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. Díaz, 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 anticoagulation, antplatelet, 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 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 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 deneeorum*, 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 PF16; 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 ticks24–26 have revealed that the saliomes (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.
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 RNA later (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 poly-merase 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 SfII/A 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 restriction enzymes, fractionated, and compared with other databases as previously described.27 The BLAST tool,28 CAP3 assembler,29 ClustalW,30 and Mega31 software were used to compare, assemble, and align sequences and to visualize alignments. For functional annotation of the transcripts, we used the tool blastx32 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 rpsblast34 was used to search for conserved protein domains in the Pfam.35

**Sequencing of the *T. matogrossensis* cDNA library.** The *T. matogrossensis* SG cDNA library was plated on LB/MgSO₄ 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 MICROTHER U-bottom plates (BD Biosciences, San José, CA) containing 75 μL of H2O 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 of 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 multiscreen 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 microtitre 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, clustered, and compared with other databases as previously described.27 The BLAST tool,28 CAP3 assembler,29 ClustalW,30 and Mega31 software were used to compare, assemble, and align sequences and to visualize alignments. For functional annotation of the transcripts, we used the tool blastx32 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 rpsblast34 was used to search for conserved protein domains in the Pfam.35
RESULTS: GENERAL DESCRIPTION OF THE SALIVARY TRANSCRIPTOME DATABASE

Description of the clustered data set/cDNA library characteristics. A total of 2,230 sequences were used to assemble a clustered 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 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). Functional annotation of the transcripts was based on 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. Phylogenetic analysis and statistical neighbor-joining bootstrap tests of the phylogenies were done with the Mega package. 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.

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 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. 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 triatomine Panstrongylus megistus, and tabeserin is a serine protease from Tabanus yao SGs that inhibits blood coagulation. Additionally, this class of enzymes could be involved in immunity, as prophenoloxidase-activating enzymes are serine proteases.

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. Members of this family were also identified in Acrystosiphon pisum, a Hemiptera that feeds on plant phloem, where metalloproteases could inactivate plant protein defenses. 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. In centipede venom, the proteases are mainly metalloproteases, showing fibrinolytic and fibrinogenolytic activities, as is the function of Ixodes scapularis metalloproteases.

Inositol polyphosphate 5-phosphatase. Transcripts coding for inositol polyphosphate 5-phosphatases (IPPases) abound in triatome and Cimex sialotranscriptomes such as Rhodius proliris, T. infestans, Triatoma dimidiata, and Cimex lectularius, however were absent in Dipetalogaster maxima. 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. 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. 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
molecules in cell signaling, regulation of cellular activation and proliferation, and other cellular events. 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.

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*. Apyrases are diphosphohydrolases responsible for hydrolysis of adenosine triphosphate (ATP) and adenosine diphosphate (ADP) into adenosine monophosphate (AMP) and Pi, inhibiting platelet function by decreasing ADP-induced platelet aggregation, thus facilitating acquisition of the blood meal by hematophagous insects. In triatomines, biochemical characterization of salivary apyrase activity was described for *R. prolixus* and *T. infestans*.

**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. Soluble AChE-coding transcripts and protein abunds in the saliome of *C. lectularius* 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. Lipocalins are a large and heterogeneous group consisting of

---

**Table 1**

<table>
<thead>
<tr>
<th>Functional classification of transcripts originating from the <em>Triatoma matogrossensis</em> sialotranscriptome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of ESTs</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Secreted</td>
</tr>
<tr>
<td>Housekeeping</td>
</tr>
<tr>
<td>Microbial/viral</td>
</tr>
<tr>
<td>Transposable</td>
</tr>
<tr>
<td>Transposable</td>
</tr>
</tbody>
</table>

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

---

**Table 2**

<table>
<thead>
<tr>
<th>Functional classification of transcripts associated with housekeeping function originating from the <em>Triatoma matogrossensis</em> sialotranscriptome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of ESTs</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Protein synthesis machinery</td>
</tr>
<tr>
<td>Unknown, conserved</td>
</tr>
<tr>
<td>Signal transduction</td>
</tr>
<tr>
<td>Metabolism, energy</td>
</tr>
<tr>
<td>Transporters/storage</td>
</tr>
<tr>
<td>Protein modification machinery</td>
</tr>
<tr>
<td>Oxidant metabolism/detoxification</td>
</tr>
<tr>
<td>Protein export machinery</td>
</tr>
<tr>
<td>Transcription machinery</td>
</tr>
<tr>
<td>Proteosome machinery</td>
</tr>
<tr>
<td>Cytoskeletal</td>
</tr>
<tr>
<td>Metabolism, lipid</td>
</tr>
<tr>
<td>Nuclear regulation</td>
</tr>
<tr>
<td>Transcription factor</td>
</tr>
<tr>
<td>Metabolism, carbohydrate</td>
</tr>
<tr>
<td>Metabolism, nucleotide</td>
</tr>
<tr>
<td>Metabolism, amino acid</td>
</tr>
<tr>
<td>Metabolism, intermediate</td>
</tr>
<tr>
<td>Extracellular matrix/cell adhesion</td>
</tr>
</tbody>
</table>

