Am. J. Trop. Med. Hyg., 86(6), 2012, pp. 1005–1014
doi:10.4269/ajtmh.2012.11-0690

An Insight into the Sialotranscriptome of *Triatoma matogrossensis*, a Kissing Bug Associated with Fogo Selvagem in South America

Teresa C. F. Assumpção, Donald P. Eaton, Van M. Pham, Ivo M. B. Francischetti, Valéria Aoki, Gunter Hans-Filho, Evandro A. Rivitti, Jesus G. Valenzuela, Luis A. Diaz, and José M. C. Ribeiro*

Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, Maryland; Department of Dermatology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; Department of Dermatology, Universidade de São Paulo, Brazil; Universidade Federal de Mato Grosso do Sul, Brazil; Wildlife Conservation Society of Brazil, Campo Grande, MS, Brazil

Abstract. *Triatoma matogrossensis* is a Hemiptera that belongs to the *oliveirai* complex, a vector of Chagas' disease that feeds on vertebrate blood in all life stages. Hematophagous insects' salivary glands (SGs) produce potent pharmacologic compounds that counteract host hemostasis, including anticlotting, antiplatelet, and vasodilatory molecules. Exposure to *T. matogrossensis* was also found to be a risk factor associated with the endemic form of the autoimmune skin disease pemphigus foliaceus, which is described in the same regions where Chagas' disease is observed in Brazil. To obtain a further insight into the salivary biochemical and pharmacologic diversity of this kissing bug and to identify possible allergens that might be associated with this autoimmune disease, a cDNA library from its SGs was randomly sequenced. We present the analysis of a set of 2,230 (SG) cDNA sequences, 1,182 of which coded for proteins of a putative secretory nature.

INTRODUCTION

The genus *Triatoma* (Hemiptera:Reduviidae) comprises various species that can be rearranged into "specific complexes" according to morphologic and molecular similarities and possible hybridizations between species.[1] *Triatoma matogrossensis*, Leite and Barbosa 1953, one of nine species forming the *oliveirai* complex, occurs in central-western and southern Brazil.[2] Besides *T. matogrossensis*, the other species included in the *oliveirai* complex are *Triatoma oliveirai*,[3] *Triatoma williami*, Galvão, Souza and Lima 1965, *Triatoma guazu*,[4] *Triatoma jurbergi*, Carcavallo, Galvão and Lent 1998, *Triatoma baratai*, Carcavallo and Jurberg 2000, *Triatoma klugi*, Carcavallo, Jurberg, Lent and Galvão 2001, *Triatoma deaneorum*, Galvão, Souza, and Lima 1967, and *Triatoma vandae* sp. n.[5] This complex is named after *T. oliveirai*, which was the first species from the group to be described.[1]

The origin of this geographically limited complex may be the Pantanal ecosystem of Mato Grosso, Brazil.[6] The species of this complex occur in the sylvatic environment, mainly in rocky formations, and are distributed throughout southwestern and southern Brazil, however *T. matogrossensis* was reported invading human dwellings—displaying a trend toward domesticity—and might act as a vector of Chagas' disease.[2]

Endemic pemphigus foliaceus (PF), also known as fogo selvagem (FS) in Brazil, is a human organ-specific autoimmune disease of the skin characterized by skin blistering and pathogenic autoantibodies against desmoglein 1 (Dsg1).[7,8] Dsg1 is an epidermal desmosomal molecule and member of the cadherin family of Ca^{++}-dependent cell adhesion molecules.[9] The anti-Dsg1 autoantibodies in FS are immunoglobulin G4 (IgG4) restricted and reproduced the human disease when passively transferred into experimental animals.[10] The disease is rare, but it is endemic in certain regions of Brazil, where Chagas' disease is also endemic.[11] It has been suggested that PF may be precipitated by exposure to environmental antigens including allergens from the saliva of hematophagous insects.[12] Simuliids in particular, have been reported to be associated with FS in case-control epidemiological studies,[11,13–15] and a transcriptome analysis of *Simulium nigrimanum* has been reported, aiming at possible identification of insect allergens that might trigger PF[16]; however, other insects including *T. matogrossensis* were also found associated with PF in the Amerindian reservation of Limão Verde, Brazil, where there is a high prevalence of PF and new cases of Chagas' disease,[13,17] thus making it possible that salivary antigens from this bug could cross-react to Dsg1, thus precipitating PF.

The salivary glands (SGs) of blood-feeding arthropods show a variety of antihemostatic compounds that are injected into their host's skin to help these organisms obtain a blood meal. Like other triatomines that have been studied,[18–22] *T. matogrossensis* is most probably capable of counteracting host hemostatic responses triggered to prevent blood loss following tissue injury, such as vasoconstriction, blood coagulation, and platelet aggregation.[23] The molecular diversity of hematophagous insect saliva presents a rich field for the discovery of novel pharmacologically active compounds and for understanding evolutionary mechanisms leading to insect adaptation to this feeding habit. Previous studies describing the sialotranscriptome of hematophagous insects and ticks[24–26] have revealed that the sialomes (from the Greek sialo = saliva) of these disease vectors are more complex than expected and contain many proteins for which we cannot yet determine a function.

In this work, we present the analysis of a set of 2,230 SG complementary DNA (cDNA) sequences, 1,182 of which code for proteins of a putative secretory nature. Most salivary proteins are described as lipocalins, comprising 76% of the transcripts coding for putative secreted proteins. *Triatoma matogrossensis* is the first species of the *oliveirai* complex to have its sialome described. We expect this work will contribute new salivary transcripts that could help in the comprehension of the role of salivary molecules in host/vector interactions, in the determination of environmental antigens

*Address correspondence to José M. C. Ribeiro, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 12735 Twinbrook Parkway, Room 2E-32D, Rockville, MD 20892. E-mail: jribeiro@niaid.nih.gov

triggering PF, and help to develop immunologic markers of *T. matogrossensis* exposure.

MATERIALS AND METHODS

***Triatoma matogrossensis* and SG cDNA library construction.** *Triatoma matogrossensis* adults and nymphs were collected April 10–11, 2009, during home searches in the Terena Amerindian community of Aldeia Limão Verde, municipality of Aquidauana, Mato Grosso do Sul state, Brazil. Within 6 hr of capture, the SGs were dissected in phosphate buffered saline, immediately transferred to 50-μL of RNAlater (Ambion, Inc., Austin, TX) and refrigerated for ~7 days. After transport to the United States, the glands were frozen and shipped to the National Institutes of Health (NIH, Rockville, MD) laboratory.

Triatoma matogrossensis SG messenger RNA (mRNA) was isolated from seven SG pairs (three females, three males, and one fifth-instar nymph) using the Micro-FastTrack mRNA isolation kit (Invitrogen, San Diego, CA). The polymerase chain reaction (PCR)-based cDNA library was made following the instructions for the SMART (switching mechanism at 5′ end of RNA transcript) cDNA library construction kit (Clontech, San Diego, CA). This kit provides a method for producing high-quality, full-length cDNA libraries from nanogram quantities of polyA+ or total RNA. It uses a specially designed oligonucleotide named SMART IV in the first-strand synthesis to generate high yields of full-length, double-stranded cDNA. *Triatoma matogrossensis* SG polyA+ RNA was used for reverse transcription to cDNA using Moloney murine leukemia virus reverse transcriptase (Clontech), the SMART IV oligonucleotide, and the CDS III/3′ primer (Clontech). Trehalose was added to the reaction that was carried out at 60°C for 1 hr, and then at 42°C for 40 min. Second-strand synthesis was performed by a long-distance PCR-based protocol using the 5′ PCR primer and the CDS III/3′ primer as sense and antisense primers, respectively. These two primers also create *Sfi1A* and *B* restriction enzyme sites at the end of the cDNA. Advantage Taq polymerase mix (Clontech) was used to carry out the long-distance PCR reaction on a Perkin Elmer GeneAmp PCR system 9700 (Perkin Elmer Corp., Norwalk, CT). The PCR conditions were 95°C for 1 min; 14 cycles of 95°C for 10 sec, 68°C for 6 min. A small portion of the cDNA was analyzed on a 1.1% agarose/EtBr (0.1 μg/mL) gel to check for the quality and range of the synthesized cDNA. Double-stranded cDNA was immediately treated with proteinase K (0.8 μg/mL) at 45°C for 20 min and washed three times with water using Amicon filters with a 100-kDa cutoff (Millipore, Billerica, MA). The clean double-stranded cDNA was then digested with *SfiI* restriction enzyme at 50°C for 2 hr followed by size fractionation on a ChromaSpin-400 drip column (Clontech). The profiles of the fractions were checked on a 1.1% agarose/EtBr (0.1 μg/mL), and fractions containing cDNA were pooled into three different groups according to their size: large, medium, or small sequences. Each group was concentrated and washed three times with water using an Amicon filter with a 100-kDa cutoff. The concentrated cDNA was then ligated into a λ TriplEx2 vector (Clontech), and the resulting ligation mixture was packaged using GigaPack Gold III Plus packaging extract (Stratagene, La Jolla, CA) according to the manufacturer's instructions. The packaged library was plated by infecting log-phase XL1-Blue *Escherichia coli* cells (Clontech). The percentage of recombinant clones was determined by performing a blue-white selection screening on LB/$MgSO_4$ plates containing X-gal/isopropyl β-D-1-thiogalactopyranoside (IPTG).

