HHS Public Access

Author manuscript

Insect Biochem Mol Biol. Author manuscript; available in PMC 2019 April 01.

Published in final edited form as:

Insect Biochem Mol Biol. 2018 April ; 95: 1–9. doi:10.1016/j.ibmb.2018.03.001.

A deep insight into the male and female sialotranscriptome of adult *Culex tarsalis* mosquitoes

José M.C. Ribeiro[a], **Ines Martin-Martin**[a], **Fernando R. Moreira**[b], **Kristen A. Bernard**[b], and **Eric Calvo**[a]

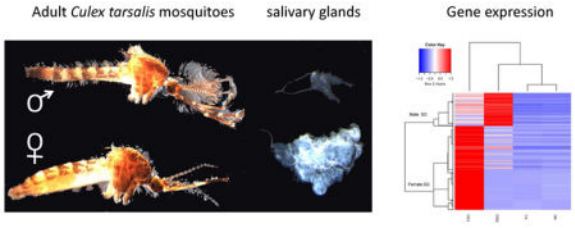
[a]Section of Vector Biology, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, 12735 Twinbrook Parkway room 2E32D, Rockville MD 20852

[b]Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin-Madison, 2015 Linden Dr., Madison WI 53706

Abstract

Previously, a Sanger-based sialotranscriptome analysis of adult female *Culex tarsalis* was published based on ~ 2,000 ESTs. During the elapsed 7.5 years, pyrosequencing has been discontinued and Illumina sequences have increased considerable in size and decreased in price. We here report an Illumina-based sialotranscriptome that allowed finding the missing apyrase from the salivary transcriptome of *C. tarsalis*, to determine several full-length members of the 34–62 kDa family, when a single EST has been found previously, in addition to identifying many salivary families with lower expression levels that were not detected previously. The use of multiple libraries including salivary glands and carcasses from male and female organisms allowed for an unprecedented insight into the tissue specificity of transcripts, and in this particular case permitting identification of transcripts putatively associated with blood feeding, when exclusive of female salivary glands, or associated with sugar feeding, when transcripts are found upregulated in both male and female glands.

Graphical Abstract



Keywords

Mosquitoes; salivary glands; transcriptome; blood-feeding

Corresponding author: José M.C. Ribeiro jribeiro@niaid.nih.gov.

1. Introduction

Saliva of adult female mosquitoes contains antihemostatic and immunomodulatory components that helps blood feeding by preventing host reactions that could affect skin blood flow or induce defensive behavior (Ribeiro et al., 2010). Male mosquitoes do not blood feed and this is reflected in their much smaller salivary glands (Juhn et al., 2011) that lack antihemostatic components but have antimicrobial and sugar-degrading enzymes that may help sugar feeding and its storage in the crop (Marinotti et al., 1990; Rossignol and Lueders, 1986). Perhaps due to their pharmacological activities, mosquito saliva components have been implicated in favoring arbovirus transmission from mosquitoes to their vertebrate hosts (Briant et al., 2014; Londono-Renteria et al., 2013; Machain-Williams et al., 2012; Machain-Williams et al., 2013; Styer et al., 2011; Thangamani et al., 2010; Wichit et al., 2016).

The mosquito *Culex tarsalis* is a vector of West Nile virus (Reisen et al., 2005) and its saliva has been implicated in altering virus transmission (Machain-Williams et al., 2013; Styer et al., 2011). A limited Sanger-based sialotranscriptome (from the Greek “sialo” = saliva) was previously described for adult females of this mosquito (Calvo et al., 2010a). We here disclose a more complete transcriptome of both male and female adult *C. tarsalis* based on the “de novo” assembly of over 266 million reads obtained by the Illumina technology. Transcriptomes were also assembled from carcasses without salivary glands of adult male and female mosquitoes. A total of 16,246 peptide and their coding sequences were publicly deposited. Comparisons of the level of reads accrued by deducted coding sequences from the different libraries allowed for identification of salivary specific transcripts, and among these, of sex-specific salivary transcripts. These results should add in the identification of salivary proteins that affect virus transmission, in the discovery of immunological markers of vector exposure, and in the identification of pharmacologically active salivary proteins of *C. tarsalis.*

2. Material and methods

This section is very similar to our previous publication (Ribeiro et al., 2016) and will be reproduced nearly verbatim for reference.

2.1. Mosquitoes and salivary gland dissection

The *C. tarsalis* colony was established in 2014 from mosquitoes collected in 2002 from the Kern National Wildlife Refuge in California (kind gift of Dr. William Reisen, University of California-Davis). Mosquitoes were maintained on a 18:6 hour light:dark cycle; food (10% sucrose) and water were provided *ad libitum*.

Seven- to ten-day-old male and female *C. tarsalis* mosquitoes were immobilized by placing them at −20°C for 5 minutes and then transferred into a cold glass petri dish on ice. Wings and legs were gently removed, and mosquitoes were dipped in 70% ethanol and transferred onto RNase-free phosphate-buffered saline (PBS) on a glass slide. Using a dissecting scope, the head was removed, and the salivary glands were dissected from the thorax by applying

pressure with dissecting pins. Four hundred female salivary glands and three hundred male salivary glands were collected and placed in 500 μL and 150 μL of RNA*later* Solution® (Ambion-ThermoFisher, Carlsbad, CA), respectively, in a 1.5 mL RNase-free Eppendorf tube. The salivary gland preparations were kept on ice during the entirety of the dissection, and upon completion maintained at −20°C for 12 hours before freezing at −80°C. For the carcasses, all lobes of both salivary glands were completely removed from ten males and ten female mosquitoes as described above. Partial thoraces, head, and abdomen (referred to as carcass) for male or female mosquitoes were placed in 500 μL RNA*later* Solution®, kept on ice during the entirety of the dissection, and maintained at 4°C for 48 hours prior to freezing at −80°C.

2.2. RNA preparation

Total RNA from male salivary glands, female salivary glands and carcasses of both sexes (abbreviated as FSG, MSG, FC and MC) were extracted using an RNeasy mini total RNA isolation kit and manufacturer's protocol (Qiagen, USA).

2.3. cDNA library construction and sequencing

Tissue samples were submitted to the North Carolina State Genomic Sciences Laboratory (Raleigh, NC, USA) for Illumina RNA library construction and sequencing. Prior to library construction, RNA integrity, purity, and concentration were assessed using an Agilent 2100 Bioanalyzer with an RNA 6000 Nano Chip (Agilent Technologies, USA). Purification of messenger RNA (mRNA) was performed using the oligo-dT beads provided in the NEBNExt Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, USA). Complementary DNA (cDNA) libraries for Illumina sequencing were constructed using the NEBNext Ultra Directional RNA Library Prep Kit (NEB) and NEBNext Mulitplex Oligos for Illumina (NEB) using the manufacturer-specified protocol. Briefly, the mRNA was chemically fragmented and primed with random oligos for first strand cDNA synthesis. Second strand cDNA synthesis was then carried out with dUTPs to preserve strand orientation information. The double-stranded cDNA was then purified, end repaired and "a-tailed" for adaptor ligation. Following ligation, the samples were selected a final library size (adapters included) of 400–550 bp using sequential AMPure XP bead isolation (Beckman Coulter, USA). Library enrichment was performed and specific indexes for each sample were added during the protocol-specified PCR amplification. The amplified library fragments were purified and checked for quality and final concentration using an Agilent 2100 Bioanalyzer with a High Sensitivity DNA chip (Agilent Technologies, USA). The final quantified libraries were pooled for sequencing on one lane of an Illumina HiSeq 2500 DNA sequencer, utilizing 125 bp single end sequencing flow cell with a HiSeq Reagent Kit v4 (Illumina, USA). To increase representation of the salivary gland libraries, and specifically of the FSG library, the libraries were sequenced with the following representation: 3 x FSG, 2 x MSG, 1 x FC and 1 x MC. Flow cell cluster generation for the HiSeq2500 was performed using an automated cBot system (Illumina, USA). The software package Real Time Analysis (RTA), version 1.18.64, was used to generate raw bcl, or base call files, which were then de-multiplexed by sample into fastq files for data submission using bcl2fastq2 software version v2.16.0.

