mestrado) – Dissertação para obtenção do título de Mestre em Ciências pelo Programa de Pós - Graduação em Ciências da Saúde do Centro de Pesquisas René Rachou. Área de concentração: Biologia Celular e Molecular.

1. Venenos de serpentes/toxicidade 2. Transcriptoma/genética 3. Metaloproteases/análise I. Título. II. Oliveira, Guilherme Corrêa (Orientação). III. Estevão-Costa, Maria Inácia (Co-orientação). IV. Sanchez, Eladio Oswaldo Flores (Co-orientação).

CDD – 22. ed. – 615.942

Ministério da Saúde
Fundação Oswaldo Cruz
Centro de Pesquisa René Rachou
Programa de Pós-graduação em ciências da Saúde

Transcritoma da glândula venenífera da serpente *Bothrops neuwiedi*: montagem,
anotação e identificação de proteínas de interesse farmacológico

Por

Valéria Gonçalves de Alvarenga

Foi avaliada pela banca examinadora composta pelos seguintes membros:

Prof. Dr. Guilherme Corrêa Oliveira (Presidente)

Prof. Dr. Marcelo Ribeiro Vasconcelos Dinis

Prof. Dra. Cristiana Ferreira Alves de Brito

Suplente: Prof. Dra. Roberta Lima Caldeira

Dissertação defendida e aprovada em: 30/04/2014

“Os teus olhos viram o meu corpo ainda informe, e no teu livro todas estas coisas foram escritas; as quais iam sendo dia-a-dia formadas, quando nem ainda uma delas havia”.

Salmos 139:16

v

AGRADECIMENTOS

A Deus, fonte de toda inspiração e sabedoria.

Aos meus pais, Brígida e Walter, meu exemplo de vida.

Ao Rodrigo, meu esposo querido, pela paciência e compreensão.

À Eloisa, minha filha, minha herança.

Aos meus irmãos Julio e Sander pelo apoio sempre presente.

À Fundação Oswaldo Cruz, ao Centro de Pesquisas René Rachou e ao Programa de Pós-graduação em Ciências da Saúde pela oportunidade ofertada.

À Fundação Ezequiel Dias pela liberação para realização deste mestrado.

Ao Dr. Guilherme pela oportunidade e orientação.

À Dra. Maria Inácia pelo incentivo.

Ao Dr. Eladio pela confiança deposita em mim.

À Dra. Laila pelas lições ensinadas.

À Dra. Larissa por tudo que me instruiu, com quem aprendi muito.

Ao Dr. Fabiano pelo enorme auxílio.

À equipe do CEBIO, Anderson, Francislon, Ângela, Fabiano, Larissa, Laura, Juliana, Mariana, Fausto, Flávio e Anna.

Aos amigos da FUNED, Gena, Paula, Rebeca, Lutiana, Ana Valentin, Patrícia, Dra. Consuelo, Siléa, Alcides, Gustavo, Jomara, Dra. Marta, Aristeu.

À Dona Fátima e Milton, Patrícia, Ana Paula, Dora Lúcia, tia Eliane, Jane, Dimas Gilberto, Mislene e a todos os que participaram deste trajeto.

À Biblioteca do CPqRR em prover acesso gratuito local e remoto à informação técnica científica em saúde custeada com recursos públicos federais, integrante do rol de referências desta dissertação, também pela catalogação e normalização da mesma.

Muito Obrigada!

Ao Apoio Financeiro da FAPEMIG (PCRH 90303/12), CAPES (23038.006280/2011-07), CNPq(482502/2012-6) e NIH (TW007012 to GO).

SUMÁRIO

Lista de Figuras.....	X
Lista de Quadros e Tabelas.....	XIII
Lista de Tabelas do Anexo 1.....	XIV
Lista de Tabelas do Anexo 2.....	XIV
Lista de abreviaturas e símbolos.....	XV
RESUMO.....	XVI
ABSTRACT.....	XVII
1 INTRODUÇÃO.....	18
1.1 Serpentes e ofidismo no Brasil.....	18
1.2 A diversidade molecular do veneno de serpentes.....	20
1.3 As Metaloproteinases de Veneno de Serpentes.....	21
2 JUSTIFICATIVA.....	25
3 OBJETIVOS.....	26
3.1 Objetivo Geral.....	26
3.2 Objetivos Específicos.....	26
4. METODOS.....	27
4.1 Extrações das glândulas de veneno de <i>Bothrops neuwiedi</i>	28
4.2 Construção da biblioteca de Cdna.....	28
4.2.1 Isolamento do RNA total.....	28
4.2.2 Depleção do RNA ribossomal.....	29
4.2.3 Preparo do cDNA para o sequenciamento.....	29
4.3 Sequenciamento do cDNA.....	30
4.4 Análise do transcrito de <i>Bothrops neuwiedi</i>	31
4.4.1 Análises da qualidade das sequências do transcrito.....	31
4.4.2 Montagem do transcrito de <i>Bothrops neuwiedi</i>	31
4.4.3 Mapeamento e quantificação das sequências	32
4.4.4 Busca por similaridades em banco de dados e anotação do transcrito.....	32
4.5 Reconstrução filogenética de SVMPS.....	32
4.5.1 Seleção de potenciais homólogos.....	33
4.5.2 Alinhamento de sequências proteicas.....	34
4.5.3 Teste de modelos evolutivos.....	35

4.5.4	Reconstrução filogenética.....	35
5	RESULTADOS.....	37
5.1	Análises do transcrito da glândula venenífera de <i>Bothrops neuwiedi</i>.....	37
5.1.1	Análise da qualidade dos experimentos.....	37
5.1.2	Transcritos identificados no veneno de <i>Bothrops neuwiedi</i>	40
5.2	Reconstrução da história evolutiva de SVMPs.....	45
5.2.1	Identificação de SVMPs de Viperidae.....	45
5.2.2	Análise das relações evolutivas de SVMPs.....	46
6	DISCUSSÃO.....	51
7	CONCLUSÕES E PERSPECTIVAS FUTURAS.....	54
8	ANEXOS.....	55
8.1	Anexo 1 - Tabelas com descrição dos principais transcritos que codificam proteínas do veneno de <i>Bothrops neuwiedi</i>	55
8.2	Anexo 2 - Tabelas da montagem dos transcritos de SVMP em cada um das bibliotecas.....	61
8.3	Anexo 3 - Alinhamento das sequências do domínio catalítico de SVMP-PI..	64
9	REFERÊNCIAS BIBLIOGRÁFICAS.....	67

Lista de Figuras

- Figura 01** - A extração do veneno de *Bothrops neuwiedi*. Foto: Leonardo Noronha..... 19
- Figura 2** - Acidentes com serpentes no Brasil em 2012. Exemplos do serpentário da Fundação Ezequiel Dias (FUNED) em Belo Horizonte, Minas Gerais. Fotos de Leonardo Noronha. Gráfico indicando os acidentes causados por serpentes da família Viperidae e Elapidae. Fonte: DATASUS/Brasil, 2013..... 19
- Figura 3** - Classificação de metaloproteínas de veneno de serpentes. Modificado de Fox e Serrano, 2008. **P**: corresponde ao peptídeo sinal; **Pro**: corresponde ao domínio pró-peptídeo presente nos transcritos de metaloproteína; **Proteinase**: corresponde ao domínio catalítico das metaloproteínas também denominado domínio reprotina; **S**: corresponde a uma porção denominada *peptideo spacer* que é um espaço estrutural entre o domínio catalítico e os demais domínios presentes na estrutura das SVMPs; **Dis**: corresponde ao domínio desintegrina nas SVMPs-P11; **Dis-Like**: corresponde ao domínio tipo desintegrina nas SVMPs-P111; **Cys-Rich**: corresponde ao domínio rico em cisteínas; **Lec**: corresponde às subunidades de lectinas tipo C que pode interagir com o domínio rico em cisteínas das SVMPs-P111..... 23
- Figura 4** - Resumo da ação das SVMPs sobre o sistema hemostático. Esquema resumido da cascata de coagulação e do sistema fibrinogênico. Representação da ação das SVMPs sobre proteínas da membrana basal e sobre a interação de receptores integrina e fibrinogênio..... 24
- Figura 5** - Fluxo de trabalho para a análise do transcrito da glândula venenifera de *Bothrops neuwiedi*. Consiste na extração das glândulas de veneno, isolamento do RNA total, depleção do RNA ribossomal e preparo da amostra para o sequenciamento, seguido pelo sequenciamento na plataforma Ion Torrent, montagem dos transcritos, mapeamento e anotação do transcrito..... 27
- Figura 6** - Fluxo de trabalho para a reconstrução da história evolutiva de SVMPs. Os principais passos são seleção dos potenciais homólogos de

SVMPs de serpentes da família Viperidae, alinhamento e edição manual das sequências de aminoácidos, teste do modelo evolutivo e reconstrução filogenética.....	33
Figura 7 - RNA total da glândula de veneno de <i>B. neuwiedi</i> . 1 – Brumadinho (RIN= 6,6); 2 e 3 – Nova Lima (RIN= 5,8 e 7,2, respectivamente); 4 – Conselheiro Lafaiete (RIN= 7,1); L – Marcador de massa molecular.....	37
Figura 8 - Resultado da análise dos dados de sequenciamento realizada pelo programa FastQC. A maioria das bases apresenta um valor de <i>phred</i> acima de 20. Na linha vertical estão os valores de <i>escores phred</i> e na horizontal, a extensão em nucleotídeos das sequências.....	38
Figura 9 - Média da qualidade das sequências originais. As setas indicam os valores de <i>phred</i> 20, a <i>trimagem</i> foi feita por este valor. Na linha vertical está o número de sequências da biblioteca e na horizontal, a medida da qualidade das sequências dada pelo <i>score phred</i> . 1 – procedente de Brumadinho, 2 e 3 – procedentes de Nova Lima, 4 – procedente de Conselheiro Lafaiete.....	39
Figura 10 - Gráfico da média de expressão dos principais transcritos que codificam proteínas do veneno. SVMP, metaloproteinase de veneno de serpente; Snaclecs, proteínas do veneno de serpentes semelhantes à Lectina tipo C; PLA2 fosfolipase A2; ANP, peptídeo natriurético atrial; SVSP, serinoproteinase de veneno de serpente; LAAO, L-aminoácido oxidase; three finger, proteínas três dedos; iPLA2, inibidor de fosfolipases A2; NGF fator de crescimento nervoso; Hial, Hialuronidase; iSVMP, inibidor de metaloproteinase de veneno de serpente.....	45
Figura 11 - Distribuição de SVMPs de Viperidae no UniProt. A , Distribuição de sequências de aminoácidos de SVMPs de diferentes gêneros de Viperidae depositadas no UniProt (acessado em 14/06/2013). B , As proteínas identificadas em cada subfamília (PI, PII e PIII) estão indicadas em cor.....	46
Figura 12 - Relações evolutivas das três subfamílias de SVMPs de Viperidae. Análise de máxima verossimilhança de 266 sequências de aminoácidos (205 sítios) do domínio catalítico (PF01421) de SVMPs de diferentes espécies de Viperidae. Proteínas das subfamílias SVMP-I (verde), SVMP-II (vermelho) e SVMP-III (azul) estão indicadas. JTT foi o modelo	

evolutivo que melhor se adequou aos dados. Valores de apoio para os nós da árvore foram estimados utilizando *Akaike Likelihood Ratio test* (aLRT). Altos valores de apoio estão destacados por círculos vermelhos..... 48

Figura 13 - Relações evolutivas da subfamília SVMP-PI de Viperidae. Análise de máxima verossimilhança de 36 sequências de aminoácidos (202 sítios) do domínio catalítico (PF01421) de SVMPs de diferentes espécies de Viperidae. Proteínas das subfamílias SVMP-I (verde), SVMP-II (vermelho) e SVMP-III (azul) estão indicadas. WAG foi o modelo evolutivo que melhor se adequou aos dados. Valores de apoio para os nós da árvore foram estimados utilizando *Akaike Likelihood Ratio test* (aLRT)..... 49

Lista de Quadros e Tabelas

Quadro 1 - Estratégia de busca de potenciais homólogos de SVMPs de Viperidae no UniProt.....	34
Tabela 1 - Cálculo estatístico da montagem dos <i>contigs</i> realizada pelo programa Trinity. ¹ Biblioteca de cDNA de glândulas de <i>Bothrops neuwiedi</i> , 1 – Brumadinho, 2 e 3 – Nova Lima, 4 – Conselheiro Lafaiete,. ² Número de <i>reads</i> geradas em milhões. ³ Número de transcritos formados com a montagem pelo programa Trinity. ⁴ O N50 é o valor do comprimento do <i>contig</i> a partir do qual, somando o comprimento dele e de todos os <i>contigs</i> maiores do que ele, o conjunto ultrapassa 50% do comprimento da montagem. ⁵ Indica o tamanho médio dos transcritos medido pelo número de bases em cada um. ⁶ Expressa o número de transcritos com tamanho maior do que 1 kilobase. ⁷ Número total de bases nos transcritos maiores do que 1 kilobase.....	40
Tabela 2 - Quantidade de transcritos montados em cada biblioteca. A Tabela apresenta ainda a soma de todos os transcritos e a quantidade de transcritos montados no transcrito referência. ¹ A média e a mediana da expressão dos transcritos foram calculadas com o intuito de permitir a avaliação do nível de expressão dos principais transcritos que codificam as proteínas que compõem o veneno desta espécie.....	41
Tabela 3 - Principais transcritos do veneno de <i>Bothrops neuwiedi</i> identificados.....	42
Tabela 4 - Principais transcritos dos domínios de SVMPs encontrados no veneno de <i>Bothrops neuwiedi</i>	43
Tabela 5 - Subfamílias de SVMPs encontradas em cada Biblioteca separadamente.....	44

Lista de Tabelas do Anexo 1

Tabela 1 - Mapeamento dos transcritos que codificam fosfolipases A2 (PLA2) identificados na montagem do transcrito referência correção feita pelo valor de RPKM.....	55
Tabela 2 - Mapeamento dos transcritos que codificam proteínas do veneno de serpentes similares às Lectinas tipo C.....	56
Tabela 3 - Mapeamento dos transcritos que codificam inibidores de PLA2.....	57
Tabela 4 - mapeamento dos transcritos que codificam peptídeos natriuréticos potencializadores de Bradicinina identificados no transcrito.....	57
Tabela 5 - Mapeamento dos transcritos que codificam serino-proteinases no veneno desta espécie.....	58
Tabela 6 - Mapeamento dos transcritos que codificam proteínas com motivos três dedos (<i>three finger</i>).....	59
Tabela 7 - Transcritos que codificam L-aminoácido oxidases no veneno.....	59
Tabela 8 - Mapeamento dos transcritos que codificam os domínios presente em metaloproteinases de veneno de serpentes. ¹ Identificador Pfam (Banco de dados público de domínios proteicos).....	60

Lista de Tabelas do Anexo 2

Tabela 1 - Transcritos de SVMPs identificados na Biblioteca 1 (espécime proveniente da cidade de Brumadinho).....	61
Tabela 2 - Transcritos de SVMPs identificados na Biblioteca 2 (espécime proveniente da cidade de Nova Lima).....	62
Tabela 3 - Transcritos de SVMPs identificados na biblioteca 3 (espécime proveniente da cidade de Nova Lima).....	62
Tabela 4 - Transcritos de SVMPs identificado na Biblioteca 4 (espécime proveniente da cidade de Conselheiro Lafaiete).....	63

Lista de abreviatura e símbolos

AIC – *Akaike information criterion*

aLRT – *approximate Likelihood Ratio Test*

ANP – Peptídeo Natriurético Atrial

BLAST – *Basic Local Alignment Search Tool*

cDNA – DNA complementar

CEBio – Centro de Excelência em Bioinformática

CPqRR – Centro de Pesquisa René Rachou

DEPC – dietilpirocarbonato

DNA – ácido desoxirribonucléico

EDTA – ácido etilenodiamino tetracético

FIOCRUZ – Fundação Oswaldo Cruz

FUNED – Fundação Ezequiel Dias

FX – Fator X

iPLA2 – inibidor de fosfolipase A2

iSVMP – inibidor de Metaloproteinase de Veneno de Serpentes

LAAO – L-aminoácido oxidase

M – molar

Mb – megabase

ML – *maximum Likelihood*; máxima Verossimilhança

mRNA – RNA mensageiro

NCBI – *National Center for Biotechnology Information*

NGF – Fator de Crescimento Nervoso

pb – pares de bases

PCR – reação em cadeia da polimerase

PLA2 – Fosfolipase A2

RNA – ácido ribonucleico

RPKM – *Read per kilobase per million*

SVMPs – metaloproteinases de veneno de serpentes

SVSP – Serino-proteinase de veneno de serpentes

TBE – Tris Borato EDTA

UV – Ultravioleta

RESUMO

Os componentes de venenos de serpentes são importantes ferramentas para a pesquisa científica, desenvolvimento de drogas, e para o diagnóstico de várias doenças. Descrevemos a montagem *de novo* e análise do transcrito da glândula de veneno de uma serpente amplamente distribuída no Brasil a *Bothrops neuwiedi*. Com base em 9.500.000 sequências identificamos várias sequências inteiras codificantes de toxinas. A mais abundante expressão foi de transcritos de metaloproteinases de veneno de serpentes (SVMPs), que apresentou o maior percentual de sequências mapeadas em um transcriptoma de referência (34,40 %), seguido por lectinas tipo C (25,20 %), fosfolipases A2 (14,20%) e serino-proteinases (7,51 %). Devido à importância fisiopatológica no envenenamento e seu potencial uso como um modelo para o desenvolvimento biotecnológico de fármacos, as SVMPs tornaram-se o principal alvo deste estudo, para o qual realizamos análises filogenéticas com o objetivo de contribuir para uma melhor compreensão da biodiversidade e dos mecanismos moleculares subjacentes a evolução destas proteínas, e também contribuir para a descoberta de novos compostos com potencial ação terapêutica.

ABSTRACT

The snake venom's components are sources of drug and physiological research tools for diagnosis of several diseases. We describe the *de novo* assembly and analysis of the venom-gland transcriptome of a widely distributed snake in Brazil (*Bothrops neuwiedi*). Based on 9.500.000 reads of quality-filtered, we identified several full-length toxin-coding sequences. The most highly expressed group of toxin was the SVMP_s (snake venom metalloproteinase), which accounted for the highest percentage of reads to the reference transcriptome (34.40%), followed by C-type lectins (25.20%), phospholipases A2 (14.20%) and serine proteinases (7.51%). Due to pathophysiological importance in the poisoning and its potential use as a model for biotechnology development, the snake venom metalloproteinases (SVMP) became the main target in this study. Therefore we performed phylogenetic analyzes of the SVMP_s in order to contribute to a better understanding of the biodiversity of these proteins underlying molecular and evolutionary mechanisms, and contribute to the discovery of new compounds with have potential in therapeutics.

1 INTRODUÇÃO

1.1 Serpentes e ofidismo no Brasil

Animais peçonhentos são conhecidos pela ampla diversidade do ponto de vista taxonômico e funcional. Embora diversos animais peçonhentos sejam conhecidos, os acidentes ofídicos recebem destaque devido à alta incidência e gravidade dos casos. Neste contexto, as principais famílias de serpentes venenosas são: 1) família Viperidae (víboras), cujo principal efeito do envenenamento resulta em lesões locais graves caracterizadas por hemorragia e mionecrose além de alterações hemostáticas; 2) família Elapidae (serpente coral e cobras najas), que possui veneno primariamente neurotóxico e cardiotoxico (Du *et al.*, 2006; Leão *et al.*, 2009; Casewell *et al.*, 2009); 3) família Hydrophiidae (serpentes marinhas), cujo veneno é principalmente miotóxico; e 4) algumas espécies da família Colubridae (serpente-rateira), cuja maioria não apresenta glândula de veneno desenvolvida.

O Brasil é um dos países com maior abundância de serpentes peçonhentas, sendo que a grande maioria pertence às famílias Viperidae (responsáveis pelos acidentes botrópico, crotálico e laquélico) e Elapidae (acidente elapídico) (Albuquerque *et al.*, 2013). Ao contrário do que se imagina, o envenenamento por picadas de serpentes é considerado pela Organização Mundial de Saúde (OMS) como sendo uma doença negligenciada que afeta principalmente habitantes de áreas rurais de países tropicais (OMS, 2013).

As serpentes da família Viperidae são amplamente distribuídas no globo terrestre e lideram a lista das principais serpentes causadoras de acidentes ofídicos (Salvador *et al.*, 2013). Dentre as serpentes da família Viperidae, o gênero *Bothrops* (**Figura 1**) é o principal responsável pela maioria dos casos de envenenamento por picada de serpentes na América Latina e no Brasil (Sanchez & Swenson, 2007, Cardoso *et al.*, 2010; Albuquerque *et al.*, 2013) liderando a produção de soro antiofídico. Das notificações feitas ao DATASUS (**Figura 2**) em 2012, 72% dos casos com picada de serpentes eram do gênero *Bothrops* (Brasil, 2013).

No Brasil são encontradas 21 espécies de serpentes do gênero *Bothrops* dentre elas, a espécie *Bothrops neuwiedi* (jararaca do rabo branco ou jararaca cruzeira) estava entre as três serpentes peçonhentas que apresentavam o maior número de recebimentos de representantes no Instituto Butantan entre 1900 e 1962 (Silva & Rodrigues, 2008).



Figura 1 - A extração do veneno de *Bothrops neuwiedi*. Foto: Leonardo Noronha.