Sequencing of the *T. matogrossensis* cDNA library. The *T. matogrossensis* SG cDNA library was plated on LB/$MgSO_4$ plates containing X-gal/IPTG to an average of 250 plaques per 150-mm Petri plate. Recombinant (white) plaques were randomly picked and transferred to 96-well MICROTEST U-bottom plates (BD BioSciences, San José, CA) containing 75 μL of H_2O per well. The phage suspension was either immediately used for PCR or stored at 4°C until use.

To amplify the cDNA using a PCR reaction, 4 μL of the phage sample were used as a template. The primers were sequences from the λ TriplEx2 vector and named PT2F1 (5′-AAG TACTCTAGCAATTGTGAGC-3′) and PT2R1 (5′-CTCTTCGCTATTACGCCAGCTG-3′), positioned at the 5′ and 3′ end of the cDNA insert, respectively. The reaction was carried out in MicroAmp 96-well PCR plates (Applied Biosystems, Fullerton, CA) using FastStart PCR Master (Roche Molecular Biochemicals, Indianapolis, IN) on a GeneAmp PCR system 9700 (Perkin Elmer Corp.). The PCR conditions were 1 hold of 75°C for 3 min, 1 hold of 94°C for 4 min, 33 cycles of 94°C for 1 min, 49°C for 1 min, and 72°C for 1 min 20 sec. The PCR products were used as a template for a cycle-sequencing reaction using the DTCS labeling kit (Beckman Coulter, Inc., Fullerton, CA). The primer used for sequencing (PT2F3) is upstream from the inserted cDNA and downstream from the PT2F1 primer. The sequencing reaction was performed on a Perkin Elmer 9700 thermocycler. Conditions were 1 hold of 75°C for 2 min, 1 hold of 94°C for 4 min, and 30 cycles of 96°C for 20 sec, 50°C for 20 sec, and 60°C for 4 min. After cycle sequencing the samples, a cleaning step was performed using the multiscreen 96-well plate cleaning system (Millipore). The 96-well multiscreening plate was prepared by adding a fixed amount (manufacturer's specification) of Sephadex-50 (Amersham Pharmacia Biotech, Piscataway, NJ) and 300 μL of deionized water. After partially drying the Sephadex in the multiscreen plate, the whole cycle-sequencing reaction was added to the center of each well, centrifuged at 2,500 rpm for 5 min, and the clean sample was collected on a sequencing microtiter plate (Beckman Coulter, Inc.). The plate was then dried on a Speed-Vac SC110 model with a microtiter plate holder (Savant Instruments, Inc., Holbrook, NY). The dried samples were immediately resuspended with 25 μL of formamide, and one drop of mineral oil was added to the top of each sample. Samples were either sequenced immediately on a CEQ 2000 DNA sequencing instrument (Beckman Coulter) or stored at −30°C. A total of 2,230 cDNA library clones was sequenced.

Bioinformatic tools and procedures. Expressed sequence tags (ESTs) were trimmed of primer and vector sequences, clusterized, and compared with other databases as previously described.[27] The BLAST tool,[28] CAP3 assembler,[29] ClustalW,[30] and Mega[31] software were used to compare, assemble, and align sequences and to visualize alignments. For functional annotation of the transcripts, we used the tool blastx[32] to compare nucleotide sequences with the non-redundant (NR) protein database of the National Center for Biotechnology Information (NCBI) and to the Gene Ontology (GO) database.[33] The tool rpsblast[34] was used to search for conserved protein domains in the Pfam,[35]

SMART,[36] KOG,[37] and conserved domains (CDD) databases.[38] We have also compared the transcripts with other subsets of mitochondrial and rRNA nucleotide sequences downloaded from NCBI and to several organism proteomes downloaded from NCBI (yeast), Flybase (*Drosophila melanogaster*), or ENSEMBL (*Anopheles gambiae*). Segments of the three-frame translations of the EST (as the libraries were unidirectional, we did not use six-frame translations) starting with a methionine in the first 100 predicted amino acids (aa)—or the predicted protein translation, in the case of complete coding sequences—were submitted to the SignalP server[39] to help identify translation products that could be secreted. *O*-glycosylation sites on the proteins were predicted with the program NetOGlyc (http://www.cbs.dtu.dk/services/NetOGlyc/).[40] Functional annotation of the transcripts was based on all the previous comparisons. Following inspection of all results, transcripts were classified as either secretory (S), housekeeping (H), or of unknown (U) function, with further subdivisions based on function and/or protein families. Sequence alignments were done with the ClustalX software package.[41] Phylogenetic analysis and statistical neighbor-joining bootstrap tests of the phylogenies were done with the Mega package.[31] Hyperlinked Excel (Microsoft, Redmond, WA) spreadsheets of the assembled ESTs and of the salivary protein database are supplied as additional files S1 and S2 at the journal site.

RESULTS: GENERAL DESCRIPTION OF THE SALIVARY TRANSCRIPTOME DATABASE

Description of the clusterized data set/cDNA library characteristics. A total of 2,230 sequences were used to assemble a clusterized database yielding 948 clusters of related sequences, 799 of which contained only one EST. The consensus sequence of each cluster is named either a contig (deriving from two or more sequences) or a singleton (deriving from a single sequence). In this work, we will use the denomination contig to address sequences deriving from both consensus sequences and from singletons. The 948 contigs were compared by the program blastx, blastn, or rpsblast[32] to the NR database of the NCBI, to the GO database,[33] to the CDD of the NCBI,[38] and to a custom-prepared subset of the NCBI nucleotide database containing either mitochondrial or rRNA sequences. Because the libraries are unidirectional, the three-frame translations of the dataset were also derived, and open reading frames starting with methionine and longer than 40 aa residues were submitted to the SignalP server[39] to help identify putative secreted proteins. The EST assembly, BLAST, and signal peptide results were transferred into an Excel spreadsheet for manual annotation.

Five categories of expressed genes derived from the manual annotation of the contigs (Table 1 and additional file S1). The S category contained 29% of the clusters and 53% of the sequences, with an average number of 4.3 sequences per cluster. The H category had 32% and 30% of the clusters and sequences, respectively, and an average of 2.2 sequences per cluster. Thirty-five percent of the clusters, containing 15% of all sequences, were classified as unknown (U) because no assignment for their function could be made; most of these consisted of singletons. Possible transposable elements created seven clusters, mostly singletons. We have also identified microbial and viral transcripts in our dataset. These data can be downloaded as an additional file S1 for the EST data.

Housekeeping (H) genes. The 308 gene contigs (comprising 672 ESTs) attributed to H genes expressed in the SGs of *T. matogrossensis* were further characterized into 16 groups, according to their possible function (Table 2 and additional file S2). According to an organ specialized in secreting polypeptides and as observed in previous sialotranscriptomes,[18,24,42,43] the two larger sets were associated with protein synthesis machinery (337 ESTs in 48 contigs) and with energy metabolism (21 contigs containing 28 ESTs). We have also included in this category a group of 79 ESTs that grouped into 57 contigs and represent conserved proteins of unknown function presumably associated with cellular metabolism. Other sequences with homology to housekeeping protein include those coding for ribosomal protein, cytochromes, and NADH-dehydrogenase, among other molecules.

