



Published in final edited form as:

*Insect Biochem Mol Biol.* 2008 May ; 38(5): 499–507.

## An insight into the sialotranscriptome of the non-blood feeding *Toxorhynchites amboinensis* mosquito

E. Calvo, V. M. Pham, and J. M. C. Ribeiro

Section of Vector Biology, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville MD 20852, USA

### Abstract

All adult mosquitoes take sugar meals, and most adult females also take blood meals to develop eggs. Salivary glands (SG) of males are thus much smaller and do not contain many of the antihemostatic and antiinflammatory compounds found in females. In the past 5 years, transcriptome analyses have identified nearly 70 different genes expressed in adult female SG. For most of these, no function can be assigned in either blood or sugar feeding. Exceptionally, *Toxorhynchites* mosquitoes are unusual in that they never feed on blood, and the SG of adults are identical in both sexes. Transcriptome analysis of the adult SG of this mosquito was performed to increase knowledge of the evolution of blood feeding—and to identify polypeptide families associated with sugar feeding—in mosquitoes.

### Keywords

Salivary glands; Transcriptome; Mosquito; Hematophagy

### Introduction

All mosquitoes take sugar meals as adults, which largely increases their lifespan as it provides metabolic energy for basal metabolism and flight. Most adult females also take blood meals for egg development, either as an absolute (anautogenous) or facultative (autogenous) requirement (Clements, 1992). In the latter case, a single egg batch is developed after the mosquito molts to adult, but blood meals are required for further egg batches. Accordingly, the salivary glands (SG) of adult mosquitoes are sexually dimorphic. Male SG are much smaller and do not have, or have much smaller amounts of, activities associated with the blood meal such as antiplatelet, anticlotting, and vasodilatory. On the other hand, both sexes have similar amounts of activities associated with sugar meals such as glycosidases and antimicrobial agents that might prevent microbial growth in the mosquito crop, where sugar meals are stored before being diverted to the midgut for further digestion and absorption (Marinotti et al., 1990; Moreira-Ferro et al., 1999; Ribeiro et al., 1984; Rossignol and Lueders, 1986). Exceptionally, the mosquito genus *Toxorhynchites* comprises species that do not take any blood meal in the adult stage. Their SG are monomorphic and have been proposed to be a good model of mosquito sugar feeding (Jariyapan et al., 2004).

---

Correspondence and proofs to J. M. C. Ribeiro, LMVR, NIAID, Twinbrook III, 12735 Twinbrook Parkway, Room 2E32D, NIH, Rockville MD 20852; Telephone: 301-496-9389; Fax: 301-480-2571.

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

Recent sialotranscriptome analysis of adult female mosquitoes has uncovered a putative cocktail of SG-secreted polypeptides numbering near 70 gene products (Arca et al., 2007; Arca et al., 2005; Calvo et al., 2007a; Ribeiro et al., 2007). For most of these putative secreted polypeptides, no functional assignment is possible based on their primary sequence. Many of these proteins are unique to mosquitoes, and some are unique to either Culicines or Anophelines (Arca et al., 2005; Calvo et al., 2004; Ribeiro et al., 2004b; Valenzuela et al., 2003). In an attempt to separate secreted salivary gene products into those assisting either blood or sugar feeding, quantitative polymerase chain reaction (PCR) or DNA microarray analysis of different tissues and sexes have been attempted for *Anopheles gambiae* (Arca et al., 2005), *Aedes aegypti* (Ribeiro et al., 2007), and *Aedes albopictus* (Arca et al., 2007), as well as a comparative analysis of male and female sialotranscriptomes, in the case of *An. gambiae* (Calvo et al., 2006b). In the present work, we compare the sialotranscriptome of adult *Toxorhynchites amboinensis* with those of other mosquitoes to better understand the molecular evolution of mosquitoes to blood feeding and to characterize SG protein families associated with sugar-feeding roles.

