

RESEARCH

Open Access



# A deeper insight into the sialome of male and female *Culex quinquefasciatus* mosquitoes

Stephen Lu<sup>1</sup>, Ines Martin-Martin<sup>1,2</sup>, Jose M. Ribeiro<sup>1</sup> and Eric Calvo<sup>1\*</sup>

## Abstract

**Introduction** During evolution, blood-feeding arthropods developed a complex salivary mixture that can interfere with host haemostatic and immune response, favoring blood acquisition and pathogen transmission. Therefore, a survey of the salivary gland contents can lead to the identification of molecules with potent pharmacological activity in addition to increase our understanding of the molecular mechanisms underlying the hematophagous behaviour of arthropods. The southern house mosquito, *Culex quinquefasciatus*, is a vector of several pathogenic agents, including viruses and filarial parasites that can affect humans and wild animals.

**Results** Previously, a Sanger-based transcriptome of the salivary glands (sialome) of adult *C. quinquefasciatus* females was published based on the sequencing of 503 clones organized into 281 clusters. Here, we revisited the southern mosquito sialome using an Illumina-based RNA-sequencing approach of both male and female salivary glands. Our analysis resulted in the identification of 7,539 coding DNA sequences (CDS) that were functionally annotated into 25 classes, in addition to 159 long non-coding RNA (LncRNA). Additionally, comparison of male and female libraries allowed the identification of female-enriched transcripts that are potentially related to blood acquisition and/or pathogen transmission.

**Conclusion** Together, these findings represent an extended reference for the identification and characterization of the proteins containing relevant pharmacological activity in the salivary glands of *C. quinquefasciatus* mosquitoes.

**Keywords** Hematophagy, Arthropod, Saliva, Endonuclease, Medical entomology, Evolution

## Introduction

Blood acquisition is a pharmacological endeavor for any hematophagous arthropod. Upon piercing the host's skin, several defensive mechanisms are deployed, including vasoconstriction, platelet activation and blood coagulation leading to a reduction of the overall blood flow at the bite site. Additionally, a complex signaling pathway

also stimulates inflammatory and immune responses that can be detrimental to the blood feeding arthropod. Over the last decades, the structural and biochemical study of salivary proteins established that hematophagous vectors convergently evolved a complex and distinct salivary mixture that can interfere with host hemostatic and immune responses, leading to the postulation that all blood feeders have at least an anticoagulant, an inhibitor of platelet aggregation and a vasodilator in their saliva [1]. In addition to their role in blood acquisition, salivary proteins from mosquitoes have been shown to facilitate pathogen transmission to vertebrate hosts [2–5] and, therefore, can be considered potential targets for vaccine development.

\*Correspondence:

Eric Calvo  
ecalvo@niaid.nih.gov

<sup>1</sup>Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA

<sup>2</sup>Laboratory of Medical Entomology, National Center for Microbiology, Instituto de Salud Carlos III, Madrid, Spain



© This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

The mosquito *Culex quinquefasciatus*, commonly referred as the southern house mosquito, is a vector of several pathogenic agents that affect both humans and wild animals including the West Nile virus, St. Louis encephalitis virus, *Wuchereria bancrofti* and *Plasmodium relictum* [6–8]. However, the salivary activity of *C. quinquefasciatus* has been less explored than other medical relevant mosquitoes such as *Anopheles* sp. and *Aedes* sp., with only a few proteins characterized [9–12]. Although the southern mosquito genome is available [13], there is a limited knowledge regarding the sequences of genes expressed in the salivary gland with only a Sanger-based sialome of *C. quinquefasciatus* females being available [14]. Therefore, we present here a more comprehensive transcriptome analysis of both male and female adult mosquitoes based on an Illumina RNA-sequence approach. We carried out a *de novo* assembly of over 100 million high quality reads and combined the extracted CDS with the ones derived from the genome assembly, resulting in an extensive reference of *C. quinquefasciatus* salivary gland contents. A total of 7,539 CDS were identified and made publicly available. Finally, the comparison between male and female transcripts allowed the identification of sex-specific transcripts that might aid in the identification and characterization of pharmacologically active salivary proteins, proteins that can affect pathogen transmission as well as determine potential biomarkers of mosquito exposure. Additionally, comparison of mosquito sialomes could also provide insights regarding the evolution of the hematophagous behaviour within the Diptera order.

## Materials and methods

### Mosquitoes rearing and salivary glands collection

A *C. quinquefasciatus* colony was established in 2015 from egg rafts collected in Hilo, Hawaii, US and maintained at the Laboratory of Malaria and Vector Research insectary, NIAID, NIH. The mosquitoes were reared using 10% Karo Syrup prior to dissection. The salivary glands were dissected from 1- to 3-day old mosquitoes (between 8 and 10 AM) in sterile PBS pH 7.4 and immediately transferred to an Eppendorf tube containing 100 µl of TRIzol (Invitrogen). Pools of 50 pairs of female salivary glands and 100 pairs of male salivary glands were stored at –80 °C until RNA extraction. More male than female salivary glands were used to account for their overall lower total RNA amount [15, 16]. Three biological replicates of male and female samples were collected.

### Library preparation, sequencing, and analysis

Total RNA from mosquito salivary glands was isolated using TRIzol (Invitrogen) according to the manufacturer instructions. RNA purity was assessed with the NanoPhotometer® spectrophotometer (IMPLEN, CA, US), and

RNA integrity and quantification were assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, US). Sequencing libraries were generated using NEBNext® UltraTMRNALibrary Prep Kit for Illumina® (NEB, US) following manufacturer's instructions. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X) or by sonication with Diagenode Bioruptor Pico for breaking RNA strands. First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEB Next Adaptor with hairpin loop structure were ligated to prepare for hybridization. To select cDNA fragments of preferentially 150–200 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, US). Three microliters of USER Enzyme (NEB, US) were used with size-selected, adaptor-ligated cDNA at 37 °C for 15 min followed by 5 min at 95 °C before PCR. PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system. The Illumina reads were trimmed of the Illumina adapters and any low-quality sequences (Q<20) using TrimGalore (<https://github.com/FelixKrueger/TrimGalore>), merged into a single file and assembled using ABySS (V2.3.1) [17] with k values from 25 to 95 (with increments of 10) in single-stranded mode and Trinity (V2.9.0) [18] in single-stranded F mode. The assemblies from ABySS and Trinity were combined and filtered using CD-HIT [19]. Coding DNA sequences (CDS) with an open reading frame (ORF) of at least 150 nucleotides were extracted based on BLASTp results to several databases, including a subset of the non-redundant protein database, the transcriptome shotgun assembly (TSA) and the Refseq-invertebrate. CDS were extracted if sequences presented at least 70% of coverage with a matching protein. Additionally, all ORFs starting with a methionine and with 40 amino acids in length were submitted to the signalP tool (V3.0). Sequences that presented a putative signal peptide were mapped to the ORFs and the most 5' methionine was selected as the starting of the transcript [20]. The extracted CDS from our *de novo* assembly were merged with the ones annotated in the genome (assembly Cqui1.0, Genbank: GCA\_015732745.1) and the redundant sequences were filtered using the CD-HIT tool (considering 95% of identity). For annotation we used an

in-house program that scans a vocabulary of ~400 words and their order of appearance in the protein matches from BLAST results against several databases (TSA, subset of the NR, Refseq-protocista, Refseq-invertebrate, Refseq-vertebrate, uniprot, MEROPS, PFAM, Smart and CDD), including their e-values and coverage. Relative quantification of CDS was performed by mapping the trimmed library reads to the extracted CDS using the RSEM tool [21]. The annotated CDSs were exported to a hyperlinked Excel spreadsheet that is currently available for download (Supplementary file 1).

### Tridimensional structure prediction

Molecular modeling was performed using AlphaFold2 [22]. The superposition of modeled proteins with protein structures deposited on the Protein Data Bank (PDB) were prepared using PyMol (2.4).