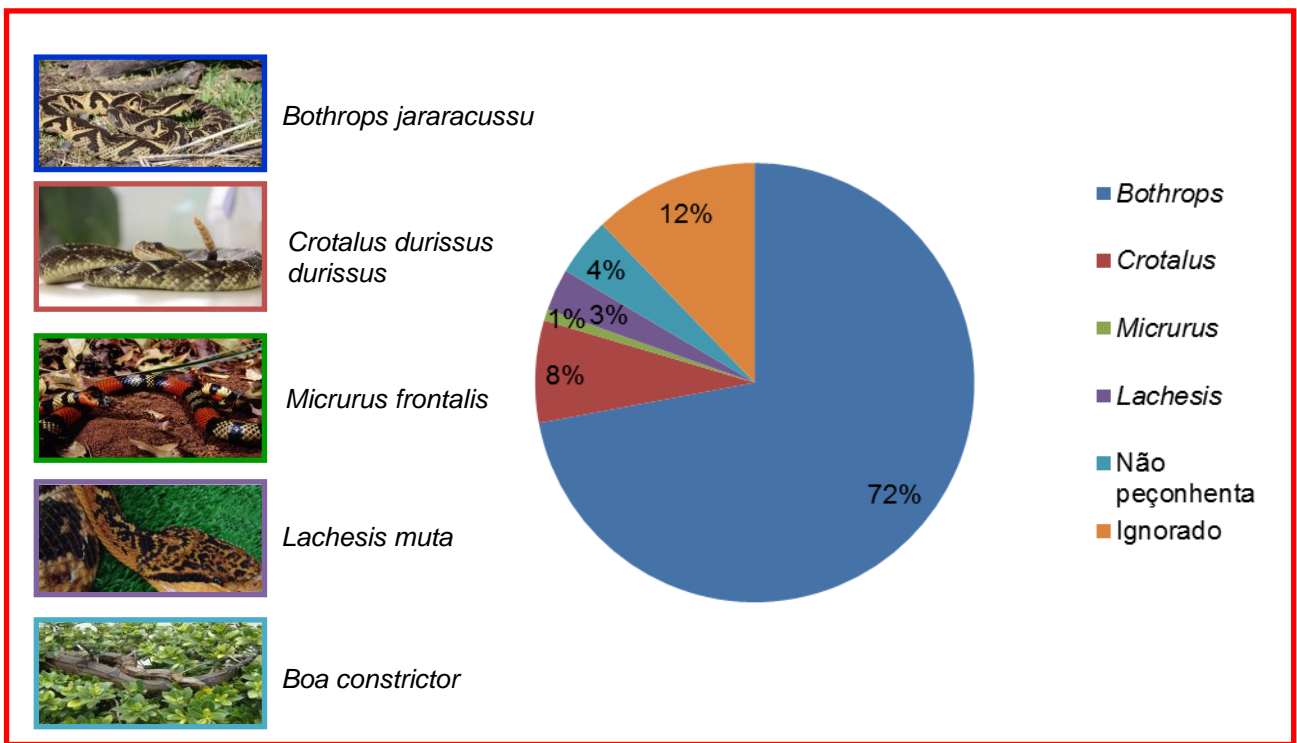


Figura 2 - Acidentes com serpentes no Brasil em 2012. Exemplos do serpentário da Fundação Ezequiel Dias (FUNED) em Belo Horizonte, Minas Gerais. Fotos de Leonardo Noronha. Gráfico indicando os acidentes causados por serpentes da família Viperidae e Elapidae. Fonte: DATASUS/Brasil, 2013.

1.2 A diversidade molecular do veneno de serpentes

O veneno é uma combinação de proteínas e peptídeos que exibem diversas funções bioquímicas e farmacológicas (Doley *et al.*, 2009), além de conter constituintes inorgânicos como cálcio, ferro, cobre, potássio, magnésio, manganês, sódio, fósforo e zinco (Friederich & Tu, 1971). O veneno das serpentes pode ser bastante diverso quanto a sua composição e atividade biológica, dependendo da espécie, idade, localização geográfica, dieta e sexo (Junqueira-de-Azevedo *et al.*, 2006).

O envenenamento por mordedura de serpentes da família Viperidae provoca hemorragia local e sistêmica, hipotensão, inflamação e oclusão vascular trombótica (Sanchez & Swenson, 2007). O veneno de serpentes do gênero *Bothrops*, que é o alvo deste estudo, pode causar: dor, edema, inflamação, hemorragia local, necrose tecidual, além de coagulopatias, hemorragia interna, choque cardiovascular e lesão renal aguda (Herrera *et al.*, 2013). Tais efeitos são causados pela ação de enzimas (serino-proteases, metaloproteinases e fosfolipases A₂). Além de enzimas existem as proteínas não enzimáticas como desintegrinas e proteínas relacionadas à lectinas tipo C (snaclecs: *snake C-type lectins*). Estas proteínas têm como principal alvo o mecanismo hemostático e produzem tantas mudanças neste sistema que podem provocar a falência do sistema circulatório (Pahari *et al.*, 2007; Kohlhoff *et al.*, 2012).

Apesar da versatilidade funcional das proteínas do veneno de serpentes, poucas famílias proteicas são frequentes no veneno. Por outro lado, alta variabilidade é observada nas sequências das proteínas de uma mesma família. A elevada diversidade é provavelmente gerada por altas taxas de substituição, o que pode refletir em novas funções no veneno (Moura-da-Silva *et al.*, 2011).

Proteínas do veneno das serpentes da família Viperidae como metaloproteinases, desintegrinas, snaclecs e serino-proteases podem inibir ou ativar vários fatores da coagulação, incluindo fibrinogênio, protrombina, fator V, fator IX, fator X e trombina (Escalante *et al.*, 2011). Algumas proteínas como snaclecs, desintegrinas e proteínas tipo desintegrinas também podem modular a função plaquetária, atuando diretamente sobre as plaquetas ou indiretamente requerendo proteínas do plasma como cofator e/ou receptores celulares (Ogawa *et al.*, 2005; Du *et al.*, 2006).

Estudos com venenos vêm contribuindo para um melhor entendimento de diversos processos fisiológicos, colaborando para o desenvolvimento de soros e

novas estratégias terapêuticas para o tratamento do envenenamento (Fox & Serrano, 2007). Além disso, os componentes do veneno são modelos promissores de agentes terapêuticos no tratamento e no diagnóstico de vários tipos de doenças que salvam milhares de vidas anualmente (Casewell *et al.*, 2009; Vaiyapuri *et al.*, 2011; Kohlhoff *et al.*, 2012). Como por exemplo, podemos citar um dos mais conhecidos medicamentos para o tratamento da hipertensão arterial, o captopril cujo protótipo foi isolado do veneno de *Bothrops jararaca* (Ferreira *et al.*, 1970). Outros medicamentos disponíveis derivados de veneno de serpentes são: Eptifibatide e Tirofiban, isolados do veneno da serpente norte americana *sistrurus miliarius barbouri* e da serpente africana *Echis pyramidum* para o tratamento da síndrome coronariana aguda; Fibrolase, isolado da serpente norte americana *Agkistrodon contortrix*, e usado como agente trombolítico; Ancrod isolado da serpente do sudeste asiático *Calloselasma rhodostoma* para o tratamento de acidente vascular cerebral (Fox & Serrano, 2007, Koh & Kini, 2012).

Neste estudo concentramos na geração de conhecimento sobre o veneno produzido pela glândula da espécie *Bothrops neuwiedi*, uma espécie que possui ampla distribuição no Brasil. A abordagem escolhida para o estudo foi a análise do transcrito da glândula venenífera. Estudos de transcrito permitem a descrição detalhada do conteúdo de genes expressos mesmo na ausência de dados sobre o genoma da espécie (Durban *et al.*, 2011) e auxiliam na exploração da diversidade molecular.

1.3 As Metaloproteínas de Veneno de Serpentes

As metaloproteínas (do inglês, *Snake Venom Metalloproteinases* - SVMPs) constituem mais de 30% do veneno das serpentes da família Viperidae (Pidd-Queiroz *et al.*, 2013). Entre as serpentes do gênero *Bothrops*, as SVMPs são o principal componente do veneno com mais de 50% de sua constituição (Souza *et al.*, 2013). Estas proteínas são responsáveis por induzir hemorragia na presa e produzir alterações notáveis no sistema hemostático (Escalante *et al.*, 2011; Leonardi *et al.*, 2014).

Metaloproteínas (EC 3.4.24.-) são enzimas que pertencem à família das metizincinas (M12), subfamília M12B (MEROPS - Rawlings *et al.*, 2014). Essas proteínas possuem como principal característica uma região ligante de zinco representada por uma sequência altamente conservada (HEXXHXXGXXH) seguida

por uma região de *loop* contendo uma metionina denominada *Met-turn* que atua como eixo para estabilização das três histidinas da região do sítio catalítico (Sajevic *et al.*, 2011; Casewell, 2011; Moura da Silva *et al.*, 2008).

As metaloproteinases de veneno de serpentes são sintetizadas como pré-proenzimas cujos transcritos contêm a sequência sinal, o pró-domínio, o domínio metaloproteinase e os domínios não catalíticos, os quais variam dependendo da subfamília à qual a proteína pertence. Como proteínas secretadas, após a síntese, elas são direcionadas para o retículo endoplasmático onde são então exportadas através da membrana com consequente perda do peptídeo sinal. No retículo endoplasmático rugoso, os polipeptídeos sofrem a formação de pontes dissulfeto, remodelamento, glicosilação, processamento proteolítico e em alguns casos montagem de estruturas multiméricas. Modificações pós-traducionais via clivagem proteolítica parecem ser importante para as SVMPs como evidenciado pelos numerosos produtos processados de precursores de SVMP encontrados no veneno maduro (Fox & Serrano, 2008).

As SVMPs são subdivididas em PI, PII e PIII (Fox & Serrano, 2008; 2009) (**Figura 3**) sendo que as SVMPs-PI possuem apenas o domínio metaloproteinase (20 - 30 kDa) que contém a sequência característica do sítio catalítico "HEXXHXXGXXH". As SVMPs-PI podem ou não provocar hemorragia (Agero *et al.*, 2007; Sanchez *et al.*, 2010). As SVMPs-II possuem o domínio metaloproteinase e o domínio desintegrina (30 kDa). Este contém o motivo RGD (arginina, glicina e ácido aspártico) ou KGD (lisina, glicina e ácido aspártico) que se liga a receptores de integrinas (GP IIb/IIIa) na superfície de plaquetas inibindo a agregação plaquetária na etapa final da cascata de coagulação. Entretanto, a proteína pode sofrer processamento pós-traducional proteolítico e liberar a desintegrina (Chen *et al.*, 2008). As SVMPs-PIII apresentam o domínio metaloproteinase, o domínio tipo desintegrina e o domínio rico em cisteína. O domínio tipo desintegrina possui outras sequências conservadas em substituição à RGD, mas com funções semelhantes. Como exemplo, podemos citar a sequência XCD com potencial de inibição de colágeno, indutor da agregação plaquetária. Por sua vez, o domínio rico em cisteína desempenha papel importante no reconhecimento do alvo das SVMPs-PIII (Mazzi *et al.*, 2007).

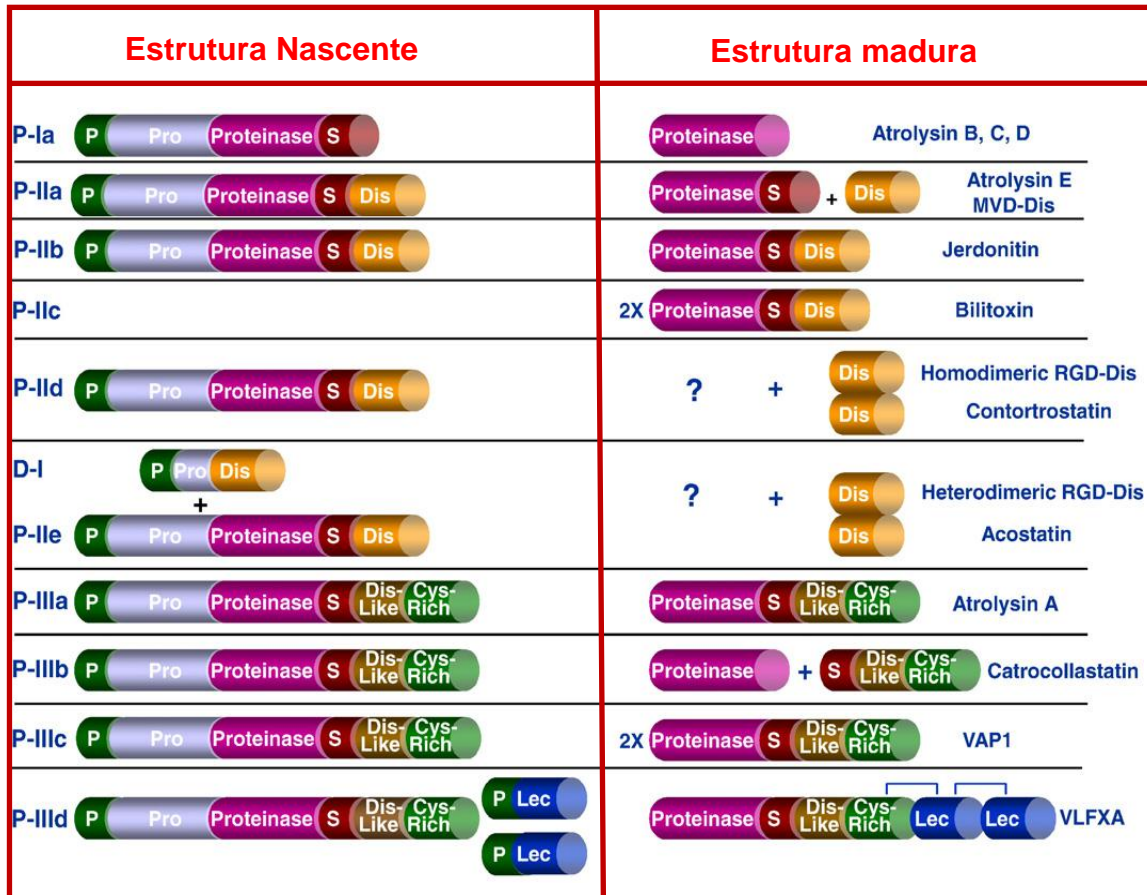


Figura 3 - Classificação de metaloproteínases de veneno de serpentes. Modificado de Fox e Serrano, 2008. **P**: corresponde ao peptídeo sinal; **Pro**: corresponde ao domínio pró-peptídeo presente nos transcritos de metaloproteínase; **Proteinase**: corresponde ao domínio catalítico das metaloproteínase também denominado domínio reprotólina; **S**: corresponde a uma porção denominada *peptide spacer* que é um espaço estrutural entre o domínio catalítico e os demais domínios presentes na estrutura das SVMPs; **Dis**: corresponde ao domínio desintegrina nas SVMPs-P-II; **Dis-Like**: corresponde ao domínio tipo desintegrina nas SVMPs-P-III; **Cys-Rich**: corresponde ao domínio rico em cisteínas; **Lec**: corresponde às subunidades de lectinas tipo C que pode interagir com o domínio rico em cisteínas das SVMPs-P-III.

Além de causar danos teciduais no local da picada, as SVMPs apresentam efeitos sobre todo o sistema hemostático. A indução de hemorragia é devido à sua ação proteolítica sobre proteínas da membrana basal do endotélio vascular tais com colágeno tipo IV e laminina e sobre fatores da coagulação (**Figura 4**) (Fox & Serrano, 2007; Mazzi *et al.*, 2007; Chen *et al.*, 2008).

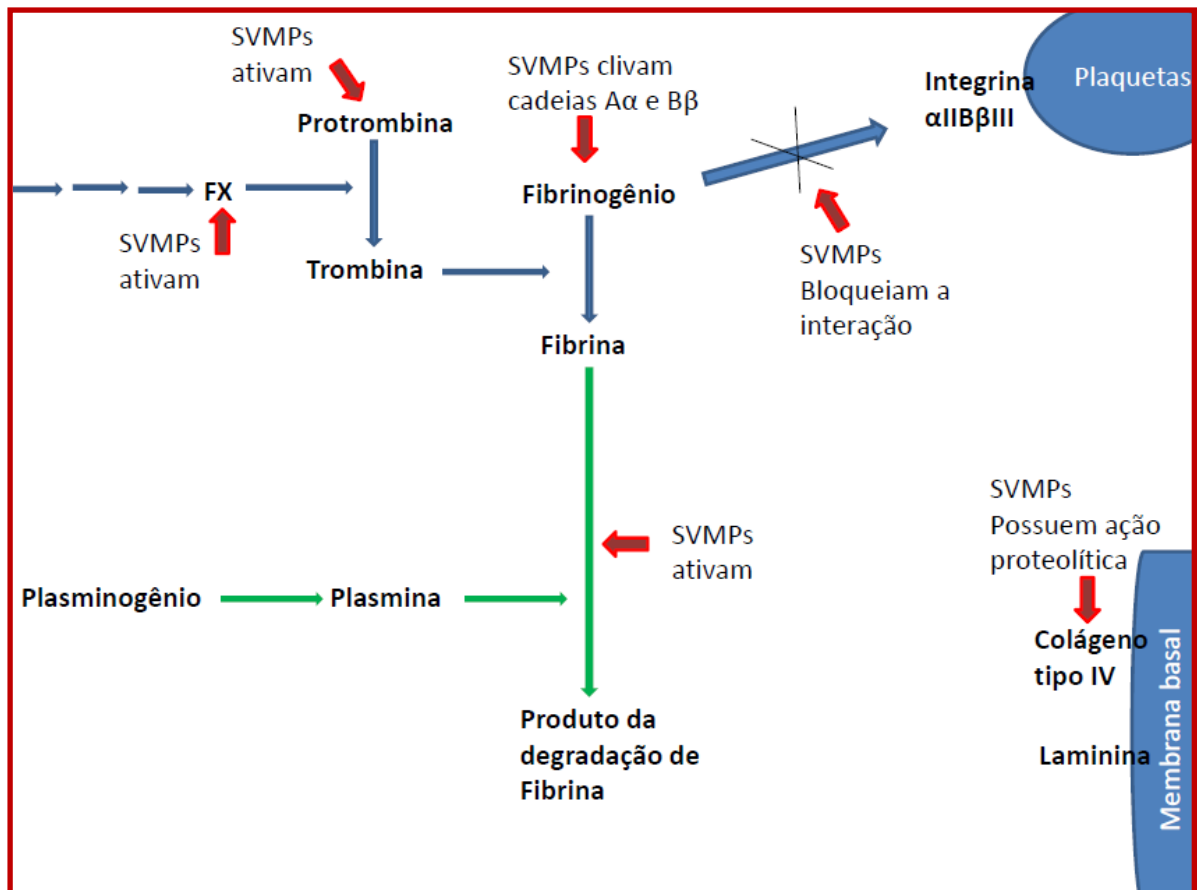


Figura 4 - A ação das SVMPs sobre o sistema hemostático. Esquema resumido da cascata de coagulação e do sistema fibrinogenolítico mostrando a ação das SVMPs sobre proteínas da membrana basal e sobre a interação de receptores integrina e fibrinogênio.

2 JUSTIFICATIVA

Neste estudo sequenciamos o transcrito da glândula de veneno de quatro indivíduos da espécie *Bothrops neuwiedi*, fêmeas, adultas, com o intuito de identificar transcritos que codificam proteínas do veneno.

Os quatro transcritomas da mesma espécie forneceu uma quantidade de dados sobre as proteínas presentes no veneno que auxiliará os estudos da estrutura e função deste veneno. Além disso, permitiu perceber a abundância das proteínas no transcritomas.

A espécie *Bothrops neuwiedi* foi escolhida devido à existência de subespécies ainda não completamente definidas e por ser uma das espécies que tem maior número de entrega de indivíduos na FUNED.

Devido à importância patofisiológica no envenenamento e por seu potencial uso com modelo para o desenvolvimento biotecnológico, as metaloproteinases de veneno de serpentes tornaram-se o alvo principal deste estudo, para o qual realizamos análises filogenéticas com o intuito de contribuir para um melhor entendimento dos mecanismos moleculares e evolutivos subjacentes à biodiversidade dessas proteínas, e colaborar para o descobrimento de novos compostos com potencial ação terapêutica.

3 OBJETIVOS

3.1 Objetivo Geral

Descrever os genes transcritos na glândula venenífera de serpentes da espécie *Bothrops neuwiedi* com o foco nos genes que codificam metaloproteinases de veneno de serpentes bem como a história evolutiva das metaloproteinases de veneno de serpentes da família Viperidae.

3.2 Objetivos Específicos

Determinar o perfil de genes transcritos na glândula de veneno dos quatro indivíduos da espécie *Bothrops neuwiedi* com o uso da abordagem de sequenciamento transcritômico.

Determinar o conjunto das metaloproteinases de veneno de serpentes (SVMPs) da família Viperidae.

Reconstruir a relação filogenética de Metaloproteinases de Veneno de Serpentes (SVMPs) para melhor entendimento das atividades fisiopatológicas desenvolvidas por estas enzimas.

4 MÉTODOS

O fluxo de trabalho realizado para a análise do transcrito desde o preparo das amostras e o uso das ferramentas de biologia computacional, pode ser visualizado na **Figura 5**.