## Material and Methods

### Mosquitoes

Adult *Tx. amboinensis* (CDC strain originating in Puerto Rico) were reared in the Insectary of the LMVR under the expert supervision of André Laughinghouse. Larvæ were fed *A. aegypti* larvæ, and adults were kept with cotton pieces containing a Karo solution. SG (50 pairs) were dissected and placed into a solution of 75% RNA-Later (Ambion Inc.), 25% 1 × PBS (RNase free), and stored in 100% RNA-Later at -20°C for isolating polyA<sup>+</sup> RNA.

### Library construction

This was achieved exactly as before (Calvo et al., 2007a). *Tx. amboinensis* SG mRNA was isolated from 50 SG pairs from adult (both males and females) mosquitoes using the Micro-FastTrack mRNA isolation kit (Invitrogen). The PCR-based cDNA library was made following the instructions for the SMART cDNA library construction kit (Clontech). SG polyA<sup>+</sup> RNA was used for reverse transcription to cDNA using PowerScript reverse transcriptase (Clontech), the SMART IV oligonucleotide, and the CDS III/3' primer (Clontech). The reaction was carried out at 42°C for 1 h. Second-strand synthesis was performed by a long-distance (LD), PCR-based protocol using the 5' PCR primer and the CDS III/3' primer as sense and anti-sense primers, respectively. These two primers also create *Sfi*I A and B restriction enzyme sites at the end of nascent cDNA. Advantage<sup>TM</sup> Taq polymerase mix (Clontech) was used to carry out the LD PCR reaction on a GeneAmp<sup>®</sup> PCR System 9700 (Perkin Elmer Corp.). The PCR conditions were: 95°C for 20 s; 24 cycles of 95°C for 5 s; 68°C for 6 min. A small portion of the cDNA was analyzed on a 1.1% agarose/EtBr (0.1 µg/ml) gel to check for the quality and the range of the cDNA synthesized. Double-stranded cDNA was immediately treated with proteinase K (0.8 µg/ml) at 45°C for 20 min. Proteinase K was removed using a Microcon YM-100 mini-column (100,000 MWCO; Millipore) following the manufacturer's recommendations.

The clean, double-stranded cDNA was then digested with *Sfi*I restriction enzyme at 50°C for 2 h, followed by size fractionation on a ChromaSpin-400 drip column (Clontech). The profiles of the fractions were checked on a 1.1% agarose/EtBr (0.1 µg/ml), and fractions containing cDNA of more than 400 bp were pooled and concentrated by mini-column as described above. The cDNAs were then ligated into a λ TriplEx2 vector (Clontech), and the resulting ligation mixture was packaged using GigaPack<sup>®</sup> III Plus packaging extract (Stratagene) according to the manufacturer's instructions. The packaged library was plated by infecting log-phase XL1-Blue *E. coli* cells (Clontech). The percentage of recombinant clones was determined by

performing a blue-white selection screening on LB/MgSO<sub>4</sub> plates containing X-gal/IPTG. Recombinants were also determined by PCR, using vector primers (5' λ TriplEx2 and 3' λ TriplEx2 sequencing primers) flanking the inserted cDNA and visualizing the products on a 1.1% agarose/EtBr gel.

### Sequencing of the *Tx. amboinensis* cDNA library

The *Tx. amboinensis* SG cDNA library was plated on LB/MgSO<sub>4</sub> plates containing X-gal/IPTG to an average of 250 plaques per 150-mm Petri dish. Recombinant (white) plaques were randomly selected and transferred to 96-well Microtest™ U-bottom plates (BD BioSciences) containing 100 μl of SM buffer (0.1 M NaCl, 0.01 M MgSO<sub>4</sub>·7 H<sub>2</sub>O, 0.035 M Tris-HCl [pH 7.5], 0.01% gelatin) per well. The plates were covered and placed on a gyrating shaker for 30 min at room temperature. The phage suspension was either immediately used for PCR or stored at 4°C for future use.

