

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA GERAL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA



TESE DE DOUTORADO

Transcriptoma da glândula de veneno do
escorpião *Tityus serrulatus* no contexto de
catálogos de transcritos de quelicerados

ORIENTADA: Érika Ramos de Alvarenga

ORIENTADOR: Prof. Dr. Evanguedes Kalapothakis

BELO HORIZONTE

Dezembro - 2012

Érika Ramos de Alvarenga

**TRANSCRIPTOMA DA GLÂNDULA DE VENENO DO ESCORPIÃO
Tityus serrulatus NO CONTEXTO DE CATÁLOGOS DE TRANSCRITOS
DE QUELICERADOS**

Tese apresentada ao Programa de Pós-graduação em Genética do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais como requisito parcial para obtenção do grau de Doutor em Genética

Orientador: Dr. Evanguedes Kalapothakis

Belo Horizonte
2012

Este trabalho foi realizado com o apoio das seguintes instituições:

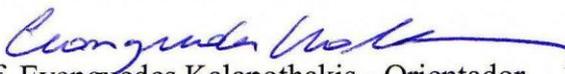
- Departamento de Biologia Geral – ICB – UFMG
- Departamento de Bioquímica e Imunologia – ICB - UFMG
- Fundação de Amparo a Pesquisa de Minas Gerais - FAPEMIG
- Conselho Nacional de Pesquisa – CNPq
- Coordenação de Aperfeiçoamento de Pessoal de Nível superior – CAPES

	<p>Pós-Graduação em Genética Departamento de Biologia Geral, ICB Universidade Federal de Minas Gerais Av. Antônio Carlos, 6627 - C.P. 486 - Pampulha - 31270-901 - Belo Horizonte - MG e-mail: pg-gen@icb.ufmg.br FAX: (+31) - 3409-2570</p>	
---	---	---

**"Transcriptoma da glândula de veneno do escorpião Tityus serrulatus
no contexto de catálogos de transcritos de quelicerados"**

Erika Ramos de Alvarenga

Tese aprovada pela banca examinadora constituída pelos Professores:


 Prof. Evangelides Kalapothakis - Orientador
 UFMG


 Rejane Maria Lira da Silva
 UFBA


 Eládio Oswaldo Flores Sanchez
 FUNED


 Liza Figueiredo Felicori Vilela
 UFMG


 Carlos Delfin Chávez Olortegui
 UFMG

Belo Horizonte, 12 de dezembro de 2012.

*Dedico essa tese a minha família, em especial, aos meus pais
(Antônio e Dalva), avós (Rosa e Líbia) e irmãos (Tiago e Dani).*

AGRADECIMENTOS

Ao meu orientador Dr. Evanguedes Kalapothakis, inicialmente pelo acolhimento e por ter propiciado uma renovação na minha experiência e um reencantamento com a Biologia. Pelos ensinamentos em Biologia Molecular, pela paciência e enorme contribuição para minha formação profissional.

Ao prof. Dr. José Miguel Ortega primeiramente pela incalculável paciência durante meu contato com a bioinformática, o que foi primordial para o desenvolvimento do trabalho. Pelas dicas, auxílios, contribuições, perguntas, discussões do trabalho e pela divertida convivência. Por ter alterado minha percepção de como processar os dados.

A toda equipe do Laboratório de Biotecnologia e Marcadores Moleculares, que direta ou indiretamente auxiliaram o desenvolvimento do trabalho ou que contribuíram para um clima harmônico e produtivo durante as atividades acadêmicas. Gostaria de agradecer, em especial, a Isabella Pena, pelo acolhimento, paciência e ensinamentos (prep., transformação) durante minha chegada ao Laboratório. À Tatiana M. Barroca pelo auxílio com os sequenciamentos. Aos ICs Bárbara F. Magalhães e Artur E. Dantas pela cooperação durante as etapas experimentais dessa tese. A Flávia S. Siqueira pelo auxílio na Bioinformática e escrita do artigo. À Carolina C. Horta pela minuciosa avaliação dos textos relacionados a essa tese. À Thais M. Mendes pela contribuição com sequenciamentos e sugestões no trabalho.

Ao estudante de doutorado em Bioinformática Rafael L. M. Guedes pela paciência e considerável contribuição nas análises de bioinformática, especialmente, com a identificação taxonômica das sequências e desenvolvimento dos scripts.

Ao pessoal do Laboratório de Genética de Microorganismos pela colaboração durante as eletroporações.

Aos membros da banca de qualificação prof. Dr. Carlos Edmundo Salas Bravo, Prof. Dr. Eduardo Martin Tarazona Santos e Dra. Liza Figueiredo Felicori Vilela, pela discussão e sugestões ao trabalho.

Aos professores e à coordenação da Pós-Graduação em Genética, aos funcionários da secretaria e aos colegas pelo auxílio, aprendizado, troca de experiências.

Às agências financiadoras CNPq, CAPES e FAPEMIG.

Aos membros que compuseram a banca examinadora, por terem aceitado o convite e pela disponibilidade na análise do trabalho.

Aos meus pais, irmãos, avós, familiares e amigos (de escolas que estudei ou dei aula, da Biologia, da Pós em Biologia Celular, dos Labs e do LAQUA) pelo incentivo e apoio durante minha formação.

À DEUS.

Por sorte a alegria da obra criada quase sempre compensa amplamente todo o estudo, a preparação, a pesquisa, o exercício, em suma, a fadiga (sempre intelectual e às vezes física) que são necessários para criá-la. A necessidade é o ponto de chegada e não de partida. E, como todo ponto de chegada, produz o prazer intenso de todas as metas atingidas. (MASI, 2005, p. 272).

SUMÁRIO

1.	INTRODUÇÃO	17
1.1	Os escorpiões	17
1.2	O escorpionismo no Brasil	18
1.3	O escorpião <i>Tityus serrulatus</i>	19
1.4	O aparato de veneno	20
1.5	Peptídeos e proteínas produzidos pela glândula de veneno de escorpiões e especificamente por <i>T. serrulatus</i>	20
1.6	Estudos de transcriptomas e ferramentas de bioinformática	23
1.6.1	<i>Banco de dados de ESTs</i>	26
1.6.2	<i>Análise de qualidade das ESTs</i>	27
1.6.3	<i>Agrupamento das sequências</i>	29
1.6.4	<i>Análise de similaridade</i>	32
1.6.5	<i>Tradução conceitual de ESTs</i>	33
1.6.6	<i>Anotação funcional</i>	34
1.7	Estudos de Transcriptoma da glândula de veneno de escorpiões	35
2.	JUSTIFICATIVA	40
3.	OBJETIVOS	42
4.	MATERIAL E MÉTODOS	44
4.1	Biblioteca de cDNA da glândula de veneno do escorpião <i>T. serrulatus</i>	44
4.2	Transformação dos plasmídeos em <i>Escherichia coli</i>	44
4.3	Extração e purificação dos plasmídeos	45
4.4	Sequenciamento dos plasmídeos	46
4.5	Análises de Bioinformática	47
4.5.1	<i>Obtenção de transcriptomas de quelicerados</i>	47
4.5.2	<i>Análise de qualidade e eliminação de contaminantes de ESTs</i>	48
4.5.3	<i>Agrupamento das sequências</i>	48
4.5.4	<i>Anotação manual das ESTs</i>	49
4.5.5	<i>Via KEGG</i>	50
4.5.6	<i>Anotação funcional usando a ferramenta Blast2GO</i>	50
4.5.7	<i>Busca de ortólogos e alinhamentos</i>	53
4.5.8	<i>Busca de ortólogos distantes: uso de PSI-BLAST e PSI-tBLASTn</i>	53
4.5.9	<i>Análise comparativa entre transcriptomas de quelicerados</i>	54
4.6	<i>Histologia da glândula de veneno</i>	55
5.	RESULTADOS	56
5.1	Sequenciamentos dos clones da biblioteca de <i>T. serrulatus</i>	56
5.2	Análise de qualidade e agrupamento das sequências de <i>T. serrulatus</i>	57
5.3	Anotação das ESTs de <i>T. serrulatus</i>	57
5.3.1	<i>Anotação manual das ESTs e unives de T. serrulatus</i>	57
5.3.2	<i>ESTs relevantes para processos celulares</i>	60
5.3.3	<i>ESTs relacionadas aos componentes do veneno</i>	63
5.3.4	<i>Anotação funcional das sequências de T. serrulatus usando a ferramenta Blast2GO</i>	65

5.4	Mapeamento dos transcritos de <i>T. serrulatus</i> usando a ferramenta KEGG PATHWAY	76
5.5	Histologia da glândula de veneno de <i>T. serrulatus</i>	76
5.6	Busca de ortólogos dos componentes do veneno encontrados no transcriptoma de <i>T. serrulatus</i>	81
5.7	Análise comparativa entre o transcriptoma de <i>T. serrulatus</i> e catálogos de transcritos de quelicerados	85
6.	DISCUSSÃO	97
6.1	O transcriptoma como ferramenta de estudo	97
6.2	ESTs relevantes para processos celulares	100
6.3	ESTs relacionadas a componentes do veneno	104
6.4	Ortólogos dos componentes do veneno de <i>T. serrulatus</i> : Inferência sobre a distribuição nos grupos de organismos	108
6.5	Análise comparativa de transcriptomas de quelicerados	114
7.	CONCLUSÕES	117
	REFERÊNCIAS BIBLIOGRÁFICAS	119
	APÊNCICE	133
	ANEXO – ARTIGO ACEITO PARA PUBLICAÇÃO	143

LISTA DE FIGURAS

Figura 1	Exemplar de <i>T. serrulatus</i>.....	19
Figura 2	Diagrama da estrutura corporal do escorpião com ênfase no télson e na glândula de veneno.	21
Figura 3	Etapas de análises de bioinformática em estudos de transcriptoma.	28
Figura 4	Distribuição das espécies de escorpião já analisadas em estudos de transcriptoma.....	39
Figura 5	Gel de Agarose 0,8% corado com GelRed demonstrando plasmídeos provenientes da biblioteca de <i>T. serrulatus</i> após amplificação, extração e purificação em placa.	56
Figura 6	Informações sobre o agrupamento e tamanho das sequências de <i>T. serrulatus</i> analisadas.	58
Figura 7	Distribuição dos transcritos sem resultado de similaridade com sequências depositadas no UniProt.	59
Figura 8	Distribuição dos transcritos em diferentes categorias de acordo com a anotação manual das sequências de <i>T. serrulatus</i>.	61
Figura 9	Distribuição dos resultados das diferentes etapas realizadas pelo programa Blast2GO.	67
Figura 10	Estatísticas do BLAST - distribuição dos <i>hits</i> obtidos do resultado do BLAST por espécie.	68
Figura 11	Estatísticas do BLAST: distribuição por espécie do melhor hit obtidos através de pesquisa BLAST.	69
Figura 12	Distribuição dos termos do GO do domínio componente celular....	70
Figura 13	Distribuição dos termos do GO do domínio processo biológico	71
Figura 14	Distribuição dos termos do GO do domínio função molecular.....	72
Figura 15	Distribuição das sequências em relação aos termos do GO do domínio componente celular nível 2.	73
Figura 16	Distribuição das sequências em relação aos termos do GO do domínio processo biológico nível 2.	74

Figura 17	Distribuição das sequências em relação aos termos do GO do domínio função molecular nível 2.	75
Figura 18	Via ribossomo fornecida pela análise das ESTs usando a ferramenta KEGG PATHWAY.	78
Figura 19	Via da Fosforilação oxidativa fornecida pela análise das ESTs usando a ferramenta KEGG PATHWAY.	79
Figura 20	Via relacionada ao processamento de proteínas no retículo endoplasmático fornecida pela análise das ESTs usando a ferramenta KEGG PATHWAY.	80
Figura 21	Características histológicas da glândula de veneno de <i>T. serrulatus</i>.	81
Figura 22	Alinhamento de sequências de NaTx e KTx	88
Figura 23	Alinhamento de Ts18 (U-BUTX-Ts1b) de <i>T. serrulatus</i> com a sequência de U1-buthitoxin-Hj1a de <i>Hottentotta judaicus</i>.....	89
Figura 24	Sequências traduzidas de AMPs e peptídeo aniônico obtidas no transcriptoma da glândula de veneno de <i>T. serrulatus</i> e os alinhamentos com ortólogos encontrados via Seed Server	89
Figura 25	Peptídeo semelhante a ponericina apresentou duas formas variantes em <i>T. serrulatus</i>.....	89
Figura 26	Alinhamento do peptídeo Pape com seus ortólogos	90
Figura 27	Sequências de hipotensinas encontradas em <i>T. serrulatus</i>.....	90
Figura 28	Sequência similar a Ts15 encontrada em <i>T. serrulatus</i>.....	90
Figura 29	Sequências de peptídeo similar a Tspept2 ou Ts12 encontrada em <i>T. serrulatus</i>	90
Figura 30	Análises de 13 transcriptomas de quelicerados quanto a distribuição dos termos do GO para os domínios componente celular, processo biológico e função molecular	96

LISTA DE TABELAS

Tabela 1	Dados epidemiológicos do escorpionismo no Brasil entre os anos de 2000 e 2011.	18
Tabela 2	Principais componetes peptídicos/protéicos descritos no veneno de <i>T. serrulatus</i>.	24
Tabela 3	Informações dos transcriptomas de glândula de veneno de escorpiões realizados até o momento.....	38
Tabela 4	Distribuição das ESTs e contigs dos componentes do veneno encontrados no transcriptoma de <i>T. serrulatus</i>.	64
Tabela 5	Vias metabólicas mais representadas na análise de ESTs e Uniques de <i>T. serrulatus</i> submetidas a ferramenta KEGG PATHWAY.	77
Tabela 6	Resultado da busca dos grupos de organismos que apresentam ortólogos de componentes do veneno de <i>T. serrulatus</i> usando Seed Server.....	82
Tabela 7	Resultados da busca dos grupos de organismos que apresentam ortólogos de componentes do veneno de <i>T. serrulatus</i> usando as ferramentas PSI-BLAST contra o UniProt e PSI-tBLASTn contra o dbEST.	91
Tabela 8	Transcriptomas amostrados demonstrando os resultados de anotação e análise comparativa entre o transcriptoma de <i>T. serrulatus</i> e outros catálogos de transcritos de quelicerados.	92
Tabela 9	Ocorrência e nível de expressão de componentes do veneno de <i>T. serrulatus</i> e seus ortólogos nos transcriptomas amostrados na análise usando a ferramenta tBLASTn.	93
Tabela 10	Ocorrência e nível de expressão de componentes do veneno de <i>T. serrulatus</i> e seus ortólogos nos transcriptomas amostrados na análise usando a ferramenta tBLASTx.	94
Tabela 11	Anotação dos transcritos que foram responsáveis por 50% da similaridade entre <i>T. serrulatus</i> e os demais transcriptomas de quelicerados.	95
Tabela 12	Componentes do veneno identificados nos transcriptomas de glândulas de veneno de escorpiões.....	110

LISTA DE QUADROS

Quadro 1	Organograma obtido pelo Taxonomy do NCBI para o grupo dos quelicerados.	49
Quadro 2	Regra de anotação (RA) do programa Blast2GO	52

LISTA DE ABREVIATURAS

AMP = *antimicrobial peptide* (peptídeo antimicrobiano)
 BLAST = *Basic Local Alignment Search Tool*
 BPP = bradykinin-potentiating peptide (peptide potencializador de bradicinina)
 Cr = foi usado algumas vezes como abreviatura da espécie *Carcinoscorpius rotundicauda*
 De = foi usado algumas vezes como abreviatura da espécie de aranha *Dysdera erythrina*
 DNA = Ácido dexossirribonucléico
 EDTA = Ácido Etileno Diaminotetracético
 EMBL = *European Molecular Biology Laboratory*
 Es = foi usado algumas vezes como abreviatura da espécie de aranha *Endeis spinosa*
 EST = *Expressed Sequence Tag* (etiqueta de sequencia expressa)
 EST/100K = ESTs por 100.000 ESTs
 GO = *Gene Ontology*
 Hp = foi usado algumas vezes como abreviatura da espécie de escorpião *Heterometrus petersii*
 Ir = foi usado algumas vezes como abreviatura da espécie de carrapato *Ixodes ricinus*
 Kb = Quilobase
 KEGG = Kyoto Encyclopedia of Genes and Genomes
 KTx = toxinas de canais para potássio
 LB = Meio de Cultura Luria-Bertani
 Ll = foi usado algumas vezes como abreviatura da espécie de aranha *Loxocoles laeta*
 Lm = foi usado algumas vezes como abreviatura da espécie de escorpião *Lynchus mucronatus*
 Lp = foi usado algumas vezes como abreviatura da espécie *Limulus polyphemus*
 M = Molar
 Mg = foi usado algumas vezes como abreviatura da espécie de escorpião *Mesobuthus gibbosus*
 mRNA = RNA mensageiro
 NaTx = toxinas de canais para sódio
 NCBI = *National Center for Biotechnology Information*
 NDBP = Peptídeo sem ponte dissulfeto
 Ne = foi usado algumas vezes como abreviatura da espécie de aranha *Nephila antipodiana*
 ng = Nanograma
 pb = Pares de bases
 PCR = Reação em cadeia da polimerase
 PHRAP = *PHRagment Assembly Program*
 PHRED = *Phil's Read Editor*
 RNA = Ácido ribonucléico
 Sj = foi usado algumas vezes como abreviatura da espécie de escorpião *Scorpiops jendeki*
 Ss = foi usado algumas vezes como abreviatura da espécie de ácaro *Sarcoptes scabiei*
 Swiss prot = Protein knowledgebase of Swiss Institute
 Taq = *Thermus aquaticus*
 Taq DNA polimerase = DNA polimerase termoestável do *Thermus aquaticus*
 TrEMBL = *Translated EMBL*
 Tris = Tris (hidroximetil) aminometano
 TsNTxP = *Tityus serrulatus non-toxic protein*
 UniProt = *The Universal Protein Resource*
 U.V. = Ultra Violeta
 V = Volts

RESUMO

O escorpião *Tityus serrulatus* pertence à família Buthidae e é a espécie que causa o maior número de acidentes com humanos no Brasil, sendo responsável por vários casos de envenenamento. No presente estudo, a análise do transcriptoma da glândula de veneno ativa desse escorpião foi realizada com 1629 ESTs e os resultados foram comparados com catálogos de transcritos de outros doze quelicerados para avaliar o padrão de expressão da glândula de veneno de escorpiões. Além de transcritos relacionados à manutenção celular, especialmente envolvidos com a síntese proteica, produção de energia e de função estrutural, foram encontradas 35 famílias de peptídeos ou proteínas que compõem o veneno na biblioteca de cDNA de *T. serrulatus*. Pelo menos três novas sequências completas de toxinas e mais de uma dezena de prováveis novos componentes do veneno foram identificados. As glândulas de veneno apresentaram um distinto padrão de distribuição dos termos do Gene Ontology (GO) quando comparadas com o padrão apresentado por catálogos de transcritos provenientes de animais inteiros ou de outros tipos de glândulas, devido à predominância de termos do GO como “extracellular region”, “response to stimulus” e “channel regulators”. A transcriptômica comparativa possibilitou uma visão mais ampla da distribuição de componentes do veneno em diferentes famílias de escorpiões e revelou aqueles que parecem ser amplamente distribuídos nesse grupo, aqueles que são específicos a uma determinada família e ainda os que foram encontrados apenas no gênero *Tityus*. Em conclusão, esse estudo representou mais um passo para a compreensão da glândula de veneno de *T. serrulatus* e apresentou alguns peptídeos bioativos que podem ser explorados para aplicações biotecnológicas. Além disso, a análise de transcriptômica comparativa revelou um padrão particular de expressão das glândulas de veneno e a distribuição das principais famílias de peptídeos/proteínas do veneno em diferentes clados.

Palavras-chave: Escorpião amarelo. Componentes do veneno, Padrão de expressão gênica. Glândula de veneno.

ABSTRACT

The scorpion *Tityus serrulatus* belongs to the family Buthidae and is considered to be the most dangerous scorpion in Brazil, being responsible for several cases of envenomation. In this study, we performed a venom gland transcriptome analysis (1,629 ESTs), and compared it with transcript catalogues of 12 other Chelicerates to characterize the expression profile of scorpion venom glands. Beyond housekeeping transcripts specially related to protein synthesis, energy supply and structural functions, we found 35 different families of venom peptides or proteins in the cDNA library of *T. serrulatus* venom gland. These transcripts included three new complete sequences of toxins and more than a dozen putative venom gland proteins/peptides. The venom glands had a clearly different distribution of Gene Ontology (GO) terms when compared to other types of glands or whole animal transcriptomes, due the predominance of GO terms such as “extracellular region”, “response to stimulus” and “channel regulators”. The comparative transcriptomic revealed that some venom components were widely distributed throughout the Scorpiones order, others were exclusives to the Buthidae family, and some components were only identified in the *Tityus* genus. This study represented an advance towards understanding the *T. serrulatus* venom gland, and presented some potential bioactive peptides that could be exploited for biotechnological application. In addition, the comparative transcriptomic analysis revealed a particular expression profile for venom glands and the distribution of the main venom component families in different clades.

Keywords: Brazilian yellow scorpion. Venom components. Gene expression profile. Venom gland.

1. INTRODUÇÃO

1.1 Os escorpiões

Evidências da relação dos seres humanos com escorpiões datam desde a antiguidade com a inserção desses animais em mitos provenientes de diferentes culturas como a Egípcia, a Grega e a Romana (Cloudsley-Thompson, 1990). O interesse inicial por esses artrópodes foi possivelmente gerado pelo fato de que algumas poucas espécies possuem toxinas capazes de causar a morte em humanos (Polis, 1990). Atualmente, os escorpiões despertam o interesse humano por diversas razões, podendo-se citar a sua importância médica, sua intrigante biologia e evolução (Polis, 1990) e a utilização do seu veneno como uma considerável fonte de proteínas bioativas de potencial interesse biotecnológico e farmacológico (Rodríguez De La Vega *et al.*, 2010).

Os primeiros escorpiões surgiram na Terra há cerca de 450 milhões de anos. Apesar do grande tempo de existência, esses animais apresentam morfologia conservada, de modo que a diversificação taxonômica desse grupo não foi acompanhada por nenhuma grande revolução na arquitetura corporal. Esses animais apresentam ampla distribuição geográfica, estando ausente apenas na Antártida. Acredita-se que o grande sucesso desses artrópodes está relacionado à plasticidade e adaptabilidade da sua história de vida, ecologia, fisiologia e comportamento (Polis, 1990).

Os escorpiões são artrópodes venenosos da classe Arachnida, ordem Scorpiones. Essa ordem conta com cerca de 1.500 espécies atuais classificadas dentro de 18 famílias: Bothriuridae, Buthidae, Chactidae, Chaerilidae, Diplocentridae, Euscorpiidae, Hemiscorpiidae, Heteroscorpionidae, Iuridae, Liochelidae, Microcharmidae, Pseudochactidae, Scorpionidae, Scorpipidae, Superstitioniidae, Troglotayosocidae, Urodacidae e Vaejovidae (Prendini e Wheeler, 2005). Dentre essas, a família Buthidae é a de maior importância médica, possuindo mais de 500 espécies distribuídas pelo mundo.

Apesar de todos os escorpiões terem glândula de veneno, cerca de 30 espécies representam potencial risco para humanos. Exceto por *Hemiscorpius lepturus* que pertence a família Hemiscorpiidae, todas essas espécies são Buthideos dos seguintes gêneros: *Androctonus*, *Buthus*, *Mesobuthus*, *Hottentotta*, *Parabuthus* e *Leiurus* localizados no norte da África, Ásia, Oriente Médio e Índia; *Centruroides*, encontrados no sul dos Estados Unidos, México e Colômbia; *Tityus* presente na América do Sul (Chippaux e Goyffon, 2008). A severidade do envenenamento causado por esses animais varia com a espécie de escorpião, a

idade e o tamanho da vítima, sendo muito mais grave em crianças. A intoxicação sistêmica é o reflexo da superestimulação do Sistema Nervoso Central, do Sistema Nervoso Simpático e Parasimpático e sua severidade varia desde a dor local até a cardiotoxicidade fatal e encefalopatia (Amitai, 1998)

1.2 O escorpionismo no Brasil

No Brasil, o escorpionismo, definido como o envenenamento provocado pela inoculação de veneno através do ferrão dos escorpiões, é um problema de saúde pública devido à elevada incidência em várias regiões, além da letalidade em alguns casos. Dados epidemiológicos de acidentes escorpiônicos no Brasil entre 2000 e 2010 são mostrados na Tabela 1. De acordo com o Ministério da Saúde, foram registrados 50.126 acidentes e 88 óbitos em 2010, sendo que a incidência do escorpionismo foi de 26,3 acidentes/100.000 habitantes, com uma letalidade de 0,2%. Essa porcentagem de letalidade dos casos no Brasil é semelhante a média de letalidade mundial causada por acidentes com escorpiões, que é de 0,27% (Chippaux e Goyffon, 2008) A incidência dos registros é maior em indivíduos entre 20 e 49 anos. Crianças entre 1 e 9 anos estão na faixa etária com maior quantidade de óbitos (39/88).

Tabela 1: Dados epidemiológicos do escorpionismo no Brasil entre os anos de 2000 e 2011. De maneira geral, o número de óbitos, o número de casos e a incidência dos mesmos aumentaram ao longo dos anos.

Critério	Região	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Óbitos por acidentes por escorpiões	Norte	1	1	3	2	1	3	5	5	6	6	6
	Nordeste	13	21	25	15	16	12	13	33	47	32	46
	Sudeste	1	19	23	31	22	29	9	25	32	54	32
	Sul	0	0	0	0	0	0	0	0	0	0	1
	Centro-oeste	1	2	6	3	3	4	1	3	2	5	3
	Brasil	16	43	57	51	42	48	28	66	87	97	88
Incidência de casos (100.000 hab)	Norte	3,2	5,1	7,0	10,0	11,4	13,0	13,6	13,1	15,3	17,9	15,4
	Nordeste	16,3	17,6	21,3	21,2	26,3	31,6	36,9	37,8	36,3	43,7	45,1
	Sudeste	5,2	10,7	13,6	15,0	17,7	20,2	18,1	17,3	20,7	24,8	25,6
	Sul	0,3	1,5	1,7	1,8	2,1	2,5	2,8	3,5	3,7	3,6	4,1
	Centro-oeste	5,3	6,2	7,4	7,1	9,9	10,5	10,2	8,4	10,9	14,0	14,7
	Brasil	7,5	10,6	13,0	13,8	16,8	19,5	20,1	20,0	21,5	25,7	26,3
Casos de acidentes	Norte	418	680	947	1.380	1.603	1.917	2.043	2.015	2.318	2.742	2.444
	Nordeste	7.802	8.515	10.402	10.478	13.132	16.142	19.051	19.746	19.276	23.416	23.929
	Sudeste	3.791	7.887	10.093	11.276	13.546	15.836	14.416	13.921	16.599	20.102	20.572
	Sul	81	373	431	479	540	677	765	980	1.010	1.011	1.114
	Centro-oeste	612	733	891	873	1.245	1.366	1.349	1.136	1.490	1.941	2.067
	Brasil	12.704	18.188	22.764	24.486	30.066	35.938	37.624	37.798	40.693	49.212	50.126

Fonte: Tabela adaptada a partir dos dados fornecidos pelo site do Ministério da Saúde, 2012, disponível em <http://portal.saude.gov.br/portal/saude/profissional/area.cfm?id_area=1537>.

No país, espécies de escorpiões do gênero *Tityus* apresentam importância médica, sendo que as principais causadoras de acidentes em humanos são: *T. serrulatus* (escorpião

amarelo), *T. bahiensis* (escorpião marrom), *T. stigmurus*, *T. obscurus* (*T. paraensis*) (escorpião-preto da amazônia) e *T. metuendus* (Brasil, MINISTÉRIO DA SAÚDE, 2012). Dentre essas, a espécie *T. serrulatus*, é a responsável pela maioria dos casos mais graves no Brasil com registro de óbitos, principalmente em crianças (Cupo *et al.*, 2003).

1.3 O escorpião *Tityus serrulatus*

De coloração amarela clara, o que originou o seu nome popular escorpião amarelo, indivíduos da espécie *T. serrulatus* (Fig. 1) medem até 7cm de comprimento (como revisado por Cologna *et al.*, 2009). A grande maioria dos estudos indica que suas populações são formadas por fêmeas, cuja reprodução se dá por partenogênese (Lourenço, 2000). Entretanto, existe relato na literatura de uma população sexual com taxa de 1 macho para 2,25 fêmeas (Lourenço e Cloudsley-Thompson, 1999). Cada fêmea de *T. serrulatus* pode ter três, quatro ou mais parições e cerca de 70 filhotes durante a vida (Matthiensen, 1971).

Originalmente, *T. serrulatus* ocupava uma região muito restrita no Estado de Minas Gerais. Entretanto, essa espécie tem se espalhado pela região do Sudeste do Brasil e nos estados do Pará, Goiás e Rio Grande do Sul (Lourenço, 2002; Torres *et al.*, 2002; Bortoluzzi *et al.*, 2007). Essa expansão tem sido atribuída à boa adaptação desses animais ao ambiente urbano, além da sua rápida e intensa proliferação, (Lourenço *et al.*, 1996). Em Belo Horizonte, a ocorrência de *T. serrulatus* e os acidentes por eles provocados representam uma situação preocupante. A ocorrência tem sido relacionada a fatores inerentes ao solo, como áreas de pedreiras, barrancos, marmorarias e também a construções antigas, serrarias e olarias (Soares *et al.*, 2002).



Figura 1: Exemplar de *T. serrulatus*. Fonte: arquivo pessoal.

1.4 O aparato de veneno

Os escorpiões contam com um aparato de veneno localizado na extremidade posterior do corpo do animal (télson), que apresenta também um acúleo utilizado para a injeção do veneno (Fig. 2). Abaixo da cutícula (exoesqueleto quitinoso) encontram-se duas glândulas de veneno, que são envolvidas por camadas musculares na porção mediana (Polis, 1990; Jarrar e Al-Rowaily, 2008; Yigit e Benli, 2008).

As glândulas de veneno dos escorpiões atuam basicamente na captura de presas e na defesa contra outros animais. Escorpiões se alimentam de insetos, aranhas e outros invertebrados pequenos. Conseqüentemente, a glândula de veneno dos escorpiões produz toxinas que efetivamente paralisam esse tipo de presa. O veneno também apresenta função de defesa, e é altamente letal para vertebrados, incluindo mamíferos (Kozminsky-Atias *et al.*, 2008). Sua composição constitui-se de uma complexa diversidade de peptídeos/proteínas sintetizados pela glândula de venenos dos escorpiões, além de nucleotídeos, lipídeos e aminas biogênicas (Almeida *et al.*, 2002, Batista *et al.*, 2006, Rodríguez de La Vega *et al.*, 2010).

Muitas questões básicas sobre a fisiologia e biologia celular desse aparato de veneno ainda permanecem abertas. A caracterização bioquímica e identificação funcional de proteínas ainda não caracterizadas pode certamente contribuir para o entendimento de processos celulares que permitam que esse órgão altamente especializado produza e secrete a variedade de componentes que constituem o veneno (Rodríguez de La Vega *et al.*, 2009). É importante salientar que a maioria dos estudos encontrados na literatura trabalha com componentes específicos do veneno. E mesmo os estudos de transcriptoma desse órgão enfocam quase exclusivamente nos constituintes tóxicos. Assim, ressalta-se a necessidade de se obter uma visão panorâmica da glândula a nível molecular, através de identificação de transcritos importantes para a função dos diferentes tecidos que a compõem e da descrição de sequências de aminoácidos de proteínas que participam da síntese e composição do veneno. Essas informações podem contribuir para a melhor compreensão dos eventos fisiológicos que ocorrem na glândula de veneno. Abaixo serão descritos os componentes moleculares comumente descritos na glândula de veneno de escorpiões.

1.5 Peptídeos e proteínas produzidos pela glândula de veneno de escorpiões e especificamente por *T. serrulatus*

Transcriptomas de glândula de veneno de escorpiões geralmente classificam os componentes produzidos nesse órgão em dois grandes grupos: (1) toxinas e demais

constituintes do veneno e, (2) componentes celulares com destaque a maquinaria proteica capaz de realizar a síntese e o processamento dos componentes do veneno (Schwartz *et al.*, 2007; Ma *et al.*, 2009; Rodríguez de La Vega *et al.*, 2010).

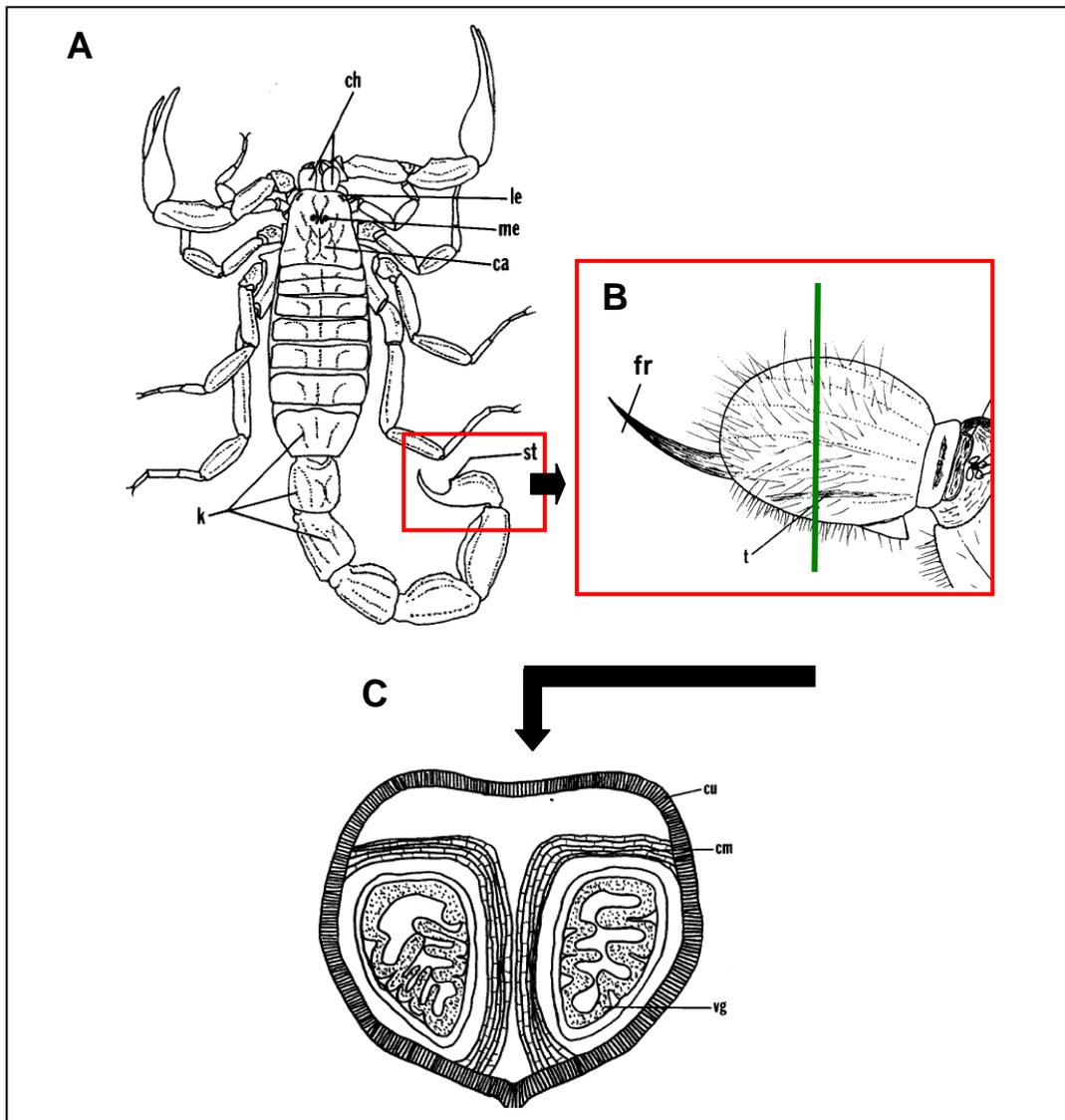


Figura 2: Diagrama da estrutura corporal do escorpião com ênfase no télson e na glândula de veneno. A) Estrutura corporal do escorpião com destaque ao télson (quadro); ch = quelíceras; le= olhos laterais ; me = olhos medianos; ca = carapaça; st = tubérculo subaculear; B) Detalhe do télson, ferrão (fr), região bulbar (t). O segmento verde indica a região de corte transversal para se obter a imagem da figura C. C) corte transversal do télson demonstrando de fora para dentro cutícula (cu), músculo compressor da glândula (cm) e a glândula de veneno (vg). Figura modificada a partir de Gary A. Polis (1990).

Em relação aos constituintes do veneno secretados pela glândula, as neurotoxinas merecem destaque. Elas são constituídas por um conjunto de pequenos peptídeos, empacotados por 2-5 pontes dissulfeto, que interagem especificamente com vários canais

iônicos (Na^+ , K^+ , Ca^{2+} e Cl^-). Tipicamente elas contêm uma assinatura estrutural definida pela presença de motivo α/β estabilizado por cisteína. Dentre as neurotoxinas, aquelas que atuam em canal para Na^+ (as NaTxS) são as mais reativas e as principais responsáveis pelos efeitos tóxicos do envenenamento por escorpiões (Gordon *et al.*, 2007). Elas são consideradas peptídeos de cadeia longa que podem ser classificadas basicamente em neurotoxinas de escorpião α e β , baseando-se em seu modo de ação e propriedade de ligação ao canal (Rodríguez de La Vega *et al.*, 2005). Toxinas de canais para K^+ (KTxs), por sua vez, constituem o grupo mais diverso de neurotoxinas, sendo formado por pequenos peptídeos de cadeia curta (23-64 resíduos de aminoácidos) contendo 2 a 4 pontes dissulfeto (Rodríguez de La Vega *et al.*, 2004, Camargos *et al.*, 2011). Embora sejam encontrados em pequena frequência, toxinas de canais para Ca^{2+} e Cl^- têm sido também obtidas do veneno de escorpiões (Zeng *et al.*, 2005; Zhijian *et al.*, 2006; Norton e McDonough, 2008).

Além de toxinas, a glândula de veneno dos escorpiões secreta enzimas e peptídeos pequenos e sem ponte dissulfeto com grande amplitude de atividade biológica, tais como a de potencializador da bradicinina, antimicrobiana, hemolítica, modulação imunológica (Pimenta e De Lima, 2005; Zeng *et al.*, 2005; Cologna *et al.*, 2009), atividades semelhantes a hormônios (Verano-Braga *et al.*, 2008) e de iniciação da sinalização celular (Zhijian *et al.*, 2006). Algumas dessas proteínas secretadas têm a função de potencializar os efeitos do veneno. Hialuronidases, por exemplo, encontradas no veneno de vários escorpiões, são consideradas fatores que facilitam a difusão das toxinas no tecido da vítima, uma vez que catalizam a quebra da matriz extracelular (Fry *et al.*, 2009).

Apesar de pouca atenção ter sido direcionada aos componentes da glândula com papéis estruturais nesse órgão e de síntese e processamento dos constituintes do veneno, alguns estudos de transcriptoma da glândula de veneno de escorpiões (Schwartz *et al.*, 2007; Ma *et al.*, 2009; Silva *et al.*, 2009) têm identificado transcritos envolvidos com essas funções. Nesses trabalhos, os autores destacam a presença de ESTs (*expressed sequence tags*) relacionadas com a transcrição e metabolismo proteico como proteínas ribossomais, com o processamento, endereçamento de proteínas e transporte celular e com a geração de energia como a presença de transcritos envolvidos no ciclo do ácido tricarboxílico e com fosforilação oxidativa (Schwartz *et al.*, 2007; Ma *et al.*, 2009; Silva *et al.*, 2009).

Similar a outros escorpiões da família Buthidae, o veneno de *T. serrulatus* é constituído principalmente por uma variedade de peptídeos neurotóxicos (especialmente toxinas de canais para Na^+ e K^+) que interagem especificamente com vários canais iônicos, peptídeos sem pontes dissulfeto como peptídeo potencializador de bradicinina (peptídeo T),

hipotensinas e proteínas como proteases, alergenicos e hialuronidasas (Rates *et al.*, 2008; Verano-Braga *et al.* 2008; Cologna *et al.*, 2009; Fletcher *et al.*, 2010). Os principais componentes do veneno identificados até o momento na glândula do escorpião *T. serrulatus* são apresentados na Tabela 2, que foi construída baseando-se no banco de dados *Animal Toxin Annotation Program* do Uniprot (Jungo *et al.*, 2012), trabalho de Cologna *et al.* (2009) e outros trabalhos da literatura. Deve-se destacar que parte dos componentes descritos são fragmentos protéicos e análise do transcriptoma da glândula de veneno dessa espécie pode contribuir com a identificação da sequência completa do peptídeo ou proteína. Devido à falta de uma padronização da nomenclatura de toxinas, observa-se uma grande confusão de nomes e símbolos, e muitas vezes, diferentes autores nomeiam toxinas distintas com o mesmo nome, ou toxinas idênticas com nomes diferentes (Cologna *et al.*, 2009). Assim, para facilitar a leitura desta tese e padronizar a nomenclatura usada, a Tabela 2 oferece os nomes das toxinas de *T. serrulatus* como proposto por Cologna *et al.* (2009) e por King *et al.* (2008) (nome entre parêntese na coluna 1) e o identificador do Uniprot (entre colchetes na coluna 1) como referência da sequência proteica.

Apesar dos dados existentes sobre os componentes do veneno do escorpião amarelo, poucas informações são disponíveis acerca de formas variantes dos transcritos que codificam esses componentes. Além disso, informações sobre peptídeos que apresentam funções estruturais, de síntese e de processamento dos peptídeos que irão constituir o veneno são escassas. Assim, a análise do transcriptoma da glândula de *T. serrulatus* poderá permitir a descoberta de novos componentes do veneno, além de possibilitar a obtenção de dados que permitam realizar uma reconstrução molecular desse órgão.

1.6 Estudos de transcriptomas e ferramentas de bioinformática

Etiquetas de sequências expressas (*Expressed sequence tag* - ESTs) são sequências curtas (geralmente de 100-800 pb de comprimento), não editadas, resultantes de apenas um sequenciamento e randomicamente selecionadas a partir de uma biblioteca de cDNA. A EST podem ser gerada a partir da extremidade 3' ou 5' do cDNA de um clone da biblioteca. Grande quantidade dessas etiquetas pode ser gerada com um custo relativamente baixo para permitir uma visão das regiões transcricionalmente ativas em qualquer organismo (Nagaraj *et al.*, 2006).

Tabela 2: Principais componetes peptídicos/protéicos descritos no veneno de *T. serrulatus* com base nos trabalhos de Cologna *et al.*, 2009 e no UniProt com suas respectivas funções.

Continua

Nome e ID Uniprot	Outros nomes	Informações a respeito da toxina
Neurotoxinas de canais para Na⁺		
Ts1 βNaTx (β-BUTX-Ts1) [P15226]	TsTX-I; toxin γ; Toxin T2-IV; Toxin Ts7; Ts VII; ToxinVII; TsTX-VII; Tityustoxin VII; Toxin III-10; Toxin II-11	Toxina beta que se liga ao sítio-4 do canal para Na ⁺ e altera a ativação para potenciais mais negativos. Toxina ativa contra insetos e mamíferos (Possani <i>et al.</i> , 1981; Bechis <i>et al.</i> , 1984; Sampaio <i>et al.</i> , 1991, Polikarpov <i>et al.</i> , 1999).
Ts2 αNaTx (δ -BUTX-Ts1a) [P68410]	TsTX-III; III-8; Tst2; Toxin II; TsTX-II; Tityustoxin II; Toxin T1-IV; beta toxin TsII	Apresenta algumas características de alfa-toxina, e outras de beta-toxinas (Cologna <i>et al.</i> , 2009).. Em ensaios de ligação ao receptor se comportou como Ts1 (Mansuelle <i>et al.</i> , 1992). Entretanto, semelhante as alfa-toxinas, prolonga a duração do potencial de ação (Sampaio <i>et al.</i> , 1991, Cologna <i>et al.</i> , 2012).
Ts3 α NaTx (δ-BUTX-Ts1b) [P01496]	TsTX; Tityustoxin; Toxin IV-5 TsIV-5; Ts IV alphatoxin; Ts IV; Toxin-4; Tityustoxin IV; Toxin IV; Ts III; Toxin-3;	Liga-se ao sítio-3 do canal para Na ⁺ dependente de voltagem e inibe a inativação de canais ativados, bloqueando a transmissão neuronal. Ativa contra mamíferos (Kirsch <i>et al.</i> , 1989; Martin-Eauclaire <i>et al.</i> , 1994; Campos <i>et al.</i> , 2004; Corona <i>et al.</i> , 1996)
Ts5 α NaTx (δ-BUTX-Ts1c) [P46115]	TsTX-V; Tityustoxin-5;Tityustoxin V; alpha toxin TsTX-V; Ts V; Toxin V	Liga-se ao sítio-3 do canal para Na ⁺ dependente de voltagem e inibe a inativação de canais ativados (Arantes <i>et al.</i> , 1994). Também afeta indiretamente os canais para K ⁺ dependentes de voltagem das células-beta, aumentando a permeabilidade a K ⁺ (Marangoni <i>et al.</i> , 1995).
Neurotoxinas de canais para K⁺		
Ts6 (λ-BUTX-Ts1b) [P59936]	TsTX-IV; α-KTx12.1; Butantoxin; BuTX; potassium channel toxin	Inibe alta condutância de canais para K ⁺ ativados por Ca ²⁺ (Novello <i>et al.</i> , 1999). Inibe fracamente canais para K ⁺ do tipo <i>Shaker B</i> .(revisado por Cologna <i>et al.</i> , 2009). Inibe a ligação de 125I-Kaliotoxina ao sítio receptor em sinaptossoma, indicando afinidade por K _v 1.3K ⁺ (Pimenta <i>et al.</i> , 2003a)
Ts7 (λ-BUTX-Ts1a) [P46114]	TstX-Kα; Tityustoxin Kalpha; Potassium channel toxin alfa-KTx 4.1; TsII-9; TsTx-K-alpha; TSK4; Toxin II-9	Bloqueia o fluxo de ⁸⁶ Rb pelos canais de K ⁺ do tipo retificador tardio não inativantes em sinaptossomos sem alterar o fluxo pelos canais de K ⁺ inativantes (Rogowski <i>et al.</i> , 1994). Bloqueia corrente de K ⁺ em K _v 1.2 (Werkman <i>et al.</i> , 1993) e K _v 1.3 (Rodrigues <i>et al.</i> , 2003).
Ts8 (κ-BUTX-ts1) [P69940]	Tityustoxin K-beta; TSK2; TsTX-K beta; TsTx-Kβ;	Seletivamente bloqueia os canais para K ⁺ não inativantes nos sinaptossomas (Rogowski <i>et al.</i> , 1994; Cologna <i>et al.</i> , 2009).
Ts9 (λ-BUTX-Ts1c) [P56219]	Ts κ; Ts Kappa; Neurotoxin Ts-kappa; TsKapa; 1TSK;_alfa-KTx 4.2	Bloqueador dos canais para K ⁺ ativados por Ca ²⁺ de baixa condutância (Legros <i>et al.</i> , 1996; Blanc <i>et al.</i> , 1997).
Ts15 (κ-BUTX-ts2) [P86270]	Tityustoxin-15	Bloqueia canais para K ⁺ ativados por voltagem, especialmente K _v 1.2 e K _v 1.3(Cologna <i>et al.</i> , 2011).
Ts16 (κ-BUTX-ts3) [P86271]	Tityustoxin-16	Apresenta similaridade a toxinas de canal para K ⁺ , entretanto, estudos fisiológicos não foram realizados.

Nome e ID Uniprot	Outros nomes	Informações a respeito da toxina
Outros peptídeos do veneno		
Ts4 [O77463]	TsTX-VI; Tityustoxin-6; TsVI NTxp, TsnTxp; Tityustoxin VI; TsTXVI; Toxin VI; NTxP TsNTxP; Non-toxic protein;	Considerada não tóxica para camundongo porque é incapaz de induzir os sintomas característicos da toxicidade produzida por outras toxinas. Causa reação alérgica, lacrimação, espasmos nas patas posteriores de camundongos e liberação dose-dependente de neurotransmissores de sinaptossomas (Marangoni <i>et al.</i> , 1990; Chavez-Olortegui <i>et al.</i> , 1997; Guatimosin <i>et al.</i> , 1999)
Ts10 [Q9TWR4]	Peptide T	Peptídeo potencializador da bradicinina: são capazes de potencializar a atividade da bradicinina; inibe a hidrólise da bradicinina pela enzima conversora de angiotensina (ACE) e a conversão da angiotensina I para angiotensina II pela kininase II (Ferreira <i>et al.</i> , 1993; Cologna <i>et al.</i> , 2009).
Ts11 [P0C174]	TsPep1; Peptide TsPep1	Função biológica desconhecida. Apresenta alta similaridade com a folha β C-terminal de algumas toxinas bem caracterizadas para canais para K^+ (Pimenta <i>et al.</i> , 2003b)
Ts12 [P0C175]	TsPep2; Peptide TsPep2	Função biológica desconhecida. Apresenta alta similaridade com a folha β C-terminal de algumas toxinas bem caracterizadas para canais para K^+ (Pimenta <i>et al.</i> , 2003b)
Ts13 [P0C176]	TsPep3; Peptide TsPep3	Função biológica desconhecida. Apresenta alta similaridade com a folha β C-terminal de algumas toxinas bem caracterizadas para canais para K^+ (Pimenta <i>et al.</i> , 2003b)
Ts14	Hypotensin-1 TsHpt-I [P84189]	Peptídeo linear enrolado randomicamente e com assinatura de aminoácido similar ao BPP. São capazes de potencializar o efeito hipotensivo da bradicinina e induzir vaso-relaxamento dependente da liberação de NO, embora não possa inibir a atividade da ACE (Verano-Braga <i>et al.</i> , 2008).
	Hypotensin-2 TsHpt-II [P84190]	Alta similaridade a TsHpt-I (Verano-Braga <i>et al.</i> , 2008)
Peptídeo Pape [P86821]		Similar a peptídeo potencializador de bradicinina encontrados em outros escorpiões (Rates <i>et al.</i> , 2008).
Hialuronidase [P8584]		Realiza a quebra do ácido hialurônico, um polissacarídeo presente na matriz extracelular do tecido conjuntivo. Hialuronidases no veneno estão relacionadas com a difusão de toxinas no tecido da vítima da picada. Pertence a família CRISP. (Pessini <i>et al.</i> , 2001)
Antarease [P86392]		Age como metaloprotease e cliva proteínas SNARE no tecido pancreático. Pode estar relacionada com pancreatite desenvolvida durante o envenenamento. (Fletcher <i>et al.</i> , 2010)
Metaloprotease [P85842]		Degradam proteínas da matriz extracelular. (Fonte: anotação do UniProt).
Alergeno [P85840]		Pertence a família CRISP, causa reação alérgica em humanos (Fonte: anotação do UniProt).
Alfa-amilase [P85843]		Realiza a endohidrólise de ligações (1->4)-alfa-D-glucosídicas em polissacarídeos contendo 3 ou mais unidades de D-glicose (1->4)-alfa (Fonte: anotação do UniProt).

O primeiro estudo de transcriptoma ocorreu em 1991, no qual esses transcritos foram utilizados como fonte de descoberta para genes humanos (Adams *et al.*, 1991). A partir daí, ocorreu um crescimento exponencial na geração e acumulação de ESTs em banco de dados públicos para diversos organismos. ESTs são versáteis e possuem múltiplas aplicações podendo auxiliar a construção de mapas do genoma (Wilcox *et al.*, 1991), a descoberta de genes (Mondego *et al.*, 2011) e a identificação de sua estrutura (Jiang *et al.*, 1998), o estabelecimento da variabilidade de transcritos alternativos (Modrek *et al.*, 2001), a caracterização de candidatos a polimorfismos de nucleotídeo único (SNPs) (Useche *et al.*, 2001), a identificação de marcadores moleculares específicos para determinados tecidos ou doença (Herbert *et al.*, 2008, Bernstein *et al.*, 2009, Romanuik *et al.*, 2009), a complementação de anotações de genomas (Rudd, 2003) e de análise de proteomas (Ma *et al.*, 2010).

Apesar das diversas possibilidades de estudos proporcionadas pelo transcriptoma, a identificação de genes codificados a partir de ESTs apresenta uma série de desafios. ESTs representam um grande número de sequências redundantes, tipicamente parciais, algumas podendo apresentar baixa qualidade, quimerismo e moderada taxa de contaminação por vetores e adaptadores (Perteau *et al.*, 2003). Para solucionar tais artefatos, os estudos de transcriptoma geralmente utilizam ferramentas de bioinformática. Existem vários passos na análise de ESTs e um grande número de ferramentas para cada um deles. Cada um desses métodos possui determinadas vantagens e buscam extrair informações biológicas a partir da sequência dos transcritos. Entretanto, existe muita confusão na escolha das ferramentas mais adequadas para cada um dos diferentes passos de análise e a subsequente anotação em nível de proteína ou DNA (Nagaraj *et al.*, 2006). Aqui serão revisadas as principais formas de análise de transcriptomas focando nos estudos que analisaram glândulas de veneno. Esses trabalhos geralmente seguem determinados passos de análise que estão ilustrados na Figura 3. Para realizar essas etapas, são utilizados diversos programas de análise bioinformática. Os programas comumente usados em cada uma dessas etapas, muitos deles também utilizados neste trabalho, serão descritos abaixo.

1.6.1 Banco de dados de ESTs

O maior depósito disponível de EST é o dbEST do NCBI (Boguski *et al.*, 1993). O artigo que descreve esse banco data de 1993, e nele é descrita a ocorrência de 22.537

ESTs. Atualmente, apresenta mais de 73 milhões de entradas provenientes de mais de 2.200 organismos (número obtido em outubro de 2012 disponível em <http://www.ncbi.nlm.nih.gov/projects/dbEST/>). O dbEST disponibiliza as sequências para download e para estudos acadêmicos.

Outro banco de dados de ESTs é o UniGene do NCBI que estoca genes únicos e representa uma série de grupos orientados de genes gerados a partir de ESTs. Para a sequência ser incluída no UniGene, o inserto do clone deve ter no mínimo 100 pb de alta qualidade sem sequências repetitivas. Além disso, a sequência deve ser ancorada na extremidade 3' dos transcritos. Evidências disso pode ser a presença de sinais de poliadenilação canônicos, presença de cauda poliA ou pelo menos duas sequências de EST obtidas do seqüenciamento com o iniciador 3' (Pontius *et al.*, 2002).

Além de banco de dados gerais que armazenam sequências de vários organismos, existem fontes especializadas de ESTs, criadas para específicos organismos. Como exemplos podem ser citados “The Institute for Genome Research”, “USA (TIGR) Gene Indices”, “The Rat EST Project” e “Cancer Genome Anatomy Project” (Nagaraj *et al.*, 2006).

1.6.2 Análise de qualidade das ESTs

Uma EST típica é uma copia curta de mRNA sequenciada apenas uma vez e, como tal, é altamente propensa ao erro, especialmente em suas extremidades. A qualidade da sequência é significativamente melhor no meio da mesma. Além disso, as sequências obtidas podem estar contaminadas com vetores, adaptadores, que precisam ser eliminados.