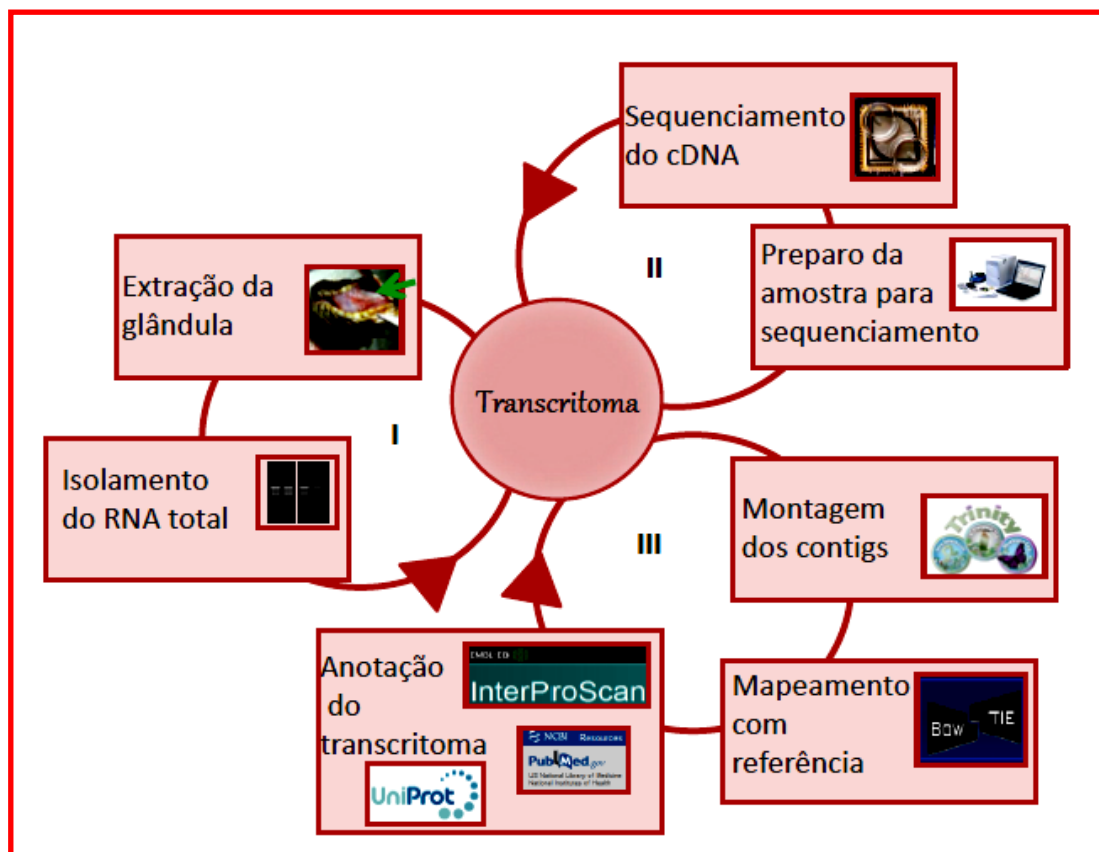


Figura 5 - Fluxo de trabalho para a análise do transcrito da glândula venenífera de *Bothrops neuwiedi*. O procedimento consiste rapidamente na extração das glândulas de veneno, isolamento do RNA total (1), depleção do RNA ribossomal e preparo da amostra para o sequenciamento, seguido pelo sequenciamento na plataforma Ion Torrent (2), montagem dos transcritos, mapeamento e anotação do transcrito (3).

4.1 Extrações das glândulas de veneno de *Bothrops neuwiedi*

Quatro serpentes da espécie *Bothrops neuwiedi* fêmeas, adultas provenientes das cidades de Nova Lima, Brumadinho e Conselheiro Lafaiete foram fornecidas pelo Serpentário da Fundação Ezequiel Dias (FUNED). O veneno das serpentes foi retirado 3 a 4 dias antes de começar o procedimento para a retirada da glândula. As serpentes foram anestesiadas com gelo seco por aproximadamente 30 minutos. As glândulas veneníferas foram retiradas com auxílio de bisturi e colocadas em criotubos previamente identificados e imediatamente foram congelados em nitrogênio líquido e armazenado a -80°C até o uso.

4.2 Construção da biblioteca de cDNA

4.2.1 Isolamento do RNA total

Uma glândula de cada uma das quatro representantes da espécie foi processada separadamente para obtenção de uma biblioteca de cDNA para cada uma. As glândulas de veneno foram pulverizadas manualmente em presença de nitrogênio líquido com utilização de gral e pistilo e o RNA total foi isolado de acordo com o método do reagente de Trizol® (GibcoBRL). O Trizol foi adicionado na proporção de 1,0ml para cada 50,0mg - 120,0mg de tecido, seguido de incubação à 30°C por 5 minutos para uma completa dissociação do complexo nucleoproteínas. Após a incubação foi adicionado clorofórmio na proporção de 0,2ml para cada 1,0ml de Trizol. Os tubos foram então agitados vigorosamente por 15 segundos e novamente incubados por 5 minutos a 30°C. Em seguida as amostras foram centrifugadas a 3500rpm durante 15 minutos à 4°C em microcentrífuga, obtendo-se uma mistura que se separou em uma fase inferior vermelha, uma fase fenol-clorofórmio e uma fase superior aquosa não colorida. O RNA, que permanece na fase aquosa, foi transferido para um novo tubo e precipitado com isopropanol, na proporção de 0,5ml para cada 1,0ml de Trizol usado na homogeneização inicial. Em seguida as amostras foram incubadas durante 10 minutos a 30°C e centrifugadas a 11.600 X g por 10 minutos a 4°C. O RNA precipitado foi lavado com 1,0ml de etanol 75% e novamente centrifugado a 4600 X g durante 15 minutos a 4°C. Após a última centrifugação, o RNA foi colocado em um suporte com a tampa aberta, em

temperatura ambiente por 15 minutos com intuito de permitir a evaporação do etanol residual. Logo depois, a ressuspensão foi feita com água tratada com dietilpirocarbonato (H₂O-DEPC).

A leitura da concentração do RNA total foi feita a 260/280nm no espectrofotômetro GeneQuant (GE Healthcare Life Sciences). A qualidade do RNA total foi verificada em gel agarose 0,8%, preparado em Tris-Borato-EDTA (TBE). Aproximadamente 0,5mg da amostra foi solubilizado em H₂O-DEPC e tampão de amostra (10,0% de sacarose, 90,0% de formamida deionizada, 0,05% de azul de bromofenol e 0,05% de xilenocianol) com adição de 2,0ml de brometo de etídio 0,1mg/ml. Após aquecimento a 65°C por 3 minutos e resfriamento por 3 minutos em gelo, esta amostra foi aplicada no gel de agarose em TBE, e a corrida foi realizada a 100 V/cm em tampão TBE por aproximadamente 50 minutos e fotografado sob iluminação UV. O RNA total da glândula de cada um dos exemplares foi conservado separadamente a -80°C.

4.2.2 Depleção do RNA ribossomal

Os kits usados para depleção do RNA ribossomal, preparo da amostra e sequenciamento próprios para o uso no sequenciador Ion Torrent Personal Genome Machine são da empresa Life Technologies (www.lifetechnologies.com).

A depleção do RNA ribossomal foi feita através do kit “RiboMinus Eukariotic for RNA-seq” que realiza uma depleção seletiva de RNA ribossomal e um enriquecimento de mRNA. Para tal foi utilizado aproximadamente 10,0µg de RNA total, tampão de hibridização e “RiboMinus probe” em um microtubo. A mistura foi aquecida a 75°C por 5 minutos e resfriado por 30 minutos a 37°C. Em seguida acrescentou-se a amostra a um preparado de esferas magnéticas “RiboMinus Magnetic beads” que foi incubado a 37°C por 15 minutos com agitação gentil neste período. Em um separador magnético as esferas formaram um complexo “RNAr-probe”, onde o RNA livre de RNA ribossomal permanece no sobrenadante. O RNA foi concentrado com glicogênio, acetato de sódio 3M e etanol 100%.

4.2.3 Preparo do cDNA para o sequenciamento

O RNA livre de RNA ribossomal foi fragmentado usando RNase III. Em seguida foi purificado com o módulo de limpeza com esferas magnéticas. O RNA

fragmentado foi misturado às esferas magnéticas ligante de ácido nucléico, e após incubação o sobrenadante foi removido, ficando os fragmentos ligados às esferas. As esferas foram então lavadas com solução de lavagem e etanol, e o sobrenadante foi novamente removido. Os fragmentos de RNA foram eluídos das esferas com uso de água livre de nucleases pré-aquecida a 37° C.

O Rendimento e a distribuição de tamanhos dos fragmentos do RNA foram avaliados com a utilização da plataforma Bioanalyzer (Agilent Technologies).

O cDNA foi preparado de acordo com o kit “Ion Total RNA-Seq” para síntese do cDNA e preparo para uso no fluxo de trabalho de sequenciamento do sequenciador Ion Torrent (lifetechnologies). Aos fragmentos de RNA foram adicionados solução de hibridização e um mix de adaptadores. A reação de hibridização foi realizada em um termociclador com temperatura a 65°C por 10 minutos e 30°C por 5 minutos. No gelo, foi adicionado à reação o mix da enzima de ligação. A reação foi incubada no termociclador a 30°C por 30 minutos. Em seguida foi efetuada a reação de transcrição reversa com o mix de enzima “SuperScript III 10X” em termociclador a 42°C por 30 minutos. O cDNA foi purificado usando novamente o módulo de limpeza com esferas magnéticas. O cDNA foi eluído das esferas com uso de água livre de nucleases pré-aquecida a 37° C. O kit “Ion OneTouch 200 system kit” foi utilizado para amplificação do cDNA através da PCR em emulsão no sistema “Ion One Touch”. A biblioteca foi enriquecida com utilização do kit “One Touch ES” e preparada para deposição em chip para o sequenciamento em larga escala.

4.3 Sequenciamento de cDNA

O kit “Ion PGM 200 Sequencing” foi utilizado para o sequenciamento. Após enriquecimento do cDNA, a próxima etapa foi o anelamento do iniciador de sequenciamento. Enquanto a amostra foi incubada para o anelamento, o chip 318, onde ocorreu o sequenciamento, foi testado para garantir o seu correto funcionamento antes da amostra ser acrescentada.

Após o anelamento foi adicionado à amostra a polimerase “Ion PGM Sequencing 200 Polymerase”, e a mistura foi incubada a temperatura ambiente por 5 minutos. A amostra foi então inserida no Chip 318 que contém 12 milhões de poços para realização de sequenciamento em larga escala gerando sequências de

tamanho aproximado de 200 pb. Ao final quatro bibliotecas de cDNA foram sequenciadas, cada uma de um representante da espécie estudada.

4.4 Análise do transcrito de *Bothrops neuwiedi*

4.4.1 Análise de qualidade das sequências do transcrito

A magnitude de dados obtida do sequenciamento em larga escala foi tratada com ferramentas de bioinformática para montagem, mapeamento e anotação das sequências. Apenas sequências com qualidade *phred* (Ewing *et al.*, 1998) superior a 20 foram consideradas.

As análises de qualidade foram realizadas por meio do *software* FASTQC v0.10.1 (www.bioinformatics.babraham.ac.uk/projects/fastqc/). O programa PRINSEQ-LITE (prinseq.sourceforge.net/) foi utilizado para fazer a *trimagem* das sequências cuja média de escore *phred* foi inferior a 20 (Ewing *et al.*, 1998). Dois arquivos foram gerados para cada biblioteca, um com sequências com escore abaixo de *phred* 20 e outro com sequências com valor igual ou acima de *phred* 20.

4.4.2 Montagem do transcrito de *Bothrops neuwiedi*

Os quatro arquivos com as sequências filtradas foram concatenados em um único arquivo para montagem dos transcritos. O programa utilizado para montagem foi o Trinity (Grabherr *et al.*, 2011), que é um programa para montagem *de novo* de transcritos a partir de sequências de RNAseq. Um *script* escrito em *Perl* (escrito por Adhemar Zerlotine Neto/CEBio-FIOCRUZ/Minas) foi utilizado para fazer as estimativas da eficiência das montagens realizadas. As métricas avaliadas foram: número total de transcritos montados, tamanho dos transcritos, número de transcritos maiores do que 1kb, número de bases em todos os transcritos, tamanho médio dos transcritos, porcentagem de transcritos maiores que 1kb e N50 (comprimento do *contig* pelo qual a soma dos maiores *contigs* do conjunto ultrapassa 50% do comprimento da montagem). Para retirar a redundância dos transcritos o programa CAP3 (Huang & Madan, 1999) foi utilizado.

4.4.3 Mapeamento e quantificação das sequências de RNAseq

A ferramenta utilizada para o mapeamento foi o Bowtie v2.1.0 (Langmead & Salzberg, 2012), que alinhou as sequências ao longo do transcrito de referência criado pelo Trinity a partir do mesmo conjunto de dados. A partir do mapeamento de cada biblioteca foi criada uma tabela para a quantificação de cada transcrito nos quatro transcriptomas. Esta tabela foi normalizada pelo valor de RPKM (*reads per kilobase per million*). Assim, por exemplo, para um transcrito hipotético, com um tamanho de 900 bases com 40 sequências mapeadas de um total de 1.200.000 possíveis, o RPKM desse transcrito seguirá a fórmula: $(40 \div 0,9) \div 1,2 = 37$ RPKM.

4.4.4 Busca por similaridades em banco de dados e anotação do transcrito.

Os transcritos montados pelo Trinity foram utilizados em pesquisas por similaridade nos bancos de dados públicos GenBank (Benson *et al.*, 2013) e UniProt (The UniProt Consortium, 2013). Um banco de dados local em MySQL (www.mysql.com) foi criado para armazenar os identificadores dos transcritos e a anotação encontrada nas pesquisas que foram feitas por meio da ferramenta BLAST (Altschul *et al.*, 1990) As sequências dos transcritos de interesse foram comparadas no banco de dados Pfam (Finn *et al.*, 2014) para melhor anotação dos seus domínios proteicos funcionais.

4.5 Reconstrução filogenética de SVMPs

O fluxo de trabalho realizado, incluindo bancos de dados e ferramentas utilizados desde a seleção de potenciais homólogos à reconstrução filogenética e anotação das árvores evolutivas de SVMPs, pode ser visualizado na **Figura 6**.

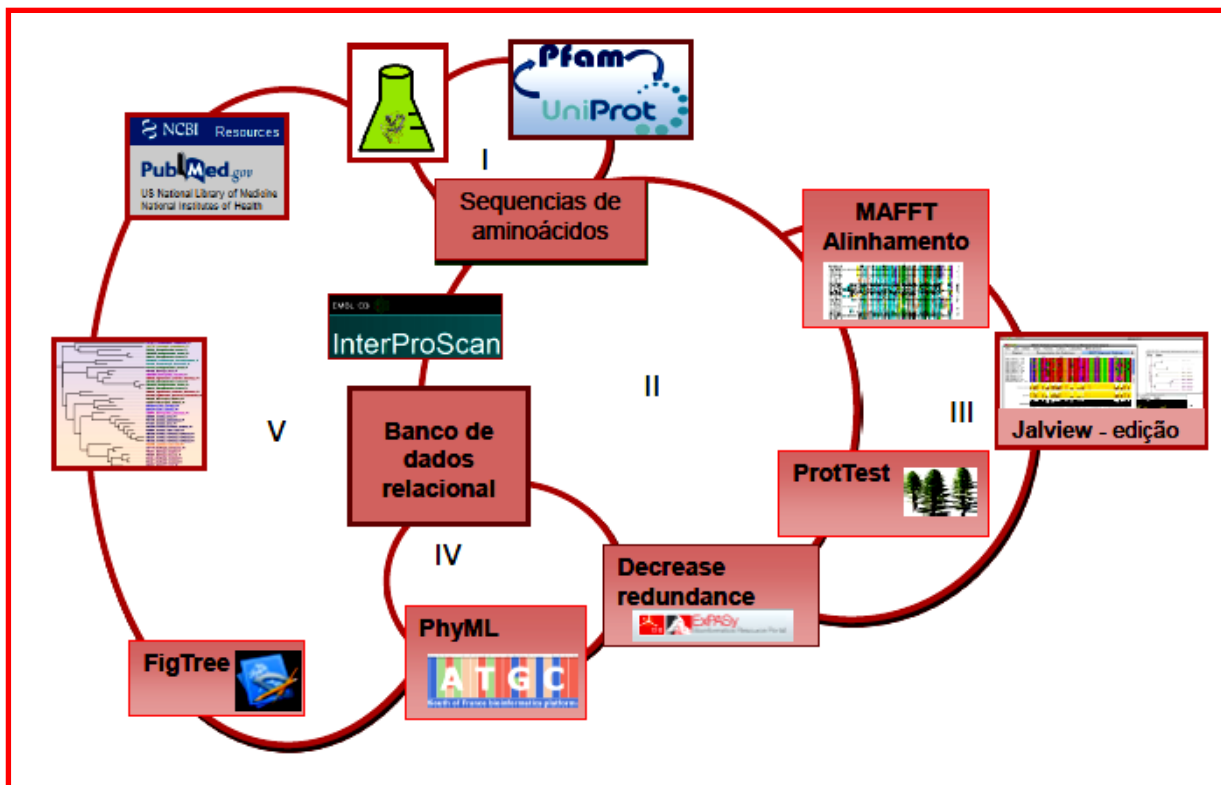


Figura 6 - Fluxo de trabalho para a reconstrução da história evolutiva de SVMPs. Os principais passos são: seleção dos potenciais homólogos de SVMPs de serpentes da família Viperidae, alinhamento e edição manual das sequências de aminoácidos, teste do modelo evolutivo e reconstrução filogenética.

4.5.1 Seleção de potenciais homólogos

As filogenias de SVMPs reconstruídas neste estudo compreendem as relações evolutivas de sequências de aminoácidos de metaloproteinasas codificadas pelo genoma nuclear de serpentes da família Viperidae. Tais enzimas pertencem à subfamília M12B conforme descrito no banco de dados MEROPS (Rawlings *et al.*, 2014). As sequências de aminoácidos de SVMPs foram identificadas através de buscas combinadas no UniProt, um banco de dados de proteínas (The UniProt Consortium, 2013). Estas buscas foram feitas usando o identificador taxonômico da família Viperidae (8689) no NCBI Taxonomy (www.ncbi.nlm.nih.gov/taxonomy) e o identificador do domínio catalítico de metaloproteinasas (PF01421) no Pfam (Finn *et al.*, 2014). Para recuperar os potenciais homólogos de cada subfamília (PI, PII e PIII) separadamente, foram realizadas buscas combinadas utilizando os identificadores

dos demais domínios proteicos encontrados nessas proteínas como descrito no **Quadro 1**.

Quadro 1 - Estratégia de busca de potenciais homólogos de SVMPs de Viperidae no UniProt.

1A			
	Domínio 1	Domínio 2	Domínio 3
SVMP P-I	PF01421	---	---
SVMP P-II	PF01421	PF00200	---
SVMP P-III	PF01421	PF00200	PF08516
1B			
	Buscas no UniProt		
SVMP P-I	(PF01421 NOT (PF00200 AND PF08516)) AND taxonomy: 8689		
SVMP P-II	(PF01421 AND PF00200 NOT PF08516) AND taxonomy: 8689		
SVMP P-III	(PF01421 AND PF00200 AND PF08516) AND taxonomy: 8689		

Quadro 1 - Busca por potenciais homólogos considerando as distintas arquiteturas proteicas dos membros das subfamílias de SVMPs (P-I, P-II e P-III). **1A**: Os números de acesso correspondem a domínios proteicos descritos no Pfam. Além do domínio catalítico (PF01421), outros domínios podem estar presentes nas SVMPs. **1B**: Comandos usados na busca avançada para recuperação de cada subfamília separadamente.

4.5.2 Alinhamento de sequências proteicas

Sequências de aminoácidos dos potenciais homólogos das subfamílias de SVMPs de Viperidae, recuperadas do UniProt (The UniProt Consortium, 2013), foram processadas para a construção de alinhamentos. A ferramenta InterProScan (Hunter *et al.*, 2009) foi utilizada para obtenção das coordenadas (início e fim) do domínio catalítico (PF01421) no conjunto de sequências selecionadas. Posteriormente, um *script* em *Perl*, foi utilizado para extrair a sequência correspondente ao domínio catalítico a partir da sequência proteica completa. Por fim, as sequências do domínio catalítico (PF01421) foram separadas em duas bases de dados, a saber: i) sequências das três subfamílias de SVMPs (PI, PII e PIII) e ii)

sequências da subfamília SVMP-PI. Os arquivos de sequências em formato FASTA foram então filtrados utilizando a opção *trim* do programa T-Coffee (versão 10.00.r1613) para remoção de sequências muito semelhantes (% máxima de identidade = 96) e muito divergentes (% acurácia média = 30) (Notredame *et al.*, 2000). Os arquivos foram ainda sujeitos à inspeção visual em que sequências muito curtas ou ricas em “XXX” foram removidas. Ambos arquivos filtrados contendo sequências de aminoácidos do domínio catalítico (PF01421) foram alinhados, separadamente, utilizando o programa MAFFT (versão 7) com refinamento iterativo pela estratégia G-INS-i (Kato & Toh, 2008). Os demais parâmetros foram mantidos no formato padrão. Os alinhamentos múltiplos foram visualizados e manualmente editados utilizando-se o programa Jalview (Waterhouse *et al.*, 2009).

4.5.3 Teste de modelos evolutivos

A partir dos alinhamentos de sequências, utilizou-se o programa ProtTest (Abascal *et al.*, 2005) para selecionar, entre as matrizes de substituição candidatas, qual o modelo evolutivo que melhor se adequava aos dados para a subsequente reconstrução das filogenias. A utilização de modelos evolutivos na reconstrução de árvores filogenéticas permite incorporar suposições sobre o processo evolutivo que originou os dados observados produzindo filogenias mais realistas (Liò & Goldman, 1998; Posada & Buckley, 2004; Kelchner & Thomas, 2007). Os modelos avaliados foram WAG, LG, mtREV, Dayhoff, DCMut, JTT, VT, Blosum62, CpREV, RtREV, MtMam, MtArt, HIVb e HIVw. Para avaliar qual modelo evolutivo melhor se ajustou aos dados, utilizamos o critério de informação de Akaike (AIC) (Akaike, 1973). Este critério é derivado dos valores de máxima verossimilhança estimados, penalizando o número de parâmetros analisados. Assim calcula-se a qualidade relativa de um modelo de substituição para um conjunto de dados estabelecido. O modelo ideal é o que apresenta menor valor de AIC.

4.5.4 Reconstrução filogenética

Com o intuito de estabelecer as relações evolutivas entre membros de diferentes subfamílias de SVMPs de Viperidae, realizamos análises usando dois distintos alinhamentos de sequências do domínio catalítico (PF01421) dessas proteínas contendo: i) sequências das três subfamílias de SVMPs (PI, PII e PIII) e ii)

sequências da subfamília SVMP-PI. Separadamente, ambos arquivos foram submetidos à reconstrução filogenética utilizando o método probabilístico de máxima verossimilhança (do inglês, *Maximum Likelihood* - ML) como implementado no PhyML (Felsenstein, 1981; Guindon & Gascuel, 2003). Valores de apoio para as diferentes partições das árvores filogenéticas foram calculados pelo teste aLRT (do inglês, *approximate likelihood ratio test*) (Anisimova & Gascuel, 2006). Esta abordagem avalia cada ramo da árvore calculando se o mesmo possui ganho de verossimilhança significativo em comparação com a hipótese nula que determina o colapso deste ramo na árvore filogenética (Anisimova & Gascuel, 2006). As árvores filogenéticas obtidas foram visualizadas e editadas em formato gráfico utilizando o programa FigTree (versão 1.4.0) (Rambaut, 2012). Em seguida, realizou-se a anotação da árvore filogenética adicionando informações estruturais e bioquímicas das SVMPs, tais como a presença ou ausência de atividade hemorrágica.