To amplify the cDNA using PCR, 4 μl of the phage sample was used as a template. The primers were sequences from the λ TriplEx2 vector and named pTEx2 5seq (5'-TCC GAG ATC TGG ACG AGC-3') and pTEx2 3LD (5'-ATA CGA CTC ACT ATA GGG CGA ATT GGC-3'), positioned at the 5' and 3' end of the cDNA insert, respectively. The reaction was carried out in 96-well flexible PCR plates (Fisher Scientific) using Platinum SuperMix (Invitrogen) on a GeneAmp® PCR system 9700. The PCR conditions were: 1 hold at 95°C for 3 min; 25 cycles of 95°C for 1 min, 61°C for 30 sec; 72°C for 2 min. Amplified products were analyzed on a 1.5% agarose/EtBr gel. cDNA library clones (1100 clones) were PCR amplified; those showing a single band were selected for sequencing. Approximately 200–250 ng of each PCR product was used for DNA sequencing. cDNA sequencing was carried out using a BigDye® Terminator v3.1 cycle sequencing kit (Applied Biosystems), and reaction products were analyzed on an ABI 3730xl DNA analyzer (Applied Biosystems).

### Bioinformatic tools and procedures used

This was done exactly as before (Calvo et al., 2007a). EST were trimmed of primer and vector sequences, clusterized, and compared with other databases as previously described (Valenzuela et al., 2003). For functional annotation of the transcripts, we used the program blastx (Altschul et al., 1997) to compare nucleotide sequences to the NR protein database of the NCBI and to the gene ontology database (Ashburner et al., 2000). The tool rpsblast (Schaffer et al., 2001) was used to search for conserved protein domains in the Pfam (Bateman et al., 2000), Smart (Letunic et al., 2002), Kog (Tatusov et al., 2003) and conserved domains databases (Marchler-Bauer et al., 2002). We also compared the transcripts with other subsets of mitochondrial and rRNA nucleotide sequences downloaded from NCBI, and to several organism proteomes downloaded from the NCBI (yeast), Flybase (*Drosophila melanogaster*), or ENSEMBL (*An. gambiae*). Segments of the three-frame translations of EST (because the libraries are unidirectional, we did not use six-frame translations) starting with a methionine found in the first 100 predicted aa, or the predicted protein translation in the case of complete coding sequences, were submitted to the SignalP server (Nielsen et al., 1997) to help identify translation products that could be secreted. O-glycosylation sites on the proteins were predicted with the program NetOGlyc (<http://www.cbs.dtu.dk/services/NetOGlyc/>) (Hansen et al., 1998). Functional annotation of the transcripts was based on all the comparisons above. Following inspection of all these results, transcripts were classified as either Secretory (S), Housekeeping (H), or of Unknown (U) function, with further subdivisions based on function and/or protein families. Sequence alignments were performed with the ClustalX (Thompson et al., 1997) software package. Phylogenetic analysis and statistical neighbor-joining bootstrap tests of the phylogenies were carried out with the Mega package (Kumar et al., 2004).

