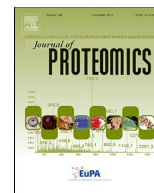
Journal of Proteomics 174 (2018) 47–60

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot



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[g], Marcelo Valle de Sousa[c], José Marcos C. Ribeiro[d], Jaime Martins de Santana[a,c]

[a] *Programa Pós-Graduação em Ciências Médicas, Faculty of Medicine, The University of Brasília, Brasília, Brazil*
[b] *Faculty of Ceilândia, The University of Brasília, Brasília, Brazil*
[c] *Department of Cell Biology, The University of Brasília, Brasília, Brazil*
[d] *Vector Biology Section, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Disease, Rockville, United States*
[e] *Department of Parasitology, The University of Goiás, Brazil*
[f] *Instituto Leônidas e Maria Deane - Fiocruz Amazônia, Manaus, AM, Brazil*
[g] *Department of Computer Science, The University of Brasília, Brasília, Brazil*

ARTICLE INFO

Keywords:
Triatoma dimidiata
Triatominae
Chagas disease
Sialome
Haematophagy
Vector biology

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 > 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 3815. Additionally, we complemented the salivary analysis through proteomics (available data *via* ProteomeXchange 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.

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.

* Corresponding author.
E-mail address: cnunes@unb.br (C.N. de Araújo).

https://doi.org/10.1016/j.jprot.2017.12.016
Received 29 September 2017; Received in revised form 20 December 2017; Accepted 22 December 2017
Available online 27 December 2017
1874-3919/ © 2017 Elsevier B.V. All rights reserved.

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 > 51 million Illumina reads assembled into 5220 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

Download English Version:

https://daneshyari.com/en/article/7633699

Download Persian Version:

https://daneshyari.com/article/7633699

Daneshyari.com