Uma primeira análise de qualidade de ESTs comumente usada é realizada com auxílio do programa Phred, desenvolvido pelos doutores Phil Green e Brent Ewing. Tal programa realiza a leitura dos cromatogramas (nos formatos SCF, ABI e ESD) gerados pelo sequenciador, interpreta e escreve as bases em um arquivo de texto e atribui valores de qualidade ("Phred scores") para cada base lida. O valor de qualidade é o logaritmo da probabilidade de erro [$Q = -10 \log_{10}(P_e)$], onde Q é a qualidade e P é a probabilidade de erro de uma base particular], esses valores vão de 4 a 60 sendo que os maiores valores correspondem a uma maior qualidade. Um valor de Phred de 20, por exemplo, significa 1% de chance de erro. Os valores de qualidade podem ser usados pelo programa para

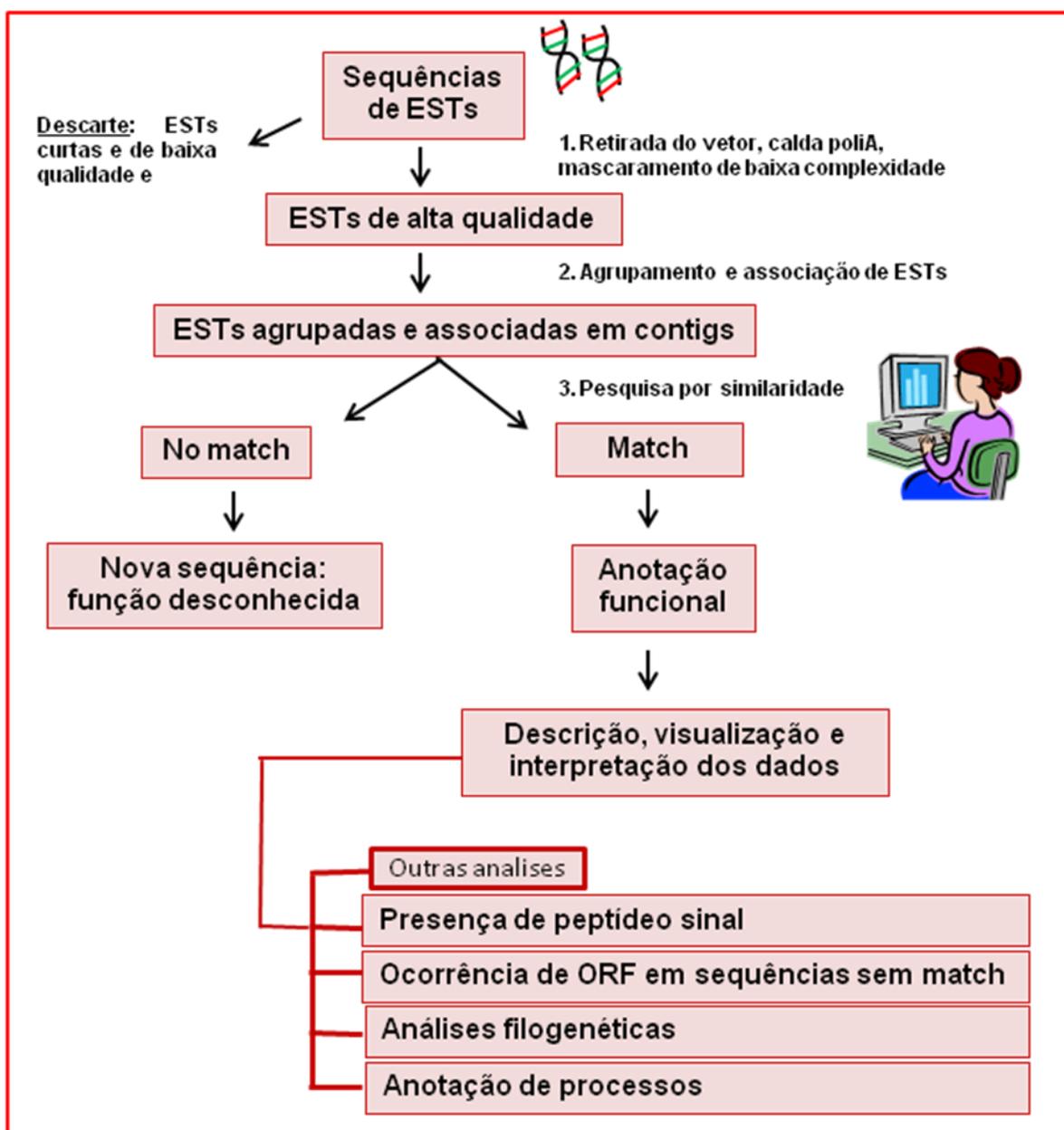


Figura 3: Etapas de análises de bioinformática em estudos de transcriptoma. Após o sequenciamento, as ESTs são avaliadas quanto a qualidade das bases e são removidos os vetores, as caudas poli A, adaptadores. Em seguida, as sequências são agrupadas utilizando programas de clusterização para formação de sequências consenso. Realiza-se, então, a busca por similaridade dos transcritos com sequências depositadas em banco de dados. Procede-se a anotação com a nomeação das ESTs e associação de suas funções baseando-se nos dados de similaridade com sequências já conhecidas. Após a descrição geral dos dados e de ESTs de interesse, como variantes de toxinas, os trabalhos apresentam diferentes formas de análise complementar, como estudos filogenéticos, avaliação de sequências sem similaridade com banco de dados em relação a ocorrência de peptídeo sinal, de janela aberta de leitura, anotação de processos e de vias biológicas presentes na glândula. Além disso, uma sequência consenso pode ser conceitualmente traduzida em um provável peptídeo e então comparada com banco de dados de proteínas.

realizar a trimagem das sequências. A alta precisão dos valores de qualidade do Phred torna essa ferramenta ideal para acessar a qualidade das sequências. O método mais frequentemente usado é a avaliação das bases com valor de Phred de 20 ou mais. Os

arquivos de saída do programa Phred podem ser usados em programas de clusterização para aumentar a acurácia das sequências agrupadas (Ewing *et al.*, 1998)

Após a análise de qualidade das sequências usando o programa Phred, as ESTs são comumente processadas para a eliminação de sequências contaminantes. A comparação das ESTs com banco de dados não redundante de vetores, tais como Univec e EMVEC utilizando ferramenteas como o BLAST ou o Cross_Match (Smith e Green, trabalho não publicado) podem identificar a contaminação com vetor para remoção do mesmo (Nagahaj *et al.*, 2006).

Um exemplo de programa que realiza a eliminação de sequências contaminantes e que foi utilizado no presente trabalho é o SeqClean (disponível em <<http://compbio.dfci.harvard.edu/tgi/software/>>). Ele consiste em um script para trimagem e validação de ESTs ou outras sequências de DNA. Essa ferramenta elimina vários contaminantes e sequências de baixa qualidade e complexidade. Esse programa foi desenvolvido primariamente para “a limpeza” de ESTs de banco de dados, quando vetores específicos e sítios de *Splicing* não estão disponíveis ou quando a eliminação de várias sequências contaminantes é desejável. O programa trabalha por processar as sequências de entrada e filtrá-las de acordo com os seguintes critérios: porcentagem de bases indeterminadas; remoção de cauda poli-A; análise geral de baixa complexidade; alinhamento com várias sequências usadas durante o processo de sequenciamento (vetores, adaptadores); alinhamento com contaminantes ou outras sequências indesejáveis como mitocondriais, ribossomais, bacterianas. No caso da limpeza de vetores/adaptadores, a sobreposição é considerada apenas se for acima de 92% de identidade, apresentar poucos gaps e serem localizadas a 30% de distância da extremidade da sequência. Como resultado, o Seqclean cria dois arquivos de saída: (1) arquivo contendo as sequências filtradas válidas e trimadas em formato fasta; (2) um relato do processo de limpeza (*cleaning report*), que fornece detalhes sobre a trimagem das sequências e descarte das mesmas (coordenadas, razões para o descarte, nome dos contaminantes, etc.).

1.6.3 Agrupamento das sequências

O agrupamento das ESTs é importante, pois elimina a redundância das sequências, aumenta o tamanho das mesmas facilitando a anotação por homologia e aumenta o nível de confiabilidade de cada EST (Perteau *et al.*, 2003, Nagahaj *et al.*, 2006). Na análise de

agrupamento, as ESTs tratadas são usadas como entrada em um programa. Tal programa deve comparar essas sequências entre si de forma a encontrar quais delas são idênticas ou que contem regiões parecidas o suficiente para serem reunidas em uma só (Nagaraj *et al.*, 2006). Assim, o programa apresenta uma saída contendo as sequências que foram agrupadas – chamadas de consenso ou contigs – e as sequências que não foram reunidas - chamadas de singlets. Cada uma das sequências resultantes do agrupamento (seja ela um singlet ou um contig) é chamada de unique. Considerando uma análise ideal cada uma das sequências unique deve representar um gene distinto. Entretanto, na prática, a presença de famílias gênicas e de genes duplicados dificulta a obtenção desse resultado ideal e, muitas vezes, a sequência unique pode representar mais de um gene. Em outras ocasiões, um mesmo gene pode estar representado por mais de um unique, sendo que um dos uniques pode corresponder, por exemplo, a extremidade 5' de um determinado gene e outro a extremidade 3' do mesmo (Prodocimi, 2003; Wang *et al.*, 2004). Os programas Phrap e CAP3 (Huang e Madan, 1999) são os mais comumente usados no agrupamento e associação de ESTs (Nagaraj *et al.*, 2006). Esses dois programas e um outro programa de agrupamento chamado TGICL, utilizado no presente estudo, serão brevemente descritos abaixo.

CAP3 constitui a terceira geração do programa de associação de sequência CAP. O algoritmo consiste de 3 fases principais: (1) Regiões de má qualidade em cada leitura são identificadas e removidas. A sobreposição entre as sequências são computadas e as falsas sobreposições identificadas; (2) Sequências são reunidas para formar contigs em ordem decrescente de escore de sobreposição (tamanho da sequência sobreposta x qualidade da região de sobreposição x escores de match/mismatch/gap). Sequências diretas e reversas são usadas para fazer correções dos contigs; (3) Alinhamentos múltiplos das leituras são construídos e uma sequência consenso juntamente com valores de qualidade para cada base é computada para cada contig. Valores de qualidade são usados para computação da sobreposição e construção de alinhamentos múltiplos (Huang e Madan, 1999).

Outro programa de agrupamento, o Phrap (Green, 1999), lê as sequências e o arquivo de qualidade, corta as regiões de homopolímero no fim das mesmas e constrói as sequências consenso. Suas principais características são: permitir o uso de sequências completas e não apenas de regiões trimadas e de excelente qualidade; uso de dados de qualidade internos e suplementados pelo usuário para melhorar a acurácia do agrupamento na presença de repetições; construção de consensos como um mosaico das partes de maior

qualidade das leituras; capacidade de lidar com grande volume de dados. Esse programa também encontra pares de sequências que tem regiões de similaridade e elimina leituras duplicadas; realiza comparações SWAT (Smith-Waterman) em pares de sequências que apresentam regiões de sobreposição e computa score de SWAT; alinha as sequências aos contigs, observa inconsistências e possíveis locais de alinhamento incorreto.

PHRAP e CAP3 usam diferentes ferramentas para gerar as sequências contigs. Comparações empíricas realizadas por Huang e Madan (1999) utilizando esses dois programas mostraram que o PHRAP produz contigs mais longos que o CAP3, enquanto este frequentemente produz consensos com menos erros que o PHRAP. Como o consenso produzido pelo PHRAP é um mosaico de leituras de regiões de maiores valores de qualidade, esse programa não utiliza a cobertura redundante como a faz o CAP3. A ferramenta de mosaico (PHRAP) pode trabalhar em uma situação onde o contig tem uma baixa cobertura, mas cada região do mesmo é coberta por uma parte da leitura com altos valores de qualidade. Por outro lado, quando os valores de qualidade não são disponíveis, é recomendável o uso do CAP3, uma vez que esse programa é capaz de usar a cobertura redundante na construção do consenso. O usuário pode usar CAP3 para produzir arcabouços com dados contendo fragmentos diretos e reversos, uma vez que esse programa utiliza tais fragmentos na associação das sequências. É recomendável o uso de ambos os programas e a comparação dos consensos resultantes provenientes pelos mesmos (Huang e Madan, 1999).

TGI Clustering tools (TGICL) foi o programa de agrupamento escolhido no presente trabalho. Ele consiste em um *pipeline* para rápido agrupamento de grande quantidade de dados de ESTs. A clusterização é realizada por uma versão ligeiramente modificada do MEGABLAST do NCBI, que realiza comparações par a par entre as sequências para agrupar aquelas que dividem regiões de significativa identidade (Perteau *et al.*, 2003). Em seguida, os grupos resultantes são associados usando o programa CAP3. TGICL começa com um grande arquivo multi-FASTA e tem o arquivo de saída com o mesmo formato do CAP3. Nesse software, a cuidadosa separação das sequências antes da associação é essencial para uma reconstrução mais confiável dos transcritos. A associação individual de cada grupo tem a vantagem de produzir uma sequência consenso maior e mais completa, enquanto elimina potenciais erros de agrupamento (Perteau *et al.*, 2003).

1.6.4 Análise de similaridade

Uma vez que as sequências consenso (provável gene) são obtidas da associação de ESTs, possíveis funções podem ser atribuídas através de anotação obtida a partir de pesquisa de similaridade, através de ferramentas e bancos de dados disponíveis livremente (Nagaraj *et al.*, 2006). A ferramenta mais utilizada nessa etapa é o BLAST (*The Basic Local Alignment Search Tool*), que encontra regiões de similaridade entre sequências. O programa compara sequências de nucleotídeos ou proteínas com sequências de banco de dados e calcula o significado estatístico da associação. BLAST pode ser usado para inferir relação evolucionária e funcional entre as sequências, bem como auxiliar na identificação de membros de famílias gênicas (Altschul *et al.*, 1997).

Diferentes tipos de programas BLAST servem como ferramentas universais para pesquisa de similaridade:

MEGABLAST: realiza pesquisa no banco de dados de nucleotídeos a partir de uma sequência de nucleotídeo. Essa ferramenta foi especificamente desenhada para descobrir longos alinhamentos de forma eficiente entre sequências muito similares.

MEGABLAST descontínuo: também realiza pesquisa no banco de dados de nucleotídeos a partir de uma sequência de nucleotídeo, entretanto essa busca considera a redundância do código genético, ou seja, foca na descoberta de match na primeira e segunda base e ignora o mismatch na terceira posição.

BLASTN: realiza pesquisa no banco de dados de nucleotídeos a partir de uma sequência de nucleotídeo. Como exemplo, pode ser usado para pesquisa de ESTs contra banco de dados de nucleotídeos. O BLASTN é melhor que o MEGABLAST na descoberta de alinhamentos com sequências de outros organismos.

BLASTX: realiza busca no banco de dados de proteínas a partir de uma sequência de nucleotídeo traduzida. Como exemplo, BLASTX traduz uma sequência consenso de ESTs (*query*) em um produto protéico nas seis janelas abertas de leitura, seguido pela comparação com banco de dados de proteína.

BLASTp: realiza busca no banco de dados de proteínas a partir de uma sequência de proteína.

tBLASTn: realiza busca no banco de dados de nucleotídeo traduzido a partir de uma sequência de proteína.

tBLASTx: realiza busca no banco de dados de nucleotídeo traduzido a partir de uma sequência de nucleotídeo traduzida.

PSI-BLAST (Position-Specific Iterated BLAST): consiste no programa BLAST mais sensível e permite a descoberta de proteínas distantemente relacionada ou novos membros de uma família proteica. A primeira rodada do PSI-BLAST consiste de um BLASTp convencional. A partir das sequências resultantes o programa constrói uma matriz de escore posição-específica, que será usada para avaliar o alinhamento na próxima iteração da pesquisa. Os novos hits são incluídos na formação de uma nova matriz para mais uma rodada de BLAST. Considera-se que o PSI-BLAST convergiu quando nenhum novo match ocorre em uma iteração subsequente.

Essas e outras informações mais detalhadas sobre a ferramenta BLAST podem ser obtidas na página: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.

1.6.5 Tradução conceitual de ESTs

Dados de ESTS podem ser relacionados com anotações centradas em proteínas devido a acurada e robusta tradução de proteínas, uma vez que os polipeptídeos são moldes melhores para identificação de domínios e motivos, para o estudo da localização de proteínas e para a designação de ontologias (GOs).

O primeiro passo de tradução de sequências de ESTs é a identificação de regiões codificadoras de proteínas ou ORFs, a partir do consenso das ESTs, acentuando o processo de descoberta do gene e predições de suas bordas. Algumas ferramentas são disponíveis para identificação de ORFs (Nagaraj *et al.*, 2006). Por exemplo, ORF Finder (*Open Reading Frame Finder*) é uma ferramenta de análise gráfica que encontra todas as janelas abertas de leitura de determinado tamanho mínimo em uma sequência. Essa ferramenta identifica as ORFs usando código genético padrão ou alternativo. A sequência de aminoácidos predita pode ser salva em vários formatos e pesquisada contra banco de dados usando BLASTP (<http://www.ncbi.nlm.nih.gov/projects/gorf/>). O ORF Finder foi o programa escolhido para a tradução de ESTs no presente trabalho.

Outro programa de tradução é o ORFPredictor, que foi desenvolvido especialmente para identificar regiões codificadoras de proteínas em sequências derivadas de ESTs. Esse programa fornece 6 frames de tradução e prediz a região codificante mais provável de todas elas (Min *et al.*, 2005). ESTScan e DECODER podem detectar e extrair regiões

codificantes de ESTs de baixa qualidade ou cDNA parcial enquanto corrige erros de mudança de frame e fornece tradução conceitual. Prot4EST é um *pipeline* de tradução de EST em polipeptídeo que incorpora o DECODER, ESTScan e BLASTX para uma predição mais acurada. Os prováveis peptídeos podem ser comparados com banco de dados de proteínas usando BLASTp (Nagaraj *et al.*, 2006).

1.6.6 Anotação funcional

Existem algumas ferramentas para anotação funcional de sequências. No presente trabalho optou-se pela utilização do *pipeline* Blast2GO (Conesa *et al.*, 2005), que inclui diferentes ferramentas de anotação, que estão detalhadas abaixo.

Blast2GO é uma ferramenta de fácil utilização para análise de um conjunto de ESTs que possibilita o entendimento do significado fisiológico de grande quantidade de genes (Conesa *et al.*, 2005). Ele usa o BLAST para descobrir homólogos no banco de dados e, em seguida, o programa extrai os termos do *Gene Ontology* (GO) de cada *hit* obtido por mapear as associações de anotações existentes. Uma regra de anotação enfim atribui o termo do GO para a sequência em questão.

A anotação funcional propiciada pelo Blast2GO é baseada na transferência de homologia e os parâmetros de anotação incluem: a escolha do banco de dados; a extensão e o número de resultados de BLAST, o comprimento da sequência do banco de dados que apresentou resultado de BLAST, a qualidade das anotações transferidas e a inclusão de motivos de anotação. As principais ferramentas de anotação incluídas no Blast2GO são os vocabulários usados pelo Gene Ontology, InterProScan e vias KEGG (Conesa e Gotz, 2008):

Gene Ontology (GO): fornece um vocabulário estruturado para domínios biológicos específicos que podem ser usados para descrever produtos gênicos em qualquer organismo (Ashburner *et al.*, 2001). A ontologia cobre três domínios: componente celular, a parte da célula ou seu ambiente extracelular; função molecular, a atividade elementar do produto gênico no nível molecular, tal como ligação e catálise; e processo biológico, operação ou série de evento molecular com início e fim definidos pertinentes ao funcionamento de unidades vivas (células, tecidos, órgãos e organismos) (Thomas *et al.*, 2007). O uso de vocabulário controlado facilita consideravelmente a troca de conhecimento biológico, além de facilitar a pesquisa computacional que lida com esse

conhecimento (Conesa e Gotz, 2008). O processo de associação entre uma molécula biológica e uma ontologia é chamado anotação ontológica. Existem três principais tipos de evidência para realizar a anotação ontológica: evidência baseada na literatura, em homologia e por outras evidências computacionais. Em última instância, a fonte de todas as anotações ontológicas é a descoberta experimental publicada na literatura científica através de um ou mais passos de inferência. O tipo de evidência é crucial para estimar a confiabilidade de diferentes anotações. A maioria das anotações ontológicas é baseada em inferências por homologia e a confiabilidade desse tipo de anotação depende da metodologia computacional usada para inferir a herança da função a partir de um ancestral comum (Thomas *et al.*, 2007).

InterProScan: Uma vez que o provável polipeptídeo é obtido, sua função pode ser predita por pesquisa de similaridade contra banco de dados de sequências de proteína, banco de dados de famílias ou motivos usando uma ferramenta integrada como o Interproscan. Interproscan escaneia uma determinada sequência de proteína contra assinaturas de proteínas presentes em banco de dados membros do InterProScan (PROSITE, PRINTS, Pfam, ProDom, SMART) (Zdobnov e Apweiler, 2001).

Kyoto Encyclopedia of Genes and Genomes PATHWAY (Vias KEGG): constitui uma coleção de mapas de vias desenvolvidas manualmente que representam o conhecimento das interações moleculares e redes de reações. Um conjunto de dados moleculares em larga escala, como informações de genomas, transcriptomas e proteomas podem ser submetidos ao mapeamento usando as Vias KEGG, o que auxilia na interpretação biológica em relação a funções sistêmicas (Ogata *et al.*, 1999).

O Blast2GO também inclui funções gráficas e estatísticas para avaliação do procedimento de anotação e resultados finais. Especialmente a abundância relativa de termos funcionais pode ser facilmente obtida e visualizada (Conesa e Gotz, 2008).

1.7 Estudos de Transcriptoma da glândula de veneno de escorpiões

Possani *et al.* (2000) estimaram a existência de cerca de 100.000 peptídeos bioativos presentes nas glândulas de veneno de escorpiões, considerando em conjunto as 1.500 espécies existentes. De acordo com os dados do *Animal Annotation Toxin Program* do Uniprot (Jungo *et al.*, 2012), mais de 60 espécies de escorpiões já tiveram seus venenos estudados, sendo obtidos 811 componentes do veneno, o que corresponde a

aproximadamente 0,05% do total de peptídeos estimados. Essa grande variedade de peptídeos encontrados na glândula de veneno tem despertado o interesse de pesquisadores na busca de moléculas com potenciais atividades farmacológicas e biotecnológicas.

O veneno de escorpiões vem sendo intensivamente estudado através de métodos baseados em PCR aplicados a bibliotecas de cDNA. Esses estudos contribuíram consideravelmente para a compreensão da composição do veneno e quando associados a trabalhos que realizam clonagem, expressão e caracterização de componentes específicos permitiu a identificação do papel de diversas moléculas do veneno. Entretanto, tais estudos são limitados pela especificidade dos *primers* usados, impossibilitando uma visão mais completa da diversidade de constituintes do veneno (Rodríguez de la Vega *et al.*, 2010; Almeida *et al.*, 2012). Uma abordagem de estudo que vem sendo utilizada para se explorar a diversidade de peptídeos produzidos pelos escorpiões é a análise do transcriptoma da glândula de veneno desses animais (Rodríguez de La Vega *et al.*, 2010). Geralmente, tais estudos resultam em um grande volume de dados que possibilitam: (1) a compreensão de mecanismos envolvidos com a síntese de peptídeos do veneno; (2) a determinação de padrões de expressão; (3) a descoberta de genes; (4) a investigação de eventos evolutivos relacionados com a diversificação de toxinas e componentes do veneno; e (5) o desenvolvimento de recursos úteis para estudos de genômica funcional (Schwartz *et al.*, 2007; Ma *et al.*, 2009; Rodríguez de La Vega *et al.*, 2010; Ruiming *et al.*, 2010).

O primeiro estudo de transcriptoma de glândula de veneno de escorpião ocorreu em 2007 (Schwartz *et al.*, 2007). Depois desse, outros onze artigos utilizando essa abordagem foram publicados até agora (Kozminsky-Atias *et al.*, 2008; D'Suze *et al.*, 2009; Ma *et al.*, 2009; Silva *et al.*, 2009; Ma *et al.*, 2010; Ruiming *et al.*, 2010; Morgenstern *et al.*, 2011; Ma *et al.*, 2012; Diego-García *et al.*, 2012, Almeida *et al.*, 2012; Rendón-Anaya *et al.*, 2012). Além desses, o trabalho de Roeding *et al.* (2009) analisou o transcriptoma de um escorpião inteiro. Esses catálogos de transcritos contém desde pouco mais de uma centena de ESTs até cerca de três milhões delas, que foram provenientes 13 espécies, pertencentes a 5 distintas famílias de escorpiões. A maioria dos trabalhos investigou a glândula de veneno considerada transcricionalmente ativa, ou seja, a extração de RNA para a construção da biblioteca foi realizada 2 a 5 dias após a extração do veneno. Três trabalhos relatam o perfil transcricional da glândula de veneno em repouso, ou seja, a extração de RNA foi realizada sem a prévia extração do veneno. Informações sobre esses trabalhos estão

sumarizadas na Tabela 3 e a distribuição dos escorpiões investigados nesses estudos é mostrada na Figura 4.

Além de uma diversidade de toxinas de canais para Na^+ , K^+ , Cl^- e Ca^{2+} , os trabalhos de transcriptomas de glândula de veneno de escorpiões identificaram e avaliaram a expressão de outros componentes do veneno como peptídeos citolíticos, antimicrobianos, aniônicos, lectinas, hialuronidases, fosfolipases, lisozimas, metaloproteases, peptídeos ricos em cisteína, peptídeos ricos em glicina (Schwartz *et al.*, 2007; Kozminsky-Atias *et al.*, 2008; D'Suze *et al.*, 2009; Ma *et al.*, 2009; Silva *et al.*, 2009; Ma *et al.*, 2010; Ruiming *et al.*, 2010; Morgenstern *et al.*, 2011; Ma *et al.*, 2012; Diego-García *et al.*, 2012, Almeida *et al.*, 2012; Rendón-Anaya *et al.*, 2012). Além disso, esses trabalhos descrevem transcritos relacionados com processos celulares que ocorrem na glândula, o que pode contribuir para a melhor compreensão da síntese do veneno. Como exemplo, a comparação do perfil transcricional entre glândula ativa *versus* glândula em repouso de *C. noxius* mostrou diferenças na transcrição de componentes celulares. Na glândula ativa houve alta representação de transcritos relacionados com metabolismo de aminoácidos, lipídeos e carboidratos, além de componentes do proteasoma. Por outro lado, na glândula em repouso ocorreu maior representação de transcritos envolvidos com processamento de informações ambientais, que incluem aqueles relacionados às vias de transdução de sinal e de transporte pela membrana (Rendón-Anaya *et al.*, 2012).

Outra contribuição de estudos de transcriptomas consiste no repertório de sequências geradas. Tais sequências são geralmente depositadas em bancos de dados mundiais, o que facilita o futuro isolamento e aplicação dessas moléculas. Interessantemente, cerca de 50% das ESTs descritas nesses trabalhos não apresentam resultado quando comparadas com as sequências depositadas em bancos de dados públicos. Isso evidencia a escassez de informação molecular sobre os escorpiões e organismos relacionados, ao mesmo tempo, que revela o grande potencial de descoberta de novas proteínas bioativas quando se realiza esse tipo de análise (Rodríguez de La Vega *et al.*, 2010).

Tabela 3. Informações dos transcriptomas de glândula de veneno de escorpiões realizados até o momento

Espécies de escorpião	Famílias	Estado da glândula (dias pós-extração)	Nº ESTs	Técnica de sequenciamento	Autores	
<i>Lynchas mucronatus</i>	Buthidae	Ativa (2)	738	Sanger	Ruiming et al., 2010; Ma et al., 2012	
<i>Isometrus maculatus</i>		Ativa (2)	743	Sanger	Ma et al., 2012	
<i>Hottentotta judaicus</i>		Repouso	537	Sanger	Morgenstem et al., 2011	
<i>Buthus occitanus</i>		Ativa (2)	450	Sanger	Kozminsky-Atlas et al., 2008	
<i>Tityus discrepans</i>		Ativa (5)	112	Sanger	D'Suze et al., 2009	
<i>T. serrulatus</i>		Ativa (2)	1629	Sanger	Alvarenga et al., 2012	
<i>T. stigmurus</i>		Repouso	540	Sanger	Almeida et al., 2012	
<i>Centruroides noxius</i>		Ativa (5) e repouso	3.008.0	NGS*	Rendón-Anaya et al., 2012	
<i>Scorpiops magerisonae</i>		Scorpiopidae	Ativa (2)	730	Sanger	Ma et al., 2012
<i>S. jendeki</i>			Ativa (2)	871	Sanger	Ma et al., 2009
<i>Heterometrus petersii</i>	Não Buthidae	Ativa (2)	486	Sanger	Ma et al., 2010	
<i>Pandinus cavimanus</i>		Ativa (2)	184	Sanger	Diego-Garcia et al., 2012	
<i>Hadruus gertschi</i>		Iuridae	147	Sanger	Schwartz et al., 2007	
<i>Opisthacanthus cayaporum</i>	Liochelidae	Ativa (4)	118	Sanger	Silva et al., 2009	

*NGS = sequenciamento de nova geração

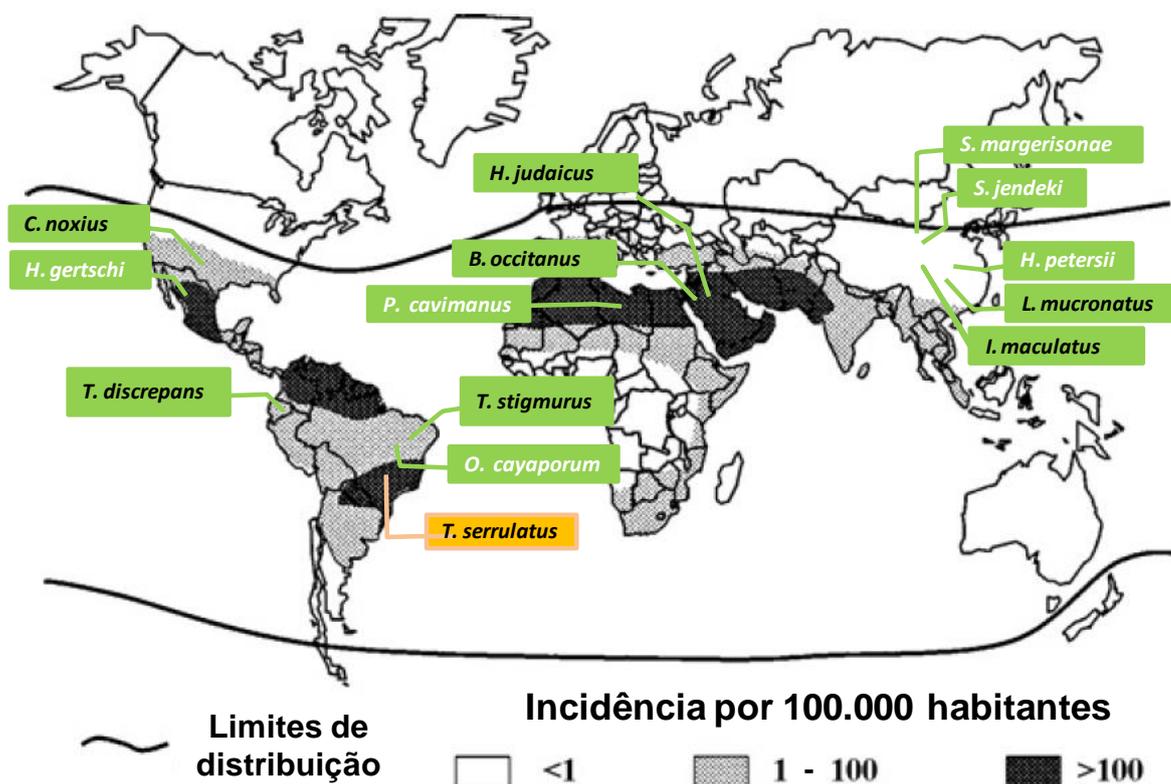


Figura 4: Distribuição das espécies de escorpião já analisadas em estudos de transcriptoma de glândula de veneno. Espécies escritas em preto pertencem à família Buthidae e as escritas em branco são não-butídeos. O mapa indica ainda a incidência de acidentes com escorpiões e o limite de distribuição desse grupo animal. Fonte: mapa adaptado de Chippaux e Goyffon (2008).

2. JUSTIFICATIVA

Em 2007, a Organização Mundial de Saúde (World Health Organization - WHO) considerou o escorpionismo uma doença negligenciada que requer uma urgente ação internacional. De fato, no Brasil, acidentes com escorpiões consistem em sério problema de saúde pública por muitas décadas, com mais de 50.000 casos descritos em 2010 (Brasil, Ministério da Saúde, 2011). A espécie de escorpião considerada a mais perigosa e a principal responsável pelos acidentes é *Tityus serrulatus* (família Buthidae).

Devido à importância médica, toxinas e outros componentes do veneno de *T. serrulatus* têm sido isolados e estudados individualmente (Possani *et al.*, 1977; Arantes *et al.*, 1989, Sampaio *et al.*, 1991; Martin-Eauclaire *et al.*, 1994; Sampaio *et al.*, 1996; Guatimosim *et al.*, 1999; Novello *et al.*, 1999; Petricevich *et al.*, 2007; Mendes *et al.*, 2008; Verano-Braga *et al.*, 2008). Entretanto, a investigação do transcriptoma da glândula de veneno dessa espécie não foi realizada. Esse tipo de estudo mais holístico da glândula de veneno é interessante porque pode revelar, por exemplo: (1) os mecanismos e genes participantes da síntese dos componentes do veneno (2) novas toxinas ou constituintes do veneno, que podem tanto constituir moléculas bioativas com potencial para aplicações em biotecnologia, quanto revelar candidatos envolvidos no envenenamento sistêmico e (3) o padrão de expressão da glândula de veneno. Assim, estudos de transcriptoma constituem uma interessante ferramenta para explorar a diversidade de moléculas produzidas na glândula de veneno de escorpiões (Rodriguez de La Veja *et al.*, 2010).

Estudos de transcriptomas de glândulas de veneno de escorpião têm sido realizados recentemente (Schwartz *et al.*, 2007; Kozminsky-Atias *et al.*, 2008; D'Suze *et al.*, 2009; Ma *et al.*, 2009; Silva *et al.*, 2009; Roeding *et al.*, 2009; Ma *et al.*, 2010; Ruiming *et al.*, 2010, Morgenstern *et al.*, 2011; Ma *et al.*, 2012; Diego-García *et al.*, 2012; Almeida *et al.*, 2012; Rendón-Anaya *et al.*, 2012). A maioria deles foca nos constituintes tóxicos do veneno e tem contribuído substancialmente para a descoberta de muitas proteínas e peptídeos sintetizados na glândula de veneno. Recentemente, o aumento de informação molecular a respeito dos escorpiões e organismos relacionados abre a possibilidade de estudos de transcriptômica comparativa, que podem gerar uma visão mais ampla sobre as glândulas de veneno e seus peptídeos e proteínas. Entretanto, de acordo com o nosso conhecimento, estudos de transcriptômica comparativa envolvendo escorpiões são escassos e, na literatura, foi encontrado apenas um trabalho comparando transcriptomas

provenientes de três espécies de escorpião pertencentes a duas diferentes famílias (Ma *et al.*, 2012). Nesse contexto, comparações entre catálogos de transcritos tornam-se necessárias para possibilitar a identificação de similaridades e diferenças entre transcriptomas de diferentes escorpiões ou mesmo entre escorpiões e organismos relacionados. Essa análise permite integrar o conhecimento já produzido, possibilitando, por exemplo, a identificação de genes exclusivos das glândulas de veneno e ortólogos em organismos relacionados, além de propiciar a caracterização da expressão gênica nesse órgão.

3. OBJETIVOS

Este trabalho apresenta basicamente dois objetivos gerais: (1) investigar a diversidade e complexidade molecular da glândula de veneno de *T. serrulatus* por meio da análise do repertório de transcritos usando como estratégia a descrição das ESTs; (2) realizar análises de transcriptômica comparativa entre catálogos de transcritos de quelícerados a fim de caracterizar o padrão de expressão de glândulas de veneno. Para tal, têm-se os seguintes objetivos específicos:

- Formação de um banco de dados contendo ESTs de excelente qualidade através do sequenciamento de cerca de 2000 clones da biblioteca de cDNA da glândula de veneno de *T. serrulatus*.

- Agrupamento das ESTs usando diferentes programas de agrupamento de sequências (Phrap, CAP3, TGICL), escolha do mais adequado e análise do número de sequências únicas e agrupamentos resultantes desse processo.

- Realização da anotação das ESTs através da pesquisa das mesmas contra o banco de dados de proteínas (UniProt e nr) através do uso de algoritmo BLASTx e anotação com termos do *Gene Ontology* usando *pipeline Blast2GO*.

- Análise de ESTs envolvidas com a maquinaria celular geral relacionadas, por exemplo, a síntese e o processamento protéico.

- Realização de análise histológica da glândula de veneno de *T. serrulatus* e comparação das características morfológicas desse órgão com os dados moleculares oriundos da análise das ESTs.

- Análise da diversidade de sequências relacionadas a componentes do veneno através da análise de formas variantes de ESTs traduzidas.

- Descrição dos componentes do veneno e sua ocorrência em termos de número de ESTs obtidas. Identificação de ortólogos dos componentes do veneno através da metodologia Seed Server e PSI-BLAST contra o UniProt. Alinhamento múltiplo dos componentes do veneno e seus ortólogos usando a ferramenta webprankster.

- Análise da ocorrência de ortólogos dos componentes do veneno de *T. serrulatus* nos grupos de organismos para inferência do grupo ancestral no qual possivelmente originou cada um dos componentes descritos. Inferência essa que será obtida através de duas análises: (1) busca de ortólogos dos componentes do veneno no universo protéico

(UniProt) usando PSI-BLAST e no universo de transcritos (dbEST) utilizando PSI-tBLASTn.

- Comparação entre catálogos de transcritos de quelicerados buscando identificar similaridades entre eles e peculiaridades no padrão de expressão de glândulas de veneno através de anotação usando BLASTx e análises de distribuição dos termos do GO realizada pelo programa Blast2GO para os transcriptomas disponíveis no dbEST das seguintes espécies *Lynchus mucronatus*, *Mesobuthus gibbosus*, *Heterometrus petersii* e *Scorpiops jendeki* (ordem Scorpiones), *Loxocles laeta*, *Nephila antipodiana* e *Dysdera erythrina* (ordem Araneae), *Ixodes ricinus* e *Sarcoptes scabiei* (ordem Acari), *Limulus polyphemus* e *Carcinoscorpius rotundicauda* (ordem Xiphosura) e *Endeis spinosa* (ordem Pycnogonida).

- Identificação de transcritos comuns entre o transcriptoma de *T. serrulatus* e os demais transcriptomas de quelicerados através de análise utilizando o programa tBLASTx.

- Avaliação da expressão de componentes do veneno encontrados em *T. serrulatus* nos demais transcriptomas amostrados usando as ferramentas de tBLASTx e tBLASTn.

4. MATERIAL E MÉTODOS

4.1 Biblioteca de cDNA da glândula de veneno do escorpião *T. serrulatus*

O Laboratório de Biotecnologia e Marcadores Moleculares já possui uma biblioteca de cDNA da glândula de veneno de *T. serrulatus*. Tal biblioteca foi construída pelo Professor Evanguedes Kalapothakis com cerca de 60 telsons. Os detalhes técnicos dessas etapas foram descritos por Kalapothakis *et al.* (2001). Abaixo serão descritas sucintamente as principais etapas da construção da biblioteca.

Os escorpiões foram submetidos à estimulação elétrica do télson para a extração do veneno. Acredita-se que tal procedimento estimule as células produtoras do veneno a sintetizarem RNAm. Dois dias após esse procedimento, foi realizada a extração total dos RNAs com a utilização do “RNA Extraction kit” (PHARMACIA). Após a extração, o RNAm poliadenilado foi purificado em coluna oligo-dT utilizando o “Poly (A) Quick mRNA Isolation kit” (STRATAGENE). O cDNA complementar foi sintetizado a partir do RNAm usando o “Zap cDNA synthesis kit” (STRATAGENE). O vetor Uni-Zap XR™ (STRATAGENE) foi usado para a montagem da biblioteca. O fagemídeo pBluescript SK⁻ está presente no vetor Uni-Zap XR™. O empacotamento do vetor Uni-Zap XR™ foi realizado através do Gigapack II packaging extract (STRATAGENE). Células hospedeiras da linhagem XL1 Blue MRF foram usadas nas etapas de clonagem para plaqueamento dos fagos. Para amplificação da biblioteca, alíquotas da mistura do empacotamento contendo cerca de 50.000 bacteriófagos foram adicionadas a células hospedeiras da linhagem XL1 Blue MRF crescidas em meio NZK. Para a subclonagem optou-se pelo vetor pBluescript SK.

4.2 Transformação dos plasmídeos em *Escherichia coli*

Os clones da biblioteca foram previamente isolados e tiveram seus plasmídios purificados e estocados como DNA precipitado. Parte dos clones já havia sido sequenciados pelo Laboratório de Biotecnologia e Marcadores Moleculares. Neste trabalho, aproximadamente 2.000 clones precipitados da biblioteca foram transformados e purificados para subsequente sequenciamento. Os procedimentos técnicos utilizados para essas etapas são descritos a seguir.

Cada clone da biblioteca foi ressuspensionado em 5 μL de água estéril e incubado com 25 μL de bactérias *E. coli* eletrocompetentes, linhagem XL1 Blue (Phoneutria Biotecnologia e Serviços). Essa solução foi transferida para uma cubeta e submetida a um pulso de 2kVa (quilovolt-ampères) em eletroporador. Posteriormente, foi adicionado 1 mL de meio 2XYT (para 1 litro de meio foi usado 16 g de extrato de levedura, 10 g de bactotriptoma e 5 g de NaCl) ou SOC (para 1 litro de meio foi usado 20 g de bactotriptona, 5g de extrato de levedura, 0,6 g de NaCl, 0,186 g de KCl, 2,14 g de MgCl_2 a 10 mM, 1,2 g de MgSO_4 a 10 mM, 3,6 g de glicose a 20 mM) as bactérias foram incubadas a 37°C por 1h. Uma vez que o plasmídeo utilizado contém o gene de resistência a ampicilina, as bactérias foram plaqueadas em meio LB ágar com esse antibiótico (concentração final de 0,1 μg de ampicilina/mL de meio – Phoneutria Biotecnologia e Serviços) para a seleção daquelas que incorporaram o plasmídeo. Uma colônia proveniente de cada clone foi selecionada para inoculação em placas de 96 poços contendo, em cada um, 1,2 mL de meio 2XYT com ampicilina (0,1 μg de ampicilina/ mL de meio). O meio com as bactérias foi mantido a 37 °C com agitação constante por cerca de 18 h. Após esse período, foi realizada a extração e purificação dos plasmídeos.

4.3 Extração e purificação dos plasmídeos

Para a extração e purificação dos plasmídeos, a placa contendo as bactérias foi centrifugada 9 min a 2.750 g e o sobrenadante descartado. Em cada poço, as células foram ressuspensionadas com 240 μL de solução I (glicose 50 mM, Tris-HCl 25 mM pH 8,0, EDTA 10 mM pH 8,0) e novamente centrifugadas a 2.750 g por 6 min. O sobrenadante foi descartado e as células ressuspensionadas com 80 μL de solução I. Esse material foi transferido para uma placa de polipropileno de 96 poços com fundo em U. Foram adicionados em cada poço 2,5 μL de RNase (10mg/ml) e 80 μL de solução II (NaOH 0,2 N, SDS 1%). Após homogeneização, 80 μL de solução III (acetato de potássio 3 M, 11,5% ácido acético glacial) foi acrescentada em cada poço. Depois de ser novamente homogeneizada, a placa foi incubada por 10 min a temperatura ambiente e, em seguida, levada a estufa a 90 °C por 30 min. A placa foi então centrifugada por 9 min a 2.750 g e o sobrenadante transferido para uma placa *Multiscreen* (com filtro hidrofílico PDVF) (MILLIPORE) acoplada sobre uma placa de polipropileno de 96 poços de fundo em V. As placas associadas foram centrifugadas por 5 min a 2.750 g. Após a centrifugação, o filtrado

foi totalmente transferido para a placa em fundo V e a placa *Multiscreen* foi descartada. Ao filtrado foi adicionado 100 μL isopropanol 100%. O material foi centrifugado a 2.750 g por 45 min, o sobrenadante descartado e 200 μL de etanol 70% foram adicionados em cada poço da placa. O material foi centrifugado a 2.750g por 5 min, o sobrenadante descartado e a placa deixada por cerca de 15 min em estufa a 37 °C para a evaporação do etanol. O DNA resultante foi ressuspensionado em 20 μL de água ultrapura e armazenado a -20 °C. Após essas etapas, o DNA foi quantificado em um gel de agarose 0,8% com coloração em GelRed para subsequente sequenciamento.

4.4 Sequenciamento dos plasmídeos

Foram realizadas reações de sequenciamento para os plasmídeos provenientes dos clones da biblioteca. Para cada reação foram utilizados cerca de 200 ng de plasmídeo, 0,5 μL de BigDye Terminator v3. 1 Cycle (Applied Biosystems), 2 μL de iniciador M13 direto ou reverso (1,6 pmol/ μL) (IDT Integrated DNA Technologies), 1,5 μL de Tampão de Diluição 5X BigDye Terminator v1.1, v3.1 da Applied Biosystems e água ultrapura suficiente para 10 μL . Cada reação foi levada para o termociclador, o qual apresenta o seguinte programa padronizado para o sequenciador 3130 Genetic Analyzer da Applied Biosystem:

96°C por 2 minutos	} 45 ciclos
96°C por 30 segundos	
50°C por 15 segundos	
60°C por 4 minutos	
Mantém a 4°C	

Após ser retirado do termociclador (Applied Biosystem), o produto da reação de sequenciamento foi precipitado com solução de precipitação (1 μl de acetato de sódio 3M, 25 μl de etanol 100%, 1 μl de EDTA 0,125 M), lavado com etanol 70% (35 μl) e armazenado em freezer a -20°C. Esse precipitado foi ressuspensionado em formamida Hi-Di (Applied Biosystem) e então encaminhado para o sequenciador ABI 3130 Genetic Analyzer.

4.5 Análises de Bioinformática

4.5.1 Obtenção de transcriptomas de quelicerados

Além das sequências de *T. serrulatus* obtidas no presente trabalho e depositadas no dbEST (dbEST_Id = 75726766-75928119 e GenBank_Accn = JK731601 - JK732954), outros conjuntos de ESTs também depositados nesse banco de dados foram analisados para servirem de base de comparação aos dados obtidos no presente estudo. Para isso, acessou-se a página do Taxonomy do NCBI (National Center for Biotechnological Information) (<http://www.ncbi.nlm.nih.gov/guide/taxonomy/#howto>), realizando-se busca com o termo Arthropoda, 3 níveis de agrupamento e opção “Nucleotide EST”. O resultado dessa busca é mostrado no Quadro 1. Do grupo dos artrópodes quelicerados obteve-se pelo menos um conjunto de EST de cada grupo que possuía coleção de transcritos (as espécies escolhidas para análise estão destacadas em vermelho no Quadro 1). Quando possível, obteve-se catálogos de transcritos provenientes do animal inteiro e de glândula em uma mesma ordem de artrópode. Essa escolha foi feita para possibilitar comparações de transcriptomas de glândulas entre si e destas com animais inteiros em relação à ocorrência de transcritos e ao padrão de expressão. Além disso, buscou-se selecionar transcriptomas com número semelhante de sequências para padronizar a representação de transcritos nos diferentes transcriptomas. No total foram analisados 12 conjuntos de transcritos depositados no NCBI, provenientes das seguintes espécies: *Lynchas mucronatus*, *Mesobuthus gibbosus*, *Heterometrus petersii* e *Scorpiops jendeki* (ordem Scorpiones), *Loxocles laeta*, *Nephila antipodiana* e *Dysdera erythrina* (ordem Araneae), *Ixodes ricinus* e *Sarcoptes scabiei* (ordem Acari), *Limulus polyphemus* e *Carcinoscorpius rotundicauda* (ordem Xiphosura) e *Endeis spinosa* (ordem Pycnogonida) (<http://www.ncbi.nlm.nih.gov/nucest>). Esses catálogos de transcritos foram submetidos ao mesmo tratamento das sequências de *T. serrulatus*, ou seja, Seqclean seguido de agrupamento com TGICL como está explicado abaixo. Nessa etapa do trabalho, as filogenias implementadas pelo UniProt e NCBI foram consideradas como referência.

4.5.2 Análise de qualidade e eliminação de contaminantes de ESTs

Os cromatogramas (2.131 sequências) providos pelo sequenciador automático foram processados pelo programa Phred para análise da qualidade das sequências de *T. serrulatus*. Realizou-se inicialmente um teste para verificar o valor de Phred mais adequado para esse estudo. Para isso, foram utilizadas as sequências dos vetores sem inserto encontradas na biblioteca de cDNA da glândula de veneno de *T. serrulatus*. Essas sequências foram submetidas a valores de Phred de 8, 9, 10, 15 e 20, o que corresponde a uma probabilidade de 16; 12,6; 10; 3,2; e de 1% de erros de base. Assim, uma mesma sequência era submetida à trimagem com diferentes valores de Phred e, posteriormente, era realizado BLASTn com filtro de baixa complexidade desativado para avaliar qual a porção da sequência apresentava *match* com o vetor do banco de dados nr/nt do NCBI. Considerando os resultados obtidos com essas análises associado ao fato de que a trimagem com valores baixos proporcionam transcritos com maior taxa de erro, optou-se por utilizar um valor de Phred de 20. Para o transcriptoma de *T. serrulatus* os cromatogramas estavam disponíveis, o que possibilitou a análise de qualidade com programa Phred. Uma vez que as sequências dos catálogos de transcritos dos demais quelicerados são disponíveis no NCBI em formato fasta, a análise com tal programa não foi possível.

Utilizou-se a ferramenta Seqclean para retirada de contaminantes das ESTs como sequências de vetor (no caso de *T. serrulatus* com vetor pBluescript), adaptador e de cauda poli A e poli T. Todos os transcriptomas de quelicerados obtidos no dbEST também passaram por esse tratamento.

4.5.3 Agrupamento das sequências

Para o agrupamento das sequências de *T. serrulatus* foram testados 3 programas: CAP3; Phrap (do pacote Phred-Phap-Consed) e TGICL. Para cada um desses programas foi avaliada a trimagem das sequências do vetor antes ou após o agrupamento. MegaBLAST das sequências contra elas mesmas também foi avaliado utilizando um valor de identidade maior ou igual a 96%. Após os testes, optou-se pela utilização TGICL. Esse programa foi desenvolvido especificamente para a clusterização de ESTs, sendo que realiza inicialmente um MEGABLAST das ESTs pré-agrupando as mesmas, seguido pela

submissão das sequências ao programa CAP3. Além do transcriptoma de *T. serrulatus*, todos os outros doze conjuntos de ESTs foram submetidos ao programa TGICL para agrupamento das mesmas.

Arthropoda (arthropods)	5.810.943	
○ Chelicerata (chelicerates)	448.860	
○ Arachnida (arachnids)	435.735	
• Acari (mites and ticks)	394.960	{ <i>Ixodes ricinus</i> <i>Sarcoptes scabiei</i>
• Amblypygi (whip spiders)		{ <i>Loxocles laeta</i> (aranha marrom) <i>Dysdera erythrina</i> <i>Nephila antipodiana</i>
• Araneae (spiders)	37.851	
• Opiliones (harvestmen)		
• Palpigradi (microwhip scorpions)		{ <i>Heterometrus petersii</i> <i>Lynchus mucronatus</i> <i>Scorpiops jendeki</i> <i>Tityus serrulatus</i>
• Pseudoscorpiones (false scorpions)		
• Ricinulei (hooded tickspiders)		
• Scorpiones (scorpions)	2.924	
• Solifugae (sun spiders)		
• Uropygi (whip scorpions)		
• unclassified Arachnida		
○ Merostomata (horseshoe crabs)	9.063	
• Xiphosura	9.063	{ <i>Limulus polyphemus</i> e <i>Carcinoscorpius retundicauda</i>
○ Pycnogonida (sea spiders)	4.062	
• Pantopoda	4.062	{ <i>Endeis spinosa</i>
○ Mandibulata (mandibulates)	5.362.083	
○ Myriapoda (myriapods)	4.461	
• Chilopoda (centipedes)		
• Diplopoda (millipedes)	4.461	
• Pauropoda		
• Symphyla (symphylans)		
○ Pancrustacea	5.357.622	
• Crustacea (crustaceans)	861.282	
Hexapoda (insects)	4.496.340	

Quadro 1: Organograma obtido pelo Taxonomy do NCBI para o grupo dos quelicerados – O nome dos grupos aparece em azul com o respectivo número de ESTs depositadas. Nos colchetes estão os nomes das espécies analisadas. Os nomes de espécies em vermelho correspondem aquelas que tiveram o conjunto de ESTs obtidos do NCBI. A espécie em preto é aquela cujo transcriptoma foi obtido no presente trabalho.

4.5.4 Anotação manual das ESTs

Para *T. serrulatus*, os agrupamentos e sequências únicas providas pelo programa TGICL foram submetidas ao programa BLASTx contra o banco de dados do UniProt. Essa análise foi realizada em março de 2011 e os parâmetros usados foram: e-value de corte de 1×10^{-10} e filtro de baixa complexidade desligado. Os dez melhores resultados de BLASTx

foram usados para fazer a anotação manual. A anotação foi realizada a partir dos unives resultantes do agrupamento uma vez que essas sequências apresentam tamanho maior do que as ESTs que os formam, o que melhora os resultados de anotação. Para facilitar essa etapa, os resultados da busca realizada pelo BLASTx foram transferidos para o sistema de gerenciamento de bancos de dados MySQL. Em seguida, realizaram-se contagens de número de transcritos e de agrupamentos correspondentes a cada anotação. Para comparação do nível de expressão entre os transcriptomas, o número de ESTs para todos os catálogos de transcritos foi padronizado para 100.000 ESTs usando regra de três. Como exemplo, no transcriptoma de *T. serrulatus* foram encontradas 18 ESTs codificando peptídeo aniônico em um total de 1.629 transcritos analisados, em 100.000 ESTs, peptídeos aniônicos corresponderiam a 1.105 ESTs. A conversão de todos os catálogos de transcritos para 100.000 ESTs permitiu, portanto, fazer comparações entre o nível de expressão de determinados genes entre eles.

4.5.5 Via KEGG

O conjunto das ESTs de *T. serrulatus* e os unives provenientes do seu agrupamento foram submetidos ao mapeamento através de KEGG PATHWAY (<http://www.genome.jp/kegg/pathway.html>). Os dados foram organizados em ordem decrescente de ocorrência das vias bioquímicas apresentadas pelo programa.

4.5.6 Anotação funcional usando a ferramenta Blast2GO

As ESTs trimadas e os unives de cada espécie em formato fasta foram utilizados como sequências de entrada do programa Blast2GO. A típica análise feita pelo Blast2GO é realizada em cinco etapas: (1) busca por similaridade usando algoritmo BLAST, (2) mapeamento, (3) anotação, (4) análise estatística e (5) visualização.

Para a realização da primeira etapa optou-se pela utilização de BLASTx com busca no Banco de Dados nr do NCBI e com um e-value de corte de 1×10^{-10} . Para os demais parâmetros oferecidos pelo programa foram utilizados os valores default, que consistem em análise dos 20 primeiros hits, comprimento mínimo de alinhamento de 33 aminoácidos e filtro de baixa complexidade ativado. Para realizar a pesquisa BLAST no NCBI fez-se o

uso do QBLAST (Queue BLAST é um serviço oferecido pelo NCBI para realizar busca para múltiplas sequências).

O mapeamento realizado pelo Blast2GO é o processo de busca de termos do GO associados com os hits obtidos após o BLAST. Essa etapa não possibilita a configuração de parâmetros e é realizada de quatro diferentes formas: (1) os acessos do resultado do BLAST são usados para coletar nomes dos genes ou símbolos por fazer o uso de dois arquivos de mapeamento fornecidos pelo NCBI (gene info, gene 2accession). Identificados os nomes dos genes, é realizada pesquisa com entradas espécie-específicas da tabela de produtos gênicos do banco de dados do Gene Ontology (GO). (2) Os identificadores GI obtidos durante a pesquisa BLAST são usados para busca de IDs no UniProt fazendo o uso do arquivo de mapeamento PIR (referência do Banco de dados de proteína não-redundante) que inclui PSD, UniProt, Swissprot, TrEMBL, RefSeq, GenPept e PDB. (3) Os acessos são pesquisados diretamente na tabela dbxref do banco de dados do GO. (4) Os acessos dos resultados de BLAST são pesquisados diretamente na tabela de produtos gênicos do banco de dados do GO.