5 RESULTADOS

5.1 Análises do transcrito da glândula venenifera de *Bothrops neuwiedi*

5.1.1 Análise da qualidade dos experimentos

Quatro glândulas de serpentes da espécie *Bothrops neuwiedi* foram usadas para a extração do RNA total empregado na construção das bibliotecas. A qualidade do RNA foi avaliada pela plataforma Bioanalyzer e os valores de RIN (número de integridade do RNA) demonstram que a integridade das quatro amostras de RNA eram adequadas para a construção de bibliotecas de cDNA e sequenciamento em larga escala (**Figura 7**).

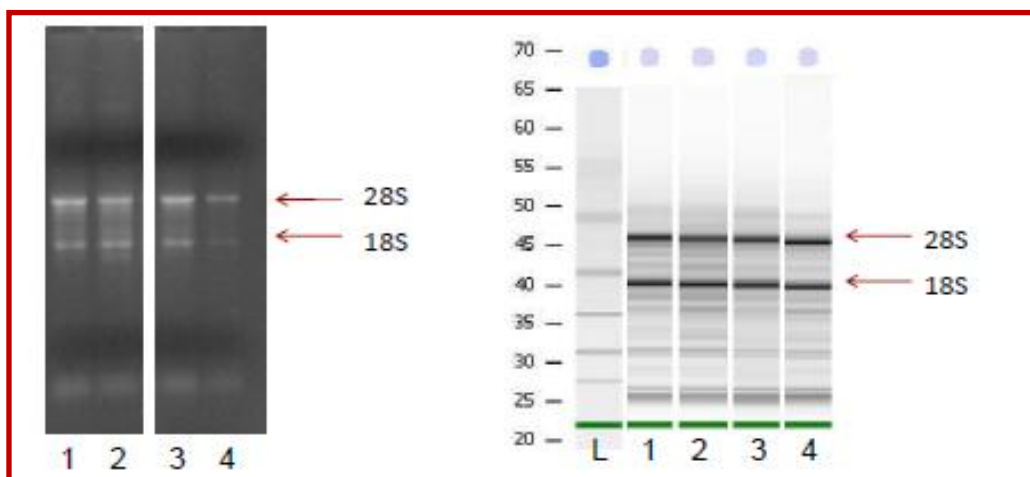


Figura 7 - RNA total de glândula de veneno de *B. neuwiedi*. 1 – Brumadinho (RIN= 6,6); 2 e 3 – Nova Lima (RIN= 5,8 e 7,2, respectivamente); 4 – Conselheiro Lafaiete (RIN= 7,1); L – Marcador de massa molecular.

A qualidade dos dados do sequenciamento foi checada pela ferramenta FastQC (www.bioinformatics.babraham.ac.uk/projects/) (**Figura 8**).

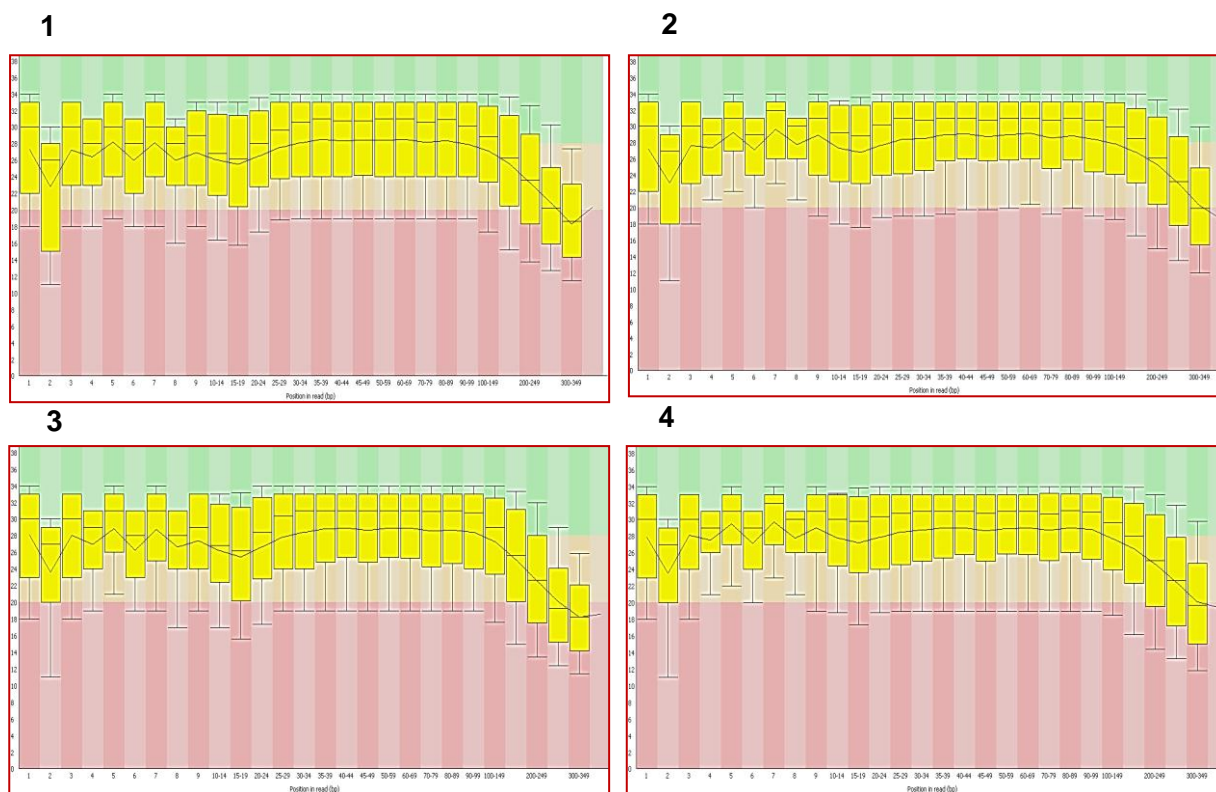


Figura 8 - Resultado da análise dos dados de sequenciamento realizada pelo programa FastQC. A maioria das bases apresenta um valor de *phred* acima de 20. Na linha vertical estão os valores de *scores phred* e na horizontal, a extensão em nucleotídeos das sequências.

Na **Figura 8**, os gráficos representam: na porção rosa os valores de qualidade abaixo de *phred* 20, na porção laranja, os valores de qualidade entre 20 e 28 e na porção verde os valores de *Phred* acima de 28 em relação a posição das bases nas *reads*. A maioria das bases apresenta um valor de *phred* acima de 20 e muitas ainda estão acima do valor de *phred* 30 o que significa uma boa qualidade dos dados, pois o valor de *phred* 20 indica que a probabilidade da base não estar correta é de 1 em 100, e *phred* 30 indica que esta probabilidade é de 1 em 1000.

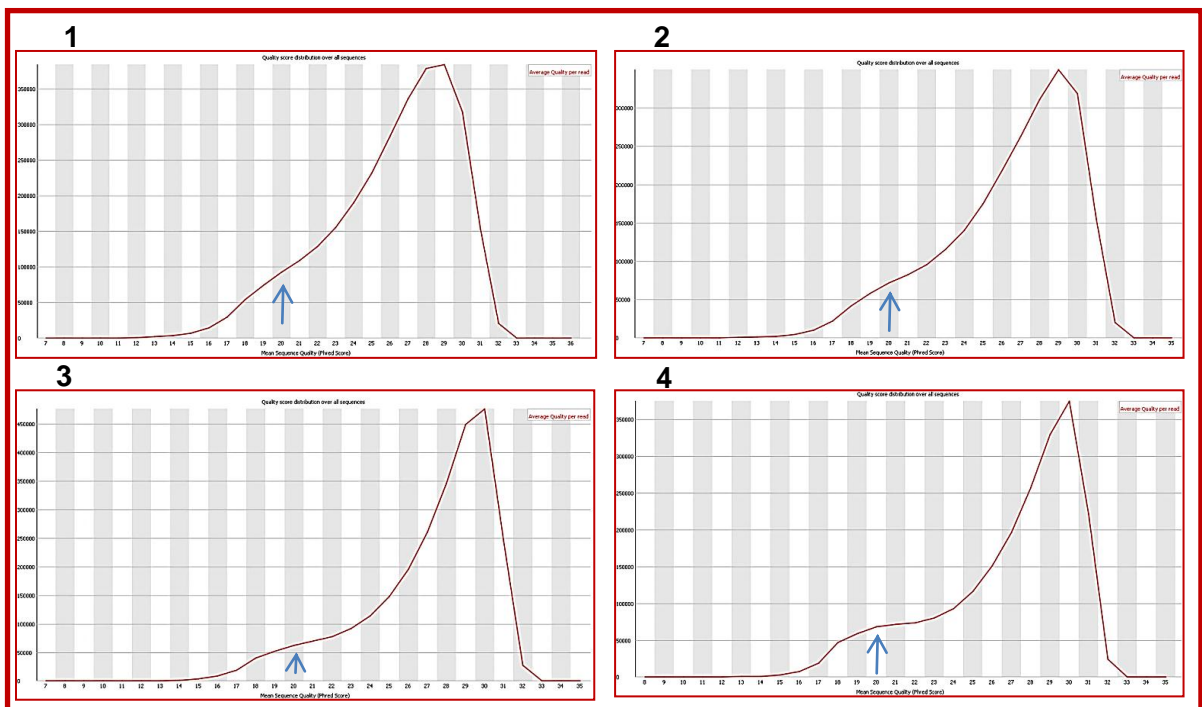


Figura 9 - Média da qualidade das sequências originais. As setas indicam os valores de *phred* 20, a *trimagem* foi feita por este valor. Na linha vertical está o número de sequências da biblioteca e na horizontal, a medida da qualidade das sequencias dada pelo *score phred*. 1 – procedente de Brumadinho, 2 e 3 – procedentes de Nova Lima, 4 – procedente de Conselheiro Lafaiete.

Ao todo foram geradas 10,4 milhões de *reads*. Aproximadamente 350.000 *reads* de cada biblioteca tiveram qualidade média de *phred* igual a 30 (**Figura 9**). Para filtrar as sequências com baixa qualidade, uma *trimagem* foi feita pelo valor de *phred* abaixo de 20. Com os arquivos filtrados pelo valor de *Phred* 20, as quatro bibliotecas foram agrupadas para a montagem do transcrito de referência pelo programa Trinity (**Tabela 1**).

Tabela 1 - Cálculo estatístico da montagem dos *contigs* realizada pelo programa Trinity. ¹Biblioteca de cDNA de glândulas de *Bothrops neuwiedi*, 1 – Brumadinho, 2 e 3 – Nova Lima, 4 – Conselheiro Lafaiete,. ²Número de *reads* geradas em milhões. ³Número de transcritos formados com a montagem pelo programa Trinity. ⁴O N50 é o valor do comprimento do *contig* a partir do qual, somando o comprimento dele e de todos os contigs maiores do que ele, o conjunto ultrapassa 50% do comprimento da montagem. ⁵Indica o tamanho médio dos transcritos medido pelo número de bases em cada um. ⁶Expressa o número de transcritos com tamanho maior do que 1 kilobase. ⁷Número total de bases nos transcritos maiores do que 1 kilobase.

Biblioteca ¹	Número de <i>reads</i> ²	Número de <i>reads</i> após trimagem	Número de transcritos ³	N50 ⁴	Tamanho médio dos transcritos ⁵	Número de transcritos > 1kb ⁶	Total de bases nos transcritos >1kb ⁷
1	3,0	2,9	1000	356	274	26	36.153
2	2,5	1,9	6.545	316	269	72	102.287
3	2,7	2,6	2.054	322	264	37	57.775
4	2,2	2,1	4.340	311	271	44	64.223
Total	10,4	9,5	13.959	-	-	-	260.438
Referência	10,4	9,5	14.339	331	275	275	383.036

5.1.2 Transcritos identificados no veneno de *Bothrops neuwiedi*

A soma de todos os transcritos gerados pelo Trinity a partir das quatro bibliotecas foi de 14.339, dos quais foram encontrados muitos transcritos alternativos. Após a montagem dos transcritos com o programa Trinity, o programa CAP3 foi utilizado para retirar a redundância da montagem. (**Tabela 2**).

Os transcritos que codificam proteínas do veneno de serpentes foram identificados nas bibliotecas de cDNA da glândula de veneno de *Bothrops neuwiedi* através de alinhamento local e busca por similaridade a partir do algoritmo BLAST (Altschul, 1990). Baseado no mapeamento das bibliotecas no transcrito de referência montado, foi possível estimar a abundância da expressão de proteínas na glândula de veneno tendo em vista a média e mediana dos transcritos expressos em cada biblioteca.

Tabela 2 - Quantidade de transcritos montados em cada biblioteca. A Tabela apresenta ainda a soma de todos os transcritos e a quantidade de transcritos montados no transcrito referêcia.

Biblioteca	Trinity	CAP3
1	1000	951
2	6545	6410
3	2054	2006
4	4340	4176
Soma	13939	13543
Referência	14339	13968

Os principais transcritos identificados neste transcrito estão apresentados na **Tabela 3**: trinta e cinco transcritos codificam fosfolipases A2 (PLA2) e o nível de expressão é bastante alto para alguns deles, se comparado à média de expressão de transcritos na biblioteca; trinta e dois transcritos codificam proteínas do veneno de serpente semelhante à lectina tipo C (Snaclecs); dezoito transcritos identificados no transcrito desta espécie codificam inibidores de fosfolipases A2; vinte e cinco transcritos codificam serino-proteinases de veneno de serpentes; vinte e cinco transcritos que codificam proteínas com motivo três dedos (*three finger*) foram identificadas na montagem do transcrito; três transcritos que codificam L-aminoácido oxidases (LAO) foram identificados no transcrito desta espécie, com um nível de produção bastante expressivo. Transcritos que codificam peptídeos natriuréticos potencializadores de bradicinina também foram identificados nesta montagem; (**Tabela 3, Anexo 1**).

Tabela 3 - Principais transcritos do veneno de *Bothrops neuwiedi* identificados.

Principais transcritos do veneno encontrados	Número de transcritos diferentes encontrados	RPKM			
		Biblioteca 1	Biblioteca 2	Biblioteca 3	Biblioteca 4
Fosfolipases A2	35	886,01	5.102,63	2.236,94	2.823,43
Snaclecs	32	816,11	6.714,73	7.434,37	4.638,96
Serino-proteinase	25	337,60	1.218,88	2.123,28	2.166,08
L-amino-ácido oxidase	3	132,87	928,74	1.207,65	1.043,57
Proteínas <i>Three finger</i>	25	395,22	852,99	308,87	616,10
Inibidores de fosfolipases A2	18	121,01	476,91	219,15	432,87

As metaloproteinase são proteínas multidomínios e de acordo com a literatura, há uma variação no nível de sequência destas enzimas entre indivíduos da mesma espécie dependendo da região de origem e da alimentação. Por este motivo não foi possível identificar SVMPs-PIII (que possuem os três domínios) através da montagem *de novo* de transcritos no transcrito de referência (4 bibliotecas), uma vez que a montagem depende da sobreposição de sequências, portanto a **Tabela 4** apresenta o mapeamento de cada um dos domínios presente nas SVMPs do transcrito desta espécie.

Tabela 4 - Principais transcritos dos domínios encontrados em SVMPs no veneno de *Bothrops neuwiedi*.

Principais transcritos dos domínios encontrados em SVMPs	Número de transcritos diferentes encontrados	Biblioteca	Biblioteca	Biblioteca	Biblioteca
		1	2	3	4
		RPKM			
Domínio catalítico	19	1.333,48	4.348,28	9.165,48	3.544,24
Domínio catalítico + Desintegrina	4	591,77	2.150,19	3.906,02	1.732,36
Domínio desintegrina	3	173,26	675,46	1.310,14	608,24
Domínio desintegrina + Domínio rico em cisteína	3	835,33	4.858,88	3.785,99	1.903,71
Domínio rico em cisteína	3	204,26	563,07	833,07	411,54

A maioria dos transcritos que codificam os domínios de SVMPs possuem uma abundância de expressão maior que a média e a mediana. Entre eles, 19 transcritos codificam o domínio catalítico de metaloproteínase (PF01421). Três transcritos codificam o domínio desintegrina (PF00200) presente em SVMP-PII e SVMP-PIII. Três transcritos codificam o domínio rico em cisteína (PF08516) que está presente em SVMP-PIII. Quatro transcritos com um nível de expressão bastante significativo codificam dois domínios proteicos presentes em SVMPs (PF01421 e PF00200). Três transcritos, um dos quais demonstra uma produção expressiva, codificam o domínio desintegrina (PF00200) juntamente com o domínio rico em cisteína (PF08516) de SVMPs.

Para identificar melhor os transcritos que codificam metaloproteinases de veneno de serpente, cada biblioteca foi analisada separadamente. Para a biblioteca1 (proveniente da cidade de Brumadinho) foram identificados 9 transcritos que codificam apenas o domínio PF01421 portanto indica que são transcritos que codificam SVMP-PI, 4 transcritos codificam o domínio catalítico (PF01421) e o domínio desintegrina (PF00200) indicando serem transcritos que codificam SVMPs-PII e 4 transcritos codificam os 3 domínios presentes em SVMPs mostrando que são transcritos que codificam SVMPs-PIII (**Tabela 5**).

Tabela 5 - Subfamílias de SVMPs encontradas em cada Biblioteca separadamente.

Bibliotecas/Procedência	SVMPs		
	PI	PII	PIII
1 – Brumadinho	9	4	4
2 – Nova Lima	3	2	4
3 – Nova Lima	5	3	1
4 – Conselheiro Lafaiete	7	4	0

Na biblioteca 2 (proveniente da cidade de Nova Lima) o número de transcritos identificados foi menor do que na Biblioteca 1, sendo 2 transcritos que codificam o domínio metaloproteinase (PF01421) indicando serem SVMP-PI, 2 transcritos que codificam os dois domínios presentes em SVMP-II e 4 transcritos codificam SVMP-PIII, além de alguns transcritos também apresentarem o domínio pro-peptídeo de metaloproteinase (PF01562). Na biblioteca 3 (Proveniente da cidade de Nova Lima) foram identificados 5 transcritos que codificam o domínio metaloproteinase (PF01421), 3 transcritos que codificam os domínios metaloproteinase e desintegrina e 1 transcrito que codifica os três domínios presentes em SVMP-PIII. E na biblioteca 4 (proveniente da cidade de Conselheiro Lafaiete) foram detectados 7 transcritos que codificam o domínio metaloproteinase (PF01421) e 4 transcritos que codificam os domínios que estão presente em SVMP-II porém não foi detectado transcritos que codificassem os três domínios presente em SVMP-PIII embora dois transcritos que codificam o domínio rico em cisteína (PF08516) tenham sido identificados (**Tabela 5, Anexo 2**).

As tabelas com dados específicos dos principais transcritos identificados neste transcrito estão descritas no **Anexo 1**. E as tabelas com os dados da montagem dos transcritos de SVMP de cada biblioteca estão descritas no **Anexo 2**.

A **Figura 10** representa a média de expressão dos principais transcritos identificados no veneno. Os transcritos que codificam metaloproteinase são os mais expressos no transcrito da glândula de veneno de *Bothrops neuwiedi*. Transcritos que codificam fator de crescimento nervoso (NGF), hialuronidase e inibidor de SVMPs também foram identificados neste transcrito, porém com um nível de

expressão menor comparado com os outros transcritos que codificam proteínas do veneno.

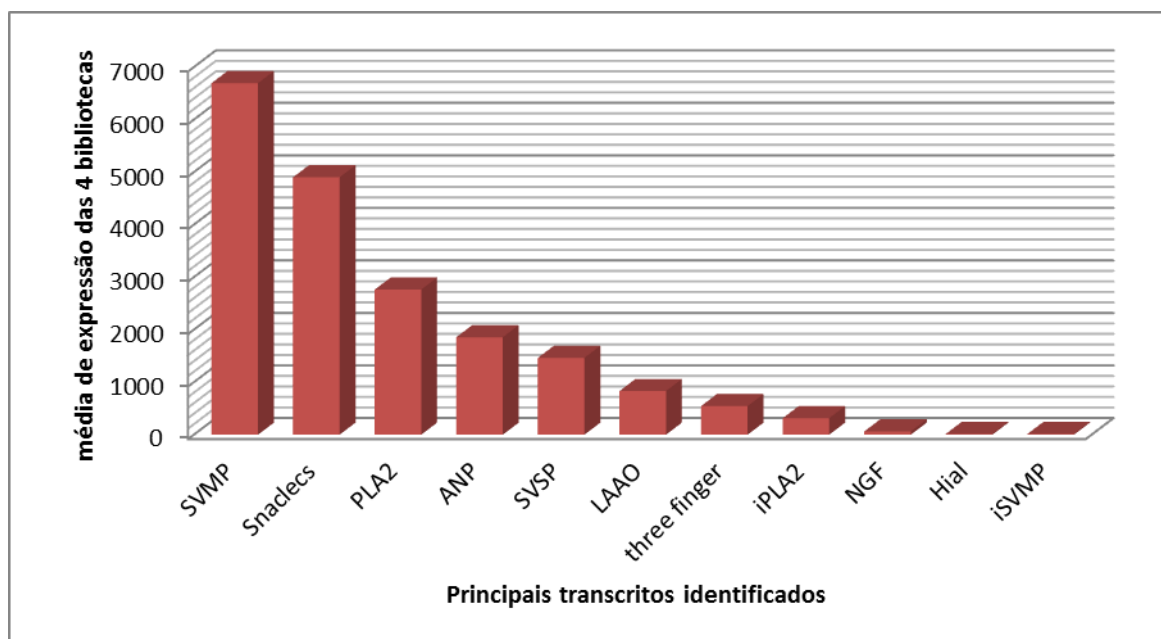


Figura 10 - Gráfico da média de expressão dos principais transcritos que codificam proteínas do veneno. SVMP, metaloproteinase de veneno de serpente; Snaclecs, proteínas do veneno de serpentes semelhantes à Lectina tipo C; PLA2 fosfolipase A2; ANP, peptídeo natriurético atrial; SVSP, serinoproteinase de veneno de serpente; LAAO, L-aminoácido oxidase; three finger, proteínas três dedos; iPLA2, inibidor de fosfolipases A2; NGF fator de crescimento nervoso; Hial, Hialuronidase; iSVMP, inibidor de metaloproteinase de veneno de serpente.