2.4. Bioinformatic analysis

Bioinformatic analyses were conducted following methods described previously (Chagas et al., 2013; Ribeiro et al., 2014), with some modifications. Briefly, the fastq files were trimmed of low quality reads (<20), removed from contaminating primer sequences and concatenated for single-ended assembly using the Abyss (using k parameters from 21–91 in 5 fold increments) (Birol et al., 2009) and Trinity (Grabherr et al., 2011) assemblers. The combined fasta files were further assembled using an iterative blast and CAP3 pipeline as previously described (Karim et al., 2011). Coding sequences (CDS) were extracted based on the existence of a signal peptide in the longer open reading frame (ORF) and by similarities to other proteins found in the Refseq invertebrate database from the National Center for Biotechnology Information (NCBI), proteins from Diptera deposited at NCBI's Genbank and from SwissProt. Reads for each library were mapped on the deducted CDS using blastn with a word size of 25, 1 gap allowed and 95% identity or better required. Up to five matches were allowed if and only if the scores were the same as the largest score. A X^2 test was performed for each CDS to detect statistically significant differences between the number of reads in paired comparisons. Bonferroni and the false discovery rate (FDR) corrections of Benjamini & Hockberg (Benjamini and Hochberg, 1995) were done using the p.adjust program from the stats package version 3.3.0 which is part of the core R package (Team, 2013). The results of these tests are mapped to hyperlinked excel sheets presented as Supplemental File S1 following column AP on worksheet named "C_tarsalis FPKM>5". The normalized ratio of the reads for paired comparisons was calculated as r1 x R2/[R1 x (r2 +1)] and r2 x R1/[R2 x (r1 +1)] where r1 and r2 are reads for libraries 1 and 2, and R1 and R2 are total number of reads from libraries 1 and 2 mapped to all CDS. One was added to the number of reads in the denominator to avoid division by zero. To compare transcript relative expression among contigs, we use the "expression index" (EI) defined as the number of reads mapped to a particular CDS multiplied by 100 and divided by the largest found number of reads mapped to a single CDS (Chagas et al., 2013), which in the case of the FSG transcriptome was a value of 3,289,242 reads mapped to the transcript coding for the salivary GQP-rich peptide, and thus having an EI=100. This same transcript yielded also a value of 100 for the MSG transcriptome. Fragments per thousand nucleotides per million reads (FPKM) (Howe et al., 2011) and transcripts per million (TPM) for the four mapped libraries were calculated and recorded in the spreadsheet (Wagner et al., 2012). Heatmap graphs were done with the program heatmap2 from the gplots package running within R package with default parameters and using Z scores for data normalization (Warnes et al., 2015).

2.5. Data availability

This project was registered at the National Center for Biotechnology Information (NCBI) under the accession PRJNA360148. The raw reads can be retrieved with the accessions SRR5149155, SRR5149158, SRR5149170 and SRR5149179. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GFDL00000000. The version described in this paper is the first version, GFDL01000000.

3. Results and discussion

3.1. General aspects of the assembly

After removal of Illumina primers and trimming of low quality bases, we obtained 115,059,061 reads for the FSG, 74,367,851 reads for the MSG, 39,297,567 reads for the FC and 37,772,770 reads for the MC libraries, with an average length and median of 124 and 125 nt, respectively. Notice that the libraries were applied to the sequencer in larger amounts for the FSG, followed by the MSG and the carcass libraries (3 × 2 × 1 × 1). These reads were assembled using Abyss with various kmer sizes and Trinity assemblers. The resulting assemblies were then assembled together using a pipeline of iterative and parallelized blastn and CAP3 where blastn with decreasing word sizes (from 300 to 60) fed the cap3 assembler through 15 iterations producing 24,772 putative coding sequences with a FPKM equal or larger than 5 (Supplemental File S1). While many of the contigs were truncated fragments, we submitted over 16 thousand sequences that are over 50% full length (based on their similarities to available mosquito proteins) to the NCBI.

3.2. Transcripts overexpressed in the salivary glands

When the combined reads of male and female salivary glands were compared to the combined reads deriving from the carcass libraries, 668 transcripts were found to be at least 10 x overexpressed in the salivary glands when compared by a X^2 test. A heat map of the relative expression of these transcripts clearly shows a subset that is overexpressed primarily on FSG, and a smaller subset that is mainly expressed in MSG (supplemental Fig. 1). As expected, the SG enriched subset has an overrepresentation of transcripts associated with secreted products (Table 1), which constitutes 55% of the enriched products, compared to 11 % of the totality of transcripts. Since there was a previous description of the *C. tarsalis* sialotranscriptome (Calvo et al., 2010a), we will highlight in this manuscript only the advances made following the deep-sequence assembly and the comparisons between male and female mosquitoes.

3.3. Transcripts overexpressed in the female salivary glands

Of the 668 transcripts that are at least 10 x overexpressed in SG, 332 of these are at least 10 x overexpressed in female SG when contrasted with male SG. Of these 332 transcripts, 218 are considered of a Secreted class while the remaining are of a putative housekeeping role, including 8 transposable elements.

3.3.1. Putative secreted transcripts overexpressed in the female salivary glands—The putative secreted gene products enriched in female salivary glands include enzymes, a serpin, antigen-5 family members, and other insect and mosquito specific families (Supplemental spreadsheet and Table 2).

Among the enzymes, at least 4 apyrase family members were detected, which were remarkably absent on the previous Sanger-based transcriptome. These four transcripts had EI values of less than 0.55, contrasting with values of 8–40 found for the homologous genes in a *Aedes aegypti* transcriptome analysis (Ribeiro et al., 2016). These low expression values may have precluded their discovery using the Sanger approach. One transcript coding for an

apyrase/5′ nucleotidase was also found expressed similarly in the salivary glands of both sexes. This transcript had a much lower EI (0.021 in females and 0.014 in males). It does not have a GPI anchor thus excluding it as a housekeeping 5′-nucleotidase. It is most similar to mosquito proteins named as apyrases. We do not know the function of this transcript in male salivary glands. Its low EI may indicate a gene that is in the path to become a pseudo-gene.