A etapa de anotação é o processo de associação de um termo funcional a uma sequência de entrada a partir de um conjunto de termos obtidos na etapa de mapeamento. O algoritmo de anotação leva em consideração a similaridade entre a sequência de entrada e o hit, a qualidade da fonte de assinaturas do GO e a estrutura do gráfico direto acíclico (DAG) do GO. A anotação com termos do GO é realizada através da aplicação de uma regra de anotação (Quadro 2). A regra busca descobrir a anotação mais específica com certo nível de confiabilidade. Para cada candidato GO uma pontuação (score) de anotação (AS) é computada. O AS é computado a partir de dois termos aditivos: (1) o DT, que representa o hit de maior similaridade do GO multiplicado por um fator correspondente ao peso do seu código de evidência (EC); e o AT, que fornece uma possibilidade de abstração. Esse é definido como a anotação de um nó parental quando vários nós derivados estão presentes na coleção de candidatos GO. Esse termo multiplica o número de GOs unificados em um nó por um peso definido pelo usuário que controla a possibilidade e extensão da anotação. Quando o peso do GO é 0, nenhuma abstração é realizada. Finalmente, a regra de anotação (RA) seleciona o menor termo por braço que obedece ao limite definido pelo usuário. Para compreender como a pontuação de anotação funciona pode-se seguir o seguinte raciocínio: quando o peso EC é 1 para todos os ECs (sem influência do EC) e o peso do GO é zero (nenhuma abstração), então o AS é igual ao valor de máxima

similaridade do hit que possui o termos do GO e a sequência será anotada com aquele termo se o score é acima de um dado limiar. A situação quando o peso de EC é menor que 1 significa que maiores similaridades são necessárias para alcançar o limiar. Se o peso GO é diferente de 0, isso significa que há possibilidade que um nó parental alcance o limiar enquanto seus vários nós derivados não alcançaram. A anotação depende da configuração dos diferentes parâmetros que compõe o AS. Tais parâmetros podem ser ajustados no “Diálogo de Configuração” e na “Configuração dos pesos dos Códigos de Evidência” oferecidos pelo programa.

$DT = \max(\text{similaridade} \times \text{peso EC})$ $AT = (\# \text{ GO} - 1) \times \text{peso GO}$ $\text{RA: menor nó [AS (DT + AT) \geq \text{limiar}]}$

Quadro 2: Regra de anotação (RA) do programa Blast2GO.

Os parâmetros ajustáveis são:

1. Filtro do E-Value do Hit: apenas termos do GO obtidos do hit com e-value menor que um determinado limiar são usados para anotação. Nesse estudo optou-se pelo uso de e-value de 1×10^{-10} .
2. Limiar de anotação. A regra de anotação seleciona o menor termo por ramo que se adequa acima do limiar. No presente estudo foi usado o parâmetro default que é igual a 55.
3. Peso do GO. Esse é o peso dado pela contribuição dos termos derivados mapeados para a anotação de termos parentais (default=5).
4. Cobertura de corte do Hsp-Hit: Corresponde a necessidade de cobertura mínima entre o hit e seu HSP. Por exemplo, um valor de 80 significa que o HSP alinhado deve cobrir no mínimo 80% da extensão do seu hit. Apenas anotações de hits que preencham esse critério irão ser consideradas para a transferência de anotação.
5. Peso de EC: podem-se alterar os pesos dos códigos de EC. Por exemplo, quando não se espera a influência dos códigos de evidência, pode-se adotar o valor 1 para todos eles. Por outro lado, quando se quer excluir um certo EC, atribui-se um valor 0 ao seu peso.

De acordo com Conesa e Götz (2008), os valores default dos parâmetros do Blast2GO foram escolhidos para otimizar a taxa entre a cobertura de anotação e a acurácia da mesma. Nesse sentido, optou-se pela manutenção dos parâmetros definidos pelo programa, exceto para o e-value de anotação.

Após as etapas de BLAST, mapeamento e anotação, os gráficos, tabelas e organogramas fornecidos pelo programa foram analisados. Para os dados de distribuição dos termos do GO fornecidos pelo programa optou-se por utilizar as tabelas com dados brutos ao invés dos gráficos fornecidos, uma vez que isso possibilitou maior flexibilidade de formatação para apresentação dos dados. Comparações entre os transcriptomas dos diferentes grupos de organismos amostrados foram realizadas.

4.5.7 Busca de ortólogos e alinhamentos

As ESTs relacionadas aos componentes do veneno foram separadas e sua sequência foi traduzida em sequência de proteína usando o programa ORF Finder (<http://www.ncbi.nlm.nih.gov/projects/gorf/>). Para prever o peptídeo sinal das ORFs encontradas foi utilizado o programa SignalP 4.0 (Petersen *et al.*, 2011). O presente trabalho seguiu a nomenclatura dos componentes do veneno sugeridas por Cologna *et al.* (2010) and King *et al.* (2008) e o UniProt. Subsequentemente, o programa Seed Server foi usado para buscar ortólogos dos componentes do veneno de *T. serrulatus*. A metodologia usada pelo Seed Server (Guedes *et al.*, trabalho não publicado) consiste basicamente no agrupamento de proteínas homólogas de interesse usando o software Seed Linkage (Barbosa-Silva *et al.*, 2008) e o UniRef50, enriquecido com grupos de ortólogos do KEGG (UEKO) construídos com procedimentos descritos por Fernandes *et al.* (2008). As sequências usadas para alinhamento foram obtidas do banco de dados do UniProt. O alinhamento múltiplo foi realizado pelo programa online webPrankster <http://www.ebi.ac.uk/goldman-srv/prank/> e o programa Jalview foi usado para a edição e visualização dos alinhamentos (<http://www.jalview.org>).

4.5.8 Busca de ortólogos distantes: uso de PSI-BLAST e PSI-tBLASTn

Para identificar ortólogos mais distantes de componentes do veneno encontrados em *T. serrulatus* e analisar em quais grupos de organismos tais sequências são encontradas,

35 uniques que codificam proteínas presentes no veneno foram submetidos ao programa PSI-BLAST contra UniProt e ao PSI-tBLASTn contra o banco de dados dbEST.

Na avaliação de PSI-BLAST contra o UniProt usou-se como parâmetros e-value de 1×10^{-10} , entretanto, para algumas sequências esse e-value foi muito baixo e não foi possível encontrar sequências homólogas. Nesses casos, utilizou-se valor de e-value de 1×10^{-5} . Para evitar a degradação da matriz durante a análise de PSI-BLAST contra UniProt, o número de iterações foram limitados da seguinte forma: para cada componente do veneno, escolheu-se a maior iteração que ainda mantivesse o valor de escore maior que 60% daquele obtido quando a sequência foi usada como entrada contra ela mesma.

Para utilizar o programa PSI-tBLASTn, foi contruída uma matriz para cada uma dos uniques usados como sequência de entrada (*query*). Essa matriz foi obtida a partir das sequências de proteínas da iteração escolhida durante a análise usando PSI-BLAST contra o UniProt. Nessa análise também se utilizou e-value de corte de 1×10^{-10} e quando não se obteve resultados usando esse valor testou-se o e-value de 1×10^{-5} .

Realizou-se a busca do identificador taxonômico de cada sequência resultante da busca realizada pelo PSI-BLAST e PSI-tBLASTn. Cada conjunto de identificadores relacionado a um componente do veneno foi usados como input para o programa online 'Taxonomy common tree' (<http://www.ncbi.nlm.nih.gov/Taxonomy/>), que controla árvores contendo os organismos correspondentes aos identificadores.

4.5.9 Análise comparativa entre transcriptomas de quelicerados

Para avaliar a proporção de transcritos comuns entre *T. serrulatus* e demais transcriptomas realizou-se análise usando tBLASTx. Nesse programa, o transcriptoma de *T. serrulatus* foi usado como entrada contra os catálogos de transcritos dos demais quelicerados. Em seguida, calculou-se a proporção que as sequências em comum ocupavam em cada um dos transcriptomas. Por exemplo, os uniques de *T. serrulatus* apresentaram *match* com 235 ESTs e 89 uniques de *L. mucronatus*. Como o transcriptoma de *L. mucronatus* possui 740 ESTs e 404 uniques, a proporção de *match* correspondeu a 32% das ESTs e 22% dos uniques de *L. mucronatus*. Os transcritos em comum foram identificados e anotados.

Os transcritos que codificaram componentes do veneno de *T. serrulatus* foram usados como sequência de entrada no programa tBLASTx contra os uniques dos demais

transcriptomas. O número de resultados obtidos para cada componente do veneno foi computado e o número de EST encontrado foi convertido para 100.000 ESTs para possibilitar a comparação do nível de expressão de tais componentes do veneno entre os transcriptomas.

Os uníques codificantes de componentes de veneno traduzidos com o auxílio do programa ORF-FINDER (<http://www.ncbi.nlm.nih.gov/projects/gorf/>) foram usados como entrada no programa tBLASTn contra os outros 12 transcriptomas de quelicerados. Foram feitas contagens do número de resultados obtidos e esses foram convertidos para 100.000 ESTs para facilitar a comparação entre os transcriptomas amostrados.

4.6 Histologia da glândula de veneno

Para a histologia da glândula de veneno de *T. serrulatus* foram utilizados 12 animais. Os escorpiões foram coletados na cidade de Belo Horizonte, Minas Gerais, Brasil. Os animais foram narcotizados com clorofórmio e em seguida o telson foi retirado, cortado transversalmente e fixado por imersão em glutaraldeído 5% tamponado com tampão fosfato 0,05 M (pH=7,2) por 24 h. Após a fixação, as amostras foram desidratadas em séries crescentes de etanol e diafanizadas em xilol. Em seguida, utilizaram-se duas técnicas distintas: 6 amostras foram incluídas em parafina e coradas com hematoxilina-eosina e as outras 6 foram incluídas em resina glicol-metacrilato e coradas com azul de toluidina 1%. As lâminas foram examinadas no microscópio de luz da Olympus BX60 (Tokyo, Japão).

5 RESULTADOS

5.1 Sequenciamentos dos clones da biblioteca de *T. serrulatus*

Oteve-se sucesso na transformação dos clones da biblioteca, extração e purificação dos plasmídeos. Além de ser mais rápida, a prep realizada em placas resultou em DNA purificado de alta qualidade (Fig. 5) e com quantidade suficiente para o sequenciamento.

Durante a realização do trabalho, obtiveram-se 1603 seqüências de clones da biblioteca de *T. serrulatus*. As seqüências obtidas foram associadas com outros sequenciamentos da biblioteca realizados anteriormente no Laboratório de Biotecnologia e Marcadores Moleculares, o que resultou em um número total de 2344 seqüências. Algumas delas consistiam-se de repetições do sequenciamento do mesmo clone, de modo que, eliminadas essas repetições, obtiveram-se 2.131 seqüências provenientes de diferentes clones.

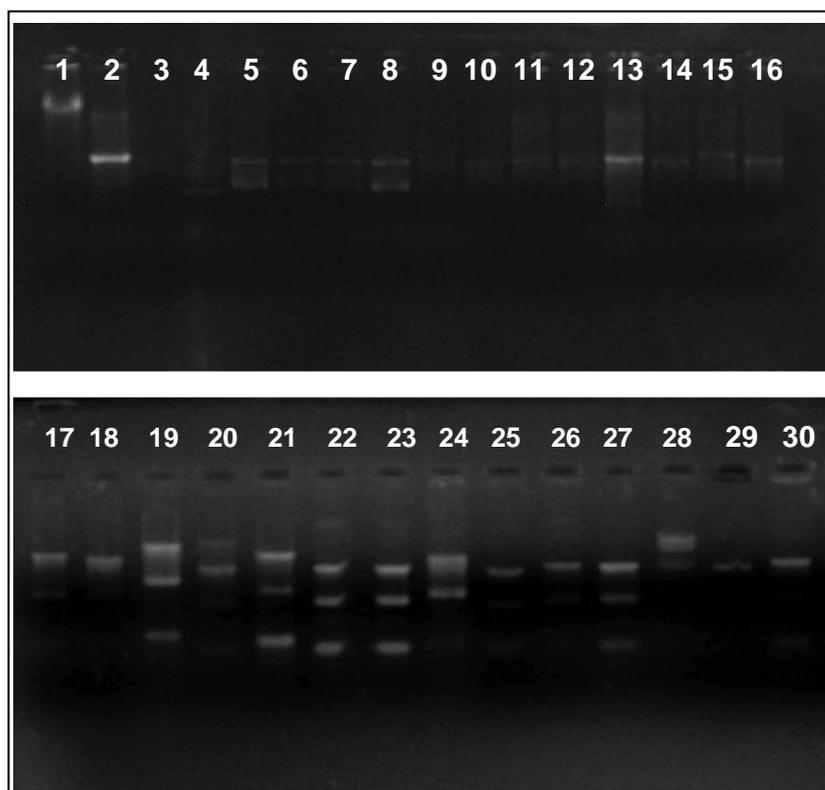


Figura 5: Gel de Agarose 0,8% corado com GelRed demonstrando plasmídeos provenientes da biblioteca de *T. serrulatus* após amplificação, extração e purificação em placa. Canaleta 1 = DNA Lambda 458 ng/µL. Canaletas de 2 a 30 representam plasmídeos provenientes de diferentes clones.

5.2 Análise de qualidade e agrupamento das sequências de *T. serrulatus*

Após a avaliação de qualidade realizada pelo programa Phred (com valor de Phred = 20) e análise das sequências com o programa SeqClean obtiveram-se 1629 ESTs de excelente qualidade. Essas ESTs foram depositadas no dbEST (75926766-75928119). O agrupamento das sequências através do programa TGICL resultou em 643 uniques, sendo eles formados por 185 contigs e 458 singlets. Cada agrupamento foi manualmente avaliado quanto à qualidade das sequências e do consenso formado pelas ESTs.

A distribuição do tamanho das ESTs e dos contigs em número de nucleotídeos é mostrada na Figura 6A. Como se pode observar o agrupamento aumentou o tamanho médio das sequências, o que geralmente resulta em melhora no processo de anotação. A análise de ocorrência de singlets e do número de EST incluídas nos contigs (Fig. 6B) mostrou prevalência de singlets (458) e agrupamentos com pequeno número de sequências (165 contigs possuíam entre 2 e 9 sequências), como também observado em outros estudos de transcriptomas de glândula de veneno (Schwartz *et al.*, 2007; D'Suze *et al.*, 2009; Ma *et al.*, 2009; Silva *et al.*, 2009; Roeding *et al.*, 2009; Ma *et al.*, 2010; Ruiming *et al.*, 2010). Foi identificado apenas um agrupamento com 138 ESTs, dois contigs contendo entre 50-99 sequências, sete possuindo entre 20-49 transcritos e dez com 10 a 19 ESTs.

5.3 Anotação das ESTs de *T. serrulatus*

Foram utilizados dois procedimentos de anotação: anotação manual e anotação funcional utilizando como ferramenta o programa Blast2GO. Os resultados obtidos com esses dois processos estão descritos separadamente abaixo.

5.3.1 Anotação manual das ESTs e uniques de *T. serrulatus*

Considerando o e-value de corte menor que 1×10^{-10} , das 1629 sequências analisadas, 370 delas não apresentaram resultado de BLASTx contra o UniProt, o que correspondeu a 295 uniques. Conforme mostrado na Figura 7, a maioria das sequências que não apresentaram *match* foram singlets ou agrupamentos com poucas sequências. As

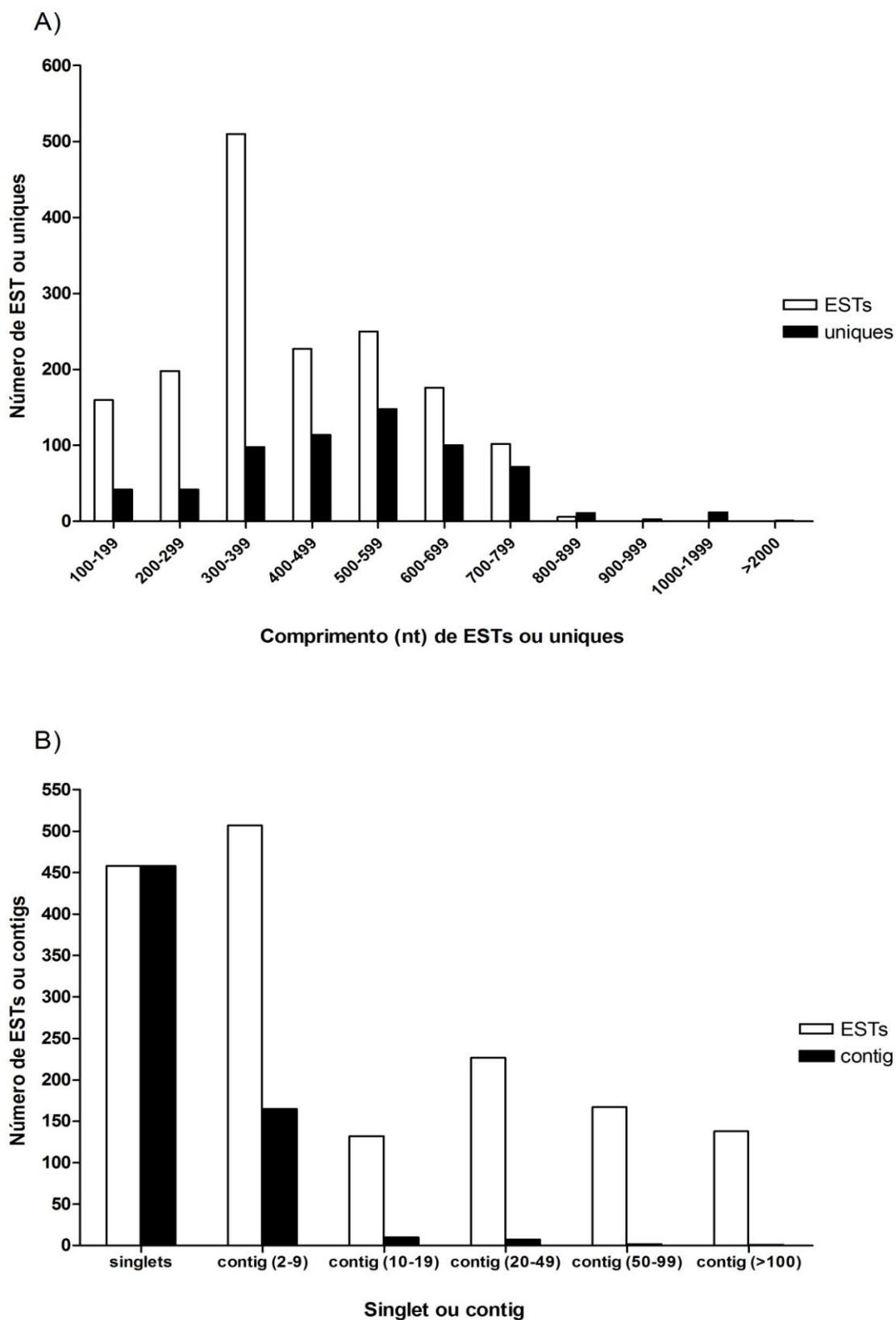


Figura 6: Informações sobre o agrupamento e tamanho das sequências de *T. serrulatus* analisadas. **A)** Distribuição do tamanho em número de nucleotídeos das ESTs e dos únicos; **B)** Ocorrência de singlets e distribuição do número de contigs contendo determinado número de EST.

sequências sem resultado de similaridades a outras depositadas no banco de dados corresponderam cerca de 46% dos únicos e 23% do total de ESTs. Essa proporção de sequências “*no match*” é comumente encontrada em estudos de transcriptoma de escorpião e pode indicar um grande potencial de descoberta de novas moléculas. Além disso,

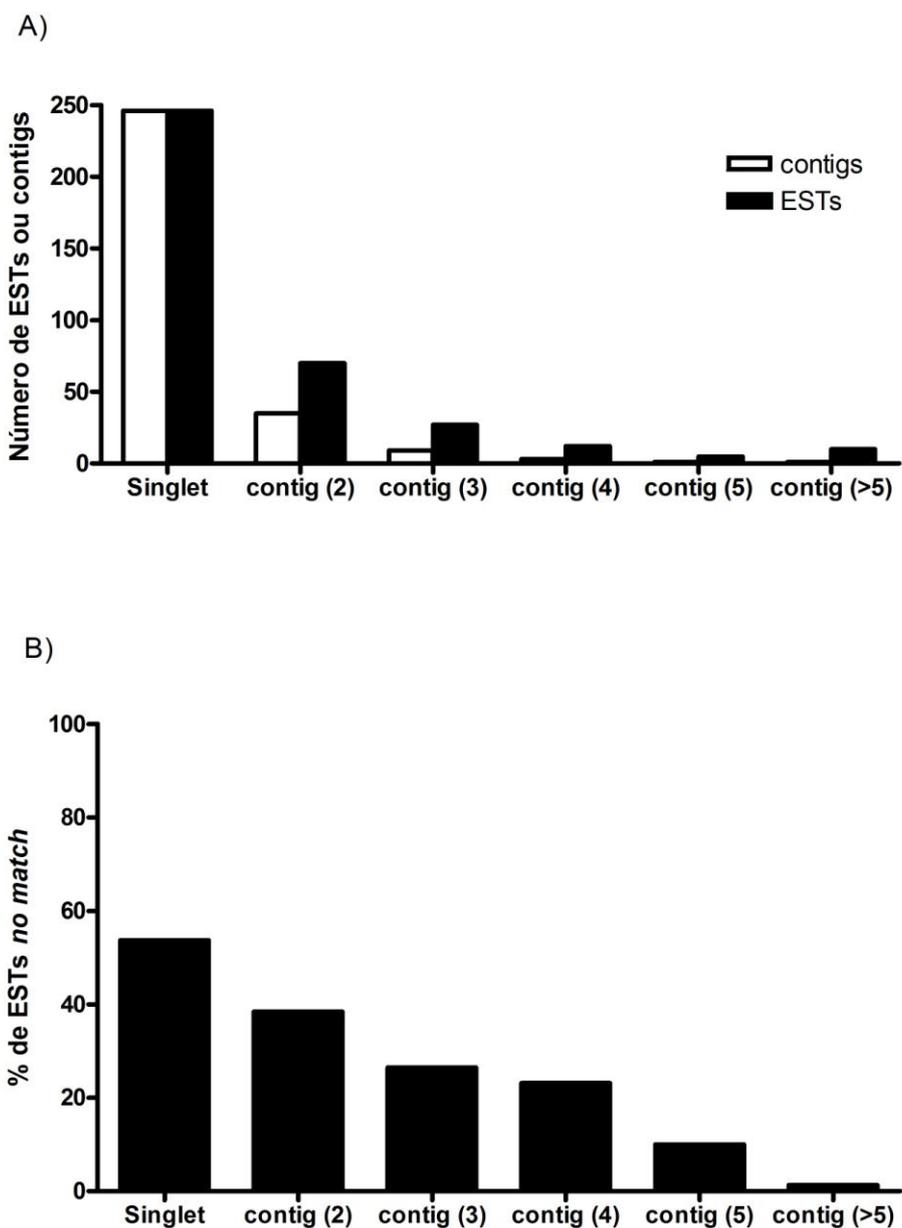


Figura 7: Distribuição dos transcritos sem resultado de similaridade com sequências depositadas no UniProt. **A)** Número de ESTs e únicos consideradas *no match*. A anotação ‘contig (2)’ significa que o agrupamento que possui 2 sequências. **B)** Percentual de ESTs sem resultado de BLAST contra o UniProt em agrupamentos compostos por diferentes números de ESTs. Observar que a predominância de ‘no match’ ocorre para singlets ou para agrupamentos com pequeno número de sequências.

considerável proporção de ESTs (cerca de 10%) corresponderam sequências similares a proteínas preditas ou não caracterizadas, o que indica que a ocorrência de *match* não significa que a função da proteína seja conhecida.

De acordo com a anotação manual realizada a partir dos 10 melhores hits, as sequências foram classificadas em proteínas similares a componentes celulares ou do veneno e em subdivisões dessas classes. A Figura 8 (A e B) mostra a distribuição dos resultados dessa anotação considerando a análise feita por EST (1629 transcritos) e por uniques (643 sequências) e destaca as sequências relacionadas aos componentes celulares. Na classificação das ESTs e uniques foram consideradas as seguintes categorias: ‘Veneno’ incluíram transcritos codificando toxinas e outros componentes secretados já descritos no veneno de escorpiões. As categorias ‘Mitocôndria’ e ‘Núcleo’ compreendem ESTs/uniques codificando proteínas localizadas nessas organelas. ‘Tradução e processamento protéico’ contém transcritos que codificaram, por exemplo, proteína ribossomal, dissulfito-isomerase e outras proteínas relacionadas a síntese protéica. ‘Componentes estruturais’ incluíram principalmente proteínas do citoesqueleto, tais como actina, miosina e tubulina. ‘Transporte’ compreende transcritos codificando proteínas envolvidas com o tráfego intracelular (por exemplo, proteínas de transporte *Copper*). ‘Proteína predita ou não caracterizada’ inclui ESTs/uniques similares a sequências previamente descritas, mas sem informação funcional. ‘Extracelular’ envolve transcritos codificando proteínas extracelulares que são, por exemplo, encontradas na matriz extracelular, como a fibronectina. A categoria ‘outros componentes celulares’ inclui uma diversidade de ESTs codificando componentes celulares que não foram incluídos nas categorias descritas acima. Na distribuição dos componentes relacionados aos constituintes do veneno mais abundantes (Fig. 8C e 8D) os transcritos que codificaram peptídeo Pape (similar a Peptídeo Potencializador de Bradicinina ou BPP), NaTx, KTx e AMPs foram grandes destaques, enquanto que as categorias NaTx, KTx, metaloproteases e sequências similares a peptídeos do veneno de outros escorpiões destacaram-se em termos de diversidade de transcritos.

5.3.2 ESTs relevantes para processos celulares

ESTs e uniques relacionados aos componentes celulares corresponderam a 31% dos transcritos (501 ESTs) e 41% dos uniques (265 uniques). Para facilitar comparações da expressão gênica, optou-se por padronizar o número de EST obtido para 100.000 ESTs

(ESTs/100.000). Dessa forma, os transcritos codificando componentes celulares representariam 30.755 ESTs/100K. Quando se considerou apenas as ESTs que apresentaram similaridade com sequências no UniProt, os componentes celulares corresponderam aproximadamente 40% das ESTs.

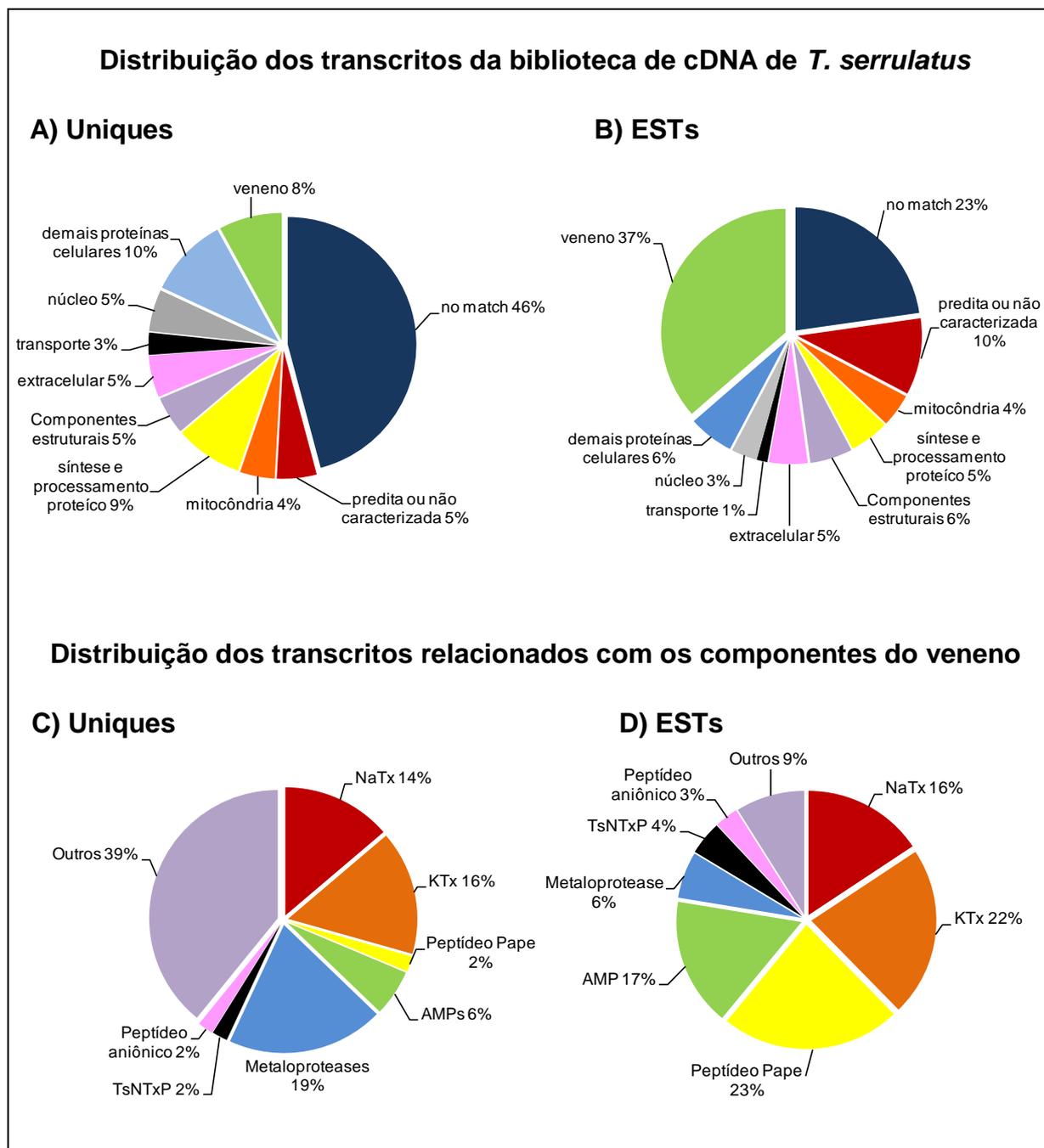


Figura 8: Distribuição dos transcritos em diferentes categorias de acordo com a anotação manual das sequências de *T. serrulatus*. Distribuição dos 643 unives (A) e das 1629 ESTs (B) em diferentes categorias com destaque para transcritos com similaridade a componentes celulares. Distribuição dos 51 unives (C) e das 594 ESTs (D) relacionados aos constituintes do veneno, com destaque aqueles que constituíram mais que 2% dos transcritos. NaTx = toxina de canal para sódio; KTx = toxina para canal para potássio; Metaloproteases incluiu zinco metaloproteases e antareases; TsNTxP = proteína não tóxica NTxP.

Os transcritos mais frequentes relacionados com a síntese e processamento proteico foram ESTs codificando proteínas ribossomais (representando 2.271 ESTs/100K), fator de iniciação (675 ESTs/100K) e alongação (430 ESTs/100K) da tradução, dissulfeto isomerase (246 ESTs/100K) e proteínas de choque térmico (HSPs – 246 ESTs/100K). Além dessas ESTs mais abundantes, foi encontrada uma EST codificante de PAM (peptidylglycine α -amidating monooxygenase - EC 1.14.17.3). Essa enzima catalisa a reação de amidação da porção C-terminal de peptídeos (Eipper *et al.*, 1993). Alguns peptídeos necessitam dessa modificação para exercerem função biológica (Jiang *et al.*, 2000), sendo que vários autores reportam a amidação C-terminal em peptídeos do veneno de escorpiões (Martin-Eauclaire *et al.*, 1994; Rates *et al.*, 2008; Diego-García *et al.*, 2005; Barona *et al.*, 2006).

Sete ESTs (representando 430 ESTs/100K) codificaram proteínas que foram relacionadas a degradação através do sistema ubiquitina-proteasoma. Esses peptídeos podem ser importantes para assegurar que proteínas anormais ou com erros de conformação sejam reconhecidas e eliminadas. Além disso, a degradação de proteínas através desse sistema pode estar relacionada com a regulação de importantes vias na glândula de veneno.

Foram obtidos abundantes componentes estruturais expressos na glândula de veneno de *T. serrulatus*, que são provavelmente envolvidos na manutenção da sua estrutura e/ou participam da atividade contrátil durante a liberação do veneno. Os transcritos mais abundantes relacionados a essa função codificaram actina (982 ESTs/100K), paramiosina (921 EST/100K), troponina (675 ESTs/100K), miosina (491 ESTs/100K), alfa-tubulina (430 ESTs/100K). Como referência, no transcriptoma de *Drosophila melanogaster* adulto, o grupamento referente a paramiosina no UniGene Dm.2704 apresentou 855 ESTs/ 100K.

A síntese do veneno é um processo considerado energeticamente custoso. De fato, cerca de 6% das ESTs e 8% dos unives que apresentaram resultados de BLASTx contra o UniProt foram transcritos mitocondriais. Os transcritos mais abundantes relacionados a produção de energia foram citocromo c oxidase (1.044 ESTs/100K), NADH-ubiquinona oxidoreductase (921 ESTs/100K), ATP sintase (675 ESTs/100K) e citocromo b (368 ESTs/100K).

5.3.3 ESTs relacionadas aos componentes do veneno

Nesse trabalho, 594 ESTs representando 36.464 ESTs/100K, 51 contigs e 35 diferentes famílias de peptídeos ou proteínas foram relacionadas com componentes do veneno. Uma lista desses componentes identificados no transcriptoma de *T. serrulatus* é mostrada na Tabela 4. Os transcritos do veneno mais abundantes codificaram peptídeo Pape (similar a peptídeo potencializador de bradicinina - BPPs – encontrados em outras espécies de escorpiões como *T. discrepans* e *L. mucronatus*), toxinas de canal para potássio (KTx) e toxinas de canal para sódio (NaTx). Grande variação entre o número de ESTs encontradas para cada toxina indica que esses componentes podem ser expressos em níveis consideravelmente diferentes na glândula de veneno.

Foram obtidos 93 transcritos para NaTx (correspondendo 5.709 ESTs/100K), que perfazem 16% das sequências relacionadas a componentes do veneno. Esses peptídeos são responsáveis pela maioria dos sintomas neurotóxicos desenvolvidos durante o envenenamento por *T. serrulatus* (Cologna *et al.*, 2009). Foram encontradas sequências de 5 distintas toxinas para canal de sódio. Ts1 e Ts2 foram as mais abundantes no transcriptoma. Considerando as NaTxs já descritas em *T. serrulatus*, a única não encontrada no transcriptoma foi a Ts5. Entretanto, foi obtida uma sequência, denominada no presente trabalho de Ts17 (U-BUTX-Ts1a) com 95% de identidade com a sequência da TbtX5 de *Tityus bahiensis*. Essa é possivelmente uma nova sequência que possui 86% e 85% de identidade com Ts3 e Ts5, respectivamente. Além dessa sequência, foi obtido um contig contendo 5 ESTs que apresentou apenas um *match* significativo com U1-Buthitoxina-Hj1a (identidade de 63%) obtido de *H. judaicus*. Essa sequência foi chamada de Ts18 ou U-BUTX-Ts1b e é uma provável toxina de canal para sódio. Essas duas novas sequências serão apresentadas juntamente com seus ortólogos na seção 5.6. Embora Ts17 e Ts18 possuem sequências com similaridade a toxinas de canal para Na⁺, ensaios farmacológicos são necessários para verificar suas funções biológicas.

Foram identificados uma grande diversidade de transcritos para KTx. Um total de 131 ESTs correspondendo a 8.042 ESTs/100K foram obtidas no transcriptoma de *T. serrulatus*. Pelo menos cinco diferentes toxinas já conhecidas e uma sequência com possível ação como KTx foram encontradas. Ts8, também chamada TsTXK-beta ou κ -BUTX-Ts1, apresentou alto nível de expressão e seus transcritos corresponderam 12% das ESTs do veneno. O segundo componente mais expresso na categoria das KTx consiste em

uma sequência que contém uma porção de 10 aminoácidos idênticos a um fragmento peptídico depositado no Uniprot denominado toxina de canal para potássio beta-Ktx2 (Uniprot ID: P69940). No presente trabalho, essa sequência aqui denominada Ts19 (U-BUTX-Ts1c) é mostrada na seção 5.6, juntamente com seus ortólogos.

Tabela 4: Distribuição das ESTs e contigs dos componentes do veneno encontrados no transcriptoma de *T. serrulatus*

Componentes do veneno	<i>T. serrulatus</i>	
	ESTs	contigs
Toxinas de canal para Na⁺		
Ts1 (β-BUTX-Ts1a) [P15226]	33	2
Ts2 (δ-BUTX-Ts1a) [P68410]	36	1
Ts3 (δ-BUTX-Ts1b) [P01496]	10	2
Ts17 (U-BUTX-Ts1b) [novo]	9	1
Ts18 (U-BUTX-Ts1c) [novo]	5	1
Toxinas de canal para K⁺		
Ts8 ou TsTXK-beta (κ-BUTX-ts1) [P69940]	73	1
Ts15 (κ-BUTX-ts2) [P86270]	7	2
Ts6 ou alfa-KTx 12.1 (λ-BUTX - Ts1b) [P59936]	5	1
Ts7 ou alfa-KTx 4.1-like (κ-BUTX - Ts1a) [P46114]	4	2
Ts16 (U-BUTX-Ts1a) [P86271]	1	1
Ts19 (U-BUTX-Ts1d) [novo]	41	1
Outros componentes		
Peptídeo Pape ou BPP-like [P86821]	138	1
AMPs [novo]	59	2
Similar a ponicina-W-like de <i>L. mucronatus</i> [novo]	40	1
Antarease * [P86392]	29	7
TsNTxP ou Ts4 [O77463]	26	1
Peptídeo aniônico [novo]	18	1
Hipotensina 1 ou Ts14 [P84189]	11	1
Hipotensina-like [P86824]	10	1
Zinco Metaloproteases *[P85842]	7	3
Similar a proteína rica em glicina de <i>L. mucronatus</i> [novo]	6	3
Tspep3 ou Ts13 [P0C176]	4	1
Similar a neuropeptídeo-3 de <i>Buthus eupeus</i> [novo]	3	1
Peptídeo rico em cisteína, similar a Ts15 [novo]	3	1
Similar a TsPep2 ou Ts12 [P0C175]	2	1
Hialuronidase* [P85841]	2	2
Similar a peptídeo semelhante a toxina de <i>B. eupeus</i> [novo]	2	1
Fosfolipase A2 [novo]	2	1
Similar a peptídeo rico em cisteína de <i>Camponotus floridanus</i> * [novo]	2	1
Similar a peptídeo rico em cisteína de <i>Hottentotta judaicus</i> [novo]	1	1
Tspep1 ou Ts11 [P0C174]	1	1
Alergeno * [P85840]	1	1
Similar a cadeia alfa-1 do peptídeo ativador de lipólise de <i>L. mucronatus</i> [novo]	1	1
Similar a proteína do veneno de <i>B. eupeus</i> [novo]	1	1
Similar a proteína secretada de <i>Opisthacanthus cayaporum</i> [novo]	1	1

* Indica que as sequências traduzidas são fragmentos e não a sequência completa. Entre parêntese utilizou-se a nomenclatura sugerida por King *et al.*, 2008. Entre colchetes estão os identificadores UniProt das sequências que foram depositadas em trabalhos anteriores ou a palavra “novo” para indicar que essa sequência não foi previamente descrita em *T. serrulatus*, ou foi depositada como fragmento (ou seja, como sequência incompleta).

Além de neurotoxinas, uma grande diversidade de transcritos codificando peptídeos e proteínas similares a componentes do veneno já descritos em outros escorpiões foram encontrados no transcriptoma de *T. serrulatus*. Alguns componentes estiveram presentes em pequeno número como fosfolipase A2 e hialuronidases, enquanto outros foram abundantes como peptídeo Pape (correspondendo a 8,5% do total de transcritos e 23% das ESTs relacionadas a componentes do veneno), peptídeos antimicrobianos (AMPs e ponericina – cerca de 17% das ESTs do veneno e 6% do total de transcritos), proteína não tóxica NTxP (TsNTxP- 4,4% das ESTs do veneno e 1,6% dos transcritos), metaloproteases (soma de antareases e zinco metaloproteases corresponderam a 6% das ESTS do veneno e 2,2% do total de ESTs) e peptídeos aniônicos (3% dos transcritos do veneneo e 1% do total de transcritos).

Alguns transcritos identificados no presente trabalho foram similares a peptídeos, proteínas ou transcritos identificados em outras espécies de escorpião, mas ainda não descritos em *T. serrulatus*. Fosfolipase A2, por exemplo, já foi encontrada no transcriptoma da glândula de veneno de outras espécies de escorpião (Schwartz *et al.*, 2007; Ruiming *et al.*, 2010; Ma *et al.*, 2010; Rendón-Anaya *et al.*, 2012) e acredita-se que ela faz parte dos constituintes do veneno. Entretanto, de acordo com o nosso conhecimento, essa proteína ainda não tinha sido descrita no veneno do escorpião amarelo. No presente estudo, foram identificadas duas ESTs que codificam essa proteína. AMPs, peptídeo aniônico, proteína rica em glicina e peptídeos ricos em cisteínas são outros exemplos de sequências identificadas no presente estudo, que ainda não tinham sido descritas em *T. serrulatus*.

5.3.4 Anotação funcional das sequências de *T. serrulatus* usando a ferramenta Blast2GO

As 1629 ESTs de *T. serrulatus* e os 643 unives em formato fasta foram utilizados como arquivos de entrada do programa Blast2GO. As informações oferecidas por esse programa em relação ao conjunto de sequências analisadas e o processo de anotação estão apresentadas no Apêndice A. Assim, gráficos da distribuição do tamanho das ESTs e unives e a distribuição da similaridade e do e-value obtido na análise de BLASTx contra nr são mostrados nos Apêndices A1 e A2, respectivamente. No apêndice A também são mostrados gráficos que indicam os bancos de dados usados para o mapeamento das

sequências (Apêndice A3), a distribuição dos códigos de evidência para os *hits* do BLAST (Apêndice A4) e por sequência (Apêndice A5), distribuição do número de anotações (Apêndice A6), porcentagem de sequências de comprimento x anotadas e número de termos do GO para sequências de comprimento x (Apêndice A7), distribuição do score de anotação (Apêndice A8), distribuição dos níveis do GO (Apêndice A9), distribuição dos resultados do Interproscan (Apêndice A10).

Informações gerais sobre as etapas realizadas pelo programa Blast2GO são mostradas na Figura 9, que descreve a proporção de sequências sem resultado de BLAST, mapeamento, e anotação para ESTs e uniques. A porcentagem de sequências com resultado de BLAST e com anotação foi maior para as ESTs do que para os uniques, provavelmente devido à ocorrência de ESTs redundantes.

O programa Blast2GO controla gráficos indicando as espécies de organismos cujas sequências foram *hits* (Fig. 10) e *top-hits* (Fig. 11) para as ESTs e uniques analisados. Interessantemente, a distribuição dos *hits* e dos *top-hits* por espécie foram diferentes para ESTs e para uniques. Ocorreu predominância de espécies de escorpião para a distribuição de *hits* e *top-hits* das ESTs. Nesse sentido, nas análises de ESTs, considerando as 10 espécies mais representativas, 7 delas são de escorpiões quando se considera todos os hits e 8 delas pertencem a esse grupo na análise dos *top-hits*. Quando se analisa os uniques, os resultados de similaridade entre as sequências de *T. serrulatus* e outras espécies estiveram mais relacionados a organismos modelos ou espécies de interesse humano. Assim, espécies como *Mus musculus*, *Aedes aegypti*, *Homo sapiens*, *Ixodes scapularis* e *Drosophila melanogaster* apareceram entre as 10 primeiras espécies listadas na análise de uniques. Isso indica que a amostragem de sequências dos bancos públicos influencia nos resultados da análise de similaridade. Além disso, os dados de ESTs refletem a expressão gênica e indicam a predominância de sequências típicas de escorpião na glândula de veneno dos mesmos.

O programa Blast2GO oferece duas opções de gráficos de pizza para a análise da distribuição dos termos do GO: (1) distribuição das sequências por nível do GO; (2) gráfico contendo múltiplos níveis, que é gerado com o menor nó de uma ramificação que se enquadra em uma determinada condição de filtro (que pode ser um determinado número de sequência ou um valor de score escolhido pelo usuário). Os gráficos multi-níveis para os três domínios do GO são mostrados nas Figuras 12, 13, 14 e aqueles que representam apenas os termos do nível 2 estão nas Figuras 15, 16 e 17. Quanto aos componentes

celulares, grande proporção de sequências foram anotadas com o termo “região extracelular”. Além deste termo, aqueles relacionados ao citoesqueleto, mitocôndria, membrana e ribossomo foram consideravelmente frequentes. Quanto ao processo biológico, produtos gênicos assinados com os termos patogênese (definido como série de processos específicos que geram a habilidade de um organismo causar doença em outro), resposta de defesa a bactéria, transporte, regulação da pressão sanguínea, foram consideravelmente frequentes. Conforme esperado, em relação à função molecular, grande proporção de sequências foi anotada como atividade inibidora de canal iônico, especialmente para os canais de sódio e potássio. Sequências anotadas com os termos atividade oxidoreductase (relacionada à respiração celular), ligação a DNA, RNA e citoesqueleto e constituinte estrutural de ribossomos estiveram consideravelmente presentes. Entretanto, a proporção dos mesmos termos nas duas análises foi consideravelmente diferente, sendo que a análise das ESTs fornece informações mais claras sobre a expressão na glândula de veneno.

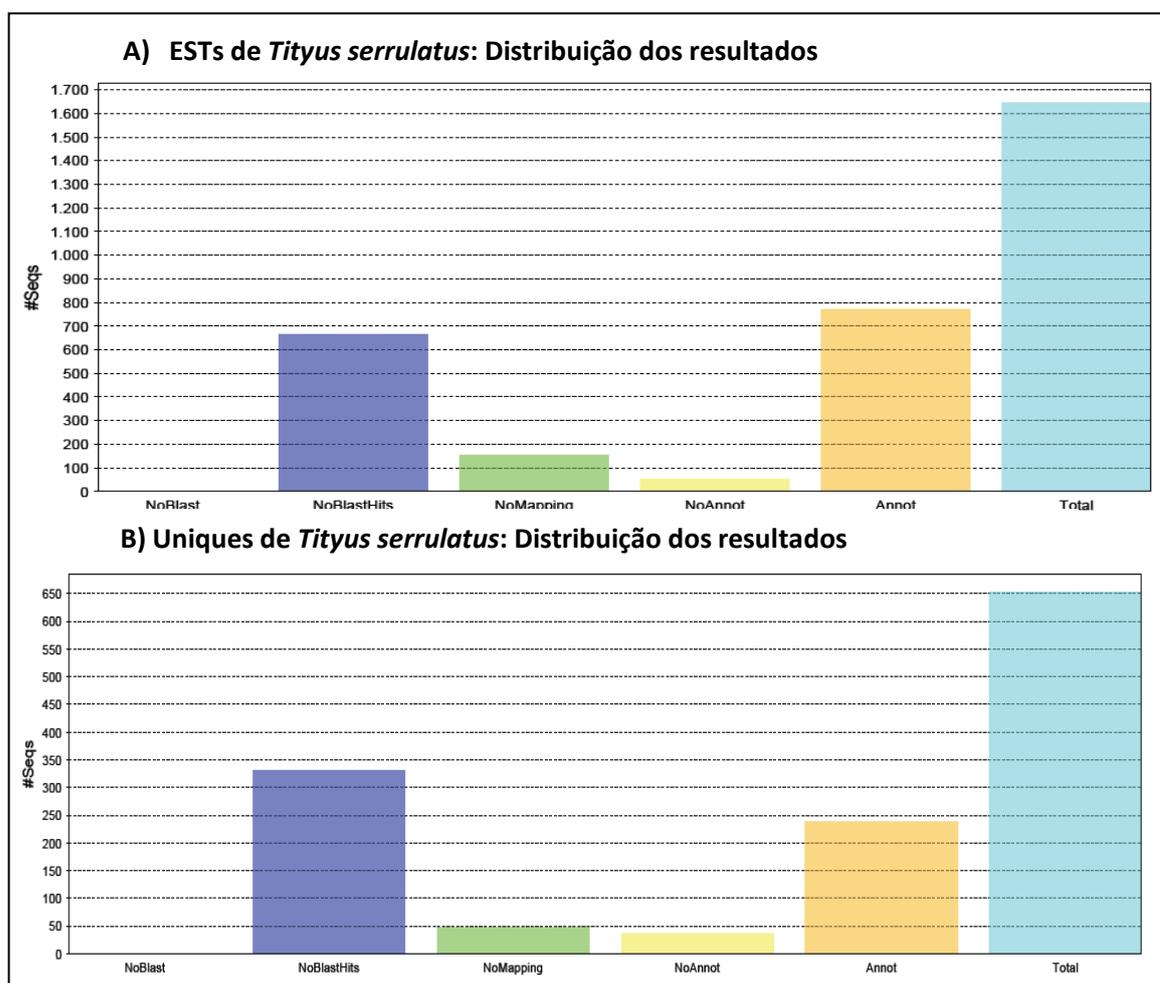


Figura 9: Distribuição dos resultados das diferentes etapas realizadas pelo programa Blast2GO: BLAST, mapeamento e anotação. **A)** análise de todas as ESTs e **B)** avaliação dos uniques.

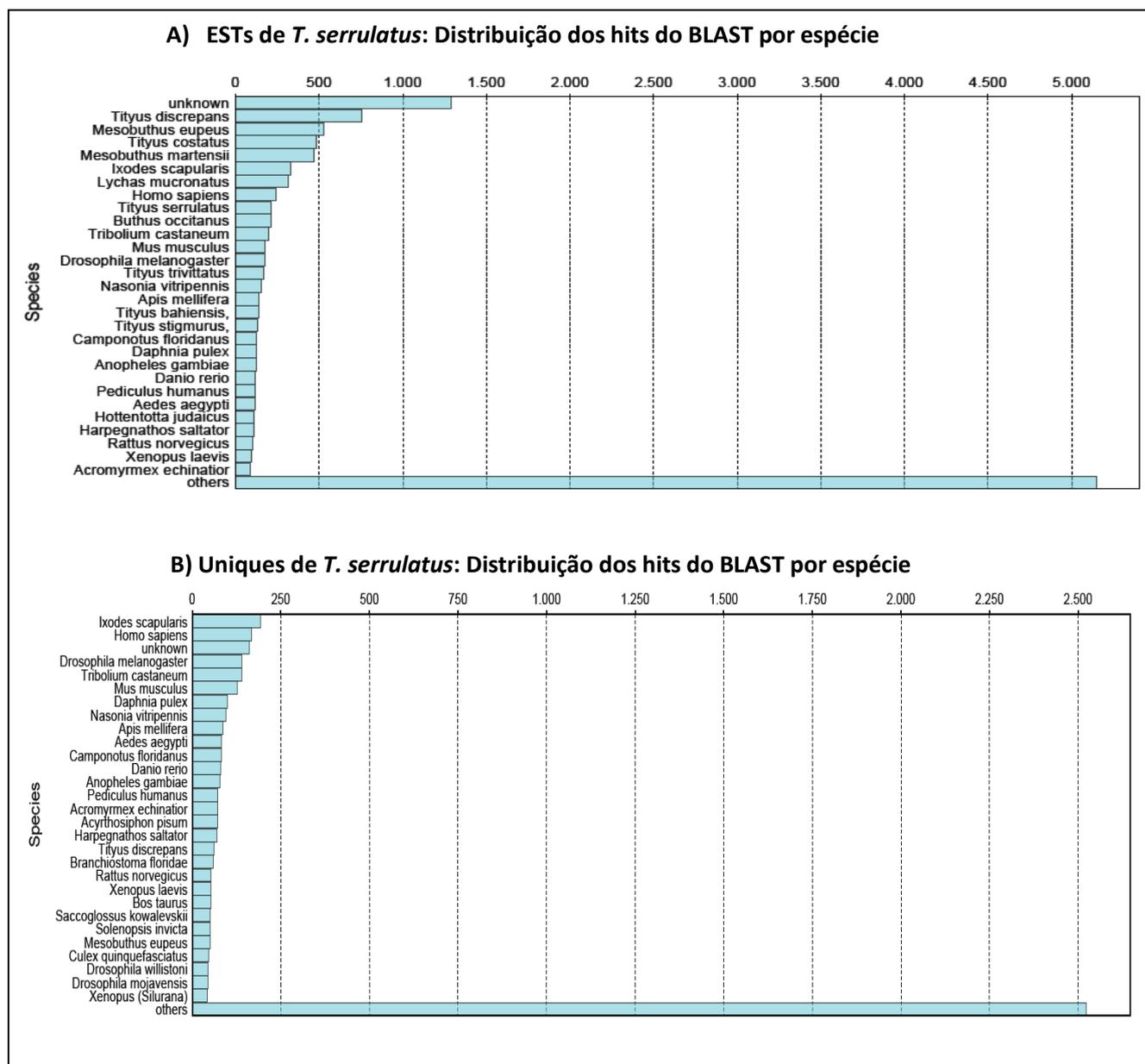


Figura 10: Estatísticas do BLAST - distribuição dos *hits* obtidos do resultado do BLAST por espécie. **A)** dados obtidos pela análise de todas as ESTs e **B)** dados dos *uniques*.

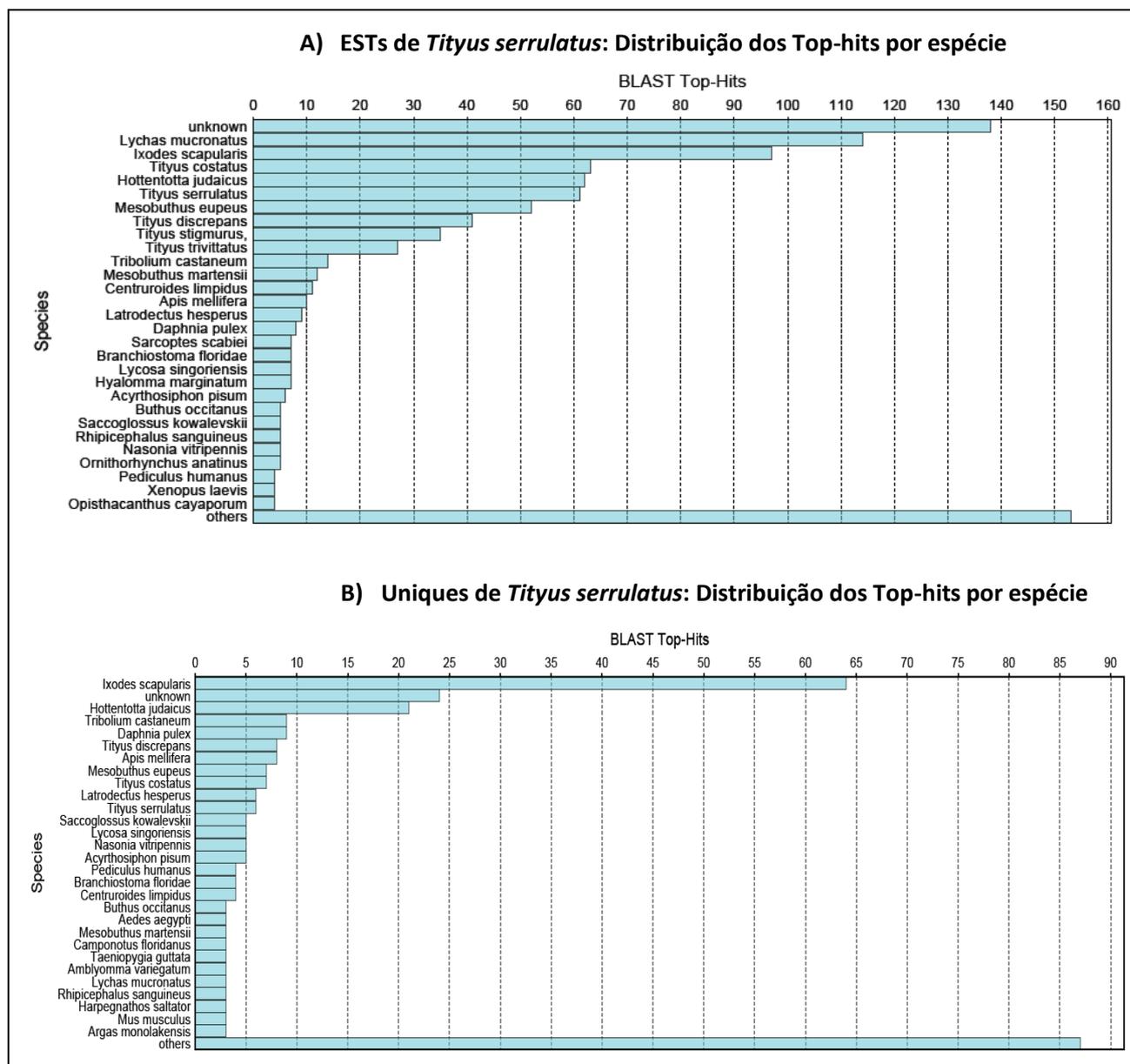
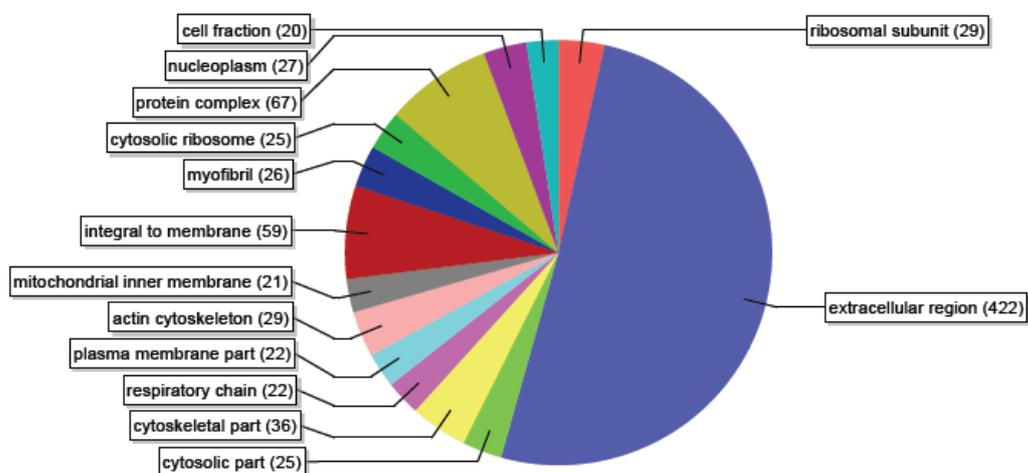


Figura 11: Estatísticas do BLAST: distribuição por espécie do melhor hit obtidos através de pesquisa BLAST. **A)** dados obtidos pela análise de todas as ESTs e **B)** dados dos únicos.