5.2 Reconstrução da história evolutiva de SVMPs

5.2.1 Identificação de SVMPs de Viperidae

Um total de 435 sequências de SVMPs de 17 diferentes gêneros de Viperidae foram recuperadas do banco de dados UniProt (The UniProt Consortium, 2013) usando a estratégia de busca indicada no **Quadro 1**. Diferentes conjuntos de dados foram obtidos para as subfamílias SVMPs-PI (63 proteínas), SVMPs-PII (164 proteínas) e SVMPs-PIII (208 proteínas), conforme ilustrado na **Figura 11**. Os gêneros *Crotalus* e *Echis* têm o maior número de SVMPs depositadas no UniProt

atualmente com 73 e 233 sequências, respectivamente. Um total de 28 sequências de SVMPs de *Bothrops*, foco deste trabalho, foram identificadas no UniProt.

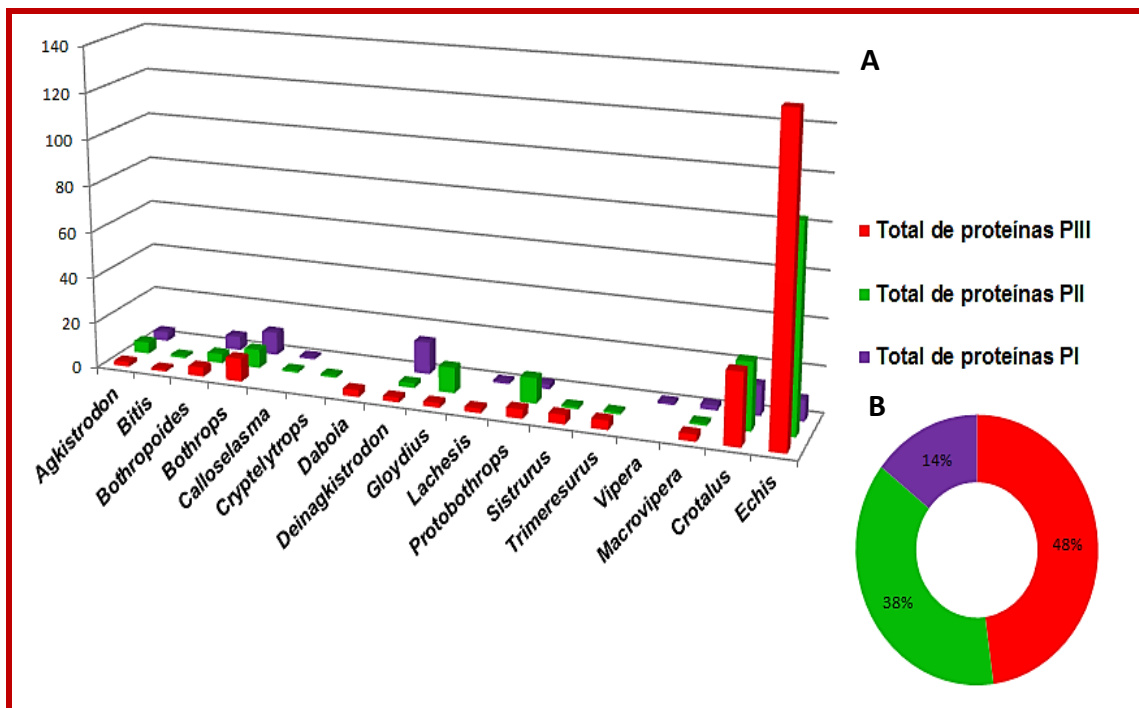


Figura 11 - Distribuição de SVMPs de Viperidae no UniProt. **A**, Distribuição de sequências de aminoácidos de SVMPs de diferentes gêneros de Viperidae depositadas no UniProt (acessado em 14/06/2013). **B**, As proteínas identificadas em cada subfamília (PI, PII e PIII) estão indicadas em cor.

5.2.2 Análise das relações evolutivas de SVMPs

Para inferir as relações evolutivas entre as metaloproteinases do veneno de serpentes da família Viperidae, uma árvore filogenética compreendendo os potenciais homólogos das três subfamílias de SVMPs (PI, PII e PIII) foi reconstruída utilizando o método de máxima verossimilhança (Felsenstein, 1981; Guindon & Gascuel, 2003). Essa filogenia foi baseada no arquivo de sequências de aminoácidos filtrado pelo T-Coffee (Notredame *et al.*, 2000) e alinhado pelo MAFFT (Kato & Toh, 2008). O alinhamento final possui um total de 266 sequências do domínio catalítico (PF01421) de SVMP-I (45 sequências), SVMP-II (69 sequências) e SVMP-III (162 sequências). O teste de modelos evolutivos realizado pelo ProtTest (Abascal *et al.*, 2005) apontou JTT (Jones *et al.*, 1992) como o modelo que melhor se adequou aos dados. Os resultados mostram que os membros das subfamílias

SVMPs-PI, SVMPs-PII e SVMPs-PIII estão distribuídos em diferentes clados, muitas vezes apoiados por alto valor de aLRT (Anisimova & Gascuel, 2006), aLRT é o teste da razão de verossimilhança aproximada do ramo que avalia cada ramo da árvore calculando se o mesmo possui ganho de verossimilhança significativo em comparação com a hipótese nula que determina o colapso deste ramo na árvore filogenética. A filogenia resultante apoia a hipótese de que não há uma origem monofilética para cada subfamília de SVMP de Viperidae (**Figura 12**).

Além da filogenia reconstruída com proteínas das três subfamílias de SVMPs, uma segunda árvore foi reconstruída contendo somente membros da subfamília SVMP-PI (**Figura 13**). Esta análise foi realizada com o arquivo de sequências de aminoácidos filtrado pelo T-Coffee (Notredame et al., 2000) e alinhado pelo MAFFT (Kato & Toh, 2008). O alinhamento final possui um total de 36 sequências do domínio catalítico (PF01421) de SVMP-I de 22 espécies de Viperidae (**Anexo 3**). Dos 202 sítios do alinhamento, 179 são polimórficos. O modelo evolutivo que melhor se adequou aos dados, segundo o ProtTest (Abascal et al., 2005), foi o WAG (Whelan & Goldman, 2001). A anotação desta árvore foi feita baseada nas informações de estrutura e função proteicas disponíveis no banco de dados UniProt (The UniProt Consortium, 2013). As proteínas para as quais existem informações estruturais e outros dados experimentais estão indicadas na árvore.

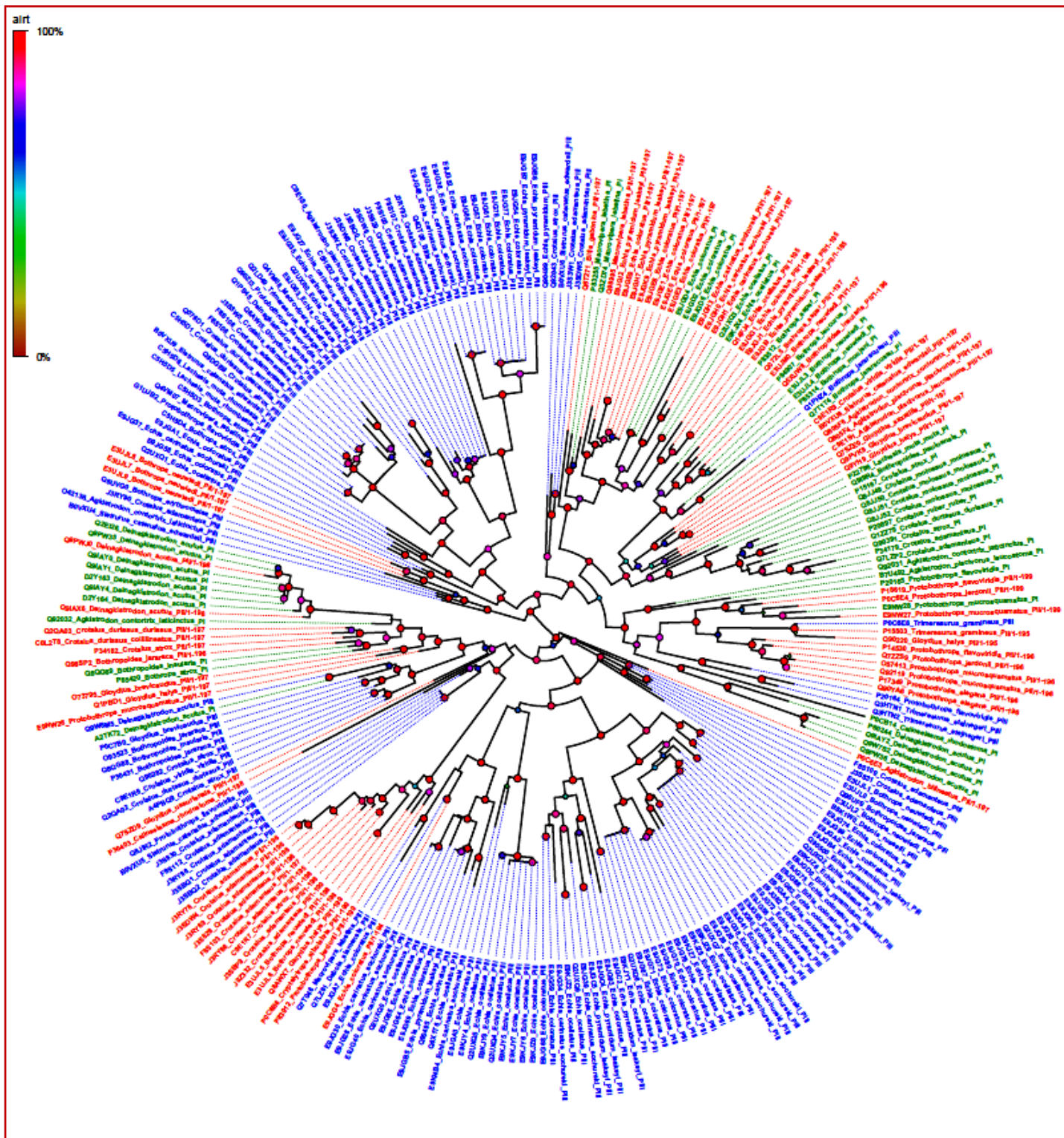


Figura 12 - Relações evolutivas das três subfamílias de SVMPs de Viperidae. Análise de máxima verossimilhança de 266 sequências de aminoácidos (205 sítios) do domínio catalítico (PF01421) de SVMPs de diferentes espécies de Viperidae. Proteínas das subfamílias SVMP-I (verde), SVMP-II (vermelho) e SVMP-III (azul) estão indicadas. JTT foi o modelo evolutivo que melhor se adequou aos dados. Valores de apoio para os nós da árvore foram estimados utilizando *Akaike Likelihood Ratio test* (aLRT). Altos valores de apoio estão destacados por círculos vermelhos.

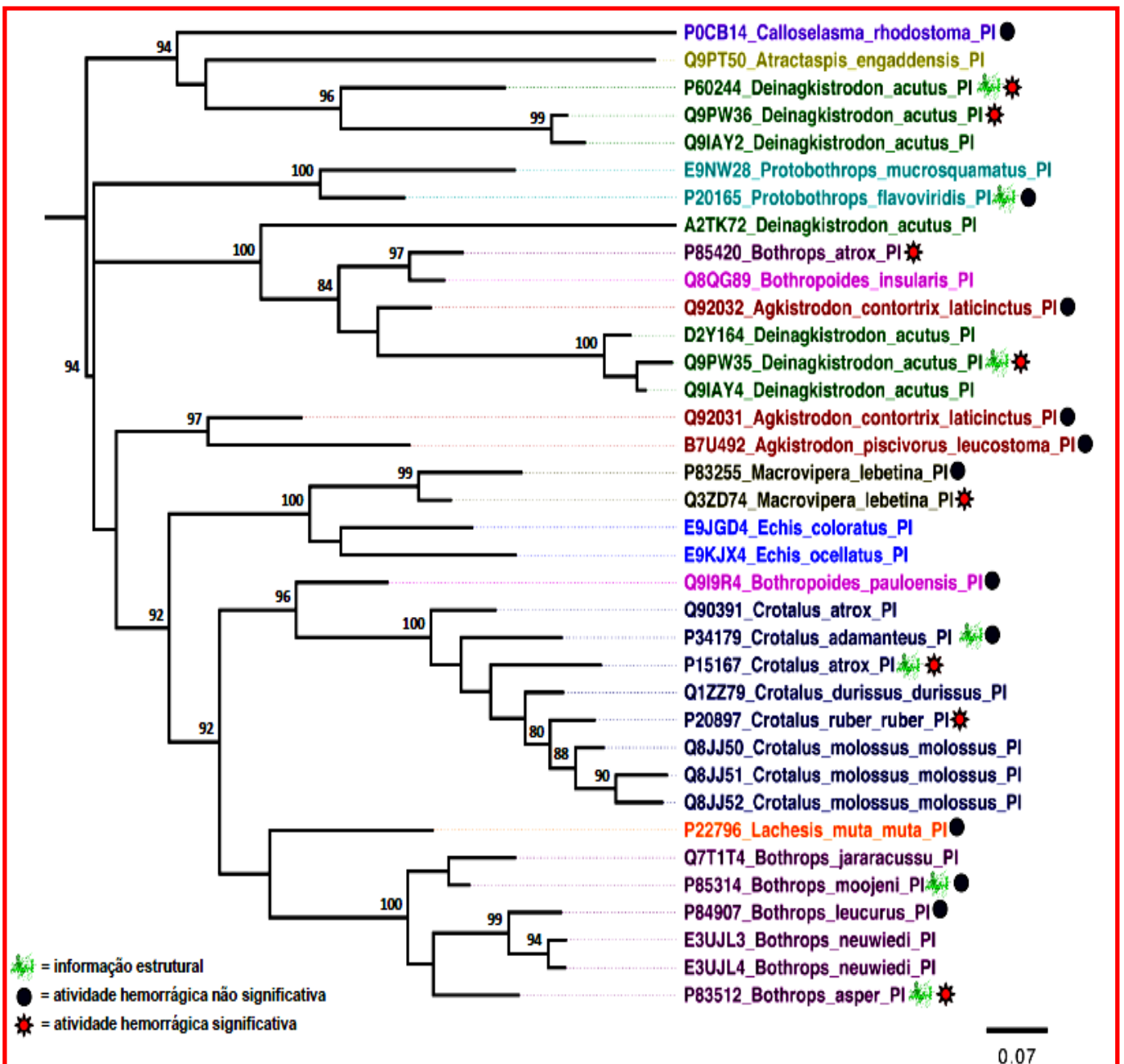


Figura 13 - Relações evolutivas da subfamília SVMP-PI de Viperidae. Análise de máxima verossimilhança de 36 sequências de aminoácidos (202 sítios) do domínio catalítico (PF01421) de SVMPs de diferentes espécies de Viperidae. WAG foi o modelo evolutivo que melhor se adequou aos dados. Valores de apoio para os nós da árvore foram estimados utilizando *Akaike Likelihood Ratio test* (aLRT).

A reconstrução dessa filogenia teve como objetivo explorar as relações evolutivas de potentes agentes fibrinolíticos com potencial aplicação em biomedicina e biotecnologia. A filogenia resultante mostra que, os homólogos de SVMP-PI não formam clados baseados em sua atividade hemorrágica, taxonomia ou distribuição geográfica dos organismos de origem (**Figura 13**). A reconstrução filogenética ainda aponta que, ao contrário do esperado, a habilidade de induzir ou não hemorragia não possui uma origem monofilética, visto que as proteínas hemorrágicas estão em clados diferentes na filogenia. Além disso, ao analisar o alinhamento não foi possível observar, por inspeção visual, assinaturas moleculares nas sequências, relacionadas à atividade hemorrágica de SVMPs (**Anexo 03**).

6 DISCUSSÃO

O mapeamento mostrou que algumas proteínas como fosfolipases A2, metaloproteinases e proteínas semelhantes à lectinas tipo C, estão presentes no transcrito desta espécie em um nível de expressão bastante alto em relação à média de expressão de transcritos na glândula de veneno. Outros peptídeos e enzimas também estão presentes em um grau de produção menos expressiva como pode ser observado na **Figura 10**. Este alto nível de expressão de algumas proteínas está relacionado às propriedades fisiológicas do veneno, que precisa dispor de função digestiva e defensiva. As PLA2, por exemplo, são enzimas esterolíticas secretadas na glândula de veneno que desempenham diversas funções na presa, entre elas, neurotoxicidade, miotoxicidade, cardiotoxicidade (Doley *et al.*, 2009), as proteínas similares à lectina tipo C do veneno que também estão altamente expressas em comparação com a média de expressão dos transcritos deste veneno, tem como principal função o desequilíbrio da homeostasia na presa, ativando ou desativando os componentes do plasma e células sanguíneas (Du & Clemetson, 2009). Contudo, a média de expressão de metaloproteinases é bem maior do que as demais proteínas do veneno sendo a principal responsável pelos sinais e sintomas observados na vítima pelo envenenamento por serpentes do gênero *Bothrops* (principalmente necrose tecidual no local da mordida, hemorragia local e sistêmica).

A enorme variedade de isoformas de transcritos que codificam os três domínios de metaloproteinases encontrados nos dados deste transcrito, é compatível com os resultados descritos por Moura da Silva e colaboradores (2011), que sugere a existência de *splicing* alternativo e recombinação de domínios pós-transcricional. Foram detectados transcritos que codificam o domínio catalítico (PF01421), transcritos que codificam os dois primeiros domínios (PF01421 e PF00200) e transcritos que codificam os três domínios (PF01421, PF00200 e PF08516), mas também foram encontrados transcritos que codificam somente o domínio desintegrina (PF00200) ou apenas o domínio rico em cisteína (PF08516). O que indica a presença de *splicing* alternativo e que a formação das SVMPs-PII e SVMPs-PIII pode se dar através de recombinação dos transcritos que codificam os domínios presente nestas proteínas.

Além de mapear e montar o transcrito da espécie *Bothrops neuwiedi*, nossa intenção era estudar especificamente as metaloproteinase e fornecer

informações relevantes associadas aos transcritos que as codificam. Com este propósito, o estudo também abrangeu a identificação de SVMPs homólogas de Viperidae em banco de dados de proteínas e a investigação de suas relações evolutivas. O número de sequências de SVMPs identificadas do banco de dados do UniProt foi bastante variado (63 para SVMP-PI, 164 para SVMP-PII e 208 para SVMP-PIII). Esta quantidade expressa tão somente o interesse das iniciativas de pesquisa em certo gênero, como é observado no número de sequências recuperadas para o gênero *Echis* (233 sequências).

Neste estudo, o interesse nas relações evolutivas de SVMPs era devido a sua importância funcional junto ao sistema hemostático, pois algumas delas não desencadeiam hemorragia, mas são fibrino(geno)líticas, o que as tornam um potencial candidato para o desenvolvimento de modelos de biofármacos para o tratamento de doenças trombóticas. Assim nossa intenção foi verificar através do estudo das relações filogenéticas se era possível distinguir aquelas que são hemorragia daquelas que não produzem hemorragia.

Segundo Gutiérrez e colaboradores (2009), a presença de SVMPs no veneno de serpentes das famílias Viperidae, Colubridae, Elapidae e Atractaspidae, sugere que um gene tipo ADAM (*A desintegrin and a metalloproteinase*) ancestral, codificador de uma metaloproteinase multidomínio (presente antes da radiação das serpentes mais recentes da superfamília Colubroidea) tenha sido recrutado para a glândula de veneno primitiva e assim a estrutura SVMP-PIII perdeu o domínio transmembrana adicional presente nas ADAMs e ausente nas SVMPs.

A constatação da ausência de uma origem monofilética para cada subfamília observada na árvore reconstruída para SVMPs PI, PII e PIII (**Figura 12**) é consistente com os estudos prévios que demonstraram que a recombinação gênica, embaralhamento de éxon, *trans-splicing* e modificações pós-traducionais podem estar envolvidos na geração da diversidade de toxinas. Sendo assim podemos verificar que:

- O recrutamento de genes para glândulas capazes de secretar o veneno resultou no surgimento de novas funções (Soto *et al.*, 2007).

- A duplicação gênica seguida pela seleção positiva de mutações não sinônimas acumuladas no genoma sob forte seleção adaptativa levou ao aparecimento de diferentes sequências; e o alto nível de variação permite uma vantagem adaptativa para diferentes ambientes e presas (Moura da Silva *et al.*, 2011).

- As modificações pós-traducionais como os episódios de alteração dos domínios e perda de domínios leva a frequentes mudanças na estrutura molecular das SVMPs (Casewell *et al.*, 2011).

A narrativa da evolução de SVMPs é que o recrutamento de uma proteína predecessora (ADAM) para um tecido especializado em excreção (glândula de veneno), seguido pela duplicação gênica, perda de domínios e recombinação de domínios no contexto de evolução acelerada (Moura-da-Silva *et al.*, 2011) permitiram a diversificação e aumentaram as probabilidades de manutenção da espécie e assim uma origem monofilética não foi possível ser constatada.

A árvore filogenética compreendendo apenas os membros de SVMP-PI (**Figura 13**) foi reconstruída para melhor entendimento das relações evolutivas destes potenciais agentes fibrino(geno)líticos. Ao contrário do que era esperada, a habilidade de produzir ou não hemorragia também não possui um origem monofilética, uma vez que proteínas hemorrágicas são encontradas em diferentes clados na filogenia. Não sendo possível identificar uma assinatura molecular relacionada à atividade hemorrágica ao nível de sequência de SVMP. No processo de evolução destas enzimas é importante considerar que SVMP-PI podem ser geradas por processos traducionais ou pós-traducionais. Estudos mais aprofundados da estrutura e função devem ser realizados para a determinação desta característica.