### Statistical analysis

Differentially expressed transcripts analysis between females and males was carried out using the edgeR package [23] for R [24]. Briefly, we generate a counting matrix of each CDS by mapping the trimmed Illumina reads to the final extracted CDS using RSEM. We then extracted the count data of CDS that presented an average TPM  $\geq 3$  in male or female samples and used this filtered matrix as input for edgeR. Finally, the count data of each filtered CDS was normalized by the library size respective library using the edgeR function `calcNormFactors()`. For the identification of differently expressed transcripts we used the generalized linear model approach of edgeR with dispersion estimated by the Cox-Reid profile-adjusted likelihood. Transcripts were considered differentially expressed when presented a  $\text{LogFC} \geq \pm 2$ , the p-value and the false discovery rate (FDR) were less than 0.05. The phylogenetic tree was constructed using the maximum likelihood model [25] with MegaX [26]. The amino acid alignments were performed with Clustal omega [27] and edited with Bioedit [28].

## Results and discussion

### Overall description of *C. quinquefasciatus* sialome

Illumina sequencing of six libraries from *C. quinquefasciatus* salivary glands resulted in 105,064,955 high quality reads. Our *de novo* assembly using ABySS and Trinity generated 34,074 sequences from which a total of 21,106 potential CDS were extracted. After merging our *de novo* CDS with those derived from the genome assembly (assembly Cqui1.0, Genbank: GCA\_015732745.1) we obtained 31,426 putative CDS and 2,100 sequences annotated in the genome as non-coding RNA (ncRNA). The quantification of each transcript was estimated using the Transcript Per Million (TPM) measurement by mapping the trimmed Illumina reads to the final list of transcripts

using the RSEM tool. The levels mapped reads were similar between our male and female samples ( $64.5\% \pm 0.6$ ), with exception of one female sample (F1), which presented only 49.4% of mapped reads, indicating a possible bias with this sample. Despite this discrepancy, the values of mapped reads observed here are within range of other sialome from blood feeding arthropods [16, 29] and indicate that no major bias is present in our pipeline. It is noteworthy that the remaining unmapped reads could be potentially from 5' and 3' UTRs or putative CDS that we failed to identify due to lack of homology to previous deposited sequences, lack of a putative signal peptide or having an ORF with less than 150 nucleotides. Initial exploration of the data using multi dimension analysis and the heatmap plot based on the relative quantification of each transcript confirmed that sample F1 was not a suitable biological replicate (Supplementary Fig. 1) and, therefore, it was excluded from the differentially expression analysis.

The putative CDS that had an average TPM of at least 3 in males or female samples were extracted, resulting in 7,539 CDS that were functionally classified into 25 groups and 159 ncRNA (Supplementary file 1). As observed in the Sanger-based *C. quinquefasciatus* and in the sialome of other blood feeding vectors [14, 15, 30, 31], the “secreted” functional class was the major component of the mosquito salivary glands homogenates (Table 1). In females this class represented 61.2% of all quantified CDS, while in males it accounted for 56%. The second most abundant classes were the “unknown” class (4.63%) and the “protein synthesis” class (8.77%) in females and males, respectively. Finally, the third most abundant class in females (4.55%) was the “protein synthesis,” while in male was the “unknown” class (5.59%). It's noteworthy that sequences classified within the “unknown” group are of particular interest, since they presented low or no identity to previously deposited sequences and therefore represent potential novel mosquito sequences.

Further inspection of the “secreted” class revealed the presence of several protein families commonly reported in the sialome of other mosquitoes, including the Antigen-5, the 30 kDa salivary allergen, D7, amylases, maltases, apyrases, peptidases, and peptidases inhibitors (Table 2) [14–16, 32]. In females, the most abundant secreted protein families were the ones implicated in blood acquisition, such as the D7, with 13 CDS accounting for 26.6% of all secreted CDS, and the aegyptin-like/30 kDa salivary allergen (3 CDS representing 18%). In males the “unknown” and “unknown, conserved” classes accounted for ~53.9% of all secreted proteins identified, highlighting the major gap in the knowledge of males' salivary glands contents.

Comparison of the relative CDS quantification between males and females using the edgeR package led to the

**Table 1** Functional classification of transcripts found in the salivary glands of *C. quinquefasciatus* mosquitoes

Class	Females		Males		
	Average No. Transcripts	Average TPM (%)	Average No. Transcripts	Average TPM (%)	No. of DEG
Cytoskeletal	198.5	0.64	199.3	0.69	9
Extracellular matrix/cell adhesion	110	0.57	106.7	0.64	21
Immunity	144	1.46	143.0	2.64	16
Metabolism, amino acid	150.5	0.87	149.0	0.96	14
Metabolism, carbohydrate	150	0.99	149.6	1.28	21
Metabolism, energy	416.5	2.77	416.3	3.32	6
Metabolism, intermediate	89	0.23	88.7	0.21	4
Metabolism, lipid	273	1.31	266.3	1.10	26
Metabolism, nucleotide	91	3.12	88.7	1.01	5
LncRNA	156.5	2.63	156.0	1.65	17
Nuclear export	18	0.02	18.0	0.02	-
Nuclear regulation	167.5	0.51	168.7	1.26	3
Oxidant metabolism/detoxification	137.5	0.65	134.0	0.57	16
Proteasome machinery	304	1.47	302.3	1.51	5
Protein export machinery	424	2.65	418.7	2.17	11
Protein modification machinery	217	2.32	214.3	1.90	10
Protein synthesis machinery	377.5	4.55	375.3	8.77	4
Secreted	847.5	61.21	817.7	56.00	173
Signal transduction	548.5	1.77	543.0	2.04	50
Storage	17	1.89	17.0	0.32	2
Transcription factor	31	0.04	30.7	0.04	1
Transcription machinery	752.5	1.60	753.7	2.20	12
Transporters	290	1.96	281.7	2.36	33
Transposable element	30	0.09	30.7	0.02	6
Unknown	1536.5	4.63	1521.7	5.59	101
Unknown, conserved	27	0.04	27.0	0.05	1

identification of 559 differentially expressed transcripts (LogFC $\geq\pm 2$ , p-value $<0.05$  and FDR $<0.05$ , Supplementary Fig. 1C), from which 398 were up-regulated and 161 down-regulated in females (Supplementary file 2). Additionally, we also identified 17 ncRNA differentially expressed between the two groups (12 up-regulated and 5 down-regulated in females). The functional classification of the differentially expressed CDS is summarized in Table 1, in which the “secreted” class showed the highest number of differentially expressed transcripts (173 CDS) between the two groups. Considering that only female mosquitoes seek blood, the identification of unique or highly abundant CDS in their salivary glands could lead to the identification of molecules containing potent pharmacology activity relevant for blood acquisition and/or related to pathogen transmission. Therefore, in the next section, we present a discussion of the putative secreted salivary protein families that were differentially expressed in female and male mosquitoes.

#### Secreted transcripts differentially expressed in the salivary glands of *C. quinquefasciatus* females and males D7 protein family

The D7 protein family is related to the odorant-binding proteins (OBP) superfamily [33], and was identified

for the first time in the salivary glands of *Aedes aegypti* [34]. Members of this protein family are subclassified based on the number of OBP-like domains they possess; short-forms contain a single domain and often present 15–20 kDa, while the long-forms D7 have two OBP-like domains and range from 24 to 30 kDa. The D7 protein family is commonly found in the salivary gland of mosquitoes where they are among the most abundant protein families [15, 16, 32, 35]. In anopheline mosquitoes the short forms are predominant, while in culicines, transcripts coding for long-D7 are the most abundant ones. The first *C. quinquefasciatus* sialome [14] reported 11 CDS belonging to this family, from which four were subclassified as short-form D7 and seven as long-forms. In the current data set we identified a total of 13 CDS coding for D7-like proteins (8 long, 5 short, Supplementary Table 1), and, as observed in other mosquitoes, most of the D7 proteins were up-regulated in females’ salivary glands (Table 2), reinforcing their role in blood acquisition.