Transcripts coding for putative secreted proteins in *Triatoma matogrossensis* salivary glands. *Possibly secreted enzymes.* Serine protease. Five contigs coding for serine proteases were detected in the sialotranscriptome of *T. matogrossensis*. A serine protease with trypsin-like activity, named triapsin, was described in *Triatoma infestans* saliva and shown to be released with ejected saliva in the active form, suggesting a role in blood feeding.[44] Serine proteases could be also involved in specific host proteolytic events that could affect clotting or the complement cascade. A trypsin-like protease with weak fibrinolytic activity was identified in the SGs of the triatominae *Panstrongylus megistus*,[45] and tabserin is a serine protease from *Tabanus yao* SGs that inhibits blood coagulation.[46] Additionally, this class of enzymes could be involved in immunity, as prophenoloxidase-activating enzymes are serine proteinases.[47]

Metalloprotease. We found two contigs and a total of five ESTs with similarity to metalloproteases, in particular the PFAM Astacin domain. This is the first time these enzymes are recognized on a triatomine sialotranscriptome. Metalloprotease coding transcripts are abundant in the SGs of ticks.[48] Members of this family were also identified in *Acyrthosiphon pisum*, a Hemiptera that feeds on plant phloem, where metalloproteases could inactivate plant protein defenses.[49] This family of proteins is an important constituent of snake venoms, where they commonly act as hemorrhagic factors degrading the extracellular matrix and preventing blood clot formation.[50] In centipede venom, the proteases are mainly metalloproteases, showing fibrinolytic and fibrinogenolytic activities,[51] as is the function of *Ixodes scapularis* metalloproteases.[52]

Inositol polyphosphate 5-phosphatase. Transcripts coding for inositol polyphosphate 5-phosphatases (IPPases) abound in triatomine and *Cimex* sialotranscriptomes such as *Rhodius prolixus*,[18] *T. infestans*,[20] *Triatoma dimidiata*,[21] and *Cimex lectularius*,[53] however were absent in *Dipetalogaster maxima*.[54] In *T. matogrossensis,* a total of 16 ESTs assembled into five contigs coding for IPPases. The IPPases dephosphorylate the 5-position of the inositol ring from phosphoinositide and inositol phosphate messenger molecules.[55] The IPPases consist of a large family of enzymes sharing a 5-phosphatase homology domain and two conserved signature motifs within this domain that function directly in catalysis.[56] The IPPases were presumed to be signal-terminating enzymes, because they inactivate second messengers; many studies have demonstrated a critical role of 5-phosphatase-hydrolyzed signaling

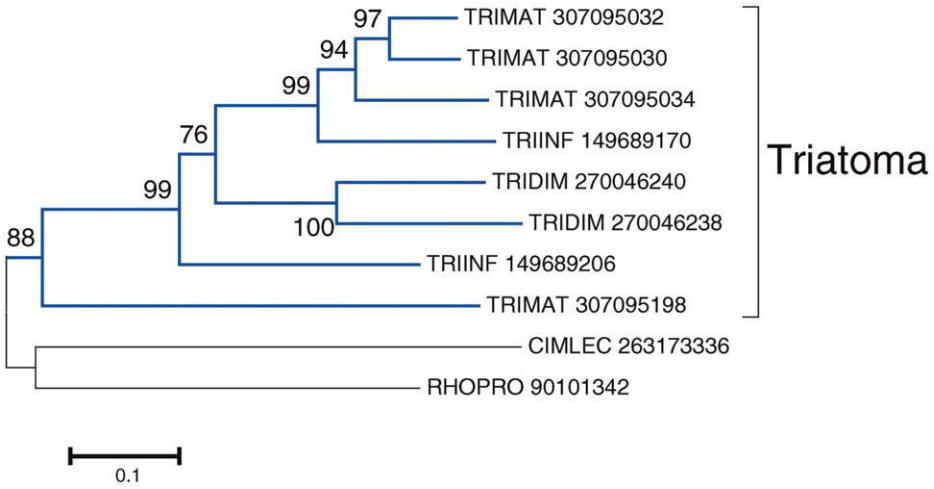

FIGURE 1. Phylogram of the salivary inositol phosphatases from bloodsucking Hemiptera. The number at the nodes indicates percentage of bootstrap support above 50% value. Sequences are represented by the first three letters of the genus name followed by the first three letters of the species name, followed by the National Center for Biotechnology Information (NCBI) accession number. The species represented are TRIMAT, *Triatoma matogrossensis*; TRIINF, *Triatoma infestans*; TRIDIM, *Triatoma dimidiata*; CIMLEC, *Cimex lectularius*; RHOPRO, *Rhodnius prolixus.* The bar at the bottom indicates 10% amino acid divergence. For other details, see text. This figure appears in color at www.ajtmh.org.

molecules in cell signaling, regulation of cellular activation and proliferation, and other cellular events.[56,57] Although shutting down the IPP pathway in their hosts could produce anti-inflammatory and immunosuppressive results, this is problematic because the pathway is intracellular, unless, as speculated before, salivary IPPases could be transported to the interior of the cells, perhaps helped by salivary hemolysins.[58] Alignment and phylogenetic analysis of the Hemipteran IPPases found in sialotranscriptomes provides for a clade with strong bootstrap support containing the sequences deriving from the *Triatoma* genus and suggests that at least three genes code for salivary IPPases in *T. matogrossensis* (Figure 1).

Apyrase. Two contigs from the *T. matogrossensis* sialotranscriptome (with a total of 27 ESTs) code for the enzymes of the 5′-nucleotidase family, which were shown to have apyrase activity in *T. infestans*.[59] Apyrases are diphosphohydrolases responsible for hydrolysis of adenosine triphosphate (ATP) and adenosine diphosphate (ADP) into adenosine monophosphate (AMP) and P_i, inhibiting platelet function by decreasing ADP-induced platelet aggregation, thus facilitating acquisition of the blood meal by hematophagous insects.[60] In triatomines, biochemical characterization of salivary apyrase activity was described for *R. prolixus*[61] and *T. infestans*.[59]

Acetylcholinesterase (AChE). Two ESTs encoding polypeptides with homology to the enzyme AChE were found in *T. matogrossensis* SGs. The AChE is a serine hydrolase essential for regulation of cholinergic neurotransmission in vertebrates and invertebrates. The main biologic role of AChE is termination of impulse transmission at cholinergic synapses by rapid hydrolysis of the neurotransmitter acetylcholine.[62] Soluble AChE-coding transcripts and protein abounds in the sialome of *C. lectularius*[53] but not of other bloodsucking Hemiptera. Its function in insect saliva remains unknown.

Protein families commonly found in triatome SG transcriptomes. Lipocalins. The most abundant group of putative secreted proteins in *T. matogrossensis* SGs is the lipocalins, corresponding to 76% of the transcripts in the S class (Table 3). This pattern has been found in all previously reported sialotranscriptomes of triatomine bugs.[19,21,22,63,64] Lipocalins are a large and heterogeneous group consisting of

TABLE 1

Functional classification of transcripts originating from the *Triatoma matogrossensis* sialotranscriptome

Number of ESTs		Class	Number of contigs ESTs/contig
Secreted	273	1182	4.3
Housekeeping	308	672	2.2
Microbial/viral	27	28	1.0
Transposable element	7	8	1.1
Unknown	333	340	1.0

Total: 948 2,230. ESTs = expressed sequence tags.

TABLE 2

Functional classification of transcripts associated with housekeeping function originating from the *Triatoma matogrossensis* sialotranscriptome

Number of ESTs		Class	Number of contigs ESTs/contig
Protein synthesis machinery	48	337	7.0
Unknown, conserved	57	79	1.4
Signal transduction	28	31	1.1
Metabolism, energy	21	28	1.3
Transporters/storage	14	27	1.9
Protein modification machinery	17	25	1.5
Oxidant metabolism/detoxification	17	25	1.5
Protein export machinery	18	22	1.2
Transcription machinery	17	17	1.0
Proteasome machinery	13	16	1.2
Cytoskeletal	12	15	1.3
Metabolism, lipid	10	12	1.2
Nuclear regulation	9	10	1.1
Transcription factor	7	8	1.1
Metabolism, carbohydrate	6	6	1.0
Metabolism, nucleotide	5	5	1.0
Metabolism, amino acid	4	4	1.0
Metabolism, intermediate	3	3	1.0
Extracellular matrix/cell adhesion	2	2	1.0

Total: 308 672. ESTs = expressed sequence tags.

TABLE 3

Functional classification of transcripts associated with secreted function originating from the *Triatoma matogrossensis* sialotranscriptome.