Apyrases are enzymes that degrade ATP and ADP and have anti-neutrophil and anti-platelet activities (Ribeiro et al., 2010; Ribeiro and Arca, 2009). In mosquitoes, these enzymes belong to the 5′-nucleotidase family (Champagne et al., 1995), and in *An. gambiae* two members are overexpressed in the salivary glands, one annotated as apyrase and the other annotated as 5′-nucleotidase (Arca et al., 1999a; Arca et al., 1999b; Lanfrancotti et al., 2002). In *Aedes, Psorophora* and *Culex quinquefasciatus* multiple members were also found (Arca et al., 2007; Arensburger et al., 2010; Chagas et al., 2013; Ribeiro et al., 2007; Ribeiro et al., 2016). Supplemental Fig. 2 shows the bootstrapped phylogram of the four *C. tarsalis* sequences and their close relatives as determined by a blastp search. The graph is suggestive of two original paralog genes, one giving origin to the salivary apyrase of *A. aegypti* (AAA99189.1) (Champagne et al., 1995; Smartt et al., 1995) and the other to a salivary 5′-nucleotidase (ABF18486.1) (Ribeiro et al., 2007).

The previous *C. tarsalis* sialome identified fragments of contigs coding for the enzymes endonuclease and adenosine deaminase, which are now fully disclosed following deep-sequencing. JAV30380.1 (EI = 0.2) is 93% identical with a *C. quinquefasciatus* endonuclease homolog, while JAV29973.1 (EI=0.2) is an adenosine deaminase 84% identical to its closest *C. quinquefasciatus* homolog. Another protein, coded by JAV26958.1 (EI=1.2) is 86% identical to another *C. quinquefasciatus* homolog, indicating the existence of two genes coding for salivary adenosine deaminases in *C. tarsalis*.

Three nearly full length contigs coding for a female SG-enriched hyaluronidase were found following deep sequencing, all with EI values smaller than 0.7. This enzyme was not previously found in the limited *C. tarsalis* sialotranscriptome, but was found in the sialome of *C. quinquefasciatus* (Ribeiro et al., 2004) and is commonly found in sand fly sialotranscriptomes (Charlab et al., 1999; Hostomska et al., 2009; Rohousova et al., 2012; Vlkova et al., 2014), where the enzyme activity was determined (Cerna et al., 2002). It has been postulated that the combined action of endonucleases and hyaluronidases decrease the viscosity of the skin matrix facilitating diffusion of salivary anti-hemostatic components, as well as facilitating pathogen transmission (Calvo and Ribeiro, 2006; Chagas et al., 2014a; Volfova et al., 2008).

Two distinct sphingomyelinases (EI values range 0.7 – 1.3) were found with over 1,000-fold enrichment in female salivary glands. This is the first finding of this enzyme on a mosquito sialome. In ticks, a salivary sphingomyelinase was implicated in driving the immune response to a Th2 phenotype by inducing CD4+ T cells to produce IL-4 (Alarcon-Chaidez et al., 2009).

The previous *C. tarsalis* sialotranscriptome identified many EST's coding for serine proteases, which is a common finding in mosquito sialomes. We currently identify the full-

length protein encoded by JAV19159.1, which is remarkably 400 x enriched in female glands, while in *Ae. aegypti* no salivary female-specific serine protease was found (Ribeiro et al., 2016). In *Tabanus yao* several salivary serine proteases have been found to have fibrinolytic activity (Ma et al., 2009). The role of *C. tarsalis* salivary serine protease remains to be determined.

Previously, a single EST was found in a *C. tarsalis* salivary gland library coding for a partial serpin. This serine protease inhibitor is here disclosed in its full length, coded by JAV25113.1 (EI=1.5). This protein may function as the salivary anti-clotting of *C. tarsalis*, similarly to the salivary serpins of *Ae. aegypti* and *Ae. albopictus* (Calvo et al., 2011; Stark and James, 1998). Serpins are often specific for their targets and this transcript product should not interfere with the *C. tarsalis* salivary serine protease described above.

Members of the aegyptin/30 kDa antigen family, as well as members of the D7, 10–14 kDa, 20.2 kDa, 62/34 kDa, gSG5, gSG8, Hyp 37, SGS1, 16.7 kDa/WRP, 4.2 kDa, 9.7 kDa, Basic tail, salivary mucin and proline rich families were found enriched in female salivary glands as contrasted with male organs, indicating a main function of these proteins in assisting hematophagy. Many of these families are represented by transcripts with large EI values, such as the D7 including EI ranging from 5–52, or aegyptin, with EI values ranging 3.5–14. Seven of these families have not been described in the previous *C. tarsalis* sialotranscriptome (Calvo et al., 2010a), and are marked with a * in Table 2. With exception of the aegyptin/30 kDa family, which inhibits collagen induced aggregation (Calvo et al., 2007; Calvo et al., 2010b; Chagas et al., 2014b; Mizurini et al., 2013), and the D7 family, which are multifunctional proteins primarily binding agonists of hemostasis (Calvo et al., 2006; Calvo et al., 2009; Mans et al., 2007), the remaining 13 families have unknown function.

The 16.7 kDa family, also known as tryptophan rich protein (WRP) was first discovered in the *C. quinquefasciatus* sialome (Ribeiro et al., 2004), representing an expanded protein family of salivary expressed proteins without similarities to any other known eukaryotic protein. However, psiblast analysis revealed similarities to bacterial proteins containing a trefoil domain similar to ricin (Ribeiro et al., 2004). Subsequent disclosure of the *C. quinquefasciatus* genome lead to identification of a total of 28 genes coding for members of this family. Interestingly, most of these genes are uniexonic, supporting a horizontal transfer hypothesis for the appearance of these unique genes in *Culex* (Arensburger et al., 2010). A sialotranscriptome of *C. tarsalis* revealed additional members of this family (Calvo et al., 2010a). Somewhat surprisingly, additional members of this family were found within a deep-sequence sialotranscriptome of the frog biting fly, *Corethrella appendiculata* (Ribeiro et al., 2014) and the mosquito *Psorophora albipes* (Chagas et al., 2013), but not deep-sequencing of the sialome of *Ae. aegypti* (Ribeiro et al., 2016). Supplemental Fig. 3 displays a phylogram of the family, where 3 super clades are evident with strong bootstrap support. The more numerous super clade I contains proteins of both *Culex* species as well as *Corethrella*, while the remaining clades are exclusive of *Culex*. Within each super clade there are strong multi-species clades, indicative of a long parallel evolutionary history.

Another mosquito salivary family possibly acquired by horizontal transfer is represented by the 62/34 kDa protein family, which are also most uniexonic in *Ae. aegypti* (Ribeiro et al.,

2007). This family was previously found only in *Aedes* mosquitoes, including *Ae. albopictus* (Arca et al., 2007), but was found later in *Psorophora albipes* (Chagas et al., 2013) and in *Corethrella appendiculata* (Ribeiro et al., 2014), from where a sequence represented a "missing link" joined the 62/34 kDa family with the 16.7 kDa family after psiblast analysis. The limited Sanger-based sialotranscriptome of *C. tarsalis* identified a single truncated transcript representing a member of the 62 kDa family. The present deep sequencing work identified 11 full length transcripts of the 62/34 kDa family, with ranging EI values of 0.2–2, which probably derive from at least eight different genes. Supplemental Fig. 4 shows a phylogram of the family with 4 major clades with strong bootstrap support. Clade I include all 62 kDa members of *Aedes* and *Psorophora* mosquitoes, each within strong sub-clades. The remaining clades are exclusive of *Culex* mosquitoes, clade II having 62 kDa members of both *Culex* species, clade III has only *C. tarsalis* sequences, and clade IV representing 34kDa members of both *Culex* species.