## Gel electrophoresis and Edman degradation studies

This was done as described in Arcà *et al.* (2006), with some modifications. SDS-PAGE of homogenized SG of *Tx. amboinensis* mosquitoes was done using a NuPAGE 4–12% bis-tris gel, 1 mm thick (Invitrogen). Briefly, 50 SG pairs of adult mosquitoes were homogenized in 50 µl of TBS (50 mM Tris, 150 mM NaCl, pH 7.4) and heated to 85°C. After 10 min, the sample was centrifuged for 30 min at 14000×g in a bench-top Eppendorf centrifuge. Supernatant and pellet were treated with NuPAGE LDS sample buffer (Invitrogen) in the presence of 1× reducing agent (Invitrogen). The gel was run with MES buffer according to the manufacturer's instructions. To estimate the molecular weight of the samples, SeeBlue™ Plus2 pre-stained protein standard from Invitrogen (myosin, phosphorylase, BSA, glutamic dehydrogenase, alcohol dehydrogenase, carbonic anhydrase, myoglobin red, lysozyme, aprotinin, and insulin, chain B) was used. The gel was stained with Coomassie blue (0.2%, v/v). For aminoterminal sequencing of the salivary proteins, the gel was transferred to polyvinylidene difluoride (PVDF) membrane using 10 mM CAPS, pH 11.0, 10% methanol as the transfer buffer on a blot module for the XCell II Mini-Cell (Invitrogen). The membrane was stained with Coomassie blue (0.02% in the absence of acetic acid). Stained bands were cut from the PVDF membrane and subjected to Edman degradation using a Procise sequencer (Perkin-Elmer Corp.). More details can be obtained in a previous publication (Francischetti *et al.*, 2002). To find the cDNA sequences corresponding to the aa sequence—obtained by Edman degradation of the proteins transferred to PVDF membranes from PAGE gels—we wrote a search program (in Visual Basic) that evaluated these aa sequences against the three possible protein translations of each cDNA sequence obtained in the mass sequencing project or against the full-length sequences derived in this work. For details, see (Valenzuela *et al.*, 2002b).

## Results and Discussion

### Preliminary characterization of the salivary proteome of *Tx. amboinensis*

To obtain information on the most abundant proteins in the SG of *Tx. amboinensis*, SG homogenates of 50 SG pairs of adult mosquitoes were separated by sodium dodecyl sulfate polyacrylamide electrophoresis (SDS-PAGE), the protein transferred to a PVDF membrane, and the stained bands submitted to Edman degradation. Nine bands yielded useful information, by being assigned either to predicted sequences from our clusterized database or to the *Æ. aegypti* proteome (in the latter case, the sequences in Fig. 1 have the name 'Ædes' indicated). These results are summarized in Fig. 1. Descriptions of the proteins thus identified are provided below.

### cDNA library characteristics

A total of 1,035 clones was used to assemble a clusterized database (Supplemental Table S1), yielding 472 clusters of related sequences, 433 of which contained only one expressed sequence tag (EST). The consensus sequence of each cluster is named either a contig (deriving from two or more sequences) or a singleton (deriving from a single sequence); in this paper, for simplicity sake, we will use the name 'cluster' to denote sequences deriving both from consensus sequences and from singletons. The 492 clusters were compared by blastx, blastn, or rpsblast (Altschul *et al.*, 1997) to the nonredundant (NR) protein database of the National Center for Biotechnology Information (NCBI), to the gene ontology database (Ashburner *et al.*, 2000), to the conserved domains database of the NCBI (Marchler-Bauer *et al.*, 2002), and to a custom-prepared subset of the NCBI nucleotide database containing either mitochondrial or rRNA sequences. Because the libraries used are unidirectional, the three-frame translations of the dataset were also derived, and open-reading frames starting with a methionine and longer than 40 amino acid (aa) residues were submitted to the SignalP server (Nielsen *et al.*, 1997) to help identify putative secreted proteins (S). The EST assembly, BLAST, and signal peptide results were transferred into an Excel spreadsheet for manual annotation.

Four categories of expressed genes derived from the manual annotation of the contigs. The S category contained 6.6% of the clusters and 53.1% of the sequences, with an average number of 17.7 sequences per cluster. While the percentage of total sequences in the S category is within the values registered for other transcriptomes, the number of sequences per cluster is larger and the percentage of clusters is smaller than in previous sialotranscriptomes, indicating a lesser complexity of secreted salivary products in *Toxorhynchites*. The housekeeping (H) category had 8.3% and 5.7% of the clusters and sequences, respectively, and an average of 1.5 sequences per cluster. Eighty-five percent of the clusters, containing 41.1% of all sequences, were classified as unknown (U) because no assignment for their function could be made; they had an average of 1.1 sequence per cluster. A good proportion of these transcripts could have derived from truncated 3' or 5' untranslated regions of genes of the above two categories, as was recently indicated for a sialotranscriptome of *An. gambiae* (Arca et al., 2005). Probable transposable elements originated one singleton representing either active transposition or, more likely, expression of transposable element regulatory transcripts in *Tx. amboinensis*. Transposable element transcripts have been a regular finding in most sialotranscriptomes to date.