			Number of contigs
Number of ESTs		Class	ESTs/contig
Lipocalins	162	900	5.6
Trialysin family	8	56	7.0
Other peptide families	35	54	1.5
Triatomine hemolysin family	5	24	4.8
Kazal family	11	16	1.5
Unknown conserved and secreted	12	15	1.3
Mys family of triatomines	5	11	2.2
Antigen 5 family	3	7	2.3
Metalloprotease	3	6	2.0
Similar to accessory gland protein	1	3	3.0
Cuticle protein	3	3	1.0
Mucins	3	3	1.0
Enzymes			
Inositol phosphate phosphatase	8	34	4.3
Apyrase	4	32	8.0
Serine protease	6	14	2.3
Acetylcholinesterase	2	2	1.0
Phosphodiesterase	1	1	1.0
Fucosidase	1	1	1.0

Total: 273 1,182. ESTs = expressed sequence tags.

small, extracellular proteins sharing several common characteristics: the binding of small, principally hydrophobic molecules; binding to specific cell-surface receptors; and formation of covalent and non-covalent complexes with other soluble macromolecules.[65] An interesting feature of the lipocalins is their well-conserved three-dimensional structure despite the low similarity between random primary sequences of different proteins.[66] Although they have been classified mainly as transport proteins, members of this family may play roles not only in retinol transport but also in cryptic coloration, olfaction, pheromone transport, prostaglandin synthesis, immune response, and cell homeostasis.[67] They have been also identified as anticoagulants, antiplatelets, and vasodilatory molecules in the SGs of bloodsucking insects.[68] The sialotranscriptome of *D. maxima* and *Triatoma brasiliensis* revealed a high content of lipocalins in their SGs, comprising 93% of the transcripts coding for putative secreted protein.[19,54] Salivary lipocalin functional studies have been reported from such blood-feeding insects as *Triatoma pallidipennis*,[69,70] *R. prolixus*,[18] *T. infestans*,[20] *T. dimidiata*,[21] *D. maxima*,[71] and also in tick saliva.[72] The most abundant salivary proteins in ticks, mosquitoes, and triatomine bugs are usually binders of histamine or serotonin, thus serving as a kratagonist[58] for these biogenic amines. The alignment and phylogenetic analysis of 35 full-length lipocalins from *T. matogrossensis* assembled from two or more ESTs shows the diversity of this family in *T. matogrossensis*, where the 35 proteins are recognized into eight distinct clades representing at least 15 different genes if we consider proteins with 10% difference in aa to derive from different genes (Figure 2).

Kazal domain-containing peptides. Ten contigs, assembled from 15 ESTs, encode for polypeptides containing Kazal domains. Kazal-type serine protease inhibitors are single- or multidomain proteins that share a conserved sequence motif, a characteristic cysteine distribution pattern, and highly similar three-dimensional structures (one central α-helix and three small antiparallel β-sheets).[73,74] Several Kazal-type family members were previously described to be present in vertebrate and invertebrate animals. Thrombin inhibitors of the

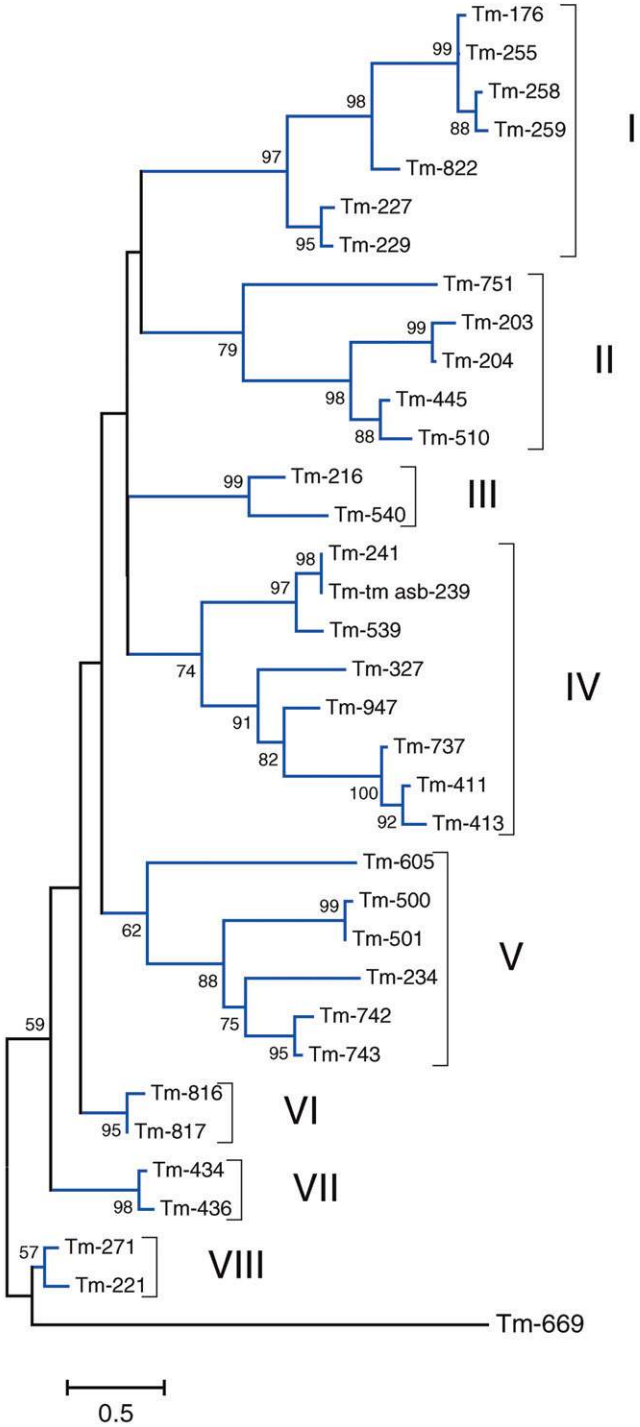

FIGURE 2. Phylogram of the salivary lipocalins of *Triatoma matogrossensis.* Roman numerals indicate clades with robust bootstrapped support. The number at the nodes indicates percentage of bootstrap support above 50% value. The bar at the bottom indicates 50% amino acid divergence. For other details, see text. This figure appears in color at www.ajtmh.org.

Kazal-type family have been found in triatomines. Rhodniin, isolated from the midgut of the triatomine *R. prolixus*, is such a Kazal-type inhibitor, composed of two domains and strongly inhibiting thrombin, the key enzyme of the blood coagulation cascade.[75] Infestin is a Kazal-type proteinase inhibitor precursor described in *T. infestans*. It seems to be similar to rhodniin