When the number of reads accrued by transcripts that are overexpressed at least 10x on FSG is computed (supplemental table 1), it is observed that near 92% of 59.5 million reads map to transcripts coding for putative secreted proteins. Of these reads, 44% map to members of the D7 protein family, followed by 32% mapping to members of the 16.7/WRP family (supplemental table 2). These results contrast with a previous Sanger-based sialotranscriptome of *C. tarsalis* (Calvo et al., 2010a) where 9% of the EST's mapped to members of the D7 family while 58% mapped to members of the 16.7/WRP family. In any case, the abundant transcription of the 16.7/WRP family, still of unknown function, suggests a kratagonist role for some relatively abundant agonist such as serotonin or histamine, as proposed before (Calvo et al., 2010a).

3.3.2 Putative housekeeping transcripts overexpressed in the female salivary glands—One hundred and six transcripts attributed to the housekeeping class were found overexpressed in female salivary glands (Table 2). These transcripts may be associated with specialized functions of the female gland, such as the presence of two subunits of the prolyl 4-hydroxylase enzyme, which may indicate the presence of modified proline-containing peptides in the salivary secretion of *C. tarsalis*. Of notice, collagen abounds with hydroxy-proline residues (Kivirikko and Pihlajaniemi, 1998). Perhaps salivary peptides mimicking collagen may also contain hydroxy proline. Several transporters of the Major Facilitator Superfamily are also identified (Quistgaard et al., 2016), and may be related with the fast secretion of saliva needed during blood feeding.

3.4. Transcripts overexpressed nearly equally in adult salivary glands

A total of 225 transcripts were found overexpressed in adult salivary glands when compared to carcass, and were similarly expressed in both sexes, indicating they belong to a salivary function shared by both sexes (Table 3). The majority (68.4%) of these transcripts are of a putative secretory nature, including enzymes, protease inhibitors, antimicrobial peptides and other immune-related products, mucins, antigen-5 proteins and nematocera-specific proteins of unknown function.

3.4.1 Putative secreted transcripts overexpressed nearly equally in adult salivary glands—Somewhat surprisingly, a member of the 5′-nucleotidase/apyrase family is represented in this group, although with a low RPKM (7.7 on Female SG and 8.0 on Male SG), and low EI (0.02%), while the female expressed enzymes of the same class had EI's ranging from 0.3–0.5%. Endonucleases were also relatively well expressed in the SG of both sexes, with EI values of 0.2–4 on females and .1–1 on male SG. Amylase/maltase are well expressed in both sexes, with large EI values (49–87%), consistent with their roles on sugar feeding. In *Ae. aegypti,* the amylase transcript had EI=100 in both sexes (Ribeiro et al., 2016). Lipases and chitinases were found similarly overexpressed in both male and female SG, as were many different serine proteases. Chitinases may be related to anti-microbial function (Theis and Stahl, 2004), as may be the peptides with TIL and thyroglobulin anti-protease motifs (Augustin et al., 2009). Anti-microbial peptides of the gambicin (Vizioli et al., 2001) and His-rich/GQP rich (De Smet and Contreras, 2005; Lai et al., 2004) families are well expressed. Remarkably, this latter peptide is the most highly expressed contig in both male and female SG, thus having an EI of 100%. Pathogen pattern-recognition proteins such as C type lectins, Gram negative binding proteins and Fred domain containing proteins may tag ingested microbes for further dealing with the mosquito innate immune system.

Peptides of the families 23.5 kDa culicine family, 56 kDa mosquito family, 41 kDa family, 30.5 kDa culicine family, hyp10-hyp12 anopheline family (also known as 7.8 kDa basic salivary peptide family), the W-rich family of culicines, the 6.8 kDa culicine family and the novel 6–8 kDa families are found overexpressed similarly in both male and female SG suggesting a role in sugar feeding or in immunity, yet to be determined. The novel family 6–8 kDa codes for peptides of mature MW ranging from 6–8 kDa, with acidic pI and having the motif M-x-S-x(4)-C-x(5)-S-x(5)-G-N-x(2)-P-x(6)-G-x-S-S-x(6)-P-x(6,7)-I-x(8)-N-x(20,40)-W-F. No similar peptides are found in other organisms.

Remarkably by its absence from this group are members of the lysozyme family. Although 10 transcripts coding for this family were found in the assembled transcriptome, none are more than 10 x overexpressed in either SG or carcass. In *Ae. aegypti* four transcripts coding for lysozyme peptides were found overexpressed in the SG, but just at a 11–12 fold level salivary overexpression (Ribeiro et al., 2016).

3.4.2 Putative housekeeping transcripts overexpressed nearly equally in adult salivary glands—Among the putative housekeeping transcripts overexpressed in the salivary glands of *C. tarsalis* irrespective of sex we highlight a bZIP transcription factor that might be related to the differentiation of adult salivary glands. The encoded protein is 48% similar to the *Drosophila melanogaster* protein cap'n'collar which has been implicated in the development of cephalic segments (Mohler et al., 1995). The *An. gambiae* homolog was also found expressed in adult female salivary glands (Arca et al., 2005).

3.5. Transcripts overexpressed in the adult male salivary glands

Only 44 transcripts were found overexpressed in male salivary glands when contrasted to their female counterparts (Table 4). Of these, only 6 were classified of a secreted nature, including one antigen-5 member. The remaining 5 transcripts have no similarities to known

proteins. Somewhat surprisingly, we found a transcript coding for a product with 94% identity to the capsid protein of *Bat sobemovirus* (Dacheux et al., 2014; Li et al., 2010). It has a relatively low expression level, with RPKM of 6.3 in male and 0.21 in female salivary glands.

3.6. Transcripts overexpressed in the adult female carcass when compared to adult male carcass, excluded of enriched SG transcripts

C. tarsalis are autogenous mosquitoes (Eberle and Reisen, 1986; Provost-Javier et al., 2010) and perhaps for this reason most of the non-salivary transcripts enriched in female bodies appear to be associated with egg development. The top transcript, annotated as a phosphatidylcholine-sterol acyltransferase may be associated with the intense lipid metabolism associated with oogenesis. The maternal effect protein Oskar (Juhn and James, 2006), which helps to stabilize the Oskar RNA at the posterior pole is highly expressed. Several histones, serine proteases and chorion peroxidase are noticeable. The supplemental spreadsheet file has a worksheet named FC 10 x MC that has 596 transcripts available for further mining.

3.7. Transcripts overexpressed in the adult male carcass when compared to adult female carcass, excluded of enriched SG transcripts

The supplemental spreadsheet file has a worksheet named MC 10 x FC with 949 transcripts that are overexpressed in MC exclusive of salivary enriched transcripts. The top most expressed transcripts include those coding for the neuropeptides allatostatin and LWamide which may be associated to male feeding and courtship behaviors. While allatostatin signaling has been studied in mosquitoes (Felix et al., 2015; Li et al., 2006; Mayoral et al., 2010; Nouzova et al., 2015), the role of LWamide neuropeptides are not known. Several transcripts code for secreted peptides having a Kazal domain. Their special role in adult male mosquitoes, if any, is not known.