7 CONCLUSÕES E PERSPECTIVAS FUTURAS

Além da importância para a produção de soro antiofídico, o veneno bothropico possui uma gama de proteínas que conhecidamente são usadas como protótipos para o desenvolvimento de fármacos e para uso biotecnológico. Neste contexto, o estudo das proteínas que compõem o veneno continua sendo muito produtivo. Com os dados gerados neste trabalho foi possível montar, mapear e anotar os transcritos que codificam proteínas que constitui o veneno, verificar sua abundancia no veneno e interpretar os transcritos de SVMPs que são enzimas detentoras de grande valor biotecnológico. O cenário evolutivo reconstruído para as metaloproteinases de serpentes da família Viperidae mostrou que algumas respostas não estão no nível de sequência, mas provavelmente ao nível de estrutura.

Perspectivas futuras:

- Aprofundar as análises do transcrito de *Bothrops neuwiedi*.
- As sequências montadas vão contribuir na identificação e caracterização das proteínas que estão sendo purificadas do veneno bruto no laboratório de Bioquímica de proteínas da Fundação Ezequiel Dias.
- As ferramentas de análises filogenéticas serão utilizadas para investigar as relações evolutivas das demais proteínas presentes no veneno.
- Algumas proteínas SVMP-PI de interesse biotecnológico serão expressas em sistemas heterólogos para estudo de suas funções e estrutura.

8 ANEXOS

8.1 Anexo 1 - Tabelas com descrição dos principais transcritos que codificam proteínas do veneno de *Bothrops neuwiedi*.

Tabela 1 - Mapeamento dos transcritos que codificam fosfolipases A2 (PLA2) identificados na montagem do transcriptoma referência correção feita pelo valor de RPKM

Transcritos que codificam PLA2	Tamanho dos transcritos	Biblioteca 1	Biblioteca 2	Biblioteca 3	Biblioteca 4
comp1339_c0_seq1	238	12,07	16,31	8,20	14,29
comp338386_c0_seq1	279	0,00	1,55	0,00	3,48
comp4256_c0_seq1	207	1,73	6,25	0,00	4,69
comp4590_c0_seq1	314	148,66	237,63	479,73	607,98
comp461392_c0_seq1	255	0,00	5,07	0,00	1,90
comp4963_c0_seq1	318	1,13	12,21	0,00	1,53
comp6055_c0_seq1	778	1,38	17,19	1,50	1,87
comp6055_c0_seq2	619	1,74	16,72	1,89	3,14
comp6217_c0_seq1	251	7,15	6,87	10,88	29,03
comp6235_c0_seq1	218	3,29	21,76	21,48	4,46
comp6235_c0_seq2	301	7,16	11,46	1,30	12,91
comp648740_c0_seq1	276	0,00	0,00	1,41	1,76
comp6643_c0_seq1	719	559,82	4151,67	1434,53	1756,60
comp6726_c0_seq1	369	10,70	18,70	27,50	34,23
comp673_c0_seq1	257	2,79	6,71	0,00	0,00
comp6803_c2_seq1	289	2,48	7,46	6,75	6,72
comp6803_c2_seq13	501	4,30	25,83	18,69	15,51
comp6803_c2_seq14	361	5,97	39,43	11,89	28,26
comp6803_c2_seq15	409	1,76	23,20	8,59	11,88
comp6803_c2_seq16	381	6,60	30,56	7,17	21,67
comp6803_c2_seq3	269	4,00	11,22	1,45	7,22
comp6803_c2_seq4	373	10,59	41,63	17,79	11,72
comp6803_c2_seq9	253	5,68	40,91	9,25	24,96
comp6890_c0_seq1	713	4,03	15,12	8,21	10,22
comp6890_c0_seq2	886	8,51	23,37	11,45	10,42
comp6890_c0_seq3	233	0,00	14,81	8,37	18,76
comp6890_c0_seq4	472	1,52	10,97	5,79	2,06
comp6890_c0_seq5	927	9,68	40,94	17,26	24,63
comp6897_c2_seq2	335	7,50	29,61	9,32	23,20
comp6897_c2_seq4	763	8,47	66,70	26,60	33,74
comp6897_c2_seq5	228	0,00	3,78	5,13	8,52
comp6897_c2_seq6	364	17,76	14,22	12,87	13,35
comp6897_c2_seq7	591	12,76	53,27	33,02	30,41
comp6897_c2_seq8	400	2,69	15,10	4,88	9,72
comp6897_c2_seq9	536	14,07	64,37	24,03	32,63
TOTAL	-	886,01	5102,62	2236,94	2823,47
Média de expressão					2762,26

Tabela 2 - Mapeamento dos transcritos que codificam proteínas do veneno de serpentes similares às Lectinas tipo C.

Transcritos que codificam snaclecs	Tamanho dos transcritos	Biblioteca 1	Biblioteca 2	Biblioteca 3	Biblioteca 4
comp407048_c0_seq1	264	1,36	4,90	2,96	5,52
comp6195_c0_seq1	288	11,22	89,85	127,37	96,14
comp6838_c0_seq1	228	14,17	191,06	160,89	85,22
comp6838_c0_seq10	1301	64,03	526,45	851,59	559,70
comp6838_c0_seq11	693	10,88	89,62	82,78	61,68
comp6838_c0_seq12	641	15,68	113,71	113,85	90,18
comp6838_c0_seq13	433	16,59	130,49	257,76	120,04
comp6838_c0_seq14	1179	49,34	387,77	594,15	281,41
comp6838_c0_seq15	433	33,17	214,16	204,59	205,30
comp6838_c0_seq16	350	89,25	587,80	502,86	370,57
comp6838_c0_seq17	667	68,91	505,67	472,16	284,03
comp6838_c0_seq18	604	14,86	103,54	69,78	70,77
comp6838_c0_seq19	571	14,46	101,97	120,29	62,10
comp6838_c0_seq2	484	32,64	249,51	181,42	156,57
comp6838_c0_seq20	453	40,42	266,59	331,67	236,99
comp6838_c0_seq21	591	21,26	159,82	225,17	87,95
comp6838_c0_seq22	609	13,56	145,89	199,93	134,00
comp6838_c0_seq23	708	11,66	99,91	110,24	80,28
comp6838_c0_seq24	504	8,55	94,99	63,49	41,44
comp6838_c0_seq25	547	34,13	276,76	230,44	151,86
comp6838_c0_seq26	626	6,88	119,88	38,03	32,59
comp6838_c0_seq27	591	20,66	140,12	184,89	103,56
comp6838_c0_seq28	571	8,80	88,38	146,94	57,85
comp6838_c0_seq29	550	18,28	203,89	320,00	148,38
comp6838_c0_seq3	524	25,35	194,25	231,62	142,76
comp6838_c0_seq4	693	9,33	72,20	92,92	52,57
comp6838_c0_seq5	679	35,96	223,59	169,55	110,17
comp6838_c0_seq6	672	16,03	177,14	242,74	130,12
comp6838_c0_seq7	708	11,16	84,07	90,95	57,63
comp6838_c0_seq8	480	16,46	154,55	212,20	130,55
comp6838_c0_seq9	571	54,08	666,22	666,36	404,95
comp6872_c2_seq1	333	26,96	249,97	134,77	86,07
TOTAL	-	816,11	6.714,73	7.434,37	4.638,96
Média de expressão					4.901,04

Tabela 3 - Mapeamento dos transcritos que codificam inibidores de PLA2.

Transcrito que codificam inibidores de PLA2	Tamanho dos transcritos	Biblioteca 1	Biblioteca 2	Biblioteca 3	Biblioteca 4
comp357886_c0_seq1	230	0,00	1,88	3,39	8,45
comp4963_c0_seq1	318	1,13	12,21	0,00	1,53
comp6658_c0_seq1	490	9,53	39,61	17,52	44,61
comp6692_c1_seq1	568	4,43	31,89	12,37	16,25
comp673_c0_seq1	257	2,79	6,71	0,00	0,00
comp6827_c0_seq1	203	1,77	4,25	5,77	7,18
comp6827_c0_seq2	352	13,26	42,89	23,28	46,92
comp6827_c0_seq3	341	9,48	43,00	22,89	32,76
comp6827_c0_seq4	340	7,39	13,95	9,18	55,72
comp6827_c0_seq5	219	4,92	23,63	8,91	44,36
comp6827_c0_seq6	351	3,07	9,83	0,00	23,53
comp6897_c2_seq2	335	7,50	29,61	9,32	23,20
comp6897_c2_seq4	763	8,47	66,70	26,60	33,74
comp6897_c2_seq5	228	0,00	3,78	5,13	8,52
comp6897_c2_seq6	364	17,76	14,22	12,87	13,35
comp6897_c2_seq7	591	12,76	53,27	33,02	30,41
comp6897_c2_seq8	400	2,69	15,10	4,88	9,72
comp6897_c2_seq9	536	14,07	64,37	24,03	32,63
TOTAL	-	121,01	476,91	219,15	432,87
Média da expressão					312,49

Tabela 4 - mapeamento dos transcritos que codificam peptídeos natriuréticos potencializador de Bradicininina identificados no transcrito.

Transcritos que codificam peptídeos natriuréticos	Tamanho dos transcritos	Biblioteca 1	Biblioteca 2	Biblioteca 3	Biblioteca 4
comp6119_c0_seq1	363	364,01	1204,80	1022,39	492,46
comp6777_c0_seq1	766	239,06	738,73	464,63	323,42
comp6777_c0_seq2	910	293,57	958,82	920,30	397,15
TOTAL		896,64	2902,35	2407,33	1213,03
Média da expressão					1854,84

Tabela 5 - Mapeamento dos transcritos que codificam serino-proteínases no veneno desta espécie.

Transcritos que codificam Serino proteinase	Tamanho dos transcritos	Biblioteca 1	Biblioteca 2	Biblioteca 3	Biblioteca 4
comp4862_c0_seq1	294	0,00	1,47	2,65	18,17
comp6608_c0_seq1	217	8,27	27,83	25,18	94,02
comp6702_c1_seq1	303	8,30	37,01	86,29	33,67
comp6702_c1_seq2	274	9,17	39,35	24,21	23,05
comp6702_c1_seq3	286	10,04	34,69	38,21	23,78
comp6702_c1_seq4	264	6,80	32,67	70,95	34,96
comp6702_c1_seq5	240	10,47	30,55	82,93	20,24
comp6702_c1_seq6	242	7,42	12,48	38,70	58,21
comp6702_c1_seq7	281	5,11	26,09	20,83	6,91
comp6713_c0_seq1	205	5,25	63,12	106,60	47,39
comp6713_c0_seq2	211	17,02	114,47	192,35	156,55
comp6713_c0_seq3	230	18,73	39,38	78,05	82,37
comp6776_c0_seq1	369	21,41	77,14	96,24	76,35
comp6776_c0_seq2	711	17,17	31,54	48,85	84,04
comp6776_c0_seq3	1024	44,18	84,24	147,49	288,90
comp6895_c1_seq10	565	12,71	138,93	265,23	175,39
comp6895_c1_seq1	882	22,80	98,29	178,31	134,38
comp6895_c1_seq2	201	12,50	19,31	52,42	87,00
comp6895_c1_seq3	2073	17,84	45,77	81,14	76,86
comp6895_c1_seq4	632	23,86	51,87	98,18	219,82
comp6895_c1_seq5	1754	14,74	42,29	76,31	79,76
comp6895_c1_seq6	665	5,40	32,43	56,34	48,21
comp6895_c1_seq7	2747	16,73	53,54	99,44	118,66
comp6895_c1_seq8	2438	14,73	48,47	90,92	126,52
comp6895_c1_seq9	984	6,93	35,94	65,44	50,85
TOTAL	-	337,60	1218,88	2123,28	2166,08
Média da expressão					1461,46

Tabela 6 - Mapeamento dos transcritos que codificam proteínas com motivos três dedos (*three finger*).

Transcritos que codificam proteínas três dedos	Tamanho dos transcritos	Bibliotecas			
		Biblioteca 1	Biblioteca 2	Biblioteca 3	Biblioteca 4
comp6803_c2_seq10	746	3,37	18,50	5,23	11,07
comp6803_c2_seq11	552	3,90	34,38	6,36	14,08
comp6803_c2_seq12	406	22,99	103,05	37,49	71,79
comp6803_c2_seq1	289	2,48	7,46	6,75	6,72
comp6803_c2_seq13	501	4,30	25,83	18,69	15,51
comp6803_c2_seq14	361	5,97	39,43	11,89	28,26
comp6803_c2_seq15	409	1,76	23,20	8,59	11,88
comp6803_c2_seq16	381	6,60	30,56	7,17	21,67
comp6803_c2_seq17	283	31,72	27,43	11,03	25,75
comp6803_c2_seq18	726	2,47	12,48	3,23	11,37
comp6803_c2_seq2	572	10,04	46,00	11,60	23,78
comp6803_c2_seq3	269	4,00	11,22	1,45	7,22
comp6803_c2_seq4	373	10,59	41,63	17,79	11,72
comp6803_c2_seq5	423	197,78	56,08	22,14	111,39
comp6803_c2_seq6	286	1,26	6,03	4,09	3,40
comp6803_c2_seq7	266	13,50	63,24	10,27	51,13
comp6803_c2_seq8	303	3,56	18,50	0,00	12,83
comp6803_c2_seq9	253	5,68	40,91	9,25	24,96
comp6897_c2_seq2	335	7,50	29,61	9,32	23,20
comp6897_c2_seq4	763	8,47	66,70	26,60	33,74
comp6897_c2_seq5	228	0,00	3,78	5,13	8,52
comp6897_c2_seq6	364	17,76	14,22	12,87	13,35
comp6897_c2_seq7	591	12,76	53,27	33,02	30,41
comp6897_c2_seq8	400	2,69	15,10	4,88	9,72
comp6897_c2_seq9	536	14,07	64,37	24,03	32,63
TOTAL	-	395,22	852,99	308,87	616,10
Média da expressão					543,29

Tabela 7 - Transcritos que codificam L-aminoácido oxidase no veneno.

Transcrito que codificam LAO	Tamanho dos transcritos	Bibliotecas			
		Biblioteca 1	Biblioteca 2	Biblioteca 3	Biblioteca 4
comp5942_c0_seq1	280	56,42	317,32	505,93	511,79
comp6811_c0_seq1	733	23,51	210,65	212,96	135,19
comp6811_c0_seq2	2571	52,93	400,77	488,76	396,59
TOTAL	-	132,87	928,74	1207,65	1043,57
Média da expressão					828,21

Tabela 8 - Mapeamento dos transcritos que codificam os domínios presente em metaloproteinases de veneno de serpentes. ¹ Identificador Pfam (Banco de dados público de domínios proteicos).

Transcritos que codificam os domínios de SVMPs	Tamanho transcrito	Serpente				Domínio ¹
		Serpente 1	Serpente 2	Serpente 3	Serpente 4	
comp13235_c0_seq1	210	17,10	260,84	302,91	94,84	
comp15932_c0_seq1	342	18,90	244,66	257,88	73,86	
comp6505_c0_seq1	320	10,22	117,26	231,71	69,83	
comp6505_c0_seq2	300	4,79	27,32	63,74	8,10	
comp6505_c0_seq3	316	0,00	13,65	8,64	0,00	
comp6505_c0_seq4	316	1,14	15,01	17,29	6,15	
comp6505_c0_seq5	386	26,05	89,39	209,28	95,64	
comp6807_c0_seq1	497	140,16	95,46	1622,25	176,91	
comp6807_c0_seq2	720	41,39	101,24	204,88	72,19	
comp6807_c0_seq3	391	228,66	583,53	1291,52	535,46	
comp6807_c0_seq4	714	92,53	301,43	1031,92	432,02	metaloproteinase (PF01421)
comp6807_c0_seq5	500	110,59	299,32	601,77	396,39	
comp6807_c0_seq6	506	95,80	242,93	163,50	69,12	
comp4716_c0_seq1	312	8,06	62,21	85,05	77,85	
comp4779_c0_seq1	261	46,77	41,31	134,57	130,28	
comp6199_c0_seq1	369	21,41	101,69	216,81	81,62	
comp6872_c1_seq2	1928	173,57	709,14	1146,06	388,76	
comp6872_c1_seq5	1946	221,42	812,52	1191,41	475,53	
comp6872_c1_seq7	393	74,92	229,37	384,29	359,69	
comp3542_c0_seq1	1186	0,61	8,36	1,32	4,10	Desintegrina (PF00200)
comp8117_c0_seq1	222	72,78	151,54	423,65	210,06	
comp5924_c0_seq1	302	99,87	515,56	885,17	394,08	
comp4715_c0_seq1	277	19,44	37,37	164,83	3,51	Domínio rico em cisteína (PF08516)
comp6336_c0_seq1	259	56,84	164,86	82,87	105,03	
comp6336_c0_seq2	202	127,98	360,84	585,37	303,00	
comp6872_c1_seq1	863	121,49	389,32	1008,41	464,94	Metaloproteinase + desintegrina (PF01421 + PF00200)
comp6872_c1_seq3	2134	186,94	745,38	1152,83	473,25	
comp6872_c1_seq4	2125	140,99	591,84	905,63	397,51	
comp6872_c1_seq6	845	142,35	423,65	839,15	396,66	
comp6190_c0_seq1	765	737,84	4593,24	3430,63	1610,97	Metaloproteinase + D. rico em cisteína (PF01421 + PF08516)
comp6354_c0_seq1	302	63,01	165,67	100,79	146,37	
comp6354_c0_seq2	302	34,48	99,97	254,57	146,37	

8.2 Anexo 2 - Tabelas da montagem dos transcritos de SVMP em cada um das bibliotecas.

Tabela 1 - Transcritos de SVMPs identificados na Biblioteca 1 (indivíduo proveniente da cidade de Brumadinho)

ID_transcritos	Domínios encontrados	Tamanho	Subfamília
Contig_55	PF01421	645	PI
Contig_56	PF01421	727	PI
Contig_59	PF01421	808	PI
Contig_60	PF01421	1118	PI
Comp909_seq7	PF01562 e PF01421	1798	PI
Comp909_seq9	PF01421	788	PI
Comp909_seq10	PF01421	808	PI
Comp909_seq13	PF01421	1098	PI
Comp909_seq16	PF01421	1118	PI
Contig_58	PF01421 e PF00200	1344	PII
Comp909_seq12	PF01421 e PF00200	1324	PII
Comp909_seq5	PF01421 e PF00200	1344	PII
Comp909_seq19	PF01562, PF01421 e PF00200	2024	PII
Comp909_seq1	PF01421, PF00200 e PF08516	1650	PIII
Comp909_seq2	PF01421, PF00200 e PF08516	1630	PIII
Comp909_seq14	PF01562, PF01421, PF00200 e Pf08516	2330	PIII
Contig_57	PF01421, PF00200 e PF08516	1650	PIII

Tabela 2 - Transcritos de SVMPs identificados na Biblioteca 2 (indivíduo proveniente da cidade de Nova Lima).

Transcritos	Domínio	Tamanho	Subfamília
Contig_192	PF01421	368	PI
Contig_280	PF01421	448	PI
Contig_324	PF01421	1969	PI
Comp_2490_seq3	PF01421 e PF00200	715	PII
Comp_2490_seq1	PF01421 e PF00200	696	PII
Comp_2662_seq1	PF01421, PF00200 e PF08516	1734	PIII
Comp_2662_seq4	PF01562, PF01421, PF00200 e PF08516	2318	PIII
Comp_2662_seq6	PF01562, PF01421, PF00200 e PF08516	2387	PIII
Contig_281	PF01561, PF01421, PF00200 e PF08516	2387	PIII

Tabela 3 - Transcritos que codificam SVMPs da biblioteca 3 (indivíduo proveniente da cidade de Nova Lima).