### H genes

The 39 clusters (comprising 59 EST) attributed to H genes expressed in the SG of *Tx. amboinensis* were further characterized into 14 subgroups according to function (Table 1). As observed in previous sialotranscriptomes (Francischetti et al., 2002; Ribeiro et al., 2004a; Ribeiro et al., 2004b), the two larger sets were associated with protein synthesis machinery (45 EST in 26 clusters) and with energy metabolism (4 clusters containing 4 EST). Additional inspection of each cluster for further information can be done online with Supplemental Table S1. Fig. 1 indicates the presence of protein bands in the SDS-PAGE experiment matching *Æ. aegypti* calreticulin and protein disulfide isomerase, which are involved in protein modification events and have been a typical find in mosquito sialotranscriptomes.; however, no transcripts coding for these two proteins were found in the transcriptome.

### S proteins and peptides

Supplemental Table S1 indicates the presence of several gene families previously described in the SG of mosquitoes, including the mucins, amylase, and members of the protein families named 56-kDa, 41.9-kDa, 30.5-kDa, 23-kDa, and antigen 5 (Ag5), as well as immunity-related products such as lysozyme, Gram-negative binding protein, and one serine protease, possibly involved in the prophenoloxidase activation cascade. Of interest, 2 clusters containing a total of 257 EST, or nearly 50% of all transcripts of the S class, code for peptides containing histidine-rich motifs that might have antimicrobial activity (Lai et al., 2004; Rothstein et al., 2002). A summary of these transcript families organized by their abundance is shown in Table 2. It is also interesting that of the 31 clusters of the S class, 8 gave a best match by blastx to the NR database to *Æ. aegypti* proteins, 6 to *Æ. albopictus*, 3 to *Culex pipiens*, 1 to *Culicoides sonorensis*, and 1 to a *Drosophila melanogaster* protein. The remaining clusters had low significance ( $< 1e-7$ ) values to proteins of these or other organisms, but no anopheline species produced a best match, in agreement with the close proximity of the *Toxorhynchites* genus to the culicines (Shepard et al., 2006).

### Analysis of the adult female *Tx. amboinensis* sialotranscriptome coding for putative secreted polypeptides

Supplemental Table S1 can be accessed online for a detailed description of the transcripts found in the SG of adult *Tx. amboinensis*, as follows (underlined blue text indicates hyperlinks with supportive information):

## Putative secreted proteins belonging to ubiquitous protein families

### Enzymes

Supplemental Table S1 presents coding sequences for the enzymes amylase (45 EST in 1 cluster), lysozyme (24 EST in 4 clusters), a serine protease (1 transcript matching a 372-aa-long serine protease of *Ae. aegypti*), and an endonuclease (2 EST in 1 cluster). All these types of transcripts were previously found in all mosquito sialotranscriptomes done to date, except for the endonuclease, which was found only on in those of *Culex pipiens quinquefasciatus* (Ribeiro *et al.*, 2004b), where the secreted recombinant enzyme was studied (Calvo and Ribeiro, 2006), and Phlebotomines (Anderson *et al.*, 2006; Oliveira *et al.*, 2006). Transcripts coding for the other enzymes are most probably associated with sugar feeding and antimicrobial activity. Endonuclease may participate in Toxorhynchites with an antimicrobial activity. Alternatively, it could have a housekeeping role, and indeed we do not know from the truncated transcripts whether the full-length coding sequence has a signal peptide indicative of secretion. The SDS-PAGE experiments revealed Edman degradation products for lysozyme and amylase in the expected regions of the gel according to their molecular masses (Fig. 1B). Note also that the band coinciding with the amylase N-terminus is the strongest protein band in the gel shown in Fig. 1B.