4. Conclusions

In 2010 we published a Sanger-based sialotranscriptome analysis of adult female *C. tarsalis* based on ~ 2,000 ESTs (Calvo et al., 2010a). In the “Conclusions” section of the article, it was considered that the number of EST’s sequenced were probably enough to identify the main salivary proteins, but possibly “more intensive sequencing of existing libraries, perhaps with newer pyrosequencing methodology, may increase the number of salivary peptides…”. During the elapsed 7.5 years pyrosequencing has been discontinued and Illumina sequences have increased considerable in size and decreased in price. Indeed, the novel technologies allowed to find the missing apyrase from the salivary transcriptome of *C. tarsalis*, to determine several full-length members of the 34–62 kDa family, when a single EST has been found previously, in addition to identifying many salivary families with lower expression levels that were not detected previously. The use of multiple libraries including SG and carcass from male and female organisms allowed for an unprecedented insight into the tissue specificity of transcripts, and in this particular case permitting identification of transcripts putatively associated with blood feeding, when exclusive of female salivary glands, or

associated with sugar feeding, when transcripts are found upregulated in both male and female glands.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge and thank Dr. William Reisen for providing *C. tarsalis* mosquitoes to establish our mosquito colony and Dr. Lyric Bartholomay for use of the insectary at the University of Wisconsin-Madison. The work at University of Wisconsin-Madison was supported by National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number R01-AI099231-01 (KAB) and a training grant supported by National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number T32AI007414 (FRM). JMCR, IMM and EC were supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases. This work utilized the computational resources of the NIH HPC Biowulf cluster. (http://hpc.nih.gov).

References

Alarcon-Chaidez FJ, Boppana VD, Hagymasi AT, Adler AJ, Wikel SK. A novel sphingomyelinase-like enzyme in *Ixodes scapularis* tick saliva drives host CD4 T cells to express IL-4. Parasite immunology. 2009; 31:210–219. [PubMed: 19292772]

Arca B, Lombardo F, Capurro M, della Torre A, Spanos L, Dimopoulos G, Louis C, James AA, Coluzzi M. Salivary gland-specific gene expression in the malaria vector Anopheles gambiae. Parassitologia. 1999a; 41:483–487. [PubMed: 10697906]

Arca B, Lombardo F, de Lara Capurro M, della Torre A, Dimopoulos G, James AA, Coluzzi M. Trapping cDNAs encoding secreted proteins from the salivary glands of the malaria vector Anopheles gambiae. Proceedings of the National Academy of Sciences of the United States of America. 1999b; 96:1516–1521. [PubMed: 9990055]

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 biochemistry and molecular biology. 2007; 37:107–127. [PubMed: 17244540]

Arca B, Lombardo F, Valenzuela JG, Francischetti IM, Marinotti O, Coluzzi M, Ribeiro JM. An updated catalogue of salivary gland transcripts in the adult female mosquito, *Anopheles gambiae*. J Exp Biol. 2005; 208:3971–3986. [PubMed: 16215223]

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 LI, Higgs S, Lanzaro GC, Lawson D, Lee NH, Muskavitch MA, Raikhel AS, Atkinson PW. Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. Science (New York, NY). 2010; 330:86–88.

Augustin R, Siebert S, Bosch TC. Identification of a kazal-type serine protease inhibitor with potent anti-staphylococcal activity as part of Hydra's innate immune system. Developmental and comparative immunology. 2009; 33:830–837. [PubMed: 19428484]

Benjamini Y, Hochberg Y. Controlling the false discovery rate: A pratical and powerful approach to multiple testing. J Royal Statistical Soc Series B. 1995; 57:289–300.

Birol I, Jackman SD, Nielsen CB, Qian JQ, Varhol R, Stazyk G, Morin RD, Zhao Y, Hirst M, Schein JE, Horsman DE, Connors JM, Gascoyne RD, Marra MA, Jones SJ. De novo transcriptome assembly with ABySS. Bioinformatics (Oxford, England). 2009; 25:2872–2877.

Briant L, Despres P, Choumet V, Misse D. Role of skin immune cells on the host susceptibility to mosquito-borne viruses. Virology. 2014; 464–465:26–32.

Calvo E, Mans BJ, Andersen JF, Ribeiro JM. Function and evolution of a mosquito salivary protein family. J Biol Chem. 2006; 281:1935–1942. [PubMed: 16301315]

Calvo E, Mans BJ, Ribeiro JM, Andersen JF. Multifunctionality and mechanism of ligand binding in a mosquito antiinflammatory protein. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:3728–3733. [PubMed: 19234127]

Calvo E, Mizurini DM, Sa-Nunes A, Ribeiro JM, Andersen JF, Mans BJ, Monteiro RQ, Kotsyfakis M, Francischetti IM. Alboserpin, a factor Xa inhibitor from the mosquito vector of yellow fever, binds heparin and membrane phospholipids and exhibits antithrombotic activity. The Journal of biological chemistry. 2011; 286:27998–28010. [PubMed: 21673107]

Calvo E, Ribeiro JM. A novel secreted endonuclease from *Culex quinquefasciatus* salivary glands. J Exp Biol. 2006; 209:2651–2659. [PubMed: 16809456]

Calvo E, Sanchez-Vargas I, Favreau AJ, Barbian KD, Pham VM, Olson KE, Ribeiro JM. An insight into the sialotranscriptome of the West Nile mosquito vector, *Culex tarsalis*. BMC genomics. 2010a; 11:51. [PubMed: 20089177]

Calvo E, Tokumasu F, Marinotti O, Villeval JL, Ribeiro JM, Francischetti IM. Aegyptin, a novel mosquito salivary gland protein, specifically binds to collagen and prevents its interaction with platelet glycoprotein VI, integrin alpha2beta1, and von Willebrand factor. The Journal of biological chemistry. 2007; 282:26928–26938. [PubMed: 17650501]

Calvo E, Tokusamu F, Mizurini D, McPhie P, Narum D, Ribeiro JM, Monteiro R, Francischetti I. Aegyptin displays high-affinity for the von Willebrand factor binding site (QGQOGVMGF) in collagen and inhibits carotid thrombus formation in vivo. The FEBS journal. 2010b; 277:413–427. [PubMed: 20015075]

Cerna P, Mikes L, Volf P. Salivary gland hyaluronidase in various species of phlebotomine sand flies (Diptera: psychodidae). Insect Biochem Mol Biol. 2002; 32:1691–1697. [PubMed: 12429121]

Chagas AC, Calvo E, Rios-Velasquez CM, Pessoa FA, Medeiros JF, Ribeiro JM. A deep insight into the sialotranscriptome of the mosquito, *Psorophora albipes*. BMC genomics. 2013; 14:875. [PubMed: 24330624]

Chagas AC, Oliveira F, Debrabant A, Valenzuela JG, Ribeiro JM, Calvo E. Lundep, a sand fly salivary endonuclease increases *Leishmania* parasite survival in neutrophils and inhibits XIIa contact activation in human plasma. PLoS Pathog. 2014a; 10:e1003923. [PubMed: 24516388]

