Published in final edited form as:

J Proteomics. 2018 March 01; 174: 47–60. doi:10.1016/j.jprot.2017.12.016.

Exploring the molecular complexity of *Triatoma dimidiata* sialome

Paula Beatriz Santiago^a, Carla Nunes de Araújo^{a,b,*}, Sébastien Charneau^c, Izabela Marques Dourado Bastos^c, Teresa Cristina F. Assumpção^d, Rayner Myr Lauterjung Queiroz^c, Yanna Reis Praça^a, Thuany de Moura Cordeiro^c, Carlos Henrique Saraiva Garcia^c, Ionizete Garcia da Silva^e, Tainá Raiol^{c,f}, Flávia Nader Motta^b, João Victor de Araújo Oliveira^c, Marcelo Valle de Sousa^c, José Marcos C. Ribeiro^d, and Jaime Martins de Santana^{a,c}

^aPrograma Pós-Graduação em Ciências Médicas, Faculty of Medicine, The University of Brasília, Brasília, Brazil

^bFaculty of Ceilândia, The University of Brasília, Brasília, Brazil

^cDepartment of Cell Biology, The University of Brasília, Brasília, Brazil

^dVector Biology Section, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Disease, Rockville, United States

^eDepartment of Parasitology, The University of Goiás, Brazil

^fInstituto Leônidas e Maria Deane - Fiocruz Amazônia, Manaus, AM, Brazil

Abstract

Triatoma dimidiata, a Chagas disease vector widely distributed along Central America, has great capability for domestic adaptation as the majority of specimens caught inside human dwellings or in peridomestic areas fed human blood. Exploring the salivary compounds that overcome host haemostatic and immune responses is of great scientific interest. Here, we provide a deeper insight into its salivary gland molecules. We used high-throughput RNA sequencing to examine in depth the *T. dimidiata* salivary gland transcriptome. From more than 51 million reads assembled, 92.21% are related to putative secreted proteins. Lipocalin is the most abundant gene family, confirming it is an expanded family in *Triatoma* genus salivary repertoire. Other putatively secreted members include phosphatases, odorant binding protein, hemolysin, proteases, protease inhibitors, antigen-5 and antimicrobial peptides. This work expands the previous set of functionally annotated sequences from *T. dimidiata* salivary glands available in NCBI from 388 to 3,815. Additionally, we complemented the salivary analysis through proteomics (available data via ProteomeXchange

*Corresponding author.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Authors' contributions

PBS CNdA JMS conceived and designed the experiments. PBS TR JVdAO JMCR worked with sequencing data tools. PBS CNdA TCFA JMCR JMS analysed the sequencing data. PBS RMLQ TMC YRP SC performed proteomic experiments. PBS CNdA SC YRP CHSG analysed the proteomic data. SC IMDB IGdS FNM MVdS JMCR JMS contributed reagents/materials/analysis tools. PBS CNdA IMDB JMCR JMS wrote the paper. All authors read and approved the final manuscript.

with identifier PXD008510), disclosing the set complexity of 119 secreted proteins and validating the transcriptomic results. Our large-scale approach enriches the pharmacologically active molecules database and improves our knowledge about the complexity of salivary compounds from haematophagous vectors and their biological interactions.

Graphical Abstract

Keywords

Triatoma dimidiata; Triatominae; Chagas disease; Sialome; Haematophagy; Vector biology

1. Introduction

Haematophagous insects are parasite vectors of debilitating diseases to humans and domestic animals. Blood-sucking triatomine species (Hemiptera: Reduviidae) are able to transmit the protozoan parasite *Trypanosoma cruzi*, the causative agent of Chagas disease, to different mammalian hosts in the wild (sylvatic cycle). In peridomestic area domestic mammals also may serve as reservoir hosts (peridomestic cycle). The vector-borne transmission to humans is commonly observed when triatomines adapt to live in human dwellings, possibly in the cracked mud walls in rustic rural houses (domestic cycle). The infection usually occurs through the contact with contaminated triatomine faeces released after vector feeding, what may occur at the bite site or through mucous membranes. Epidemiologically less notified, additional mechanisms of transmission include ingestion of contaminated food, blood transfusion, organ transplantation and congenital transmission [1]. Despite international programs against domestic vectors achieved some progress in Chagas disease control, currently, new scenarios are emerging in different areas, such as domestic re-infestation, resistance to insecticides and sylvatic invasion of urban areas [1, 2].

Triatoma dimidiata is widely distributed in Americas, being found from Mexico, through Central America into north of South America [3]. Central America is noteworthy the region where *T. dimidiata* is highlighted as the main vector of the disease. In fact, it has great capability to colonize human dwellings [4–6], insects caught inside homes or in peridomestic areas in Central America were investigated and presented human and dog blood, showing their potential for domestic adaptation [7]. *T. dimidiata* has become the main

target of control measures in Central American countries where *Rhodnius prolixus* has been eliminated [7]. *T. dimidiata* shows different phenetic and genetic features throughout its geographical distribution, and it can be found in sylvatic, peridomestic and domestic ecotopes, influencing the effectiveness of vector control activities. Additionally, *T. dimidiata* exhibits variations in domestication, in infestation, and in rates of infection with *T. cruzi*, which in turn are factors that directly affect the epidemiological importance of this Chagas disease vector at the different regions [3]. All these issues make this vector's eradication an extremely difficult process.

The five nymph stages and adults of both sexes feed on blood, and in order to achieve a blood vessel, the triatomines insert the long and flexible mouthparts into vertebrate host skin causing tissue injury and bleeding. To avoid blood loss, the host responds with haemostasis (vasoconstriction, platelet aggregation and coagulation), inflammation and immunity. As a strategy to obtain a successful meal, the blood sucker releases a wide range of salivary molecules that act as platelet aggregation inhibitors, vasodilators and anticoagulants to counteract a number of specific targets in the haemostatic and wound responses, altering the physiology of the bite site and maintaining a continuous blood flow [8–10]. The description of haematophagous saliva composition, the sialomes, improves the understanding on the dynamics of the interactions between the vector and the vertebrate host, as well as how salivary molecules can affect disease transmission. In addition, sialome studies provide a greater insight on the evolution of blood-feeding habit. In this regard, distinctive sets of salivary protein families with specialized functions are associated with the blood-feeding life style adaptation, a result of independently evolution [11]. As such salivary proteins target the haemostatic and immune system of vertebrate hosts, the investigation of these molecules pursues as biotechnological interest.

A cDNA library of salivary glands (SGs) from *T. dimidiata* has been previously reported to contain 388 sequences, of which 53 are related to saliva [12]. Here, aiming to obtain a further insight into *T. dimidiata* salivary repertoire, we used RNA-seq to obtain its salivary transcriptome. We obtained more than 51 million Illumina reads assembled into 5,220 coding sequences (CDS), expanding the gene families related to secretory nature in *T. dimidiata* SGs transcriptome. Our results enabled the discovery of complete gene sequences and gene families not reported in this species before, offering additional data for the haematophagous salivary protein data banks, and enriching significantly the existing transcriptomic database of *T. dimidiata* [12]. Moreover, we also reported on the proteomic profile of this triatomine saliva.

2. Materials and methods

2.1 Triatomines and transcriptome salivary glands preparation

T. dimidiata triatomines were reared in standard insectary conditions at the University of Brasília, Brazil (27±1°C, 70–75% relative humidity and a 12 h/12 h light/dark cycle). The blood source of these insects was *Gallus gallus domesticus*. 5th instar nymphs and adults, starved for 5, 12 and 24 days, were used to capture the majority of transcripts differentially expressed in both stages. They were dissected under a stereomicroscope to obtain their SGs.

Pools containing thirty SGs pairs in cold Trizol reagent (Invitrogen, Carlsbad, CA, USA) were kept at -80°C until used for RNA extraction.

2.2 Salivary glands RNA extraction, library preparation and sequencing

Total RNA from SGs was isolated using Trizol and treated with turbo-DNase (Ambion, Austin, TX, USA) following the manufacturer's protocols. Isolated mRNA sample quality was checked by lab-on-chip analysis using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), and the sample was sent to the Federal District High-Performance Genome Center (DF, Brazil) for Illumina cDNA library construction and next generation sequencing. The SG library was constructed with standard protocols using TruSeq RNA kit, v2 (Illumina, San Diego, CA). The RNA-seq sequencing step was performed on an Illumina MiSeq sequencer (Illumina, San Diego, CA, USA). *T. dimidiata* SGs cDNA library was run on a single lane with another library, distinguished by bar coding. Paired-end reads of 300 nucleotides in length were generated. A total of 51,782,206 reads were generated and checked for data quality using FASTQC (Babraham Institute, Cambridge, UK). The nominal length of the sequences was 301 nt. Following trimming of low quality bases (quality 20 or lower), the average length was 233.6 bases, the median was 293 nt and L50 was 243 nt. Sequences smaller than 25 nt or with average quality < 20 were rejected.

2.3 Bioinformatic tools for transcriptomic analysis

Bioinformatic analyses were performed as described before [13]. A *de novo* assembly of the reads was done with Abyss [14] and Soapdenovo Trans [15] softwares using different kmer (k) values (from 20 to 95 at fivefold intervals). The data were joined by an iterative BLAST and cap3 assembler [16]. Sequence contamination among bar-coded libraries were identified and removed when their sequence identities were over 98%. CDS were extracted based on similarities to known proteins or on the existence of a signal peptide as evaluated by version 3.0 of the SignalP software [17]. CDS and their protein sequences were mapped into a hyperlinked Excel spreadsheet. To estimate the transcripts abundance, reads were mapped back into the CDS using blastn [18] with a word size of 25, masking homonucleotide decamers and allowing mapping up to five different CDS if the BLAST results had the same score values. Mapping of the reads was also included in the Excel spreadsheet. RPKM values for each coding sequence were also mapped to the spreadsheet. To compare the relative expression of transcripts, we used the term "expression index" (EI), which is defined as the number of reads mapped to a particular CDS divided by the highest found number of reads mapped to a single CDS. In the case of this transcriptome, a value of 4,415,811 reads mapped to a single lipocalin CDS was normalized to 100 and, then, used to calculate EI for each transcript. To functionally classify the protein sequences, an automated annotation was made based on a vocabulary of nearly 250 words found in matches to the different databases, including Swissprot, Gene Ontology, KOG, PFAM, SMART and a subset of the non-redundant protein database of NCBI. Additional manual annotation was made as required. Alignment analyses were done with Bioedit software [19] after multiple sequence alignment performed by ClustalW [20]. The sequences used in the alignments with *T. dimidiata* CDS were obtained from the nonredundant protein database of the NCBI and each of them is represented by six letters (derive from the first three letters from the genus and the first three

letters from the species name) followed by the NCBI gi accession number. *T. dimidiata* sequences are named with initials TD. The sequences were ClustalW aligned and the phylograms were done with the Mega package [21] after 1,000 bootstrap iterations using the neighbor-joining (NJ) algorithm.