Over the last decade several members of the D7 protein family have been functionally and structurally characterized and shown to act as *kratagonists*, proteins capable of binding and trapping small agonists relevant for host homeostasis such as biogenic amines and leukotrienes

**Table 2** Families of putative secreted proteins within the sialotranscriptome of *C. quinquefasciatus*

Family	Female		Male		No. of DEG	Female TPM / (1 + Male TPM)	Male TPM / (1 + Female TPM)
	No. CDS	Avg.TPM	No. CDS	Avg.TPM			
7.7 kDa	1	1,506.50	1	154.8	1	9.67	0.1
7.8 kDa	3	2,125.20	3	0.3	3	1,585.97	0
9.6 kDa	1	3,202.90	1	0.9	1	1,694.66	0
9.7 kDa	1	484.3	2	0.2	1	413.93	0
13.1 kDa	1	68.7	-	0	1	68.7	0
15.3 kDa	4	34,705.80	2	2.8	4	9,109.13	0
16 kDa	3	74	2	20.3	1	3.47	0.27
16.8 kDa	1	26	1	3.1	-	6.28	0.12
17.5 kDa	1	1,022.30	1	0.1	1	896.75	0
20.2 kDa	1	3,664.50	1	141	1	25.81	0.04
23.4 kDa	1	527.6	1	1,142.40	-	0.46	2.16
27 kDa	4	333.4	4	385.4	-	0.86	1.15
30 kDa	3	97,499.30	3	431.7	3	225.33	0
30.5 kDa	2	2,719.90	2	8,791.10	-	0.31	3.23
34 kDa	1	1,076.10	1	8.2	1	116.97	0.01
41 kDa	1	3,398.50	1	2,898.20	-	1.17	0.85
56 kDa	1	7,906.30	1	13,738.00	-	0.58	1.74
62 kDa	1	837.1			1	837.1	0
Antigen-5	7	39,914.50	7	12,271.70	7	3.25	0.31
Basic-tail	1	942.3	1	30.1	1	30.27	0.03
D7	13	144,368.0	13	281.1	11	511.76	0
gSG1	1	101.7	1	0.1	1	96.86	0
gSG5	2	3,161.20	2	3.5	2	702.49	0
Hormone binding	9	354.8	9	448.9	2	0.79	1.26
Immunity	1	5,062.70	1	14,028.50	-	0.36	2.77
Enzymes							
Acid phosphatase	1	119.5	1	31.5	1	3.68	0.26
Apyrase	4	231.1	4	43.3	2	5.22	0.19
ACE	1	1.5	1	8.6	-	0.15	3.47
Cysteine peptidase	12	2,100.80	15	2,246.70	-	0.93	1.07
Dipeptidyl peptidase	1	95.1	1	53.5	-	1.74	0.56
Prolyl endopeptidase	1	18.8	1	11.1	-	1.55	0.56
Furin	2	12.4	2	13.4	-	0.86	1
Metalloprotease	5	91.3	4	47	1	1.9	0.51
Serine peptidase	124	5,838.40	124	9,498.70	11	0.61	1.63
Other proteases	27	459.6	27	827.1	3	0.56	1.8
Phospholipase	10	538.2	10	536.1	3	1	0.99
Esterases/Lipases	13	4,691.60	13	16,007.60	3	0.29	3.41
Sphingomyelinase	5	2,094.10	5	661.9	-	3.16	0.32
Amylase	6	13466.4	6	53882.4	1	0.25	4
Maltase	4	30051.7	4	97,402	-	0.313	3.24
Protease Inhibitor						0	0
Cystatin	1	32.4	3	81.2	-	0.39	2.43
Kazal	4	320.1	4	1,027.00	-	0.31	3.2
Pacifastin	1	165	1	12.8	1	11.96	0.08
Serpin	19	15,256.00	19	579.1	4	26.3	0.04
TIL	2	8,561.50	2	21,387.40	-	0.4	2.5
TIMP	0	0	1	4.5	-	0	4.53
Leu-rich	10	317.4	10	112	3	2.81	0.35
Mucin	16	21,894.40	16	76,449.60	4	0.29	3.49
OBP	18	481.4	18	4,789.40	8	0.1	9.93

**Table 2 (continued)**

Family	Female		Male		No. of DEG	Female TPM / (1 + Male TPM)	Male TPM / (1 + Female TPM)
	No. CDS	Avg.TPM	No. CDS	Avg.TPM			
Pro-rich	2	3,377.50	2	55.9	1	59.36	0.02
Scoloptoxin	4	343.3	4	353.9	-	0.97	1.03
SGS3	1	4.7	1	0.4	-	3.38	0.07
Unknown, conserved	227	56,737.20	228	71,011.20	48	0.8	1.25
Unknown	211	38,304.40	216		22	0.25	3.92
WRP	12	24,892.40	6	7.2	12	3,031.96	0

[36–40]. Recently, two long-D7 from *C. quinquefasciatus* were described as kratagonists [10]. The CxD7L1 (reported here as XP\_038114211.1) was shown to specifically bind to adenine, adenosine, AMP, ADP, and ATP, while CxD7L2 (reported here as XP\_001865413.2) was shown to interact with biogenic amines (serotonin, histamine, and epinephrine) in addition to cysteinyl leukotrienes (LTC<sub>4</sub>, LTD<sub>4</sub> and LTE<sub>4</sub>). Currently, CxD7L1 is the only D7 protein shown to bind to ADP and ATP [10]. Considering the low levels of apyrases observed *C. quinquefasciatus* salivary glands when compared to other mosquitoes, it is possible that they selected a kratagonist as a platelet aggregation inhibitor, as they would be more efficient in removing low concentrations of ADP (1 μM) than apyrases [41].

Putative short-D7 proteins were also reported to be up-regulated in the salivary glands of females culicine mosquitoes when compared to males [14, 15]. Here we found three short-forms highly expressed in females' salivary glands (LogFC 12.9–13.6, Supplementary file 2), suggesting a potential role in blood acquisition. It's noteworthy that recombinant short-D7 from *Ae. aegypti* and *Ae. albopictus* did not presented the biogenic binding activity found in the long-D7 of culicines [42]. Currently, no short-D7 has been functionally characterized in culicine mosquitoes and their role in mosquito physiology remains elusive. However, it was previously speculated that culicines short-D7 may not interact with biogenic amines since they lack the residues important for such interaction [43].

In addition to their sequestering activity, the role of D7 proteins during viral infection has also been explored. An initial transcriptome comparing the salivary glands of WNV-infected and non-infected *C. quinquefasciatus* females identified a long-D7 down-regulated 14- and 21-days post-infection [44]. In a second study, mice vaccinated with a recombinant long-D7 protein from *C. tarsalis* showed enhanced pathogenesis and higher mortality rates when compared to the control groups when exposed to WNV-infected mosquitoes. Additionally, passive immunization of mice with sera from D7-vaccinated mice also resulted in increased pathogenesis, indicating that such phenotype is related to antibodies targeting the

D7 proteins [45]. The authors speculated that blocking the sequestering activity of D7 proteins would result in increased probing time, and, therefore, more viral particles are injected into the host. Thus, it appears that D7 proteins are not suitable targets for a pathogen-blocking vaccine.

Mapping of the D7 CDS to the genome assembly indicated that most D7 sequences are clustered in the chromosome 2 (CM024711.1, GCA\_015732765.1), with exception of three CDS (XP\_038120412.1, Cq-contig\_13280 and XP\_038111256.1) which are mapped in chromosome 3 (Supplementary file 1). Interesting, XP\_038120412.1 was found to be similar to the juvenile hormone binding D7 from *Ae. aegypti* (Supplementary Fig. 2) [46] and presented an overall low expression in the salivary glands of both males (TPM=27.54) and females (TPM=70.82). Taken together, it's possible the XP\_038120412.1 is also acting as a juvenile hormone binding protein.

### 30 kDa salivary allergen

Members of this protein family are usually among the most abundant CDS in mosquitoes sialomes [47, 48]. In the present study, they were the second most abundant salivary protein family. The previous *C. quinquefasciatus* sialome reported two CDS belonging to this protein family [14], while three CDS were identified in the current dataset, including the most abundant secreted transcript (XM\_001845231) in the female salivary gland with an average TPM of 88,429 (Supplementary file 1). Together, these three CDS accounted for almost 18% of all secreted CDS identified in females (Table 2) and all three were found up-regulated in females (LogFC 3.3–10.8) (Supplementary file 2).