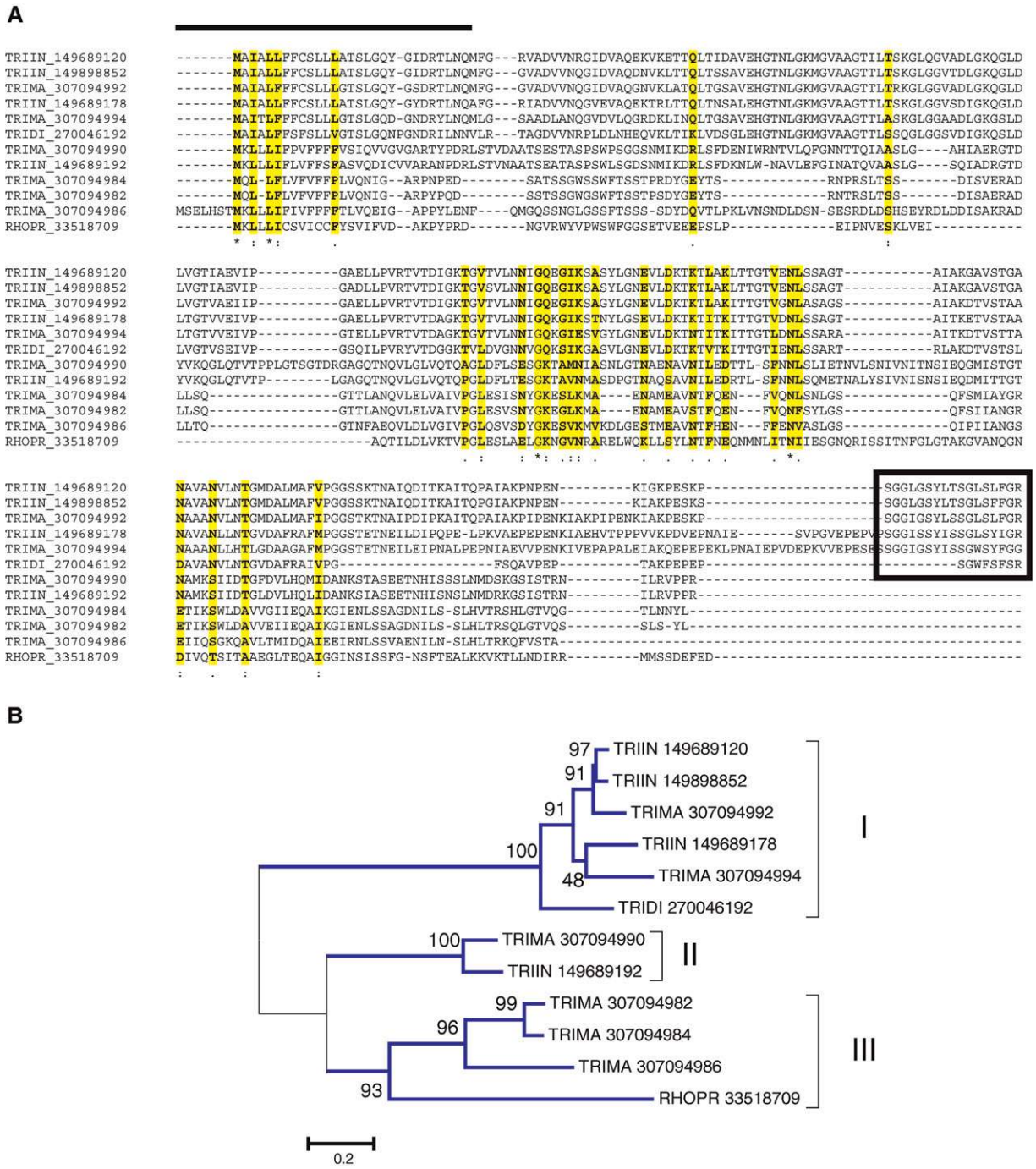

FIGURE 3. The MYS3/hemolysin family of triatomines. (**A**) Clustal alignment showing conserved residues. The bar at the top indicates the region of the signal peptide indicative of secretion. The box at the end of the sequences indicates a glycine/serine-rich region found in a subset of the sequences. Sequences are represented by the first three letters of the genus name followed by the first two letters of the species name, followed by the National Center for Biotechnology Information (NCBI) accession number. The symbols below the alignment indicates (*) identity, (:) conserved, and (.) less conserved sites. (**B**) Consensus bootstrapped phylogenetic tree based on the alignment in **A**, showing three robust clades of proteins. The numbers at the nodes indicate percentage of bootstrap support. The line at the bottom indicates 20% amino acid diversity. This figure appears in color at www.ajtmh.org.

and dipetalogastin precursors but has more Kazal-type domains.[76] Comparison of the Kazal domain proteins of *T. matogrossensis* with their best BLAST match on the NR database shows amino acid identities varying from 30–54%, all from either *T. infestans* or *T. brasiliensis*, indicating the relatively fast evolution of these proteins within these closely related species.

Trialysin. We identified six contigs (assembled from 54 ESTs that matched the protein named trialysin from *T. infestans*[77,78]) having between 38% and 88% identity at the aa level with their best match in the NR database, all from *T. infestans.* Trialysin, which belongs to a multigene family, is a pore-forming molecule present in the SGs of *T. infestans*. This 22-kDa protein is synthesized as a precursor and

processed by limited proteolysis. Trialysin possesses a basic amphipathic lytic motif in the N-terminal region containing 27 aa residues, similar to antimicrobial lytic peptides, and a protein portion that increases the lytic specificity toward eukaryotic cells. It exerts its potent cytolytic activity on a large variety of cell types, from bacteria to mammalian cells. The structural features of the protein and charge distribution, as well as the composition of the target membrane, are important determinants for protein/membrane interaction and lysis specificity.[77,78]

Antigen 5 family. This family of secreted proteins belongs to the CAP family (cysteine-rich secretory proteins; Ag5 proteins of insects; pathogenesis-related protein 1 of plants).[79] The CAP family is related to venom allergens in social wasps and ants[80,81] and to defense response in plants.[82] Snake venom proteins of the same family have been shown to contain smooth muscle-relaxing activity.[83,84] Members of this protein family are found in the sialotranscriptomes of most bloodsucking insects. The function of most Ag5 proteins in the saliva of blood-feeding arthropods is still unknown, but an antigen-5 protein from the saliva of a tabanid fly was shown to inhibit platelet aggregation through incorporation of an RGD (Arg-Gly-Asp) domain, avoiding fibrinogen binding to platelets and, thus, their aggregation.[46,85] Furthermore, a recombinant member of this family was shown to interact with the F_c portion of immunoglobulins and may play a role in evading host immune response.[86] We report antigen 5 transcripts found in *T. matogrossensis* SGs, distributed in two contigs that share 60% and 40% identity to their *T. infestans* best matches (additional file S2), indicating the divergence of this family within closely related species.

The hemolysin/MYS3 family of triatomine proteins. Two previously described unique protein families from triatomines were based on the *T. infestans* hemolysin salivary protein and the *R. prolixus* MYS3 protein.[20,64] No members of these families have been functionally characterized. Interestingly, the *T. matogrossensis* Tm-166 protein produces similarities not only to the *T. infestans* salivary hemolysin but also to *T. matogrossensis* proteins producing strong similarities to *Rhodnius* MYS3 protein, indicating that Tm-166 could be a missing link to connect these two protein families. PSI-BLAST of Tm-166 against the NR database retrieves only triatomine proteins previously identified in sialotranscriptomes, including the *Rhodnius* MYS3 protein (additional file S3). Alignment of these proteins (Figure 3A) indicates only three identical amino acid positions and 41 conserved sites on a total of 240 sites. All family members have a signal peptide indicative of secretion. Members of Clade I have a unique carboxyterminal region rich in Gly and Ser, marked with a box in Figure 3B. This Ser-rich region is not indicated as possibly glycosylated by the NetOglyc server. The phylogenetic tree based on the alignment (Figure 3B) indicates three robust clades, including one containing the *R. prolixus* protein and three *T. matogrossensis* proteins (marked as clade III in Figure 3B). The sialotranscriptome of *T. matogrossensis* thus produced a link that allowed joining two unique families of triatomine proteins.

Deorphanized Hemiptera protein family. Two ESTs assembled into Tm-507 that produces 85% identity to a protein previously reported from the *T. infestans* sialotranscriptome (National Center for Biotechnology Information [NCBI] accession gi|149689042).[63] This protein in turn produces 30% identity to gb|ACY69900.1|, a protein previously identified in the sialotranscriptome of the bed bug *C. lectularius*.[53] Tm-507 thus helped to consolidate a unique protein family found in sialotranscriptomes of bloodsucking Hemiptera.

Secreted proteins of conserved families. Nineteen predicted secreted proteins produce significant matches to non-Hemipteran sequences deposited in the NR database and may represent housekeeping proteins or proteins co-opted by *T. matogrossensis* to be expressed in its saliva. Among these, three cuticular proteins were found that may represent structures associated with the salivary ducts. Tm-485, assembled from three ESTs, is 47% identical to a *Gryllus* accessory gland protein and may represent an antimicrobial peptide.[87] Tm-138 is 57% identical to insect proteins having the pheromone binding domain.

Putative secreted proteins of orphan status. Additional file S2 shows 21 other putative secreted proteins, including a possible mucin fragment with 29 predicted galactosylation sites. These putative secreted polypeptides have no significant matches or poor matches to bacterial proteins in the NR database. Future transcriptomes from Hemiptera may deorphanize these proteins.

Possible similarity of *T. matogrossensis* proteins to desmoglein. When the deducted proteins from the *T. matogrossensis* sialotranscriptome were compared with the human desmoglein type 1 sequence (GenBank accession X56654.1) and its five cadherin repeats, some triatomine sequences revealed (additional file S2, worksheet named Desmog) relatively large stretches of similarity but with a relatively low degree of identity, such as Tm-166, a protein belonging to the hemolysin superfamily and having a 165-aa segment producing 21% identity/39% similarity, or Tm-255, a lipocalin with a stretch of 80 aa producing 41% aa similarity to desmoglein 1. Several salivary proteins produce matches of limited similarity to the cadherin repeats, such as the lipocalin Tm-364 (54% similarity on 50 aa) or the lipocalin Tm-264 (56% similarity to the fifth repeat on a stretch of 25 aa). Ultimately, western blots using two-dimensional gels probed with antibodies to desmoglein followed by mass spectrometric determination of the tryptic digest of the reacting gel proteins may identify insect salivary proteins that may be associated with triggering fogo selvagem. The protein database reported here will be useful in this endeavor.