A) Distribuição das ESTs por termo do GO referente ao domínio componente celular



B) Distribuição dos unives por termo do GO referente ao domínio componente celular

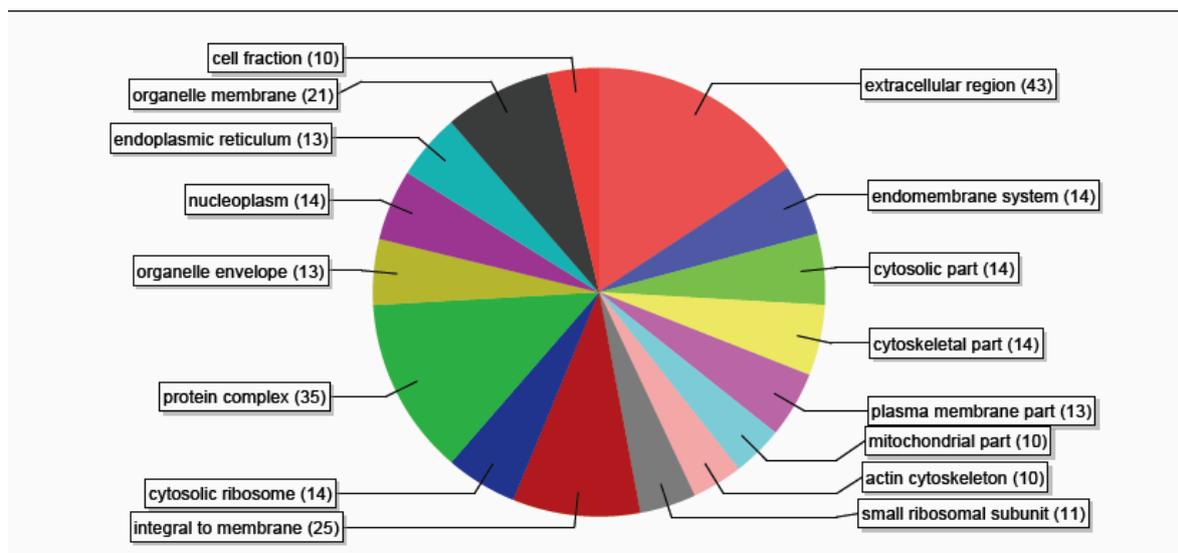


Figura 12: Distribuição dos termos do GO do domínio componente celular. Resultados são mostrados para as ESTs (A) e os unives (B). Para as ESTs foram considerados apenas os termos que foram atribuídos para no mínimo 20 seqüências e para os unives, no mínimo, 10 seqüências.

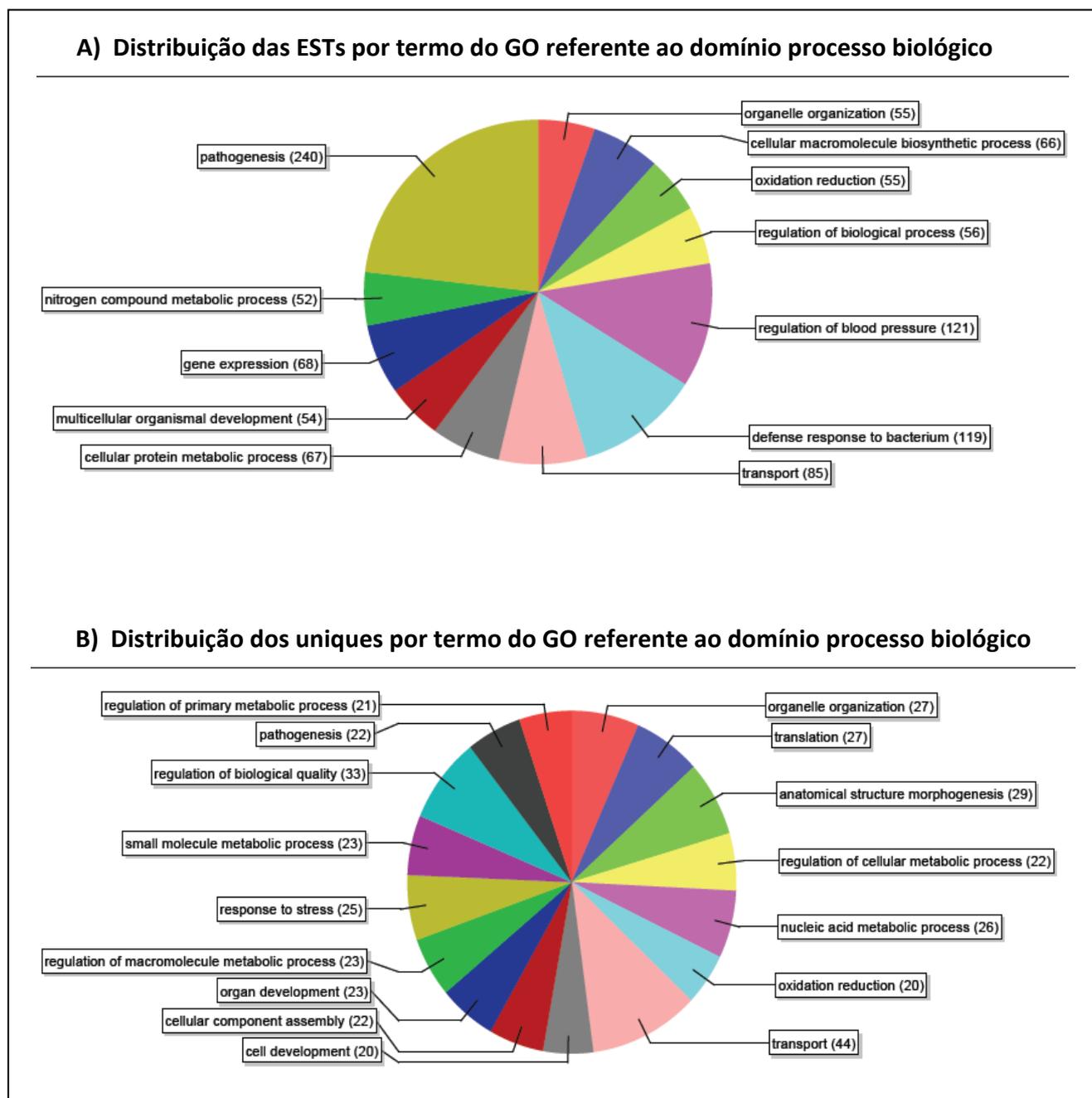
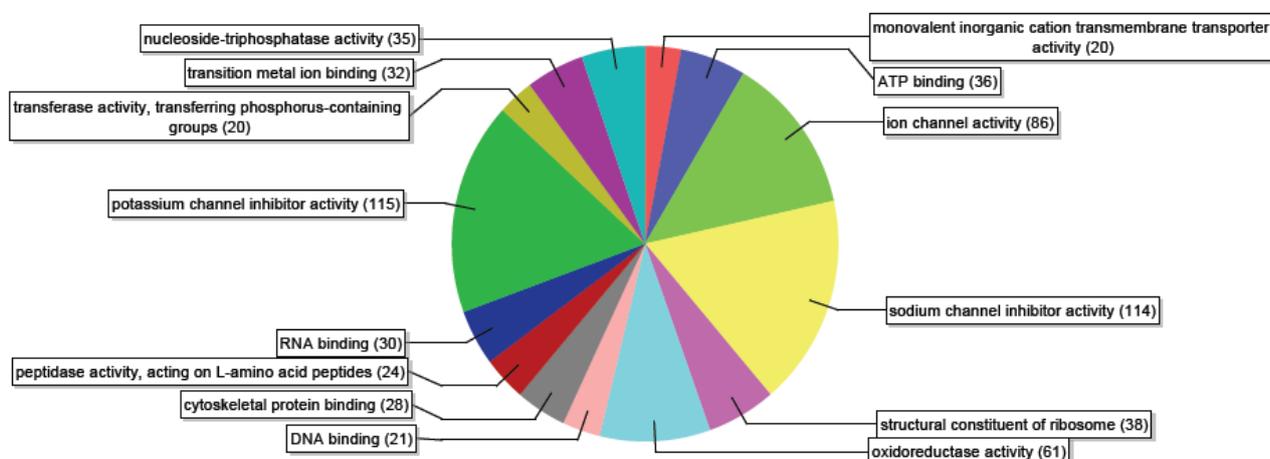


Figura 13: Distribuição dos termos do GO do domínio processo biológico. Resultados são mostrados para ESTs (**A**) e os uniques (**B**). Para as ESTs foram considerados apenas os termos que foram atribuídos para no mínimo 50 seqüências e para os uniques, no mínimo, 20 seqüências.

A) Distribuição das ESTs por termo do GO referente ao domínio função molecular



A) Distribuição dos unives por termo do GO referente ao domínio função molecular

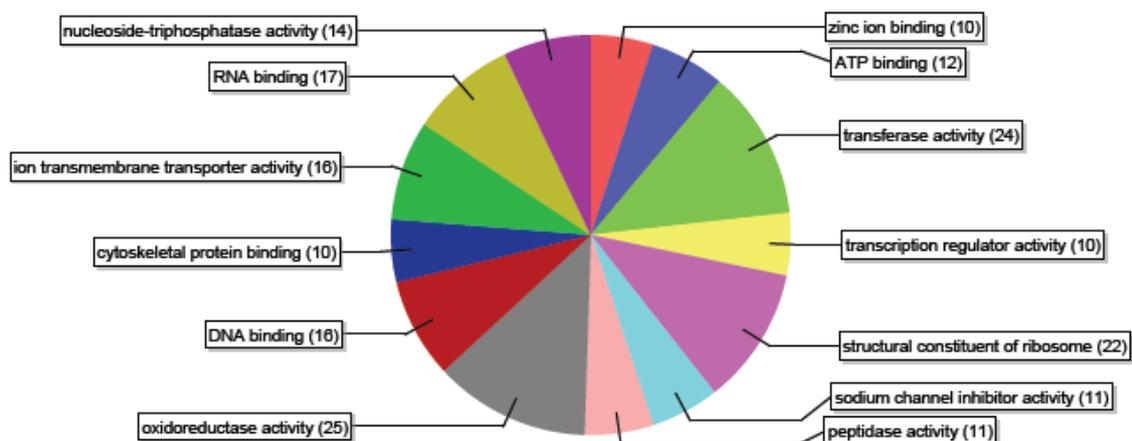


Figura 14: Distribuição dos termos do GO do domínio função molecular. Resultados são mostrados para ESTs (A) e os unives (B). Para as ESTs foram considerados apenas os termos que foram atribuídos para no mínimo 20 sequências e para os unives, no mínimo, 10 sequências.

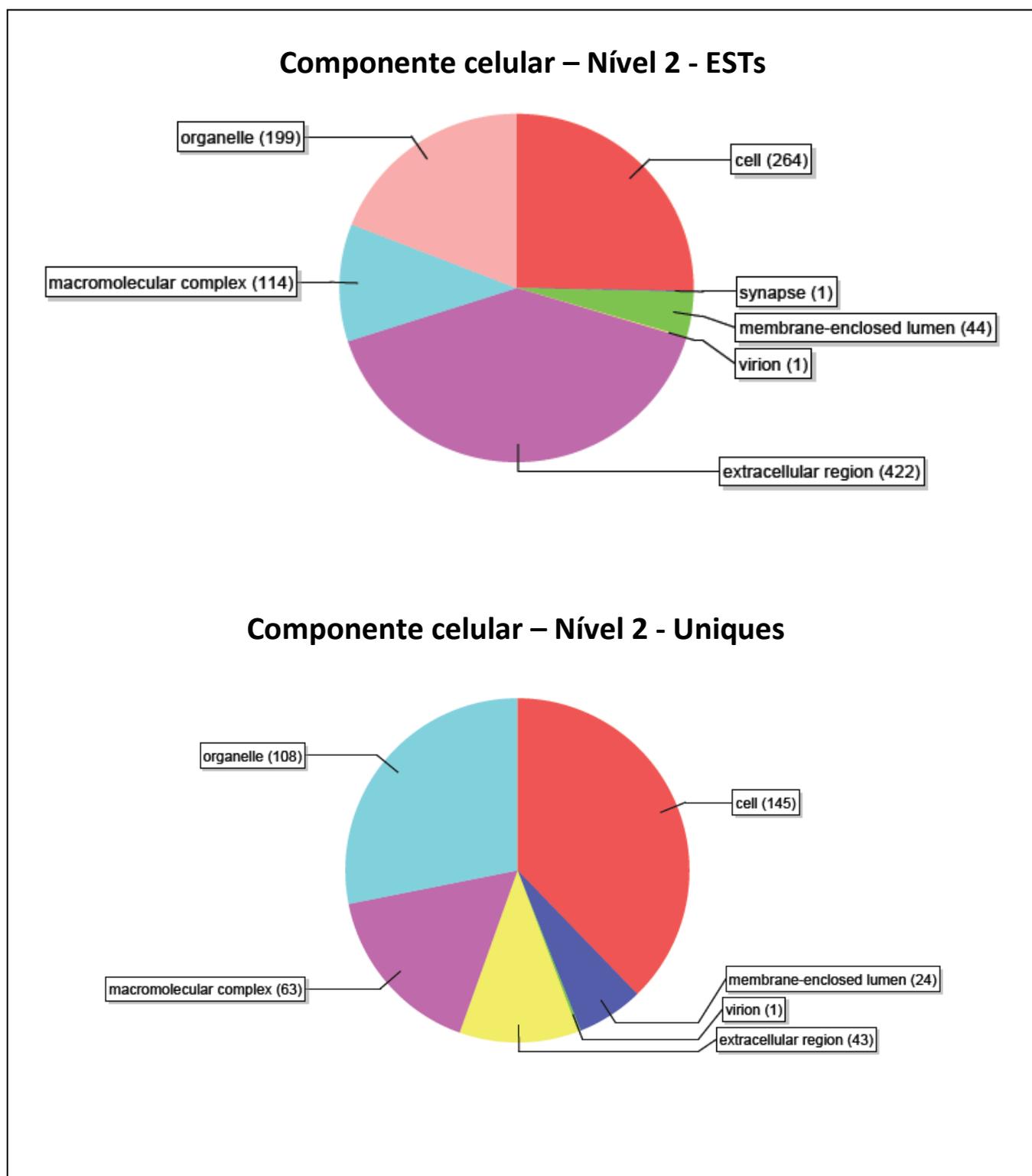


Figura 15: Distribuição das sequências em relação dos aos termos do GO do domínio componente celular nível 2. Resultados são mostrados considerando as ESTs (A) e os uniques (B).

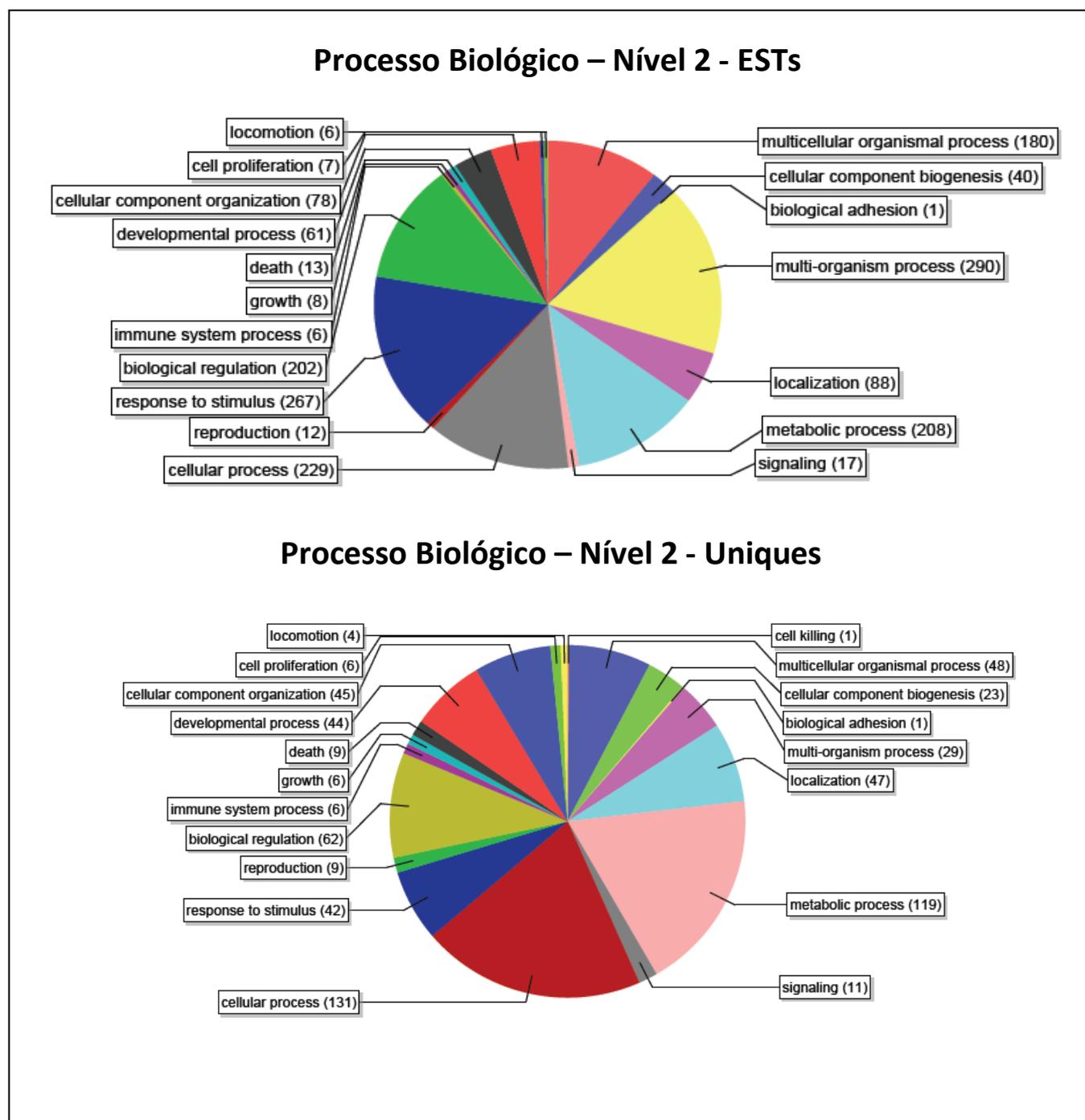


Figura 16: Distribuição das sequências em relação dos aos termos do GO do domínio processo biológico nível 2. Resultados são mostrados considerando as ESTs (A) e os únicos (B).

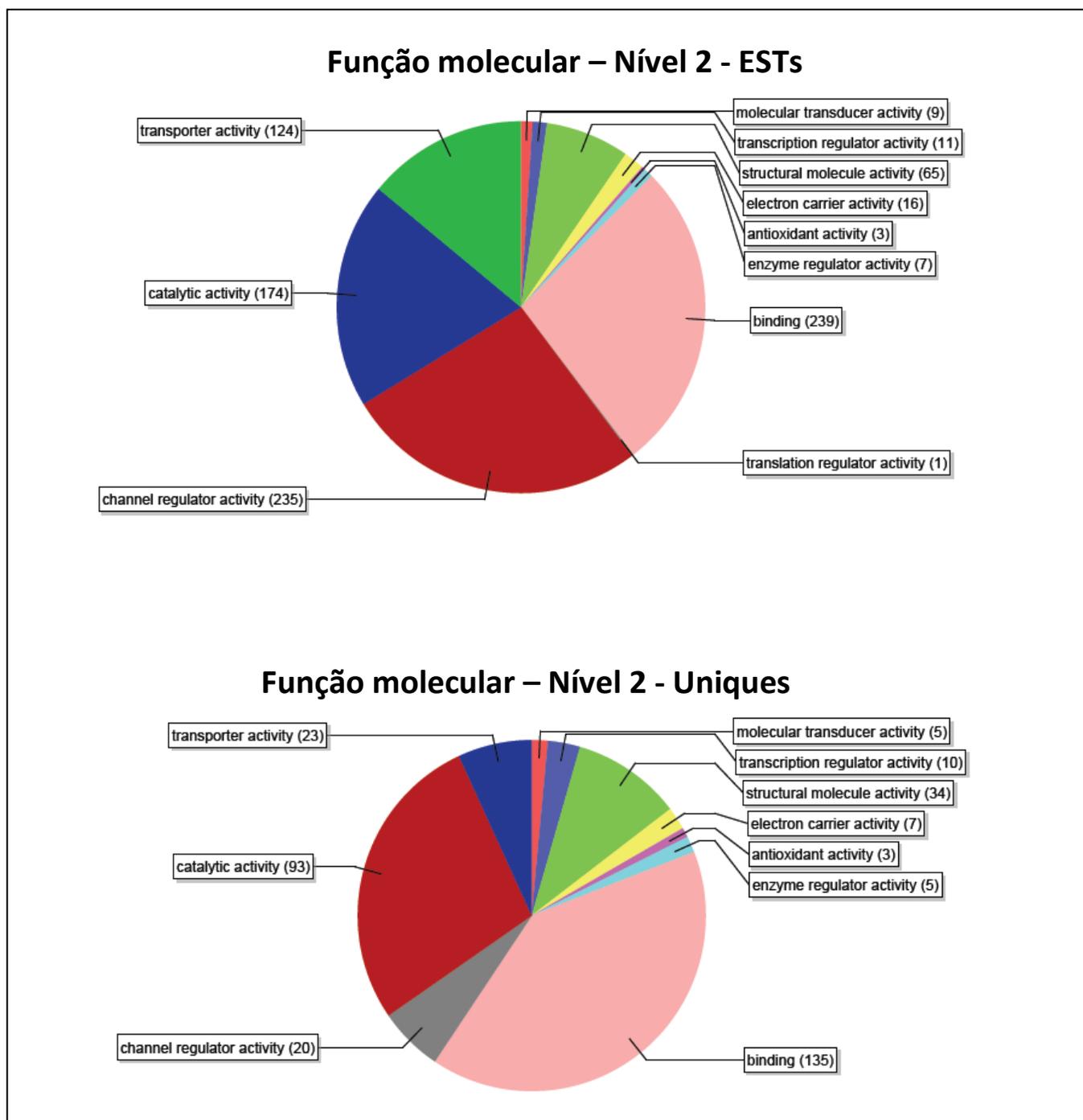


Figura 17: Distribuição das sequências em relação aos termos do GO do domínio função molecular nível 2. Resultados são mostrados considerando as ESTs (A) e os uniques (B).

5.4 Mapeamento dos transcritos de *T. serrulatus* usando a ferramenta KEGG PATHWAY

De maneira geral, ESTs relacionadas as funções celulares foram muito diversas, mas com poucos representantes cada uma. As vinte e cinco vias bioquímicas que apresentaram maior número de sequências de ESTs e unives de *T. serrulatus* na análise através da ferramenta KEGG PATHWAY são mostradas na Tabela 5. As vias ribossomo, fosforilação oxidativa, processamento de proteína no retículo endoplasmático, contração de músculo cardíaco e digestão e absorção de proteínas foram aquelas que apresentaram maior número de sequências. Os mapas metabólicos das três vias mais representadas com transcritos de *T. serrulatus* é mostrado nas Figuras 18, 19 e 20.

Conforme descrito nas seções anteriores, transcritos relacionados à síntese e processamento protéico e produção de energia foram muito representados na glândula de veneno do escorpião amarelo. Interessantemente, contração de músculo cardíaco apareceu como a quarta via que apresentou maior número de sequências. Esses transcritos são provavelmente provenientes da camada muscular que envolve a glândula.

O transcriptoma de *T. serrulatus* possui genes conservados relacionados ao processo de secreção, uma vez que nele apareceram componentes das vias secreção pancreática, de bile e salivar. Os transcritos identificados no presente estudo e que fazem parte dessas vias são: protease, caveolina (proteína essencial para formação das cavéolas), fosfolipase A2 (lipase pancreática), aquaporina, bomba de Na^+/K^+ .

5.5 Histologia da glândula de veneno de *T. serrulatus*

Um par de glândulas de veneno, uma em cada lado do corpo é encontrada na vesícula do telson (Polis, 1990). Histologicamente, na análise do telson de fora para dentro observam-se a cutícula seguida por uma espessa camada de músculos compressores que envolvem a glândulas de veneno (Fig. 21A). Em *T. serrulatus*, o epitélio secretório demonstra extensivo dobramento e é suportado por uma delgada camada de tecido conjuntivo frouxo. As células epiteliais são altamente polarizadas, com núcleo basal contendo cromatina finamente granulada (Fig. 21B), típico de célula em intensa atividade metabólica. Especialmente em torno do núcleo, o citoplasma é granular e basofílico. Essas características das células epiteliais são típicas de células com intensa atividade de síntese

protéica. Interessantemente, foram observadas diferentes morfologias e padrões de coloração dos produtos secretados (Fig. 21C). Assim, foram observados grânulos de diferentes diâmetros e coloração (acidófilos, basófilos e com distinta intensidade de coloração). Grânulos com padrão de coloração heterogêneo sugere a presença de compartimentalização da síntese de diferentes componentes do veneno.

Tabela 5: Vias bioquímicas mais representadas na análise de ESTs e Uniques de *T. serrulatus* submetidas a ferramenta KEGG PATHWAY.

Vias KEGG				
	Entrada	Via metabólica	nº de uniques	nº of ESTs
1	03010	Ribosome	21	22
2	00190	Oxidative phosphorylation	13	16
3	04141	Protein processing in endoplasmic reticulum	9	9
4	04260	Cardiac muscle contraction	7	9
5	04974	Protein digestion and absorption	5	5
6	00010	Glycolysis / Gluconeogenesis	4	0
7	03013	RNA transport	4	4
8	04010	MAPK signaling pathway	4	4
9	04910	Insulin signaling pathway	4	4
10	04972	Pancreatic secretion	4	4
11	00620	Pyruvate metabolism	3	3
12	00680	Methane metabolism	3	3
13	00330	Arginine and proline metabolism	3	3
14	03040	Spliceosome	3	3
15	04020	Calcium signaling pathway	3	3
16	04145	Phagosome	3	3
17	04810	Regulation of actin cytoskeleton	3	4
18	04510	Focal adhesion	3	3
19	04530	Tight junction	3	3
20	04612	Antigen processing and presentation	3	3
21	04270	Vascular smooth muscle contraction	3	3
22	04970	Salivary secretion	3	3
23	04971	Gastric acid secretion	3	3
24	04976	Bile secretion	3	3
25	04978	Mineral absorption	3	3

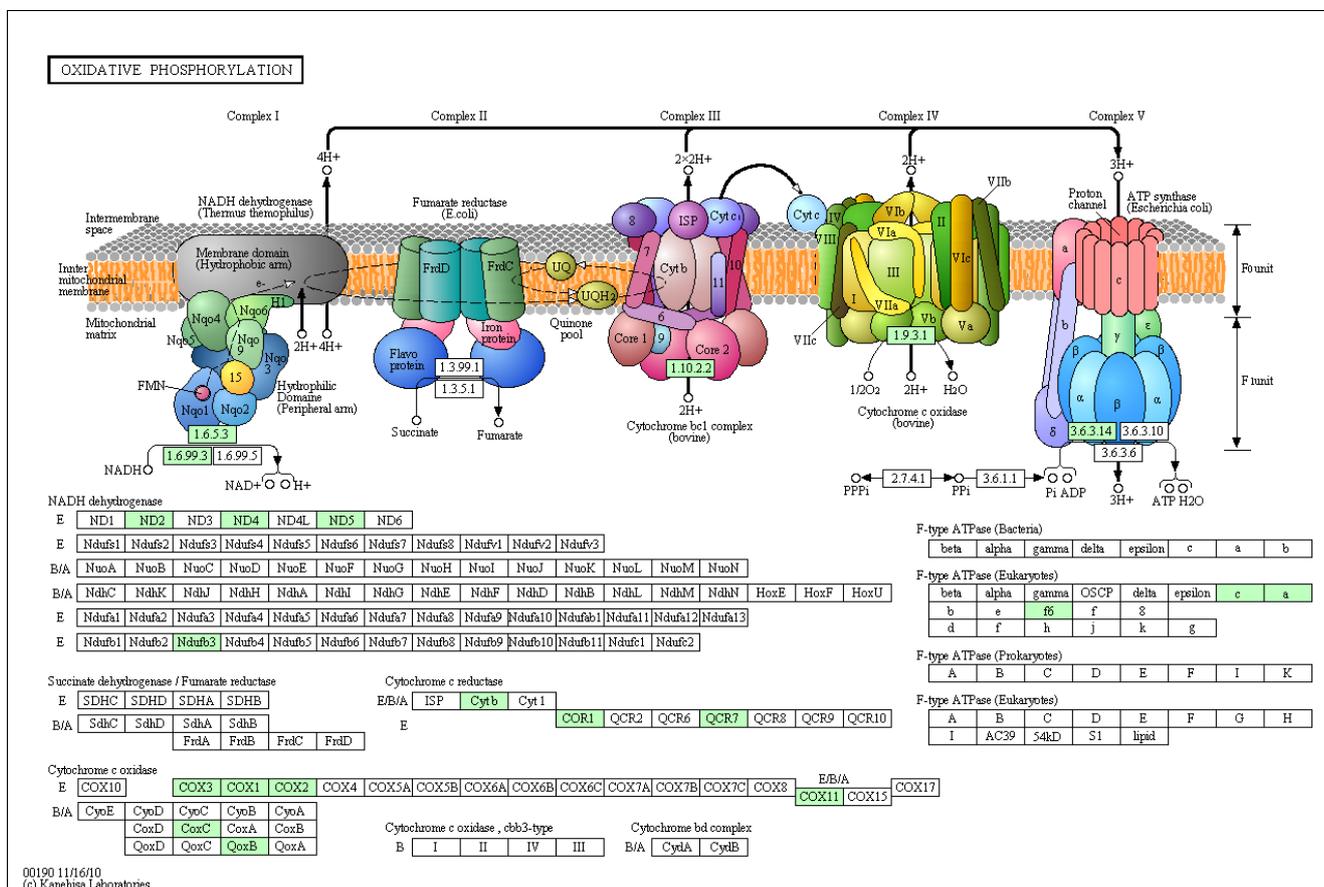


Figura 19: Via da Fosforilação oxidativa fornecida pela análise das ESTs usando a ferramenta KEGG PATHWAY. A produção de veneno é um processo considerado energeticamente custoso. Nesse sentido, diversas sequências similares a proteínas presentes na via de fosforilação oxidativa foram encontradas. Na parte superior da figura são mostrados os complexos proteicos envolvidos na produção de energia na mitocôndria. Na parte inferior são mostradas as subunidades que constituem os complexos proteicos com os componentes encontrados no transcriptoma de *T. serrulatus* destacados em verde.

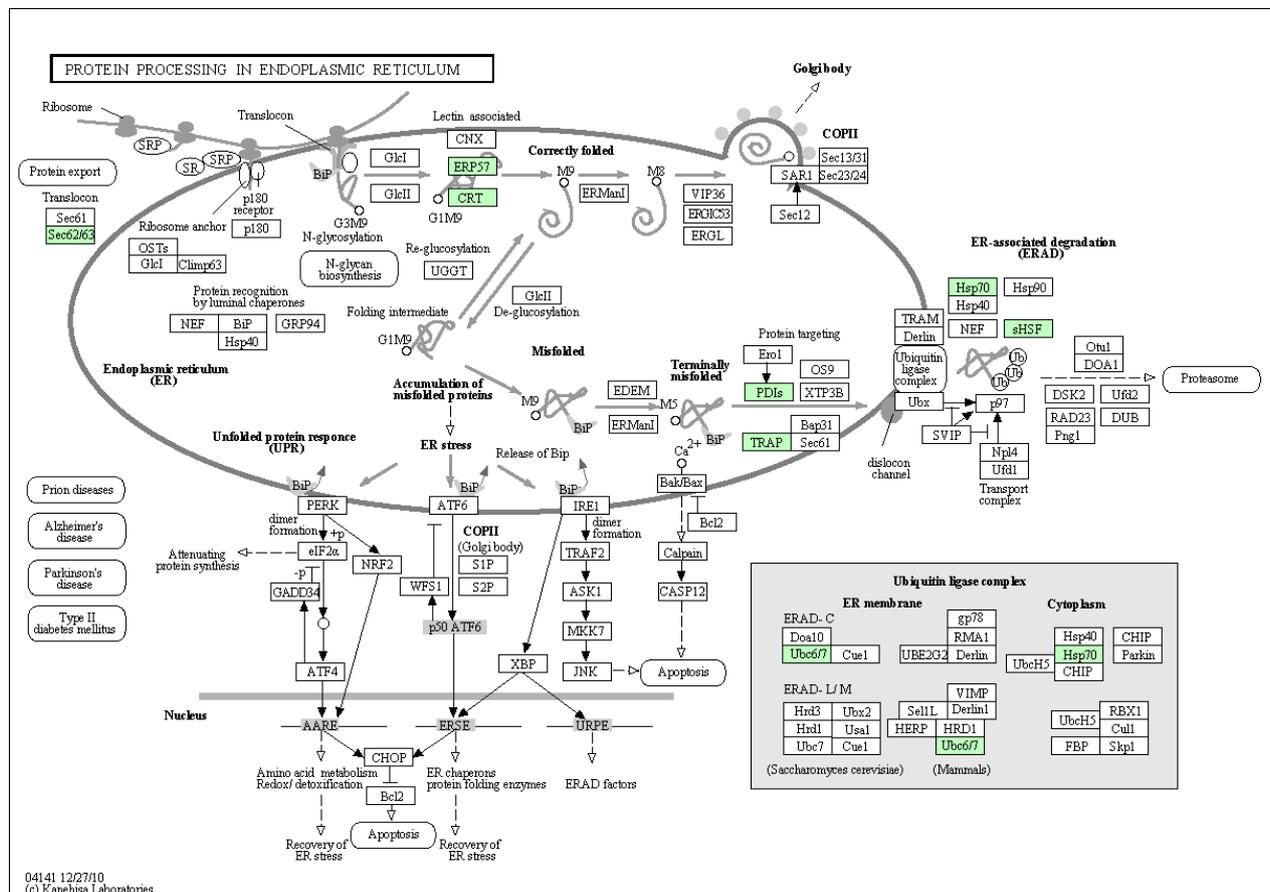


Figura 20: Via relacionada ao processamento de proteínas no retículo endoplasmático fornecida pela análise das ESTs usando a ferramenta KEGG PATHWAY. A glândula de veneno é um órgão especializado na síntese de proteínas e, nesse sentido, foram encontrados diferentes componentes da via de processamento de proteínas no transcriptoma de *T. serrulatus*. As proteínas que apresentaram similaridade com ESTs do presente estudo foram: CRT = calreticulina; ERP57 = proteína dissulfeto isomerase família A, membro 3; Sec62/63 = Proteína de translocação Sec62; PDIs = proteína dissulfeto isomerase A1; TRAP = Proteína associada a translocon; Hsp70 = proteína de choque térmico de 70 kDa; sHSF = cristalin alfa A; Ubc6/7 enzima E2 J1 de conjugação a ubiquitina. Interessante notar que foram encontradas 4 ESTs com similaridade a dissulfeto isomerasas. Essas enzimas, que são importantes para a formação de pontes dissulfeto e devem desempenhar um papel importante na síntese de toxinas, uma vez que muitas delas apresentam pontes dissulfeto em sua estrutura.

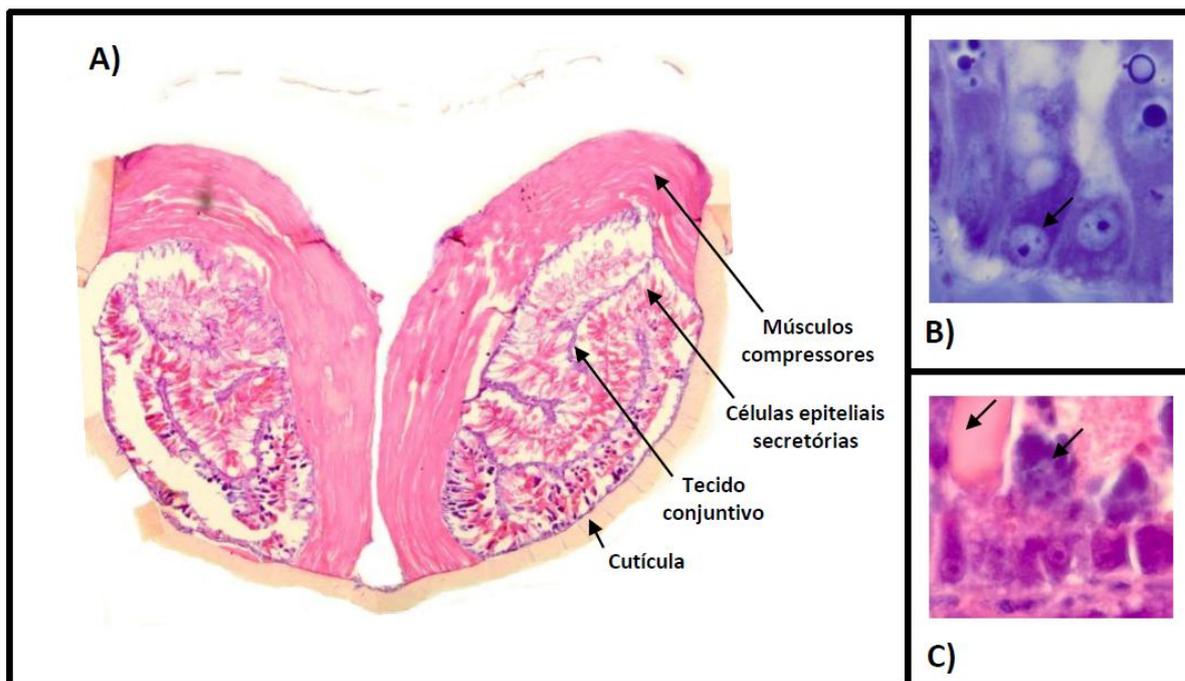


Figura 21: Características histológicas da glândula de veneno de *T. serrulatus*. **A)** A cutícula separa a estrutura da glândula do meio externo. Na porção médio-dorsal, uma espessa camada muscular envolve a glândula de veneno. O epitélio secretório demonstra extensivas dobras, o que possivelmente favorece a produção de uma maior quantidade de veneno. Esse epitélio secretório é suportado por uma delgada camada de tecido conjuntivo frouxo. **B)** Detalhe das células epiteliais demonstrando sua morfologia prismática, com núcleo basal (seta) contendo cromatina fina e com nucléolo bem evidente. Em volta do núcleo, observa-se aspecto granuloso, típico de célula contendo grande número de ribossomos. Essa figura foi obtida da glândula de veneno incluída em resina glicol-metacrilato, corada com azul de toluidina, o que favorece a obtenção de imagens de melhor resolução para visualização dos componentes celulares quando comparado com a inclusão em parafina e coloração com hematoxilina-eosina (HE). **C)** Detalhe das células epiteliais, com destaque a variedade de grânulos sintetizados por elas. Podem ser observados grânulos basófilos e acidófilos (setas) e com diferentes granulometrias presentes em células epiteliais adjacentes. Essa imagem foi obtida a partir de material incluído em parafina e corado com HE que também permitiu visualizar as características das células epiteliais típicas de células em intensa atividade de síntese protéica, como núcleo contendo nucléolo evidente e cromatina fina e citoplasma basófilo em torno de núcleo.

5.6 Busca de ortólogos dos componentes do veneno encontrados no transcriptoma de *T. serrulatus*

As sequências dos únicos considerados componentes do veneno foram traduzidas e utilizadas para a busca de ortólogos próximos com auxílio da ferramenta Seed Server e ortólogos distantes utilizando PSI-BLAST contra o UniProt e PSI-tBLASTn contra dbEST.

As análises com o Seed Server foram realizadas com todos os peptídeos e proteínas listados na Tabela 4 que apresentaram sequência completa. Utilizando essa ferramenta resultados não foram encontrados para alguns componentes como Ts18, peptídeo similar a ponicina-W-like, proteína ricos em glicina. Os grupos de organismos nos quais foram encontrados ortólogos usando o Seed Server são mostrados na Tabela 6. Deve-se ressaltar

Tabela 6: Resultados da busca dos grupos de organismos que apresentam ortólogos de componentes do veneno de *T. serrulatus* usando Seed Server.

Componentes do veneno	gênero	família	ordem	classe	filo	grupo	grupo	grupo	grupo	grupo	
	Tityus	Butidae	Scorpiones	Arachnida	Arthropoda	Coelomata	Bilateria	Eumetazoa	Metazoa	Eukaryota	Organismo celular
Toxinas de canal para Na⁺											
Ts1 (β-BUTX-Ts1a)*											
Ts2 (δ-BUTX-Ts1b) [P68410]											
Ts3 (δ-BUTX-Ts1)											
Ts17 (U-BUTX-Ts1a) [novo]											
Toxinas de canal para K⁺											
Ts8 ou TsTXK-beta (κ-BUTX-ts1)											
Ts15 (κ-BUTX-ts2)											
Ts6 ou alfa-KTx 12.1 (λ-BUTX - Ts1b)											
Ts7 ou alfa-KTx 4.1-like (λ-BUTX - Ts1a)											
ts16 (κ-BUTX-ts3)											
Ts19 (U-BUTX-Ts1c) [new]											
Outros componentes											
Peptídeo Pape ou BPP-like [P86821]											
AMPs [novo]											
Peptídeo aniônico [novo]											
Hipotensina 1 ou Ts14 [P84189]											
Hipotensina-like [novo]											
Tssep3 ou Ts13 [P0C176]											
Peptídeo rico em cisteína, similar a Ts15 [novo]											
Similar a TsPep2 ou Ts12 [P0C175]											
Tssep1 ou Ts11 [P0C174]											

* Entre parêntese, a nomenclatura de toxinas conforme sugerida por King et al. (2008)

Legenda

Seed Server

que essa ferramenta é bastante rigorosa, uma vez que exige cobertura mínima de 50% da sequência usada como entrada, 50% de identidade, dentre outros pré-requisitos (Barbosa-Silva *et al.*, 2008; Fernandes *et al.*, 2008). Assim, essa busca não objetivou encontrar toda gama de ortólogos da sequência usada como entrada, mas sim identificar ortólogos muito próximos para uma anotação baseada em similaridade mais criteriosa. A comparação entre a Tabela 6 e 7 evidência o maior rigor do Seed Server na identificação de ortólogos.

As sequências de Ts17 e Ts19 (Fig. 22) e AMP e peptídeo aniônico (Fig. 23), cujas sequências completas não foram ainda descritas em *T. serrulatus* são apresentadas juntamente com seus ortólogos. De acordo com as buscas realizadas pelo programa Seed Server, ortólogos de Ts17, Ts19, AMPs e peptídeo aniônico de *T. serrulatus* foram encontrados apenas em membros da família Buthidae. Esses peptídeos de *T. serrulatus* e seus ortólogos foram alinhados usando a ferramenta webPrankster (Löytynoja e Goldman, 2010).

Grande número de ortólogos foram identificados para Ts17 (Fig. 22A). Desde que uma grande parte desses ortólogos consistem em α -NaTx, Ts17 teria possivelmente essa ação. A análise da sequência de Ts17 indica que ela seria amidada na porção C-terminal devido a presença do motivo GKK. Estudos eletrofisiológicos e a identificação do peptídeo no veneno são necessários para confirmar essas hipóteses.

Existe evidência da ocorrência de Ts19 no veneno de *T. serrulatus* desde que um peptídeo contendo uma sequência de 10 aminoácidos idênticos a Ts19 foi identificado em estudo de proteoma (Rates *et al.*, 2008). Essa sequência (Fig. 22B) possui 37% de identidade com a β -KTX Ts8. Devido a presença de 6 cisteínas, o peptídeo maduro contém provavelmente 3 pontes dissulfeto. Diego-Garcia *et al.* (2007) identificou peptídeos similares a Ts19 em espécies do gênero *Tityus* e considerou que esses componentes formariam um grupo distinto das β -KTX. Entre os ortólogos identificados, apenas BmTXK-beta de *Mesobuthus martensi* foi submetida a estudo funcional, que indicou essa toxina bloquearia canais para K^+ (Zhu *et al.*, 1999).

Ts18 não apresentou resultado utilizando Seed Server. Por se tratar de sequência nova em *T. serrulatus*, o alinhamento dela com a única sequência obtida pelo BLASTx é mostradas na Figuras 23. Embora Ts18 tenha sido anotada como provável toxina de canal para Na^+ , dada a sua similaridade ao peptídeo proveniente de *H. judaicus* também anotado dessa forma, ela possui algumas características de toxinas de canal para K^+ . Ou seja, ela é mais curta do que as NaTxs típicas e possui provavelmente 3 pontes dissulfeto, que são

mais comuns em KTx. Assim, estudos eletrofisiológicos são necessários para confirmação da atividade da Ts18.

Três famílias de prováveis peptídeos antimicrobianos foram identificados no transcriptoma de *T. serrulatus*: AMPs, peptídeo aniônico e Ponericina-like (Figs. 24 e 25). Exceto para peptídeos aniônicos, que até o presente momento não tiveram sua atividade antimicrobiana investigada, ortólogos de AMPs e ponericina-like já foram submetidos a estudos funcionais que confirmaram sua atividade. Uma característica comum a esses três peptídeos é a ocorrência de aminoácidos hidrofóbicos e aminoácidos carregados, que dariam ao peptídeo um caráter anfipático.

Utilizando ferramentas de proteômica, peptídeo Pape e hipotensinas já foram descritos no veneno de *T. serrulatus*. No presente estudo, as sequências completas de seus precursores, que inclui o peptídeo sinal foram identificadas e mostradas nas Figuras 26 e 27. Além dessas, sequências muito similares a Ts12 e Ts15 já descritas anteriormente foram identificadas e essas sequências variantes são mostradas nas Figuras 28 e 29. Uma vez que esses peptídeos foram diretamente sequenciados pela metodologia de degradação de Edman, o transcriptoma trouxe como novidade a sequência do peptídeo sinal. Uma outra observação interessante é que para todos esses componentes, foram encontradas pró-regiões eliminadas no peptídeo maduro. A importância dessas pró-regiões no processamento dos peptídeos ainda precisa ser investigada.

Na busca por ortólogos distantes optou-se por mostrar os grupos de organismos nos quais foram identificadas espécies contendo sequências ortólogas aos componentes do veneno de *T. serrulatus*, uma vez que, para alguns componentes, um grande número de ortólogos foram obtidos pela análise de PSI-BLAST. Assim, para obter essa informação utilizou-se os identificadores taxonômicos de cada resultado de PSI-BLAST. Os resultados dessas buscas considerando o universo protéico (PSI-BLAST contra UniProt) e o universo de transcritos (PSI-tBLASTn contra dbEST) são mostrados na Tabela 7.

Independente da ferramenta usada (Seed Server, PSI-BLAST ou PSI-tBLASTn), NaTx apenas foram encontradas em espécies da família Buthidae, enquanto que as KTx apresentaram distribuição mais ampla e seus ortólogos foram encontrados na ordem Scorpiones.

A categoria “outros componentes do veneno” inclui tanto genes que possivelmente surgiram no gênero *Tityus* (como hipotensina 1, Tspept3), quanto proteínas que surgiram

em grupos basais e são amplamente distribuídos nos organismos (como metaloproteases e hialuronidases) e foram recrutadas para serem expressas nas glândulas de veneno.

Para alguns componentes do veneno foram observadas divergências de resultados quanto aos grupos de organismos nos quais os ortólogos foram encontrados. Como era esperado, a busca por ortólogos realizadas pelo Seed Server foi mais rigorosa. Assim, enquanto esse programa encontrou ortólogos de peptídeo aniônico apenas em membros da família Buthidae, o PSI-tBLASTn identificou sequências relacionadas a peptídeo aniônico na Ordem Scorpiones. Em geral, os resultados de PSI-tBLASTn e PSI-BLAST foram semelhantes e as divergências encontradas podem estar relacionadas com o depósito da proteína sem depósito correspondente do transcrito ou vice-versa.

5.7 Análise comparativa entre o transcriptoma de *T. serrulatus* e catálogos de transcritos de quelicerados

Com objetivo de identificar diferenças e similaridades na ocorrência de transcritos e padrão de expressão das glândulas de veneno foi realizada análise comparativa entre transcriptomas obtidos de diferentes espécies de quelicerados. Dados a respeito do número de ESTs analisados, do número de uniques obtidos após o processo de agrupamento, proporção de resultados de BLASTx de uniques e ESTs contra o UniProt e nr e proporção de resultados de tBLASTx dos uniques de *T. serrulatus* contra os catálogos de transcritos dos demais quelicerados são mostrados na Tabela 8.

Uma porcentagem relativamente alta de sequências sem match foi uma característica comum na maioria dos transcriptomas de quelicerados analisados. Cerca de 27–56% dos Uniques e 22–58% das ESTs não apresentaram *hits* significativos com sequências depositadas nos bancos de dados. Exceto para o transcriptoma de *N. antipodiana*, no qual 74% dos uniques e 79% das ESTs apresentaram resultado da análise de BLASTx, quase a metade das ESTs não apresentou *match* com sequências depositadas em banco de dados públicos. Esse resultado indica a relativa escassez de informação molecular dos escorpiões e de organismos relacionados. A alta porcentagem de resultados na análise através da ferramenta BLASTx do transcriptoma de *N. antipodiana* pode estar relacionado ao pequeno número de sequências nesse catálogo de transcritos ou ao considerável interesse pelas glândula de seda como fonte de biomateriais (Prosdocimi *et al.*, 2011). Vale ressaltar que a ocorrência de *match* não significa que a função da proteína seja conhecida e uma considerável proporção das sequências desses transcriptomas pode

ter sido alinhada com proteínas preditas ou não caracterizadas, como de fato ocorreu para o transcriptoma de *T. serrulatus*, no qual 10% das ESTs foram anotadas desta forma.

A comparação do transcriptoma de *T. serrulatus* contra catálogos de transcritos de outros quelicerados revelaram que as espécies mais próximas de *T. serrulatus* apresentaram transcriptomas com maior similaridade ao catálogo de transcrito desenvolvido no presente estudo (veja a coluna ‘resultados de tBLASTx usando Uniques’ na Tabela 8). O menor grau de similaridade foi de 5%, quando se comparou *T. serrulatus* com *E. spinosa*. Essa similaridade subiu para 22%, quando a comparação foi realizada entre o transcriptoma da glândula de veneno de *T. serrulatus* com o de outro escorpião também pertencente à família Buthidae (*L. mucronatus*). Quando a mesma análise comparativa foi realizada contra as ESTs, esse padrão não foi observado devido à redundância de transcritos nos transcriptomas de *I. ricinus*, *S. scabiei* e *L. polyphemus* (observe coluna ‘resultados de tBLASTx usando ESTs’ na Tabela 8). Como exemplo, em *I. ricinus*, 47% da similaridade com *T. serrulatus* foi causada por apenas 5 agrupamentos contendo dezenas de ESTs que codificaram subunidades de citocromo c oxidase (3 agrupamentos), uma subunidade ATP sintase (1 agrupamento) e uma proteína hipotética (1 agrupamento). A similaridade entre *T. serrulatus* e outros transcriptomas de glândulas (salivar e de seda) não foi necessariamente maior do que a similaridade entre o transcriptoma de *T. serrulatus* e de animais inteiros. A avaliação da ocorrência e do nível de expressão de ortólogos dos componentes de veneno de *T. serrulatus* em outros transcriptomas usando o programa tBLASTn e tBLASTx são mostrados nas Tabelas 9 e 10, respectivamente. Com exceção do transcriptoma de *L. mucronatus*, que compartilhou muitas sequências e padrão de expressão com *T. serrulatus*, os outros catálogos de transcritos apresentaram um pequeno número de ESTs similares a componentes do veneno de *T. serrulatus*.

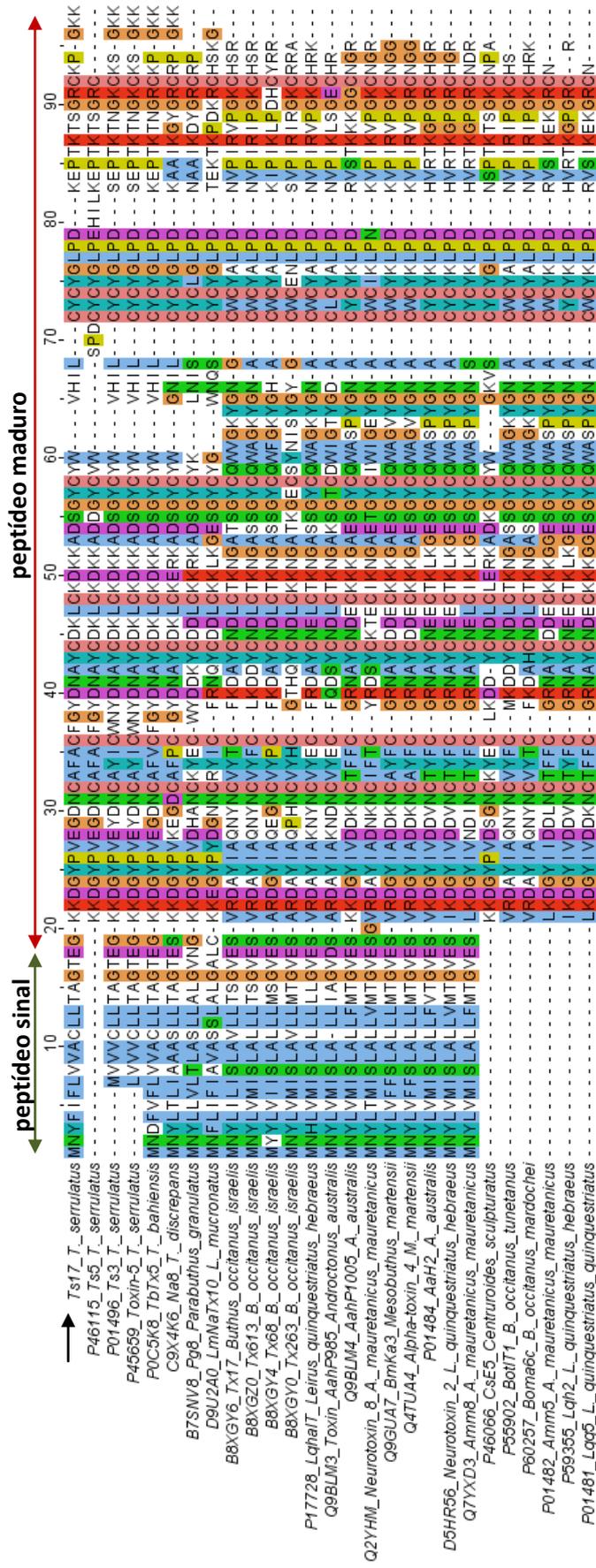
Especialmente em transcriptomas proveniente de espécies não pertencentes à ordem Scorpiones, a incidência e nível de expressão de transcritos similares a componentes de veneno de *T. serrulatus* foram muito baixos. De maneira geral, os resultados da busca utilizando os programas tBLASTx e tBLASTn foram muito semelhantes. Embora o programa tBLASTx tenha resultado em maior número de *hits*, os resultados do tBLASTn são mais confiáveis quando se busca uma determinada proteína, uma vez que o usuário oferece a sequência de aminoácidos como sequência de entrada. Uma vez que o tBLASTx traduz a sequência de nucleotídeos de entrada nas seis janelas de leitura, o programa pode

encontrar uma outra sequência de nucleotídeo que não necessariamente codifica a sequência de aminoácidos correspondente a da proteína pesquisada. Isso pode explicar, por exemplo, a ocorrência de sequência similar a uma provável KTx em *E. spinosa* na análise por tBLASTx.