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*. Proceedings of the National Academy of Sciences of the United States of America. 2014b; 111:6946–6951. [PubMed: 24778255]

Champagne DE, Smartt CT, Ribeiro JM, James AA. 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; 92:694–698. [PubMed: 7846038]

Charlab R, Valenzuela JG, Rowton ED, Ribeiro JM. 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; 96:15155–15160. [PubMed: 10611354]

Dacheux L, Cervantes-Gonzalez M, Guigon G, Thiberge JM, Vandenbogaert M, Maufrais C, Caro V, Bourhy H. A preliminary study of viral metagenomics of French bat species in contact with humans: identification of new mammalian viruses. PLoS ONE. 2014; 9:e87194. [PubMed: 24489870]

De Smet K, Contreras R. Human antimicrobial peptides: defensins, cathelicidins and histatins. Biotechnology letters. 2005; 27:1337–1347. [PubMed: 16215847]

Eberle MW, Reisen WK. Studies on autogeny in Culex tarsalis: 1. Selection and genetic experiments. Journal of the American Mosquito Control Association. 1986; 2:38–43. [PubMed: 3507468]

Felix RC, Trindade M, Pires IR, Fonseca VG, Martins RS, Silveira H, Power DM, Cardoso JC. Unravelling the Evolution of the Allatostatin-Type A, KISS and Galanin Peptide-Receptor Gene Families in Bilaterians: Insights from Anopheles Mosquitoes. PLoS ONE. 2015; 10:e0130347. [PubMed: 26135459]

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. Nature biotechnology. 2011; 29:644–652.

Hostomska J, Volfova V, Mu J, Garfield M, Rohousova I, Volf P, Valenzuela JG, Jochim RC. Analysis of salivary transcripts and antigens of the sand fly *Phlebotomus arabicus*. BMC genomics. 2009; 10:282. [PubMed: 19555500]

Howe EA, Sinha R, Schlauch D, Quackenbush J. RNA-Seq analysis in MeV. Bioinformatics (Oxford, England). 2011; 27:3209–3210.

Juhn J, James AA. oskar gene expression in the vector mosquitoes, Anopheles gambiae and Aedes aegypti. Insect molecular biology. 2006; 15:363–372. [PubMed: 16756555]

Juhn J, Naeem-Ullah U, Maciel Guedes BA, Majid A, Coleman J, Paolucci Pimenta PF, Akram W, James AA, Marinotti O. Spatial mapping of gene expression in the salivary glands of the dengue vector mosquito, *Aedes aegypti*. Parasit Vectors. 2011; 4:1. [PubMed: 21205315]

Karim S, Singh P, Ribeiro JM. A deep insight into the sialotranscriptome of the gulf coast tick, *Amblyomma maculatum*. PLoS ONE. 2011; 6:e28525. [PubMed: 22216098]

Kivirikko KI, Pihlajaniemi T. Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. Adv Enzymol Relat Areas Mol Biol. 1998; 72:325–398. [PubMed: 9559057]

Lai R, Takeuchi H, Lomas LO, Jonczy J, Rigden DJ, Rees HH, Turner PC. A new type of antimicrobial protein with multiple histidines from the hard tick, *Amblyomma hebraeum*. Faseb J. 2004; 18:1447–1449. [PubMed: 15247144]

Lanfrancotti A, Lombardo F, Santolamazza F, Veneri M, Castrignano T, Coluzzi M, Arca B. Novel cDNAs encoding salivary proteins from the malaria vector *Anopheles gambiae*. FEBS Lett. 2002; 517:67–71. [PubMed: 12062411]

Li L, Victoria JG, Wang C, Jones M, Fellers GM, Kunz TH, Delwart E. Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. Journal of virology. 2010; 84:6955–6965. [PubMed: 20463061]

Li Y, Hernandez-Martinez S, Fernandez F, Mayoral JG, Topalis P, Priestap H, Perez M, Navare A, Noriega FG. Biochemical, molecular, and functional characterization of PISCF-allatostatin, a regulator of juvenile hormone biosynthesis in the mosquito Aedes aegypti. The Journal of biological chemistry. 2006; 281:34048–34055. [PubMed: 16968697]

Londono-Renteria B, Cardenas JC, Cardenas LD, Christofferson RC, Chisenhall DM, Wesson DM, McCracken MK, Carvajal D, Mores CN. Use of Anti-Aedes aegypti Salivary Extract Antibody Concentration to Correlate Risk of Vector Exposure and Dengue Transmission Risk in Colombia. PLoS ONE. 2013; 8:e81211. [PubMed: 24312537]

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; 8:2071–2079. [PubMed: 19531497]

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 immunology. 2012; 34:15–22. [PubMed: 21995849]

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 immunology. 2013; 26:84–92. [PubMed: 23362833]

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*. The Journal of biological chemistry. 2007; 282:36626–36633. [PubMed: 17928288]

Marinotti O, James A, Ribeiro JMC. Diet and salivation in female *Aedes aegypti* mosquitoes. J Insect Physiol. 1990; 36:545–548.

Mayoral JG, Nouzova M, Brockhoff A, Goodwin M, Hernandez-Martinez S, Richter D, Meyerhof W, Noriega FG. Allatostatin-C receptors in mosquitoes. Peptides. 2010; 31:442–450. [PubMed: 19409436]

Mizurini DM, Francischetti IM, Monteiro RQ. Aegyptin inhibits collagen-induced coagulation activation in vitro and thromboembolism in vivo. Biochem Biophys Res Commun. 2013; 436:235–239. doi:210.1016/j.bbrc.2013.1005.1082. Epub 2013 May 1030. [PubMed: 23726920]

Mohler J, Mahaffey JW, Deutsch E, Vani K. Control of Drosophila head segment identity by the bZIP homeotic gene cnc. Development (Cambridge, England). 1995; 121:237–247.

Nouzova M, Rivera-Perez C, Noriega FG. Allatostatin-C reversibly blocks the transport of citrate out of the mitochondria and inhibits juvenile hormone synthesis in mosquitoes. Insect biochemistry and molecular biology. 2015; 57:20–26. [PubMed: 25500428]

Provost-Javier KN, Chen S, Rasgon JL. Vitellogenin gene expression in autogenous Culex tarsalis. Insect molecular biology. 2010; 19:423–429. [PubMed: 20456510]

Quistgaard EM, Low C, Guettou F, Nordlund P. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. Nat Rev Mol Cell Biol. 2016; 17:123–132. [PubMed: 26758938]

Reisen WK, Fang Y, Martinez VM. Avian host and mosquito (Diptera: Culicidae) vector competence determine the efficiency of West Nile and St. Louis encephalitis virus transmission. Journal of medical entomology. 2005; 42:367–375. [PubMed: 15962789]

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. [PubMed: 17204158]

Ribeiro JM, Chagas AC, Pham VM, Lounibos LP, Calvo E. An insight into the sialome of the frog biting fly, *Corethrella appendiculata*. Insect biochemistry and molecular biology. 2014; 44:23–32.