The function of the 30 kDa salivary allergen was initially described in other mosquitoes, *Ae. aegypti* (aegyptin) [49, 50] and *Anopheles stephensi* (AAPP) [51], as molecules capable of interacting with collagen at different binding sites, including the Von Willebrand factor, the glycoprotein VI and the integrin α2β1 [52] binding domains, inhibiting platelet aggregation. Members of this protein family were later characterized in other blood feeding Diptera such as the black fly, *Simulium*

*nigrimanum*. Simplagrin was shown to bind to the collagen von Willebrand factor binding site, however, it failed to block GPVI and integrin  $\alpha 2\beta 1$  binding to collagen [53]. Similarly, the *Ae. albopictus* aegyptin-like protein named AlALP was shown to have an anticoagulant activity by prolonging the activate partial thromboplastin time (APTT), prothrombin time (PT) and thrombin time (TT), however its molecular mechanism remains unknown [54].

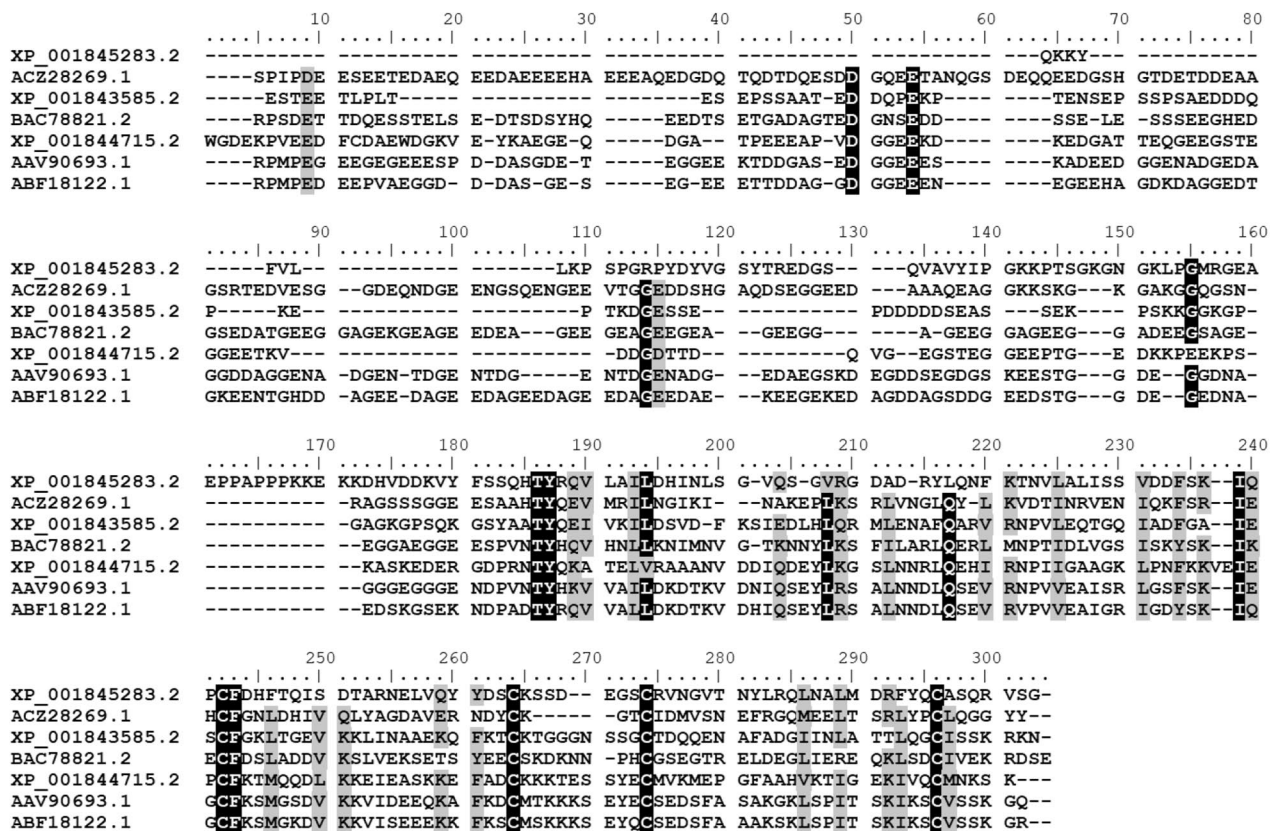
When comparing *C. quinquefasciatus* sequences to those from other Diptera vectors, we found that the predicted mature sequences have shorter N-termini, while some degree of similarity was found in the C-terminal region (Fig. 1). Considering that the C-terminal region is responsible for its collagen binding activity [55], it is likely that *C. quinquefasciatus* aegyptin-like proteins have a similar activity. It is important to note, however, that collagen binding activity has not been demonstrated in *C. quinquefasciatus* saliva or salivary glands homogenates.

Interestingly, although classified within this family, XP\_001845283.2 displays distinct features when compared to other 30 kDa salivary allergens. The predicted mature sequence has a theoretical isoelectric point of

8.94, while other aegyptin-like proteins have rich acidic N-termini conferring them a pI near 4. Tridimensional structure prediction using AlphaFold revealed that the C-terminal is arranged as four packed  $\alpha$ -helices, which is also found in the salivary complement inhibitors of anopheline mosquitoes [56] (Supplementary Fig. 3), while the rest of the protein appears to be intrinsically disordered. In addition to the  $\alpha$ -helices, XM\_001845231 model presents three anti-parallel  $\beta$ -sheets that are not found in other members of this family (Supplementary Fig. 4). Finally, phylogenetic analysis of the 30 kDa salivary allergens positioned the XP\_001845283.2 near simplagrin in a distinct clade from the *Ae. aegypti* and *Ae. albopictus* sequences (Supplementary Fig. 5), suggesting a distant evolutionary relationship between mosquito sequences.

Further biochemical studies are currently underway to uncover its function.

Similar to other salivary proteins, the possible involvement of the 30 kDa salivary allergen during viral transmission has been recently explored. A proteome comparison of DENV-2 infected and non-infected *Ae. aegypti* revealed that aegyptin is 3.5-fold less expressed



**Fig. 1** Amino acid alignment of members of the 30 kDa salivary allergen protein family from *C. quinquefasciatus* (XP\_001845283.2, XP\_001843585.2, XP\_001844715.2), *Simulium nigrimanum* (ACZ28269.1), *Ae. albopictus* (AAV90693.1), *Ae. aegypti* (ABF18122.1) and *An. stephensi* (BAC78821.2). Identical and similar residues are highlighted

in the salivary glands of infected mosquitoes [57]. Additionally, co-inoculation of recombinant aegyptin and DENV-2 in mice resulted in a reduction of DENV-2 particles at 2- and 3-days post-infection [58]. Although their exact mechanism is still unknown, these studies highlight their importance during viral transmission.

### ***β-trefoil Salivary proteins***

Proteins classified within this family had a predicted structure similar to the  $\beta$ -trefoil domain found in ricin (Supplementary Fig. 6). It is important to note that the CDS unified in this section were originally classified into different proteins families such as the tryptophan-rich proteins (WRP), 13.1 kDa, 15.3 kDa, 16 kDa and the 17 kDa protein families. Together, they represented the third most abundant secreted salivary protein group in females.

Among the protein families contained in this section, the 15.3 kDa and WRP protein families were the 6th and 7th most abundant secreted salivary protein families in females, accounting for 6.4% and 4.6%, respectively (Table 2). It is noteworthy that the CDS XM\_038255158.1, classified here as a member of the 15.3 kDa family, was the fourth most abundant CDS in female salivary glands with an average TPM of 32,096 (Supplementary file 1), while only a small fraction was observed in male mosquitoes (average TPM=2.7), suggesting a potential role in blood feeding.