Transcritos	Domínios	Tamanho	Subfamília
Contig_37	PF01421	373	PI
Comp_1447_seq4	PF01421	484	PI
Comp_1447_seq5	PF01421	497	PI
Comp_1449_seq3	PF01562 e PF01421	1163	PI
Comp_38505_seq1	PF01421	376	PI
Contig_104	PF01562, PF01421 e PF00200	2032	PII
Contig_105	PF01421 e PF00200	518	PII
Comp_1449_seq2	PF01421 e PF00200	940	PII
Comp_1449_seq9	PF01421, PF00200 e PF08516	670	PIII

Tabela 4 - Transcritos que codificam SVMPs identificado na Biblioteca 4 (individuo proveniente da cidade de Conselheiro Lafaiete)

Transcritos	Domínio	Tamanho	Subfamília
Contig_172	PF01421	466	PI
Contig_188	PF01562 e PF01421	1093	PI
Comp_923_seq1	PF01421	425	PI
Comp_1867_seq1	PF01421	364	PI
Comp_2145_Seq1	PF01421	451	PI
Comp_2145_seq3	PF01562 e PF01421	1093	PI
Comp_2145_seq6	PF01562 e PF01421	949	PI
Contig_120	PF01421 e PF00200	339	PII
Contig_121	PF01421 e PF00200	341	PII
Contig_187	PF01421 e PF00200	768	PII
Comp_2145_seq2	PF01421 e PF00200	768	PII
Comp_2001_seq1	PF08516	455	domínio rico em cisteína
Comp_2129_seq2	PF00200 e Pf08516	891	domínio rico em cisteína

8.3 Anexo 3 - Alinhamento das sequências do domínio catalítico de SVMP-PI.

	10	20	30	40	50	60																																																																	
<i>P85420_Bothrops_atrox</i>	R	Y	V	D	L	F	I	V	V	D	H	G	M	F	M	K	Y	N	G	S	D	K	I	R	R	R	I	H	Q	M	V	N	I	M	K	E	A	Y	S	T	M	Y	I	D	I	L	L	T	G	V	E	I	W	S	N	K	D	L	I	N	V	Q	P	A	A	P	Q	T	L	D	
<i>Q8QG89_Bothropoides_insularis</i>	R	Y	I	E	L	F	I	V	V	D	H	G	M	F	M	K	Y	N	G	S	D	K	I	R	R	R	I	H	Q	M	V	N	I	M	K	E	A	Y	R	Y	L	I	D	I	A	L	T	G	V	E	I	W	S	N	K	D	M	I	N	V	Q	P	A	A	P	Q	T	L	D		
<i>Q92032_Agkistrodon_contortrix_laticinctus</i>	R	Y	V	E	L	V	T	V	V	D	H	G	M	Y	T	K	Y	N	G	S	D	K	I	R	Q	W	V	H	Q	M	V	N	T	M	K	E	S	Y	R	Y	M	Y	I	D	I	S	L	A	G	V	E	I	W	S	N	K	D	L	I	D	V	Q	P	A	A	R	H	T	L	D	
<i>Q9PW35_Deinagkistrodon_acutus</i>	R	Y	M	E	I	V	I	V	V	D	H	S	M	V	K	K	Y	N	G	S	D	K	I	K	A	W	V	Y	E	M	I	N	T	I	T	E	S	Y	S	Y	L	I	D	I	L	S	G	L	E	I	W	S	E	K	D	L	I	N	V	E	T	S	A	E	N	T	L	K			
<i>Q9IAY4_Deinagkistrodon_acutus</i>	R	Y	M	E	I	V	I	V	V	D	H	S	M	Y	T	K	Y	K	G	D	S	D	K	I	K	A	W	V	Y	E	M	I	N	T	S	E	S	Y	R	Y	L	I	D	I	V	S	A	L	E	M	W	S	E	K	D	L	I	N	V	E	T	S	A	E	N	T	L	K			
<i>D2Y164_Deinagkistrodon_acutus</i>	R	Y	M	E	I	V	I	V	V	G	H	S	M	Y	V	K	Y	N	E	D	S	D	K	I	K	A	W	V	Y	Q	M	I	N	T	L	R	E	S	Y	R	Y	L	I	D	T	S	V	A	A	I	E	I	W	S	E	K	D	L	I	N	V	E	T	S	A	K	N	T	L	E	
<i>P34179_Crotalus_adamanteus</i>	R	Y	I	E	L	V	V	V	A	D	R	R	V	F	M	K	Y	N	S	D	L	N	I	R	T	R	V	H	E	I	V	N	I	N	G	F	Y	R	S	L	N	I	D	V	S	L	V	N	L	E	I	W	S	Q	D	P	L	T	I	Q	S	S	S	S	N	T	L	N			
<i>Q90391_Crotalus_atrox</i>	R	Y	I	E	L	V	V	V	A	D	H	R	V	F	M	K	Y	N	S	D	L	N	I	R	K	R	V	H	E	L	V	N	T	I	N	G	F	Y	R	S	L	N	I	D	V	S	L	T	D	L	E	I	W	S	D	D	F	I	T	V	Q	S	S	A	K	N	T	L	N		
<i>P20897_Crotalus_ruber_ruber</i>	-	Y	I	E	L	V	V	V	A	D	H	R	F	M	K	Y	N	S	D	L	N	T	I	R	T	R	V	H	E	I	V	N	F	I	N	G	F	Y	R	S	L	N	I	R	V	S	L	T	D	L	E	I	W	S	D	Q	D	F	I	T	V	Q	S	S	A	E	N	T	L	H	
<i>Q8JJ50_Crotalus_molossus_molossus</i>	-	Y	I	E	L	V	V	V	A	D	R	R	L	F	M	K	Y	K	S	D	L	N	T	I	R	T	R	V	H	E	I	V	N	F	I	N	G	F	Y	R	S	L	N	I	R	V	S	L	T	D	L	E	I	W	S	D	Q	D	Y	I	T	V	Q	S	S	A	E	N	T	L	H
<i>Q1ZZ79_Crotalus_durissus_durissus</i>	R	Y	I	E	L	V	I	V	A	D	H	R	V	F	M	K	Y	N	S	D	L	N	I	R	T	R	V	H	E	I	V	N	F	I	N	G	F	Y	R	S	L	N	I	D	V	S	L	T	D	L	E	I	W	S	D	Q	D	F	I	T	V	Q	S	S	A	E	N	T	L	H	
<i>Q8JJ51_Crotalus_molossus_molossus</i>	R	Y	V	E	L	V	I	V	A	D	H	R	M	F	M	K	Y	K	S	D	L	F	S	I	C	S	R	V	H	D	I	V	N	F	I	N	W	F	Y	R	S	L	N	I	R	V	S	L	T	D	L	G	I	W	S	D	Q	D	Y	I	T	V	Q	S	S	A	E	N	T	L	H
<i>Q8JJ52_Crotalus_molossus_molossus</i>	-	Y	I	E	L	V	V	V	A	D	H	R	L	F	M	R	Y	K	S	D	L	F	T	I	R	S	R	V	H	D	I	F	N	F	I	N	W	F	Y	R	S	L	N	I	R	A	A	L	T	D	L	G	I	W	S	D	Q	D	Y	I	T	V	Q	S	S	A	K	Y	T	L	H
<i>P15167_Crotalus_atrox</i>	R	Y	I	E	L	V	V	V	A	D	H	R	V	F	M	K	Y	N	S	D	L	N	T	I	R	T	R	V	H	E	I	V	N	F	I	N	G	F	Y	R	S	L	N	I	H	V	S	L	T	D	L	E	I	W	S	N	E	D	Q	I	N	I	Q	S	A	S	S	D	T	L	N
<i>Q9I9R4_Bothropoides_pauloensis</i>	R	Y	I	E	L	V	I	V	A	D	R	M	Y	T	K	Y	N	S	D	S	N	K	I	R	T	R	V	H	E	L	V	N	T	V	N	G	F	F	R	S	M	N	V	D	A	S	L	A	N	L	E	V	W	S	K	K	D	L	I	K	V	E	K	D	S	S	K	T	L	T	
<i>P83512_Bothrops_asper</i>	R	Y	I	E	L	A	V	V	A	D	H	G	I	F	T	K	Y	N	S	N	L	N	T	I	R	T	R	V	H	E	M	L	N	T	V	N	G	F	Y	R	S	V	D	V	H	A	P	L	A	N	L	E	V	W	S	K	Q	D	L	I	K	V	Q	K	D	S	S	K	T	L	K
<i>Q7T1T4_Bothrops_jararacussu</i>	R	Y	I	E	V	A	V	V	A	D	H	R	M	F	K	K	Y	N	S	N	L	N	T	I	R	K	W	V	H	E	M	V	N	S	M	N	G	V	Y	R	S	M	D	V	H	L	S	L	A	N	L	E	V	W	S	K	K	D	L	I	N	V	Q	K	D	S	R	E	T	L	K
<i>P85314_Bothrops_moojeni</i>	R	Y	I	E	L	V	V	V	A	D	H	G	M	F	K	K	Y	N	S	N	L	N	T	I	R	K	W	V	H	E	M	V	N	S	M	N	G	F	Y	R	S	V	D	V	T	A	S	L	A	N	L	E	V	W	S	K	K	D	L	I	N	V	Q	K	D	S	R	E	T	L	K
<i>P84907_Bothrops_leucurus</i>	R	Y	I	E	L	V	V	V	A	D	H	G	M	F	K	K	Y	N	S	N	L	N	T	I	R	K	W	V	H	E	M	L	N	T	V	N	G	F	F	R	S	M	N	V	D	A	S	L	V	N	L	E	V	W	S	K	K	D	L	I	K	V	E	K	D	S	S	K	T	L	T
<i>E3UJL3_Bothrops_neuwiedi</i>	R	Y	V	E	L	A	V	V	A	D	N	G	M	F	T	K	Y	N	S	N	L	N	T	I	R	T	R	V	H	E	M	V	N	T	V	N	G	F	F	R	S	M	N	V	D	A	S	L	A	N	L	E	V	W	S	K	K	D	L	I	K	V	E	K	D	S	S	K	T	L	T
<i>E3UJL4_Bothrops_neuwiedi</i>	K	Y	I	E	L	A	V	V	A	D	H	G	M	F	T	K	Y	N	S	N	V	N	T	I	R	T	W	H	E	M	V	N	S	L	N	G	F	F	R	S	M	N	V	D	A	S	L	V	N	L	E	V	W	S	K	K	D	L	I	K	V	E	K	D	S	S	K	T	L	T	
<i>P22796_Lachesis_muta_muta</i>	K	Y	I	E	L	V	V	V	A	D	H	G	M	F	T	K	Y	N	C	N	L	N	T	I	R	T	R	V	H	E	I	V	N	T	L	N	G	F	Y	R	S	L	N	I	L	I	S	L	T	D	L	E	I	W	S	N	Q	D	L	I	N	V	Q	S	A	A	N	D	T	L	K
<i>Q92031_Agkistrodon_contortrix_laticinctus</i>	R	Y	V	E	L	V	I	V	A	D	H	R	M	N	T	K	Y	N	G	D	S	D	K	I	R	Q	W	V	H	Q	I	V	N	T	I	N	E	I	Y	R	P	L	N	I	R	F	A	L	V	G	L	E	I	W	S	N	Q	D	L	I	T	V	T	S	V	S	H	D	T	L	A
<i>B7U492_Agkistrodon_piscivorus_leucostoma</i>	R	Y	V	E	L	V	I	I	A	D	H	R	M	Y	M	K	Y	K	R	D	S	N	K	I	T	Q	W	V	H	Q	M	V	N	T	I	N	E	I	Y	R	P	L	N	I	Q	F	A	L	V	G	L	E	I	W	S	N	Q	D	L	I	T	V	T	S	V	S	D	D	T	L	I
<i>P83255_Macrovipera_lebetina</i>	R	Y	I	E	L	V	I	V	A	D	H	S	V	A	T	K	Y	N	D	N	V	T	A	I	L	S	W	V	H	Q	L	V	N	N	I	I	L	F	Y	R	D	L	N	V	H	F	T	L	S	A	V	E	V	W	S	N	G	D	L	I	N	V	Q	P	E	A	T	V	T	L	N
<i>Q3ZD74_Macrovipera_lebetina</i>	R	Y	I	E	L	V	I	V	A	D	H	S	M	V	T	K	Y	D	G	D	L	A	A	I	R	T	W	A	H	Q	L	V	N	N	I	I	V	F	Y	R	D	L	N	V	H	I	T	L	S	A	V	E	V	W	T	N	G	D	L	I	N	V	Q	P	A	A	S	V	T	L	N
<i>E9JGD4_Echis_coloratus</i>	R	Y	I	Q	L	I	V	A	D	P	R	M	F	V	K	Y	N	S	D	L	T	A	I	R	T	L	A	H	Q	I	V	N	Y	I	M	Y	R	D	L	N	I	H	I	T	L	A	A	I	E	T	W	S	N	G	D	L	I	T	V	T	S	S	H	D	T	L	N				
<i>E9KJX4_Echis_ocellatus</i>	R	Y	I	Q	L	V	I	V	A	D	H	A	M	V	T	K	K	H	N	D	L	T	A	I	T	T	W	I	H	Q	I	V	N	D	M	I	V	M	Y	I	D	L	N	I	H	I	T	L	A	A	V	E	I	W	X	N	G	D	L	I	T	V	T	S	A	R	E	I	L	N	
<i>A2TK72_Deinagkistrodon_acutus</i>	K	Y	V	E	T	V	F	V	V	D	K	A	M	V	T	K	Y	N	G	D	L	D	K	I	K	T	R	M	Y	E	A	A	N	N	M	N	E	M	Y	R	Y	M	F	F	R	V	V	M	V	G	L	I	I	W	T	E	E	D	K	I	T	V	K	P	D	V	D	Y	T	L	N
<i>E9NW28_Protobothrops_mucrosquamatus</i>	R	Y	V	K	L	A	I	V	V	D	Y	R	M	Y	I	K	Y	N	R	D	S	N	K	I	T	V	R	V	H	E	M	V	N	H	V	N	E	M	Y	K	P	L	N	V	A	I	T	L	S	L	L	R	I	W	S	T	R	D	L	I	T	V	Q	S	D	S	K	V	T	L	G
<i>P20165_Protobothrops_flavoviridis</i>	R	Y	I	E	L	A	I	V	V	D	H	G	M	Y	K	K	Y	N	Q	N	S	D	K	I	K	V	R	V	H	Q	M	V	N	H	I	N	E	M	Y	R	P	L	N	I	A	I	S	L	N	R	L	Q	I	W	S	K	K	D	L	I	T	V	K	S	A	S	N	V	T	L	E
<i>Q9PW36_Deinagkistrodon_acutus</i>	--	V	N	L	Q	L	I	V	D	H	S	M	Y	A	K	Y	N	S	N	S	E	K	I	T	K	T	V	Q	E	R	V	N	I	M	K	E	I	F	K	P	L	N	F	D	I	T	L	S	G	I	E	M																			

	80	90	100	110	120	130	140																																																																	
<i>P85420_Bothrops_atrox</i>	S	F	G	E	W	R	K	T	D	L	L	N	R	K	S	H	D	N	A	Q	L	L	T	S	T	D	F	N	G	P	T	I	G	L	A	Y	V	G	S	M	C	D	P	K	R	S	T	G	V	I	Q	D	H	S	E	Q	D	L	M	V	A	I	T	M	A	H	E	L	G	H	N	
<i>Q8QG89_Bothropoides_insularis</i>	S	F	G	E	W	R	K	T	D	L	L	N	R	K	S	H	D	N	A	Q	L	L	T	N	T	D	F	D	G	P	T	I	G	L	A	Y	V	G	T	M	C	D	P	K	L	S	T	G	V	I	Q	D	H	S	P	I	N	L	L	V	A	V	T	M	A	H	E	L	G	H	N	
<i>Q92032_Agkistrodon_contortrix_laticinctus</i>	S	F	G	E	W	R	E	R	D	L	L	H	R	I	S	H	D	N	A	Q	L	L	T	S	T	D	F	D	G	P	T	I	G	L	A	Y	V	G	T	M	C	D	P	K	L	S	T	G	V	V	E	D	H	S	K	I	N	F	L	V	A	V	T	M	A	H	E	M	G	H	N	
<i>Q9PW35_Deinagkistrodon_acutus</i>	S	F	G	E	W	R	A	K	D	L	I	H	R	I	S	H	D	N	A	Q	L	L	T	A	T	D	F	D	G	P	T	I	G	L	A	Y	V	A	S	M	C	D	P	K	R	S	V	G	V	V	Q	D	H	S	S	V	N	R	L	V	A	I	T	L	A	H	E	M	A	H	N	
<i>Q9IAY4_Deinagkistrodon_acutus</i>	S	F	G	E	W	R	A	K	D	L	I	H	R	I	S	H	D	N	A	Q	L	L	T	A	T	D	F	D	G	P	T	I	G	L	A	Y	V	A	S	M	C	D	P	K	R	S	V	G	V	V	Q	D	H	S	S	V	N	H	L	V	A	I	T	L	A	H	E	I	A	H	N	
<i>D2Y164_Deinagkistrodon_acutus</i>	S	F	G	E	W	R	A	R	D	L	I	H	R	I	S	H	D	N	A	Q	L	L	T	A	T	D	L	D	G	P	T	I	G	L	A	Y	V	A	S	M	C	D	P	K	R	S	V	G	V	V	Q	D	H	S	S	V	N	H	L	V	A	I	T	L	A	H	E	I	A	H	N	
<i>P34179_Crotalus_adamanteus</i>	S	E	G	L	W	R	E	K	V	L	L	N	K	K	K	D	N	A	Q	L	L	T	A	I	E	F	K	C	E	T	L	G	K	A	Y	L	N	S	M	C	N	P	R	S	S	V	G	I	V	K	D	H	S	P	I	N	L	L	V	A	V	T	M	A	H	E	L	G	H	N		
<i>Q90391_Crotalus_atrox</i>	S	F	G	E	W	R	E	A	D	L	L	R	R	K	S	H	D	N	A	Q	L	L	T	A	I	N	F	E	G	K	I	G	R	A	Y	T	S	S	M	C	N	P	R	K	S	V	G	I	V	K	D	H	S	P	I	N	L	L	V	G	V	T	M	A	H	E	L	G	H	N		
<i>P20897_Crotalus_ruber_ruber</i>	S	F	G	E	W	R	K	S	V	L	L	N	R	K	R	H	D	N	A	Q	L	L	T	A	I	V	L	D	D	Y	T	L	G	L	A	Y	L	N	S	M	C	H	P	R	N	S	V	G	L	I	Q	D	H	S	P	I	N	L	L	M	G	V	T	M	A	H	E	L	G	H	N	
<i>Q8JJ50_Crotalus_molossus_molossus</i>	S	F	G	E	W	R	E	S	V	L	L	N	R	K	R	H	D	N	A	Q	L	L	T	A	I	V	L	D	E	D	T	L	G	L	A	Y	L	S	S	M	C	N	P	R	T	S	V	G	I	I	Q	D	H	S	P	I	N	L	L	M	G	V	T	M	A	H	E	L	G	H	N	
<i>Q1ZZ79_Crotalus_durissus_durissus</i>	S	F	G	E	W	R	K	S	V	L	L	N	R	K	R	H	D	N	A	Q	L	L	T	A	I	V	L	D	E	D	T	L	G	L	A	Y	L	S	S	M	C	H	P	K	H	S	V	G	I	V	Q	D	H	S	P	I	N	L	L	M	G	V	T	M	A	H	E	L	G	H	N	
<i>Q8JJ51_Crotalus_molossus_molossus</i>	S	F	G	E	W	G	K	S	V	L	L	K	Q	K	R	H	D	N	A	Q	L	L	T	A	I	V	L	D	E	D	T	L	G	L	A	Y	L	S	S	M	C	N	P	W	T	S	V	E	I	I	Q	D	H	S	P	I	N	L	L	M	G	V	T	M	A	H	E	L	G	H	N	
<i>Q8JJ52_Crotalus_molossus_molossus</i>	S	F	G	E	W	R	K	S	V	L	L	K	H	K	R	H	D	N	A	Q	L	L	T	A	I	V	L	D	E	D	T	L	G	L	A	Y	L	S	S	M	C	N	P	R	T	S	V	E	I	I	Q	D	H	S	Q	I	N	L	L	M	G	V	T	M	A	H	E	L	G	H	N	
<i>P15167_Crotalus_atrox</i>	A	F	A	E	W	R	E	T	D	L	L	N	R	K	S	H	D	N	A	Q	L	L	T	A	I	E	L	D	E	E	T	L	G	L	A	P	L	G	T	M	C	D	P	K	L	S	I	G	I	V	Q	D	H	S	P	I	N	L	L	M	G	V	T	M	A	H	E	L	G	H	N	
<i>Q9I9R4_Bothropoides_pauloensis</i>	S	F	G	E	W	R	E	R	D	L	L	R	R	K	S	H	D	N	A	Q	L	L	T	A	I	D	F	N	G	N	T	I	G	R	A	Y	L	G	S	M	C	N	P	K	R	S	V	G	I	V	Q	D	H	S	P	I	N	L	L	V	G	V	T	M	A	H	E	L	G	H	N	
<i>P83512_Bothrops_asper</i>	S	F	G	E	W	R	E	R	D	L	L	P	R	I	S	H	D	N	A	Q	L	L	T	A	V	V	F	D	G	N	T	I	G	R	A	Y	T	G	G	M	C	D	P	R	H	S	V	G	V	V	R	D	H	S	K	N	N	L	W	V	A	V	T	M	A	H	E	L	G	H	N	
<i>Q7T1T4_Bothrops_jararacussu</i>	S	F	G	E	W	R	E	R	D	L	L	P	R	I	S	H	D	N	A	Q	L	L	T	A	V	V	F	D	Q	Q	T	I	G	R	A	Y	I	A	G	M	C	D	P	R	H	S	V	G	V	V	M	D	H	S	K	E	N	L	Q	V	A	V	T	M	A	H	E	L	G	H	N	
<i>P85314_Bothrops_moojeni</i>	S	F	G	E	W	R	E	R	D	L	L	P	R	I	S	H	D	N	A	Q	L	L	T	A	I	V	F	D	G	H	T	I	G	R	A	Y	T	G	G	M	C	D	P	R	H	S	V	G	V	V	M	D	H	S	P	K	N	L	Q	V	A	V	T	M	A	H	E	L	G	H	N	
<i>P84907_Bothrops_leucurus</i>	S	F	G	E	W	R	E	R	D	L	L	P	R	I	S	H	D	N	A	Q	L	L	T	V	I	F	L	D	E	E	T	I	G	I	A	Y	T	A	G	M	C	D	L	S	Q	S	V	A	V	V	M	D	H	S	K	K	N	L	R	V	A	V	T	M	A	H	E	L	G	H	N	
<i>E3UJL3_Bothrops_neuwiedi</i>	S	F	G	E	W	R	E	R	D	L	L	P	R	I	S	H	D	N	A	Q	L	L	T	T	I	V	F	D	Q	Q	T	I	G	M	A	Y	T	A	G	M	C	D	P	S	Q	S	V	A	V	V	M	D	H	S	K	K	N	I	R	V	A	V	T	M	A	H	E	L	G	H	N	
<i>E3UJL4_Bothrops_neuwiedi</i>	S	F	G	E	W	R	E	R	D	L	L	P	R	I	S	H	D	N	A	Q	L	L	T	T	I	V	F	D	Q	Q	T	I	G	I	A	Y	T	A	G	M	C	D	P	S	Q	S	V	A	V	V	M	D	H	S	K	K	N	I	R	V	A	V	T	M	A	H	E	L	G	H	N	
<i>P22796_Lachesis_muta_muta</i>	T	F	G	E	W	R	E	R	V	L	L	N	R	I	S	H	D	N	A	Q	L	L	T	A	I	D	L	A	D	N	T	I	G	I	A	Y	T	G	G	M	C	Y	P	K	N	S	V	G	I	V	Q	D	H	S	P	K	T	L	L	I	A	V	T	M	A	H	E	L	G	H	N	
<i>Q92031_Agkistrodon_contortrix_laticinctus</i>	S	F	G	N	W	R	E	T	D	L	L	R	R	Q	R	H	D	N	A	Q	L	L	T	A	I	D	F	D	G	D	T	V	G	L	A	Y	V	G	G	M	C	L	K	H	S	T	G	V	I	Q	D	H	S	A	I	N	L	L	V	A	L	T	M	A	H	E	L	G	H	N		
<i>B7U492_Agkistrodon_piscivorus_leucostoma</i>	S	F	A	N	W	R	E	T	V	L	L	R	R	K	S	H	D	N	A	Q	L	L	T	A	I	V	F	D	E	N	T	I	G	R	A	P	L	A	G	M	C	D	P	N	R	S	V	G	T	V	Q	D	H	S	K	I	N	F	R	V	A	I	M	A	H	E	I	G	H	N		
<i>P83255_Macrovipera_lebetina</i>	L	F	G	E	W	R	E	R	D	L	L	N	R	R	M	H	D	N	A	Q	L	L	T	N	N	V	A	L	D	D	N	T	I	G	L	A	Y	D	E	G	M	C	D	P	K	Y	S	V	G	I	V	K	D	H	S	A	I	N	R	M	V	A	A	T	M	A	H	E	I	G	H	N
<i>Q3ZD74_Macrovipera_lebetina</i>	L	F	G	E	W	R	E	R	D	L	L	N	R	R	M	H	D	N	A	Q	L	L	T	A	I	N	L	D	D	N	T	I	G	L	A	Y	N	E	G	M	C	D	P	K	Y	S	V	G	I	V	Q	D	H	S	A	I	N	R	M	V	A	A	T	M	A	H	E	I	G	H	N	
<i>E9JGD4_Echis_coloratus</i>	S	F	G	E	W	R	E	R	V	L	L	N	R	R	T	H	D	N	A	Q	L	F	T	N	I	N	F	S	G	N	V	I	G	R	A	Y	L	N	G	M	C	D	P	R	R	S	V	G	I	V	T	D	H	S	T	E	R	I	L	V	A	A	T	M	A	H	E	M	G	H	N	
<i>E9KJX4_Echis_ocellatus</i>	S	F	G	E	W	R	Q	R	D	L	V	N	R	I	T	H	D	N	A	Q	L	L	T	A	V	N	L	X	G	D	T	I	G	R	A	Y	M	E	S	M	C	D	P	K	K	S	V	G	I	N	Q	D	H	S	T	V	H	L	L	V	A	V	T	M	A	H	E	L	G	H	N	
<i>A2TK72_Deinagkistrodon_acutus</i>	A	F	A	E	W	R	K	T	Y	L	L	A	E	K	K	H	D	N	A	Q	L	I	T	G	I	D	F	R	G	S	I	G	Y	A	I	G	S	M	C	H	P	K	R	S	V	G	I	I	Q	D	Y	S	P	I	N	L	V	L	A	V	I	M	A	H	E	M	G	H	N			
<i>E9NW28_Protobothrops_mucrosquamatus</i>	S	F	G	D	W	R	K	T	V	L	L	K	Q	Q	S	H	D	N	A	Q	L	L	T	D	I	T	F	T	K	N	V	I	G	V	A	Y	K	K	G	M	C	D	P	K	L	S	V	G	L	V	Q	D	Y	S	N	V	F	V	A	A	I	M	T	H	E	L	G	H	N			
<i>P20165_Protobothrops_flavoviridis</i>	S	F	G	N	W	R	E	T	V	L	L	K	Q	Q	N	D	C	A	H	L	L	T	A	T	N	L	N	D	N	T	I	G	L	A	Y	K	K	G	M	C	N	P	K	L	S	V	G	L	V	Q	D	Y	S	P	N	V	F	M	V	A	V	T	M	T	H	E	L	G	H	N		
<i>Q9PW36_Deinagkistrodon_acutus</i>	L	F	A	K	W	R	Q	T	D	L	L	K	R	I	D	N	D	N	A	Q	L	Q	T	A	V	D	F	D	G	E	T	V	G	L	A	F	K	S	T	M	C	D	K	R	Y	S	A	G	I	I	Q	D	H	S	A	I	P	L														