Considerando que a maioria dos componentes de veneno foi exclusiva das glândulas de veneno, genes relacionados a homeostase e processos celulares gerais foram os principais responsáveis pela similaridade encontrada entre os transcriptomas. De fato, isso pôde ser observado tanto na análise da anotação das sequências responsáveis por 50% da similaridade com o transcriptoma de *T. serrulatus* (Tabela 11) quanto na avaliação da distribuição dos termos do GO entre os transcriptomas de quelicerados (Fig. 30).

Quanto à distribuição dos termos do GO para os diferentes transcriptomas, pôde-se observar que os catálogos de transcritos de glândulas de veneno apresentaram alguns aspectos particulares. Por exemplo, maior porcentagem de sequências foi assinada com o termo ‘extracellular region’ em transcriptomas de glândulas de veneno. Assim, enquanto 40% das anotações dos transcritos de *T. serrulatus* constituiu-se no termo ‘extracellular region’, apenas 3% das anotações das ESTs pertencentes à *M. gibbosus* (transcriptoma de animal inteiro) foi realizada com esse termo. Além disso, a porcentagem de sequências assinadas com o termo ‘multi-organism process’, ‘response to stimulus’ e ‘channel regulator activity’ foram claramente maiores em transcriptomas de glândulas de veneno do que em transcriptomas de animais inteiros. Em particular, *T. serrulatus* e *L. mucronatus*, ambos butídeos, apresentaram padrão muito semelhante de distribuição de termos do GO.

A) Provável toxina de canal para Na⁺: Ts17 (U-BUTX-Ts1a)



B) Provável toxina de canal para K⁺ (U-BUTX-Ts1c)

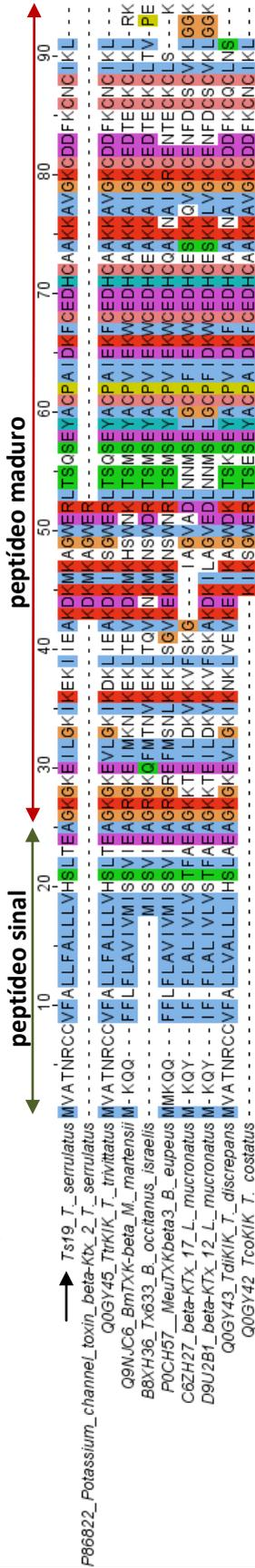


Figura 22: Alinhamento de seqüências de NaTx (A) e KTx (B). As setas pretas indicam seqüências obtidas no presente estudo. Em A, a seqüência traduzida de Ts17 (U-BUTX-Ts1a) e seus ortólogos encontrados por meio da ferramenta Seed Server. O peptídeo sinal (seta verde) foi previsto usando SignalP4.0 e o sítio de clivagem foi identificado entre os aminoácidos 19 e 20. Para efeito de comparação, os parálogos Ts3 e Ts5 também são mostrados. Em B, a seqüência completa de Ts19 e o alinhamento com os ortólogos encontrados pelo Seed Server. O sítio de clivagem do peptídeo sinal está entre o aminoácido 25 e 26. Como referência, o fragmento peptídico depositado da toxina de canal para K⁺ beta Ktx2 foi apresentado.

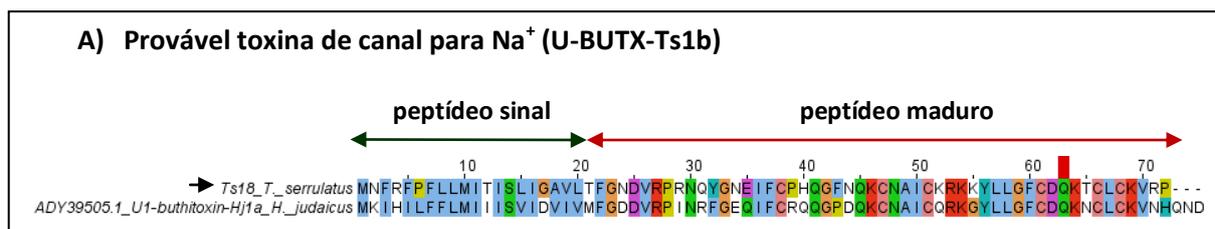


Figure 23: Alinhamento de Ts18 (U-BUTX-Ts1b) de *T. serrulatus* com a sequência de U1-buthitoxin-Hj1a de *Hottentotta judaicus*. O peptídeo sinal (seta verde) e sítio de clivagem são mostrados (entre o aminoácido 21-22).

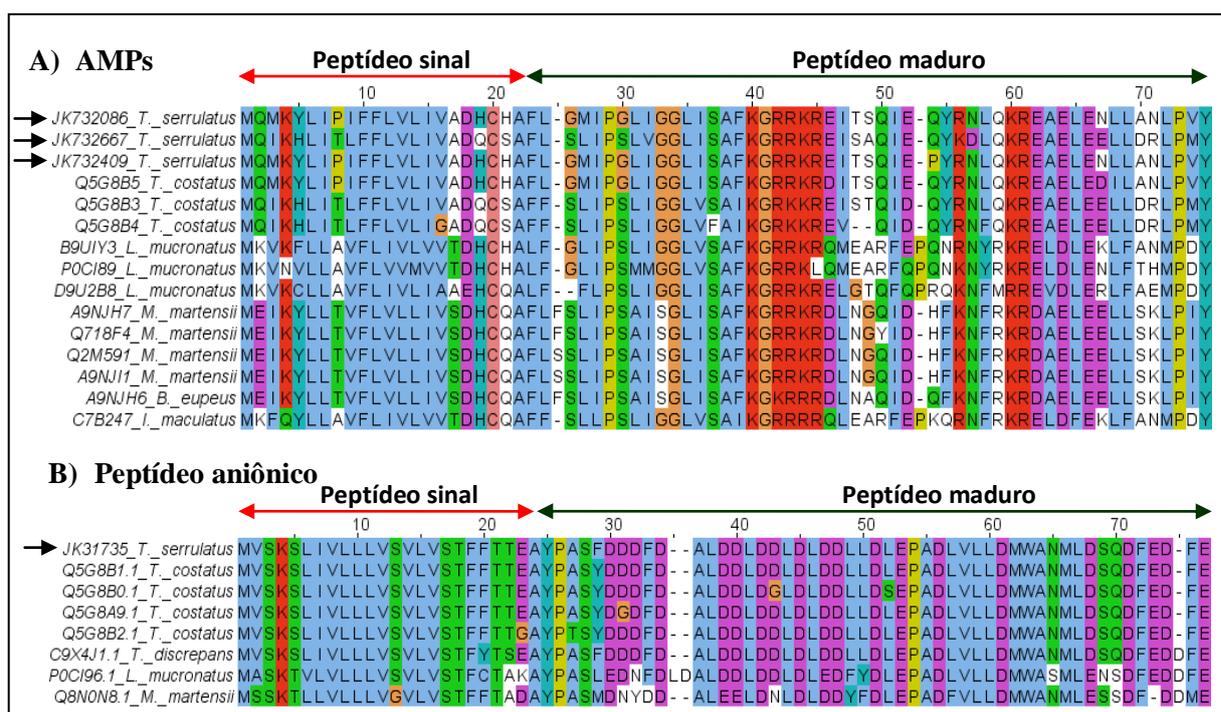


Figure 24: Sequências traduzidas de AMPs e peptídeo aniônico obtidas no transcriptoma da glândula de veneno de *T. serrulatus* e os alinhamentos com ortólogos encontrados via Seed Server. Em **A**, observe que 3 sequências diferentes são observadas em *T. serrulatus*, *T. costatus*, *L. mucronatus* e *Mesobuthus martensii*. O sítio de clivagem do peptídeo sinal é localizado entre os aminoácidos 22 e 23 (**B**) A sequência de peptídeo aniônico não apresentou formas variantes em *T. serrulatus* e o alinhamento com seus ortólogos demonstra que essa sequência é relativamente conservada. O peptídeo sinal com o sítio de clivagem entre os aminoácidos 24 e 25 é mostrado. Todos os alinhamentos foram realizados no programa webprankster (Löytynoja e Goldman, 2010) e a edição feita com auxílio do programa Jalview (Waterhouse *et al.*, 2009).

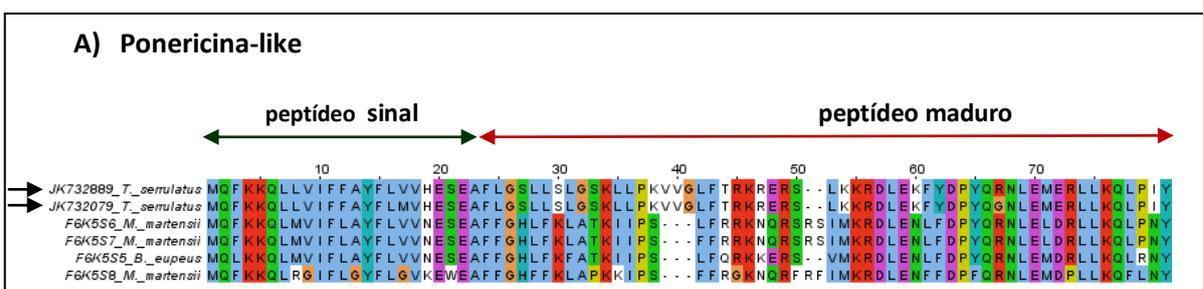


Figure 25: Peptídeo semelhante a ponericina apresentou duas formas variantes em *T. serrulatus*. Setas pretas indicam as sequências identificadas no presente estudo. O peptídeo sinal (seta verde) e o sítio de clivagem entre os aminoácidos 23 e 24 são mostrados.

Tabela 7: Resultados da busca dos grupos de organismos que apresentam ortólogos de componentes do veneno de *T. serrulatus* usando as ferramentas PSI-BLAST contra o UniProt e PSI-tBLASTn contra o dbEST.

Componentes do veneno	gênero	família	ordem	classe	filo	grupo	grupo	grupo	grupo	grupo	
Toxinas de canal para Na ⁺	<i>Tityus</i>	Buthidae	Scorpiones	Arachnida	Arthropoda	Coelomata	Bilateria	Eumetazoa	Metazoa	Eucarioto	Organismo celular
Toxinas de canal para Na⁺											
Ts1 (β-BUTX-Ts1a) [P15226]											
Ts2 (δ-BUTX-Ts1a) [P68410]											
Ts3 (θ-BUTX-Ts1b) [P01496]											
Ts17 (U-BUTX-Ts1b) [novo]											
Ts18 (U-BUTX-Ts1c) [novo]											
Toxinas de canal para K⁺											
Ts8 ou TsTXK-beta (κ-BUTX-ts1) [P69940]											
Ts15 (κ-BUTX-ts2) [P86270]											
Ts6 ou alfa-KTx 12.1 (λ-BUTX - Ts1b) [P59936]											
Ts7 ou alfa-KTx 4:1-like (κ-BUTX - Ts1a) [P46114]											
Ts16 (U-BUTX-Ts1a) [P86271]											
Ts19 (U-BUTX-Ts1d) [novo]											
Outros componentes											
Peptídeo Pape ou BPP-like [P86821]											
AMPs [novo]											
Similar a ponicrina-W-like de <i>L. mucronatus</i> [novo]											
Antarease * [P86392]											
TsNTXP ou Ts4 [O77463]											
Peptídeo aniônico [novo]											
Hipotensina 1 ou Ts14 [P84189]											
Hipotensina-like [novo]											
Zinco Metaloproteases * [P85842]											
Similar a proteína rica em glicina de <i>L. mucronatus</i> [novo]											
Tssep3 ou Ts13 [P0C176]											
Similar a neuropeptídeo-3 de <i>Buthus eupeus</i> [novo]											
Peptídeo rico em cisteína, similar a Ts15 [novo]											
Similar a TsPep2 ou Ts12 [P0C175]											
Hialuronidase* [P85841]											
Similar a peptídeo semelhante a toxina de <i>B. eupeus</i> [novo]											
Fosfolipase A2* [novo]											
Similar a peptídeo rico em cisteína de <i>Camponotus floridanus</i> * [novo]											
Similar a peptídeo rico em cisteína de <i>Hottentotta judaicus</i> * [novo]											
Tssep1 ou Ts11 [P0C174]											
Alérgeno * [P85840]											
Similar a cadeia alfa-1 do peptídeo ativador de lipólise de <i>L. mucronatus</i> * [novo]											
Similar a proteína do veneno de <i>B. eupeus</i> * [novo]											
Similar a proteína secretada de <i>Opis/thacanthus cayaporum</i> [novo]											

Legenda:

Apenas resultados de PSI-tBLASTn (versus dbEST)

Apenas resultados de PSI-BLAST (versus UniProt)

Ambos resultados: PSI-tBLASTn (versus dbEST) e PSI-BLAST (versus UniProt)

* Sequência com ORF incompleta

** Entre parêntese, nomenclatura sugerida por King et al. (2008)

Tabela 8: Transcriptomas amostrados demonstrando os resultados de anotação e análise comparativa entre o transcriptoma de *T. serrulatus* e outros catálogos de transcritos de quelicerados. O resultado de tBLASTx indica a porcentagem de cada transcriptoma que apresentou match com seqüências de *T. serrulatus*. Como exemplo, 32% das ESTs de *L. mucronatus* (totalizando 235 ESTs) tiveram match com transcritos de *T. serrulatus*. Em cada coluna, quanto mais escura for a célula, maior é o valor que está contido nela.

Espécie	Ordem	Transcriptoma	n° ESTs	n° únicos	Resultado de BLASTx		Resultado de tBLASTx	
					Uniques X UniProt	ES Ts x nr	usando únicos	usando ESTs
<i>T. serrulatus</i>	Scorpiones	glândula de veneno	1629	643	54,60%	59,05%		
<i>L. mucronatus</i>	Scorpiones	glândula de veneno	740	404	54,70%	69,86%	22%	32%
<i>M. gibbosus</i>	Scorpiones	animal inteiro	982	560	61,25%	56,42%	15%	32%
<i>H. petersii</i>	Scorpiones	glândula de veneno	262	198	62,00%	62,21%	15%	20%
<i>S. jendeki</i>	Scorpiones	glândula de veneno	872	304	59,87%	58,72%	17%	20%
<i>N. antipodiana</i>	Araneae	glândula de seda	512	334	73,65%	78,71%	14%	18%
<i>D. erythrina</i>	Araneae	animal inteiro	726	274	56,57%	46,83%	12%	17%
<i>L. laeta</i>	Araneae	glândula de veneno	1358	1206	53,15%	55,01%	9%	12%
<i>L. polyphemus</i>	Xiphosura	animal inteiro	8488	4100	55,88%	70,77%	8%	32%
<i>C. rotundicauda</i>	Xiphosura	hemócito e hepatopâncreas	575	512	44,73%	46,09%	8%	11%
<i>S. scabiei</i>	Acari	animal inteiro	1039	807	51,67%	42,83%	8%	20%
<i>I. ricinus</i>	Acari	glândula salivar	1969	1303	44,44%	75,52%	6%	18%
<i>E. spinosa</i>	Pantopoda	animal inteiro	4062	2640	50,87%	69,35%	5%	13%

Tabela 9: Ocorrência e nível de expressão de componentes do veneno de *T. serrulatus* e seus ortólogos nos transcriptomas amostrados na análise usando a ferramenta tBLASTn. Para facilitar a comparação, o número de transcritos de todos os transcriptomas foi padronizado para número de ESTs/100.000.

Componentes do veneno	Resultados de tBLASTn										Pantopoda	
	Escorpiões/Buthidae		Escorpiões/não-Buthidae		Araneae		Xiphosura		Acari			
	Ts	Lm	Mg	Hp	Sj	Li	Na	De	Cr	Ir	Ss	Es
toxinas de canal para Na⁺												
Ts1 (β-BUTX-Ts1a) [P15226]	2.026	541										
Ts2 (δ-BUTX-Ts1a) [P68410]	2.210											
Ts3 (δ-BUTX-Ts1b) [P01496]	614	541										
Ts17 (U-BUTX-Ts1b) [novo]	552	946										
Ts18 (U-BUTX-Ts1c) [novo]	301											
toxinas de canal para K⁺												
Ts8 ou TsTXK-beta (κ-BUTX-Ts1) [P69940]	4.481	541	204	382	1.376							
Ts15 (κ-BUTX-ts2) [P86270]	430											
Ts6 ou alfa-KTx 12.1 (λ-BUTX - Ts1b) [P59936]	307											
Ts7 ou alfa-KTx 4.1-like (κ-BUTX - Ts1a) [P46114]	246											
Ts16 (U-BUTX-Ts1a) [P86271]	61											
Ts19 (U-BUTX-Ts1d) [novo]	2.517	811										
Outros componentes												
Peptídeo Pape ou BPP-like [P86821]	8.471	2.973										
AMPs [novo]	3.622	3.514										
Similar a ponericina-W-like de <i>L. mucronatus</i> [novo]	2.455	1.081		763						152		
Antitasease * [P86392]	1.780	676										
TsNTXP ou Ts4 [O77463]	1.596	1.351										
Peptídeo aniônico [novo]	1.105											
Hipotensina 1 ou Ts14 [P84189]	675	614										
Hipotensina-like [novo]	614											
Zinco Metaloproteases * [P85842]	430	1.081										
Similar a proteína rica em glicina de <i>L. mucronatus</i> [novo]	368											
Tssep3 ou Ts13 [P0C176]	246											
Similar a neuropeptídeo-3 de <i>Buthus eupeus</i> [novo]	184											
Peptídeo rico em cisteína, similar a Ts15 [novo]	184											
Similar a TsPep2 ou Ts12 [P0C175]	123											
Hialuronidase* [P85841]	123											
Similar a peptídeo semelhante a toxina de <i>B. eupeus</i> [novo]	123											
Fosfolipase A2* [novo]	123			1.731								
Similar a peptídeo rico em cisteína de <i>Camponotus floridanus</i> * [novo]	123											
Similar a peptídeo rico em cisteína de <i>Hottentotta judaicus</i> * [novo]	61											
Tssep1 ou Ts11 [P0C174]	61											
Alergeno * [P85840]	61	135										
Similar a cadeia alfa-1 do peptídeo ativador de lipólise de <i>L. mucronatus</i> * [novo]	61	2.432										
Similar a proteína do veneno de <i>B. eupeus</i> * [novo]	61											
Similar a proteína secretada de <i>Opisthacanthus cayaporum</i> [novo]	61	135										

* Sequências com ORF incompleta.

** Entre parenteses, a nomenclatura das toxinas como sugerido por King et al. (2008).

Ts = *Tityus serrulatus*; Lm = *Lynchus mucronatus*; Mg = *Mesobuthus gibbosus*; Hp = *Heterometrus petersii*; Sj = *Scorpiops jendeki*; Li = *Loxocoeles laeta*; Na = *Nephila antipodiana*; De = *Dysdera erythrina*; Ir = *Ixodes ricinus*; Ss = *Sarcoptes scabiei*; Lp = *Limulus polyphemus*; Cr = *Carcinoscorpius rotundicauda*; Es = *Endeis spinosa*.

Escola
0-100
101-500
501-1.000
1.001-2.000
2.001-3.000
>3.000

Tabela 10: Ocorrência e nível de expressão de componentes do veneno de *T. serrulatus* e seus ortólogos nos transcriptomas amostrados na análise usando a ferramenta tBLASTx. Para facilitar a comparação, o número de transcritos de todos os transcriptomas foi padronizado para número de ESTs/100.000.

Componentes do veneno	Resultados de tBLASTx										Escala	
	Ts	Ln	Mg	Hp	Escorpiões/não-Buthidae	Araneae	De	Xiphosura	Acari	Pantopoda		
toxinas de canal para Na⁺												
Ts1 (β-BUTX-Ts1a) [P15226]	2.026	541										
Ts2 (δ-BUTX-Ts1a) [P68410]	2.210											
Ts3 (δ-BUTX-Ts1b) [P01496]	614	946										
Ts17 (U-BUTX-Ts1b) [novo]	552	946										
Ts18 (U-BUTX-Ts1c) [novo]	301											
toxinas de canal para K⁺												
Ts8 ou TsTXK-beta (κ-BUTX-Ts1) [P69940]	4.481	541	204	382								
Ts15 (κ-BUTX-Ts2) [P86270]	430											
Ts6 ou alpha-KTx 12.1 (λ-BUTX - Ts1b) [P59936]	307	405										
Ts7 ou alpha-KTx 4.1-like (κ-BUTX - Ts1a) [P46114]	246											
Ts16 (U-BUTX-Ts1a) [P86271]	61											
Ts19 (U-BUTX-Ts1d) [novo]	2.517	1.081										
Outros componentes												
Peptídeo Pape ou BPP-like [P86821]	8.471	2.973										
AMPs [novo]	3.622	3.378										
Similar a pomicrina-W-like de <i>L. mucronatus</i> [novo]	2.455	1.081		763							254	
Antartase * [P86392]	1.780	946										
TsNTXP ou Ts4 [O77463]	1.596	2.027										
Peptídeo aniônico [novo]	1.105	2.568										
Hipotensina 1 ou Ts14 [P84189]	675											
Hipotensina-like [novo]	614											
Zinco Metaloproteases [P85842]	430	946				74						
Similar a proteína rica em glicina de <i>L. mucronatus</i> [novo]	368	270				74						
Similar a neuropeptídeo-3 de <i>Buthus eupeus</i> [novo]	246											
Similar a TsPep2 ou Ts12 [POC175]	184											
Similar a TsPep3 ou Ts13 [POC176]	184											
Similar a TsPep1 ou Ts11 [POC174]	123											
Similar a TsPep2 ou Ts12 [POC175]	123											
Hialuronidase* [P85841]	123											
Similar a peptídeo semelhante a toxina de <i>B. eupeus</i> [novo]	123											
Fosfolipase A2* [novo]	123	135										
Similar a peptídeo rico em cisteína de <i>Camponotus floridanus</i> * [novo]	123											
Similar a peptídeo rico em cisteína de <i>Hottentotta judaicus</i> * [novo]	61											
TsPep1 ou Ts11 [POC174]	61											
Alergeno* [P85840]	61											
Similar a cadeia alfa-1 do peptídeo ativador de lipólise de <i>L. mucronatus</i> * [novo]	61											
Similar a proteína do veneno de <i>B. eupeus</i> * [novo]	61	2.888	1.629									
Similar a proteína secretada de <i>Opisthacanthus cayaporum</i> [novo]	61											
Similar a TsPep3 ou Ts13 [POC176]	61	135										
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											
Similar a TsPep2 ou Ts12 [POC175]	61											
Similar a TsPep3 ou Ts13 [POC176]	61											
Similar a TsPep1 ou Ts11 [POC174]	61											

Tabela 11: Anotação dos transcritos que foram responsáveis por 50% da similaridade entre *T. serrulatus* e os demais transcriptomas de quelicerados. Para cada transcriptoma selecionou-se os contigs contendo o maior número de ESTs e que juntos representaram a metade do número de transcritos que deram *match* com seqüências de *T. serrulatus*. O Contig foi anotado através de pesquisa contra o nr usando a ferramenta BLASTx. Na tabela, cada x representa um contig e o número entre parêntese indica o número de ESTs incluídos nele. O número total de match entre *T. serrulatus* e cada um dos demais transcriptomas aparece no cabeçalho entre parêntese ao lado da sigla indicativa da espécie a partir da qual foi feito o transcriptoma.

	Lm (235)	Mg (312)	Hp (53)	Sj (204)	Na (95)	De (127)	Li (159)	Lp (2719)	Cr (63)	Ss (204)	Ir (345)	Es (524)
Produção de energia												
ATP sintase	x (5)		x (2)		x (3)	x (7)	x (3)	x (45)	x (2)	x (21)	x (29)	x (14)
Citocromo c oxidase					x (3)	xx (30+9)	xxx (3+2+2)	x (168)	xxx (5+4+3)	xxxx (23+18+15+14)	xxx (37+31+29)	xx (47+11)
Citocromo b					x (3)		x (2)					x (22)
NADH desidrogenase					xx (5+4)		xxxx (2+2+2+2)	xx (107+35)				
Translocase ADP/ATP					x (8)							
Síntese, processamento e degradação proteica												
Fator de alongação da tradução					xxx(9+4+3)			x (62)			x (23)	xx(23+22)
Fator de iniciação da tradução 5A					x (4)							x (11)
Proteína disulfeto isomerase												
Proteína de choque térmico												
Proteína ribossomal 60S												x (18)
Ubiquitina C												
Componentes estruturais e/ou envolvidos com o movimento celular												
Actina	x (4)				x (4)			xx (76+33)				xxx (25+24+11)
Cadeia pesada da miosina	x (4)							x (35)				
Cadeia leve da miosina								x (43)				
Paramiosina												
Proteína contendo domínio LIM								xxx (255+38+31)				x (15)
Proteína de cutícula 14												x (10)
Proteína 20 específica do músculo		x (22)						xx (257+40)				
Tropomiosina								x (33)				
Troponina								xx (90+31)				
Outros componentes												
Alfa-B-cristalina												
Antígeno CD63												
Arginina cinase												
Defensina		x (17)										
Desidrogenase retinal												
Enzima conversora de angiotensina												
Hemocianina												
Hipotética proteína de <i>Thermobia domestica</i>												
Hipotética proteína de <i>T. discrepans</i>												
Peptídeo semelhante a orcoquina												
Protease semelhante a quimotripsina												
Proteína de defesa I(2)34Fc		x (12)										
Proteína de ligação a ácidos graxos												
Proteína semelhante a elastase												
Provável proteína F-box L168		x (27)										
Serina protease												
XP_001988913.1 GH11424 [Drosophila grimshawi]												
no match.		xxx (77+17+16)					xxxxxx (4+3+3+2+2)					
Componentes do veneno												
Alérgeno do veneno												
Cadeia alfa do peptídeo ativador de lipólise		xx (7 + 8)										
Fosfolipase A2												
Hipotética proteína secretada de <i>H. judaicus</i>												
Neuropeptídeo-3 do veneno de <i>M. eupeus</i>												
Neurotoxina LmNaTx11.1												
Peptídeo aniônico		x (15)										
Peptídeo antimicrobiano		x (24)										
Peptídeo potencializador de bradicinina		x (22)										
Ponericina-W-like 32.1		x (8)										
Proteína do veneno rica em cisteína		x (5)										

Alérgeno do veneno

Cadeia alfa do peptídeo ativador de lipólise

Fosfolipase A2

Hipotética proteína secretada de *H. judaicus*

Neuropeptídeo-3 do veneno de *M. eupeus*

Neurotoxina LmNaTx11.1

Peptídeo aniônico

Peptídeo antimicrobiano

Peptídeo potencializador de bradicinina

Ponericina-W-like 32.1

Proteína do veneno rica em cisteína

Ts = *Tiroyus serrulatus*; Lm = *Mesobuthus gibbosus*; Mg = *Heterometrus petersii*; Sp = *Scorpiops jendeki*; Li = *Loxocelles laeta*; Na = *Nephila antipodiana*; De = *Dysdera erythrina*; Ir = *Ixodes ricinus*; Ss = *Sarcopites scabiei*; Lp = *Limulus polyphemus*; Cr = *Carciniscorpius rotundicauda*; Es = *Endeis spinosa*.

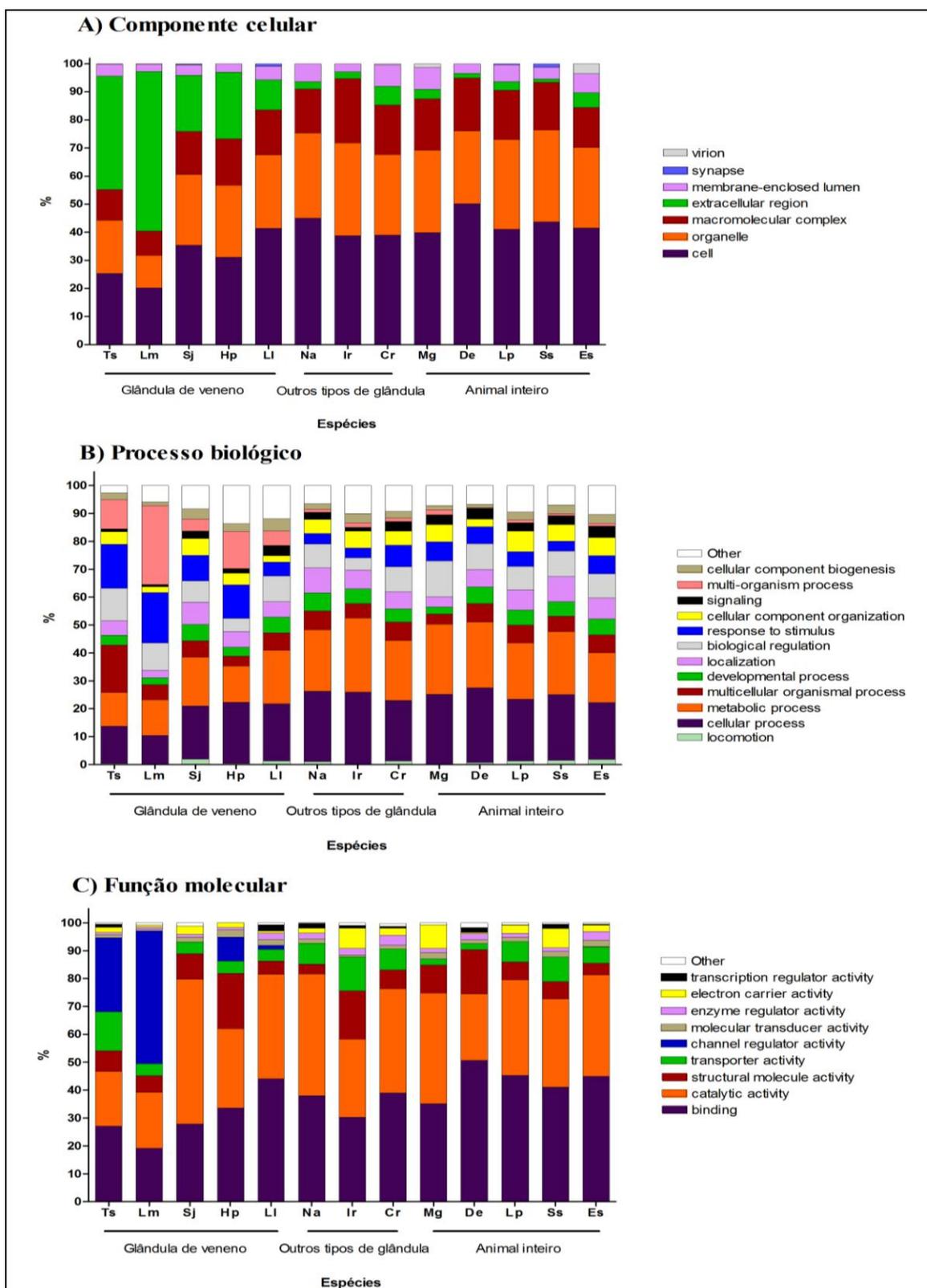


Figura 30: Análises de 13 transcriptomas de quelicerados quanto a distribuição dos termos do GO para os domínios componente celular (A), processo biológico (B) e função molecular (C). Como exemplo, a seguinte distribuição de anotações realizada pelo programa Blast2GO quanto ao domínio componente celular foi obtida para o transcriptoma de *T. serrulatus*: cell = 25%; organelle = 19%, macromolecular complex = 11%; extracellular region = 40%; and membrane-enclosed lumen = 4%. Ts = *Tityus serrulatus*; Lm = *Lynchas mucronatus*; Mg = *Mesobuthus gibbosus*; Hp = *Heterometrus petersii*; Sp = *Scorpiops jendeki*; LI = *Loxocelés laeta*; Na = *Nephila antipodiana*; De = *Dysdera erythrina*; Lp = *Limulus polyphemus*; Cr = *Carcinoscorpius rotundicauda*; Ir = *Ixodes ricinus*; Ss = *Sarcoptes scabiei*; Es = *Endeis spinosa*.

6. DISCUSSÃO

6.1 O transcriptoma como ferramenta de estudo

O presente trabalho buscou estudar a diversidade e complexidade molecular da glândula de veneno de *T. serrulatus* por meio da avaliação do repertório de transcritos usando como estratégia a descrição das ESTs. Assim, nessa análise foram obtidas 1.629 ESTs de excelente qualidade. A análise dos catálogos de transcritos tem proporcionado uma visão mais holística sobre as glândulas de veneno e, dessa forma, essa metodologia tem contribuído para acelerar a descoberta de peptídeos e proteínas sintetizados nesse órgão (Rodriguez de La Vega *et al.*, 2010). De acordo com o nosso conhecimento, existem onze estudos de transcriptoma da glândula de veneno até o momento (Schwartz *et al.*, 2007; Kozminsky-Atias *et al.*, 2008; D'Suze *et al.*, 2009; Ma *et al.*, 2009; Silva *et al.*, 2009; Ma *et al.*, 2010; Ruiming *et al.*, 2010, Morgenstern *et al.*, 2011, Ma *et al.*, 2011, Rendón-Anaya *et al.*, 2012; Almeida *et al.*, 2012). Entre eles, apenas o transcriptoma de *C. noxius* foi realizado com sequenciamento de nova geração (pirosequenciamento) no qual mais de 3 milhões de sequências foram produzidas. Nos demais trabalhos, o número de ESTs analisadas varia de 127 na espécie *T. discrepans* a 871 em *Scorpiops jendeki*. No conjunto desses estudos, foram avaliadas em média cerca de 450 sequências e, dessa forma, o presente estudo fornece em média cerca de 4 vezes mais sequências descritas para uma espécie, comparando aos demais trabalhos. Assim, esse estudo representa o maior catálogo de transcrito de glândula de veneno de escorpiões obtidos por sequenciamento de Sanger. Esse maior número de ESTs analisadas apresenta a vantagem de possibilitar a identificação de transcritos que ocorrem em baixa frequência na glândula. Assim, a opção de avaliar um grande volume de dados contribui para a descoberta de novos genes. Além do maior número de ESTs analisadas, o presente trabalho tem como diferencial a análise comparativa entre transcriptomas de quelicerados. O aumento do número de transcriptomas disponíveis em banco de dados possibilitou a análise comparativa e inferências sobre a evolução molecular nesse subfilo. Tal avaliação permitiu, por exemplo, identificar similaridades e particularidades da transcrição de glândulas de veneno.

Neste trabalho, 35 diferentes famílias de peptídeos codificados por 594 ESTs foram relacionadas com componentes do veneno. Muitas toxinas e componentes do veneno de *T. serrulatus* tem sido isolados e estudados individualmente (Possani *et al.*, 1977; Arantes *et*

al., 1989, Sampaio *et al.*, 1991; Martin-Eauclaire *et al.*, 1994; Sampaio *et al.*, 1996; Guatimosim *et al.*, 1999; Novello *et al.*, 1999; Petricevich *et al.*, 2007; Mendes *et al.*, 2008; Verano-Braga *et al.*, 2008). Além disso, o veneno desse escorpião já foi explorado por análises proteômicas (Pimenta *et al.*, 2001; Rates *et al.*, 2008). Assim, apesar do veneno de *T. serrulatus* ser um dos mais estudados, a análise do transcriptoma ainda revelou pelo menos três sequências completas de potenciais toxinas e mais de uma dezena de novos possíveis componentes do veneno. Essa é possivelmente uma lista não conclusiva de componentes do veneno dessa espécie, uma vez que um grande número de sequências sem resultado na busca por similaridade (ou seja, ESTs sem *match*) foi obtido. Tais transcritos estão sendo investigados pelo nosso laboratório para identificar novos componentes do veneno.

Análises de transcriptoma e proteoma têm sido normalmente usadas para explorar a complexidade dos componentes do veneno (Ma *et al.*, 2009). Análise do proteoma do veneno de *T. serrulatus* foi realizada por Pimenta *et al.* (2001a). Nesse trabalho, o veneno foi submetido à fitração em gel que separou os constituintes de acordo com o peso molecular. Como resultado, duas frações, a II e a III, corresponderam a 90 e 10% de todo o efeito tóxico do veneno, respectivamente. A massa molecular da fração II esteve entre 6.500 a 7.500 Da, que corresponde a toxinas de cadeias longas que agem principalmente em canais para Na⁺ dependentes de voltagem. A fração III foi mais complexa e predominantemente continha moléculas com massas entre 2.500 a 5.000 Da, que corresponde à família de toxinas de cadeias curtas, que agem em canais para K⁺, além de outros peptídeos desconhecidos. Esses autores obtiveram a massa molecular de 380 componentes presentes nas duas frações investigadas. No presente estudo, foram identificadas 35 diferentes famílias de componentes secretados codificados por 594 ESTs que possivelmente correspondem aos constituintes do veneno. Eles somaram 36% do total de ESTs e 47% dos transcritos com função identificada. Esse número de constituintes do veneno foi bem menor do que o descrito no estudo do proteoma (380 componentes). Comparações entre proteoma e transcriptoma dos escorpiões *T. discrepans* (Batista *et al.*, 2006; D'Suze *et al.*, 2009), *O cayaporum* (Schwartz *et al.*, 2008; Silva *et al.*, 2009) e *P. cavimanus* (Diego-Garcia *et al.*, 2012) encontraram a mesma tendência. Essa diferença pode estar relacionada ao fato de que uma parte das massas moleculares não identificadas no estudo de proteoma corresponde a fragmentos espúrios derivados de proteínas maiores, ou que o processo de agrupamento das ESTs que associam transcritos 95-100% similares

gera uma subestimativa do número de sequências precursoras (Rodríguez de La Vega *et al.*, 2010). Alternativamente, modificações pós-traducionais de uma mesma cadeia polipeptídica resultaria em uma expansão no número de sequências no proteoma (Pimenta *et al.*, 2001). Além disso, deve-se considerar que essa lista de componentes do veneno obtida no estudo do transcriptoma é possivelmente incompleta e muitos constituintes podem não ter sido caracterizados e estão incluídos nas sequências consideradas ‘*no match*’. Outros estudos realizados com animais venenosos demonstram que análises de proteoma e transcriptoma também divergem em relação à abundância relativa de diferentes famílias de toxinas (Calvete *et al.*, 2007; Sanz *et al.*, 2008)

Apesar das ferramentas de proteoma fornecerem uma visão precisa sobre a diversidade molecular dos venenos de escorpião, o reconhecimento de massas moleculares sem identificação prévia representa o principal desafio dessa tecnologia, uma vez que grande parte das proteínas do veneno em questão não possui similaridade com sequências de banco de dados. Nesse sentido, estudos de transcriptoma auxiliariam na obtenção uma visão mais completa da diversidade molecular dos venenos de escorpião (Ma *et al.*, 2010; Rodríguez de La Vega *et al.*, 2010). De acordo com Junqueira-de-Azevedo *et al.* (2006) e Schwartz *et al.* (2007), a estratégia baseada na descrição das ESTs tem se mostrado efetiva na descoberta da diversidade real da composição do veneno, permitindo caracterizar não apenas sequências de toxinas conhecidas, mas também moléculas atípicas expressas pela glândula de veneno (Ma *et al.*, 2009). Por essa razão, neste trabalho, a ferramenta de descrição e análise das ESTs da glândula de veneno de *T. serrulatus* foi adotada como estratégia de se produzir uma visão geral dos transcritos presentes nesse órgão.

Embora incompletos, os catálogos de transcritos de glândulas de veneno de escorpião fornecem uma figura geral dos processos celulares que ocorrem no estado fisiológico e ontológico no momento em que a biblioteca de cDNA é formada (Rodríguez de la Vega *et al.*, 2010). Deve-se salientar que, dois dias antes da extração de RNA, os escorpiões utilizados neste trabalho foram submetidos à estimulação elétrica do télson para coleta de veneno. Essa prática também foi realizada no estudo de transcriptoma de outras espécies com variações do número de dias entre a extração do veneno e a do RNAm. Assim, o intervalo de dois dias foi utilizado nos trabalhos de Kozminsky-Atias *et al.* (2008), Ma *et al.* (2009; 2010) e de Ruiming *et al.* (2010), de quatro dias no estudo de Silva *et al.* (2009) e de cinco dias nos trabalhos de Schwartz *et al.* (2007), D’Suze *et al.* (2009) e Rendón-Anaya *et al.* (2012). Existem três trabalhos de transcriptoma de glândula

de veneno de escorpião, realizados com as espécies *H. judaicus* (Morgenstern *et al.*, 2011), *T. stigmurus* (Almeida *et al.*, 2012) e *C. noxius* (Rendón-Anaya *et al.*, 2012) nos quais a extração do veneno antes da montagem da biblioteca de cDNA não foi realizada. A extração prévia do veneno faz com que, no momento da obtenção do RNAm da glândula, esse órgão apresente um padrão de expressão gênica que representa o seu estado ativado. De acordo com específicas observações do curso do tempo de produção do transcrito da toxina, o nível máximo de RNA mensageiro ocorre 24 a 48 horas após a extração do veneno (Alami *et al.*, 2001; Zeng *et al.*, 2002). De fato, a proporção dos componentes do veneno em *H. judaicus* (não submetido a extração prévia) foi de 24% do total de transcritos, um valor consideravelmente mais baixo do que o obtido em outros butídeos. Os transcritos para possíveis constituintes do veneno de *T. serrulatus* correspondem a no mínimo 36% do total das ESTs. Esse valor encontra-se intermediário aos obtidos para *S. jendeki* (50% das ESTs corresponderam a toxinas do veneno) (Ma *et al.*, 2009) e *H. gertschi* (aproximadamente 30% das ESTs) (Schwartz *et al.*, 2007). Embora as diferenças na proporção de constituintes do veneno podem ser em parte influenciadas pela metodologia utilizada, muitos autores relacionam essas diferenças de expressão a variações genéticas entre as espécies e mesmo entre populações de uma mesma espécie (Batista *et al.*, 2006; Ruiming *et al.*, 2010).

6.2 ESTs relevantes para processos celulares

De maneira geral, a diversidade e a natureza dos transcritos anotados nos estudos de transcriptoma fornecem uma visão sobre os processos moleculares que ocorrem nas células da glândula de veneno (Schwartz *et al.*, 2007). Assim, um grande número de transcritos apresentando a mesma anotação ou sendo anotados como componentes de uma mesma via metabólica ou ainda a presença de diferentes transcritos em uma mesma categoria do *Gene Ontology* podem sugerir que tal via metabólica/categoria é muito importante dentro do processo biológico da glândula de veneno do escorpião. Por exemplo, intuitivamente a glândula de veneno deve dar suporte a intensa síntese e secreção proteicas para produzir grande quantidade de proteínas do veneno (Rodríguez de La Vega, 2010). De fato, diferentes metodologias como análise histológica, anotação manual, anotação funcional utilizando o programa Blast2GO, mapeamento das ESTs e unives por meio das Vias KEGG indicaram os processos celulares de destaque na glândula de veneno, que foram:

síntese proteica, produção de energia e componentes do citoesqueleto possivelmente relacionados à manutenção da estrutura da glândula e contração durante a injeção do veneno. Apesar da indiscutível relevância de tais processos biológicos na glândula de veneno, eles constituem vias comuns a maioria das células. Assim, análises comparativas entre transcriptomas são necessárias para avaliar se esses processos são realmente mais acentuados na glândula de veneno e quais são as particularidades desse órgão quanto a expressão gênica.

Quando os catálogos de transcritos não estão disponíveis em banco de dados públicos, a única forma de se comparar os resultados dos transcriptomas é extrair os dados descritos na publicação. Entretanto, como as metodologias de análises empregadas (valores de corte de qualidade e de E-value, método de anotação, dentre outros) em diferentes estudos não são as mesmas, o exercício de comparação torna-se limitado. Dessa maneira, confiáveis generalizações a cerca da ocorrência e expressão dos transcritos geralmente não podem ser feitas. Assim, no presente estudo, além da comparação com os dados da literatura, optou-se por analisar diversos transcriptomas de quelicerados, e a partir daí, realizar as comparações com o catálogo de transcrito de *T. serrulatus* a fim de obter uma visão mais holística da transcrição em glândulas de veneno.

Uma vez que a glândula de veneno é um órgão especializado na síntese de proteínas e peptídeos, esperava-se uma considerável proporção de transcritos envolvidos nesse processo. De fato, 5% do total de transcritos e 20% dos únicos e 17% das ESTs anotadas como componentes celulares codificaram proteínas envolvidas diretamente com a síntese e processamento proteico. Para uma outra espécie de escorpião, *H. gertschi*, 8,2% do total de transcritos e 16,2% de sequências únicas apresentaram similaridade com componentes ribossomais ou proteínas envolvidas no tráfego celular (Schwartz *et al.*, 2007). Resultados similares foram obtidos por Fernandes-Pedrosa *et al.* (2008) em um estudo de transcriptoma da glândula de veneno da aranha *L. laeta*, no qual ESTs relacionadas ao processo de transcrição e tradução foram as mais abundantes dentro da categoria “proteínas celulares” (corresponderam a 8%). Interessantemente, comparando os dados obtidos de diferentes transcriptomas de glândulas de veneno, percebe-se que a proporção de transcritos relacionados com a síntese proteica foi consideravelmente semelhante nas diferentes espécies analisadas. Análises da anotação dos transcriptomas com os termos do GO e anotação manual não encontraram consideráveis diferenças na proporção de transcritos relacionados com a síntese e processamento proteico entre os transcriptomas de

glândula de veneno e outras glândulas ou animais inteiros. Aparentemente, tais transcritos não estão aumentados nas glândulas de veneno. Para exemplificar, transcritos que codificam fatores de alongação da tradução corresponderam a 0,4, 0,7 e 1,1% do total de ESTs nos transcriptomas de *T. serrulatus* (glândula de veneno), *L. polyphemus* (animal inteiro) e *I. ricinus* (glândula salivar), respectivamente.

A importância do correto processamento proteico no contexto da glândula de veneno de escorpião é enfatizada pela presença de transcritos codificando proteínas relacionadas ao dobramento adequado de peptídeos, processamento pós-traducional ou degradação dependente de proteossoma (Schwartz *et al.*, 2007). Em *T. serrulatus*, ESTs relacionadas diretamente com esses processos corresponderam a 3,4% dos transcritos incluídos na categoria proteínas celulares. Processamento, direcionamento e degradação de proteínas somaram 5% das ESTs de *L. laeta*, sendo essa categoria formada principalmente por proteínas de choque térmico (Hsp), dissulfeto isomerase e ubiquitina (Fernandes-Pedrosa *et al.*, 2008). No estudo do transcriptoma de *H. gertschi*, Schwartz e *et al.* (2007) deram destaque a dois grupos de ESTs envolvidos com o processamento pós-traducional de proteínas. Um deles apresentou *match* com enzimas de glicosilação β -1,4-manosiltransferases, que é consistente com a presença de proteínas glicosiladas no veneno de escorpiões (Valdez-Cruz *et al.*, 2004). E o outro, com similaridade a aminopeptidases da família M17, é possivelmente secretado como componente do veneno e poderia ser importante para modificações pós-traducionais de componentes da glândula de veneno. No presente estudo, um transcrito codificando PAM (peptidylglycine α -amidating monooxygenase) foi identificado. Alguns estudos indicam que tal enzima pode estar envolvida com a amidação C-terminal de peptídeos do veneno de escorpiões (Rates *et al.*, 2008; Diego-García *et al.*, 2005; Barona *et al.*, 2006), modificação essa que é possivelmente necessária para que o peptídeo exerça sua função biológica (Jiang *et al.*, 2000). De fato, a provável NaTx identificada no presente estudo (Ts17) apresenta na sua porção C-terminal a sequência GKK, que é clivada e o resíduo de glicina fornece o grupo amina para a amidação da prolina que precede esses três aminoácidos. Esse mesmo processamento é realizado para TdNa8 de *T. discrepans* e TbTx5 de *T. bahiensis*, nas quais a prolina também é amidada e para Ts3 de *T. serrulatus*, na qual a serina precedente aos três aminoácidos C-terminais é amidada (Schwartz *et al.*, 2007; D'Suze *et al.*, 2009).

Os processos de síntese e secreção proteica são considerados energeticamente custosos para a célula. Nesse contexto, escorpiões que tiveram seus venenos artificialmente

depletados demonstraram aumento do consumo de oxigênio e estima-se que 40% de toda taxa metabólica do escorpião *Parabuthus transvaalicus* esteja relacionada com a produção do veneno (Nisani *et al.*, 2007). No presente estudo, considerável parcela das ESTs codificou proteínas mitocondriais (4% de todas as ESTs e 14% das ESTs correspondentes a proteínas celulares). Porcentagens semelhantes foram obtidas para *H. gertschi*, no qual 4,1% de todas as ESTs e 7,4% de precursores de proteínas identificadas apresentaram similaridade com componentes da mitocôndria (Schwartz *et al.*, 2007). Apesar desse considerável percentual de transcritos para proteínas mitocondriais, alguns componentes envolvidos com a produção de energia não estão aumentados em glândulas de veneno quando comparados a outras glândulas ou a animais inteiros. Como exemplo, transcritos codificantes de ATP sintase corresponderam 0,2, 0,7 e 0,8 % das ESTs provenientes dos transcriptomas de glândulas de veneno de *L. laeta*, *T. serrulatus* e *H. petersii*, respectivamente, enquanto nos catálogos de transcritos dos animais inteiros *D. erythrina* e *S. scabiei*, tais componentes corresponderam a 1 e 2% do total de ESTs. Em outro exemplo, os níveis de citocromo b foram muito semelhantes para transcriptomas de glândula de seda (*N. antipodiana*), glândula de veneno (*T. serrulatus*) e de animal inteiro (*E. spinosa*) e corresponderam a 0,5, 0,4 e 0,5% das ESTs, respectivamente.

A análise histológica da glândula de veneno de *T. serrulatus* mostra uma espessa camada muscular envolvendo esse órgão. Transcritos provavelmente expressos nesse tecido associados com outros responsáveis pela manutenção da estrutura da glândula corresponderam a 18% das ESTs relacionadas com componentes celulares. A comparação da expressão de transcritos pertencentes a essa categoria não foi muito diferente entre os transcriptomas analisados. Como exemplo, transcritos para actina perfizeram 1% das ESTs de *T. serrulatus*, 0,5% de *L. mucronatus* (glândula de veneno), 0,8% de *N. antipodiana* (glândula da seda), 1,2% de *L. polyphemus* (animal inteiro) e 1,4% de *E. spinosa* (animal inteiro).

Conforme descrito acima, o nível de expressão de genes envolvidos com a homeostase celular é relativamente semelhante entre os diferentes transcriptomas de quelicerados amostrados. As análises usando a ferramenta Blast2GO também mostraram um padrão de distribuição dos termos do GO bastante similar entre os catálogos de transcritos estudados. A comparação entre os transcriptomas evidenciou as particularidades das glândulas de veneno, que se destacaram dos demais catálogos de transcritos devido as ESTs que codificam componentes secretados e que irão constituir as proteínas/peptídeos

do veneno. Corroborando esses dados, a análise comparativa do perfil transcricional do telson e do escorpião inteiro da espécie *C. noxius* indicou que a maioria dos transcritos (73%) foi encontrada tanto no corpo quanto no telson, sendo 3,5% específicos do corpo do escorpião e 23,8% exclusivos do telson (Rendón-Anaya *et al.*, 2012).

6.3 ESTs relacionadas a componentes do veneno

Considerando as informações já descritas sobre a composição do veneno (Pimenta *et al.*, 2001; Rates *et al.*, 2008; Cologna *et al.*, 2009), alto nível de expressão de neurotoxinas era esperado. De fato, os principais agentes responsáveis pelos efeitos tóxicos (as NaTx) durante o envenenamento por *T. serrulatus* apresentaram alto nível de expressão (correspondendo a 6% do total de transcritos). Em particular, grande número de transcritos foi observado para Ts1 e Ts2. No transcriptoma de *T. stigmurus*, outro escorpião que causa acidentes no Brasil, a proporção de NaTx foi de 1,3% do total de transcritos (Almeida *et al.*, 2012). Uma vez que esses dois escorpiões do gênero *Tityus* apresentam toxinas bastante similares, a maior proporção de NaTx em *T. serrulatus* pode estar relacionado a maior letalidade dessa espécie quando comparada a *T. stigmurus*. Além das sequências de toxinas mais comuns, dois uniques distintos chamados no presente trabalho de Ts17 e Ts18 apresentaram alta similaridade com NaTxs de outros escorpiões e podem constituir novas toxinas cujas as atividades biológicas precisam ser investigadas.

KTxs também apresentaram grande número de ESTs (correspondendo a 8% do total de transcritos), entretanto, essa alta expressão foi observada apenas para uma KTx descrita (Ts8, também chamada Tityustoxin K-beta, TsTX-K beta, TSK2), que correspondeu cerca de 13% dos componentes do veneno. Um unique com alta similaridade a KTx que foi aqui denominado Ts19 apresentou expressão relativamente alta na glândula de veneno de *T. serrulatus* e pode consistir em uma nova sequência completa de toxina. Apesar da alta expressão, a proporção de KTxs no transcriptoma de *T. serrulatus* foi menor do que a encontrada em *T. stigmurus* (KTx=13,5% do total de transcritos) (Almeida *et al.*, 2012). A diferença encontrada entre a proporção de NaTx e KTx entre esses dois escorpiões pode ser espécie-específica ou refletir o perfil transcricional da glândula de veneno em repouso em *T. stigmurus* e ativa em *T. serrulatus*. De fato, no transcriptoma da glândula em repouso de *H. judaicus*, a proporção de NaTx foi subrepresentada e as toxinas mais abundantes foram KTxs. Assim, parece ocorrer uma expressão diferencial entre KTxs

e NaTxS nos estados ativos e em repouso na glândula de veneno de butídeos. Uma possível implicação biológica dessa observação seria que dada a importância das NaTxS para a sobrevivência dos escorpiões, a reposição dessas toxinas após a utilização ou extração do veneno seria rápida.

Além das neurotoxinas, a glândula de veneno de *T. serrulatus* apresentou intensa expressão de peptídeo Pape, AMPs, metaloproteases, peptídeo aniônico e hipotensinas. Considerando essa alta expressão, esses constituintes podem apresentar importantes funções na glândula de veneno de escorpiões e serão discutidos abaixo.