2.4 Data access

The raw reads were deposited in the Sequence Read Archive (SRA) of NCBI, identifier SRP065810, under bioproject PRJNA300554. A total of 3,815 CDS were deposited in DDBJ/EMBL/GenBank through the Transcriptome Shotgun Annotation portal under the accession GECL00000000.

2.5 LC-MS/MS protein identification

Two independent biological replicates of fifteen pairs of SGs dissected from Vth instar nymphs and adults at 5, 12 and 24 days post blood meal (five pairs of each stage) were performed in order to increase the total number of identified proteins. The SGs were carefully punctured at 4 °C. Following centrifugation (16.000 × g, 15 min, 4 °C), the soluble protein fraction from each homogenate was ethanol/acetone precipitated. Resuspended proteins were consecutively reduced with 20 mM dithiothreitol at 56 °C for 45 min, alkylated with 40 mM iodoacetamide at room temperature in the dark for 60 min, digested overnight at 37 °C with modified trypsin (Promega, Madison, USA) and acidified to a final concentration of 0.1% trifluoroacetic acid (TFA) and desalted using homemade 2 cm long microcolumns of Poros Oligo R3 resin (Applied Biosystems) packed in p200 tips. Samples were subjected to LC-MS/MS analysis as previously described [22]. The concentration of the peptide solution was determined by Qubit® 2.0 quantification assay (Thermo Fischer). The tryptic peptides were first loaded onto a 2 cm fused silica trap column (150 µm inner diameter) packed in-house with reverse phase capillary column ReproSil-Pur C18-AQ 5 µm resin (Dr. Maisch GmbH, Germany) and separated in a 35 cm reverse phase capillary column ending with a fused silica emitter packed in-house with ReproSil-Pur 120 Å C18-AQ 3 µm resin (Dr. Maisch GmbH, Ammerbuch, Germany) using a Dionex Ultimate™ 3000 RSLC nano system (Thermo Fisher Scientific, Waltham, MA) coupled online with an LTQ-Orbitrap Elite™ mass spectrometer (Thermo Scientific, San Jose, CA). Briefly, peptides solubilized in 0.1% (v/v) formic acid were injected into the trap column and washed with 98% solvent A (0.1% formic acid in water) and 2% solvent B (0.1% formic acid in acetonitrile) under flow rate of 3 µL/min. Peptides were resolved over 200 minutes with 250 nL/min flow rate in the analytical column through use of gradients: 5–25% solvent B until 130 min; and then 45–85% until 181 min; 85% until 191 min; and re-equilibrated at 2% until 200 min. Peptides were ionized by Nanospray Flex™ ion source (Thermo Fisher Scientific, Waltham, MA, USA) at 2.5 kV and a capillary temperature of 275 °C. MS1 spectra were recorded in the Orbitrap mass analyzer with 120,000 resolution at m/z of 400, in the m/z range of 300 to 1800. After ion fragmentation using collision-induced dissociation (CID) with normalized collision energy of 35%, MS/MS spectra of the 15 most intense ions were acquired. Raw files were generated and used for protein identification using Proteome Discoverer v.2.1 (Thermo Scientific, Waltham, USA) with in-house SequestHT algorithm for *T. dimidiata* SG transcriptome and contaminant proteins (several human keratins, BSA and porcine trypsin). All contaminant proteins identified were manually

removed from the result lists. The searches were performed with the following parameters: MS accuracy 10 ppm, MS/MS accuracy 0.5 Da, trypsin digestion with up to 2 missed cleavage allowed, fixed carbamidomethyl modification of cysteine and variable modification of oxidized methionine. Number of proteins, protein groups and number of peptides were filtered for the false discovery rate less than 1%, peptide rank of 1 and at least 2 peptides per protein.

2.6 Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [23, 24] via the PRIDE [25] partner repository with the dataset identifier PXD008510.

3. Results and discussion

3.1 *T. dimidiata* carries a diverse and secreted-rich sialotranscriptome

In order to identify CDS from *T. dimidiata* SGs, total RNA was used to construct a cDNA library sequenced via Illumina-based next generation sequencing. A set of 5,220 CDS was extracted and classified into five categories of expressed genes as follows: putative Secreted, Housekeeping, Unknown, Transposable element, and Viral product, condensed in Table 1. Summing up a total of 564 CDS, the putative Secreted category, which are associated with haematophagy, mapped with the majority of the reads, 92.21%, and had an average of 84,405 reads per CDS. The Housekeeping category comprised 6.28% of total reads, with 4,012 CDS and an average of 808 reads per CDS. Together, these classes covered 4,576 CDS and over 98% of the mapped reads. The transcriptomic analysis and functional classification of CDS were based on sequence homology and expression index (EI) levels. The EI gives an overview about the resulting reads assembly distribution and the relative expression of each CDS. Based on previous results, CDS with values $EI > 0.5$ are considered of high expression [26].

The Housekeeping CDS were further divided into 24 subcategories accordingly to their possible biological function (Table 2). Interestingly, the Oxidant metabolism/detoxification subclass was the most abundant, accounting for 17.37% of housekeeping reads. This percentage is mainly due to the expression of cytochrome P450 gene family. Thirteen CDS were assembled from 539,482 reads, five of them had high expression indexes, $0.72 < EI < 3.03$. Cytochrome P450 members are also abundant in the sialotranscriptome of *Triatoma rubida* [27], and detected in *Rhodnius prolixus* [28] and *Triatoma matogrossensis* [29].

Insects develop resistance to insecticides through multiple metabolic resistance mechanisms and detoxification occurs primarily by the action of cytochrome P450 enzymes, glutathione S-transferases, and carboxylesterases [30]. In this regard, through molecular modelling and docking of a candidate cytochrome P450, Mamidala *et al.* (2011) compared and contrasted pesticide-resistant and pesticide-susceptible strains of the bed bug *Cimex lectularius*, revealing a flexibility of the deduced protein to metabolize a broad range of insecticide

substrates. Future studies may provide insights into the basis for insecticide resistance in blood-sucking bugs as well.

Finally, considering other Housekeeping reads, 9.84% mapped to protein synthesis machinery, followed by 8.99% that mapped to signal transduction, both sets give rise to molecules of important functions of saliva.

The Secreted category was also divided into different subcategories accordingly to their putative function (Table 3). Lipocalin is the most abundant secreted gene family, comprising almost 90% of the reads. The remaining reads are distributed among other ubiquitous secreted expressed genes, highlighting inositol polyphosphate phosphatases (IPPs), protease inhibitors, antigen-5, apyrases, immunity related molecules, serine proteases, hemolysin and odorant binding protein (OBP). The following section describes the functional classification of putative secreted CDS from *T. dimidiata* SGs based on homology, evolutionary approaches, and EI levels.

3.2 Members of the *T. dimidiata* sialotranscriptome required for blood feeding

3.2.1 Small molecule binding domains

Lipocalins: Present in a variety of organisms, the lipocalins are a widespread functional diverse family of typically small and extracellular members that usually act through the binding of small and mostly hydrophobic molecules, but may also bind to cell-surface receptors or form complexes with other soluble macromolecules [31–33]. Here, the transcriptomic analysis reveals lipocalin is the most expressed salivary gene family in *T. dimidiata* SGs, totalizing 78 CDS assembled from 42,637,563 reads (Table 3). The lipocalin abundance is a common feature reported before in other *Triatoma* sialotranscriptomes [29, 34–36]. Remarkably, lipocalins are also abundantly expressed in ticks [37, 38].

Accordingly, the seventeen most expressed transcripts belong to this family, as shown by their expression indexes: the CDS TD_32097, TD_31156, TD_31150 and TD_31157, which are triabin-like lipocalins, possess EI=100, 94.28, 83.94 and 75.05 respectively, and the next thirteen lipocalins show EI between 16.95<EI>62.09. From the 78 annotated lipocalin CDS, 26 are members of the triabin family. A phylogenetic tree was constructed by ClustalW alignment (Fig. 1). The result shows sequences grouped within different clades containing members of triabin, pallidipin, triatin, triafestin, procalin and apolipoprotein-D subfamilies. The expanded repertoire of this gene family may be the result of specific lineage expansion through gene duplication events during millions of years of *T. dimidiata* haematophagic habit evolution and adaptation to their host physiological characteristics [39]. In this regard, while members from the same clade may share similar functions, components from different clades may have acquired novel functions over generations, even though all of them are deeply related to evade host responses. Therefore, the multiple copies of salivary lipocalin genes in *T. dimidiata*, as well as in other triatomines, may create a genetic redundancy that seems to be fundamental to its salivary biological function. The set of salivary genes available during adaptation to a blood-feeding habit may have influenced this feature [40, 41].

It has been suggested that in haematophagous arthropods an expansion of the lipocalin family would have occurred twice, independently, in ticks and bugs [37, 38] [42]. The biological function of lipocalins in tick saliva is diverse, such as histamine and serotonin binding [43, 44], immunomodulation [45], anti-complement activity [46] and toxic properties [38]. On the same hand, many salivary lipocalins have been functionally characterized in triatomines. In *Rhodnius* spp., the nitrophorin is an abundant nitric oxide (NO)-carrying hemeprotein able to induce vasodilation and block platelet aggregation [42]. Although typical of *Rhodnius* genus, nitrophorin CDS has already been reported among the salivary transcripts of *Triatoma brasiliensis* [47] and *Triatoma pallidipennis* [48], and here, from 78 lipocalin CDS, 5 belong to nitrophorins.

Thrombin is a key serine protease to haemostasis and thrombosis as it activates coagulation factors, cleaving fibrinogen to fibrin and stimulating platelet aggregation [49]. Salivary lipocalin members from *Triatoma* spp. acting as thrombin inhibitors have been studied previously. In *T. pallidipennis*, triabin is an anticoagulant molecule that inhibits thrombin activation, forming a noncovalent complex at a molar ratio of 1:1 [50]. From the same species, pallidipin specifically inhibits platelet aggregation induced by collagen [51]. *Dipetalogaster maxima* salivary dipetalodipin shares strong sequence similarity to pallidipin, being suggested it has multifunctional anti-haemostatic properties, reducing platelet aggregation and vasoconstriction, besides blocking angiogenesis [52]. Others triatoma lipocalins have been described. Procalin is the major allergen protein present in *Triatoma protracta* saliva [53]. From *Triatoma infestans*, triafestin exhibits both anti-inflammatory activity by inhibiting the kallikrein-kinin system, and anticoagulant activity by interacting with coagulation pathway factor XII [54]. Finally, triplatin inhibits platelet aggregation induced by low doses of collagen and thromboxane A₂, but not by ADP, and seem to interact with PGF (2 α) modulating vessel tonus [55].