Total: 308 672. 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. An interesting feature of the lipocalins is their well-conserved three-dimensional structure despite the low similarity between random primary sequences of different proteins. 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. They have been also identified as anticoagulants, antiplatelets, and vasodilatory molecules in the SGs of bloodsucking insects. 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. Salivary lipocalin functional studies have been reported from such blood-feeding insects as *Triatoma pallidipennis*, *R. prolixus*, *T. infestans*, *T. dimidiata*, *D. maxima*, and also in tick saliva. The most abundant salivary proteins in ticks, mosquitoes, and triatomine bugs are usually binders of histamine or serotonin, thus serving as a kratagonist 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).

Table 3

<table>
<thead>
<tr>
<th>Number of ESTs</th>
<th>Class</th>
<th>ESTs/contig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipocalins</td>
<td>162</td>
<td>900</td>
</tr>
<tr>
<td>Trypsin family</td>
<td>8</td>
<td>56</td>
</tr>
<tr>
<td>Other peptide families</td>
<td>35</td>
<td>54</td>
</tr>
<tr>
<td>Trypsin-hemolysin family</td>
<td>5</td>
<td>24</td>
</tr>
<tr>
<td>Kazal family</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>Unknown conserved and secreted</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>Mys family of triatomines</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Antigen 5 family</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Metalloprotease</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Similar to accessory gland protein</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Cuticle protein</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Mucins</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>Number of ESTs</th>
<th>Class</th>
<th>ESTs/contig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inositol phosphate phosphatase</td>
<td>8</td>
<td>34</td>
<td>4.3</td>
</tr>
<tr>
<td>Apyrase</td>
<td>4</td>
<td>32</td>
<td>8.0</td>
</tr>
<tr>
<td>Serine protease</td>
<td>6</td>
<td>14</td>
<td>2.3</td>
</tr>
<tr>
<td>Acetylcholinesterase</td>
<td>2</td>
<td>2</td>
<td>1.0</td>
</tr>
<tr>
<td>Phosphodiesterase</td>
<td>1</td>
<td>1</td>
<td>1.0</td>
</tr>
<tr>
<td>Fucosidase</td>
<td>1</td>
<td>1</td>
<td>1.0</td>
</tr>
</tbody>
</table>

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

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). Several Kazal-type family members were previously described to be present in vertebrate and invertebrate animals. Thrombin inhibitors of the 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. Infestin is a Kazal-type proteinase inhibitor precursor described in *T. infestans*. It seems to be similar to rhodniin.
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.

Triaylsin. We identified six contigs (assembled from 54 ESTs that matched the protein named triaylsin 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 \(T. infestans\). Triaylsin, 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

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 line at the bottom indicates 20% amino acid diversity. This figure appears in color at www.ajtmh.org.
processed by limited proteolysis. Trialyisin 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 potnet 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 ants80,81 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.36,37 Furthermore, a recombinant member of this family was shown to interact with the Fc portion of immunoglobulins and may play a role in evading host immune response.85 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 deduced 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 Triatominae proteins were deorphanized.

The publicly available sequences from T. matogrossensis should facilitate identification of environmental antigens.
trigging fogo selvagem and development of immunological markers of exposure to this insect, as was done for T. infestans, where a recombinant protein has been used as a marker of low-level exposure to triatomines. 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.


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 R01-AR32599 and R01-AR32081 to Luis A. Diaz.

Disclaimer: The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the government of the United States of America.

Disclosure: Because TCFA, VMP, IMBF, JGV, and JMCJR are government employees and this is a government work, the work is in the public domain in the United States. Notwithstanding any other agreements, the NIH reserves the right to provide the work to PubMedCentral for display and use by the public, and PubMedCentral may tag or modify the work consistent with its customary practices. You can establish rights outside of the United States subject to a government use license.

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: assumpcao@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: ksedeaton@yahoo.com. Valéria Aoki and Evandro A. Ivitti, Department of Dermatology, Universidade de São Paulo, Brazil. E-mails: valeria.aoki@gmail.com and evandro.ivitti@hc.fmsp.usp.br. Gislan 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


51. Undheim EA, King GF, 2011. On the venom system of centipedes (Chilopoda), a neglected group of venomous animals. Toxicon 57: 512–524.


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 elas-