ESTs com similaridade a BPP (Peptídeo Pape) tiveram alta ocorrência no transcriptoma de *T. serrulatus*, correspondendo a 23% das ESTs relacionadas a componentes do veneno. Considerável presença desse peptídeo também foi detectada pelo estudo do proteoma da glândula de veneno de *T. serrulatus*, no qual os BPPs perfizeram a metade das massas moleculares identificadas (Rates *et al.*, 2008; Verano-Braga *et al.*, 2008). Sequências homólogas a obtida no transcriptoma de *T. serrulatus* foram descritas em diversas espécies como *T. costatus* (Diego-García *et al.*, 2005), *T. stigmurus* (Batista *et al.*, 2007), *T. discrepans* (Batista *et al.*, 2006; D'Suze *et al.*, 2009), *L. mucronatus* (Ruiming *et al.*, 2010), *M. martensii* (Zeng *et al.*, 2000) e *Buthus occitanus* (Meki *et al.*, 1995).

Rates *et al.* (2008) indicaram a ocorrência de peptídeos similares a Ponericina-L1 e Ponericina-L2 na análise do peptidoma do veneno de *T. serrulatus*. No presente trabalho, foram identificadas a sequência completa desse peptídeo e outras três sequências de AMPs. AMPs tem sido descritas no veneno de várias espécies de escorpiões (Schwartz *et al.*, 2007; Ma *et al.*, 2009; Silva *et al.*, 2009; Ma *et al.*, 2010; Ruiming *et al.*, 2010), sendo observado alto nível de expressão em algumas delas (Ma *et al.*, 2009; Ma *et al.*, 2010). Ainda não se tem um consenso do papel desses peptídeos no veneno de escorpiões, entretanto, algumas funções têm sido sugeridas. AMPs podem proteger os escorpiões contra infecção bacteriana via o par de ductos que diretamente conecta a glândula de veneno ao ambiente externo. Além disso, AMPs podem despolarizar células neurais induzindo imobilização da presa permitindo que os escorpiões a capture e podem potencializar ações de outras neurotoxinas encontradas no veneno (Carballar-Lejarazú *et al.*, 2008). Recentemente, peptídeos antimicrobianos tem se tornado foco da toxinologia, uma vez que novas substâncias e arcabouços moleculares são necessários para dar resposta à ocorrência de cepas de bactérias resistentes (Pimenta e De Lima, 2005). Por exemplo, o

AMP chamado scorpina, foi produzido em sistema heterólogo e demonstrou atividade contra diferentes patógenos: (1) efeitos antibacterianos contra *Bacillus subtilis* e *Klebsiella pneumoniae*; (2) mortalidade de estágios sexuais de *Plasmodium berghei* e *P. falciparum*; e (3) inibição da replicação do vírus da dengue-2 em células de mosquito C6/36 (Carballar-Lejarazú *et al.*, 2008). Outro AMP, denominado mucroporina, foi capaz de causar efeitos de inibição bacteriana (Dai *et al.*, 2008). Esse peptídeo causou morte de *Staphylococcus aureus* muito rapidamente, e pode efetivamente inibir alguns patógenos resistentes a antibióticos (Dai *et al.*, 2008). Dessa forma, apesar de ensaios microbiológicos serem necessários, os AMPs descritos no presente trabalho podem ser considerados potenciais candidatos a antibióticos.

Morgenstern *et al.* (2011) demonstraram grande abundância de proteases como metaloproteases na glândula de veneno em repouso do Buthidae *H. judaicus*. Interessantemente, todos os animais usados no presente trabalho tiveram o veneno extraído antes da extração de RNAm e ainda assim o nível de expressão de metaloproteases foi tão alto como o de alguns transcritos codificantes de NaTx e KTx. Prosdocimi *et al.* (2011) encontrou alta expressão de metaloproteases da família *astacin* no transcriptoma da glândula de seda de *Gasteracantha cancriformis* e sugeriram que tais proteínas desempenham um papel no processo de remodelamento durante a deposição da fibra da seda. Uma grande expressão de proteases foi também observada no veneno de *L. mucronatus* sugerindo que essas proteínas podem desempenhar função central no veneno de escorpiões (Ruiming *et al.*, 2010). De fato, certas metaloproteases, chamadas antareases, purificadas a partir do veneno de *T. serrulatus* clivaram proteínas SNARE dentro do tecido pancreático e podem estar relacionadas com a pancreatite desenvolvida durante o envenenamento por escorpiões (Fletcher *et al.*, 2010). Alternativamente, proteases como metaloproteases podem desempenhar papéis específicos na glândula de veneno. Um grande número de fragmentos derivados de peptídeos maiores tem sido obtido em estudos de proteoma do veneno de escorpiões (Pimenta *et al.*, 2001; Rates *et al.*, 2008). Aparentemente diferentes fragmentos derivados do mesmo peptídeo podem apresentar atividade biológica distinta (Diego-García *et al.*, 2005). Assim, pode-se especular que a fragmentação de peptídeos realizada por proteases pode causar enorme aumento na diversidade de peptídeos do veneno e conseqüentemente, amplificar a possibilidade de alvos biológicos.

Peptídeos aniônicos são ricos em resíduos de ácido aspártico e glutâmico (verificar a Fig. 23). Aparentemente, peptídeos aniônicos são altamente expressos e conservados entre escorpiões pertencentes à família Buthidae (D'Suze *et al.*, 2009; Ma *et al.*, 2011). Embora a função desses peptídeos ainda não tenha sido esclarecida (Diego-García *et al.*, 2005), alguns autores sugerem que eles podem ter atividade antimicrobiana (Harris *et al.*, 2009) ou desempenhar importante papel no balanço de pH, uma vez que o seu caráter ácido contrapõe as neurotoxinas que são peptídeos básicos (Ma *et al.*, 2009; Ruiming *et al.*, 2010).

Hipotensinas foram identificadas pela primeira vez no veneno de *T. serrulatus* através de ferramentas de proteômica (Verano-Braga *et al.*, 2008). No presente estudo, o precursor desse componente do veneno foi obtido. Esse peptídeo linear é caracterizado pela presença de uma assinatura de aminoácidos de peptídeo potencializador de bradicinina e exerce seu efeito anti-hipertensivo de maneira dependente de óxido nítrico e independente da enzima conversora de angiotensina (ECA) (Verano-Braga *et al.*, 2008). Das três sequências de hipotensinas encontradas no *Animal toxin annotation program* do UniProt (hipotensina-1, UniProt ID: P84189; hipotensina-2, UniProt ID: P84190; hipotensina-like peptide, UniProt ID: P86824), a sequência completa dos precursores de hipotensina-like e hipotensina-1 foram identificadas (Fig. 27).

Além dos grupos de peptídeos citados acima, o transcriptoma revelou um grande número de transcritos do veneno com frequência de ocorrência menor que 1% do total de ESTs. Esses transcritos são representados por fosfolipases A2, alérgenos, hialuronidases, e diversos peptídeos com similaridade com componentes do veneno de escorpiões de outras espécies ou mesmo de outros organismos. Acredita-se que esses outros componentes possam mediar respostas tóxicas ou potencializar a ação das toxinas. Assim, a fosfolipase A2, por exemplo, é possivelmente mediadora de várias respostas tóxicas como a citotoxicidade, neurotoxicidade, miotoxicidade, edema, distúrbios na coagulação sanguínea, além de estarem envolvidas com a quebra da estrutura de membranas celulares e o espalhamento do veneno (Zamudio *et al.*, 1997; Kuhn-Nentwing, 2003, Schwartz *et al.*, 2007). Apesar de vários transcriptomas identificarem uma diversidade de componentes do veneno, para uma considerável parte desses, não existem informações sobre sua implicação biológica. Assim, não se sabe, por exemplo, o papel de peptídeos ricos em glicina e prolina e peptídeos semelhantes a La1 encontrados no veneno de escorpiões (Ma *et al.*, 2012). Após a descrição dos componentes, são necessários estudos que realizam a expressão ou o

isolamento dos peptídeos/proteínas e subsequente análise quanto seus efeitos biológicos. Isso não só permite uma melhor compreensão do envenenamento, como também possibilita a identificação de moléculas úteis para o desenvolvimento de fármacos.

É interessante notar que alguns componentes do veneno de *T. serrulatus* identificados em estudos anteriores não foram encontrados na análise do transcriptoma. Como exemplos podem ser citados Peptide T (UniProt ID: Q9TWR4), Ts5 (UniProt ID: P45659) e alfa amilase (UniProt ID: P85843). É interessante notar que para o Peptídeo T, dos 13 aminoácidos que o compõem, 9 deles se alinham perfeitamente com a porção N-terminal da Ts3 e da Ts17 de *T. serrulatus*, o que levaria a se especular que esse peptídeo seria resultante de modificações dessas sequências. De qualquer forma, a ausência de transcritos dos componentes do veneno em transcriptomas não é surpreendente. De fato, mesmo o catálogo de transcritos de *C. noxius*, no qual a plataforma de pirosequenciamento foi utilizada resultando em mais de três milhões de sequências, não foi capaz de detectar a completa coleção de toxinas já descrita para a espécie (Rendón-Anaya *et al.*, 2012). Algumas hipóteses poderiam explicar a ausência de transcritos: (1) a análise de transcriptoma não fornece uma completa cobertura das ESTs; (2) genes de toxinas apresentavam nenhuma ou baixa expressão no momento de extração de RNA ou sofreram degradação mediada por microRNA durante o processamento, como sugerido por Rendón-Anaya *et al.* (2012); (3) as proteínas/peptídeos do veneno sofrem modificações pós-traducionais que significativamente as diferenciam do transcrito original, como descrito por Pimenta *et al.* (2001).

6.4 Ortólogos dos componentes do veneno de *T. serrulatus*: Inferência sobre a distribuição nos grupos de organismos

A maioria dos trabalhos de transcriptoma de glândula de veneno de escorpiões encontrados na literatura tem como foco os componentes do veneno e, dessa forma, muitos tipos de peptídeos e proteínas têm sido obtidos de diversas espécies de escorpião (Rodríguez de La Vega *et al.*, 2010). Alguns deles são amplamente distribuídos entre as espécies, como os pertencentes à família das α -KTxs, enquanto outros são aparentemente restritos a uma linhagem de escorpião em particular. A presença de moléculas do veneno comuns e incomuns entre diferentes linhagens reflete ao processo evolutivo dinâmico do arsenal do veneno (Ma *et al.*, 2009). Na tentativa de se obter uma visão panorâmica das famílias de peptídeos e proteínas que foram consideradas componentes do veneno em

estudos de transcriptomas, analisaram-se os 12 trabalhos disponíveis na literatura e as informações das glândulas de veneno das 14 espécies (incluindo *T. serrulatus*) foram organizadas e são apresentadas na Tabela 12.

O aumento da disponibilidade de informação molecular nos últimos anos, aliado ao desenvolvimento de ferramentas de bioinformática permitiu avaliar a distribuição dos ortólogos dos componentes do veneno de *T. serrulatus* nos grupos de organismos tanto no universo protéico quanto no de transcritos, o que possibilitou inferir em quais grupos de organismos determinado componentes do veneno está presente, o que indicaria aonde possivelmente ele surgiu. A busca de ortólogos de componentes do veneno de *T. serrulatus* no universo proteico e de transcritos através da ferramenta PSI-BLAST mostrou que cerca de 70% das famílias de peptídeos/proteínas do veneno foram recrutados em um desses três níveis: ordem Scorpiones, família Buthidae ou gênero *Tityus*. Esse percentual é predominantemente formado por peptídeos ricos em cisteína. Os 30% restantes apresentam ortólogos mais antigos e amplamente distribuídos ao longo dos grupos de organismos, e é constituído por enzimas (metaloproteases, hialuronidases e fosfatases) e por outros componentes pouco caracterizados similares a peptídeo/proteínas encontrados no veneno de outros escorpiões.

Para validar as análises de bioinformática, os resultados foram comparados com os dados disponíveis na literatura. A busca de ortólogos utilizando PSI-BLAST (Tabela 7) e a avaliação da ocorrência de componentes em transcriptomas de glândula de veneno da ordem Scorpiones (Tabela 12) foi coincidente para os seguintes constituintes: KTx (ordem Scorpiones), NaTx (família Buthidae), peptídeo aniônico (ordem Scorpiones), peptídeo ativador de lipólise (família Buthidae), hipotensinas (gênero *Tityus*), peptídeo rico em glicina e fosfolipase A2 (distribuição ampla nos grupos de organismos). A distribuição das KTx e NaTx é condizente com a história evolutiva dessas neurotoxinas, de acordo com a qual, as KTx surgiram na ordem Scorpiones e apresentam um ancestral comum com defensas de artrópodes. As NaTx, presentes apenas em membros da família Buthidae, seriam derivadas das KTx (Froy e Gurevitz, 2004). Embora NaTx tenham sido anteriormente identificada no veneno de *Anuroctonus phaiodactylus* pertencente a família Iuridae (Valdez-Cruz *et al.*, 2004), a busca realizada no presente trabalho contra banco de dados de proteína (UniProt) e ESTs (dbEST) não identificou tal peptídeo como ortólogo das sequências de *T. serrulatus*. Conforme também sugerido por outros autores (Ma *et al.*,

Tabela 12: Componentes do veneno identificados nos transcriptomas de glândulas de veneno de escorpiões.

Continua...

	Buthidae									
	<i>L. mucronatus</i>	<i>I. maculatus</i>	<i>H. judaicus</i>	<i>B. occitanus</i>	<i>T. discrepans</i>	<i>T. serrulatus</i>	<i>T. stigmurus</i>	<i>C. noxius</i>		
neurotoxinas	KTx X*	X	X	X	X	X	X	X	X*	X*
	NaTx	X	X	X*	X	X	X	X	X	X
	CaTx	X	X	X	X	X	X	X	X	X
	CITx	X	X	X	X	X	X	X	X	X
Scorpiinas										
Peptídeos com atividade antimicrobiana	NDBP3 ¹	X	X	X	X	X*	X*	X*	X*	X
	NDBP4 ²	X	X*	X*	X	X	X	X	X	X
	NDBP5 ³	X	X	X	X	X	X*	X*	X*	X
Outros componentes	Peptídeo anionico	X	X	X	X*	X	X*	X*	X*	X
	Peptídeo La-1	X	X	X	X	X	X	X	X	X
	Serina protease	X	X	X	X	X	X	X	X	X
	Metaloprotease	X	X*	X*	X	X	X	X	X	X
	Fosfolipase	X	X	X	X	X	X	X	X	X
	Inibidor de Serina Protease	X	X	X	X	X	X	X	X	X
	Hialuronidase	X	X	X	X	X	X	X	X	X
	Peptídeo rico em glicina	X	X	X	X	X	X	X	X	X
	Peptídeo rico em cisteína	X	X	X	X	X	X	X	X	X
	Peptídeo ativador de lipólise	X	X	X	X	X	X	X	X	X
	Hipotensina	X	X	X	X	X	X	X	X	X
	Alergeno	X	X	X	X	X	X	X	X	X
	Proteínas hipotéticas	X	X	X	X	X	X	X	X	X
Particularidade	variações intra-específicas no veneno	Peptídeo rico em prolina	pseudogênes identificados	evolução de neurotoxinas	Td defensiva	transcriptômica comparativa	Lectinas	Lipase		
Autores	Ruiming et al., 2010; Ma et al., 2012	Ma et al., 2012	Morgenstem et al., 2011	Kozminsky-Atias et al., 2008	D'Suze et al., 2009	Alvarenga et al., 2012	Almeida et al., 2012	Rendón-Anaya et al., 2012		

* indica família de peptídeo/proteína mais expressa na glândula de veneno
 NDBP: Peptídeo sem ponte dissulfeto.

¹ NDBP 3: inclui peptídeos como Pandimina1, BPPs, peptídeo Pape

² NDBP4: inclui, por exemplo, AMPs de *T. serrulatus*, Pandimina 2 e BmKb1

³ NDBP5: inclui peptídeos citolíticos como aqueles similares a IsCT, IsCT2, BmKn2 e BmKn2.

	Euscorpidae		Scorpionidae		Iuridae		...Conclusão.
	<i>S. marginisanae</i>	<i>S. jendeki</i>	<i>H. petersii</i>	<i>P. cavimanus</i>	<i>H. gertschi</i>	<i>O. cayaporum</i>	
neurotoxinas	KTx	x	x	x*	x*	x*	
	NaTx						
	CaTx	x	x	x		x	
	CITx						
Peptídeos com atividade antimicrobiana	Scorpinas	x	x	x	x	x	x
	NDBP3 ¹	x	x	x		x	
	NDBP4 ²		x	x	x	x	x
	NDBP5 ³	x	x*	x*		x	
Outros componentes	Peptídeo aniónico	x	x		x		x
	Peptídeo La-1	x	x	x	x		
	Serina protease	x	x	x			
	Metaloprotease						
	Fosfolipase			x		x	x
	Inibidor de Serina Protease					x	
	Hialuronidase						
	Peptídeo rico em glicina			x			
	Peptídeo rico em cisteína				x		
	Peptídeo ativador de lipólise						
Hipotensina							
Alergeno							
Proteínas hipotéticas			x	x			
Particularidade	Margarina*	Lizosina e jendinas	Peptídeo diurético e fosfatase ácida	Associação transcriptômica e proteômica	1º transcriptoma de glândula de veneno de	Comparação com o proteoma já realizado	
Autores	Ma et al., 2012	Ma et al., 2009	Ma et al., 2010	Diego-Garcia et al., 2012	Schwartz et al., 2007	Silva et al., 2009	

* indica família de peptídeo/proteína mais expressa na glândula de veneno
 NDBP: Peptídeo sem ponte dissulfeto.

¹ NDBP 3: inclui peptídeos como Pandinina1, BPPs, peptídeo Pape

² NDBP4: inclui, por exemplo, AMPs de *T. serrulatus*, Pandinina 2 e BmKb1

³ NDBP5: inclui peptídeos citolíticos como aqueles similares a IsCT, IsCT2, BmKn2 e BmKn2.

2012), um arcabouço molecular diferente daquele encontrado para as toxinas de butídeos pode explicar a ausência de NaTx em escorpiões não-butídeos.

Para alguns componentes do veneno não houve concordância entre as análises de PSI-BLAST e os dados da literatura. Essa discrepância ocorreu de duas formas: 1) a busca por ortólogos através do PSI-BLAST identificou componentes do veneno em um grupo mais amplos do que o encontrado nos trabalhos de transcriptomas; 2) de modo oposto, para alguns peptídeos, a análise de bioinformática encontrou ortólogos em um grupo mais restrito do que o relatado nos estudos de catálogos de transcritos. Metaloproteases, hialuronidases e alergenos se enquadram na situação 1, enquanto peptídeo sem ponte dissulfeto pertencente as famílias 3 e 4 (NDBP3 e NDBP4) na situação 2.

Os estudos de transcriptoma de glândula de veneno de não-butídeos não relataram a ocorrência de metaloproteases, hialuronidases e alergenos. Essas três famílias de proteínas são amplamente distribuídas nos grupos de organismos e estão presentes em venenos de aranhas, cobras e escorpiões (More *et al.*, 2004; Calvete *et al.*, 2009; Froy *et al.*, 2009; Madero *et al.*, 2009; Gremski *et al.*, 2010; Trevisan-Silva *et al.*, 2010). A ausência de transcritos que codificam essas proteínas nos transcriptomas de não-butídeos poderia indicar que elas não foram recrutadas para a glândula de veneno desses animais. Entretanto, as análises de tBLASTx e tBLASTn (Tabelas 9 e 10) indicaram a ocorrência de transcritos similares a metaloprotease e alergeneno de *T. serrulatus* no transcriptoma de *S. jendeki*, o que é um indício de que essas duas proteínas podem ser expressas na glândula de veneno de não-butídeos. De fato, atividade de hialuronidases e metaloproteases foi identificada no veneno de *Hemiscorpius lepturus*, um escorpião pertencente à família Hemiscorpiidae (Seyedian *et al.*, 2010). Transcriptomas de baixa cobertura poderiam explicar a ausência da descrição das metaloproteases, alergenos e hialuronidases em não-butídeos. Outra possibilidade seria que os autores não consideraram essas moléculas como componentes do veneno.

NDBP3 e NDBP4 são famílias de peptídeos sem ponte dissulfeto que inclui moléculas com atividade antimicrobiana ou potencializadora de bradicinina (Zeng *et al.*, 2005). O peptídeo Pape e as AMPs de *T. serrulatus* se enquadram, respectivamente, nessas duas categorias. A comparação entre transcriptomas indica que os membros dessas famílias de peptídeos são ubíquos em escorpiões e sua expressão é relativamente alta nas glândulas de veneno. Entretanto, utilizando como sequências de entrada o peptídeo Pape e os AMPs, as análises com o PSI-BLAST, identificou ortólogos desses componentes apenas na família

Buthidae. De modo similar ao resultado obtido no presente trabalho através das análises de bioinformática, Ma *et al.* (2012) observaram a ocorrência de BPP (pertencente aos NDBP3), peptídeo similar a pandindina-2 e BmKb1 (pertencentes aos NDBP4) apenas em Buthídeos. Algumas hipóteses poderiam explicar a incongruência na distribuição desses peptídeos sem pontes dissulfeto: 1) a alta diversificação das sequências prejudicaria a identificação de ortólogos; 2) diferentes membros dessas famílias de peptídeos seriam específicos para determinadas famílias de escorpiões.

Toxinas de canal para cálcio (CaTx), toxinas de canal para cloro (ClTx), serina proteases, peptídeos sem ponte dissulfeto da família NDBP5, peptídeo La1 (e similares), escorpinas e inibidor de serina-proteases não foram identificados no transcriptoma de *T. serrulatus* e, portanto, a busca por ortólogos usando PSI-BLAST não foi realizada para essas famílias de componentes do veneno. A análise comparativa dos transcriptomas indica que CaTx é distribuído em diferentes famílias de escorpiões, mas sua expressão na glândula de veneno é relativamente baixa. Como exemplos, esses transcritos representaram 0,4% em *L. mucronatus*; 1,1% em *I. maculatus* e 0,1% em *S. magerisonae*. Diferente de CaTx, as ClTx foram identificadas apenas em transcriptomas de membros da família Buthidae, o que concorda com dados da literatura (Lippens *et al.*, 1995; Fu *et al.*, 2007; Thompson *et al.*, 2009; Rjeibi *et al.*, 2011; Tekook *et al.*, 2012). A expressão desses componentes também foi baixa, correspondendo 0,1 e 0,9% dos transcritos de *B. occitanus* e *H. judaicus*, respectivamente. Como mostrado na comparação dos transcriptomas e reforçado por Ma *et al.* (2012), até o momento, escorpinas foram identificadas apenas em não-butídeos. Peptídeos da família NDBP5, inibidor de serina proteases e peptídeos La1 já foram encontrados em escorpiões butídeos e não buthídeos (Zeng *et al.*, 2005; Schwartz *et al.*, 2007; Ma *et al.*, 2010; 2012; Rendón-Anaya *et al.*, 2012). As análises dos transcriptomas da glândula de veneno de escorpiões indicam que esses componentes apresentam maior expressão em não-butídeos. Serina-proteases foram encontradas em transcriptomas de escorpiões de diferentes famílias. Aparentemente, a glândulas de veneno em repouso possui expressão mais acentuada dessas proteases e a glândula de veneno no estado ativo apresenta expressão consideravelmente baixa. Apesar de existirem evidências da ocorrência de serina proteases no veneno de *T. serrulatus* e *T. bahiensis* (Almeida *et al.*, 2002), a expressão dessas proteínas não foi identificada até o momento no transcriptoma de espécies do gênero *Tityus*. A baixa expressão de CaTx, ClTx, peptídeos da família NDBP5, inibidor de serina proteases, peptídeos La1, e serina proteases gera a dúvida se a falta de

identificação desses componentes em algumas espécies, incluindo *T. serrulatus*, seria decorrente ou baixa cobertura dos transcriptomas ou da não expressão desses constituintes no momento da extração de RNAm para a construção da biblioteca de cDNA.

Como mostrado na Tabela 12, muitos trabalhos de transcriptoma de glândula de veneno relatam a ocorrência de peptídeos ricos em cisteínas e proteínas hipotéticas. Entretanto, a análise da distribuição dessas classes de proteínas e peptídeos em escorpiões é difícil de ser realizada, uma vez que dentro de cada uma dessas categorias há uma grande diversidade de componentes. Além disso, para essas moléculas, há poucos dados na literatura que evidenciam sua ocorrência no veneno.

Os dados apresentados até aqui indicam que o veneno de cada espécie de escorpião representa uma combinação de componentes cuja composição tem sido elaborada durante eventos de recrutamento que ocorreram nos níveis de ordem, família, gênero e espécie. Assim, alguns componentes do veneno são amplamente distribuídos nas diferentes famílias de escorpiões como as KTxs e peptídeo aniônico, enquanto outros foram encontrados apenas em algumas espécies, como as magerinas, jendinas e peptídeo rico em prolina. Múltiplos mecanismos podem estar envolvidos na geração da diversidade dos componentes do veneno, como os processos de duplicação gênica, seleção positiva, edição de RNAsm, *trans-splicing* e *splicing* alternativo (Zhijian *et al.*, 2006; D'Suzi *et al.*, 2009). Isso seria benéfico para a sobrevivência do escorpião, uma vez que um arsenal de veneno cada vez mais complexo poderia ir ao encontro com a demanda de interação com a presa, predador e competidor (Ma *et al.*, 2009).

6.5 Análise comparativa de transcriptomas de quelicerados

No presente estudo, o transcriptoma de *T. serrulatus* foi analisado no contexto dos catálogos dos transcritos de quelicerados. Cerca de 5–22% dos unives dos transcriptomas amostrados foram similares a transcritos de *T. serrulatus*. Os principais transcritos responsáveis pela similaridade foram genes de manutenção celular (*housekeeping genes*), principalmente envolvidos com a produção de energia, com a síntese protéica e com a manutenção da estrutura e o movimento celular (verificar a Tabela 11). O único transcriptoma analisado que dividiu considerável parcela de constituintes do veneno com *T. serrulatus* foi o catálogo de transcrito de uma glândula de veneno pertencente a outra espécie de escorpião (*L. mucronatus*) que também está incluída na família Buthidae. A

porcentagem de sequências compartilhadas entre transcriptomas de glândula de veneno de escorpiões e demais catálogos de transcritos foi relativamente pequena, indicando que a transcrição da glândula de veneno é altamente especializada.

Quando sequências de transcriptomas são anotadas com termos do GO e a distribuição desses termos é descrita, geralmente se observa um padrão relativamente semelhante de distribuição de termos, mesmo quando se compara catálogos de transcritos provenientes de animais ou órgão bem diferentes. Essa observação ficou evidenciada no presente trabalho, uma vez que transcriptomas de animais inteiros e de outras glândulas (como glândula de seda, salivar e hepatopâncreas) apresentaram padrão similar de distribuição de termos do GO, mesmo quando se comparou catálogos de transcritos provenientes de amostras biológicas tão diversas como glândula de seda de aranha e um animal inteiro da espécie *L. polyphemus*. Contrapondo essa semelhança, transcriptomas de glândulas de veneno apresentam um típico padrão de distribuição de termos GO. Alguns termos como ‘*channel regulator activity*’, ‘*multi-organism process*’, ‘*response to stimulus*’ e ‘*extracellular region*’ foram exclusivos e/ou ocuparam uma proporção bastante acentuada na anotação de transcriptomas de glândulas de veneno. Uma vez que para a comparação a anotação com os termos do GO foi realizada com termos mais genéricos (nível 2), pode-se inferir que as diferenças entre o padrão de distribuição dos termos do GO só puderam ser observadas quando se comparou transcriptomas altamente especializados, como aqueles de glândulas de veneno.

Quando os peptídeos e proteínas de *T. serrulatus* considerados componentes do veneno foram comparados com outros transcriptomas pôde-se verificar que o número de sequências similares foi muito restrito em transcriptomas de animais inteiros ou outros tipos de glândulas. Assim, apenas componentes similares a antarease, a alergeno, a proteína secretada de *O. cayaporum* e a peptídeo rico em cisteína de *H. judaicus* foram comuns entre o transcriptoma de *T. serrulatus* e os provenientes de animais inteiros ou demais glândulas de espécies não escorpiônicas. Como esperado, transcriptomas de glândulas de veneno compartilharam um maior número de transcritos. Quando se compara Buthidae *versus* não-Buthidae, os responsáveis pela semelhança entre os transcriptomas foram uníques consideradas “outros componentes”, e não as principais neurotoxinas. Considerável diferença tem sido descrita entre a composição do veneno de butídeos e não-butídeos. Ortólogos dos principais componentes do veneno de *T. serrulatus* não foram encontrados em escorpiões não-butídeos ou quando presentes foram expressos em baixos

níveis. Corroborando esses resultados, Ma *et al.* (2011) demonstraram diferenças quanto a ocorrência e nível de expressão de transcritos entre transcriptomas de escorpiões butídeos e euscorpídeos, sendo que alguns componentes do veneno como o peptídeo Pandinina-1-like e magerinas apenas foram detectados no transcriptoma de euscorpídeo. Os dois transcriptomas provenientes de glândula de veneno de espécies de escorpiões da família Buthidae apresentaram considerável semelhança quanto a ocorrência de transcritos e padrão de expressão (verificar nas Tabelas 9 e 10). Essa semelhança pode ser ainda maior uma vez que os componentes do veneno de *T. serrulatus* não encontrados em *L. mucronatus* consistiram principalmente daqueles que apresentaram baixa expressão em *T. serrulatus* e podem não ter ocorrido na outra espécie de butídeos por questão de amostragem, dado o menor tamanho desse catálogo de transcrito.

As análises comparativas realizadas aqui mostram que as análises de transcriptomas potencialmente revelam a existência de tipos de toxinas comuns e exclusivas em diferentes espécies de escorpião. Assim, a obtenção de informação molecular proveniente de diferentes espécies de escorpiões contribui consideravelmente para se contruir um cenário mais claro do arsenal de veneno e das trajetórias evolutivas dos seus componentes, além de propiciar uma visão mais holística das glândulas de veneno desses quelicerados.

7.0 CONCLUSÕES

O principal foco dos estudos de transcriptomas é a descrição da grande diversidade de proteínas e peptídeos sintetizados na glândula de veneno. Nesse sentido, o presente trabalho realizou tal função, contribuindo com 1629 sequências (ESTs) que poderão ser exploradas em estudos futuros. Outro papel do transcriptoma de glândulas de veneno é revelar transcritos secretados pouco expressos ou diferentes das neurotoxinas mais comuns, visto que o sequenciamento é realizado de forma não guiada, ou seja, sem a utilização de iniciadores (*primers*) específicos. Novamente, o presente trabalho cumpriu essa função e, mesmo para uma espécie de escorpião intensivamente estudada como o *T. serrulatus*, essa ferramenta se mostrou útil na identificação de componentes ainda não descritos no veneno da espécie (como os AMPs, Ts17, Ts18).

Um diferencial do presente trabalho foi a análise comparativa entre os transcriptomas de quelicerados. Essa análise revelou que as glândulas de veneno de escorpiões apresentam transcrição altamente especializada para a síntese dos componentes do veneno. Aparentemente, espécies de escorpiões pertencentes à mesma família apresenta ocorrência de transcritos e nível de expressão dos componentes do veneno bastante semelhante. Apesar de glândulas de veneno de escorpiões de famílias diferentes apresentarem alguns componentes do veneno comuns, elas geralmente não compartilharam os constituintes considerados principais ou o nível de expressão dos mesmos é muito diferente entre as glândulas. Quase a totalidade da similaridade entre o transcriptoma de *T. serrulatus* e catálogos de transcritos provenientes de animais inteiros ou outros tipos de glândula esteve relacionada a componentes envolvidos com a homeostase celular e tais componentes apresentaram nível de expressão relativamente semelhante em todos os conjuntos de transcritos dos quelicerados. Isso não seria muito surpreendente, desde que a glândula de veneno é composta por diferentes tecidos (muscular, epitelial, conjuntivo e nervoso), o que resultaria em uma composição de transcritos similar a de outras partes do corpo.

A transcriptômica comparativa também indicou que a composição do veneno de *T. serrulatus* envolve genes que tiveram origem em diferentes clados. As KTxS possivelmente surgiram na ordem Scorpiones e apresentam uma distribuição ampla em diversas espécies de escorpiões. As NaTxS, principais responsáveis pelos efeitos tóxicos do veneno, provavelmente surgiram na família Buthidae, na qual também se originou outros

importantes constituintes do veneno de *T. serrulatus* como os AMPs e BPPs. Alguns componentes do veneno, especialmente peptídeos ricos em cisteína aparentemente são mais recentes e seus ortólogos não foram identificados em outro gênero além de *Tityus*. Proteínas presentes no veneno, em especial enzimas como antareases, hialuronidasas, alergenos e fosfolipase A2 constituem componentes conservados, amplamente distribuídos nos grupos animais e que foram recrutadas para serem expressas na glândula de veneno de escorpiões.

Vale ressaltar que as análises de bioinformática são essenciais na realização das comparações entre os transcriptomas e auxiliam muito na obtenção de uma visão mais holística dos componentes do veneno de escorpiões. Entretanto, é muito importante a validação dos resultados obtidos com esse tipo de análise através de ensaios biológicos.

Como perspectiva, a análise de transcriptoma realizada nesse trabalho abre diversas possibilidades de análises. Por descrever uma ampla variedade de transcritos do veneno, o transcriptoma constitui um passo importante no conhecimento da composição das secreções dos escorpiões, além de abrir diversas possibilidades de estudos e aplicações desse conhecimento. Como exemplo, novos transcritos poderão ser clonados, expressos e terem sua ação farmacológica estudada, o que poderia resultar na descoberta de novos fármacos. Especialmente, os peptídeos antimicrobianos descritos aqui, poderão ser testados através de ensaios microbiológicos para avaliar sua capacidade de inibição bacteriana. Outra possibilidade seria de se testar a capacidade de componentes do veneno, especialmente aqueles não tóxicos, em induzir produção de anticorpos, o que poderia proporcionar o desenvolvimento de vacinas e soros mais eficientes contra escorpiões. Outra interessante investigação seria o estudo de mecanismos de processamento proteico e modificações pós-traducionais dos peptídeos/proteínas secretados pela glândula de veneno, uma vez que o proteoma e o transcriptoma estão disponíveis. Além disso, considerando a grande variedade de peptídeos/proteínas presentes na glândula de veneno, esses transcritos podem ser alvos de estudos evolutivos que visam compreender os mecanismos que levaram à diversificação dos componentes do veneno.

REFERÊNCIAS

- Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, *et al.* (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252:1651–6.
- Alami M, Ouafik L, Céard B, Legros C, Bougis PE, Martin-Eauclaire MF (2001). Characterisation of the gene encoding the alpha-toxin Amm V from the scorpion *Androctonus mauretanicus mauretanicus*. *Toxicon*, 39(10):1579-85.
- Almeida DD, Scortecci KC, Kobashi LS, Agnez-Lima LF, Medeiros SR, Silva-Junior AA, Junqueira-de-Azevedo I de L, Fernandes-Pedrosa M de F. (2012) Profiling the resting venom gland of the scorpion *Tityus stigmurus* through a transcriptomic survey. *BMC Genomics*, 13, 362.
- Almeida FM, Pimenta AM, De Figueiredo SG, Santoro MM, Martin-Eauclaire MF, Diniz CR, De Lima ME. (2002). Enzymes with gelatinolytic activity can be found in *Tityus bahiensis* and *Tityus serrulatus* venoms. *Toxicon*, 40(7):1041-5.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25: 3389-3402.
- Arantes EC, Prado WA, Sampaio SV, Giglio JR (1989). A simplified procedure for the fractionation of *Tityus serrulatus* venom: isolation and partial characterization of TsTX-IV, a new neurotoxin. *Toxicon*, 27(8):907-16.
- Arantes EC, Riccioppo Neto F, Sampaio SV, Vieira CA, Giglio JR. (1994). Isolation and characterization of TsTX-V, a new neurotoxin from *Tityus serrulatus* scorpion venom which delays the inactivation of Na⁺ channels. *Biochim Biophys Acta.*, 1199(1):69-75.
- Amitai Y (1998). Clinical manifestations and management of scorpion envenomation. *Public Health Rev.*, 26(3): 257-63.
- Barbosa-Silva A, Satagopam V, Schneider R, Ortega JM (2008). Clustering of cognate proteins among distinct proteomes derived from multiple links to a single seed sequence. *BMC Bioinformatics*, 9:141.
- Barona J, Batista CVF, Zamudio FZ, Gomez-Lagunas F, Wanke E, Otero R, Possani LD (2006). Proteomic analysis of the venom and characterization of toxins specific for Na⁺- and K⁺-channels from the Colombian scorpion *Tityus pachyurus*. *Biochim Biophys Acta.*, 1764(1):76-84.
- Batista CV, D'Suze G, Gómez-Lagunas F, Zamudio FZ, Encarnación S, Sevcik C, Possani LD. (2006). Proteomic analysis of *Tityus discrepans* scorpion venom and amino acid sequence of novel toxins. *Proteomics*, 6(12):3718-27.

- Batista CV, Román-González SA, Salas-Castillo SP, Zamudio FZ, Gómez-Lagunas F, Possani LD (2007). Proteomic analysis of the venom from the scorpion *Tityus stigmurus*: biochemical and physiological comparison with other *Tityus* species. *Comp Biochem Physiol C Toxicol Pharmacol.*, 146(1-2):147-57.
- Bechis G, Sampieri F, Yuan PM., Brando T, Martin MF, Diniz CR, Rochat H. (1984). Amino acid sequence of toxin VII, a beta-toxin from the venom of the scorpion *Tityus serrulatus*. *Biochem. Biophys. Res. Commun.* 122:1146-53.
- Bernstein SL, Guo Y, Peterson K, Wistow G. (2009). Expressed sequence tag analysis of adult human optic nerve for NEIBank: identification of cell type and tissue markers. *BMC Neurosci.* 10:121.
- Bissinger BW, Donohue KV, Khalil SM, Grozinger CM, Sonenshine DE, Zhu J, Roe RM (2011). Synganglion transcriptome and developmental global gene expression in adult females of the American dog tick, *Dermacentor variabilis* (Acari: Ixodidae). *Insect Mol Biol.* 20(4):465-91.
- Blanc E, Lecomte C, Martin-Eauclaire M-F, van Rietschoten J, Sabatier J-M, Darbon H. (1997). Solution structure of TsKapa, a charybdotoxin-like scorpion toxin from *Tityus serrulatus* with high affinity for apamin-sensitive Ca(2+)-activated K⁺ channels. *Proteins* 29:359-369.
- Boguski MS, Lowe TM, Tolstoshev CM.(1993). dbEST--database for "expressed sequence tags". *Nat Genet.*, 4(4):332-3.
- Bortoluzzi LR, Querol MVM, Querol E (2007). Notas sobre a ocorrência de *Tityus serrulatus* Lutz & Mello, 1922 (Scorpiones, Buthidae) no oeste do Rio Grande do Sul. *Brasil Biota Neotro.*, 7, 357–359.
- Brasil (2009). Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. Manual de controle de escorpiões / Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Vigilância Epidemiológica. – Brasília: Ministério da Saúde, 72p.
- Bulet P, Stocklin R, Menin L (2004). Anti-microbial peptides: from invertebrates to vertebrates. *Immunol Rev.*, 198:169-184.
- Calvete JJ, Marcinkiewicz C, Sanz L (2007). Snake venomomics of *Bitis gabonica gabonica* Protein family composition, subunit organization of venom toxins, and characterization of dimeric disintegrins bitisgabonin-1 and bitisgabonin-2. *J.Proteome res.*, 6 (1):326-336.
- Calvete JJ, Sanz L, Angulo Y, Lomonte B, Gutiérrez JM. (2009). Venoms, venomomics, antivenomics. *FEBS Lett.*, 583(11):1736-43.
- Camargos TS, Restano-Cassulini R, Possani LD, Peigneur S, Tytgat J, Schwartz CA, Alves EM, de Freitas SM, Schwartz EF. (2011). The new kappa-KTx 2.5 from the scorpion *Opisthacanthus cayaporum*. *Peptides*, 32(7):1509-17.
- .

Campos FV, Coronas FIV, Beirao PSL (2004). Voltage-dependent displacement of the scorpion toxin Ts3 from sodium channels and its implication on the control of inactivation. *Br. J. Pharmacol.*, 142:1115-1122.

Carballar-Lejarazú R, Rodríguez M, de la Cruz Hernández-Hernández F, Ramos-Castañeda J, Possani L, Zurita-Ortega M, Reynaud-Garza E, Hernández-Rivas R, Loukeris T, Lycett G, Lanz-Mendoza H (2008). Recombinant scorpine: a multifunctional antimicrobial peptide with activity against different pathogens. *Cell Mol Life Sci.*, 65:3081-3092.

Cardoso KC, Da Silva MJ, Costa GG, Torres TT, Del Bem LE, Vidal RO, Menossi M, Hyslop S. (2010). A transcriptomic analysis of gene expression in the venom gland of the snake *Bothrops alternatus* (urutu). *BMC Genomics*, 26;11:605

Chávez-Olórtegui C, Kalapothakis E, Ferreira AM, Ferreira AP, Diniz CR (1997). Neutralizing capacity of antibodies elicited by a non-toxic protein purified from the venom of the scorpion *Tityus serrulatus*. *Toxicon*, 35(2): 213-21.

Chippaux, J.-P., Goyffonb, M. Epidemiology of scorpionism: A global appraisal. (2008) *Acta Trop.* 107, 71–79.

Cloudsley-Thompson (1990). Scorpions in Mythology, Folklore, and History. In: Polis, GA (Ed). *The Biology of Scorpions*. Stanford University Press, 587p.

Cologna CT, Marcussi S, Giglio JR, Soares AM, Arantes EC (2009). *Tityus serrulatus* scorpion venom and toxins: an overview. *Protein Pept Lett.*, 16(8):920-32.

Cologna CT, Peigneur S, Rosa JC, Selistre-de-Araujo HS, Varanda WA, Tytgat J, Arantes EC. (2011). Purification and characterization of Ts15, the first member of a new α -KTX subfamily from the venom of the Brazilian scorpion *Tityus serrulatus*. *Toxicon*. 58(1):54-61.

Cologna CT, Peigneur S, Rustiguel JK, Nonato MC, Tytgat J, Arantes EC. (2012). Investigation of the relationship between the structure and function of Ts2, a neurotoxin from *Tityus serrulatus* venom. *FEBS J.* , 279(8):1495-504.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, (18):3674-6.

Consortium TU. (2010). The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.*, 38:D142-D148.

Corona M, Zurita M, Possani LD, Becerril B. (1996). Cloning and characterization of the genomic region encoding toxin IV-5 from the scorpion *Tityus serrulatus* Lutz and Mello. *Toxicon*, 34(2):251-6.

- Cupo P, Azevedo-Marques MM, Hering SE (2003). Acidentes por animais peçonhentos: *Escorpiões e aranhas. Medicina, Ribeirão Preto*, 36: 490-497.
- Dai C, Ma Y, Zhao Z, Zhao R, Wang Q, Wu Y, Cao Z, Li W (2008): Mucroporin, the first cationic host defense peptide from the venom of *Lychas mucronatus*. *Antimicrob Agents Chemother*, 52: 3967-3972.
- De Masi, D. (2005). Criatividade e grupos criativos. Tradução Léa Manzi e Yadyr Figueiredo – Rio de Janeiro: Sextante, v. 2.
- Diego-Garcia E, Batista CV, Garcia-Gomez BI, Lucas S, Candido DM, Gomez-Lagunas F, Possani LD (2005). The Brazilian scorpion *Tityus costatus* Karsch: genes, peptides and function. *Toxicon*, 45: 273–283.
- Diego-García E, Schwartz EF, D'Suze G, González SA, Batista CV, García BI, de la Vega RC, Possani LD (2007). Wide phylogenetic distribution of Scorpine and long-chain beta-KTx-like peptides in scorpion venoms: identification of "orphan" components. *Peptides*, 28(1):31-7.
- Diego-García E, Peigneur S, Clynen E, Marien T, Czech L, Schoofs L, Tytgat J (2012). Molecular diversity of the telson and venom components from *Pandinus cavimanus* (Scorpionidae Latreille 1802): Transcriptome, venomomics and function. *Proteomics*, 12(2):313-28.
- D'Suze G, Schwartz EF, García-Gómez BI, Sevcik C, Possani LD (2009). Molecular cloning and nucleotide sequence analysis of genes from a cDNA library of the scorpion *Tityus discrepans*. *Biochimie.*, (8): 1010-9.
- Eipper BA, Milgram SL, Husten EJ, Yun HY, Mains RE (1993). Peptidylglycine alpha-amidating monooxygenase: a multifunctional protein with catalytic, processing, and routing domains. *Protein Sci*, 2:489-497.
- Ellis KC, Tenenholz TC, Jerng H, Hayhurst M, Dudlak CS, Gilly WF, Blaustein MP, Weber DJ. (2001). Interaction of a toxin from the scorpion *Tityus serrulatus* with a cloned K⁺ channel from squid (sqKv1A). *Biochemistry*, 40:5942-5953.
- Ewing B, Hillier L, Wendl M, Green P (1998). Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.*, 8: 175-185.
- Fernandes GR, Barbosa DVC, Prosdocimi F, Pena IA, Santana-Santos L, Coelho Junior O, Barbosa-Silva A, Velloso HM, Mudado MA, Natale DA, *et al* (2008). A procedure to recruit members to enlarge protein family databases - the building of UECOG (UniRef-Enriched COG Database) as a model. *Genet Mol Res* , 7 (3):910-924.
- Fernandes-Pedrosa MF, Junqueira-de-Azevedo IL, Gonçalves-de-Andrade RM, Kobashi LS, Almeida DD, Ho PL, Tambourgi DV (2008). Transcriptome analysis of *Loxosceles laeta* (Araneae, Sicariidae) spider venomous gland using expressed sequence tags. *BMC Genomics*, 9:279.

Ferreira LAF, Alves EW, Henriques OB. (1993). Peptide T, a novel bradykinin potentiator isolated from *Tityus serrulatus* scorpion venom. *Toxicon* 31:941-947.

Fletcher PLJr, Fletcher MD, Weninger K, Anderson TE, Martin BM (2010). Vesicle-associated membrane protein (VAMP) cleavage by a new metalloprotease from the Brazilian scorpion *Tityus serrulatus*. *J Biol Chem.*, 285 (10):7405-16.

Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JD, King GF, Nevalainen TJ, Norman JA, Lewis RJ, Norton RS, Renjifo C, Rodríguez de la Vega C (2009). The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu Rev Genomics Hum Genet.*, 10:483-511.

Fu YJ, Yin LT, Liang AH, Zhang CF, Wang W, Chai BF, Yang JY, Fan XJ (2007). Therapeutic potential of chlorotoxin-like neurotoxin from the Chinese scorpion for human gliomas. *Neurosci Lett.*, 412(1):62-7.

Gordon D, Karbata I, Ilana N, Cohena L, Kahna R, Gillesb N, Dongc K, Stühmerd W, Tytgat J, Gurevitza M (2007). The differential preference of scorpion α -toxins for insect or mammalian sodium channels: Implications for improved insect control. *Toxicon*, 49:452–472.

Götz S, García-Gómez JM, Terol J, Williams TD, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.*, 36(10): 3420–3435.

Gremski LH, da Silveira RB, Chaim OM, Probst CM, Ferrer VP, Nowatzki J, Weinschutz HC, Madeira HM, Gremski W, Nader HB, Senff-Ribeiro A, Veiga SS. (2010). A novel expression profile of the *Loxosceles intermedia* spider venomous gland revealed by transcriptome analysis. *Mol Biosyst.*, 6(12):2403-16.

Guatimosim SC, Prado VF, Diniz CR, Chávez-Olórtegui C, Kalapothakis E (1999). Molecular cloning and genomic analysis of TsNTxp: an immunogenic protein from *Tityus serrulatus* scorpion venom. *Toxicon*, 37: 507-517.

Guatimosim SC, Kalapothakis E, Diniz CR, Chávez-Olórtegui C (2000). Induction of neutralizing antibodies against *Tityus serrulatus* toxins by immunization with a recombinant nontoxic protein. *Toxicon*, 38 (1):113-21.

Harris F, Dennison SR, Phoenix DA. (2009). Anionic antimicrobial peptides from eukaryotic organisms. *Curr. Protein Pept. Sci.*, 10(6), 585-606.

Herbert JM, Stekel D, Sanderson S, Heath VL, Bicknell R. (2008). A novel method of differential gene expression analysis using multiple cDNA libraries applied to the identification of tumour endothelial genes. *BMC Genomics.*, 7;9:153.

Huang X, Madan A (1999). CAP3: A DNA Sequence Assembly Program. *Genome Res.* 9(9):868-77.

Jarrar BM, Al-Rowaily MA (2008). Histology and Histochemistry of the venom apparatus of the black scorpion *Androctonus crassicauda* (OLIVIER, 1807) (Scorpiones: Buthidae). *J. Venom. Anim. Toxins incl. Trop. Dis.*, 14 (3): 514-526.

Jiang J, Jacob HJ (1998). EbEST: an automated tool using expressed sequence tags to delineate gene structure. *Genome Res* 8:268–75.

Jiang N, Kolhekar AS, Jacobs PS, Mains RE, Eipper BA, Taghert PH (2000). PHM is required for normal developmental transitions and for biosynthesis of secretory peptides in *Drosophila*. *Dev Biol.*, 226:118-136.

Jinjun C, Zhao L, Jiang L, Meng E, Zhang Y, Xiong X, Liang S. (2008). Transcriptome analysis revealed novel possible venom components and cellular processes of the tarantula *Chilobrachys jingzhao* venom gland. *Toxicon*, 52: 794–806

Jungo F, Bougueleret L, Xenarios I, Poux S. (2012). The UniProtKB/Swiss-Prot Tox-Prot program: A central hub of integrated venom protein data. *Toxicon*, 60(4):551-7.

Junqueira-de-Azevedo IL, Ching AT, Carvalho E, Faria F, Nishiyama MY Jr, Ho PL, Diniz MR (2006). *Lachesis muta* (Viperidae) cDNAs reveal diverging pit viper molecules and scaffolds typical of cobra (Elapidae) venoms: implications for snake toxin repertoire evolution. *Genetics*, 173 (2): 877-889.

Kalapothis E, Jardim S, Magalhães AC, Mendes TM, De Marco L, Afonso LC, Chávez-Olórtegui C (2001). Screening of expression libraries using ELISA: identification of immunogenic proteins from *Tityus bahiensis* and *Tityus serrulatus* venom. *Toxicon*, 39(5):679-85.

King GF, Gentz MC, Escoubas P, Nicholson GM (2008). A rational nomenclature for naming peptide toxins from spiders and other venomous animals. *Toxicon*, 52(2):264-76.

Kirsch GE, Skattebol A, Possani LD, Brown AM. (1989). Modification of Na channel gating by an alpha scorpion toxin from *Tityus serrulatus*. *J. Gen. Physiol.* 93:67-83.

Kozminsky-Atias A, Bar-Shalom A, Mishmar D, Zilberberg N (2008). Assembling an arsenal, the scorpion way. *BMC Evol Biol.*, 8:333.

Legros C, Oughideni R, Darbon H, Rochat H, Bougis PE, Martin-Eauclaire M-F (1996). Characterization of a new peptide from *Tityus serrulatus* scorpion venom which is a ligand of the apamin-binding site. *FEBS Lett.*, 390:81-84.

Lourenço, WR (2000). Reproduction in scorpions, with special reference to parthenogenesis. *European Arachnology*, 71-85.

Lourenço WR. (2002). *Scorpions of Brazil*. Ed. de l'If, Paris, 308 p.

Lourenço, WR, Cloudsley-Thompson JL, Cuellar O, Von Eickstedt VRD, Barraviera B, Knox, MB. (1996). The evolution of scorpionism in Brazil in recent years. *J. Venom. Anim. Toxins*, 2 (2):121-134.

- Lourenço WR, Cloudsley-Thompson JL. (1999). Discovery of a sexual population of *Tityus serrulatus*, one of the morphs within the complex *Tityus stigmurus* (scorpiones, buthidae). *J Arachn.* 27:154–158.
- Löytynoja A, Goldman N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A.*, 102(30):10557-62
- Ma Y, Zhao R, He Y, Li S, Liu J, Wu Y, Cao Z, Li W (2009). Transcriptome analysis of the venom gland of the scorpion *Scorpiops jendeki*: implication for the evolution of the scorpion venom arsenal. *BMC Genomics*, 10: 290.
- Ma Y, Zhao Y, Zhao R, Zhang W, He Y, Wu Y, Cao Z, Guo L, Li W (2010). Molecular diversity of toxic components from the scorpion *Heterometrus petersii* venom revealed by proteomic and transcriptome analysis. *Proteomics*, 10(13):2471-85.
- Ma Y, He Y, Zhao R, Wu Y, Li W, Cao Z (2011). Extreme diversity of scorpion venom peptides and proteins revealed by transcriptomic analysis: Implication for proteome evolution of scorpion venom arsenal. *J Proteomics* 75(5):1563-76.
- Madero MF, Gámez C, Madero MA, Fernández-Nieto M, Sastre J, del Pozo V. (2009) Characterization of allergens in four South American snake species. *Int Arch Allergy Immunol.*, 150(3):307-10.
- Mansuelle P, Martin-Eauclaire M-F, Chavez-Olortegui C, de Lima ME, Rochat H, Granier C. (1992). The beta-type toxin Ts II from the scorpion *Tityus serrulatus*: amino acid sequence determination and assessment of biological and antigenic properties. *Nat. Toxins*, 1:119-125.
- Marangoni S, Ghiso J, Sampaio SV, Arantes EC, Giglio JR, Oliveira B, Frangione B. (1990). The complete amino acid sequence of toxin TsTX-VI isolated from the venom of the scorpion *Tityus serrulatus*. *J. Protein Chem.*, 9(5), 595-601.
- Martin-Eauclaire M-F, Ceard B, Ribeiro AM, Diniz CR, Rochat H, Bougis PE. (1994). Biochemical, pharmacological and genomic characterisation of Ts IV, an alpha-toxin from the venom of the South American scorpion *Tityus serrulatus*. *FEBS Lett.*, 342:181-184.
- Matthiensen FA (1971). The breeding of *T. serrulatus* Lutz & Mello, 1927 in captivity (Scorpiones, Butthidae). *Revista Brasileira de Pesquisas Médicas e Biológicas*, 4: 299-300.
- Mendes TM (2007). Antígenos para a produção de soro contra o veneno do escorpião *Tityus serrulatus*. Tese (Doutorado em Fisiologia e Farmacologia) – Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, 242p.
- Mendes TM, Dias F, Horta CC, Pena IF, Arantes EC, Kalapothakis E (2008). Effective *Tityus serrulatus* anti-venom produced using the Ts1 component. *Toxicon*, 52 (7): 787-93.
- Meki AR, Nassar AY, Rochat H (1995). A bradykinin-potentiating peptide (peptide K12) isolated from the venom of Egyptian scorpion *Buthus occitanus*. *Peptides*, 16 (8):1359-65.

- Min, XJ, Butler, G, Storms, R and Tsang, A (2005). OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res.*, (Web Server issue):W677-80.
- Ministério da saúde, Brasil, 2012.
[http://portal.saude.gov.br/portal/saude/profissional/visualizar_texto.cfm?idtxt=31510]
- Miranda DM, Romano-Silva MA, Kalapothakis E, Diniz CR, Cordeiro MN, Santos TM, Prado MA, Gomez MV (1998). *Phoneutria nigriventer* toxins block tityustoxin-induced calcium influx in synaptosomes. *Neuroreport.*, 9(7):1371-3.
- Modrek B, Resch A, Grasso C, Lee C (2001). Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res.*, 29:2850–9.
- Mondego JM, Vidal RO, Carazzolle MF, Tokuda EK, Parizzi LP, Costa GG, Pereira LF, Andrade AC, Colombo CA, Vieira LG, Pereira GA; Brazilian Coffee Genome Project Consortium. (2011). An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. *BMC Plant Biol.* 11:30.
- More D, Nugent J, Hagan L, Demain J, Schwertner H, Whisman B, Freeman T (2004). Identification of allergens in the venom of the common striped scorpion. *Ann Allergy Asthma Immunol.*, 93(5):493-8.
- Moreira-Ferreira AM, Kalapothakis E, Diniz CR, Chávez-Olórtegui C (1998). In vivo protection against *Tityus serrulatus* scorpion toxins by immunization of mice with a non-toxic protein. *Toxicon*, 36(2):333-9
- Morgenstern D, Rohde BH, King GF, Tal T, Sher D, Zlotkin E (2011). The tale of a resting gland: Transcriptome of a replete venom gland from the scorpion *Hottentotta judaicus*. *Toxicon*, 57:695-703.
- Nagaraj SH, Gasser RB, Ranganathan S (2007). A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief Bioinform*, 8(1):6-21.
- Nascimento DG, Rates B, Santos DM, Verano-Braga T, Barbosa-Silva A, Dutra AA, Biondi I, Martin-Eauclaire MF, De Lima ME, Pimenta AM (2006). Moving pieces in a taxonomic puzzle: venom 2D-LC/MS and data clustering analyses to infer phylogenetic relationships in some scorpions from the Buthidae family (Scorpiones). *Toxicon*, 47:628-639.
- Nie Q, Fang M, Jia X, Zhang W, Zhou X, He X, Zhang X (2011). Analysis of muscle and ovary transcriptome of *Sus scrofa*: assembly, annotation and marker discovery. *Insect Mol Biol.* 18(5):343-51.
- Nisani Z, Dunbar SG, Hayes WK (2007). Cost of venom regeneration in *Parabuthus transvaalicus* (Arachnida: Buthidae). *Comp BiochemPhysiol A.*, 147: 509-513.