Mysterious protein 3 (MYS3)/hemolysin: MYS members were firstly reported in *R. prolixus* sialotranscriptome [28] and later on in *T. brasiliensis* [56] and *T. matogrossensis* [29] as hemolysin-like proteins. Common in microorganisms such as *Escherichia coli* [57], *Bacillus cereus* [58] and *Staphylococcus aureus* [59], the name hemolysin is based on their actions as red blood cell lytic toxins. Their toxicity has been also demonstrated to occur on epithelial, endothelial and immune cells, and on platelets [59]. The hemolysins are water-soluble monomers that undergoes conformational changes on the surface of susceptible cells to form β -barrel oligomers in the transmembrane space, forming cytotoxic pores in lipid bilayers that culminate in cellular lysis [57, 59, 60]. Roles of MYS/hemolysin in triatomine saliva have been proposed: erythrocyte lysis to help the early steps of digestion process [34] and a non-blood-feeding role, possibly antimicrobial [11, 28]. In *T. dimidiata* salivary transcriptome analysis, representing 0.09 % of the total reads from Secreted category, eight CDS were assigned as MYS/hemolysin members (Table 3), one of them presenting EI=0.84.

Odorant binding protein (OBP): OBP is a small soluble polypeptide highly concentrated in olfactory organs. The molecule displays binding activity responsible to detect chemical signals, activating specific chemoreceptors through yet unknown mechanisms [61, 62]. Common both in the lymph of chemosensilla in insects and in the nasal apparatus of

vertebrates, the OBPs are not homologous between these taxa. However, they are crucial at the first step of olfactory signal transmission, transforming an extracellular chemical signal into an intracellular electronic signal [62, 63], and thus present the same physiological function. Insects OBPs are composed by a pattern of six α -helical domains arranged to enclose a hydrophobic binding pocket stabilized by three disulfide bridges [64]. OBP members have been reported in the saliva of *Rhodnius*, *Triatoma* and *Cimex* genus [11]. In blood-feeding Diptera, a molecule distantly related to the OBP family, termed D7, is abundantly found in the saliva and counteracts biogenic amines, representing a possible example of convergent evolution [65].

The analysis of *T. dimidiata* sialotranscriptome resulted in the extraction of nine OBP CDS (Table 3), with maximum EI=0.02. Alignment analysis of TD_3904, TD_4743 with their matching OBP sequences from the NR database reveals the six cysteine residues involved in disulfide bridges formation and the functional domain of OBPs (smart00708; pfam01395) (Fig. 2). Olfaction enables insects to recognize outer signals ensuring their survival and reproduction. During the evolution of blood-feeding habit, triatomine OBPs have been probably adapted to facilitate the meal acquisition process.

Others small binding molecules: Up to 15,491 reads are related to juvenile hormone binding and pheromone-hormone binding transcripts (Table 3). Their products act as carriers supplying hormones to the target tissues to regulate insect development, metamorphosis and reproduction [66, 67]. It is feasible that in saliva, they might bind to some host molecules associated with haemostasis.

3.2.2 Antigen-5—The CAP protein superfamily (cysteine-rich secretory proteins (CRISP), antigen-5, and pathogenesis-related 1 proteins) is composed of secreted proteins found in a wide range of organisms. The CAP domain forms a unique three layers alpha-beta-alpha fold that is common to all members of the CAP superfamily (ipr014044). Some members of this family may have the CRISP domain, which contains 10 conserved cysteines especially in their C-terminal third involved in disulphide bonds, and is structurally related to the toxins BgK and ShK that are ion channel inhibitors [68]. Together with phospholipase A1, antigen-5 is the major protein component of wasps, hornets and fire ants venom [69, 70]. As new sialome studies are unrevealed, antigen-5 protein shows to be ubiquitous in haematophagous insect saliva. *T. infestans* and *D. maxima* antigen-5 have been reported as molecules that modulate platelet aggregation induced by low doses of collagen [71].

In this *T. dimidiata* sialotranscriptome, antigen-5 sequences were extracted from 338, 467 reads distributed in 6 CDS (Table 3). The sequence TD_32570 exhibits a high EI=15.25. Interestingly, different from the other *T. dimidiata* antigen-5 sequences, which possess only the CAP domain (pfam 00188), TD_13979 is a longer sequence presenting both a CAP and a CRISP domain (pfam 08562). As a consequence, TD_13979 product is a multidomain protein, where both domains may work in cooperation, strategy that could improve the effector function of this protein increasing the binding affinity and/or target specificity. This feature is common in antigen-5 members from reptile venoms [72, 73]. The phylogram reveals the sequence TD_13979 grouped with diptera species and *Zootermopsis nevadensis*, a termite species, in Clade I (Fig. 3). Another sequence, TD_5315 possesses two CAP

domains and is phylogenetically closer to sequences sharing this feature, for instance those from *Cimex lectularius* and *Halyomorpha halys*. Other three sequences from *T. dimidiata* grouped in Clade II, together with *T. infestans* antigen-5, containing only one CAP domain. These results indicate that *T. dimidiata* contains three different lineages of antigen-5 proteins.

3.2.3 Enzymes

Phosphatases

Inositol polyphosphate 5-phosphatase (IPP): IPP is an enzyme family whose members selectively remove the phosphate from the 5'-position of the inositol ring from both lipidic and non-lipidic substrates. IPPs are associated with different cellular processes such as signal transduction, secretion and cytoskeletal structure [74]. In bed bug and triatomine sialotranscriptomes, they are widely expressed [11]. The presence of salivary IPPs in the genus *Triatoma* can be related to anti-haemostatic responses facilitating blood meal acquisition [29, 34].

Secreted IPPs members have been described in *C. lectularius* [75, 76] and *R. prolixus* [77]. *C. lectularius* IPPs are associated with a heme moiety that stores and transports NO, promoting vasodilation and inhibiting platelet aggregation at the bite site [75, 76]. *R. prolixus* salivary type II IPP (IPPRp) shows selectivity for soluble and lipid substrates [77]. IPPRp acts hydrolysing inositol phosphate through three to four allosteric sites. Inositol phosphate concentration changes in platelet plasma membrane could affect the cytoskeletal architecture or induce changes in membrane trafficking and vesicle secretion, a possible regulatory mechanism to counteract platelet host response [77].

In this work, the IPPs are the most expressed transcripts after lipocalins, comprising 5.37% of the reads from Secreted category. Five CDS were assembled from 2,560,519 reads (Table 3), all of them exhibit EI between 2.33<EI>14.43. Phylogenetic analysis of *T. dimidiata* IPPs resulting from their alignment with their sequences matches shows Clade I containing *T. dimidiata* and *T. infestans* sequences, divergent from *R. prolixus* and other arthropods sequences grouped in Clade II (Fig. 4).

Apyrase/5' nucleotidase: During host's haemostasis response, activated platelets release ADP and serotonin from dense granules, agonists that recruit more platelets, amplifying the activation pathways that induce their aggregation to damaged vessel walls [78]. Among the inhibitory mechanisms of platelet aggregation already reported, an important strategy is performed by apyrases, which hydrolyze ADP and ATP to AMP and Pi [79]. The enzyme is ubiquitous in the saliva of haematophagous arthropods, reinforcing its fundamental role in blood meal acquisition. Based on substrate specificities and amino acid sequences, three genetic families have been reported in haematophagous saliva: (1) the 5'-nucleotidase family, described in mosquitoes [80], triatomines [81, 82] and ticks [83]; (2) the Cimex family, found in bed bugs and flies [84–86]; and (3) the CD39 family, a cell-surface apyrase, reported in fleas [87]. It has been suggested elsewhere that even distantly related arthropod species can share the same apyrase family, a result of convergent evolution [88].

Here, three CDS containing a signal peptide of secretion were assembled from 190,433 reads (Table 3), one of them with an EI=1.15. TD_31776 and TD_31514 CDS were aligned with their matches demonstrating that these sequences are characterized by an N-terminal metallophosphatase domain (pfam 00149) and a C-terminal 5'-nucleotidase domain (pfam 02872), as in *T. infestans* apyrase (gi 34481604) (Fig. 5, A). Notably, the conserved residues are those related to the active site even with sequences not derived from haematophagous insects, suggesting conservation from an ancestral lineage. The phylogram analysis confirms that *T. dimidiata* CDS are closely related to *T. infestans* salivary apyrases grouped in Clade III, which are divergent from those sequences of non-haematophagous arthropods, Clade I (Fig. 5, B). *D. maxima* [89] and *T. matogrossensis* [29] sialotranscriptomes also present 5'-nucleotidase family members CDS. *T. brasiliensis* [56] and *T. rubida* [27] sialotranscriptomes present a single CDS, suggesting a low level of 5'-nucleotidase expression in these species. Schwarz *et al.* (2014) extended the sialotranscriptome database of *T. infestans* using RNA-seq strategy [90] and identified additional expressed transcripts 66% identical to the described *T. infestans* salivary apyrases [82, 91].

3.2.4 Proteases—Of 30 CDS derived from 109,266 reads, *T. dimidiata* SGs transcriptomic analysis disclosed a variable repertoire of proteases (Table 3). The majority codes for serine proteases, a large family of peptidases containing a typical catalytic triad composed by His/Asp/Ser residues in the polypeptide chain [92]. In insects, proteases are responsible for protein catabolism and digestive processes. In hemipterans, a high protease activity in digestive tract is due mainly to cysteine and aspartic proteases [93, 94].

Although commonly found in arthropod sialotranscriptomes [56, 95, 96], few serine proteases displaying fibrinolytic activity have been characterized in triatomines [97, 98]. Remarkably, in *Rhodnius neglectus* sialotranscriptome the serine proteases have been found as a bulk of transcripts from SGs, thereby with possible roles in hematophagy and immunity [99]. Proteolytic activities also detected in SGs from non-haematophagous heteropterans [100, 101] suggest haematophagous salivary proteases derive from an ancestral lineage, which during adaptative pressure to blood-feeding habit, they may have been selected to play different functions.

In the current transcriptome, 12 CDS code for serine proteases, nine of them contain the CDD (cd00190) and Smart (smart00020) Tryp-SPc trypsin domain, similar to *R. neglectus* SGs transcripts [99]. TD_32288, a homolog of *T. infestans* salivary trypsin, presented higher EI value (EI=0.5).

The phylogram shows different specific clades containing the serine proteases sequences and reveals this is an expanded gene family in the *T. dimidiata* SGs (Fig. 6), as observed in *R. neglectus* sialotranscriptome [99]. There is a clade composed only by triatomine salivary serine proteases, suggesting a common ancestral lineage among them. Venom and snake venom serine protease (SVSP) sequences are grouped in another clade, composed by bumblebee, honeybee and paper wasp, indicating they make part of the set of allergenic molecules from these hymenoptera insects. Notably, bee venom serine protease (Bi-VSP) and SVSP exhibit fibrin(ogen)olytic activity in mammals [102, 103]. Characteristically these proteases are typical trypsins with a single serine protease domain [104]. ClustalW

alignment indicates TD_4045 shares the catalytic triad (His, Asp and Ser) in the sequence with BiVSP and SVSP, which is surrounded by conserved residues (Fig. 7).