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 biochemistry and molecular biology. 2004; 34:543–563. [PubMed: 15147756]

Ribeiro JM, Mans BJ, Arca B. An insight into the sialome of blood-feeding Nematocera. Insect biochemistry and molecular biology. 2010; 40:767–784. [PubMed: 20728537]

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. [PubMed: 26999592]

Ribeiro JMC, Arca B. From sialomes to the sialoverse: An insight into the salivary potion of blood feeding insects. Adv Insect Physiol. 2009; 37:59–118.

Rohousova I, Subrahmanyam S, Volfova V, Mu J, Volf P, Valenzuela JG, Jochim RC. Salivary gland transcriptomes and proteomes of *Phlebotomus tobbi* and *Phlebotomus sergenti*, vectors of leishmaniasis. PLoS neglected tropical diseases. 2012; 6:e1660. [PubMed: 22629480]

Rossignol PA, Lueders AM. Bacteriolytic factor in the salivary glands of *Aedes aegypti*. Comp Biochem Physiol. 1986; 83B:819–822.

Smartt CT, Kim AP, Grossman GL, James AA. The *Apyrase* gene of the vector mosquito, *Aedes aegypti*, is expressed specifically in the adult female salivary glands. Exp Parasitol. 1995; 81:239–248. [PubMed: 7498420]

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; 273:20802–20809. [PubMed: 9694825]

Styer LM, Lim PY, Louie KL, Albright RG, Kramer LD, Bernard KA. Mosquito saliva causes enhancement of West Nile virus infection in mice. Journal of virology. 2011; 85:1517–1527. [PubMed: 21147918]

Team, R.C. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2013.

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. [PubMed: 20711354]

Theis T, Stahl U. Antifungal proteins: targets, mechanisms and prospective applications. Cell Mol Life Sci. 2004; 61:437–455. [PubMed: 14999404]

Vizioli J, Bulet P, Hoffmann JA, Kafatos FC, Muller HM, Dimopoulos G. Gambicin: a novel immune responsive antimicrobial peptide from the malaria vector *Anopheles gambiae*. Proc Natl Acad Sci U S A. 2001; 98:12630–12635. [PubMed: 11606751]

Vlkova M, Sima M, Rohousova I, Kostalova T, Sumova P, Volfova V, Jaske EL, Barbian KD, Gebre-Michael T, Hailu A, Warburg A, Ribeiro JM, Valenzuela JG, Jochim RC, Volf P. Comparative analysis of salivary gland transcriptomes of *Phlebotomus orientalis* sand flies from endemic and non-endemic foci of visceral leishmaniasis. PLoS neglected tropical diseases. 2014; 8:e2709. [PubMed: 24587463]

Volfova V, Hostomska J, Cerny M, Votypka J, Volf P. Hyaluronidase of bloodsucking insects and its enhancing effect on leishmania infection in mice. PLoS neglected tropical diseases. 2008; 2:e294. [PubMed: 18820742]

Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory in biosciences = Theorie in den Biowissenschaften. 2012; 131:281–285. [PubMed: 22872506]

Warnes, GR., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, WHA., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B. R package version 2.14.1 2. 2015. gplots: Various R programming tools for plotting data.

Wichit S, Ferraris P, Choumet V, Misse D. The effects of mosquito saliva on dengue virus infectivity in humans. Current opinion in virology. 2016; 21:139–145. [PubMed: 27770704]

- Transcriptomes of male and female salivary glands and carcasses of *Culex tarsalis* mosquitoes were compared.
- 668 transcripts were found enriched in the salivary glands when compared to the carcasses.
- 332 of the 668 transcripts were at least 10x enriched in the female salivary glands.
- 218 of the 332 transcripts are of a putative secreted nature and likely involved in blood feeding.
- Over 16,000 coding sequences have been publicly submitted to GenBank.

Table 1

Functional classification of transcripts from *Culex tarsalis*

Class	Total Number of Contigs	% total	Number of contigs overexpressed in the salivary glands	% Total
Secreted	2,753	11.11	366	54.79
Unknown	3,256	13.14	131	19.61
Unknown, conserved	4,095	16.53	48	7.19
Transposable element	486	1.96	23	3.44
Immunity	394	1.59	15	2.25
Protein modification machinery	666	2.69	14	2.10
Metabolism, nucleotide	278	1.12	13	1.95
Transporters/storage	1,059	4.27	12	1.80
Metabolism, carbohydrate	471	1.90	11	1.65
Extracellular matrix/cell adhesion	443	1.79	11	1.65
Signal transduction	2,711	10.94	7	1.05
Metabolism, lipid	762	3.08	5	0.75
Metabolism, energy	658	2.66	3	0.45
Metabolism, amino acid	327	1.32	3	0.45
Transcription machinery	1,577	6.37	2	0.30
Proteasome machinery	614	2.48	1	0.15
Protein synthesis machinery	488	1.97	1	0.15
Transcription factor	350	1.41	1	0.15
Viral	15	0.06	1	0.15
Protein export machinery	1,017	4.11	0	0.00
Cytoskeletal	815	3.29	0	0.00
Nuclear regulation	806	3.25	0	0.00
Oxidant metabolism/detoxification	393	1.59	0	0.00
Metabolism, intermediate	115	0.46	0	0.00
Nuclear export	95	0.38	0	0.00
Signal transduction, apoptosis	77	0.31	0	0.00
Storage	51	0.21	0	0.00
Total	24,772	100	668	100

Table 2

Classification of transcripts 10 x overexpressed in female SG when contrasted with male SG, from a subset of transcripts enriched in the SG as compared to carcasses of both sexes.

Class	Number of contigs	Average RPKM female SG	Average RPKM male SG	Ratio Female/male	Average RPKM female carcass	Average RPKM male carcass
Putative secreted						
Enzymes						
Apyrase *	7	184.44	0.19	985.82	0.12	0.13
Adenosine deaminase	2	280.98	0.50	560.28	0.75	2.69
Endonuclease	2	399.55	0.48	839.11	0.30	1.23
Hyaluronidase *	3	193.89	0.15	1,326.69	0.08	0.09
Sphingomyelinase *	5	416.28	0.36	1,150.09	0.35	0.40
Serine protease	1	1,109.21	2.77	400.43	0.63	1.03
Protease inhibitors						
Serpins	1	618.31	0.59	1,039.77	0.41	0.40
Ubiquitous protein families						
Antigen 5 family	3	3,323.33	2.70	1,229.73	2.85	3.22
Mosquito specific families						
D7 protein family	19	5,956.98	4.36	1,366.88	3.93	4.47
Aegyptin/30kDa antigen family	5	3,273.15	3.87	845.19	2.85	3.88
10–14 kDa family - also in Corethrella *	6	4,844.78	22.07	219.53	3.60	4.22
20.2 kDa family - Culex and Corethrella *	4	278.51	7.34	37.95	0.30	0.27
62/34 kDa culicine family	11	287.72	0.34	851.31	0.22	0.19
gSG5 family *	4	394.11	0.27	1,466.63	0.33	0.34
gSG8 family *	4	155.17	1.14	136.05	0.10	0.14
Hyp37 *	6	1,478.03	0.99	1,495.81	0.88	1.30
SGS1 family	8	60.21	0.10	627.36	0.52	0.22
16.7 kDa/WRP family	70	2,260.61	1.90	1,187.89	1.53	1.86
4.2 kDa peptide	5	1,697.78	1.61	1,052.29	0.67	1.10
9.7 kDa family	19	603.36	0.59	1,019.43	0.42	0.48