CONCLUSIONS

In an attempt to improve our understanding of the variety of proteins and transcripts expressed in *T. matogrossensis* SGs, we constructed a cDNA library using mRNA from this same tissue. We described the set of cDNAs present in the SGs of *T. matogrossensis*. The derived sequences confirm the multigene status of many of the salivary products such as the lipocalins, IPPases, antigen-5, serine proteases, Kazal domain-containing proteins, and the MYS3/hemolysin family. For a discussion of the role of gene duplication in the evolution of sialomes, see Mans and others.[88] The *T. matogrossensis* sialotranscriptome also allowed the consolidation of the MYS3/hemolysin family by the discovery of an intermediate sequence between those two previously independent families exclusive of triatomines. Other orphan *Triatoma* proteins were deorphanized.

The publicly available sequences from *T. matogrossensis* should facilitate identification of environmental antigens

triggering fogo selvagem and development of immunological markers of exposure to this insect, as was done for *T. infestans*,[89,90] where a recombinant protein has been used as a marker of low-level exposure to triatomines.[91] This recombinant *T. infestans* protein was based on the sequence gi149689094, coding for a protein of unknown function, having 90% identity to Tm-727. The fast divergence of salivary proteins may lead to identification of novel species-specific exposure markers such as those grouped in the "secreted orphan proteins" category or to the many lipocalins sharing < 50% identity to their best closest *T. infestans* homolog.

Received November 3, 2011. Accepted for publication January 6, 2012.

Note: Supplemental tables appear at www.ajtmh.org.

Acknowledgments: We thank NIAID intramural editor Brenda Rae Marshall for assistance.

Financial support: This work was supported in part by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health, and by extramural NIAID grants RO1-AR32599 and RO1-AR32081 to Luis A. Diaz.

Authors' addresses: Teresa C. F. Assumpção, Van M.Pham, Ivo M. B. Francischetti, Jesus G. Valenzuela, and José M. C. Ribeiro, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, MD, E-mails: assumpcaot@niaid.nih.gov, vpham@niaid.nih.gov, ifrancischetti@niaid.nih.gov, jvalenzuela@niaid.nih.gov, and jribeiro@niaid.nih.gov. Donald P. Eaton, Wildlife Conservation Society, Campo Grande, MS, Brazil, E-mail: ksadeaton@yahoo.com. Valéria Aoki and Evandro A. Rivitti, Department of Dermatology, Universidade de São Paulo, Brazil, E-mails: valeria.aoki@gmail.com and evandro.rivitti@ gmail.com. Gunter Hans-Filho, Department of Dermatology, Universidade Federal de Mato Grosso do Sul, Campo Grande, Brazil, E-mail: ghansfilho@hotmail.com. Luis A. Diaz, Department of Dermatology, University of North Carolina at Chapel Hill, Chapel Hill, NC, E-mail: luis_diaz@med.unc.edu.

Reprint requests: José M. C. Ribeiro, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 12735 Twinbrook Parkway, Room 2E-32D, Rockville, MD, E-mail: jribeiro@niaid.nih.gov.

REFERENCES

1. Carcavallo RU, Jurberg J, Lent H, Galvao C, Steindel M, Carvalho Pinto CJ, 2001. A new species of the *oliveirai* complex (new designation for matogrossensis complex) from the State of Rio Grande do Sul, Brazil. *Mem Inst Oswaldo Cruz 96:* 71–79.
2. Noireau F, dos Santos SM, Gumiel M, Dujardin JP, dos Santos Soares M, Carcavallo RU, Galvao C, Jurberg J, 2002. Phylogenetic relationships within the *oliveirai* complex (Hemiptera: Reduviidae:Triatominae). *Infect Genet Evol 2:* 11–17.
3. Neiva A, Pinto C, Lent H, 1939. Notas sobre triatomídeos do Rio Grande do Sul e descrição de uma nova espécie. *Mem Inst Oswaldo Cruz 34:* 607–610.
4. Lent H, Wygodzinsky P, 1979. Revision of the Triatominae (Hemiptera:Reduviidae) and their significance as vectors of Chagas disease. *Bull Am Mus Nat Hist 163:* 127–520.
5. Carcavallo RU, Jurberg J, Rocha Dda S, Galvao C, Noireau F, Lent H, 2002. *Triatoma vandae* sp.n. of the *oliveirai* complex from the State of Mato Grosso, Brazil (Hemiptera: Reduviidae: Triatominae). *Mem Inst Oswaldo Cruz 97:* 649–654.
6. Carcavallo RU, Jurberg J, Lent H, Noireau F, Galvão C, 2000. Phylogeny of the Triatominae (Hemiptera: Reduviidae). *Entomolgía y Vectores 7:* 1–99.
7. Culton DA, Qian Y, Li N, Rubenstein D, Aoki V, Filhio GH, Rivitti EA, Diaz LA, 2008. Advances in pemphigus and its endemic pemphigus foliaceus (Fogo Selvagem) phenotype: a paradigm of human autoimmunity. *J Autoimmun 31:* 311–324.
8. Stanley JR, Klaus-Kovtun V, Sampaio SA, 1986. Antigenic specificity of fogo selvagem autoantibodies is similar to North American pemphigus foliaceus and distinct from pemphigus vulgaris autoantibodies. *J Invest Dermatol 87:* 197–201.
9. Buxton RS, Cowin P, Franke WW, Garrod DR, Green KJ, King IA, Koch PJ, Magee AI, Rees DA, Stanley JR, Steinberg M, 1993. Nomenclature of the desmosomal cadherins. *J Cell Biol 121:* 481–483.
10. Rock B, Martins CR, Theofilopoulos AN, Balderas RS, Anhalt GJ, Labib RS, Futamura S, Rivitti EA, Diaz LA, 1989. The pathogenic effect of IgG4 autoantibodies in endemic pemphigus foliaceus (fogo selvagem). *N Engl J Med 320:* 1463–1469.
11. Diaz LA, Sampaio SA, Rivitti EA, Martins CR, Cunha PR, Lombardi C, Almeida FA, Castro RM, Macca ML, Lavrado C, Filho GH, Borges PC, Chaul A, Minelli L, Empinotti JC, Friedman H, Campbell I, Labib RS, Anhalt FJ, 1989. Endemic pemphigus foliaceus (Fogo Selvagem): II. Current and historic epidemiologic studies. *J Invest Dermatol 92:* 4–12.
12. Diaz LA, Arteaga LA, Hilario-Vargas J, Valenzuela JG, Li N, Warren S, Aoki V, Hans-Filho G, Eaton D, dos Santos V, Nutman TB, de Mayolo AA, Qaqish BF, Sampaio SA, Rivitti EA, 2004. Anti-desmoglein-1 antibodies in onchocerciasis, leishmaniasis and Chagas disease suggest a possible etiological link to Fogo selvagem. *J Invest Dermatol 123:* 1045–1051.
13. Aoki V, Millikan RC, Rivitti EA, Hans-Filho G, Eaton DP, Warren SJ, Li N, Hilario-Vargas J, Hoffmann RG, Diaz LA, 2004. Environmental risk factors in endemic pemphigus foliaceus (fogo selvagem). *J Investig Dermatol Symp Proc 9:* 34–40.
14. Eaton DP, Diaz LA, Hans-Filho G, Santos VD, Aoki V, Friedman H, Rivitti EA, Sampaio SA, Gottlieb MS, Giudice GJ, Lopez A, Cupp EW, 1998. Comparison of black fly species (Diptera: Simuliidae) on an Amerindian reservation with a high prevalence of fogo selvagem to neighboring disease-free sites in the State of Mato Grosso do Sul, Brazil. The Cooperative Group on Fogo Selvagem Research. *J Med Entomol 35:* 120–131.
15. Lombardi C, Borges PC, Chaul A, Sampaio SA, Rivitti EA, Friedman H, Martins CR, Sanches Junior JA, Cunha PR, Hoffmann RG, Diaz LA, Cooperative Group for Fogo Selvagem Research, 1992. Environmental risk factors in endemic pemphigus foliaceus (Fogo selvagem). *J Invest Dermatol 98:* 847–850.
16. Ribeiro JM, Valenzuela JG, Pham VM, Kleeman L, Barbian KD, Favreau AJ, Eaton DP, Aoki V, Hans-Filho G, Rivitti EA, Diaz LA, 2010. An insight into the sialotranscriptome of *Simulium nigrimanum*, a black fly associated with fogo selvagem in South America. *Am J Trop Med Hyg 82:* 1060–1075.
17. Hans-Filho G, dos Santos V, Katayama JH, Aoki V, Rivitti EA, Sampaio SA, Friedman H, Moraes JR, Moraes ME, Eaton DP, Lopez AL, Hoffman RG, Fairley JA, Giudice GJ, Diaz LA, 1996. An active focus of high prevalence of fogo selvagem on an Amerindian reservation in Brazil. Cooperative Group on Fogo Selvagem Research. *J Invest Dermatol 107:* 68–75.
18. Ribeiro JM, Andersen J, Silva-Neto MA, Pham VM, Garfield MK, Valenzuela JG, 2004. Exploring the sialome of the blood-sucking bug *Rhodnius prolixus*. *Insect Biochem Mol Biol 34:* 61–79.
19. Santos A, Ribeiro JM, Lehane MJ, Gontijo NF, Veloso AB, Sant'Anna MR, Nascimento Araujo R, Grisard EC, Pereira MH, 2007. The sialotranscriptome of the blood-sucking bug *Triatoma brasiliensis* (Hemiptera, Triatominae). *Insect Biochem Mol Biol 37:* 702–712.