### Gram-negative bacteria binding protein

One truncated transcript codes for a 43-aa stretch, with 81% identity, of the carboxy terminus of *Ae. aegypti* Gram-negative bacteria-binding protein. Together with the serine protease, this protein may participate in tagging bacteria to surface activation of the prophenoloxidase cascade.

### Ag5 family

This is a family of secreted proteins that belong to the CAP family (cysteine-rich secretory proteins; Ag5 proteins of insects; pathogenesis-related protein 1 of plants) (Megraw *et al.*, 1998). The CAP family is related to venom allergens in social wasps and ants (Hoffman, 1993; King and Spangfort, 2000) and to antifungal proteins in plants (Stintzi *et al.*, 1993; Szyperki *et al.*, 1998). Members of this protein family are found in the SG of many blood-sucking insects and ticks (Francischetti *et al.*, 2002; Li *et al.*, 2001; Valenzuela *et al.*, 2002b). In *An. gambiae*, four such proteins were identified in sialotranscriptomes, but only one (putative gVAG protein precursor) had coding transcripts enriched in adult female SG (Arca *et al.*, 2005). In *Ae. aegypti* and *Ae. albopictus*, three and four gene products of this protein family were described, respectively, but only one member of the family appears to be female specific (Arca *et al.*, 2007; Ribeiro *et al.*, 2007). Two truncated transcripts were found in the *Tx. amboinensis* sialotranscriptome, being 56% identical to an *Ae. aegypti* and *Ae. albopictus* Ag5 proteins on a 65-aa stretch of the carboxy terminus of these proteins. This finding is in accordance with the Ag5 family in mosquitoes comprising both female-specific and female-nonspecific salivary transcripts.

## Secreted protein families found exclusively in mosquitoes

### Mucins

Transcripts coding for putative mucins constitute near 20% of secreted transcripts, with 101 EST in 4 clusters. The derived full-length polypeptides are rich in serine and threonine residues that can contain O-linked galactosamine. Mucins may be important for lining the chitinous salivary ducts, in addition to other possible roles. Among this class of transcripts, we here include the full-length sequence of TX-12, homologous to *Culex* and *Aedes* proteins and containing 16 predicted glycosylation sites in its mature aminoterminal domain; TX-10, similar to *Aedes* mucins and peritrophins; and TX-4, also uniquely similar to *Aedes* salivary proteins.

Supplemental Table S2 presents detailed data on these and other full-length salivary transcripts of *Tx. amboinensis*. Evidence for expression of TX-10 was found in a region of the gel compatible with the predicted mature molecular mass (12 kDa), indicating this protein may not be heavily glycosylated.

### 56-kDa family

This protein family is found exclusively in the SG of culicines and anophelines. In *An. gambiae*, a single gene is apparent for this protein (Arca et al., 2005), and this appears to be the case for *Ae. aegypti* and most likely also for other mosquito species including *Ae. albopictus* and *Cx. p. quinquefasciatus*. In *An. gambiae*, *Ae. aegypti*, and *Ae. albopictus*, the gene coding for the 56-kDa protein was found expressed in female SG and in male mosquitoes (Arca et al., 2005). This protein family is abundantly expressed in *Tx. amboinensis* SG, where 30 EST in 2 clusters were found. The larger cluster, with 23 sequences, produces a truncated product matching one half of *Cx. p. quinquefasciatus* 56-kDa protein, with 40% identity. It is possible that the 56-kDa protein family plays a role as an antimicrobial agent in the crop-stored sugar meal. Although we did not obtain any full-length transcripts coding for *Tx. amboinensis*, the SDS-PAGE experiment in Fig. 1B shows a band near the 62-kDa marker that matches the N-terminal sequence of *Ae. aegypti* 56-kDa protein, suggesting expression of this protein.