Norton RS, McDonough SI (2008). Peptides targeting voltage-gated calcium channels. *Curr Pharm Des.*, 14(24): 2480-91.

Novello JC, Arantes EC, Varanda WA, Oliveira B, Giglio JR, Marangoni S. (1999). TsTX-IV, a short chain four-disulfide-bridged neurotoxin from *Tityus serrulatus* venom which acts on Ca²⁺-activated K⁺ channels. *Toxicon*, 37:651-660.

Lippens G, Najib J, Wodak SJ, Tartar A (1995). NMR sequential assignments and solution structure of chlorotoxin, a small scorpion toxin that blocks chloride channels. *Biochemistry*, 34(1):13-21.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, 27(1), 29-34.

Orivel J, Redeker V, Le Caer JP, Krier F, Revol-Junelles AM, Longeon A, Chaffotte A, Dejean A, Rossier J. (2001). Ponericins, new antibacterial and insecticidal peptides from the venom of the ant *Pachycondyla goeldii*. *J Biol Chem.*, 276(21):17823-9.

ORF Finder [<http://www.ncbi.nlm.nih.gov/projects/gorf/>]

Pertea G, Xiaoqiu H, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J (2003). TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*, 19 (5): 651–652.

Pessini AC, Takao TT, Cavalheiro EC, Vichnewski W, Sampaio SV, Giglio JR, Arantes EC. (2001). A hyaluronidase from *Tityus serrulatus* scorpion venom: isolation, characterization and inhibition by flavonoids. *Toxicon*, 39(10):1495-504.

Petersen TN, Brunak S, von Heijne G, Nielsen H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, 8:785-786.

Petricevich VL, Hernández Cruz A, Coronas FI, Possani LD (2007). Toxin gamma from *Tityus serrulatus* scorpion venom plays an essential role in immunomodulation of macrophages. *Toxicon*, 50(5):666-75.

Pimenta AM, Stöcklin R, Favreau P, Bougis PE, Martin-Eauclaire MF (2001a). Moving pieces in a proteomic puzzle: mass fingerprinting of toxic fractions from the venom of *Tityus serrulatus* (Scorpiones, Buthidae). *Rapid Commun Mass Spectrom.*, 15(17):1562-72.

Pimenta AM, Martin-Eauclaire M, Rochat H, Figueiredo SG, Kalapothakis E, Afonso LC, De Lima ME (2001b). Purification, amino-acid sequence and partial characterization of two toxins with anti-insect activity from the venom of the South American scorpion *Tityus bahiensis* (Buthidae). *Toxicon*, (7): 1009-19.

- Pimenta AM, Mansuelle P, Diniz CR, Martin-Eauclaire MF.(2003a). Covalent structure and some pharmacological features of native and cleaved alpha-KTx12-1, a four disulfide-bridged toxin from *Tityus serrulatus* venom. *J Pept Sci*, 9(2):132-40.
- Pimenta AMC, Legros C, Almeida FM, Mansuelle P, De Lima ME, Bougis PE, Martin-Eauclaire MF (2003b). Novel structural class of four disulfide-bridged peptides from *Tityus serrulatus* venom. *Biochem. Biophys. Res. Commun.* 301:1086-1092.
- Pimenta AM, De Lima ME (2005). Small peptides, big world: biotechnological potential in neglected bioactive peptides from arthropod venoms. *J Pept Sci.*, 11(11): 670-6.
- Polikarpov I, Junior MS, Marangoni S, Toyama MH, Teplyakov A.(1999). Crystal structure of neurotoxin Ts1 from *Tityus serrulatus* provides insights into the specificity and toxicity of scorpion toxins. *J Mol Biol.* 2;290(1):175-84.
- Polis, GA (1990). *The Biology of Scorpions*. Stanford University Press, 587p.
- Pontius JU, Wagner L, Schuler GD (2002). UniGene: a unified view of the transcriptome. In: The NCBI Handbook. Bethesda (MD): National Center for Biotechnology Information (US).
- Possani LD, Alagón AC, Fletcher PL Jr, Erickson BW (1977) Purification and properties of mammalian toxins from the venom of Brazilian Scorpion *Tityus serrulatus* Lutz and Mello. *Arch Biochem Biophys*, 180(2):394-403.
- Possani LD, Martin BM, Mochca-Morales J, Svendsen IB (1981). Purification and chemical characterization of the major toxins from the venom of the Brazilian scorpion *Tityus serrulatus* Lutz and Mello. *Carlsberg Res. Commun.* 46: 195-205.
- Possani LD, Martin BM, Fletcher MD, Fletcher PL Jr. (1991). Discharge effect on pancreatic exocrine secretion produced by toxins purified from *Tityus serrulatus* scorpion venom. *Biol. Chem.* 266:3178-3185.
- Possani LD, Merino E, Corona M, Bolivar F, Becerril B (2000). Peptides and genes coding for scorpion toxins that affect ion-channels. *Biochimie.*, 82(9-10):861-8.
- Prendini L., Wheeler WC. (2005) Scorpion higher phylogeny and classification, taxonomic anarchy, and standards for peer review in online publishing. *Cladistics*, 21, 446–494.
- Prosdocimi CF. Análise dos genes mais expressos e do status atual do transcriptoma de *Schistosoma mansoni* utilizando ferramentas de bioinformática. 2003. 106 p. Dissertação (mestrado) - Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte.
- Prosdocimi F, Bittencourt D, da Silva FR, Kirst M, Motta PC, Rech EL (2011). Spinning gland transcriptomics from two main clades of spiders (order: Araneae) -insights on their molecular, anatomical and behavioral evolution. *PLoS One.*, 6(6):e21634.

Rates B, Ferraz KK, Borges MH, Richardson M, De Lima ME, Pimenta AM (2008). *Tityus serrulatus* venom peptidomics: assessing venom peptide diversity. *Toxicon*, 52(5): 611-8.

Rendón-Anaya M., Delaye L, Possani LD, Herrera-Estrella, A. (2012) Global transcriptome analysis of the scorpion *Centruroides noxius*: New toxin families and evolutionary insights from an ancestral scorpion species. *PLoS One*, 7(8), e43331.

Rjeibi I, Mabrouk K, Mosrati H, Berenguer C, Mejdoub H, Villard C, Laffitte D, Bertin D, Ouafik L, Luis J, Elayeb M, Srairi-Abid N (2011). Purification, synthesis and characterization of AaCtx, the first chlorotoxin-like peptide from *Androctonus australis* scorpion venom. *Peptides*, 32(4):656-63.

Rodrigues ARA, Arantes EC, Monje F; Stühmer W, Varanda WA. (2003). Tityustoxin-K(alpha) blockade of the voltage-gated potassium channel Kv1.3. *Br. J. Pharmacol.*, 139(6), 1180-86.

Rodríguez De La Vega RC, Possani LD (2004). Current views on scorpion toxins specific for K⁺-channels. *Toxicon*, 43 (8): 865-75.

Rodríguez De La Vega RC, Possani LD (2005). Overview of scorpion toxins specific for Na⁺ channels and related peptides: biodiversity, structure-function relationships and evolution. *Toxicon*. 46(8): 831-44.

Rodríguez De La Vega RC, Schwartz EF, Possani LD (2010). Mining on scorpion venom biodiversity. *Toxicon*, 56(7): 1155-61.

Roeding F, Borner J, Kube M, Klages S, Reinhardt R, Burmester T (2009). A 454 sequencing approach for large scale phylogenomic analysis of the common emperor scorpion (*Pandinus imperator*). *Mol Phylogenet Evol.*, 53 (3): 826-34.

Rogowski RS, Krueger BK, Collins JH, Blaustein MP. (1994). Tityustoxin K alpha blocks voltage-gated noninactivating K⁺ channels and unblocks inactivating K⁺ channels blocked by alpha-dendrotoxin in synaptosomes. *Proc Natl Acad Sci U S A.*, 91(4):1475-9.

Romanuik TL, Ueda T, Le N, Haile S, Yong TM, Thomson T, Vessella RL, Sadar MD. (2009). Novel biomarkers for prostate cancer including noncoding transcripts. *Am J Pathol.*, 175(6):2264-76.

Rowlands DS, Thomson JS, Timmons BW, Raymond F, Fuerholz A, Mansourian R, Zawhlen MC, Métairon S, Glover EI, Stellingwerff T *et al.* (2001). The transcriptome and translational signalling following endurance exercise in trained skeletal muscle: Impact of Dietary Protein. *DNA Res.* 43(17):1004-20.

Rudd S. (2003) Expressed sequence tags: alternative or complement to whole genome sequences? *Trends Plant Sci.*, 8:321–9.

Ruiming Z, Yibao M, Yawen H, Zhiyong D, Yingliang W, Zhijian C, Wenxin L (2010). Comparative venom gland transcriptome analysis of the scorpion *Lychas mucronatus*

reveals intraspecific toxic gene diversity and new venomous components. *BMC Genomics*, 11: 452.

Sampaio SV, Arantes EC, Prado WA, Riccioppo Neto F, Giglio JR. (1991). Further characterization of toxins T1IV (TsTX-III) and T2IV from *Tityus serrulatus* scorpion venom. *Toxicon* 29:663-672.

Sampaio SV, Coutinho-Netto J, Arantes EC, Marangoni S, Oliveira B, Giglio JR. (1996). Isolation of toxin TsTX-VI from *Tityus serrulatus* scorpion venom. Effects on the release of neurotransmitters from synaptosomes. *Biochem Mol Biol Int*, 39(4):729-40.

Sanz L, Escolano J, Ferretti M, Biscoglio MJ, Rivera E, Crescenti EJ, Angulo Y, Lomonte B, Gutierrez JM, Calvete JJ (2008). Snake venomomics of the South and Central American Bushmasters. Comparison of the toxin composition of *Lachesis muta* gathered from proteomic versus transcriptomic analysis. *J Proteomics*. 71(1): 46-60.

Schneider MC, Cella DM (2010). Karyotype conservation in 2 populations of the parthenogenetic scorpion *Tityus serrulatus* (Buthidae): rDNA and its associated heterochromatin are concentrated on only one chromosome. *J Hered.* 101(4):491-6.

Scholte RG, Caldeira RL, Simões MC, Stutz WH, Silva LL, Carvalho Odos S, Oliveira G (2009). Inter- and intrapopulational genetic variability of *Tityus serrulatus* (Scorpiones, Buthidae). *Acta Trop.*, 112(2):97-100.

Schwartz EF, Diego-Garcia E, Rodríguez de la Vega RC, Possani LD (2007). Transcriptome analysis of the venom gland of the Mexican scorpion *Hadrurus gertschi* (Arachnida: Scorpiones). *BMC Genomics*, 8:119.

Schwartz EF, Camargos TS, Zamudio FZ, Silva LP, Bloch C Jr, Caixeta F, Schwartz CA, Possani LD (2008). Mass spectrometry analysis, amino acid sequence and biological activity of venom components from the Brazilian scorpion *Opisthacanthus cayaporum*. *Toxicon*, 51(8):1499-508.

SeqClean [<http://compbio.dfci.harvard.edu/tgi/software/>].

Seyedian R, Pipelzadeh MH, Jalali A, Kim E, Lee H, Kang C, Cha M, Sohn ET, Jung ES, Rahmani AH, Mirakabady AZ. (2010). Enzymatic analysis of *Hemiscorpius lepturus* scorpion venom using zymography and venom-specific antivenin. *Toxicon*, 56(4):521-5.

Soares MRM, Azevedo CS, De Maria M (2002). Escorpionismo em Belo Horizonte, MG: um estudo retrospectivo. *Rev. Soc. Bras. Med. Trop.*, 35 (4):359-363.

Silva EC, Camargos TS, Maranhão AQ, Silva-Pereira I, Silva LP, Possani LD, Schwartz EF (2009). Cloning and characterization of cDNA sequences encoding for new venom peptides of the Brazilian scorpion *Opisthacanthus cayaporum*. *Toxicon.*, 54(3):252-61.

Tekook MA, Fabritz L, Kirchhof P, König S, Müller FU, Schmitz W, Tal T, Zlotkin E, Kirchhefer U (2012). Gene construction, expression and functional testing of an inotropic

peptide from the venom of the black scorpion *Hottentotta judaicus*. *Toxicon*. 15;60(8):1415-27.

The Gene Ontology Consortium (2001). Creating the Gene Ontology Resource: Design and Implementation. *Genome Res.* 1425-1433.

Thomas PD, Mi H, Lewis S. (2007) Ontology annotation: mapping genomic regions to biological function. *Curr Opin Chem Biol.* 11(1):4-11.

Thompson CH, Olivetti PR, Fuller MD, Freeman CS, McMaster D, French RJ, Pohl J, Kubanek J, McCarty NA (2009). Isolation and characterization of a high affinity peptide inhibitor of CIC-2 chloride channels. *J Biol Chem.*, 284(38):26051-62.

Torres, JB, Marques M, Martini R, Borges C (2002). Acidente por *Tityus serrulatus* e suas implicações epidemiológicas no Rio Grande do Sul. *Rev. Saúde Pública*, 36: 631–633.

Trevisan-Silva D, Gremski LH, Chaim OM, da Silveira RB, Meissner GO, Mangili OC, Barbaro KC, Gremski W, Veiga SS, Senff-Ribeiro A (2010). Astacin-like metalloproteases are a gene family of toxins present in the venom of different species of the brown spider (genus *Loxosceles*). *Biochimie.* 92(1):21-32.

Useche FJ, Gao G, Harafey M, Rafalski A (2001). Highthroughput identification, database storage and analysis of SNPs in EST sequences. *Genome Inform.*, 12:194–203.

Valdez-Cruz NA, Batista CV, Possani LD (2004). Phaiodactylipin, a glycosylated heterodimeric phospholipase A from the venom of the scorpion *Anuroctonus phaiodactylus*. *Eur J Biochem.*, 271:1453-1464.

Verano-Braga T, Rocha-Resende C, Silva DM, Ianzer D, Martin-Eauclaire MF, Bougis PE, de Lima ME, Santos RA, Pimenta AM (2008). *Tityus serrulatus* Hypotensins: a new family of peptides from scorpion venom. *Biochem Biophys Res Commun.*, 371(3):515-20.

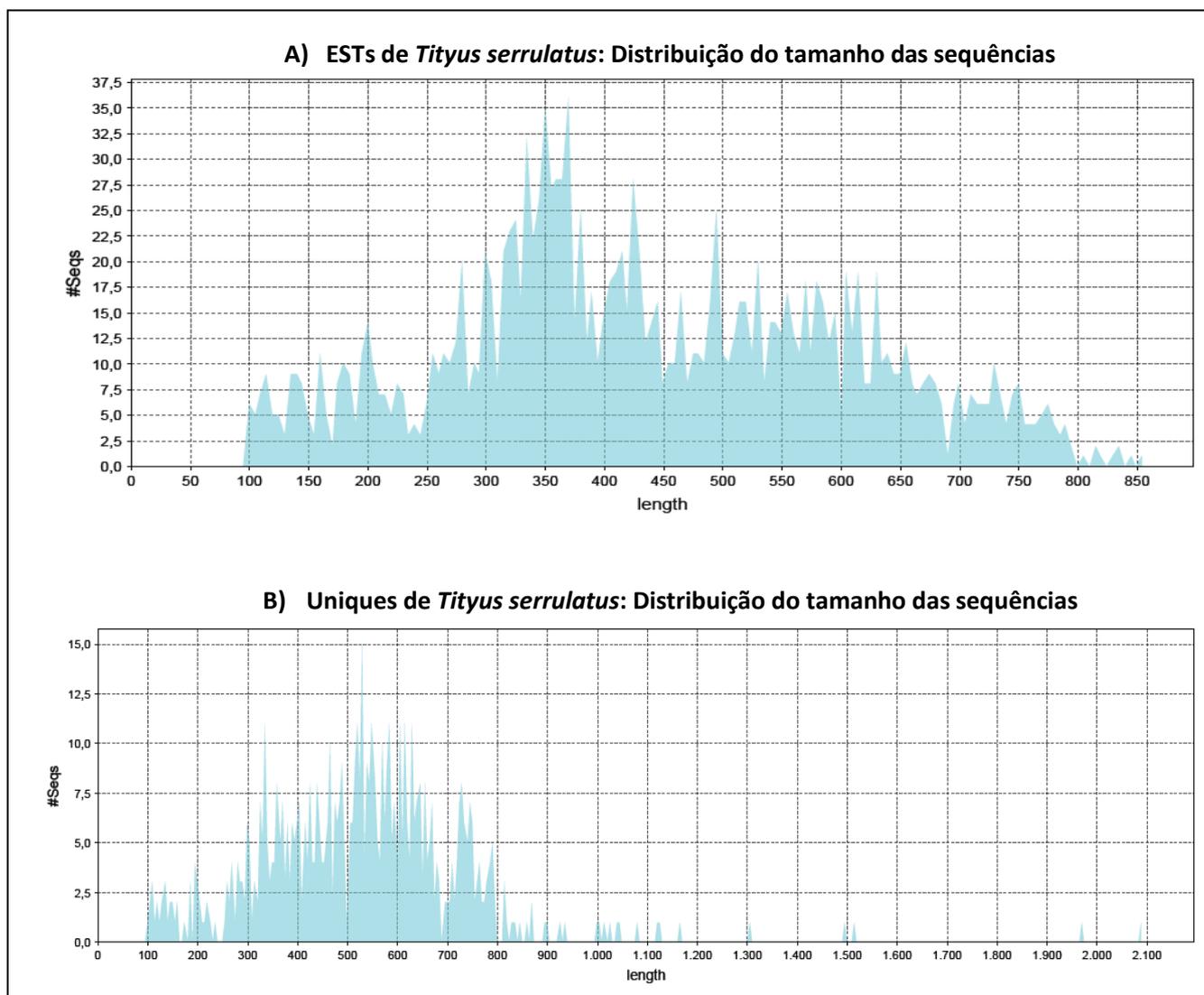
Wang JP, Lindsay BG, Leebens-Mack J, et al. (2004). EST clustering error evaluation and correction. *Bioinformatics*; 20:2973–84.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M., Barton, G.J. (2009). Jalview version 2: A Multiple Sequence Alignment and Analysis Workbench. *Bioinformatics* 25(9):1189-91.

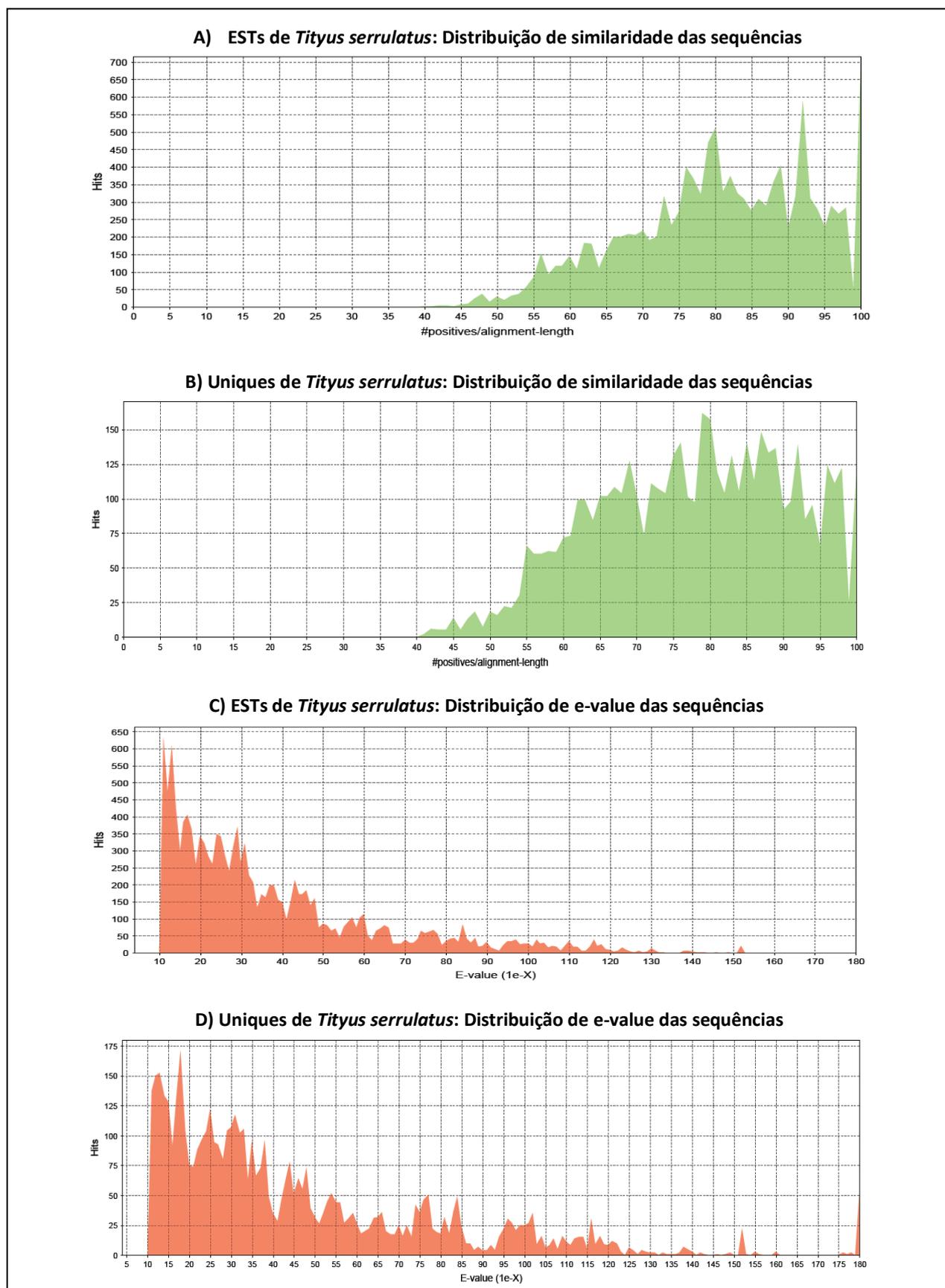
Werkman TR, Gustafson TA, Rogowski RS, Blaustein MP, Rogawski MA.(1993) Tityustoxin-K alpha, a structurally novel and highly potent K⁺ channel peptide toxin, interacts with the a-dendrotoxin binding site on a cloned Kv1.2 K⁺ channel. *Mol. Pharmacol.*, 44(2), 430-436.

Wilcox AS, Khan AS, Hopkins JA, Sikela JM. (1991). Use of 30 untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome. *Nucleic Acids Res.*, 19:1837–43.

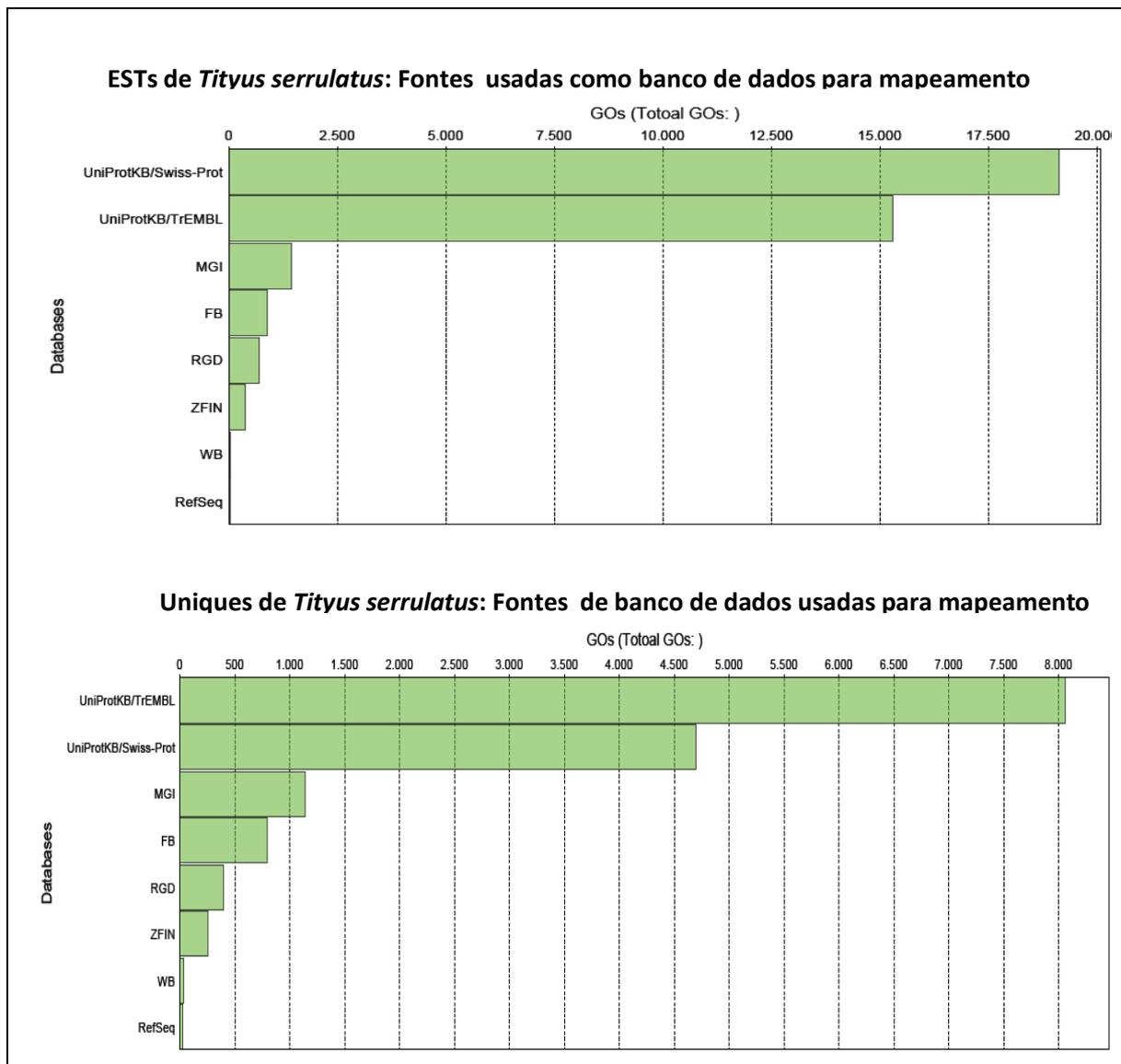
- Yigit N, Benli M (2008). The venom gland of the scorpion species *Euscorpium mingrelicus* (Scorpiones: Euscorpidae): morphological and ultrastructural characterization. *J. Venom. Anim. Toxins incl. Trop. Dis.*, 14 (3): 514-528.
- Zamudio FZ, Conde R, Arévalo C, Becerril B, Martin BM, Valdivia HH, Possani LD (1997). The mechanism of inhibition of ryanodine receptor channels by imperatoxin I, a heterodimeric protein from the scorpion *Pandinus imperator*. *J Biol Chem.*, 272(18): 11886-94.
- Zasloff M (2002). Antimicrobial peptides of multicellular organisms. *Nature*, 415(6870): 389-395.
- Zeng XC, Li WX, Zhu SY, Peng F, Zhu ZH, Wu KL, Yiang FH (2000). Cloning and characterization of a cDNA sequence encoding the precursor of a chlorotoxin-like peptide from the Chinese scorpion *Buthus martensii* Karsch. *Toxicon*, 38(8):1009-14.
- Zeng XC, Wang SX, Li WX (2002). Identification of BmKAPi, a novel type of scorpion venom peptide with peculiar disulfide bridge pattern from *Buthus martensii* Karsch. *Toxicon*, 40(12):1719-22.
- Zeng XC, Wang SX, Zhu Y, Zhu SY, Li WX (2004). Identification and functional characterization of novel scorpion venom peptides with no disulfide bridge from *Buthus martensii* Karsch. *Peptides*, 25(2):143-50.
- Zeng XC, Corzo, G., Hahin, R (2005). Scorpion venom peptides without disulfide bridges. *Life*, 57 (1): 13-21.
- Zhang B, Liu Q, Yin W, Zhang X, Huang Y, Luo Y, Qiu P, Su X, Yu J, Hu S, Yan G (2006). Transcriptome analysis of *Deinagkistrodon acutus* venomous gland focusing on cellular structure and functional aspects using expressed sequence tags. *BMC genomics*, 7:152-152.
- Zhijian C, Feng L, Yingliang W, Xin M, Wenxin L (2006). Genetic mechanisms of scorpion venom peptide diversification. *Toxicon*, 47(3): 348-55.
- Zhu S, Li W, Zeng X, Jiang D, Mao X, Liu H. (1999). Molecular cloning and sequencing of two 'short chain' and two 'long chain' K(+) channel-blocking peptides from the Chinese scorpion *Buthus martensii* Karsch. *FEBS Lett*, 457:509-14.

APÊNDICE A - ESTATÍSTICAS DO Blast2GO

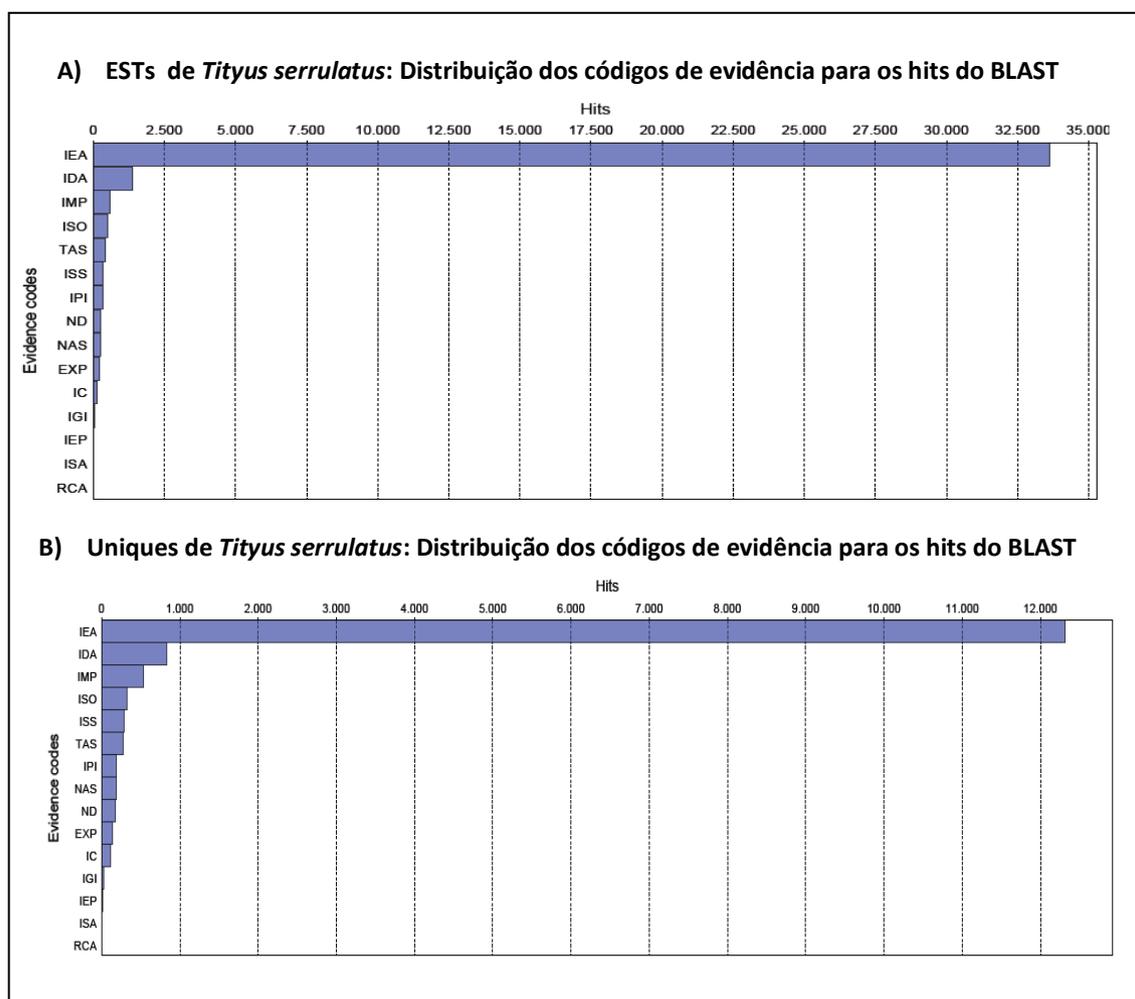
Apêndice A1 -Distribuição do tamanho das sequências A) dados obtidos pela análise de todas as ESTs e B) dados dos uniques. Conforme o observado, o agrupamento das sequências resultou em aumento do tamanho das mesmas. Assim, enquanto as ESTs tiveram tamanho entre 100 e 850 nucleotídeos, o comprimento dos uniques esteve entre 100 e 2.100 nucleotídeos. O maior tamanho da sequência pode favorecer a busca de resultados de similaridade e, portanto, facilita a anotação.



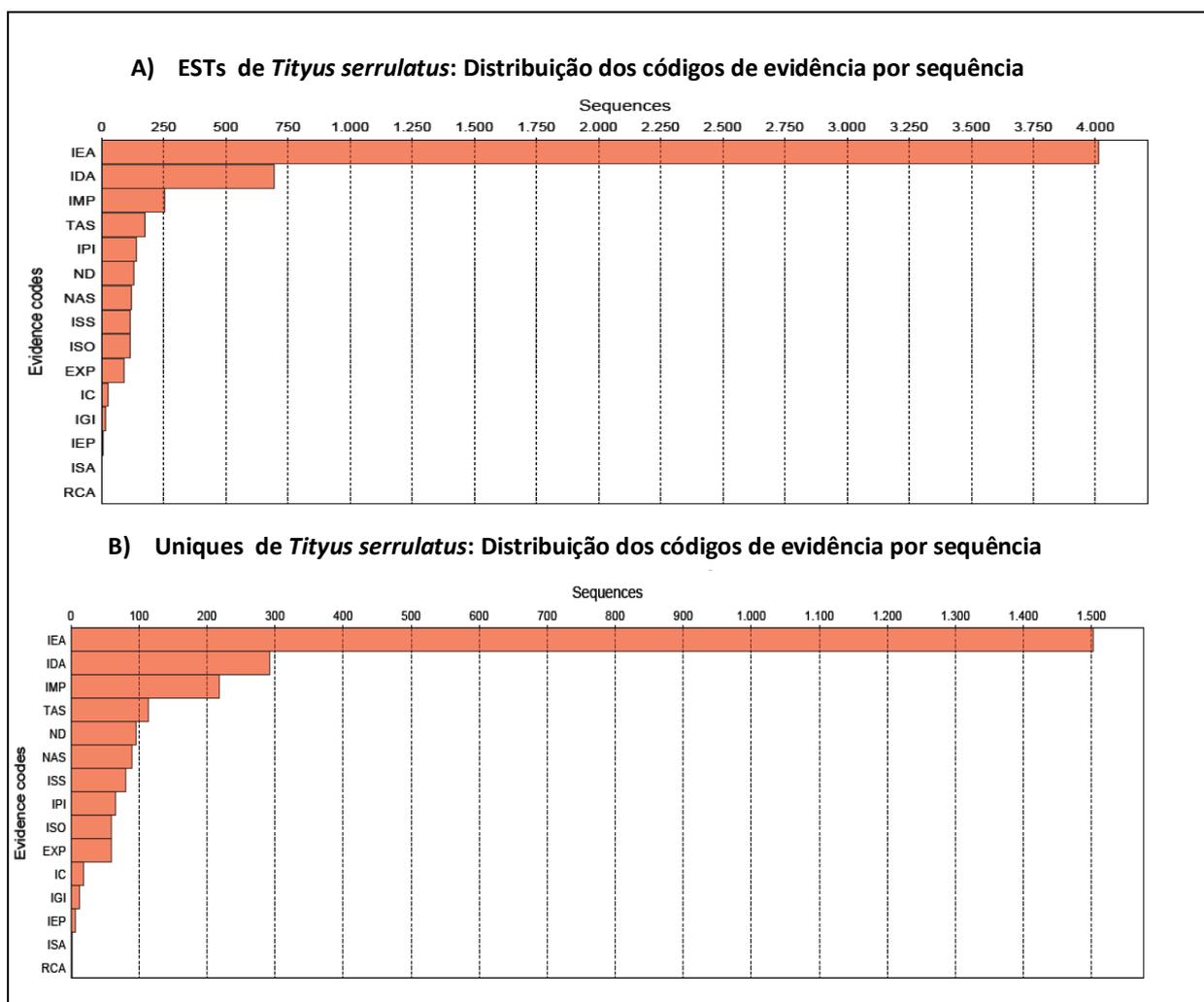
Apêndice A2: Estatísticas do Blast2GO: distribuição de similaridade e e-value. A e B) correspondem a distribuição de similaridade das ESTs e uniques respectivamente; C) indica a distribuição de e-value das ESTs e D) dos uniques. Os resultados de similaridade estiveram acima de 40%. Quanto ao e-value, maiores valores foram alcançados quando os uniques foram analisados, o que está possivelmente associado ao maior tamanho das sequências quando elas são agrupadas.



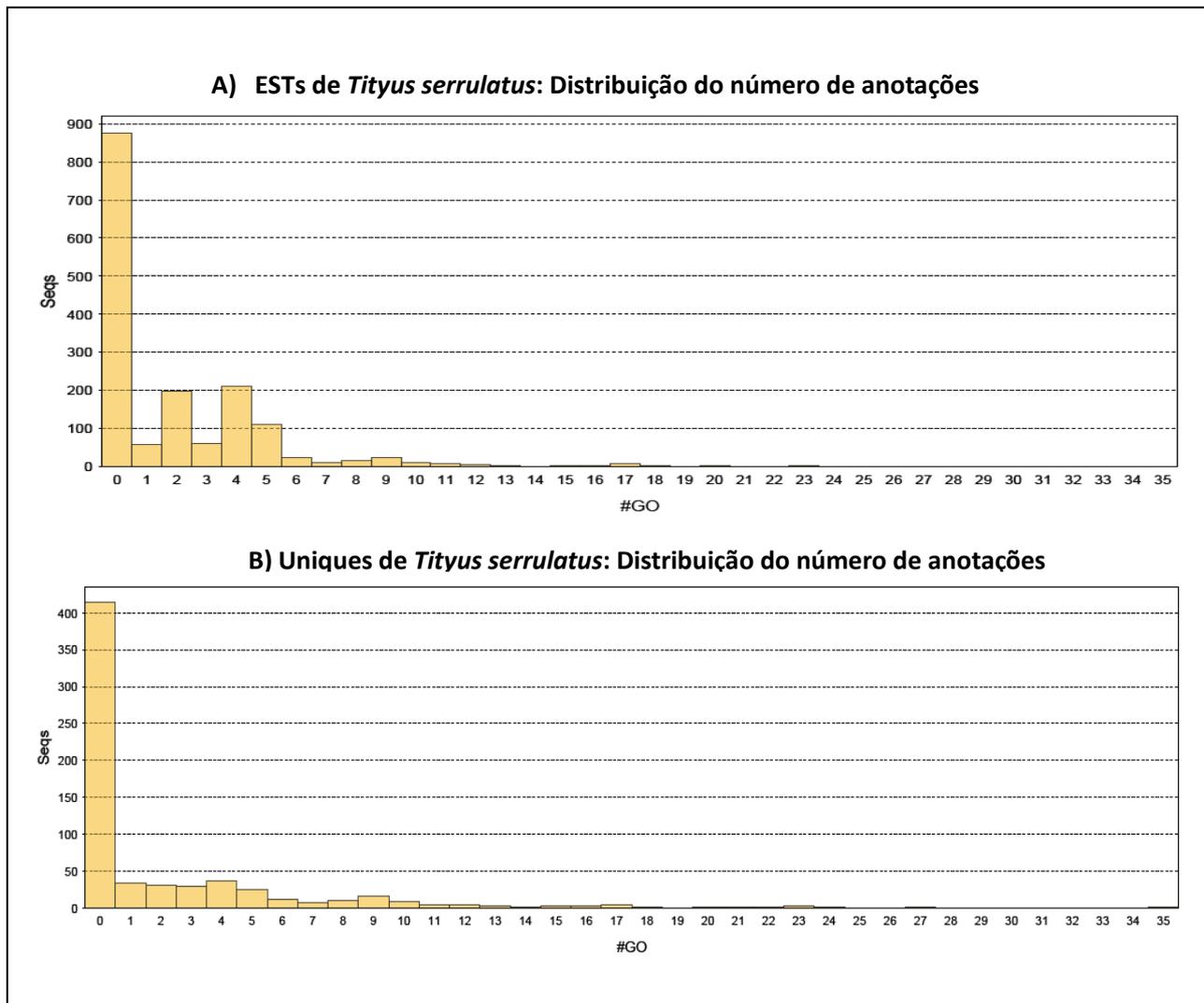
Apêndice A3: Estatísticas do processo de mapeamento: banco de dados usados para o mapeamento. A) dados obtidos pela análise de todas as ESTs e B) dados das uniques. UniProtKB/TrEMBL: seção do UniProtKB contendo coleção de informações funcionais de proteínas não revisadas e anotadas automaticamente; UniProtKB/Swiss-Prot : seção do UniProtKB contendo coleção de dados de proteínas revisados e manualmente anotados; MGI: Mouse Genome Informatics, RGD: Rat Genome Database; FB: Flybase; ZFIN: The Zebrafish Model Organism Database; WB: Wormbase; Refseq: seqüências de referência do NCBI.



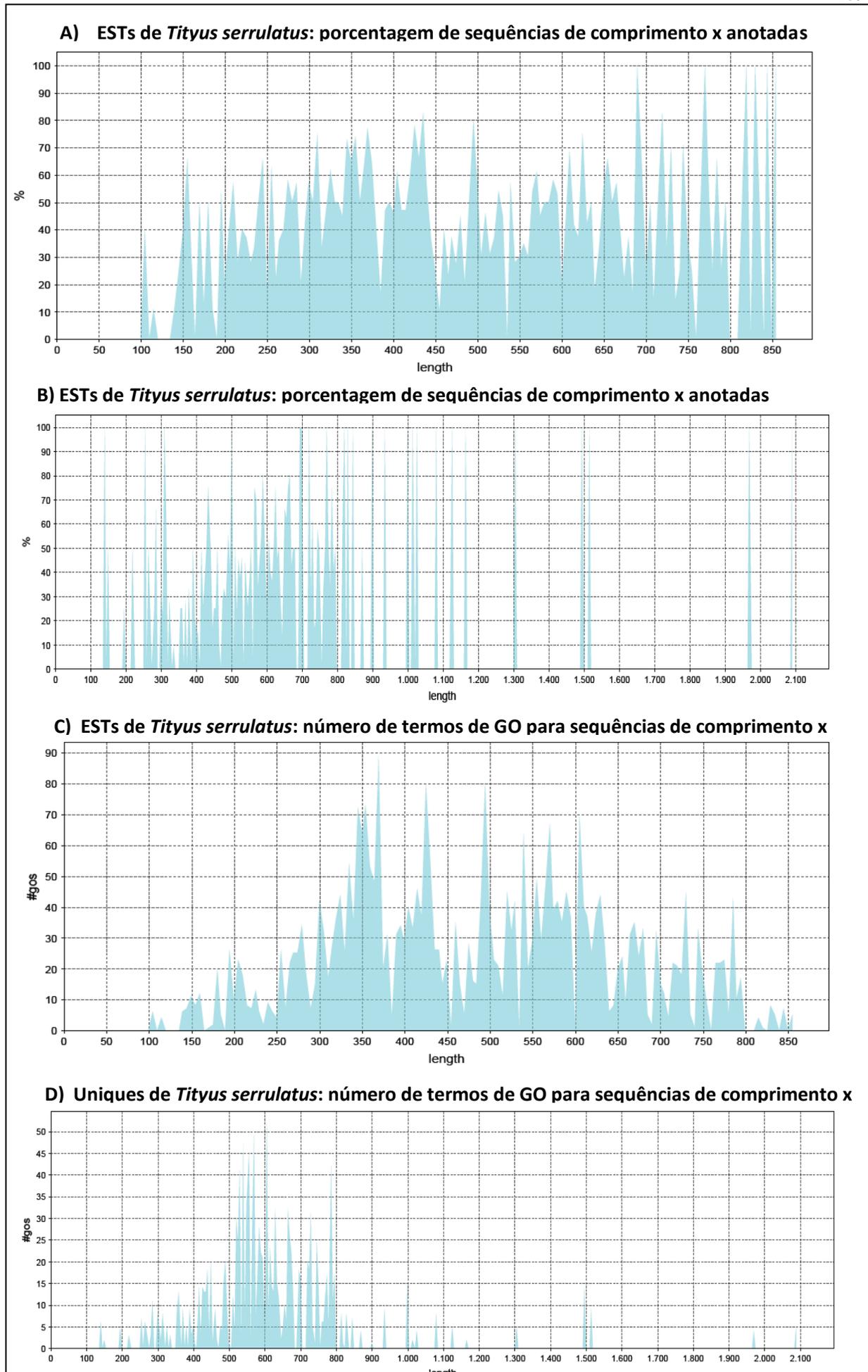
Apêndice A4: Dados do processo de anotação realizado pelo programa Blast2GO: distribuição dos códigos de evidência por *hits* do BLAST. Dados das ESTs (A) e dos uniques (B). Códigos de evidência: inferência por dado experimental (EXP), por ensaio direto (IDA), por interação física (IPI), por fenótipo mutante (IMP); por interação genética (IGI), por padrão de expressão (IEP), por similaridade de sequência ou estrutural (ISS), por ontologia de sequência (ISO), por alinhamento de sequência (ISA), por análise computacional revisada (RCA); por declaração rastreável do autor (TAS), por declaração não rastreável de autor (NAS), por curador (IC), por anotação eletrônica (IEA), dado biológico não disponível (ND). Conforme o que é geralmente observado (Conesa e Gotz, 2009), a anotação eletrônica (IEA) foi a mais freqüente. Entretanto, a contribuição dessa anotação para a assinatura da sequência pode ser modulada na etapa de anotação no programa Blast2GO. Assim, no presente estudo utilizaram-se os parâmetros default deste programa no qual a anotação eletrônica tem um peso de 0,7, enquanto evidências experimentais (EXP) tem peso 1 para a anotação final.



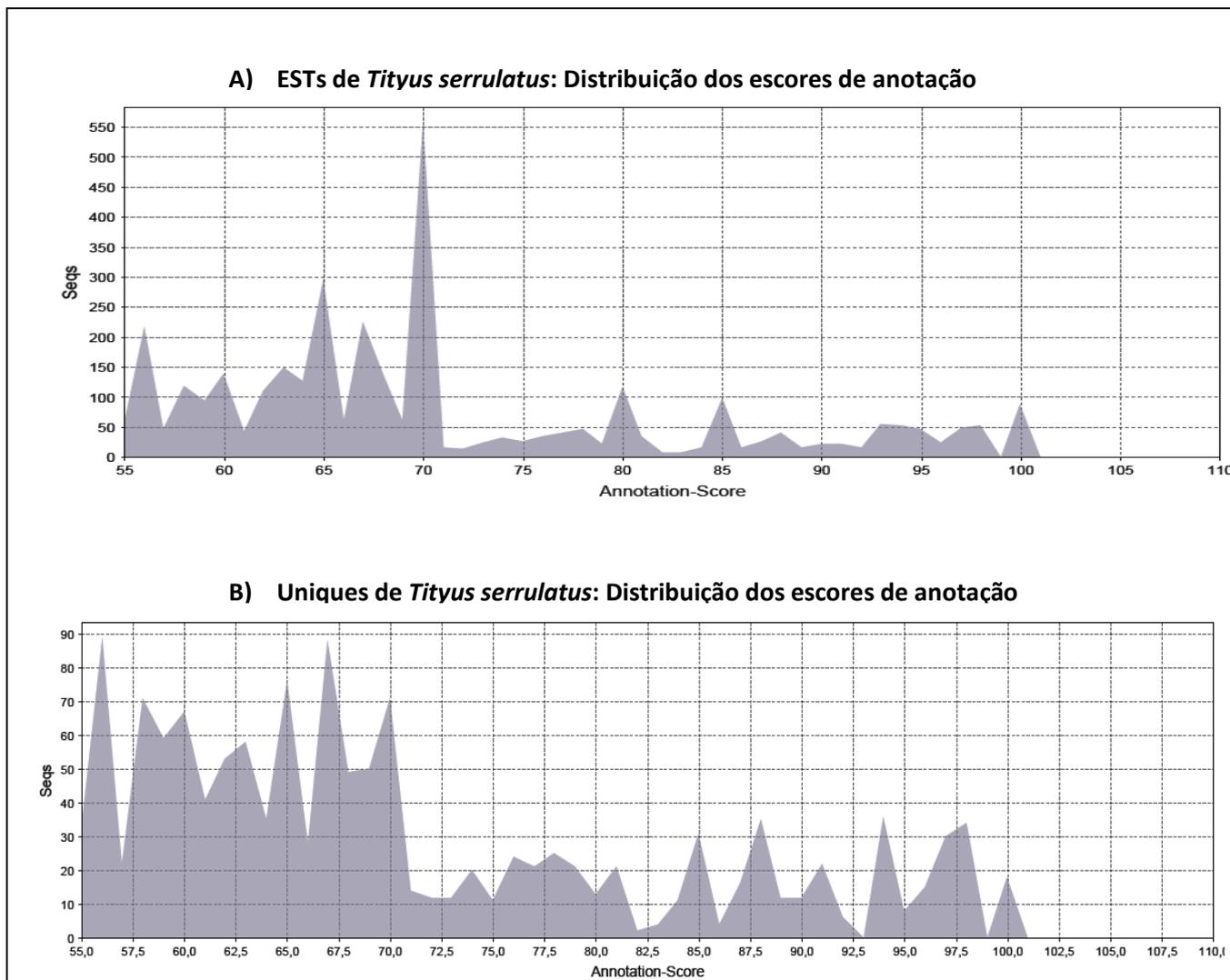
Apêndice A5: Dados do processo de anotação realizado pelo programa Blast2GO: distribuição dos códigos de evidência por sequência. Dados das ESTs (A) e dos uniques (B). Códigos de evidência: inferência por dado experimental (EXP), por ensaio direto (IDA), por interação física (IPI), por fenótipo mutante (IMP); por interação genética (IGI), por padrão de expressão (IEP), por similaridade de sequência ou estrutural (ISS), por ontologia de sequência (ISO), por alinhamento de sequência (ISA), por análise computacional revisada (RCA); por declaração rastreável do autor (TAS), por declaração não rastreável de autor (NAS), por curador (IC), por anotação eletrônica (IEA), dado biológico não disponível (ND).



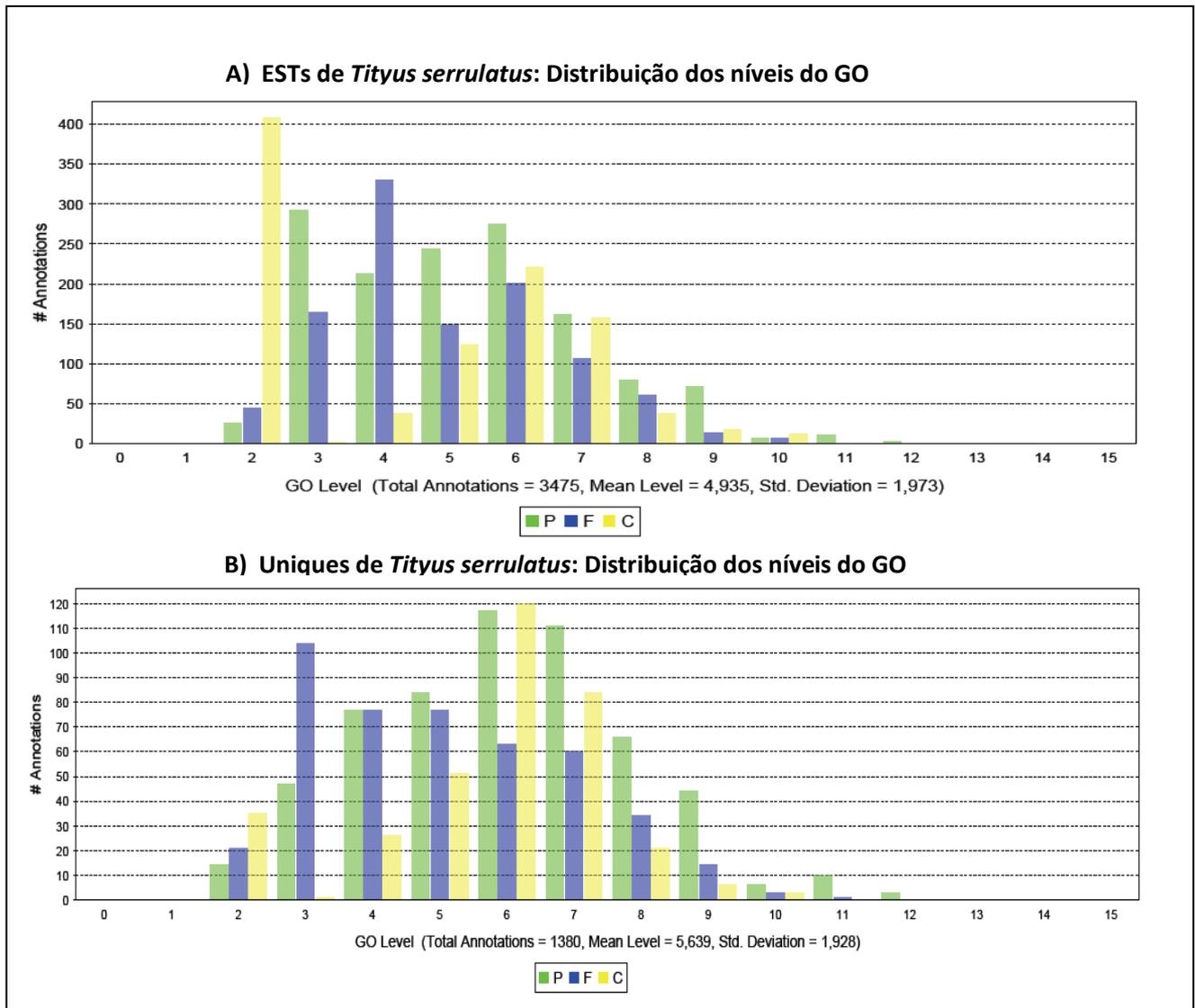
Apêndice A6: Dados do processo de anotação realizado pelo programa Blast2GO: distribuição do número de termos de GO. a) dados obtidos pela análise de todas as ESTs e b) dados dos uniques. para as ESTs, sequências anotadas com 1 a 5 termos foram mais frequentes e ocorreu uma variação acentuada entre esses números de termos do GO. Por outro lado, para os uniques, foi observada homogeneidade da distribuição do número de sequências anotadas com 1 a 10 termos.