20. Assumpção TC, Francischetti IM, Andersen JF, Schwarz A, Santana JM, Ribeiro JM, 2008. An insight into the sialome of the blood-sucking bug *Triatoma infestans*, a vector of Chagas' disease. *Insect Biochem Mol Biol 38:* 213–232.
21. Kato H, Jochim RC, Gomez EA, Sakoda R, Iwata H, Valenzuela JG, Hashiguchi Y, 2010. A repertoire of the dominant transcripts from the salivary glands of the blood-sucking bug, *Triatoma dimidiata*, a vector of Chagas disease. *Infect Genet Evol 10:* 184–191.
22. Assumpcao TC, Charneau S, Santiago PB, Francischetti IM, Meng Z, Araujo CN, Pham VM, Queiroz RM, de Castro CN, Ricart CA, Santana JM, Ribeiro JM, 2011. Insight into the salivary transcriptome and proteome of *Dipetalogaster maxima*. *J Proteome Res 10:* 669–679.
23. Ribeiro JM, 1995. Blood-feeding arthropods: live syringes or invertebrate pharmacologists? *Infect Agents Dis 4:* 143–152.
24. Francischetti IM, Valenzuela JG, Pham VM, Garfield MK, Ribeiro JM, 2002. Toward a catalog for the transcripts and proteins (sialome) from the salivary gland of the malaria vector *Anopheles gambiae*. *J Exp Biol 205:* 2429–2451.
25. Valenzuela JG, Pham VM, Garfield MK, Francischetti IM, Ribeiro JM, 2002. Toward a description of the sialome of the adult female mosquito *Aedes aegypti*. *Insect Biochem Mol Biol 32:* 1101–1122.
26. Valenzuela JG, Francischetti IM, Pham VM, Garfield MK, Mather TN, Ribeiro JM, 2002. Exploring the sialome of the tick *Ixodes scapularis*. *J Exp Biol 205:* 2843–2864.
27. Valenzuela JG, Francischetti IM, Pham VM, Garfield MK, Ribeiro JM, 2003. Exploring the salivary gland transcriptome and proteome of the *Anopheles stephensi* mosquito. *Insect Biochem Mol Biol 33:* 717–732.
28. Altschul SF, Gish W, 1996. Local alignment statistics. *Methods Enzymol 266:* 460–480.
29. Huang X, Madan A, 1999. CAP3: a DNA sequence assembly program. *Genome Res 9:* 868–877.
30. Thompson JD, Higgins DG, Gibson TJ, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22:* 4673–4680.
31. Kumar S, Tamura K, Nei M, 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform 5:* 150–163.
32. Altschul SF, Madden TL, Schaffer 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.
33. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G, 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet 25:* 25–29.
34. Schaffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF, 2001. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res 29:* 2994–3005.
35. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL, 2000. The Pfam protein families database. *Nucleic Acids Res 28:* 263–266.
36. Letunic I, Goodstadt L, Dickens NJ, Doerks T, Schultz J, Mott R, Ciccarelli F, Copley RR, Ponting CP, Bork P, 2002. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res 30:* 242–244.
37. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA, 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics 4:* 41.
38. Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH, 2002. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res 30:* 281–283.
39. Nielsen H, Engelbrecht J, Brunak S, von Heijne G, 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng 10:* 1–6.
40. Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S, 1998. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj J 15:* 115–130.
41. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG, 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res 25:* 4876–4882.
42. Ribeiro JM, Charlab R, Pham VM, Garfield M, Valenzuela JG, 2004. An insight into the salivary transcriptome and proteome of the adult female mosquito *Culex pipiens quinquefasciatus*. *Insect Biochem Mol Biol 34:* 543–563.
43. Calvo E, Dao A, Pham VM, Ribeiro JM, 2007. An insight into the sialome of *Anopheles funestus* reveals an emerging pattern in anopheline salivary protein families. *Insect Biochem Mol Biol 37:* 164–175.
44. Amino R, Tanaka AS, Schenkman S, 2001. Triapsin, an unusual activatable serine protease from the saliva of the hematophagous vector of Chagas' disease *Triatoma infestans* (Hemiptera: Reduviidae). *Insect Biochem Mol Biol 31:* 465–472.
45. Meiser CK, Piechura H, Meyer HE, Warscheid B, Schaub GA, Balczun C, 2010. A salivary serine protease of the haematophagous reduviid *Panstrongylus megistus*: sequence characterization, expression pattern and characterization of proteolytic activity. *Insect Mol Biol 19:* 409–421.
46. Xu X, Yang H, Ma D, Wu J, Wang Y, Song Y, Wang X, Lu Y, Yang J, Lai R, 2008. Toward an understanding of the molecular mechanism for successful blood feeding by coupling proteomics analysis with pharmacological testing of horsefly salivary glands. *Mol Cell Proteomics 7:* 582–590.
47. Söderhäll K, Cerenius L, 1998. Role of the prophenoloxidase-activating system in invertebrate immunity. *Curr Opin Immunol 10:* 23–28.
48. Francischetti IM, Sá-Nunes A, Mans BJ, Santos IM, Ribeiro JM, 2009. The role of saliva in tick feeding. *Front Biosci 14:* 2051–2088.
49. Carolan JC, Fitzroy CI, Ashton PD, Douglas AE, Wilkinson TL, 2009. The secreted salivary proteome of the pea aphid *Acyrthosiphon pisum* characterized by mass spectrometry. *Proteomics 9:* 2457–2467.
50. Gutiérrez JM, Rucavado A, 2000. Snake venom metalloproteinases: their role in the pathogenesis of local tissue damage. *Biochimie 82:* 841–850.
51. Undheim EA, King GF, 2011. On the venom system of centipedes (Chilopoda), a neglected group of venomous animals. *Toxicon 57:* 512–524.
52. Francischetti IM, Mather TN, Ribeiro JM, 2003. Cloning of a salivary gland metalloprotease and characterization of gelatinase and fibrin(ogen)lytic activities in the saliva of the Lyme disease tick vector *Ixodes scapularis*. *Biochem Biophys Res Commun 305:* 869–875.
53. Francischetti IM, Calvo E, Andersen JF, Pham VM, Favreau AJ, Barbian KD, Romero A, Valenzuela JG, Ribeiro JM, 2010. Insight into the Sialome of the Bed Bug, *Cimex lectularius*. *J Proteome Res 9:* 3820–3831.
54. Assumpção TC, Charneau S, Santiago PB, Francischetti IM, Meng Z, Araújo CN, Pham VM, Queiroz RM, de Castro CN, Ricart CA, Santana JM, Ribeiro JM, 2011. Insight into the salivary transcriptome and proteome of *Dipetalogaster maxima*. *J Proteome Res 10:* 669–679.
55. Erneux C, Govaerts C, Communi D, Pesesse X, 1998. The diversity and possible functions of the inositol polyphosphate 5-phosphatases. *Biochim Biophys Acta 1436:* 185–199.
56. Zhang X, Majerus PW, 1998. Phosphatidylinositol signalling reactions. *Semin Cell Dev Biol 9:* 153–160.
57. Mitchell CA, Gurung R, Kong AM, Dyson JM, Tan A, Ooms LM, 2002. Inositol polyphosphate 5-phosphatases: lipid phosphatases with flair. *IUBMB Life 53:* 25–36.
58. Ribeiro JM, Arca B, 2009. From sialomes to the sialoverse: an insight into the salivary potion of blood feeding insects. *Adv Insect Physiol 37:* 59–118.
59. Faudry E, Lozzi SP, Santana JM, D'Souza-Ault M, Kieffer S, Felix CR, Ricart CA, Sousa MV, Vernet T, Teixeira AR, 2004. *Triatoma infestans* apyrases belong to the 5′-nucleotidase family. *J Biol Chem 279:* 19607–19613.