At the bottom of the phylogenetic tree there is a divergent clade represented by serine carboxypeptidases (SCPs) members of the S28 family of serine exopeptidases. Among members of this family, the lysosomal_prolylcarboxypeptidase (PRCP) is implicated in the regulation of vascular tone, inducing vasodilation dependent on endothelium-derived nitric oxide [105]. In *T. brasiliensis*, SCPs from the S10 family function as digestive enzymes [106]. The exactly role of salivary proteases remains to be investigated and functionally characterized, but this protease mixture may enhance the anti-haemostatic or immunomodulatory effect of saliva, and may also present an allergenic role, as observed in hymenoptera venom [104].

Four CDS coding for di and tripeptidyl peptidases were identified (Table 3), among them, a dipeptidylpeptidase Angiotensin Converting Enzyme (ACE) CDS, a member of the M2-metalloprotease family (pfam 01401). *T. pallidipennis* SGs also present ACE [48]. In mammals, this enzyme is known to regulate blood pressure through the conversion of angiotensin I to the active vasoconstrictor angiotensin II, as well as through the inactivation of the vasodilator bradykinin [107, 108]. So far, ACE homologous have been found in several dipteran insect tissues and may play a role in reproduction, fertility [109] and hormone biosynthesis [110].

Additionally, this *T. dimidiata* sialotranscriptomic analysis display metalloproteases and cathepsins. Metalloproteases have been associated with fibrinogenolytic and anti-angiogenic activities in *I. scapularis* saliva [111], while the biological function of the cathepsins present in saliva remain unknown, but it is possible they serve a constitutive housekeeping role, acting in cellular organelles such as lysosomes or the endoplasmic reticulum [96].

3.2.5 Protease inhibitors—In order to maintain a continuous blood flow and success in feeding, the salivary proteins target different enzymes in host's coagulation cascade, inhibiting proteolysis and acting as anti-haemostatic proteins. From 974,943 reads, 26 CDS were extracted as serine protease inhibitors such as kazal-type, serpin, pacifastin and cystatin families (Table 3).

Kazal-type family: The function of protease inhibitors in haematophagous saliva is frequently associated with anticoagulant activity through inactivation of SPs. A different role was proposed to the blood-feeding horsefly *Tabanus yao*, the protease inhibitor vasotab acts as a vasodilator [112, 113], and may also inhibit platelet aggregation through its KGD (Lys-Gly-Asp) motif [114].

Comprising most of the reads from the protease inhibitors group (a total of 919,924 reads), four from sixteen CDS matching with members from kazal-family are well expressed, with an EI $7.8 < EI < 10.54$.

The kazal-type sequences phylogram is shown in Fig. 8. The sequences TD_30859, TD_17576 and TD_21160 are grouped in clade II with vasotab-like sequences containing only one kazal domain (cd00104; smart00280). In addition TD_33228 and TD_28148,

which contain three and two kazal domains respectively, are grouped in clade I with protease inhibitors containing multi kazal domains isolated from triatomine gut, such as dipetalogastin from *D. maxima* [115], infestin from *T. infestans* [116], and brasiliensin from *T. brasiliensis* [117]. This result suggests that different kazal-type serine protease inhibitors in *T. dimidiata* saliva might work in more than one manner. In homologous sequences to dipetalogastin each domain may bind independently to different surface sites of thrombin, inhibiting blood coagulation. In the *Triatoma* genus, vasodilator molecules are not commonly reported, but the relaxation of a blood vessel to speed the ingestion of the blood meal is an important strategy and may be represented in *T. dimidiata* saliva.

Serpin (Serine protease inhibitor): Serpin is a superfamily widely distributed among eukaryotes that presents a consensus tertiary structure containing a bundle of nine α -helices and three β -sheets [118, 119]. They show a reactive center loop (RCL) that interacts with the target protease, leading to catalytic site disruption [119, 120]. Among a diversity of functions, serpins regulate the activity of serine proteases in the vertebrate blood coagulation pathway [121]. Accordingly, it has been proposed serpins from tick saliva act as anticoagulant and immunosuppressive molecules [122–125]. Not usually described in triatomine and diptera, a salivary serpin from the Dengue vector, *Aedes aegypti*, has been characterized as a potent reversible inhibitor of the host's coagulation factor Xa [126, 127]. Thus, in haematophagous saliva, serpins may be considered among the remarkable diverse group of salivary anticoagulants targeting factors of blood coagulation cascade.

In *T. dimidiata* sialotranscriptome three CDS are related to serpins. The transcripts presented a relatively low EI (EI=0.0, EI=0.3 and EI=0.36), although all of them have been predicted to have signal peptide. The phylogenetic tree shows TD_28419 and TD_33209 grouped in Clade I with sequences containing just one serpin domain (cd00172; pfam00079; smart00093), whereas TD_28621 grouped in Clade II, with sequences containing two serpin domains (Fig. 9).

Pacifastin: Pacifastin is a serine protease inhibitor family firstly described in the plasma of the crayfish *Pacifastacus leniusculus* that inhibits both trypsin and chymotrypsin in an efficient manner [128]. This crustacean inhibitor proteinase is heterodimeric and possess a consensus inhibitory subunit composed of six cysteine residues, C1 X9–12 C2 N X C3 X C4 X2–3 G X3–4 C5 T X3 C6, arranged to form a pattern with three disulfide bonds [129, 130]. Studies reported this inhibitor presence in insects and crustaceans [131–135]. Pacifastin-like molecules present in blood-sucking insects are up-regulated in the fat body after a blood meal, suggesting a role in insect's immune system as a manner to counteract microorganisms acquired during the meal [136]. Although little is known about their physiological role and target proteases, they might be related to arthropod innate immune system.

Pacifastin members have been described in *R. neglectus* [99], *P. megistus* [96] and *T. infestans* [90] sialotranscriptomes. Here, 2,057 reads and four CDS (Table 3) are related to pacifastin, all products display signal peptide prediction and the five domains from the molecular signature of the family (pfam05375) (Fig. 10).

3.2.6 Immunity related families—Immune mechanisms play a key role in the survival of insects protecting them from external invaders. Due to pathogen invasion and microbial components recognition by pattern recognition receptors, the immune humoral response culminates in production and release of antimicrobial peptides (AMPs) [137]. Three examples of AMPs found in *T. dimidiata* sialotranscriptome are defensins, lysozymes and trialysins, which cause bacterial cell wall disruption/permeabilization [138–140]. Trialysin is a *T. infestans* pore forming lytic salivary protein that may also have a function in blood feeding since it is able to permeate mammalian cells [141]. One of the two detected *T. dimidiata* trialysin CDS is well expressed, with an EI of 2.11.

Despite a positive signal peptide prediction in defensins and lysozymes CDS, none of them present good EI value, suggesting salivary function for their proteins may be selectively up regulated as a strategy to prevent microbial infection. Defensins have been reported in *T. infestans* [95] and *P. megistus* [96] SGs transcriptomes.

3.3 A panoramic view of the transcriptome results

Sialotranscriptomes from four triatomine genera (*Rhodnius*, *Triatoma*, *Dipetalogaster* and *Panstrongylus*) were reported. It is possible to observe that when increasing the depth analysis using NGS, the salivary transcriptomic screening expands, enabling the evaluation of novel gene families and their abundance in each species. Our results indicate that the *T. dimidiata* sialotranscriptome shares with other triatomine sialotranscriptomes a core set of putative secreted genes that may be expressed to ensure blood feeding. They code for proteins involved in anti-haemostatic, anti-inflammatory and antimicrobial functions, as well as modulation of immune responses. These genes may also be roughly shared among other haematophagous arthropods. Our data also revealed IPP transcripts as the second most abundants, a feature that may represent an expansion and specialization of this gene family in *T. dimidiata* compared with related triatomine species sialotranscriptomes.

3.4 The proteome analysis from *T. dimidiata* saliva reveals a positive correlation with the sialotranscriptome

As a strategy to obtain a large-scale proteomic analysis of the *T. dimidiata* saliva, the SGs contents were tryptic digested and subjected to LC/MS-MS mass spectrometry. Of 119 secreted proteins identified (see Supplementary Table 1 for more information), 47 are members of the lipocalin family, including triabin, triatin, pallidipin and nitrophorin (Table 4). Other soluble proteins identified in the saliva are: OBP, pheromone-hormone binding protein, hemolysin, IPP, apyrase, ACE, trypsin, protease inhibitors and antigen-5 among other putative secreted proteins (Table 4).

To address an integrated analysis of transcriptomic and proteomic data, we investigated the correlation among reads and protein abundance. The result reveals there is a positive correlation, $r = 0.64$, among the abundance of SGs reads from secretory nature and the PSM's from salivary proteins identified in the proteomic analysis (Fig. 11). Considering the 20 enzymes and the 47 lipocalins identified, the correlation coefficient are even high, 0.75 and 0.73 respectively, indicating compatible changes in mRNA/protein pairs and a convergent regulation of transcription and translational processes for these gene families

(Fig. 11). In the other hand, contradictory changes in mRNA/protein pairs with negative correlation are observed among OBP ($r = -0.37$) and protease inhibitor members ($r = -0.29$), suggesting divergent regulation of transcription and translational processes. This weak correlation may be due to biological factors affecting translational processes [142]. Altogether, these results suggest that haematophagous salivation is a thoroughly regulated process. Other proteins may reveal a modest level of contamination that might have occurred during SGs dissection and saliva collection, not reflecting true salivary proteins.

4. Conclusions

Here we achieved a deeper sequencing coverage in comparison to the first *T. dimidiata* sialotranscriptome, expanding from 388 to 3,815 CDS deposited on NCBI databank. Furthermore, we reported products not described in the first *T. dimidiata* SGs transcriptomic analysis, such as, IPPs, mys/hemolysins, proteases, protease inhibitors (for instance, Kazal domain containing peptides) and antimicrobial peptides (trialysins, defensins, among other), suggesting different biological functions and complementing the first information. The greater complexity of *T. dimidiata* salivary glands transcripts shown here may be due to our high-throughput RNA sequencing strategy, leading to a more comprehensive analysis in comparison with the first *T. dimidiata* transcriptome report where, accordingly to the authors, a small number of cDNA clones were analysed [12]. In addition, we provide the salivary proteomic analysis, which disclosed the diversity of *T. dimidiata* salivary proteins that may be secreted during the feeding time, a set of molecules able to disarrange hosts' inflammation and haemostasis and neutralize the immune system.

We expect this work will help future investigations regarding vector biology and disease transmission, besides functional roles of salivary proteins and their exploitation for identification and selection new targets for disease control strategies and/or novel promising pharmacologically active compounds.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are very grateful to Dr. Jaques Miranda Ferreira de Souza, Ana Cristina Gomes and Lucas Barbosa de Aguiar for technical assistance.