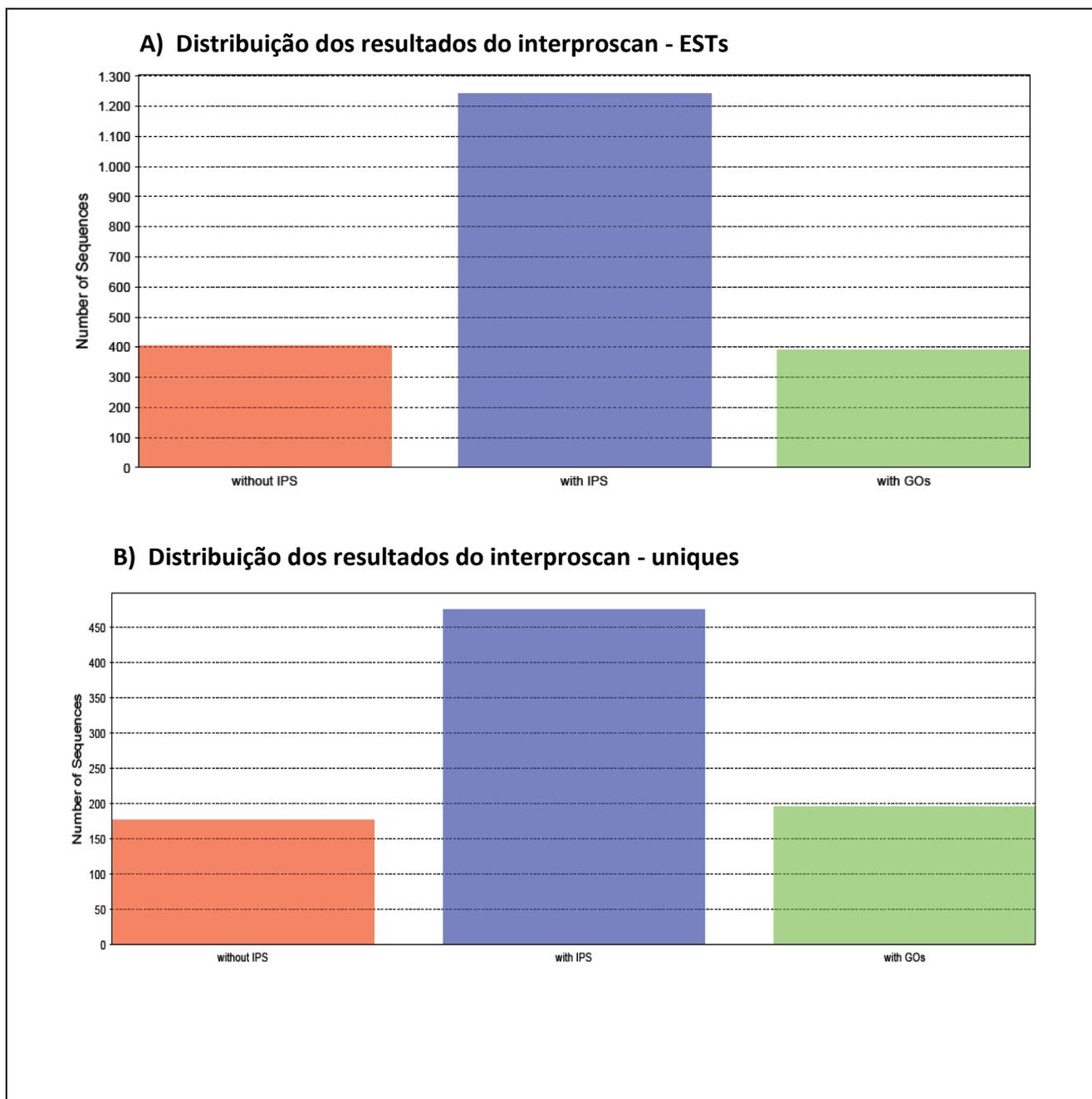
Apêndice A7: Dados do processo de anotação realizado pelo programa Blast2GO: porcentagem de seqüências de comprimento x anotadas considerando todas as ESTs (A) e por uniques (B); número de termos de GO para seqüências de comprimento x para as ESTs (C) e para os uniques (D).



Apêndice A8: Dados do processo de anotação realizado pelo programa Blast2GO: distribuição dos escores de anotação das ESTs (A) e dos unques (B).



Apêndice A9: Dados do processo de anotação realizado pelo programa Blast2GO: Distribuição dos níveis do GO para as ESTs (A) e para os uniques (B).



Apêndice A10: Distribuição dos resultados do Interproscan para as ESTs (A) e para os unives (B). O programa Blast2GO permite o uso da ferramenta InterProScan que possibilita a busca de informações de domínios protéicos nas sequências. Grande maioria das sequências apresentaram resultado com InterProScan tanto na análise de unives quanto de ESTs.

ANEXO: ARTIGO ACEITO PARA A PUBLICAÇÃO

Transcriptome analysis of the *Tityus serrulatus* scorpion venom gland

Érika R. Alvarenga¹, Thaís M. Mendes¹, Bárbara F. Magalhães¹, Flávia F. Siqueira¹, Arthur E. Dantas¹, Tatiana M. Barroca¹, Carolina C. Horta¹, Evanguedes Kalapothakis¹

¹Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
 Email: ekalapo@icb.ufmg.br

Received Day Month Year (2012).

ABSTRACT

The *Tityus serrulatus* scorpion is considered the most dangerous scorpion in Brazil and is responsible for several cases of human envenomation annually. In this study, we performed transcriptome profiling of the *T. serrulatus* venom gland. In addition to transcripts with housekeeping functions, such as those related to protein synthesis, energy supply and structural processes, transcripts from thirty-five families of venom peptides or proteins were identified. These transcripts included three new complete sequences of toxins and more than a dozen putative venom gland proteins/peptides. The venom gland transcriptome profile was verified by comparison with the previously determined proteomic profile. In conclusion, this transcriptome data provides novel insights into the putative mechanisms underlying the venomous character of *T. serrulatus*. The collected data of scorpion transcripts and proteins/peptides described herein may be an important resource for identifying candidate targets of molecular therapies and preventative measures.

Keywords: Scorpions, Antimicrobial peptides, Neutoxins, Venom glands, Brazilian yellow scorpion.

1. INTRODUCTION

In 2007, the World Health Organization (WHO) officially designated human envenomation by scorpion sting as a neglected public health issue and recommended urgent international action with a focus on the tropical regions where these venomous animals are abundant. In Brazil, scorpion stings have been recognized a serious public health threat for many decades; yet, the annual incidence of accidental stings remains high, with more than 50,000 cases reported in 2010 [1]. The most frequent culprit among the cases of human envenomation by Brazilian scorpions is the *Tityus serrulatus* of the

Buthidae family [2, 3]. As such, this species (known commonly as the Brazilian yellow scorpion) has been the subject of extensive research efforts to isolate and characterize its toxins and other venom components for potential clinical benefit [4-13].

Analyses of *T. serrulatus* venom have revealed a complex mixture of functionally diverse molecules, including neurotoxins, proteases, hyaluronidases, and hypotensins [3-14]. Of these, the neurotoxins are believed to play a key role in the pathogenesis of scorpionism through their interactions with sodium channels and potassium channels (as reviewed by Cologna *et al.* [3]). However, to gain a more in-depth and comprehensive understanding of the molecular components of the *T. serrulatus* venom gland, the gene expression profiling is necessary.

Transcriptomic analyses of the scorpion venom gland have been performed recently for species belonging to various scorpion families, including Buthidae (*T. stigmurus* and *Centruroides noxius*), Scorpiones (*Scorpiops magerisonae* and *S. jendeki*), Scorpiones (*Pandinus cavimanus* and *Heterometrus petersii*), Iuridae (*Hadrurus gertschi*), and Liochelidae (*Opisthacanthus cayaporum*) [15-26]. Transcriptomes have provided important insights into the biological processes that are taking place inside venom gland cells [15, 18, 19, 26], as well as identified new genes and expression patterns that form the basis of functional genomics studies [27-28]. Such data for the venom glands of different species represents a resource of potential toxicological and clinical relevance [29-30].

In the current study, the molecular complexity of *T. serrulatus* venom was investigated by transcriptome profiling of the venom gland. This approach identified the main cellular components and revealed new putative venom constituents, some of which may be candidate targets of new therapeutic strategies to help promote the health of sting victims.

2. MATERIAL AND METHODS

2.1. Library construction

A cDNA library was constructed from active venom glands of 60 *T. serrulatus* scorpions that had been milked two days prior to the RNA extraction, as previously described [31].

2.2. Expressed sequence tag (EST) sequencing, data processing, and bioinformatic analysis

For large-scale DNA sequencing (EST generation), random clones were grown in antibiotic selective medium for approximately 18 h. The plasmid DNA was then isolated using the standard alkaline lysis method and sequenced on an ABI 3130 sequencer with reagents from the BigDye Sequencing Kit (Applied Biosystems Inc., Foster City, CA, USA) and the standard M13 forward or reverse primers.

The resultant trace files of the sequenced clones were applied to the Phred program for base calling and quality scoring using a Phred score cut-off value of 20 [32]. The nucleotide sequences corresponding to vector, adaptors, and *Escherichia coli* DNA, and any short transcripts (<100 pb) were removed by the SeqClean program [33] (<http://compbio.dfci.harvard.edu/tgi/software>). The final sequences were deposited in GenBank under accession IDs JK731601 – JK732954. The TGICL program [34] was used to assemble high-quality ESTs into contigs (overlapping ESTs that together represent a consensus sequence). Any ESTs without significant similarity to any other ESTs were classified as singlets. Considering the diversity of scorpion toxins, those clusters putative to encode venom peptides were re-examined manually to pick out individual different isoforms. The electropherogram of each putative isoform was visually inspected to confirm the sequencing quality in the polymorphic region.

To annotate the contigs, the set of contigs plus singlets (collectively designated as 'uniques') were searched against the UniProt database [35] (March 2011) using the BLASTx [36] (e-values < 1×10^{-10}) and nr database using BLASTx (e-values < 1×10^{-10}) in the BLAST2GO program [37]. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis [38] was also performed. The toxin nomenclature was assigned according to the recommendations of King *et al.* [39] and Cologna *et al.* [3].

Amino acid sequence of the transcripts was deduced using the open reading frame (ORF) Finder program (<http://www.ncbi.nlm.nih.gov/projects/gorf/>). Subsequently, the Seed Server (unpublished, Guedes *et al.*) was used to search for venom component orthologs. Briefly, the Seed Server methodology uses a protein of interest to

group homologues by means of Seed Linkage software [40] and UniRef50 Enriched KEGG Orthology (UEKO) clusters built with the procedure described by Fernandes *et al.* [41]. The sequences used for alignment analysis were retrieved from the UniProt database and the alignment was performed by webPrank [42] and visualised by the Jalview program [43]. Peptide signal prediction was carried out for putative venom components by using the SignalP 4.0 program [44].

3. RESULTS

3.1. EST sequencing and clustering

A total of 1629 high-quality ESTs were generated from the *T. serrulatus* venom glands. The average length of these ESTs was 421 bp. TGICL-based clustering of the 1629 ESTs yielded 185 contigs (= 1171 ESTs). The average length of these contigs was 625 pb. Four-hundred-and-fifty-eight ESTs showed no significant similarity to any other ESTs in the database and were identified as singlets.

3.2. *T. serrulatus* venom gland transcript profile

The BLASTx searches against the UniProt database indicated that among the 643 uniques (185 contigs and 458 singlets), 54% (348 uniques representing 1259 ESTs) encoded precursors of known proteins. The remaining 46% (295 uniques representing 370 ESTs) had e-values > 1×10^{-10} and were thus designated as having no match. These no-match transcripts may represent new proteins or peptides.

Cellular localisation and function analyses classified the uniques- and ESTs-deduced products into 10 categories; the distributions of which are shown in Figure 1A and B. The most frequently represented functional categories among the cellular components in *T. serrulatus* venom gland were: protein synthesis and processing, structural function, and energy supply.

Table 1. The most well represented KEGG pathways in the *T. serrulatus* transcriptome

	KEGG Pathway		Annotated ESTs
	Entry	Name	
1	03010	Ribosome	22
2	00190	Oxidative phosphorylation	16
3	04141	Protein processing in endoplasmic reticulum	9
4	04260	Cardiac muscle contraction	9
5	04974	Protein digestion and absorption	5
6	03013	RNA transport	4
7	04010	MAPK signalling pathway	4
8	04810	Regulation of actin cytoskeleton	4
9	04910	Insulin signalling pathway	4
10	04972	Pancreatic secretion	4

3.3. Housekeeping genes

When considering only the ESTs that presented similarity with sequences in the database, about 40% of ESTs were related to cellular components. KEGG analysis (corroborated with manual annotation; Table 1) indicated that the most frequently represented pathways were also involved in protein synthesis and processing (ribosome and protein processing in the endoplasmic reticulum), structural processes (cardiac muscle contraction – important to venom release; regulation of actin cytoskeleton), and energy supply (oxidative phosphorylation). The automated KEGG pathway analysis and manual annotation method yielded similar but not identical results, likely due to the different rules of each approach used to categorize the transcript products.

By manual annotation, the most frequently represented functional categories were protein synthesis and processing (37 ESTs coding ribosomal proteins), eukaryotic translation initiation (11 ESTs), elongation factors (seven ESTs), disulfite isomerases (four ESTs), and heat shock proteins (HSPs; four ESTs). Seven ESTs encoded peptides that were related to degradation through the ubiquitin-proteasome system. These transcripts might play a role in ensuring that misfolded or otherwise abnormal proteins are recognised and eliminated, or may regulate various molecular pathways in the venom gland, as suggested by Rendón-Anaya *et al.* [26].

Several transcripts expressed in the venom gland of *T. serrulatus* were identified as structural components, which may be involved in the maintenance of gland structure and/or contractile activity that mediates venom release. In particular, transcripts were found that encoded actin (16 ESTs), troponin (11 ESTs), myosin (eight ESTs), alpha tubulin (seven ESTs), paramyosin (15 ESTs), and five other ESTs specifically related to cytoskeleton organisation.

The synthesis of venom is considered an energetically costly process. About 5% of the ESTs (8% of uniques) aligned with mitochondrial proteins encoded by nuclear DNA. The majority of these transcripts were related to energy production; in particular transcripts were found that encoded cytochrome c oxidase (17 ESTs), NADH-ubiquinone oxidoreductase (15 ESTs), ATP synthase (11 ESTs), and cytochrome b (six ESTs).

3.4. ESTs related to venom components

Five-hundred-and-ninety-four of the ESTs comprised 51 clusters and coded for 35 different families of peptides or proteins related to venom components (Table 2). Except for the incomplete sequences (indicated by asterisks in Table 2), all transcripts had signal peptides. The most abundant and diverse venom transcripts (Figure 1D and E) encoded neurotoxins (sodium channel toxins or NaTxS and potassium channel toxins or KTxs). In addition,

a wide variety of transcripts encoded secreted peptides and proteins. The components that were abundantly expressed included the Pape peptide, similar to bradikinin-potentiating peptide or BPP (UniProt ID: P86821), antimicrobial peptide (AMP), metalloproteases (zinc metalloproteases [UniProt ID: P85842] plus antareases [UniProt ID: P86392]), non-toxic protein NTxP (TsNTxP; UniProt ID: O77463), anionic peptide and hypotensin (UniProt IDs: P84189 and P86824); however, phospholipase A2, hyaluronidase (UniProt ID: P86821), allergen (UniProt ID: P85840) and some cysteine-rich peptides were less frequently expressed.

Table 2. ESTs and contigs of *T. serrulatus* venom components.

Venom components	<i>T. serrulatus</i>	
	ESTs	Contigs
Na⁺ channel toxins		
Ts1 (β-BUTX-Ts1a) [P15226]	33	2
Ts2 (δ-BUTX-Ts1a) [P68410]	36	1
Ts3 (δ-BUTX-Ts1b) [P01496]	10	2
Ts17 (U-BUTX-Ts1b) [new]	9	1
Ts18 (U-BUTX-Ts1c) [new]	5	1
K⁺ channel toxins		
Ts8 (κ-BUTX-ts1) [P69940]	73	1
Ts15 (κ-BUTX-ts2) [P86270]	7	2
Ts6 (λ-BUTX - Ts1b) [P59936]	5	1
Ts7 (κ-BUTX - Ts1a) [P46114]	4	2
Ts16 (U-BUTX-Ts1a) [P86271]	1	1
Ts19 (U-BUTX-Ts1d) [new]	41	1
Other components		
Pape peptide or BPP-like [P86821]	138	1
AMPs [new]	59	2
Ponericin-W-like of <i>L. mucronatus</i> [new]	40	1
Antarease* [P86392]	29	7
TsNTxP or Ts4 [O77463]	26	1
Anionic peptide [new]	18	1
Hypotensin 1 or Ts14 [P84189]	11	1
Hypotensin-like [P86824]	10	1
Zinc Metalloproteases* [P85842]	7	3
Similar to glycine-rich protein <i>L. mucronatus</i> [new]	6	3
TsPep3 or Ts13 [P0C176]	4	1
Similar to venom neuropeptide-3 of <i>Buthus eupeus</i> [new]	3	1
Cysteine-rich peptide [new]	3	1
Similar to TsPep2 or Ts12 [P0C175]	2	1
Hyaluronidase* [P85841]	2	2
Similar to venom toxin-like peptide of <i>B. eupeus</i> [new]	2	1
Phospholipase A2* [new]	2	1
Similar to cysteine-rich peptide of <i>Camponotus floridanus</i> * [new]	2	1
Similar to cysteine-rich peptide of <i>Hottentotta judaicus</i> * [new]	1	1
TsPep1 or Ts11 [P0C174]	1	1
Allergen* [P85840]	1	1
Similar to lipolysis-activating peptide 1-alpha chain of <i>L. mucronatus</i> * [new]	1	1
Similar to venom protein of <i>B. eupeus</i> * [new]	1	1
Similar to secreted protein of <i>Opisthacanthus cayanorum</i> [new]	1	1

* indicates sequences with incomplete ORF. The toxin nomenclature suggested by King *et al.* [39] is presented in parentheses.

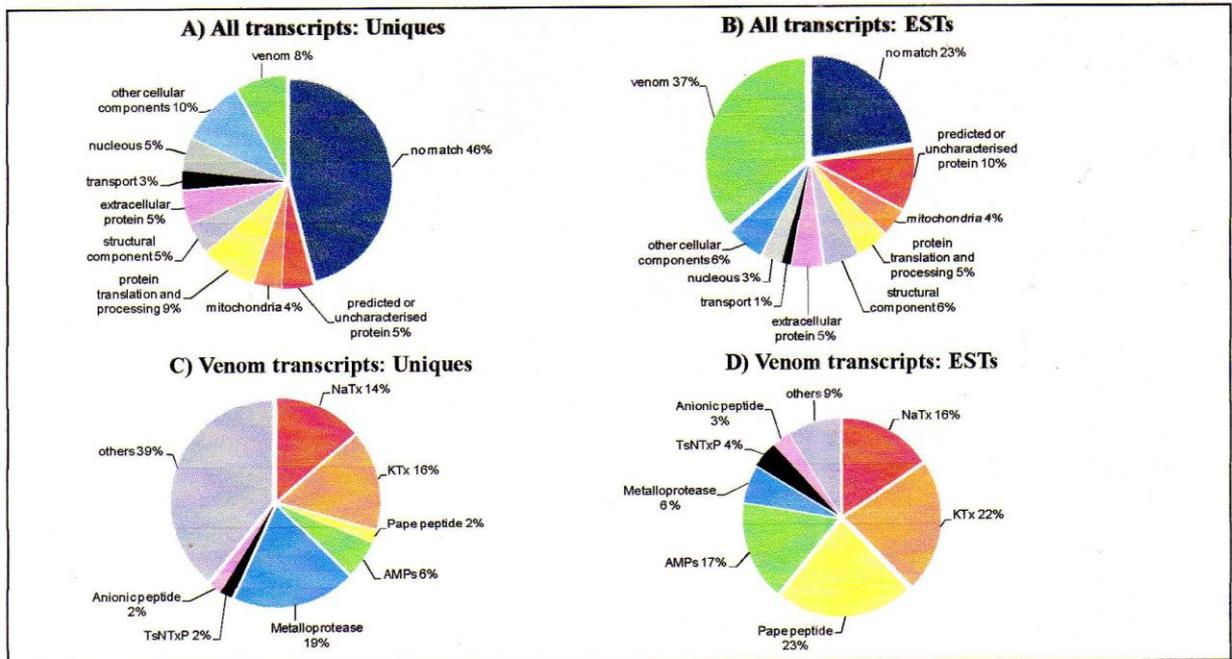


Figure 1. Molecular characteristics of the *T. serrulatus* venom gland transcriptome. (A) Relative proportion of each category in the uniques. ‘Venom’ includes transcripts encoding toxins and other secreted components previously described in scorpion venom. ‘Mitochondrial’ and ‘Nuclear’ categories comprise ESTs coding conserved proteins located in these cellular organelles. ‘Protein translation and processing’ contains transcripts encoding, for instance, ribosomal protein, disulfide-isomerase, and other proteins related to protein synthesis. ‘Structural components’ includes mainly cytoskeleton proteins, such as actin, myosin, and tubulin. ‘Transport’ comprises transcripts encoding proteins involved in intracellular trafficking, such as the copper transport protein. ‘Predicted or uncharacterized protein’ includes ESTs similar to previously described sequences that have no functional assessment. ‘Extracellular’ comprises transcripts encoding extracellular proteins that are, for instance, found in the extracellular matrix, such as fibronectin. ‘Other cellular components’ includes ESTs encoding cellular components that were not included in any of the other categories. (B) Distribution of ESTs in the same categories described in (A). (C) Relative proportion of uniques encoding different classes of venom components. (D) Abundance of transcripts in the main peptide/protein families of *T. serrulatus* venom from Table 2.

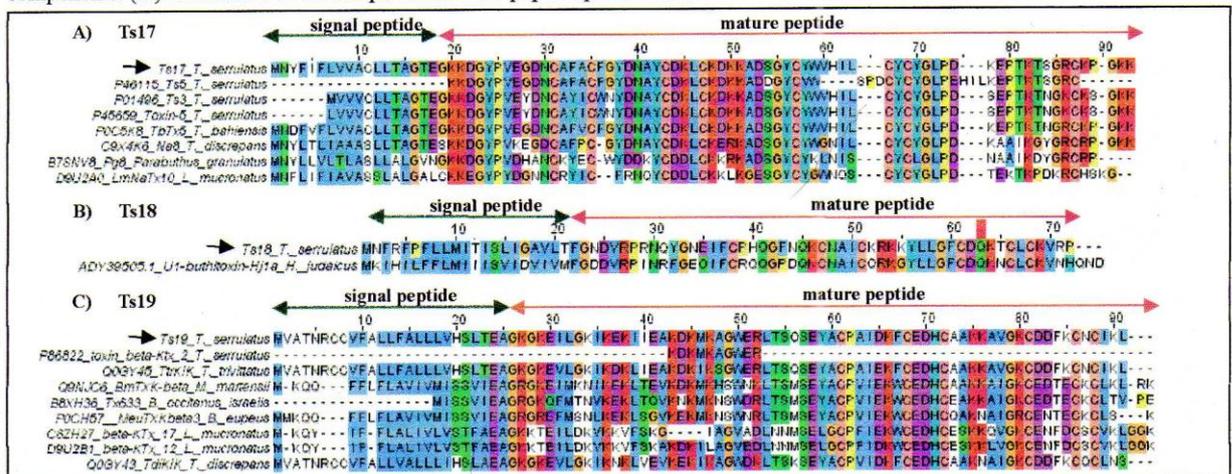


Figure 2. Sequence alignments of NaTxs (A and B) and KTx (C). (A) The translated sequence of Ts17 and orthologs. For comparison, the paralogous sequences of Ts3 and Ts5 are also presented. The UniProt ID and the corresponding scorpion species are indicated. (B) The Ts18 from the *T. serrulatus* cDNA library and the related sequence of U1-buthitoxin-Hj1a from *Hottentotta judaicus*, a predicted NaTx [22]. (C) Complete sequence of Ts19 (KTxs) and ortholog alignment. The UniProt IDs and corresponding scorpion species from genus *Tityus*, *Mesobuthus*, *Lynchus*, *Buthus* are indicated. The deposited fragment of the potassium channel toxin beta-KTx 2 is presented for comparison. The signal peptide (green arrow) and cleavage sites are shown for Ts17 (between 19-20 amino acid residues), Ts18 (between 21-22 amino acid residues) and Ts19 (25-26 amino acid residues). Black arrows indicate sequences identified in the current study. Dots represent gaps introduced to improve alignment.

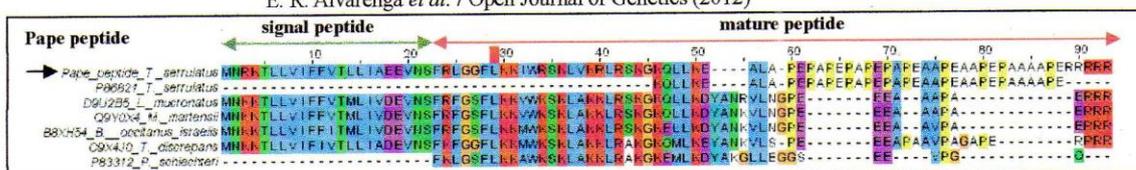


Figure 3. Alignment of the Pape peptide with its orthologs. The black arrow indicates the sequence identified in the current study and P86821 is the Pape peptide previously identified. The signal peptide (green arrow) and the cleavage site (between 22-23 amino acid residues) are shown. The UniProt ID and the corresponding scorpion species are indicated. Dots represent gaps introduced to improve alignment.

3.5. New venom components of *T. serrulatus*

Although *T. serrulatus* venom has been extensively studied, the current transcriptome analysis revealed at least three complete sequences of new potential toxins and more than a dozen new venom components (Table 2). The complete sequence of some of these are detailed below and presented in Figures 2-5. It is important to note that still other new venom components may be represented by those sequences that were designated as no-match (see the Discussion for more detail).

3.5.1. Sodium channel toxins

Three of the known major NaTxs of *T. serrulatus* venom, Ts1, Ts2 and Ts3, were found in the current transcriptome analysis. Although the Ts5 toxin was not found in this study, a probable new NaTx that is similar to Ts5 (P45659; identity = 86%) was found (designated as Ts17; U-BUTX-Ts1a). Figure 2A shows the alignment of the Ts17 sequence with Ts3 (UniProt ID: P01496), Ts5, and other orthologs. In addition, five ESTs matched significantly with the U1-buthitoxin-Hj1a (identities = 63%), a predicted NaTx derived from *Hottentotta judaicus*, and the *T. serrulatus* sequence was named Ts18 (U-BUTX-Ts1b) (Figure 2B).

3.5.2. Potassium channel toxin

Among the KTx identified in *T. serrulatus* venom, Ts6, Ts7, Ts8, Ts15 and Ts16 had been previously identified. However, the Ts9 was not found in this study. The potassium channel toxin beta-KTx 2 (UniProt ID: P69940) had been previously identified by proteomic analysis and deposited as peptide fragment [30]; however, our analysis identified its precursor (designated as Ts19; U-BUTX-Ts1c) and its orthologs (Figure 2C).

3.5.3. Other components

The category "other components" included non-neurotoxic venom components that had been previously described for *T. serrulatus* as well as potential new venom components harbouring signal peptides and/or showing similarity to venom constituents of other scorpion species (Table 2). However, more evidence is needed to confirm the occurrence of these protein and/or

peptide in *T. serrulatus* venom. It is possible that the expression of some of these secreted components is restricted to the venom gland (for example, secreted components in connective tissue of this organ). Here, we summarize the data of new sequences for highly-expressed transcripts, including the Pape peptide, AMPs, anionic peptides, and hypotensins.

Second only to the neurotoxins, the Pape peptide was the most highly expressed transcript in the *T. serrulatus* venom gland transcriptome. The Pape peptide (UniProt ID: P86821) was previously identified by proteomic analysis and deposited as a peptide fragment [30]. The complete sequence of the peptide precursor and its orthologs are shown in Figure 3. While the signal peptide is highly conserved among these peptides, the N-terminal region is moderately conserved and the C-terminal region shows very little conservation.

Sequences of AMPs and anionic peptides were abundant in the venom gland transcriptome of *T. serrulatus*, indicating that these components may play important roles in the function of this organ. The complete sequences of these venom components are shown in Figure 4. Three different sequences coded an AMP highly similar to the putative AMP of *T. costatus* (Figure 4A). In addition, two poneracin-like sequences were found and presumed to be antimicrobial peptide (Figure 4B). In this study, 18 ESTs encoded the same sequence of anionic peptide (Figure 4C). Anionic peptides have been previously reported as highly expressed and conserved among the Buthidae scorpion species [17, 23]. Although the function of these peptides remains unknown [45], some researchers have suggested that they might play antimicrobial activity [46] or an important role in pH balance, since neurotoxins are basic peptides [18, 22].

Hypotensins were identified in *T. serrulatus* venom gland transcriptome. These random-coiled linear peptides are characterized by the bradykinin-potentiating peptide amino acid signature [13]. Of the three sequences in the UniProt database (hypotensin-1, UniProt ID: P84189; hypotensin-2, UniProt ID: P84190; hypotensin-like peptide, UniProt ID: P86824), the complete sequence of hypotensin-like and hypotensin-1 precursors were identified (Figure 5).

4. DISCUSSION

Transcriptome profiling has provided comprehensive

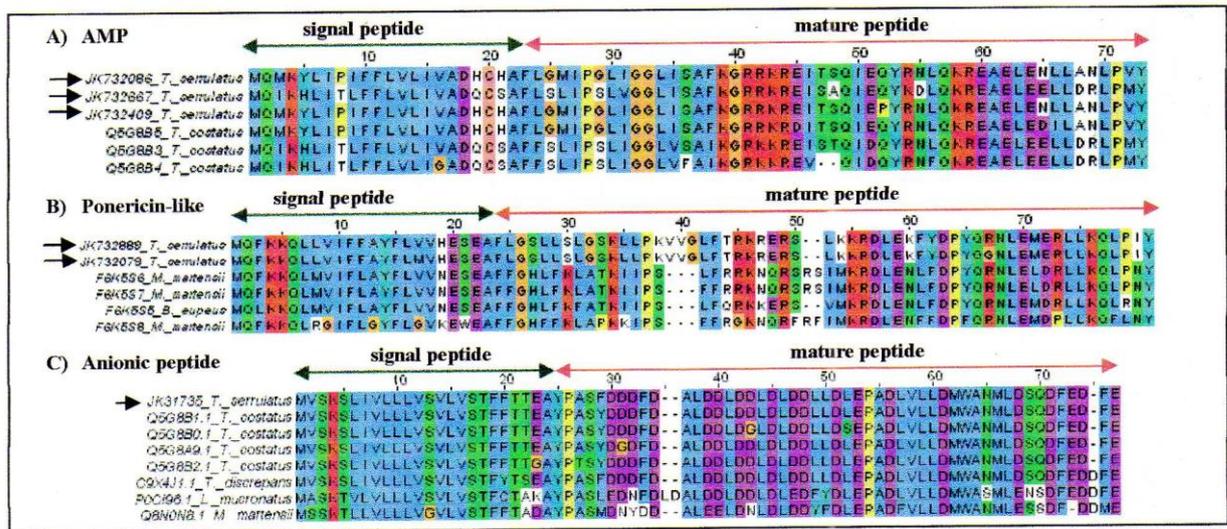


Figure 4. Sequence alignment of putative antimicrobial peptides and orthologs. (A) Three sequences of AMP exist in the Buthidae scorpions *T. serrulatus* and *T. costatus*. (B) Two Ponericin-like peptide sequences, presumed to be an antimicrobial peptide, were found in *T. serrulatus* transcriptome. (C) Sequence of anionic peptide found in *T. serrulatus* and alignment with its relatively conserved orthologs from other Buthidae scorpions. The signal peptide (green arrow) and cleavage sites are shown for AMP (between 22-23 amino acid residues), Ponericin-like peptide (between 23-24 amino acid residues) and anionic peptide (24-25 amino acid residues). Black arrows indicate sequences identified in the current study.

information of venom glands, accelerating the discovery of their peptides and proteins [47]. Currently, 12 studies in the publicly available literature have reported data of the venom glands from various scorpion species [15-26]. The study described herein not only provides the first transcriptome data for the *T. serrulatus* venom gland but also the largest transcript catalogue for scorpion venom glands obtained by Sanger sequencing to date. This work also focused on the discovery of new genes (using the EST approach) which when added to the larger databases will allow for better anchoring of transcriptomic short reads generated by most of the next-generation sequencing (NGS) platforms.

In this study, 35 different peptide families, coded by 594 ESTs, were identified as related to venom components. Previous proteomic analysis of *T. serrulatus* venom [14, 30] has identified many toxins, in addition, venom components that have since been isolated and studied individually [4-9, 13]. The Animal Toxin Annotation Program (UniProt) [48], which annotates the secreted proteins in animal venoms, has already identified 32 peptides/proteins in *T. serrulatus* venom. The current transcriptome analysis identified three precursors of potential new toxins (Ts17, Ts18, and Ts19) and more than a dozen new venom components in *T. serrulatus*. However, this list is certainly not exhaustive, since a large number of sequences without matches (295 uniques representing 370 transcripts) were found. We intend to continue to explore the entire dataset to identify other new

venom components and to verify the biological function and clinical relevance in envenomation for any promising factors.

Some of the venom components previously identified were not identified by the current transcriptome analysis, such as Peptide T (UniProt ID: Q9TWR4), Ts5 (UniProt ID: P45659) and alpha amylase (UniProt ID: P85843). However, the absence of venom components at the transcript level is not a surprising observation. In fact, a previous study in *C. noxius* indicated that the powerful pyrosequencing platform, producing over three million reads, was unable to detect the entire panel of known toxins [26]. Some hypotheses to explain the transcript absence are: (1) the transcriptomic analysis performed did not provide complete coverage; (2) toxin genes were down-regulated at the time of RNA extraction or had undergone microRNA-mediated degradation during processing, as suggested by Rendón-Anaya *et al.* [26]; and (3) the venom protein/peptide had undergone post-translational modifications that significantly differentiated it from the intact form, as described by Pimenta *et al.* [14].

Considering the previously reported protein/peptide compositions of *T. serrulatus* venom [3, 14, 30], we expected to find to a high level of neurotoxin expression. Indeed, NaTxs, the main agents responsible for the toxic effects of *T. serrulatus* envenomation [49], presented high expression; in particular, this was observed for Ts1 and Ts2. In the current study, NaTx represented about

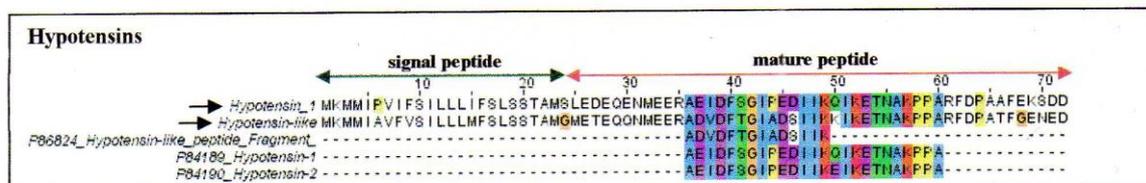


Figure 5. Sequence of hypotensins found in *T. serrulatus*. Black arrows indicate sequences identified in the current study. The signal peptide (green arrow) with the cleavage site between 24 and 25 amino acid residues is shown.

6% of the total transcripts. However, transcriptome analysis of another scorpion Brazilian species, *T. stigmurus*, venom gland [25], showed lower expression levels of NaTx (1.3% of the total transcripts). The higher expression of NaTx in *T. serrulatus* may be related to the higher lethality of this scorpion, as compared to *T. stigmurus* [25]. KTx also represented a high number of ESTs in *T. serrulatus*, representing about 8% of the total transcripts. Specifically, Ts8 and Ts19 were responsible for the higher expression level. However, the overall expression level of KTx in *T. serrulatus* was lower than in *T. stigmurus*, for which KTx represented 13.5%. The differences in expression level of NaTx and KTx between *T. serrulatus* and *T. stigmurus* may be species-specific or reflect the differences of transcriptional profiles for active and resting venom glands. The *T. stigmurus* cDNA library was performed for resting venom glands [25], while the current study used active venom glands.

Another important finding from the current study is the robust expression of some particular components, such as the Pape peptide, AMPs, anionic peptide, and metalloproteases. Although further analysis is required to uncover the precise functions of these venom components in the active venom gland of *T. serrulatus*, their observed abundance suggests an important role in the biological function of this species' venom gland. Indeed, the Pape peptide and a peptide similar to Ponericin-L1 and Ponericin-L2 were identified in the previous proteomic study of *T. serrulatus* venom performed by Rates *et al.* [30]. The current transcriptome analysis identified precursors of both peptides. Interestingly, the Pape peptide represented 8.5% of the total transcripts. This peptide was similar to the BPP found in other scorpion species and Parabutoporin, an antimicrobial peptide identified in *Parabuthus schlechteri* (Figure 3). Despite its higher expression level, the function of the Pape peptide in *T. serrulatus* venom remains unknown. Besides the Ponericin-like peptide, sequence similar to AMP was found. Various AMPs have been previously identified in the venom of several scorpion species [15, 18-19] and appear to be highly expressed in the *S. jendeki* [18], *H. petersii* [20], *Isometrus maculatus* [23] and *T. stigmurus*

[26] species. While no consensus has yet been reached about the precise functions of AMP in scorpion venom, it is theorized they may act as protectants against bacterial infection or potentiators of neurotoxin action [50]. The demonstrated antibacterial activities of AMPs in animals, plants, and insects have indicated their potential for use as antimicrobial agents [46, 50-52]. Hence, the AMPs described in the current study might represent potential candidates for anti-infective drugs.

Morgenstern *et al.* [22] showed a higher abundance of proteases and metalloproteases in the resting venom gland of the Buthidae scorpion. All animals used in the present study had been milked prior to the mRNA extraction, yet the expression level of metalloproteases was as high as some of the NaTx and KTx sequences. Prosdocimi *et al.* [53] found a similar high expression of metalloproteases (astacin family) in the *Gasteracantha cancriformis* spider's spinning gland transcriptome, and suggested that these proteins may play a role in the remodelling processes of silk fibre deposition. A higher expression of these proteases was also observed in the *L. mucronatus* venom, suggesting that these proteins might play a central role in scorpion venom as well [21]. Indeed, several antarease members of the metalloprotease family have been purified from the *T. serrulatus* venom gland and shown to selectively cleave the essential SNARE protein within mammalian pancreatic tissue; it has been suggested that this function may be responsible for the pancreatitis that develops in some patients following scorpion envenomation [54]. Alternatively, the metalloproteases may play a specific non-toxic role in the scorpion's venom gland. A large number of fragments derived from larger peptides have been found by the proteomic studies of venom glands [14, 30]. It appears that different fragments derived from the same peptide might have distinct biological functions [55]. Hence, we speculate that the protease-mediated fragmentation may act to exponentially increase the diversity of venom peptides and their biological targets.

The scorpion's venom is mainly used for prey capture and defence. The venom gland is a complex organ where a large number of substances are synthesized. Transcriptomic tools have accelerated the description of

venom components, greatly expanding the publicly available sequence databases. A large diversity of venom components were found using this approach: NaTx, KTxs, scorpines, calcines, AMPs, BPPs, anionic peptides, proteases, glycine-rich peptides, phospholipases, lectins, hypotensins [15-26]. In the current transcriptome study – the first of its kind to examine the active venom gland of *T. serrulatus* – several of these components were found as well as several new components, including potential toxins, AMPs, and cysteine-rich peptides. Another important finding was the surprisingly high transcription levels of some non-neurotoxic components, such as the Pape peptide, AMPs, anionic peptides, metalloproteases, and hypotensins. Overall, this analysis provides a novel and more comprehensive insight into the venom arsenal of *T. serrulatus*. These data may act as an important resource for future investigations of the evolution of the scorpion venom arsenal, envenomation mechanisms, and discovery of bioactive peptides and proteins.

5. CONCLUSION

This work represents a step towards better understanding the gene expression profile of the active *T. serrulatus* venom gland, expanding our knowledge of its peptide and protein content. The transcriptome analysis revealed at least three new precursors of toxins and more than a dozen potential venom components. Besides the more common toxins (NaTx and KTx), some transcripts, such as the Pape peptide, metalloproteases, AMPs, anionic peptides, and hypotensins, presented high expression level. The functions of these components may help to advance our understanding of the biological and molecular processes of the venom gland. The gene expression profile of the venom gland agreed with the activated state of this organ and revealed the activities of protein synthesis and processing and energy supply. Some potential bioactive proteins/peptides (such as Ts17, Ts18, Ts19, and AMPs) described in this work may be an important resource for the investigation and characterization of molecules applicable in pharmaceutical research and biotechnology.

6. ACKNOWLEDGEMENTS

We wish to thank the Brazilian National Council for Research (CNPq), FAPEMIG and CAPES (Edital Toxinologia 063/2010), for financial support. The FAPEMIG scholarship awarded to Érika Ramos de Alvarenga is also gratefully acknowledged. The authors also are grateful to Dr. J. Miguel Ortega and Raphael L. M. Guedes for help in the bioinformatics analysis.

REFERENCES

- [1] Brazilian Ministry of Health (2011). http://portal.saude.gov.br/portal/saude/profissional/visualizar_texto.cfm?idtxt=31519
- [2] Chippaux, J.P., Goyffon, M. (2008) Epidemiology of scorpionism: a global appraisal. *Acta Tropica*, **107**, 71-9. doi:10.1016/j.actatropica.2008.05.021
- [3] Cologna, C.T., Marcussi, S., Giglio, J.R., Soares, A.M., Arantes EC. (2009) *Tityus serrulatus* Scorpion Venom and Toxins: An Overview. *Protein and Peptide Letters*, **16**, 920-932. doi:10.2174/092986609788923329
- [4] Possani, L.D., Alagón, A.C., Fletcher, P.L. Jr, Erickson, B.W. (1977) Purification and properties of mammalian toxins from the venom of Brazilian Scorpion *Tityus serrulatus* Lutz and Mello. *Archives of Biochemistry and Biophysics*, **180**(2), 394-403. doi:10.1016/0003-9861(77)90053-4
- [5] Arantes, E.C., Prado, W.A., Sampaio, S.V., Giglio, J.R. (1989) A simplified procedure for the fractionation of *Tityus serrulatus* venom: isolation and partial characterization of TsTX-IV, a new neurotoxin. *Toxicon*, **27**(8), 907-16. doi:10.1016/0041-0101(89)90102-5
- [6] Sampaio, S.V., Arantes, E.C., Prado, W.A., Riccioppo Neto, F., Giglio, J.R. (1991) Further characterization of toxins T1IV (TsTX-III) and T2IV from *Tityus serrulatus* scorpion venom. *Toxicon*, **29**(6), 663-72.
- [7] Martin-Eauclaire, M.F., Céard, B., Ribeiro, A.M., Diniz, C.R., Rochat, H., Bougis, P.E. (1994) Biochemical, pharmacological and genomic characterisation of Ts IV, an alpha-toxin from the venom of the South American scorpion *Tityus serrulatus*. *FEBS Letters*, **342**, 181-184.
- [8] Sampaio, S.V., Coutinho-Netto, J., Arantes, E.C., Marangoni, S., Oliveira, B., Giglio, J.R. (1996) Isolation of toxin TsTX-VI from *Tityus serrulatus* scorpion venom. Effects on the release of neurotransmitters from synaptosomes. *Biochemistry and molecular biology international*, **39**(4), 729-40. doi:10.1080/15216549600201811
- [9] Guatimosim, S.C., Prado, V.F., Diniz, C.R., Chávez-Olórtegui, C., Kalapothakis, E. (1999) Molecular cloning and genomic analysis of TsNTxp: an immunogenic protein from *Tityus serrulatus* scorpion venom. *Toxicon*, **37**(3), 507-17. doi:10.1016/S0041-0101(98)00187-1
- [10] Novello, J.C., Arantes, E.C., Varanda, W.A., Oliveira, B., Giglio, J.R., Marangoni, S. (1999) TsTX-IV, a short chain four-disulfide-bridged neurotoxin from *Tityus serrulatus* venom which acts on Ca²⁺-activated K⁺ channels. *Toxicon*, **37**, 651-660. doi:10.1016/S0041-0101(98)00206-2
- [11] Petricevich, V.L., Hernández Cruz, A., Coronas, F.I., Possani, L.D. (2007) Toxin gamma from *Tityus serrulatus* scorpion venom plays an essential role in immunomodulation of macrophages. *Toxicon*, **50**(5), 666-75. doi:10.1016/j.toxicon.2007.06.001
- [12] Mendes, T.M., Dias, F., Horta, C.C., Pena, I.F., Arantes, E.C., Kalapothakis, E. (2008) Effective *Tityus*

- serrulatus* anti-venom produced using the T_s1 component. *Toxicon*, **52**, 787-793. doi:10.1016/j.toxicon.2008.08.005
- 3] Verano-Braga, T., Rocha-Resende, C., Silva, D.M., Ianzer, D., Martin-Eauclaire, M.F., et al (2008) *Tityus serrulatus* Hypotensins: a new family of peptides from scorpion venom. *Biochemical and Biophysical Research Communications*, **371**, 515-520. doi:10.1016/j.bbrc.2008.04.104
 - 4] Pimenta, A.M., Stöcklin, R., Favreau, P., Bougis, P.E., Martin-Eauclaire, M.F. (2001) Moving pieces in a proteomic puzzle: mass fingerprinting of toxic fractions from the venom of *Tityus serrulatus* (Scorpiones, Buthidae). *Rapid Communications in Mass Spectrometry*, **15**, 1562-1572. doi:10.1002/rcm415
 - 5] Schwartz, E.F., Diego-Garcia, E., Rodríguez de la Vega, R.C., Possani, L.D. (2007) Transcriptome analysis of the venom gland of the Mexican scorpion *Hadrurus gertschi* (Arachnida: Scorpiones). *BMC genomics*, **8**, 119-119. doi:10.1186/1471-2164-8-119
 - 6] Kozminsky-Atias, A., Bar-Shalom, A., Mishmar, D., Zilberberg, N. (2008) Assembling an arsenal, the scorpion way. *BMC Evolutionary Biology*, **8**, 333-333. doi:10.1186/1471-2148-8-333
 - 7] D'Suze, G., Schwartz, E.F., García-Gómez, B.I., Sevcik, C., Possani, L.D. (2009) Molecular cloning and nucleotide sequence analysis of genes from a cDNA library of the scorpion *Tityus discrepans*. *Biochimie*, **91**(8), 1010-9. doi:10.1186/1471-2148-8-333
 - 8] Ma, Y., Zhao, R., He, Y., Li, S., Liu, J., Wu, Y., et al. (2009) Transcriptome analysis of the venom gland of the scorpion *Scorpiops jendeki*: implication for the evolution of the scorpion venom arsenal. *BMC genomics*, **10**, 290-290. doi:10.1186/1471-2164-10-290
 - 9] Silva, E.C., Camargos, T.S., Maranhão, A.Q., Silva-Pereira, I., Silva, L.P., Possani, L.D., et al. (2009) Cloning and characterization of cDNA sequences encoding for new venom peptides of the Brazilian scorpion *Opisthacanthus cayaporum*. *Toxicon*, **54**, 252-261. doi:10.1016/j.toxicon.2009.04.010
 - 0] Ma, Y., Zhao, Y., Zhao, R., Zhang, W., He, Y., Wu, Y., et al. (2010) Molecular diversity of toxic components from the scorpion *Heterometrus petersii* venom revealed by proteomic and transcriptome analysis. *Proteomics*, **10** (13), 2471-85. doi:10.1002/pmic.200900763
 - 1] Ruiming, Z., Yibao, M., Yawen, H., Zhiyong, D., Yingliang, W., Zhijian, C., et al. (2010) Comparative venom gland transcriptome analysis of the scorpion *Lychas mucronatus* reveals intraspecific toxic gene diversity and new venomous components. *BMC genomics*, **11**, 452-452. doi:10.1186/1471-2164-11-452
 - 2] Morgenstern, D., Rohde, B.H., King, G.F., Tal, T., Sher, D., Zlotkin, E. (2011) The tale of a resting gland: Transcriptome of a replete venom gland from the scorpion *Hottentotta judaicus*. *Toxicon*, **57**, 695-703. doi:10.1016/j.toxicon.2011.02.001
 - 3] Ma, Y., He, Y., Zhao, R., Wu, Y., Li, W., Cao, Z. (2011) Extreme diversity of scorpion venom peptides and proteins revealed by transcriptomic analysis: Implication for proteome evolution of scorpion venom arsenal. *Journal of Proteomics*, **75**, 1563-76. doi:10.1016/j.jpro.2011.11.029
 - 4] Diego-García, E., Peigneur, S., Clynen, E., Marien, T., Czech, L., Schoofs, L., et al. (2012) Molecular diversity of the telson and venom components from *Pandinus cavimanus* (Scorpionidae Latreille 1802): Transcriptome, venomics and function. *Proteomics*, **12**(2), 313-28. doi:10.1002/pmic.201100409
 - 5] Almeida, D. D., Scortecci, K.C., Kobashi, L.S., Agnez-Lima, L.F., Medeiros, S.R., Silva-Junior, A.A., et al. (2012) Profiling the resting venom gland of the scorpion *Tityus stigmurus* through a transcriptomic survey. *BMC Genomics*, **13**, 362. doi:10.1186/1471-2164-13-362
 - 6] Rendón-Anaya, M., Delaye, L., Possani, L.D., Herrera-Estrella, A. (2012) Global Transcriptome Analysis of the Scorpion *Centruroides noxius*: New Toxin Families and Evolutionary Insights from an Ancestral Scorpion Species. *PLoS One*, **7**(8), e43331. doi:10.1371/journal.pone.0043331
 - 7] Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, **252**, 1651-1656. doi:10.1126/science.2047873
 - 8] Takasuga, A., Hirotsune, S., Itoh, R., Jitohzono, A., Suzuki, H., Aso, H., et al. (2001) Establishment of a high throughput EST sequencing system using poly(A) tail-removed cDNA libraries and determination of 36 000 bovine ESTs. *Nucleic Acids Research*, **29**, e108-e108. doi:10.1093/nar/29.22.e108
 - 9] Zhang, B., Liu, Q., Yin, W., Zhang, X., Huang, Y., Luo, Y., et al. (2006) Transcriptome analysis of *Deinagkistrodon acutus* venomous gland focusing on cellular structure and functional aspects using expressed sequence tags. *BMC genomics*, **7**, 152-152. doi:10.1186/1471-2164-7-152
 - 0] Rates, B., Ferraz, K.K., Borges, M.H., Richardson, M., De Lima, M.E., Pimenta, A.M. (2008) *Tityus serrulatus* venom peptidomics: assessing venom peptide diversity. *Toxicon*, **52**(5), 611-8. doi:10.1016/j.toxicon.2008.07.010
 - 1] Kalapothakis, E., Jardim, S., Magalhães, A.C., Mendes, T.M., De Marco, L., Afonso, L.C., et al. (2001) Screening of expression libraries using ELISA: identification of immunogenic proteins from *Tityus bahiensis* and *Tityus serrulatus* venom. *Toxicon*, **39**, 679-85. doi:10.1016/S0041-0101(00)00194-X
 - 2] Ewing, B., Hillier, L., Wendl, M.C., Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, **8**:175-185. doi:10.1101/gr.8.3.175
 - 3] SeqClean [http://compbio.dfci.harvard.edu/tgi/software/]
 - 4] Perte, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., et al. (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*, **19**, 651-652. doi:10.1093/bioinformatics/btg034
 - 5] Consortium TU. (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Research*, **38**, D142-D148. doi:10.1093/nar/ekp846
 - 6] Altschul, S., Madden, T., Schäffer, A., Zhang, J., Zhang, Z., Miller, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.

- Nucleic Acids Research*, **25**, 3389-3402.
doi:10.1093/nar/25.17.3389
- [37] Conesa, A., Götz, S., García-Gómez, J., Terol, J., Talón, M., Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674-3676. doi:10.1093/bioinformatics/bti610
- [38] Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., Kanehisa, M. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, **27(1)**, 29-34. doi:10.1093/nar/27.1.29
- [39] King, G.F., Gentz, M.C., Escoubas, P., Nicholson, G.M. (2008) A rational nomenclature for naming peptide toxins from spiders and other venomous animals. *Toxicon*, **52(2)**, 264-76. doi:10.1016/j.toxicon.2008.05.020
- [40] Barbosa-Silva, A., Satagopam, V., Schneider, R., Ortega, J.M. (2008) Clustering of cognate proteins among distinct proteomes derived from multiple links to a single seed sequence. *BMC Bioinformatics*, **9**, 141-141. doi:10.1186/1471-2105-9-141
- [41] Fernandes, G.R., Barbosa, D.V.C., Prosdocimi, F., Pena, I.A., Santana-Santos, L., Coelho Junior, *et al.* (2008) A procedure to recruit members to enlarge protein family databases - the building of UECOG (UniRef-Enriched COG Database) as a model. *Genetics and Molecular Research*, **7(3)**, 910-924. doi:10.4238/vol7-3X-Meeting008
- [42] Löytynoja, A., Goldman, N. (2010) webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics*, **11**, 579-579. doi:10.1186/1471-2105-11-579
- [43] Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G. J. (2009) Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25(9)**, 1189-1191. doi:10.1093/bioinformatics/btp033
- [44] Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, **8**, 785-786. doi:10.1038/nmeth.1701
- [45] Diego-García, E., Batista, C.V., García-Gómez, B.I., Lucas, S., Candido, D.M., Gómez-Lagunas, *et al.* (2005) The Brazilian scorpion *Tityus costatus* Karsch: genes, peptides and function. *Toxicon*, **45**, 273-283. doi:10.1016/j.toxicon.2004.10.014
- [46] Harris, F., Dennison, S.R., Phoenix, D.A. (2009) Anionic antimicrobial peptides from eukaryotic organisms. *Current Protein and Peptide Science*, **10(6)**, 585-606. doi:10.2174/138920309789630589
- [47] Rodríguez de la Vega, R.C., Schwartz, E.F., Possani, L.D. (2010) Mining on scorpion venom biodiversity. *Toxicon*, **56(7)**, 1155-61. doi:10.1016/j.toxicon.2009.11.010
- [48] Jungo F., Estreicher A., Bairoch A., Bougueleret L., Xenarios I. (2010) Animal Toxins: How is Complexity Represented in Databases? *Toxins*, **2**, 262-282. doi:10.3390/toxins2020261
- [49] Kalapothakis, E., Chávez-Olortegui, C. (1997) Venom variability among several *Tityus serrulatus* specimens. *Toxicon*, **35(10)**, 1523-1529. doi:10.1016/S0041-0101(97)00017-2
- [50] Carballar-Lejarazú, R., Rodríguez, M., de la Cruz Hernández-Hernández, F., Ramos-Castañeda, J., Possani, L., Zurita-Ortega, *et al.* (2008) Recombinant scorpine: a multifunctional antimicrobial peptide with activity against different pathogens. *Cellular and Molecular Life Science*, **65**, 3081-3092. doi:10.1007/s00018-008-8250-8
- [51] Pimenta, A.M., De Lima, M.E. (2005) Small peptides, big world: biotechnological potential in neglected bioactive peptides from arthropod venoms. *Journal of Peptide Science*, **11**, 670-676. doi:10.1002/psc.701
- [52] Dai, C., Ma, Y., Zhao, Z., Zhao, R., Wang, Q., Wu, Y., *et al.* (2008) Mucroporin, the first cationic host defense peptide from the venom of *Lychas mucronatus*. *Antimicrobial Agents and Chemotherapy*, **52**, 3967-3972. doi:10.1128/AAC.00542-08
- [53] Prosdocimi, F., Bittencourt, D., da Silva, F.R., Kirst, M., Motta, P.C., Rech, E.L. (2011) Spinning gland transcriptomics from two main clades of spiders (order: Araneae) -insights on their molecular, anatomical and behavioral evolution. *PLoS One*, **6(6)**, e21634. doi:10.1371/journal.pone.0021634
- [54] Fletcher, P.L., Fletcher, M.D., Weninger, K., Anderson, T.E., Martin, B.M. (2010) Vesicle-associated membrane protein (VAMP) cleavage by a new metalloprotease from the Brazilian scorpion *Tityus serrulatus*. *The Journal of Biological Chemistry*, **285**, 7405-7416. doi:10.1074/jbc.M109.028365
- [55] Diego-García, E., Abdel-Mottaleb, Y., Schwartz, E.F., de la Vega, R.C., Tytgat, J., Possani, L.D. (2008) Cytolytic and K⁺ channel blocking activities of beta-KTx and scorpine-like peptides purified from scorpion venoms. *Cellular and Molecular Life Sciences*, **65(1)**, 187-200.