60. Ribeiro JM, Francischetti IM, 2003. Role of arthropod saliva in blood feeding: sialome and post-sialome perspectives. *Annu Rev Entomol 48:* 73–88.
61. Sarkis JJ, Guimaraes JA, Ribeiro JM, 1986. Salivary apyrase of *Rhodnius prolixus*. Kinetics and purification. *Biochem J 233:* 885–891.
62. Dvir H, Silman I, Harel M, Rosenberry TL, Sussman JL, 2010. Acetylcholinesterase: from 3D structure to function. *Chem Biol Interact 187:* 10–22.
63. Assumpcao TC, Francischetti IM, Andersen JF, Schwarz A, Santana JM, Ribeiro JM, 2008. An insight into the sialome of the blood-sucking bug *Triatoma infestans,* a vector of Chagas' disease. *Insect Biochem Mol Biol 38:* 213–232.
64. Ribeiro JM, Andersen J, Silva-Neto MA, Pham VM, Garfield MK, Valenzuela JG, 2004. Exploring the sialome of the blood-sucking bug *Rhodnius prolixus*. *Insect Biochem Mol Biol 34:* 61–79.
65. Flower DR, North AC, Sansom CE, 2000. The lipocalin protein family: structural and sequence overview. *Biochim Biophys Acta 1482:* 9–24.
66. Flower DR, 1995. Multiple molecular recognition properties of the lipocalin protein family. *J Mol Recognit 8:* 185–195.
67. Flower DR, 1996. The lipocalin protein family: structure and function. *Biochem J 318:* 1–14.
68. Andersen JF, Gudderra NP, Francischetti IM, Ribeiro JM, 2005. The role of salivary lipocalins in blood feeding by *Rhodnius prolixus*. *Arch Insect Biochem Physiol 58:* 97–105.
69. Noeske-Jungblut C, Kratzschmar J, Haendler B, Alagon A, Possani L, Verhallen P, Donner P, Schleuning WD, 1994. An inhibitor of collagen-induced platelet aggregation from the saliva of *Triatoma pallidipennis*. *J Biol Chem 269:* 5050–5053.
70. Noeske-Jungblut C, Haendler B, Donner P, Alagon A, Possani L, Schleuning WD, 1995. Triabin, a highly potent exosite inhibitor of thrombin. *J Biol Chem 270:* 28629–28634.
71. Assumpção TC, Alvarenga PH, Ribeiro JM, Andersen JF, Francischetti IM, 2010. Dipetalodipin, a novel multifunctional salivary lipocalin that inhibits platelet aggregation, vasoconstriction, and angiogenesis through unique binding specificity for TXA2, PGF2alpha, and 15(S)-HETE. *J Biol Chem 285:* 39001–39012.
72. Paesen GC, Adams PL, Nuttall PA, Stuart DL, 2000. Tick histamine-binding proteins: lipocalins with a second binding cavity. *Biochim Biophys Acta 1482:* 92–101.
73. Schlott B, Wohnert J, Icke C, Hartmann M, Ramachandran R, Guhrs KH, Glusa E, Flemming J, Gorlach M, Grosse F, Ohlenschlager O, 2002. Interaction of Kazal-type inhibitor domains with serine proteinases: biochemical and structural studies. *J Mol Biol 318:* 533–546.
74. Stubbs MT, Morenweiser R, Sturzebecher J, Bauer M, Bode W, Huber R, Piechottka GP, Matschiner G, Sommerhoff CP, Fritz H, Auerswald EA, 1997. The three-dimensional structure of recombinant leech-derived tryptase inhibitor in complex with trypsin. Implications for the structure of human mast cell tryptase and its inhibition. *J Biol Chem 272:* 19931–19937.
75. van de Locht A, Lamba D, Bauer M, Huber R, Friedrich T, Kroger B, Hoffken W, Bode W, 1995. Two heads are better than one: crystal structure of the insect derived double domain Kazal inhibitor rhodniin in complex with thrombin. *EMBO J 14:* 5149–5157.
76. Lovato DV, Nicolau de Campos IT, Amino R, Tanaka AS, 2006. The full-length cDNA of anticoagulant protein infestin revealed a novel releasable Kazal domain, a neutrophil elastase inhibitor lacking anticoagulant activity. *Biochimie 88:* 673–681.
77. Amino R, Martins RM, Procopio J, Hirata IY, Juliano MA, Schenkman S, 2002. Trialysin, a novel pore-forming protein from saliva of hematophagous insects activated by limited proteolysis. *J Biol Chem 277:* 6207–6213.
78. Martins RM, Sforca ML, Amino R, Juliano MA, Oyama S Jr, Juliano L, Pertinhez TA, Spisni A, Schenkman S, 2006. Lytic activity and structural differences of amphipathic peptides derived from trialysin. *Biochemistry 45:* 1765–1774.
79. Megraw T, Kaufman TC, Kovalick GE, 1998. Sequence and expression of *Drosophila* Antigen 5-related 2, a new member of the CAP gene family. *Gene 222:* 297–304.
80. Hoffman DR, 1993. Allergens in Hymenoptera venom. XXV: the amino acid sequences of antigen 5 molecules and the structural basis of antigenic cross-reactivity. *J Allergy Clin Immunol 92:* 707–716.
81. King TP, Spangfort MD, 2000. Structure and biology of stinging insect venom allergens. *Int Arch Allergy Immunol 123:* 99–106.
82. Stintzi A, Heitz T, Prasad V, Wiedemann-Merdinoglu S, Kauffmann S, Geoffroy P, Legrand M, Fritig B, 1993. Plant pathogenesis-related proteins and their role in defense against pathogens. *Biochimie 75:* 687–706.
83. Yamazaki Y, Koike H, Sugiyama Y, Motoyoshi K, Wada T, Hishinuma S, Mita M, Morita T, 2002. Cloning and characterization of novel snake venom proteins that block smooth muscle contraction. *Eur J Biochem 269:* 2708–2715.
84. Yamazaki Y, Morita T, 2004. Structure and function of snake venom cysteine-rich secretory proteins. *Toxicon 44:* 227–231.
85. Peter K, Schwarz M, Ylanne J, Kohler B, Moser M, Nordt T, Salbach P, Kubler W, Bode C, 1998. Induction of fibrinogen binding and platelet aggregation as a potential intrinsic property of various glycoprotein IIb/IIIa (alphaIIbbeta3) inhibitors. *Blood 92:* 3240–3249.
86. Ameri M, Wang X, Wilkerson MJ, Kanost MR, Broce AB, 2008. An immunoglobulin binding protein (antigen 5) of the stable fly (Diptera: Muscidae) salivary gland stimulates bovine immune responses. *J Med Entomol 45:* 94–101.
87. Otti O, Naylor RA, Siva-Jothy MT, Reinhardt K, 2009. Bacteriolytic activity in the ejaculate of an insect. *Am Nat 174:* 292–295.
88. Mans BJ, Louw AI, Neitz AW, 2002. Evolution of hematophagy in ticks: common origins for blood coagulation and platelet aggregation inhibitors from soft ticks of the genus *Ornithodoros*. *Mol Biol Evol 19:* 1695–1705.
89. Schwarz A, Medrano-Mercado N, Billingsley PF, Schaub GA, Sternberg JM, 2010. IgM-antibody responses of chickens to salivary antigens of *Triatoma infestans* as early biomarkers for low-level infestation of triatomines. *Int J Parasitol 40:* 1295–1302.
90. Schwarz A, Sternberg JM, Johnston V, Medrano-Mercado N, Anderson JM, Hume JC, Valenzuela JG, Schaub GA, Billingsley PF, 2009. Antibody responses of domestic animals to salivary antigens of *Triatomainfestans* as biomarkers for low-level infestation of triatomines. *Int J Parasitol 39:* 1021–1029.
91. Schwarz A, Helling S, Collin N, Teixeira CR, Medrano-Mercado N, Hume JC, Assumpcao TC, Marcus K, Stephan C, Meyer HE, Ribeiro JM, Billingsley PF, Valenzuela JG, Sternberg JM, Schaub GA, 2009. Immunogenic salivary proteins of *Triatoma infestans:* development of a recombinant antigen for the detection of low-level infestation of triatomines. *PLoS Negl Trop Dis 3:* e532.