The WRP protein family was reported for the first time in the previous *C. quinquefasciatus* sialome as a novel protein family with no identity to any other known eukaryotic sequence. However, PSI-BLAST analysis indicated their similarities to the bacterial proteins containing the trefoil domain [14]. The later sequencing and annotation of *C. quinquefasciatus* genome revealed the presence of 28 genes belonging to this family [13], and additional sequences over-expressed in female salivary glands were also reported within the recent *C. tarsalis* sialome [30]. Currently, BLASTp of these CDS against the Diptera taxonomy group fails to return significant matches from *Aedes* and *Anopheles* mosquitoes, although similar sequence can be found in the frog biting fly *Corethrella appendiculata* [59] and in the mosquito *Psorophora albipes* [60]. Functionally, only two members of the WRP proteins have been characterized. The CqDVP-2 and CqDVP-4 transcripts were overexpressed in the salivary glands of *C. quinquefasciatus* females. Additional Western blot and immunofluorescence assays confirmed the presence of the native protein in the distal part of the lateral lobes of the salivary gland tissue, while mass spectrometry analysis of the mosquito saliva identified unique peptides from both proteins, confirming that they are indeed secreted into the host. The three-dimensional structure of both proteins was also solved by X-ray

crystallography, confirming the presence of the  $\beta$ -trefoil fold and an initial functional screen suggests that they can bind to carbohydrates [11], however their biological function in blood feeding remains to be elucidated.

### ***Antigen-5***

Antigen-5-like proteins belong to the larger family of cysteine-rich extracellular proteins (CRISP) that are ubiquitously found in animals and plants [61] and are usually the main component of vespid venoms [62, 63]. In the previous *C. quinquefasciatus* sialome, three CDS classified as antigen-5-like proteins were identified [14]. Moreover, Edman degradation confirmed the presence of at least one protein in the salivary gland homogenates, suggesting that it is potentially secreted into the saliva and, consequently, into the host. In our current dataset we observed seven CDS belonging to this family and together they represented the fourth most abundant protein family in female salivary glands (TPM=7.36%) and the 7th most abundant family in males (TPM= $\sim$ 3%). The differential expression gene (DEG) analysis revealed that all seven CDS are differentially expressed between the two groups, with two (XM\_001841639.2 and XM\_001868651.2) transcripts up-regulated and the other five down-regulated in females.

Functionally, the role of salivary antigen-5-like proteins in mosquitoes remains elusive. Their presence in both male and female salivary glands and in the salivary glands of non-blood-feeding mosquitoes indicates they are not uniquely related to blood acquisition [14, 15, 30, 64]. However, evidence that the antigen-5-like proteins can contribute to blood acquisition has been demonstrated elsewhere. In the horse fly, *Tabanus yao*, salivary antigen-5 proteins interfered with platelet aggregation and thrombus formation [65, 66]; however, this was achieved by the acquisition of a disintegrin RGD domain. In *Dipetalogaster maxima* and *Triatoma infestans*, an antigen-5-like protein also inhibited platelet aggregation [67] and possessed a superoxide dismutase activity. Recently, an antigen-5-like salivary protein from *Ae. aegypti* was shown to facilitate Zika virus transmission in a mice model by inducing autophagy [68]. A second study demonstrated that the same protein can bind to the Zika virus envelope protein with high affinity [69], suggesting that salivary antigen-5 proteins might favor viral transmission by different molecular mechanisms.

### ***Serine peptidase inhibitors***

Different classes of serine peptidase inhibitors are commonly described in the sialome of mosquitoes with a high range of TPM values [15, 16, 30]. Functionally, such inhibitors have been associated with the disruption of several biological processes such as host complement activation and blood clotting cascade [1]. Studies with



salivary glands extracts of culicine mosquitoes demonstrate the presence of a specific factor Xa inhibitor that was later isolated and identified as a serpin [70–72]. In the previous *C. quinquefasciatus* sialome two CDS classified as serpins were reported [14], while in the current dataset we identified 19 sequences coding for putative serpins in both male and female mosquitoes (Table 2). Our DEG analysis revealed that four sequences were differentially expressed, from which three were up-regulated (LogFC 4.7–8.4) and one (Cq-contig\_10400) was down-regulated in females (LogFC = -4.1). The latter should be carefully interpreted since the average TPM values of Cq-contig\_10400 in male salivary glands were somewhat low ( $15.2 \pm 6.26$ , Supplementary file 2), and, therefore, the protein might not be present in relevant physiological concentrations.

Interestingly, when comparing the overall expression of serpins in the salivary glands of female mosquitoes, we found that *C. quinquefasciatus* have a higher average TPM values than other culicine mosquitoes such as *Ae. aegypti*, *Ochlerotatus triseriatus* and *C. tarsalis* [15, 16, 30]. Although biochemical characterization is currently unavailable for *C. quinquefasciatus* serpin's, it is very likely that at least one of them acts as an anti-clotting molecule by inhibiting factor Xa activity. Considering that serpins often display high specificity for their targets, it is possible that other serine peptidases are disrupted by the mosquito salivary serpins.

### Apyrases

Apyrases are enzymes that target ATP and ADP catalyzing their degradation to AMP plus orthophosphate [73]. Members of this protein family are currently subclassified into three types: the *Cimex*-type apyrases, the homologs to the human B cell antigen CD-39 and the 5'-nucleotidase family [74]. In mosquitoes, such enzymes belong to the 5'-nucleotidase subfamily [1]. Functionally, apyrases have been shown to inhibit ADP-induced platelet aggregation, thus favoring blood acquisition [75]. Moreover, phylogenetic analysis of apyrases from different blood-feeding arthropods indicates that members of this protein family were convergently selected, highlighting their relevance for the hematophagic behavior [74].

In the current dataset we found three full-length CDS coding for putative apyrases/5'nucleotidases and one truncated CDS in the salivary gland of both males and females *C. quinquefasciatus* (Table 2), while only a single sequence was found in the previous sialome study [14]. The DEG analysis revealed that two of the four CDS were up-regulated in female mosquitoes (LogFC 2.57 and 5.53). Survey of salivary glands activity from different mosquitoes' species (*C. quinquefasciatus*, *Ae. aegypti* and *An. albimanus*) revealed that *C. quinquefasciatus* had the least abundant ADPase activity among them [76].

In line with this observation, we found that the overall expression levels of apyrases from *C. quinquefasciatus* (TPM=231.1) were drastically lower when compared to other female mosquitoes such *Ae. aegypti* (TPM=1,049) [15] and *O. triseriatus* (TPM=6,191)[16], which could explain why only one member of this family was found in the Sanger-based sialome.

Low levels of apyrases were also reported in the recent *C. tarsalis* sialome (RPKM=184.44)[30], suggesting that apyrases activity is also reduced. Considering the preference of *Culex* mosquitoes for avian hosts [77] and that avian thrombocytes (analogous to the mammalian platelets) do not aggregate upon ADP or ATP exposure [78], it is possible that the apyrase genes in culicids are in a path to become a pseudo-gene [30].

### Unknown and unknown conserved

Members of this “functional class” represent CDS that we failed to classify in known protein families. The *unknown* class accounts for CDS that had low or no sequence identity to any previously deposited sequence and, therefore, can be considered potential novel sequences. While the *unknown conserved* class comprises sequences that presented high similarities to sequences of no known function reported elsewhere. In the current dataset both classes combined represented, approximately, 17.5% of all secreted CDS in females and 53.8% in males (Table 2), highlighting the significant knowledge gap between male and female salivary gland composition and function. DEG analysis revealed that 39 CDS were up-regulated in female salivary glands (Supplementary file 2). Among those sequences XP\_038104183.1 was the 9th most abundant transcript coding for a putative salivary protein in female with an average TPM of 10,316.4 (Supplementary file 1). XP\_038104183.1 was found to be almost identical to sequences from the frog biting fly *C. appendiculata* [59] and from *C. pipiens pallens* (Fig. 2). Currently, no function has been attributed to these proteins. However, their abundance in female salivary glands suggests a potential role in blood acquisition and/or pathogen transmission.

It is important to note that since potential novel sequences classified under the *unknown* class were obtained by our *de novo* assembly strategy, we cannot fully reject the possibility that they are artefacts or *chimeric* sequences generated by miss assembly of the Illumina reads. Therefore, the data regarding these CDS must be carefully interpreted and further validated.