### Proteins found to date only in culicine SG transcriptomes, unknown function

This class of transcripts was previously found only in *Aedes* and *Culex* mosquitoes (Arca et al., 2007; Ribeiro et al., 2007; Ribeiro et al., 2004b), despite extensive transcriptome coverage of *An. gambiae* (Arca et al., 2005) as well as genomic searches in the same mosquito. Finding these transcripts in Toxorhynchites indicates that this genus shared a common ancestor with Culicines after the divergence of anophelines, in support of current mosquito phylogeny (Shepard et al., 2006).

### 41.9-kDa basic protein family

Members of this family are found in *Cx. p. quinquefasciatus*, *Ae. aegypti*, and *Ae. albopictus*. No other significant similarities occur when these sequences are compared with the NR protein database. Previous RT-PCR expression analysis of this protein in *Ae. albopictus* indicated expression of this gene both in female glands and in adult males (Arca et al., 2007), consistent with the finding of 9 transcripts in *Tx. amboniensis* sialotranscriptome producing 41% identity to the *Ae. aegypti* homolog over a 236-aa stretch. The function of members of this family is presently unknown.

### 30.5-kDa family

Two genes coding for proteins of ~30.5 kDa (not to be confused with the 30-kDa/GE-rich protein family) were found as a tandem repeat in the genome of *Ae. aegypti* (Ribeiro et al., 2007). Similarly, two transcripts were also found in *Ae. albopictus*, named 27-kDa family, unfortunately (Arca et al., 2007). Both genes in *Ae. aegypti* are significantly overtranscribed in the sialotranscriptome when compared with other transcriptomes; additionally, RT-PCR studies in *Ae. aegypti* (Ribeiro et al., 2007) indicate enrichment in the female SG or exclusive expression in the same organ. Toxorhynchites cluster TX-contig\_18, with 6 EST, codes for the last 100 aa of its *Ae. aegypti* homolog, indicating this protein family has a role in sugar feeding or that, in Toxorhynchites, it has evolved to function differently from *Aedes* counterparts.

### 23.4-kDa family

This protein family was found in salivary transcriptomes of *Cx. p. quinquefasciatus*, *Ae. aegypti*, and *Ae. albopictus*. Five truncated EST in the Toxorhynchites sialotranscriptome

assembled into TX-contig\_20 to provide matches to the three known Culicine proteins. In *Aedes*, the family was found transcribed in female SG and in male mosquitoes by RT-PCR (Arca et al., 2007; Ribeiro et al., 2007). In accordance with its being found in Toxorhynchites, this protein family may assist sugar feeding as an antimicrobial agent.

### Other putative secreted proteins

This group includes transcripts coding for novel polypeptides or polypeptides not ubiquitously found in other mosquito sialotranscriptomes.

#### Histidine-rich 8-kDa peptides

Two clusters of EST—representing one quarter of the total transcriptome and nearly half of all transcripts coding for putative secreted products—code for two peptides with mature mass of 8.1 and 8.6 kDa and containing domains rich in histidine, mostly GH or GHH repeats found in metal chelators or metal transporters. TX-3, devoid of cysteines, is similar to a previously described salivary protein from *C. sonorensis*, as it shares a similar signal peptide and GHH repeats located in the aminoterminal region of the mature peptide. TX-1 is similar to a peptide of similar length found previously only in *Ae. albopictus* sialotranscriptomes, the GHGHGH-rich region in the Toxorhynchites sequence being substituted for a HHHHH stretch in *Aedes*. As mentioned above, these peptides may belong to a class of histidine-rich antimicrobial peptides (Lai et al., 2004; Rothstein et al., 2002). They may work by sequestering Ni or Zn, which are trace metals essential for bacterial growth. These two polypeptides appear to be abundantly expressed, as indicated in Fig. 1A. The polypeptide TX-3, which has a predicted molecular mass of 8.1 kDa (Edman result if Fig. 1A begins with 'ARF...'), shows mobility in SDS-PAGE indicative of > 20 kDa, indicative of some post-translation modification or polymerization. This peptide is associated with the strongest staining protein band in Fig. 1A. TX-1 is also detected in the gel band with Edman degradation beginning with 'HGHHG...' in Fig. 1A, with a gel mobility consistent with a 20-kDa mass, more than double that predicted of 8.6 kDa. These two peptides have one and four predicted O-galactosylation sites that could explain their gel mobility.