Declarations

Availability of data and material

Sequence data that support the findings of this study have been deposited in the Sequence Read Archive (SRA) of NCBI, identifier SRP065810, under bioproject PRJNA300554.

Sequence data that support the findings of this study have been deposited in DDBJ/EMBL/GenBank through the Transcriptome Shotgun Annotation portal under the accession GECL00000000.

Funding

This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, grant number 23038.005298/2011-83); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grant

numbers 79/2013-MCTI/CNPq/FNDCT, 407730/2013-3 PRO-CENTRO-OESTE and 476252/2012-1); Fundação de Amparo à Pesquisa do Distrito Federal (FAPDF, grant numbers 193.001.076/2015, 193.001.053/2015 and 193.000.822/2015); Instituto Nacional de Ciência e Tecnologia de Estratégias de Interação Patógeno-Hospedeiro (INCT, grant number: 465771/2014-9); Financiadora de Estudos e Projetos (FINEP) and the Intramural Research Program of the National Institute of Allergy and Infectious Diseases (Vector-Borne Diseases: Biology of Vector Host Relationship, grant number Z01 AI000810-18 to JMCR). Because JMCR is a government employee 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 U.S. subject to a government use license. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Bern C. Chagas' Disease. *N Engl J Med*. 2015; 373(5):456–66. [PubMed: 26222561]
- Andrade DV, Gollob KJ, Dutra WO. Acute chagas disease: new global challenges for an old neglected disease. *PLoS Negl Trop Dis*. 2014; 8(7):e3010. [PubMed: 25077613]
- Dorn PL, Monroy C, Curtis A. *Triatoma dimidiata* (Latreille, 1811): a review of its diversity across its geographic range and the relationship among populations. *Infect Genet Evol*. 2007; 7(2):343–52. [PubMed: 17097928]
- Ramírez CJ, et al. Genetic structure of sylvatic, peridomestic and domestic populations of *Triatoma dimidiata* (Hemiptera: Reduviidae) from an endemic zone of Boyaca. Colombia *Acta Trop*. 2005; 93(1):23–9. [PubMed: 15589794]
- Guhl F. Chagas disease in Andean countries. *Mem Inst Oswaldo Cruz*. 2007; 102(Suppl 1):29–38.
- Ponce C. Current situation of Chagas disease in Central America. *Mem Inst Oswaldo Cruz*. 2007; 102(Suppl 1):41–4.
- Coura JR. The main sceneries of Chagas disease transmission. The vectors, blood and oral transmissions—a comprehensive review. *Mem Inst Oswaldo Cruz*. 2015; 110(3):277–82. [PubMed: 25466622]
- Ribeiro JM. Blood-feeding arthropods: live syringes or invertebrate pharmacologists? *Infect Agents Dis*. 1995; 4(3):143–52. [PubMed: 8548192]
- Champagne DE. Antihemostatic molecules from saliva of blood-feeding arthropods. *Pathophysiol Haemost Thromb*. 2005; 34(4–5):221–7. [PubMed: 16707932]
- Champagne DE. Antihemostatic strategies of blood-feeding arthropods. *Curr Drug Targets Cardiovasc Haematol Disord*. 2004; 4(4):375–96. [PubMed: 15578959]
- Ribeiro JMC, Assumpção TC, Francischetti IMB. An Insight into the Sialomes of Bloodsucking Heteroptera. *Psyche*. 2012; 2012(470436):16.
- Kato H, et al. 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*. 2010; 10(2):184–91. [PubMed: 19900580]
- Ribeiro JM, et al. An insight into the sialome of the frog biting fly, *Corethrella appendiculata*. *Insect Biochem Mol Biol*. 2014; 44:23–32.
- Simpson JT, et al. ABySS: a parallel assembler for short read sequence data. *Genome Res*. 2009; 19(6):1117–23. [PubMed: 19251739]
- Luo R, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*. 2012; 1(1):18. [PubMed: 23587118]
- Karim S, Singh P, Ribeiro JM. A deep insight into the sialotranscriptome of the gulf coast tick, *Amblyomma maculatum*. *PLoS ONE*. 2011; 6(12):e28525. [PubMed: 22216098]
- Nielsen H, Brunak S, von Heijne G. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng*. 1999; 12(1):3–9. [PubMed: 10065704]
- Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997; 25(17):3389–402. [PubMed: 9254694]
- Hall, TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT 1999. *Oxford University Press*; p. 95-98. *Nucleic Acids Symposium Series*

20. Larkin MA, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007; 23(21):2947–8. [PubMed: 17846036]
21. Tamura K, et al. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*. 2007; 24(8):1596–9. [PubMed: 17488738]
22. Queiroz RM, et al. Comprehensive proteomic analysis of *Trypanosoma cruzi* epimastigote cell surface proteins by two complementary methods. *J Proteome Res*. 2013; 12(7):3255–63. [PubMed: 23682730]
23. Deutsch EW, et al. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res*. 2017; 45(D1):D1100–D1106. [PubMed: 27924013]
24. Vizcaíno JA, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol*. 2014; 32(3):223–6. [PubMed: 24727771]
25. Vizcaíno JA, et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res*. 2016; 44(D1):D447–56. [PubMed: 26527722]
26. Ribeiro JM, et al. An insight into the sialome of the horse fly, *Tabanus bromius*. *Insect Biochem Mol Biol*. 2015; 65:83–90. [PubMed: 26369729]
27. Ribeiro JM, et al. An insight into the sialotranscriptome of *Triatoma rubida* (Hemiptera: Heteroptera). *J Med Entomol*. 2012; 49(3):563–72. [PubMed: 22679863]
28. Ribeiro JM, et al. Exploring the sialome of the blood-sucking bug *Rhodnius prolixus*. *Insect Biochem Mol Biol*. 2004; 34(1):61–79. [PubMed: 14976983]
29. Assumpcao TC, et al. An insight into the sialotranscriptome of *Triatoma matogrossensis*, a kissing bug associated with fogo selvagem in South America. *Am J Trop Med Hyg*. 2012; 86(6):1005–14. [PubMed: 22665609]
30. Mamidala P, Jones SC, Mittapalli O. Metabolic Resistance in Bed Bugs. *Insects*. 2011; 2(1):36–48. [PubMed: 26467498]
31. Flower DR. Multiple molecular recognition properties of the lipocalin protein family. *J Mol Recognit*. 1995; 8(3):185–95. [PubMed: 8573354]
32. Flower DR, North AC, Sansom CE. The lipocalin protein family: structural and sequence overview. *Biochim Biophys Acta*. 2000; 1482(1–2):9–24. [PubMed: 11058743]
33. Grzyb J, Latowski D, Strzałka K. Lipocalins - a family portrait. *J Plant Physiol*. 2006; 163(9):895–915. [PubMed: 16504339]
34. Assumpcao TC, et al. An insight into the sialome of the blood-sucking bug *Triatoma infestans*, a vector of Chagas' disease. *Insect Biochem Mol Biol*. 2008; 38(2):213–32. [PubMed: 18207082]
35. Ribeiro JM, et al. An insight into the sialotranscriptome of *Triatoma rubida* (Hemiptera: Heteroptera). *J Med Entomol*. 2012; 49(3):563–72. [PubMed: 22679863]
36. Ribeiro JMC, Schwarz A, Francischetti IMB. A deep insight into the sialotranscriptome of the Chagas disease vector, *Panstrongylus megistus* (Hemiptera: Heteroptera). *J Med Entomol*. 2015; 52(3):351–358. [PubMed: 26334808]
37. Mans BJ, Neitz AW. Exon-intron structure of outlier tick lipocalins indicate a monophyletic origin within the larger lipocalin family. *Insect Biochem Mol Biol*. 2004; 34(6):585–94. [PubMed: 15147759]
38. Mans BJ, Louw AI, Neitz AW. The major tick salivary gland proteins and toxins from the soft tick, *Ornithodoros savignyi*, are part of the tick Lipocalin family: implications for the origins of tick toxicoses. *Mol Biol Evol*. 2003; 20(7):1158–67. [PubMed: 12777525]
39. Schofield CJ, Galvão C. Classification, evolution, and species groups within the Triatominae. *Acta Trop*. 2009; 110(2–3):88–100. [PubMed: 19385053]
40. Mans BJ, Neitz AW. Adaptation of ticks to a blood-feeding environment: evolution from a functional perspective. *Insect Biochem Mol Biol*. 2004; 34(1):1–17. [PubMed: 14723893]
41. Mans BJ, Ribeiro JM. Function, mechanism and evolution of the moubatin-clade of soft tick lipocalins. *Insect Biochem Mol Biol*. 2008; 38(9):841–52. [PubMed: 18694828]
42. Montfort WR, Weichsel A, Andersen JF. Nitrophorins and related antihemostatic lipocalins from *Rhodnius prolixus* and other blood-sucking arthropods. *Biochim Biophys Acta*. 2000; 1482(1–2): 110–8. [PubMed: 11058753]