### Putative non-coding RNA differentially expressed

In addition to the 559 CDS differentially expressed, our DEG analysis revealed the presence of 12 sequences up-regulated in females and five up-regulated in male mosquitoes currently annotated in *C. quinquefasciatus*



**Fig. 2** Amino acid alignment of unknown proteins from *C. quinquefasciatus* (XP\_038104183.1), *C. pipiens pallens* (XP\_039448873.1) and *C. appendiculata* (XP\_039448873.1). Identical and similar residues are boxed

genome as putative ncRNA (Table 1). The function of ncRNA in mosquitos' physiology can be considered a recent area of study that is undergoing a rapid expansion. The advances of next-generation sequencing technologies in the last decades allowed the identification of thousands of putative ncRNA in different mosquito species, including *An. gambiae* [79], *Ae. aegypti* [80] and *C. quinquefasciatus* [81]. Though much of their function remains largely elusive, it is thought that they have regulatory roles in several biological processes such as transcriptional and post-transcriptional gene regulation, regulation of genomic stability and epigenic regulation [82–84].

The potential role of ncRNA during virus transmission is also under investigation. MicroRNA, 18–24 nucleotides ncRNA, were identified in *Aedes* spp. saliva and found to modulate viral replication in mosquito and mammalian cell cultures [85]. Likewise, microRNAs were also observed in the saliva and salivary glands of *An. coluzzii*, suggesting a role in the vector-host interface [86]. It is noteworthy that the ncRNA reported here are classified as long ncRNA since they are longer than 200 nucleotides. However, due to their abundance in female salivary glands (Supplementary file 1), we cannot discard their possible involvement in blood acquisition and/or pathogen transmission.

## Conclusion

In 2004 a Sanger-based sialome study of adult female *C. quinquefasciatus* was published and reported 503 protein sequences in which 284 were classified as putative secreted proteins [14]. In this work using an Illumina-based approach we reported 7,539 protein sequences from which 898 were classified as secreted proteins, providing a higher resolution of *C. quinquefasciatus* salivary gland contents. The differential expression analysis between males and females showed here allowed the identification of female-enriched putative secreted proteins. The present findings represent a comprehensive reference for future studies focused on the characterization of mosquito salivary proteins assisting in the

identification and characterization of pharmacologically active proteins, proteins that can affect viral transmission and potential biomarkers of mosquito exposure.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09236-1>.

Additional file 1.

## Acknowledgements

We thank Brian Bonilla for excellent technical support and mosquito rearing.

## Author contribution

Stephen Lu: Formal analysis, Investigation, Writing – Original draft & Editing. Ines Martin-Martin: Methodology, Investigation, Writing – Reviewer & Editing. Jose M. Ribeiro: Formal analysis, Investigation, Writing – Reviewer & Editing. Eric Calvo: Conceptualization, Methodology, Writing – Reviewer & Editing, Funding acquisition.

## Funding

This work was supported by the Intramural Research Program of the Division of Intramural Research (AI001246 and AI000810), National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH). This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>). Open Access funding provided by the National Institutes of Health (NIH) Open Access funding provided by the National Institutes of Health (NIH)

## Data availability

The transcriptome data was deposited at the National Center for Biotechnology Information (NCBI) under Bioproject PRJNA892574 and Biosample accession SAMN31387522. The raw reads were deposited at the Short Reads Archive of the NCBI under accessions SRR22053549 - SRR22053554. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GKDO00000000. The version described in this paper is the first version, GKDO01000000.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

All authors have approved the manuscript for submission.

### Competing interests

The author declare that they have no competing interests.

Received: 17 January 2023 / Accepted: 8 March 2023

Published online: 20 March 2023

## References

- Ribeiro JMC, a. A. B. (2009) From Sialomes to the Sialoverse: An Insight into Salivary Potions of Blood-Feeding Insects. in *Advances in Insect Physiology*, ELSEVIER. pp 59–118
- Machain-Williams C, Reagan K, Wang T, Zeidner NS, Blair CD. Immunization with *Culex tarsalis* mosquito salivary gland extract modulates West Nile virus infection and disease in mice. *Viral Immunol.* 2013;26:84–92.
- Thangamani S, Higgs S, Ziegler S, Vanlandingham D, Tesh R, Wikel S. Host immune response to mosquito-transmitted chikungunya virus differs from that elicited by needle inoculated virus. *PLoS ONE.* 2010;5:e12137.
- Styer LM, Lim PY, Louie KL, Albright RG, Kramer LD, Bernard KA. Mosquito saliva causes enhancement of West Nile virus infection in mice. *J Virol.* 2011;85:1517–27.
- Machain-Williams C, Mammen MP Jr, Zeidner NS, Beaty BJ, Prenni JE, Nisalak A, Blair CD. Association of human immune response to *Aedes aegypti* salivary proteins with dengue disease severity. *Parasite Immunol.* 2012;34:15–22.
- Farajollahi A, Fonseca DM, Kramer LD, Marm Kilpatrick A. Bird biting\* mosquitoes and human disease: a review of the role of *Culex pipiens* complex mosquitoes in epidemiology. *Infect Genet Evol.* 2011;11:1577–85.
- Lai CH, Tung KC, Ooi HK, Wang JS. Competence of *Aedes albopictus* and *Culex quinquefasciatus* as vector of *Dirofilaria immitis* after blood meal with different microfilarial density. *Vet Parasitol.* 2000;90:231–7.
- Turell MJ. Members of the *Culex pipiens* complex as vectors of viruses. *J Am Mosq Control Assoc.* 2012;28:123–6.
- Calvo E, Ribeiro JMC. A novel secreted endonuclease from *Culex quinquefasciatus* salivary glands. *J Exp Biol.* 2006;209:2651–9.