Class	Number of contigs	Average RPKM female SG	Average RPKM male SG	Ratio Female/male	Average RPKM female carcass	Average RPKM male carcass
Basic tail protein *	1	1,085.96	1.79	606.43	1.09	1.02
Mucins	5	658.27	0.89	741.70	0.68	1.17
Proline rich salivary peptide *	4	212.57	0.22	972.77	0.62	0.32
Unknown secreted proteins	23	191.37	6.05	31.65	0.38	0.96
Putative Housekeeping						
Extracellular matrix	21	39.52	0.55	72.15	0.04	0.15
Amino acid metabolism	1	322.89	20.41	15.82	7.27	33.60
Protein modification	5	36.80	0.29	126.50	0.02	2.26
Signal transduction	3	14.98	0.17	87.36	0.80	1.07
Transporters	7	158.03	8.17	19.34	1.99	5.26
Unknown conserved	22	2,259.62	2.13	1,059.58	1.37	1.75
Unknown	47	77.88	0.73	106.28	0.51	0.98
Transposable elements	8	14.44	0.11	126.70	0.09	0.12
Total or average	332	1,027.77	2.93	681.39	1.12	2.38

* Not found on previous limited *C. tarsalis* sialotranscriptome (Calvo et al., 2010).

Table 3

Classification of transcripts overexpressed nearly equal in both male and female salivary glands

Class	Number of contigs	Average RPKM female SG	Average RPKM male SG	Ratio Female/male	Average RPKM female carcass	Average RPKM male carcass
Putative secreted						
Enzymes						
Apyrase/5′ nucleotidase	1	7.70	8.01	0.96	0.27	1.19
Endonuclease	13	897.19	526.00	1.71	0.46	0.63
Lipases	6	195.12	316.89	0.62	0.23	0.79
Chitinase	1	20.18	24.74	0.82	0.07	0.49
Amylase/Maltase	4	18,428.54	27,652.44	0.67	23.00	37.39
Chitinase	4	3,309.88	3,223.10	1.03	3.11	4.92
Serine proteases	14	157.37	242.60	0.65	0.74	0.91
Protease inhibitors						
TIL domain	4	265.82	624.71	0.43	0.23	0.17
Thyroglobulin repeats	1	8.24	1.70	4.85	0.50	0.36
Antimicrobial peptides/Immunity						
Gambicin	2	1,226.23	1,226.93	1.00	35.75	163.29
H rich peptides	2	40,934.49	105,148.02	0.39	58.42	55.73
C-type lectins	6	1,843.81	2,834.35	0.65	6.82	50.30
Fred domain containing proteins	7	527.70	217.05	2.43	15.29	12.63
Gram negative binding protein	3	634.77	883.67	0.72	1.18	1.89
Ubiquitous protein families						
Nematocera mucin I family	11	5,102.16	12,243.32	0.42	5.77	6.01
Mucin II superfamily	5	441.98	656.03	0.67	0.52	0.71
Perithrophin-like	4	1,120.15	265.03	4.23	65.96	1.91
Orphan mucins	10	78.16	125.33	0.62	0.14	0.14
Antigen 5 related	6	1,167.20	2,437.53	0.48	1.21	1.77
Nematocera protein families						
23.5 kDa culicine family	2	369.03	140.17	2.63	0.32	0.50
56 kDa mosquito family	4	7,550.37	7,165.22	1.05	5.67	6.66
41 kDa culicine family	1	1,122.98	708.71	1.58	0.94	1.10

Class	Number of contigs	Average RPKM female SG	Average RPKM male SG	Ratio Female/male	Average RPKM female carcass	Average RPKM male carcass
30.5 kDa culicine protein family	6	1,693.44	4,198.73	0.40	2.32	2.36
hyp10 hyp12 anopheline family	3	8,589.98	15,849.59	0.54	12.11	12.14
W rich peptides of culicines	7	2,548.26	4,070.01	0.63	2.72	2.96
6.8 kDa culicine family	3	755.70	1,352.77	0.56	0.45	0.81
Novel 6–8 kDa family	5	1,911.70	1,089.59	1.75	0.90	1.22
Culicoides epididymal protein	1	2.83	8.49	0.33	0.00	0.00
Conserved secreted proteins	4	4,579.51	9,838.11	0.47	5.40	4.82
Novel secreted proteins	14	4,915.41	8,994.77	0.55	10.83	14.91
Putative housekeeping						
Extracellular matrix	5	26.66	3.92	6.80	0.03	0.02
Amino acid metabolism	1	7,257.57	15,697.46	0.46	30.72	23.18
Energy metabolism	1	14,724.35	17,802.51	0.83	85.81	158.25
Lipid metabolism	1	16.21	4.89	3.32	1.55	0.25
Protein modification	3	3,755.52	6,665.16	0.56	30.18	26.82
Protein synthesis machinery	1	642.13	381.94	1.68	15.82	15.47
Signal transduction	4	92.66	223.95	0.41	1.43	5.13
Transcription factor	1	171.97	497.85	0.35	21.53	25.36
Transcription machinery	1	339.90	193.03	1.76	17.25	14.76
Transporters and channels	6	126.19	123.48	1.02	6.89	8.98
Unknown conserved	5	51.63	48.25	1.07	0.19	2.95
Unkown conserved membrane protein	4	7.81	6.31	1.24	0.29	1.14
Unknown product	32	1,026.63	2,244.85	0.46	2.42	3.55
Transposable element	6	218.84	692.20	0.32	7.28	21.76
Total or average	225	3,156.00	5,833.17	1.23	10.97	15.83

Table 4

Classification of transcripts overexpressed in the adult male salivary glands

Class	Number of contigs	Average RPKM female SG	Average RPKM male SG	Ratio male/female	Average RPKM female carcass	Average RPKM male carcass
Putative secreted proteins	6	81.58	2,048.36	25.11	38.33	58.66
Putative housekeeping						
Immunity	2	73.94	4,568.98	61.79	30.66	79.20
Amino acid metabolism	1	134.47	7,135.76	53.07	70.81	158.03
Energy metabolism	2	299.63	12,949.26	43.22	193.63	383.32
Lipid metabolism	1	50.26	1,330.44	26.47	21.93	56.11
Nucleotide metabolism	2	130.09	7,529.86	57.88	120.23	154.98
Proteasome machinery	1	277.29	11,976.76	43.19	133.50	269.72
Signal transduction	1	115.04	6,111.54	53.12	51.55	116.54
Transporters and channels	1	783.64	48,431.45	61.80	314.03	801.90
Unknown conserved	6	7.14	267.77	37.51	2.08	6.22
Unknown product	11	5.96	133.36	22.39	0.87	1.89
Viral product	1	0.21	6.38	30.64	0.19	0.24
Transposable element	9	18.47	595.47	32.24	3.85	10.38
Total or average	44	152.13	7,929.64	42.19	75.51	161.32

Calvo, E., Sanchez-Vargas, I., Favreau, A.J., Barbian, K.D., Pham, V.M., Olson, K.E., Ribeiro, J.M., 2010. An insight into the sialotranscriptome of the West Nile mosquito vector, *Culex tarsalis*. BMC genomics 11, 51.