43. Paesen GC, et al. Tick histamine-binding proteins: isolation, cloning, and three-dimensional structure. *Mol Cell*. 1999; 3(5):661–71. [PubMed: 10360182]
44. Sangamnatdej S, et al. A high affinity serotonin- and histamine-binding lipocalin from tick saliva. *Insect Mol Biol*. 2002; 11(1):79–86. [PubMed: 11841505]
45. Preston SG, et al. Novel immunomodulators from hard ticks selectively reprogramme human dendritic cell responses. *PLoS Pathog*. 2013; 9(6):e1003450. [PubMed: 23825947]
46. Nunn MA, et al. Complement inhibitor of C5 activation from the soft tick *Ornithodoros moubata*. *J Immunol*. 2005; 174(4):2084–91. [PubMed: 15699138]
47. Gurgel-Gonçalves R, et al. Enzootic transmission of *Trypanosoma cruzi* and *T. rangeli* in the Federal District of Brazil. *Rev Inst Med Trop Sao Paulo*. 2004; 46(6):323–30. [PubMed: 15654478]
48. Hernández-Vargas MJ, et al. Proteomic and transcriptomic analysis of saliva components from the hematophagous reduviid *Triatoma pallidipennis*. *J Proteomics*. 2017; 162:30–39. [PubMed: 28442446]
49. Kalafatis M, et al. The regulation of clotting factors. *Crit Rev Eukaryot Gene Expr*. 1997; 7(3): 241–80. [PubMed: 9399073]
50. Noeske-Jungblut C, et al. Triabin, a highly potent exosite inhibitor of thrombin. *J Biol Chem*. 1995; 270(48):28629–34. [PubMed: 7499380]
51. Noeske-Jungblut C, et al. An inhibitor of collagen-induced platelet aggregation from the saliva of *Triatoma pallidipennis*. *J Biol Chem*. 1994; 269(7):5050–3. [PubMed: 8106481]
52. Assumpcao TC, et al. 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*. 2010; 285(50):39001–12. [PubMed: 20889972]
53. Paddock CD, et al. Identification, cloning, and recombinant expression of procalin, a major triatomine allergen. *J Immunol*. 2001; 167(5):2694–9. [PubMed: 11509613]
54. Isawa H, et al. Identification and characterization of plasma kallikrein-kinin system inhibitors from salivary glands of the blood-sucking insect *Triatoma infestans*. *FEBS J*. 2007; 274(16):4271–86. [PubMed: 17645545]
55. Ma D, et al. Triplatin, a platelet aggregation inhibitor from the salivary gland of the triatomine vector of Chagas disease, binds to TXA(2) but does not interact with glycoprotein PVI. *Thromb Haemost*. 2012; 107(1):111–23. [PubMed: 22159626]
56. Santos A, et al. The sialotranscriptome of the blood-sucking bug *Triatoma brasiliensis* (Hemiptera, Triatominae). *Insect Biochem Mol Biol*. 2007; 37(7):702–12. [PubMed: 17550826]
57. Bhakdi S, et al. Staphylococcal alpha-toxin, streptolysin-O, and *Escherichia coli* hemolysin: prototypes of pore-forming bacterial cytolysins. *Arch Microbiol*. 1996; 165(2):73–9. [PubMed: 8593102]
58. Ramarao N, Sanchis V. The pore-forming haemolysins of *Bacillus cereus*: a review. *Toxins (Basel)*. 2013; 5(6):1119–39. [PubMed: 23748204]
59. Berube BJ, Wardenburg J Bubeck. Staphylococcus aureus α -toxin: nearly a century of intrigue. *Toxins (Basel)*. 2013; 5(6):1140–66. [PubMed: 23888516]
60. Montoya M, Gouaux E. Beta-barrel membrane protein folding and structure viewed through the lens of alpha-hemolysin. *Biochim Biophys Acta*. 2003; 1609(1):19–27. [PubMed: 12507754]
61. Pelosi P. Odorant-binding proteins. *Crit Rev Biochem Mol Biol*. 1994; 29(3):199–228. [PubMed: 8070277]
62. Pelosi P, et al. Soluble proteins of chemical communication: an overview across arthropods. *Front Physiol*. 2014; 5:320. [PubMed: 25221516]
63. Fan J, et al. An overview of odorant-binding protein functions in insect peripheral olfactory reception. *Genet Mol Res*. 2011; 10(4):3056–69. [PubMed: 22180039]
64. Pelosi P, et al. Structure and biotechnological applications of odorant-binding proteins. *Appl Microbiol Biotechnol*. 2014; 98(1):61–70. [PubMed: 24265030]
65. Calvo E, et al. Function and evolution of a mosquito salivary protein family. *J Biol Chem*. 2006; 281(4):1935–42. [PubMed: 16301315]

66. Gäde G, Hoffmann KH, Spring JH. Hormonal regulation in insects: facts, gaps, and future directions. *Physiol Rev.* 1997; 77(4):963–1032. [PubMed: 9354810]
67. Zalewska M, et al. Juvenile hormone binding protein traffic – Interaction with ATP synthase and lipid transfer proteins. *Biochim Biophys Acta.* 2009; 1788(9):1695–705. [PubMed: 19426705]
68. Gibbs GM, Roelants K, O'Bryan MK. The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins—roles in reproduction, cancer, and immune defense. *Endocr Rev.* 2008; 29(7):865–97. [PubMed: 18824526]
69. Hoffman DR. 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.* 1993; 92(5):707–16. [PubMed: 8227862]
70. King TP, Spangfort MD. Structure and biology of stinging insect venom allergens. *Int Arch Allergy Immunol.* 2000; 123(2):99–106. [PubMed: 11060481]
71. Assumpção TC, et al. Salivary antigen-5/CAP family members are Cu²⁺-dependent antioxidant enzymes that scavenge O₂ · and inhibit collagen-induced platelet aggregation and neutrophil oxidative burst. *J Biol Chem.* 2013; 288(20):14341–61. [PubMed: 23564450]
72. Yamazaki Y, Morita T. Structure and function of snake venom cysteine-rich secretory proteins. *Toxicon.* 2004; 44(3):227–31. [PubMed: 15302528]
73. Nobile M, et al. Helothermine, a lizard venom toxin, inhibits calcium current in cerebellar granules. *Exp Brain Res.* 1996; 110(1):15–20. [PubMed: 8817251]
74. Erneux C, et al. The diversity and possible functions of the inositol polyphosphate 5-phosphatases. *Biochim Biophys Acta.* 1998; 1436(1–2):185–99. [PubMed: 9838104]
75. Valenzuela JG, Walker FA, Ribeiro JM. A salivary nitrophorin (nitric-oxide-carrying hemoprotein) in the bedbug *Cimex lectularius*. *J Exp Biol.* 1995; 198(Pt 7):1519–26. [PubMed: 7658188]
76. Valenzuela JG, Ribeiro JM. Purification and cloning of the salivary nitrophorin from the hemipteran *Cimex lectularius*. *J Exp Biol.* 1998; 201(Pt 18):2659–64. [PubMed: 9716517]
77. Andersen JF, Ribeiro JM. A secreted salivary inositol polyphosphate 5-phosphatase from a blood-feeding insect: allosteric activation by soluble phosphoinositides and phosphatidylserine. *Biochemistry.* 2006; 45(17):5450–7. [PubMed: 16634626]
78. Jurk K, Kehrel BE. Platelets: physiology and biochemistry. *Semin Thromb Hemost.* 2005; 31(4):381–92. [PubMed: 16149014]
79. Ribeiro JM, et al. Salivary apyrase of *Aedes aegypti*: characterization and secretory fate. *Comp Biochem Physiol B.* 1984; 79(1):81–6. [PubMed: 6094095]
80. Champagne DE, et al. The salivary gland-specific apyrase of the mosquito *Aedes aegypti* is a member of the 5'-nucleotidase family. *Proc Natl Acad Sci U S A.* 1995; 92(3):694–8. [PubMed: 7846038]
81. Faudry E, et al. Kinetics of expression of the salivary apyrases in *Triatoma infestans*. *Insect Biochem Mol Biol.* 2004; 34(10):1051–8. [PubMed: 15475299]
82. Faudry E, et al. *Triatoma infestans* apyrases belong to the 5'-nucleotidase family. *J Biol Chem.* 2004; 279(19):19607–13. [PubMed: 14985353]
83. Stutzer C, et al. *Ornithodoros savignyi*: soft tick apyrase belongs to the 5'-nucleotidase family. *Exp Parasitol.* 2009; 122(4):318–27. [PubMed: 19393241]
84. Valenzuela JG, et al. Purification, cloning, and expression of an apyrase from the bed bug *Cimex lectularius*. A new type of nucleotide-binding enzyme. *J Biol Chem.* 1998; 273(46):30583–90. [PubMed: 9804829]
85. Valenzuela JG, et al. The salivary apyrase of the blood-sucking sand fly *Phlebotomus papatasi* belongs to the novel *Cimex* family of apyrases. *J Exp Biol.* 2001; 204(Pt 2):229–37. [PubMed: 11136609]
86. Charlab R, et al. Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly *Lutzomyia longipalpis*. *Proc Natl Acad Sci U S A.* 1999; 96(26):15155–60. [PubMed: 10611354]
87. Andersen JF, et al. An insight into the sialome of the oriental rat flea, *Xenopsylla cheopis* (Rots). *BMC Genomics.* 2007; 8:102. [PubMed: 17437641]

88. Hughes AL. Evolution of the salivary apyrases of blood-feeding arthropods. *Gene*. 2013; 527(1): 123–30. [PubMed: 23791653]
89. Assumpção TC, et al. Insight into the salivary transcriptome and proteome of *Dipetalogaster maxima*. *J Proteome Res*. 2011; 10(2):669–79. [PubMed: 21058630]
90. Schwarz A, et al. An updated insight into the Sialotranscriptome of *Triatoma infestans*: developmental stage and geographic variations. *PLoS Negl Trop Dis*. 2014; 8(12):e3372. [PubMed: 25474469]
91. Faudry E, et al. Salivary apyrases of *Triatoma infestans* are assembled into homo-oligomers. *Biochem J*. 2006; 396(3):509–15. [PubMed: 16542158]
92. Polgár L. The catalytic triad of serine peptidases. *Cell Mol Life Sci*. 2005; 62(19–20):2161–72. [PubMed: 16003488]
93. Terra WR, Ferreira C. Insect digestive enzymes: properties, compartmentalization and function. *Comp Biochem Physiol*. 1994; 109B(1):1–62.
94. Santiago PB, et al. Proteases of haematophagous arthropod vectors are involved in blood-feeding, yolk formation and immunity - a review. *Parasit Vectors*. 2017; 10(1):79. [PubMed: 28193252]
95. Assumpção TC, et al. An insight into the sialome of the blood-sucking bug *Triatoma infestans*, a vector of Chagas' disease. *Insect Biochem Mol Biol*. 2008; 38(2):213–32. [PubMed: 18207082]
96. Ribeiro JM, Schwarz A, Francischetti IM. A Deep Insight Into the Sialotranscriptome of the Chagas Disease Vector, *Panstrongylus megistus* (Hemiptera: Heteroptera). *J Med Entomol*. 2015; 52(3):351–8. [PubMed: 26334808]
97. Amino R, Tanaka AS, Schenkman S. Triapsin, an unusual activatable serine protease from the saliva of the hematophagous vector of Chagas' disease *Triatoma infestans* (Hemiptera: Reduviidae). *Insect Biochem Mol Biol*. 2001; 31(4–5):465–72. [PubMed: 11222956]
98. Meiser CK, et al. A salivary serine protease of the haematophagous reduviid *Panstrongylus megistus*: sequence characterization, expression pattern and characterization of proteolytic activity. *Insect Mol Biol*. 2010; 19(3):409–21. [PubMed: 20345395]
99. Santiago PB, et al. A Deep Insight into the Sialome of *Rhodnius neglectus*, a Vector of Chagas Disease. *PLoS Negl Trop Dis*. 2016; 10(4):e0004581. [PubMed: 27129103]
100. Cohen AC. Organization of Digestion and Preliminary Characterization of Salivary Trypsin-like Enzymes in a Predaceous Heteropteran, *Zelus renardii*. *J Insect Physiol*. 1993; 39(10):823–829.
101. Colebatch G, East P, Cooper P. Preliminary characterisation of digestive proteases of the green mirid, *Creontiades dilutus* (Hemiptera: Miridae). *Insect Biochem Mol Biol*. 2001; 31(4–5):415–23. [PubMed: 11222951]
102. Choo YM, et al. Dual function of a bee venom serine protease: prophenoloxidase-activating factor in arthropods and fibrin(ogen)olytic enzyme in mammals. *PLoS One*. 2010; 5(5):e10393. [PubMed: 20454652]
103. Matsui T, Fujimura Y, Titani K. Snake venom proteases affecting hemostasis and thrombosis. *Biochim Biophys Acta*. 2000; 1477(1–2):146–56. [PubMed: 10708855]
104. Winningham KM, et al. Hymenoptera venom protease allergens. *J Allergy Clin Immunol*. 2004; 114(4):928–33. [PubMed: 15480337]
105. Mallela J, Yang J, Shariat-Madar Z. Prolylcarboxypeptidase: a cardioprotective enzyme. *Int J Biochem Cell Biol*. 2009; 41(3):477–81. [PubMed: 18396440]
106. Waniek PJ, et al. Serine carboxypeptidases of *Triatoma brasiliensis* (Hemiptera, Reduviidae): Sequence characterization, expression pattern and activity localization. *J Insect Physiol*. 2014; 63:9–20. [PubMed: 24548612]
107. Corvol P, Williams TA, Soubrier F. Peptidyl dipeptidase A: angiotensin I-converting enzyme. *Methods Enzymol*. 1995; 248:283–305. [PubMed: 7674927]
108. Zhao Y, Xu C. Structure and function of angiotensin converting enzyme and its inhibitors. *Sheng Wu Gong Cheng Xue Bao*. 2008; 24(2):171–6. [PubMed: 18464595]
109. Isaac RE, et al. Insect angiotensin-converting enzyme. A processing enzyme with broad substrate specificity and a role in reproduction. *Ann N Y Acad Sci*. 1999; 897:342–7. [PubMed: 10676461]