- Martin-Martin I, Paige A, Valenzuela Leon PC, Gittis AG, Kern O, Bonilla B, Chagas AC, Ganesan S, Smith LB, Garboczi DN, Calvo E. ADP binding by the *Culex quinquefasciatus* mosquito D7 salivary protein enhances blood feeding on mammals. *Nat Commun.* 2020;11:2911.
- Kern O, Valenzuela Leon PC, Gittis AG, Bonilla B, Cruz P, Chagas AC, Ganesan S, Ribeiro JMC, Garboczi DN, Martin-Martin I, Calvo E. The structures of two salivary proteins from the West Nile vector *Culex quinquefasciatus* reveal a beta-trefoil fold with putative sugar binding properties. *Curr Res Struct Biol.* 2021;3:95–105.
- Suthangkornkul R, Sirichaiyakul P, Sungvornyothin S, Thepouyporn A, Svasti J, Arthan D. Functional expression and molecular characterization of *Culex quinquefasciatus* salivary alpha-glucosidase (Mali). *Protein Expr Purif.* 2015;110:145–50.
- Arensburger P, Megy K, Waterhouse RM, Abrudan J, Amedeo P, Antelo B, Bartholomay L, Bidwell S, Caler E, Camara F, Campbell CL, Campbell KS, Casola C, Castro MT, Chandramouliswaran I, Chapman SB, Christley S, Costas J, Eisenstadt E, Feschotte C, Fraser-Liggett C, Guigo R, Haas B, Hammond M, Hansson BS, Hemingway J, Hill SR, Howarth C, Ignell R, Kennedy RC, Kodira CD, Lobo NF, Mao C, Mayhew G, Michel K, Mori A, Liu N, Naveira H, Nene V, Nguyen N, Pearson MD, Pritham EJ, Puiu D, Qi Y, Ranson H, Ribeiro JM, Roberston HM, Severson DW, Shumway M, Stanke M, Strausberg RL, Sun C, Sutton G, Tu ZJ, Tubio JM, Unger MF, Vanlandingham DL, Vilella AJ, White O, White JR, Wondji CS, Wortman J, Zdobnov EM, Birren B, Christensen BM, Collins FH, Cornel A, Dimopoulos G, Hannick L, Higgs S, Lanzaro GC, Lawson D, Lee NH, Muskavitch MA, Raikhel, A. S., and, Atkinson. P. W. (2010) Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science* 330, 86–88
- Ribeiro JM, Charlab R, Pham VM, Garfield M, Valenzuela JG. An insight into the salivary transcriptome and proteome of the adult female mosquito *Culex pipiens quinquefasciatus*. *Insect Biochem Mol Biol.* 2004;34:543–63.
- Ribeiro JM, Martin-Martin I, Arca B, Calvo E. A deep insight into the sialome of male and female *Aedes aegypti* Mosquitoes. *PLoS ONE.* 2016;11:e0151400.
- Lu S, Martin-Martin I, Ribeiro JM, Calvo E. A deeper insight into the sialome of male and female *Ochlerotatus triseriatus* mosquitoes. *Insect Biochem Mol Biol.* 2022;147:103800.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. ABySS: a parallel assembler for short read sequence data. *Genome Res.* 2009;19:1117–23.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29:644–52.
- Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012;28:3150–2.
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol.* 2004;340:783–95.
- Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohli SAA, Balard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596:583–9.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26:139–40.
- Team RC. (2020) R: A Language and Environment for Statistical Computing.
- Nei MaKS. *Molecular Evolution and Phylogenetics.* Oxford University Press; 2000.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol.* 2018;35:1547–9.
- Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 2019;47:W636–41.
- Hall TA. (1990) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. in *Nucleic acids symposium series*, Oxford University Press
- Tirloni L, Lu S, Calvo E, Sabadin G, Di Maggio LS, Suzuki M, Nardone G, da Silva Vaz I Jr, Ribeiro JMC. Integrated analysis of sialotranscriptome and sialoproteome of the brown dog tick *Rhipicephalus sanguineus* (s.l.): insights into gene expression during blood feeding. *J Proteom.* 2020;229:103899.
- Ribeiro JMC, Martin-Martin I, Moreira FR, Bernard KA, Calvo E. A deep insight into the male and female sialotranscriptome of adult *Culex tarsalis* mosquitoes. *Insect Biochem Mol Biol.* 2018;95:1–9.
- Lu S, Andersen JF, Bosio CF, Hinnebusch BJ, Ribeiro JMC. Integrated analysis of the sialotranscriptome and sialoproteome of the rat flea *Xenopsylla cheopis*. *J Proteom.* 2022;254:104476.
- Francischetti IM, Valenzuela JG, Pham VM, Garfield MK, Ribeiro JM. Toward a catalog for the transcripts and proteins (sialome) from the salivary gland of the malaria vector *Anopheles gambiae*. *J Exp Biol.* 2002;205:2429–51.
- Hekmat-Scafe DS, Dorit RL, Carlson JR. Molecular evolution of odorant-binding protein genes OS-E and OS-F in *Drosophila*. *Genetics.* 2000;155:117–27.
- James AA, Blackmer K, Marinotti O, Ghosn CR, Racioppi JV. Isolation and characterization of the gene expressing the major salivary gland protein of the female mosquito, *Aedes aegypti*. *Mol Biochem Parasitol.* 1991;44:245–53.
- Arca B, Lombardo F, Francischetti IM, Pham VM, Mestres-Simon M, Andersen JF, Ribeiro JM. An insight into the sialome of the adult female mosquito *Aedes albopictus*. *Insect Biochem Mol Biol.* 2007;37:107–27.
- Alvarenga PH, Francischetti IM, Calvo E, Sa-Nunes A, Ribeiro JM, Andersen JF. The function and three-dimensional structure of a thromboxane A2/cysteinyl leukotriene-binding protein from the saliva of a mosquito vector of the malaria parasite. *PLoS Biol.* 2010;8:e1000547.
- Calvo E, Mans BJ, Ribeiro JM, Andersen JF. Multifunctionality and mechanism of ligand binding in a mosquito anti-inflammatory protein. *Proc Natl Acad Sci U S A.* 2009;106:3728–33.
- Martin-Martin I, Smith LB, Chagas AC, Sa-Nunes A, Shrivastava G, Valenzuela-Leon PC, Calvo E. (2020) *Aedes albopictus* D7 Salivary Protein Prevents Host Hemostasis and Inflammation. *Biomolecules* 10
- Martin-Martin I, Kern O, Brooks S, Smith LB, Valenzuela-Leon PC, Bonilla B, Ackerman H, Calvo E. Biochemical characterization of AeD7L2 and its physiological relevance in blood feeding in the dengue mosquito vector, *Aedes aegypti*. *FEBS J.* 2021;288:2014–29.
- Smith LB, Duge E, Valenzuela-Leon PC, Brooks S, Martin-Martin I, Ackerman H, Calvo E. Novel salivary antihemostatic activities of long-form D7 proteins from the malaria vector *Anopheles gambiae* facilitate hematophagy. *J Biol Chem.* 2022;298:101971.
- Francischetti IMB, Andersen JF, Ribeiro JMC. Biochemical and functional characterization of recombinant *Rhodnius prolixus* platelet aggregation inhibitor 1 as a novel lipocalin with high affinity for adenosine diphosphate and other adenine nucleotides. *Biochemistry-U S.* 2002;41:3810–8.