	150	160	170	180	190	200	
<i>P85420_Bothrops_atrox</i>	LGISHDTG	SCSCG	GYS CIMS PVL	SEPSKYFSDCSYIQ	CWDFIMKENP	QCILNK-	
<i>Q8QG89_Bothropoides_insularis</i>	LGISHDTD	SCSCG	GYS CIMAP EIS	SEPSKYFSDCSYIQ	CWDFIMKENP	QCILNK-	
<i>Q92032_Agkistrodon_contortrix_laticinctus</i>	LGMHRDTG	SCSCG	GYS CIMS PVI	SDDSPKYFSNCSYIQ	CWDFIMKENP	QCILNKP	
<i>Q9PW35_Deinagkistrodon_acutus</i>	LGVRRDEG	SCSCG	GYTCIMS PVI	NSEVIKYFSDCSYIQ	CREYISKENP	PCILNKP	
<i>Q9IAY4_Deinagkistrodon_acutus</i>	LGVVHDEG	SCSCG	GYTCIMS PVI	NSEVIKYFSDCSYIQ	CREYISKENP	PCILNKP	
<i>D2Y164_Deinagkistrodon_acutus</i>	LGVVHDEG	SCSCG	GYTCIMS PVI	NSEVIKYFSDCSYIQ	CREYISKENP	PCILNKP	
<i>P34179_Crotalus_adamanteus</i>	LGMEDGK	DCLRG	ASLCIMR PGL	T PGRS YEF SDDSMG	YYQKFLNQYK	PQCILNKP	
<i>Q90391_Crotalus_atrox</i>	LGMNH DGD	KCLRG	ASLCIMR PGL	T PGRS YEF SDDSMG	YYQSFLNQYK	PQCILNKP	
<i>P20897_Crotalus_ruber_ruber</i>	LGMEDGK	DCLRG	ASLCIMR PGL	T PGRS YEF SDDSMR	YYQKFLDQYK	PQCILNKP	
<i>Q8JJ50_Crotalus_molossus_molossus</i>	LGMEDGK	DCLRG	ASLCIMR PGL	T PGRS YEF SDDSMH	YYERFLKQYN	PQCILNKP	
<i>Q1ZZ79_Crotalus_durissus_durissus</i>	LGMKH DGD	ECLRG	ASLCIMR PGL	T PGRS YEF SDDSMG	YYQSFLNQYK	PQCILNKP	
<i>Q8JJ51_Crotalus_molossus_molossus</i>	LGMKH DEK	DCLRG	ATLCIMR PGL	T PGRS YEF SDDSMG	YYQSFLDQYK	PQCILNKP	
<i>Q8JJ52_Crotalus_molossus_molossus</i>	LGMEDGK	DCLRG	ATLCIMR PGL	T PGRS YEF SDDSMH	YYESFLKQYK	PQCILNKP	
<i>P15167_Crotalus_atrox</i>	LGMEDGK	DCLRG	ASLCIMR PGL	T KGRS YEF SDDSMH	YYERFLKQYK	PQCILNKP	
<i>Q919R4_Bothropoides_pauloensis</i>	LGMEDGK	DCLCG	ASLCIMS PGL	T DGPS YEF SDCSKD	YYQTFLTNHN	PQ-----	
<i>P83512_Bothrops_asper</i>	LGIHHD TG	SCSCG	AKS CIMS VLS	KVLS YEF SDCSQN	QYET YL TNHN	PQCILNKP	
<i>Q7T1T4_Bothrops_jararacussu</i>	LGMEDEN	QCHCD	APSCVMAS VLS	VVLS YEF SDCSQN	QYQT YL TKHN	PQCILNEP	
<i>P85314_Bothrops_moojeni</i>	LGMHHDGN	QCHCD	AASCIMAD SLS	VVLS YEF SDCSQN	QYQT YL TKHN	PQCILNEP	
<i>P84907_Bothrops_leucurus</i>	LGMHR DGN	QCHCN	APSCIMAD TLS	KGLS YEF SDCSQN	QYQT YL TKHN	PQCILNKP	
<i>E3UJL3_Bothrops_neuwiedi</i>	LGMHDH D	TCTCG	AKSCIMAS TIS	KGLS FEF SDCSQN	QYQT YVTKHN	PQCILNKP	
<i>E3UJL4_Bothrops_neuwiedi</i>	LGMHDH D	TCTCG	AKSCIMAS TIS	KGLS FEF SDCSQN	QYQT YVTKHN	PQCILNKP	
<i>P22796_Lachesis_muta_muta</i>	LGMKH DEN	HCHCS	ASF CIMP S IS	SEGPS YEF SDCSKD	YYQMFLTKRK	PQCILNKP	
<i>Q92031_Agkistrodon_contortrix_laticinctus</i>	LGMNH DGN	QCHCG	ANSCVMP S VLS	DQPS KLF SDCSKK	DYQTFLPVNN	PQCILNKP	
<i>B7U492_Agkistrodon_piscivorus_leucostoma</i>	LGMGH DDN	SCTCG	GYS CIML PRL	S KQPS KLF SDCSKK	DYLTFLKVKN	PQCILNKP	
<i>P83255_Macrovipera_lebetina</i>	LGMNH DGS	QCNCG	GNG CVMS AVL	MQQHS YQF SDCSKD	EYQR YL TNHN	PQCILNQP	
<i>Q3ZD74_Macrovipera_lebetina</i>	LGMHDH DGN	QCNCG	ANG CVMS AVI	TQQRS YQF SDCSKN	KYQTYL TNHN	PQCILNQP	
<i>E9JGD4_Echis_coloratus</i>	LGMNH DGN	QCNCG	ANG CVMS AML	VRQRS YQF SDCSMN	EYQS YL TNHN	PQCILNQP	
<i>E9KJX4_Echis_ocellatus</i>	LGMHDH DEN	QCNCG	ASG CVMS GEL	RAQDS YHF SDCSKN	EYQS YIATYN	PQCILNQP	
<i>A2TK72_Deinagkistrodon_acutus</i>	LGIHHD DG	YCYCG	GYP CIMG P S IS	SEPS KFF SDCSYI	QCWDFIMNHN	PECIDNEP	
<i>E9NW28_Protobothrops_mucrosquamatus</i>	LGMEDH DED	ENGKKCK	CD---TCIMS	PAIS DPPAQLFSDCS	KNDYHTFLTNSK	PQCILNAP	
<i>P20165_Protobothrops_flavoviridis</i>	LGMEDH DDK	KCKCE	---ACIMS	DVIS DKPSKLFSDCS	KNDYQTFLTKNY	PQCILNAP	
<i>Q9PW36_Deinagkistrodon_acutus</i>	LGMHDH DDT	S---KCN	---VCIMAP	R LNTNPSKTFSDCS	NNDYQKFLTDKK	PKCI-HKK	
<i>Q9IAY2_Deinagkistrodon_acutus</i>	LGMHDH DDT	Y---KCN	---VCIMP	R LNTNPSKTFSDCS	NNDYQKFLTDKK	PKCI-HKK	
<i>P60244_Deinagkistrodon_acutus</i>	LGMNH DGA	---GCS	---TCIMAP	VLS SGP AKSFSDCS	KHDYQSFLT IHK	PQCLLN---	
<i>P0CB14_Calloselasma_rhodostoma</i>	LGMEDH DEA	---NCK	---ACVMAP	EVN NNP T KKFSDCS	RNYYQKFLKDR	PKPECLFKKP	
<i>Q9PT50_Atractaspis_engaddensis</i>	LGIHHD DT	---ACN	---SNS	CIMS PEIGDD	PLNEFSSCS	LQEHQR YL INN	MPCCLLNKP

9 REFERÊNCIAS BIBLIOGRÁFICAS

Abascal F, Zardoya R, Posada D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 2005, 9: 2104-5.

Agero U, Arantes RM, Lacerda-Queiroz N, Mesquita ON, Magalhães A, Sanchez EF, Carvalho-Tavares J. Effect of Mutalysin II on vascular recanalization after thrombosis induction in the ear of the hairless mice model. *Toxicon* 2007, 50: 698-706.

Akaike H: Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csàki F, editors. *The 2nd International Symposium on Information Theory 1973*. Budapest: Akademia kiado;1973. P. 267-81.

Albuquerque PL, Jacinto CN, Silva Junior GB, Lima JB, Veras Mdo S, Daher EF. Acute kidney injury caused by *Crotalus* and *Bothrops* snake venom: a review of epidemiology, clinical manifestations and treatment. *Rev Inst Med Trop Sao Paulo* 2013, 55(5):295-301.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990, 215:403-410.

Anisimova M, Gascuel O. Approximate likelihoodratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol* 2006, 55: 539-52.

Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res* 2014, 42:D32-7.

Brasil. Ministério da Saúde. Acidentes por animais peçonhentos. Notificações registradas no sistema de agravos de notificação – SINAN. Disponível em dtr2004.saude.gov.br/sinanweb/tabnet/dh?sinannet/animaisp/bases/animaisbrnet.def acessado em julho de 2013.

Cardoso KC, Silva MJ, Costa GGL, Torres TT, Del Bem EV, Vidal RO, et al. A transcriptomic analysis of gene expression in the venom gland of the snake *Bothrops alternatus* (urutu). *BMC Genomics* 2010, 11: 605.

Casewell NR, Harrison RA, Wüster W, Wagstaff SC. Comparative venom gland transcriptoma surveys of the saw-scaled vipers (Viperidae: *Echis*) reveal substantial

intra-family gene diversity and novel venom transcripts. *BCM Genomics* 2009, 10:564.

Casewell NR, Wagstaff SC, Harrison RA, Renjifo C, Wüster W. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. *Mol Biol Evol* 2011, 28(9), 2637-49.

Chen HS, Tsai HY, Wang YM, Tsai IH. P-III hemorrhagic metalloproteinases from Russell's viper venom: Cloning, characterization, phylogenetic and functional site analyses. *Biochimie* 2008, 1486-98.

Doley R, Xingding Z, Kini RM. Snake venom phospholipase A2 enzymes. In: Mackessy, SP, editors. *Handbook of venom and toxins of reptiles*. New York: CRC Press; 2009. P. 173-206.

Du XY, Sim DS, Lee WH, Zhang Y. Blood cells as targets of snake toxin. *Blood Cells Mol Dis* 2006, 36:414-21.

Du XY & Clemetson KJ. Reptile C-Type Lectins. IN: Mackessy, SP, editors. *Handbook of venom and toxins of reptiles*. New York: CRC Press; 2009. P 358 -70.

Durban J, Juárez P, Angulo Y, Lomonte B, Flores-Diaz M, Alape-Girón A, et al. Profiling the venom gland transcriptomes of Costa Rica snakes by 454 pyrosequencing. *BMC genomics* 2011; 12:259.

Escalante T, Rucavado A, Fox JW, Gutiérrez JM. Key events in microvascular damage induced by snake venom hemorrhagic metalloproteinases. *J.prot* 2011, 74: 1781-94

Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998, 8(3):175-85.

Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 1981, 17(6):368-76.

Ferreira, S. H., Greene, L. J., Alabaster, Valerie A., Bakhle, Y. S. & Vane, J. R.. Activity of various fractions of bradykinin potentiating factor against angiotensin I converting enzyme. *Nature* 1970,225, 379-380.

Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. Pfam: the protein families database. *Nucleic Acids Res* 2014, 42: D222-30.

Fox JW, Serrano SM. Approaching the golden age of natural product pharmaceuticals from venom libraries: an overview of toxins and toxin-derivatives currently involved in therapeutic or diagnostic applications. *Curr Pharm Des* 2007, 13: 2927-34.

Fox JW, Serrano SM. Insights into and speculation about snake venom metalloproteinase (SVMP) synthesis, folding and disulfide bond formation and their contribution to venom complexity. *FEBS Journal* 2008, 275; 3016-30.

Friederick C, Tu TA. Role of metals in snake venoms for hemorrhagic, esterase and proteolytic activities. *Biochem Pharmacol* 1971, 20:1549-56.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011, 29(7):644-52.

Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003, 52: 696-704.

Gutiérrez JM, Rucavado A, Escalante T. Snake venom metalloproteinases: biological roles and participation in the pathophysiology of envenomation. In: Mackessy, SP, editors. *Handbook of venom and toxins of reptiles*. New York: CRC Press; 2009. P. 115-38.

Herrera C, Rucavado A, Warrel DA, Gutiérrez JM. Systemic effects induced by the venom of the snake *Bothrops caribbaeus* in a murine model. *Toxicon* 2013, 63:19-31.

Huang X, Madan A. CAP3: A DNA sequence assembly program. *Genome Res* 1999, 9: 868-77.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: the integrative protein signature database. *Nucleic Acids Res* 2009, 37:D211-15

Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 1992,8(3):275-82.

Junqueira-de-Azevedo ILM, Ching ATC, Carvalho E, Faria F, Nishiyama Jr MY, Ho PL, Diniz MRV. *Lachesis muta* (Viperidae) cDNAs reveal diverging pitpiver molecules and scaffolds typical of cobra (Elapidae) venoms: implications in snake toxin repertoire evolution. *Genetics* 2006, 173(2): 877-89.

Katoh K, Toh H: Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 2008, 9(4):286–98.

Kelchner SA, Thomas MA. Model use in phylogenetics: nine key questions. *Trends Ecol Evol* 2007, 22: 87-94.

Koh CY, Kini RM. From snake venom toxins to therapeutics--cardiovascular examples. *Toxicon* 2012, 59(4):497-506.

Kohlhoff M, Borges MH, Yarleque A, Cabezas C, Richardson M, Sanchez EF. Exploring the proteomes of the venoms of the Peruvian pit vipers *Bothrops atrox*, *B. barnetti* and *B. pictus*. *J Proteomics* 2012, 75: 2181-95.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012, 9(4):357-9.

Leão L, Ho PL, Junqueira-de-Azevedo ILM. Transcriptomic basis for an antiserum against *Micrurus corallines* (coral snake) *BMC Genomics* 2009, 10:112.

Leonardi A, Sajevic T, Kovačič L, Pungercar J, Lang Balija M, Halassy B, Trampuš Bakija A, Križaj I. Hemorrhagin VaH4, a covalent heterodimeric P-III metalloproteinase from *Vipera ammodytes ammodytes* with a potential antitumour activity. *Toxicon* 2014, 77:141-55.

Liò P, Goldman N. Models of molecular evolution and phylogeny. *Genome Res.* 1998, 8(12):1233-44.

Mazzi MV, Magro AJ, Amui SF, Oliveira CZ, Ticli FK, Stábeli RG, et al. Molecular characterization and phylogenetic analysis of BjussuMP-I: a RGD-P-III class hemorrhagic metalloprotease from *Bothrops jararacussu* snake venom. *J Mol Graph Model.* 2007, (1):69-85.

Moura da Silva AM, Ramos OHP, Baldo C, Niland S, Hansen U, Ventura JS, Furlan S, Butera D, Della-Casa MS, Tanjoni I, Clissa PB, Fernandes I, Chudzinski-Tavassi AM, Eble JA. Collagen binding is a key factor for the hemorrhagic activity of snake venom metalloproteinases. *Biochimie* 2008, 90: 484-92.

Moura da Silva AM, Furlan MS, Caporrino MC, Grego KF, Portes-Junior JA, Clissa PB, et al. Diversity of metalloproteinases in *Bothrops neuwiedi* snake venom transcripts: evidences for recombination between different classes of SVMPs. *BMC Genomics* 2011, 12:94.

Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 2000, 302(1):205-17.

Ogawa T, Chijiwa T, Oda-Ueda N, Ohno M. Molecular diversity and accelerated evolution of C-type lectin like proteins from snake venom. *Toxicon* 2005, 45: 1-14.

Organização Mundial de Saúde. Neglected tropical diseases. Disponível em www.who.int/neglected_diseases/diseases/snakebites/en/ acessado em setembro de 2013.

Pahari S, Mackessy SP, Kini RM. The venom gland transcriptome of the Desert Massasauga Rattlesnake (*Sistrurus catenatus edwardii*): towards an understanding of venom composition among advanced snakes (Superfamily Colubroidea). *BMC Mol Biol* 2007, 8:115.

Pidde-Queiroz G, Magnoli FC, Portaro FC, Serrano SM, Lopes AS, Paes Leme AF, et al. P-I snake venom metalloproteinase is able to activate the complement system by direct cleavage of central components of the cascade. *PLoS Negl Trop Dis*. 2013, e2519.

Posada D, Buckley TR. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst Biol* 2004, 53: 793-808.

Rambaut A: Tree Figure Drawing Tool, version 1.4.0. 2012 [tree.bio.ed.ac.uk/software/figtree]

Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* 2014, 42:D503-9.

Sajevic T, Leonardi A, Krizaj I. Haemostatically active proteins in snake venoms. *Toxicon* 2011, 57: 627-45.

Salvador GHM, Fernandes CAH, Magro AJ, Marchi-Salvador DP, Cavalcante WLG, Fernandez RM, et al. Structural and Phylogenetic Studies with MjTX-I Reveal a Multi-Oligomeric Toxin – a Novel Feature in Lys49-PLA2s Proteins Class. *Plos One* 2013, e60610.

Sanchez EF, Swenson S. Proteases from south american snake venoms affecting fibrinolysis. *Curr Pharm Analysis* 2007, 3:147-57.

Sanchez EF, Schneider FS, Yarleque A, Borges MH, Richardson M, Figueiredo SG, Evangelista KS, Eble JA. The novel metalloproteinase atroxlysin-I from Peruvian *Bothrops atrox* (Jergon) snake venom acts both on blood vessel ECM and platelets. *Arch Biochem Biophys* 2010, 496: 9-20.

Silva VX, Rodrigues MT. Taxonomic revision of the *Bothrops neuwiedi* complex (Serpentes, Viperidae) with description of a new species. *Phyllomedusa* 2008, 7: 45-90.

Soto JG, White SA, Reyes SR, Regalado R, Sanchez EE, Perez JC. Molecular evolution of PIII-SVMP and RGD disintegrin genes from the genus *Crotalus*. *Gene* 2007, 66-72.

Souza LJ, Nicolau CA, Peixoto PS, Bernardoni JL, Oliveira SS, Portes-Junior JA, et al. Comparison of phylogeny, venom composition and neutralization by antivenom in diverse species of *Bothrops* Complex. *Plos Negl trop Dis* 2013, e2442.

The UniProt Consortium. Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res* 2013, 41: D43-7.

Vaiyapuri S, Wagstaff SC, Harisson RA, Gibbins JM, Hutchinson EG. Evolutionary Gland Transcriptome of *Bitis gabonica rhinoceros*. *Plos One* 2011, 6: e21532.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009, 25:1189-91.

Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 2001, 18(5):691-9.

World Health Organization. Snake antivenoms. 2013. Disponível em www.who.int/mediacentre/factsheets/fs337/en/