110. Macours N, Hens K. Zinc-metalloproteases in insects: ACE and ECE. *Insect Biochem Mol Biol*. 2004; 34(6):501–10. [PubMed: 15147752]
111. Francischetti IM, Mather TN, Ribeiro JM. 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*. 2003; 305(4):869–75. [PubMed: 12767911]
112. Takac P, et al. Vasotab, a vasoactive peptide from horse fly *Hybomitra bimaculata* (Diptera, Tabanidae) salivary glands. *J Exp Biol*. 2006; 209(Pt 2):343–52. [PubMed: 16391356]
113. Ma D, et al. Anti-thrombosis repertoire of blood-feeding horsefly salivary glands. *Mol Cell Proteomics*. 2009; 8(9):2071–9. [PubMed: 19531497]
114. Zhang Z, et al. A potent anti-thrombosis peptide (vasotab TY) from horsefly salivary glands. *Int J Biochem Cell Biol*. 2014; 54:83–8. [PubMed: 25025626]
115. Mende K, et al. Dipetalogastin, a potent thrombin inhibitor from the blood-sucking insect. *Dipetalogaster maximus* cDNA cloning, expression and characterization. *Eur J Biochem*. 1999; 266(2):583–90. [PubMed: 10561601]
116. Campos IT, et al. Infestin, a thrombin inhibitor presents in *Triatoma infestans* midgut, a Chagas' disease vector: gene cloning, expression and characterization of the inhibitor. *Insect Biochem Mol Biol*. 2002; 32(9):991–7. [PubMed: 12213235]
117. Araujo RN, et al. Brasiliensin: A novel intestinal thrombin inhibitor from *Triatoma brasiliensis* (Hemiptera: Reduviidae) with an important role in blood intake. *Int J Parasitol*. 2007; 37(12):1351–8. [PubMed: 17575982]
118. Gettins PG. Serpin structure, mechanism, and function. *Chem Rev*. 2002; 102(12):4751–804. [PubMed: 12475206]
119. Irving JA, et al. Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function. *Genome Res*. 2000; 10(12):1845–64. [PubMed: 11116082]
120. Huntington JA, Read RJ, Carrell RW. Structure of a serpin-protease complex shows inhibition by deformation. *Nature*. 2000; 407(6806):923–6. [PubMed: 11057674]
121. Rau JC, et al. Serpins in thrombosis, hemostasis and fibrinolysis. *J Thromb Haemost*. 2007; 5(Suppl 1):102–15.
122. Leboulle G, et al. Characterization of a novel salivary immunosuppressive protein from *Ixodes ricinus* ticks. *J Biol Chem*. 2002; 277(12):10083–9. [PubMed: 11792703]
123. Imamura S, et al. A serine protease inhibitor (serpin) from *Haemaphysalis longicornis* as an anti-tick vaccine. *Vaccine*. 2005; 23(10):1301–11. [PubMed: 15652673]
124. Prevot PP, et al. Anti-hemostatic effects of a serpin from the saliva of the tick *Ixodes ricinus*. *J Biol Chem*. 2006; 281(36):26361–9. [PubMed: 16672226]
125. Mulenga A, Kim T, Ibelli AM. *Amblyomma americanum* tick saliva serine protease inhibitor 6 is a cross-class inhibitor of serine proteases and papain-like cysteine proteases that delays plasma clotting and inhibits platelet aggregation. *Insect Mol Biol*. 2013; 22(3):306–19. [PubMed: 23521000]
126. Stark KR, James AA. Isolation and characterization of the gene encoding a novel factor Xa-directed anticoagulant from the yellow fever mosquito, *Aedes aegypti*. *J Biol Chem*. 1998; 273(33):20802–9. [PubMed: 9694825]
127. Stark KR, James AA. A factor Xa-directed anticoagulant from the salivary glands of the yellow fever mosquito. *Aedes aegypti* *Exp Parasitol*. 1995; 81(3):321–31. [PubMed: 7498429]
128. Hergenhausen HG, Aspan A, Söderhäll K. Purification and characterization of a high-Mr proteinase inhibitor of pro-phenol oxidase activation from crayfish plasma. *Biochem J*. 1987; 248(1):223–8. [PubMed: 3124809]
129. Liang Z, et al. Pacifastin, a novel 155-kDa heterodimeric proteinase inhibitor containing a unique transferrin chain. *Proc Natl Acad Sci U S A*. 1997; 94(13):6682–7. [PubMed: 9192625]
130. Mer G, et al. Solution structure of PMP-C: a new fold in the group of small serine proteinase inhibitors. *J Mol Biol*. 1996; 258(1):158–71. [PubMed: 8613985]

131. Liu Y, et al. PtPLC, a pacifastin-related inhibitor involved in antibacterial defense and prophenoloxidase cascade of the swimming crab *Portunus trituberculatus*. *Fish Shellfish Immunol.* 2015; 43(1):36–42. [PubMed: 25542376]
132. Breugelmans B, et al. A lepidopteran pacifastin member: cloning, gene structure, recombinant production, transcript profiling and in vitro activity. *Insect Biochem Mol Biol.* 2009; 39(7):430–9. [PubMed: 19364530]
133. Simonet G, et al. Bacterial production and purification of SGPI-1 and SGPI-2, two peptidic serine protease inhibitors from the desert locust, *Schistocerca gregaria*. *Protein Expr Purif.* 2003; 31(2): 188–96. [PubMed: 14550636]
134. Simonet G, et al. Characterization of two novel pacifastin-like peptide precursor isoforms in the desert locust (*Schistocerca gregaria*): cDNA cloning, functional analysis and real-time RT-PCR gene expression studies. *Biochem J.* 2005; 388(Pt 1):281–9. [PubMed: 15631618]
135. Boigegrain RA, et al. Insect immunity: two proteinase inhibitors from hemolymph of *Locusta migratoria*. *Biochem Biophys Res Commun.* 1992; 189(2):790–3. [PubMed: 1472051]
136. de Marco R, et al. The first pacifastin elastase inhibitor characterized from a blood sucking animal. *Peptides.* 2010; 31(7):1280–6. [PubMed: 20381560]
137. Kurata S. Recognition and elimination of diversified pathogens in insect defense systems. *Mol Divers.* 2006; 10(4):599–605. [PubMed: 16969720]
138. Araújo CA, et al. Sequence characterization and expression patterns of defensin and lysozyme encoding genes from the gut of the reduviid bug *Triatoma brasiliensis*. *Insect Biochem Mol Biol.* 2006; 36(7):547–60. [PubMed: 16835020]
139. Daffre S, et al. The lysozyme locus in *Drosophila melanogaster*: an expanded gene family adapted for expression in the digestive tract. *Mol Gen Genet.* 1994; 242(2):152–62. [PubMed: 8159165]
140. Bulet P, Stöcklin R. Insect antimicrobial peptides: structures, properties and gene regulation. *Protein Pept Lett.* 2005; 12(1):3–11. [PubMed: 15638797]
141. Amino R, et al. Trialysin, a novel pore-forming protein from saliva of hematophagous insects activated by limited proteolysis. *J Biol Chem.* 2002; 277(8):6207–13. [PubMed: 11751887]
142. Wang J, et al. Integrated Analysis of Transcriptomic and Proteomic Datasets Reveals Information on Protein Expressivity and Factors Affecting Translational Efficiency. *Methods Mol Biol.* 2016; 1375:123–36. [PubMed: 25762301]

Significance

Several haematophagous triatomine species can transmit *Trypanosoma cruzi*, the etiological agent of Chagas disease. Due to the reemergence of this disease, new drugs for its prevention and treatment are considered priorities. For this reason, the knowledge of vector saliva emerges as relevant biological finding, contributing to the design of different strategies for vector control and disease transmission. Here we report the transcriptomic and proteomic compositions of the salivary glands (sialome) of the reduviid bug *Triatoma dimidiata*, a relevant Chagas disease vector in Central America. Our results are robust and disclosed unprecedented insights into the notable diversity of its salivary glands content, revealing relevant anti-haemostatic salivary gene families. Our work expands almost ten times the previous set of functionally annotated sequences from *T. dimidiata* salivary glands available in NCBI. Moreover, using an integrated transcriptomic and proteomic approach, we showed a correlation pattern of transcription and translation processes for the main gene families found, an important contribution to the research of triatomine sialomes. Furthermore, data generated here reinforces the secreted proteins encountered can greatly contribute for haematophagic habit, *Trypanosoma cruzi* transmission and development of therapeutic agent studies.

Highlights

- We provide a *Triatoma dimidiata* sialome analysis.
- We were able to extract 3,815 CDS from transcriptomic analysis.
- 119 proteins, of which, lipocalin, odorant binding protein, hemolysin, phosphatases, proteases, proteases inhibitors and antigen-5 were identified in proteomic analysis.
- The results disclosed relevant salivary molecules to the vector-host interaction for *Triatoma dimidiata* species.

Figure 1. Phylogram of the lipocalin family from *Triatoma dimidiata* salivary glands transcriptome

Phylogenetic tree showing the distance between members of the family is derived from the alignment of *T. dimidiata* CDS and other lipocalin sequences from Hemiptera identified as described in Methods section. The number at the nodes indicates percentage of bootstrap score above 70% value. The bar at the bottom represents 10% amino acid substitution. The colored circles identify the species whose sequences were used: blue, *T. dimidiata*; red, *Triatoma infestans*; green, *Meccus pallidipennis*; magenta, *Triatoma brasiliensis*; grey, *Triatoma protracta*; yellow, *Cimex lectularius*; cyan, *Halyomorpha halys*.