42. Calvo E, Mans BJ, Andersen JF, Ribeiro JM. Function and evolution of a mosquito salivary protein family. *J Biol Chem*. 2006;281:1935–42.
43. Mans BJ, Calvo E, Ribeiro JM, Andersen JF. The crystal structure of D7r4, a salivary biogenic amine-binding protein from the malaria mosquito *Anopheles gambiae*. *J Biol Chem*. 2007;282:36626–33.
44. Girard YA, Mayhew GF, Fuchs JF, Li H, Schneider BS, McGee CE, Rocheleau TA, Helmy H, Christensen BM, Higgs S, Bartholomay LC. Transcriptome changes in *Culex quinquefasciatus* (Diptera: Culicidae) salivary glands during West Nile virus infection. *J Med Entomol*. 2010;47:421–35.
45. Reagan KL, Machain-Williams C, Wang T, Blair CD. (2012) Immunization of mice with recombinant mosquito salivary protein D7 enhances mortality from subsequent West Nile virus infection via mosquito bite. *PLoS Negl Trop Dis* 6, e1935
46. Kim IH, Pham V, Jablonka W, Goodman WG, Ribeiro JMC, Andersen JF. A mosquito hemolymph odorant-binding protein family member specifically binds juvenile hormone. *J Biol Chem*. 2017;292:15329–39.
47. Ribeiro JM, Arca B, Lombardo F, Calvo E, Phan VM, Chandra PK, Wikel SK. An annotated catalogue of salivary gland transcripts in the adult female mosquito, *Aedes aegypti*. *BMC Genomics*. 2007;8:6.
48. Calvo E, Dao A, Pham VM, Ribeiro JM. An insight into the sialome of *Anopheles funestus* reveals an emerging pattern in anopheline salivary protein families. *Insect Biochem Mol Biol*. 2007;37:164–75.
49. Calvo E, Tokumasu F, Marinotti O, Villevial JL, Ribeiro JMC, Francischetti IMB. Aegyptin, a novel mosquito salivary gland protein, specifically binds to collagen and prevents its interaction with platelet glycoprotein VI, integrin  $\alpha 2\beta 1$ , and von Willebrand factor. *J Biol Chem*. 2007;282:26928–38.
50. Chagas AC, Ramirez JL, Jasinskiene N, James AA, Ribeiro JM, Marinotti O, Calvo E. Collagen-binding protein, aegyptin, regulates probing time and blood feeding success in the dengue vector mosquito, *Aedes aegypti*. *Proc Natl Acad Sci U S A*. 2014;111:6946–51.
51. Yoshida S, Sudo T, Niimi M, Tao L, Sun B, Kambayashi J, Watanabe H, Luo E, Matsuoka H. Inhibition of collagen-induced platelet aggregation by anopheline antiplatelet protein, a saliva protein from a malaria vector mosquito. *Blood*. 2008;111:2007–14.
52. Calvo E, Tokumasu F, Mizurini DM, McPhie P, Narum DL, Ribeiro JM, Monteiro RQ, Francischetti IM. Aegyptin displays high-affinity for the von Willebrand factor binding site (RGQQGVMGF) in collagen and inhibits carotid thrombus formation in vivo. *FEBS J*. 2010;277:413–27.
53. Chagas AC, McPhie P, San H, Narum D, Reiter K, Tokomasu F, Brayner FA, Alves LC, Ribeiro JM, Calvo E. Simiplagin, a platelet aggregation inhibitor from *Simulium nigricornis* salivary glands specifically binds to the Von Willebrand factor receptor in collagen and inhibits carotid thrombus formation in vivo. *PLoS Negl Trop Dis*. 2014;8:e2947.
54. Li XP, Lin D, Zhang Y, Chen SQ, Bai HQ, Zhang SN, Liu WQ, Liang SH. Expression and characterization of anticoagulant activity of salivary protein aALP from asian tiger mosquito *Aedes albopictus*. *Trop Biomed*. 2020;37:116–26.
55. Hayashi H, Kyushiki H, Nagano K, Sudo T, Iyori M, Matsuoka H, Yoshida S. Identification of the active region responsible for the anti-thrombotic activity of anopheline anti-platelet protein from a malaria vector mosquito. *Platelets*. 2013;24:324–32.
56. Strayer EC, Lu S, Ribeiro J, Andersen JF. Salivary complement inhibitors from mosquitoes: structure and mechanism of action. *J Biol Chem*. 2021;296:100083.
57. Chisenhall DM, Londono BL, Christofferson RC, McCracken MK, Mores CN. Effect of dengue-2 virus infection on protein expression in the salivary glands of *Aedes aegypti* mosquitoes. *Am J Trop Med Hyg*. 2014;90:431–7.
58. McCracken MK, Christofferson RC, Grasperge BJ, Calvo E, Chisenhall DM, Mores CN. *Aedes aegypti* salivary protein “aegyptin” co-inoculation modulates dengue virus infection in the vertebrate host. *Virology*. 2014;468–470:133–9.
59. Ribeiro JMC, Chagas AC, Pham VM, Lounibos LP, Calvo E. An insight into the sialome of the frog biting fly, *Corethrella appendiculata*. *Insect Biochem Molec*. 2014;44:23–32.
60. Chagas AC, Calvo E, Rios-Velasquez CM, Pessoa FAC, Medeiros JF, Ribeiro JMC. (2013) A deep insight into the sialotranscriptome of the mosquito, *Psorophora albipes*. *Bmc Genomics* 14
61. Schreiber MC, Karlo JC, Kovalick GE. A novel cDNA from *Drosophila* encoding a protein with similarity to mammalian cysteine-rich secretory proteins, wasp venom antigen 5, and plant group 1 pathogenesis-related proteins. *Gene*. 1997;191:135–41.
62. Lu G, Villalba M, Coscia MR, Hoffman DR, King TP. Sequence analysis and antigenic cross-reactivity of a venom allergen, antigen 5, from hornets, wasps, and yellow jackets. *J Immunol*. 1993;150:2823–30.
63. Monsalve RI, Gutierrez R, Hoof I, Lombardero M. Purification and molecular characterization of phospholipase, antigen 5 and hyaluronidases from the venom of the asian hornet (*Vespa velutina*). *PLoS ONE*. 2020;15:e0225672.
64. Calvo E, Pham VM, Ribeiro JM. An insight into the sialotranscriptome of the non-blood feeding *Toxorhynchites amboinensis* mosquito. *Insect Biochem Mol Biol*. 2008;38:499–507.
65. Ma D, Wang Y, Yang H, Wu J, An S, Gao L, Xu X, Lai R. Anti-thrombosis repertoire of blood-feeding horsefly salivary glands. *Mol Cell Proteomics*. 2009;8:2071–9.
66. Ma D, Gao L, An S, Song Y, Wu J, Xu X, Lai R. A horsefly saliva antigen 5-like protein containing RTS motif is an angiogenesis inhibitor. *Toxicol*. 2010;55:45–51.
67. Assumpcao TCF, Ma D, Schwarz A, Reiter K, Santana JM, Andersen JF, Ribeiro JMC, Nardone G, Yu LL, Francischetti IMB. Salivary antigen-5/CAP family members are Cu<sup>2+</sup>-dependent antioxidant enzymes that scavenge O<sub>2</sub>(<sup>-</sup>) and inhibit collagen-induced platelet aggregation and neutrophil oxidative burst. *J Biol Chem*. 2013;288:14341–61.
68. Sun P, Nie KX, Zhu YB, Liu Y, Wu P, Liu ZW, Du SY, Fan HH, Chen CH, Zhang RL, Wang PH, Cheng G. (2020) A mosquito salivary protein promotes flavivirus transmission by activation of autophagy. *Nature Communications* 11
69. Valenzuela-Leon PC, Shrivastava G, Martin-Martin I, Cardenas JC, Londono-Renteria B, Calvo E. (2022) Multiple Salivary Proteins from *Aedes aegypti* Mosquito Bind to the Zika Virus Envelope Protein. *Viruses* 14
70. Stark KR, James AA. Salivary gland anticoagulants in culicine and anopheline mosquitoes (Diptera: Culicidae). *J Med Entomol*. 1996;33:645–50.
71. 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;273:20802–9.
72. Calvo E, Mizurini DM, Sa-Nunes A, Ribeiro JM, Andersen JF, Mans BJ, Monteiro RQ, Kotsyfakis M, Francischetti IM. Albo-serpin, a factor xa inhibitor from the mosquito vector of yellow fever, binds heparin and membrane phospholipids and exhibits antithrombotic activity. *J Biol Chem*. 2011;286:27998–8010.
73. Plesner L. Ecto-ATPases: identities and functions. *Int Rev Cytol*. 1995;158:141–214.
74. Hughes AL. Evolution of the salivary apyrases of blood-feeding arthropods. *Gene*. 2013;527:123–30.
75. Sun D, McNicol A, James AA, Peng Z. Expression of functional recombinant mosquito salivary apyrase: a potential therapeutic platelet aggregation inhibitor. *Platelets*. 2006;17:178–84.
76. Ribeiro JMC. Blood-feeding in mosquitoes: probing time and salivary gland anti-haemostatic activities in representatives of three genera (*Aedes*, *Anopheles*, *Culex*). *Med Vet Entomol*. 2000;14:142–8.
77. Gonzalez MA, Prosser SW, Hernandez-Triana LM, Alarcon-Elbal PM, Goiri F, Lopez S, Ruiz-Arondo I, Hebert PDN, Garcia-Perez AL. (2020) Avian Feeding Preferences of *Culex pipiens* and *Culiseta* spp. Along an Urban-to-Wild Gradient in Northern Spain. *Front Ecol Evol* 8
78. Stiller RA, Belamarich FA, Shepro D. Aggregation and release in thrombocytes of the duck. *Am J Physiol*. 1975;229:206–10.
79. Jenkins AM, Waterhouse RM, Muskavitch MAT. (2015) Long non-coding RNA discovery across the genus *Anopheles* reveals conserved secondary structures within and beyond the *Gambiae* complex. *Bmc Genomics* 16
80. Etebari K, Asad S, Zhang GM, Asgari S. (2016) Identification of *Aedes aegypti* Long Intergenic Non-coding RNAs and Their Association with *Wolbachia* and Dengue Virus Infection. *Plos Neglect Trop D* 10
81. Azlan A, Halim MA, Mohamad F, Azzam G. Identification and characterization of long noncoding RNAs and their association with acquisition of blood meal in *Culex quinquefasciatus*. *Insect Sci*. 2021;28:917–28.
82. Xu Y, Dong YQ, Xu YZ, Lai ZT, Jin BB, Hao YQ, Gao YH, Sun Y, Chen XG, Gu JB. (2019) Differentiation of Long Non-Coding RNA and mRNA Expression Profiles in Male and Female *Aedes albopictus*. *Front Genet* 10
83. Belavilas-Trovas A, Gregoriou ME, Tastsoglou S, Soukia O, Giakountis A, Mathiopoulou K. (2022) A species-specific lncRNA modulates the reproductive ability of the asian tiger mosquito. *Front Bioeng Biotech* 10
84. Mysore K, Hapairai LK, Li P, Roethele JB, Sun L, Igiede J, Misenti JK, Duman-Scheel M. A functional requirement for sex-determination M/m locus region lncRNA genes in *Aedes aegypti* female larvae. *Sci Rep*. 2021;11:10657.
85. Maharaj PD, Widen SG, Huang J, Wood TG, Thangamani S. (2015) Discovery of Mosquito Saliva MicroRNAs during CHIKV Infection. *Plos Neglect Trop D* 9
86. Arca B, Colantoni A, Fiorillo C, Severini F, Benes V, Di Luca M, Calogero RA, Lombardo F. MicroRNAs from saliva of anopheline mosquitoes mimic human endogenous miRNAs and may contribute to vector-host-pathogen interactions. *Sci Rep*. 2019;9:2955.

**Publisher's note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.