Figure 2. The odorant binding protein (OBP) from *Triatoma dimidiata* salivary glands transcriptome

ClustalW alignment of OBP members from *T. dimidiata* salivary transcriptome (TD_3904 and TD_4743) and other OBP sequences from Hemiptera identified as described in Methods section: *Triatoma brasiliensis* (gi 1026942666), *T. brasiliensis* (gi 1026942662), *Apolygus lucorum* (gi 1007629701), *Adelphocoris lineolatus* (gi 270000360), *Euschistus heros* (gi 299474028) and *Lygus lineolaris* (gi 573006010). The alignment indicates conserved residues in black and similar residues in grey background. The blue bar indicates the signal peptide indicative of secretion and the hashtag above indicate the position of the conserved cysteine residues.

Figure 3. Phylogram of the antigen-5 family from *Triatoma dimidiata* salivary glands transcriptome

Phylogenetic tree showing the distance between members of the family is derived from the alignment of *T. dimidiata* CDS and other antigen-5 sequences from Insecta identified as described in Methods section. The number at the nodes indicates percentage of bootstrap score above 70% value. The bar at the bottom represents 20% amino acid substitution.

Figure 4. Phylogram of the inositol polyphosphate 5-phosphatase (IPP) family from *Triatoma dimidiata* salivary glands transcriptome

Phylogenetic tree showing the distance between members of the family is derived from the alignment of *T. dimidiata* CDS and other IPPs sequences from Insecta identified as described in Methods section. The number at the nodes indicates percentage of bootstrap score above 70% value. The bar at the bottom represents 10% amino acid substitution.

Figure 5. Apyrase/5' nucleotidase from *Triatoma dimidiata* salivary glands transcriptome
 (A) ClustalW alignment of apyrase/5' nucleotidase members from *T. dimidiata* salivary transcriptome (TD_31776 and TD_31514) and other apyrase/5' nucleotidase sequences from Insecta: *T. infestans* (gi 34481604), *Cyphomyrmex costatus* (gi 1009376035), *Atta colombica* (gi 1009353741), *Trachymyrmex cornetzi* (gi 1009393714), *Trachymyrmex zeteki* (gi 1012981877) and *Harpegnathos saltator* (gi 307202233). The alignment indicates conserved residues in black and similar residues in grey background. The blue bar indicates the signal peptide indicative of secretion, while the green indicates the 5' nucleotide

superfamily located in C terminal domain. The hashtag shows the metal binding site residues and the asterisk shows the active site residues. **(B)** Phylogenetic tree showing the distance between members of the family is derived from the alignment of *T. dimidiata* CDS and other apyrase/5' nucleotidase sequences from Insecta identified as described in Methods section. The number at the nodes indicates percentage of bootstrap score above 70% value. The bar at the bottom represents 10% amino acid substitution.

Figure 6. Phylogram of the serine proteases from *Triatoma dimidiata* salivary glands transcriptome

Phylogenetic tree showing the distance between members of the family is derived from the alignment of *T. dimidiata* CDS and other serine proteases sequences from Insecta identified as described in Methods section. The number at the nodes indicates percentage of bootstrap score above 70% value. The bar at the bottom represents 20% amino acid substitution. The colored circles identify the species whose sequences were used: blue, *T. dimidiata*; red, Hemiptera order; and green, Hymenoptera order.

Figure 7. The serine protease from *Triatoma dimidiata* salivary glands transcriptome ClustalW alignment of serine protease member from *T. dimidiata* salivary gland transcriptome (TD_4045) and other serine proteases sequences from Insecta: *Cimex lectularius* (gi 939277320), *Halyomorpha halys* (gi 939641366), *Diuraphis noxia* (gi 985385408), *Copidosoma floridanum* (gi 936576698) and *Apis dorsata* (gi 572262138). The alignment indicates conserved residues in black and similar residues in grey background. The blue bar indicates the signal peptide indicative of secretion. The triangle indicates the cleavage site, the circles indicate the active site and the diamonds indicate the substrate binding sites.

Figure 8. Phylogram of the kazal-type protease inhibitors from *Triatoma dimidiata* salivary glands transcriptome

Phylogenetic tree showing the distance between members of the family is derived from the alignment of *T. dimidiata* CDS and other kazal-type protease inhibitors sequences from Insecta identified as described in Methods section. The number at the nodes indicates percentage of bootstrap score above 70% value. The bar at the bottom represents 20% amino acid substitution.

Figure 9. Phylogram of the serpin inhibitors from *Triatoma dimidiata* salivary glands transcriptome

Phylogenetic tree showing the distance between members of the family is derived from the alignment of *T. dimidiata* CDS and other serpin inhibitors sequences from Insecta identified as described in Methods section. The number at the nodes indicates percentage of bootstrap score above 70% value. The bar at the bottom represents 10% amino acid substitution.

Figure 10. The pacifastin inhibitors from *Triatoma dimidiata* salivary glands transcriptome
 Arrangement of pacifastin domain showing the residues position of the conserved inhibitory subunit, composed by C1 X9–12 C2 N X C3 X C4 X2–3 G X3–4 C5 T X3 C6, containing the six cysteines residues responsible for the intra-chain disulfide bridges: C1:C4, C2:C6, C3:C5. The *T. dimidiata* CDS contain pacifastin domain are TD_27731, TD_354, TD_3748 and TD_7118.

Figure 11. *Triatoma dimidiata* salivary transcriptome/proteome correlation

The diagram shows positive linear relationship between reads and proteins abundance ($p < 0.0001$).

Table 1

Functional classification and expression levels of the CDS extracted from the *de novo* assembly of *Triatoma dimidiata* sialotranscriptome

Class	Number of CDS	Number of mapped reads	Average number of reads/CDS	% Reads
Secreted	564	47,604,907	84,405	92.21
Housekeeping	4,012	3,241,995	808	6.28
Unknown	509	713,097	1,400	1.38
Transposable element	132	64,221	486	0.12
Viral product	3	1,154	384	0.00
Total	5,220	51,625,374		100

Table 2

Functional classification and expression levels of the CDS attributed to the Housekeeping class extracted from the *de novo* assembly of *Triatoma dimidiata* sialotranscriptome

Housekeeping subclass	Number of CDS	Number of reads	Average number of reads/CDS	% Reads
Oxidant metabolism/Detoxification	35	563,079	16,087	17.37
Protein synthesis machinery	187	319,088	1,706	9.84
Signal transduction	526	291,509	554	8.99
Cytoskeletal proteins	151	221,092	1,464	6.82
Transcription machinery	479	216,994	453	6.69
Protein export	269	205,218	762	6.33
Energy metabolism	297	188,405	634	5.81
Lipid metabolism	182	175,491	964	5.41
Transporters and channels	210	151,314	720	4.67
Proteasome machinery	189	146,380	774	4.52
Protein modification	196	142,266	725	4.39
Unknown conserved	390	126,538	324	3.90
Carbohydrate metabolism	122	97,475	798	3.01
Nuclear regulation	190	72,214	380	2.23
Amino acid metabolism	70	69,088	987	2.13
Unknown conserved membrane protein	84	41,160	490	1.27
Extracellular matrix	90	40,351	448	1.24
Immunity	46	37,723	820	1.16
Transcription factor	97	34,093	351	1.05
Detoxification	37	32,038	865	0.99
Nucleotide metabolism	80	31,259	390	0.96
Intermediary metabolism	47	17,385	369	0.54
Nuclear export	27	12,837	475	0.40
Storage	11	8,998	818	0.28
Total	4,012	3,241,995		100

Table 3

Functional classification and expression levels of the CDS attributed to the secreted class extracted from the *de novo* assembly of *Triatoma dimidiata* sialotranscriptome

Secreted class	Number of CDS	Number of reads	Average number of reads/CDS	% Reads
<i>Small molecule binding domains</i>				
Lipocalins	78	42,637,563	546,635	89.39
Odorant binding protein	9	2,951	327	0.01
Juvenile hormone binding	6	1,233	205	0.00
Pheromone-hormone binding	8	14,258	1,782	0.03
Mys/Hemolysin	8	44,218	5,527	0.09
<i>Enzymes</i>				
<i>Phosphatases</i>				
Inositol polyphosphate phosphatases	5	2,560,519	512,103	5.37
Apyrase/5' nucleotidase	3	190,433	63,477	0.40
<i>Proteases</i>				
Di and tripeptidylpeptidases	4	29,077	7,269	0.06
Serine proteases	12	71,829	5,985	0.15
Cathepsins	6	4,827	804	0.01
Metalloproteases	4	164	41	0.00
Other peptidases	4	3,369	842	0.01
<i>Lipases</i>	10	7,596	759	0.02
<i>Endonuclease</i>	3	796	265	0.00
<i>Protease inhibitors</i>				
Kazal-type	16	919,294	57,455	1.93
Serpins	3	52,799	17,599	0.11
Pacifastin	4	2,057	514	0.00
Cystatin	2	197	98	0.00
Inter-alpha trypsin inhibitor	1	596	596	0.00
<i>Antimicrobial peptides</i>				
Defensins	5	298	59	0.00
Lysozyme	4	740	185	0.00
Pathogen pattern recognition proteins	7	63,679	9,097	0.13
Prokineticin domain	1	59	59	0.00
Trialysin	2	95,978	47,989	0.20
<i>Antigen 5</i>	6	338,467	56,411	0.71
<i>Mucins</i>	12	11,822	985	0.02
<i>Phe rich peptides, possible artifacts</i>	24	28,454	1,185	0.06
<i>Other putative secreted proteins</i>	315	614,243	1,95	1.29
Total	566	47,700,885		100

Table 4

Classification and abundance of proteins from the salivary proteome of *Triatoma dimidiata* based on LC-MS/MS

Protein identification	Number of identified proteins	# PSM's
<i>Small molecule binding domains</i>		
<i>Lipocalin</i>		
Salivary lipocalin	22	2019
Triabin	14	1743
Triatin	5	2108
Pallidipin	4	72
Nitrophorin	2	13
<i>Odorant binding protein</i>	4	38
<i>Pheromone-hormone binding</i>	1	2
<i>MYS/Hemolysin</i>	4	43
<i>Enzymes</i>		
<i>Phosphatases</i>		
Inositol polyphosphate phosphatase	2	508
Apyrase/5' nucleotidase	2	212
<i>Proteases</i>		
Trypsin	3	553
Angiotensin converting enzyme	1	53
<i>Esterase and lipase</i>	4	23
<i>Lysozyme</i>	1	23
<i>Protease inhibitors</i>		
Kazal-type	4	44
Serpin	2	19
Pacifastin	1	35
Inter-alpha trypsin inhibitor	1	6
<i>Antigen-5</i>	1	61
<i>Other putative secreted proteins</i>	41	675

PSM's: peptide spectrum matches.