#### PolyQ secreted protein

TX-contig\_7, with 37 EST, codes for a low complexity product with a signal peptide indicative of secretion. The predicted mature basic protein (pI = 8.5) has a long polyglutamine repeat on its mature amino terminus and is rich in serine and threonine in its carboxy terminus, with up to 12 predicted glycosylation sites. Polyglutamine peptides are known to form peculiar water-filled nanotubes (Perutz *et al.*, 2002; Sharma *et al.*, 2005). TX-7 thus is an unusual protein with one mucous and one nanotube ends. Polyglutamine repeats are normally found in large intracellular proteins associated with signal transduction or transcription factor activities. To attempt further characterization of this unique product, we scanned 3,952,268 sequences of the NR protein database for proteins starting with a methionine and having at least eight consecutive glutamines residues. A total of 8,240 proteins satisfied this requirement, of which 633 had a signal peptide indicative of secretion and only 42 had molecular weight of the mature peptide lower than 20 kDa, indicating the rarity of this type of polypeptide.

#### Unremarkable secreted peptide

TX-contig\_13 was assembled from 11 transcripts and codes for a secreted peptide with mature mass of 7.3 kDa. It has no cysteines or glycosylation sites and produces no significant matches to the NR database even without the filter for low complexity regions. It is unremarkable, and its function is unknown.



### 30 kDa-like 6.5-kDa peptide

TX-contig\_14, assembled from 10 transcripts, produces blast matches to an *Æ. albopictus* member of the 30-kDa salivary antigen representing 29% identity and 48% similarity to the first 72 aa of the > 200 aa-long protein. On close inspection, the majority of the matches are to the signal peptide region of the *Ædes* protein. TX-14, however, is a 6.5-kDa peptide, much shorter than the members of the 30-kDa antigen family, which is uniquely found in adult female sialotranscriptomes (Arca et al., 2007; Arca et al., 2005; Calvo et al., 2006b; Ribeiro et al., 2007). The 30 kDa antigen in *Ae. aegypti* was recently shown to inhibit collagen-induced platelet aggregation (Calvo et al., 2007b). This Toxorhynchites peptide may have originated from a member of the 30-kDa antigen family to produce a product with a different function.

### Concluding remarks

The accumulating information on mosquito sialotranscriptomes—including male and female *An. gambiae* as well as female *Anopheles stephensi*, *Anopheles funestus*, *Anopheles darlingi*, *Æ. aegypti*, *Æ. albopictus*, *Cx. p. quinquefasciatus*, and now *Tx. amboinensis*—is helping to distinguish the unique salivary composition of mosquitoes. In the present case, the sialotranscriptome of *Tx. amboinensis* presents useful information both on the matches to previous salivary proteins and also for some remarkable absences. Finding proteins such as amylase, serine proteases, Gram-negative binding protein, and lysozyme is not surprising but reinforces the presumed functions ascribed in sugar feeding. On the other hand, finding Toxorhynchites homologs of unique mosquito families of unknown function, namely members of the 56-kDa, 41.9-kDa, 30.5-kDa, and 23.4 kDa families, indicates a role for these proteins as relatively large antimicrobial agents, or they may represent some yet-unknown structure with glycosidase activity. Expression of these proteins followed by antimicrobial and enzymatic assays on carbohydrate substrates should be the natural course of this research. Unique polypeptides were also described: a novel mucin containing a polyglutamine domain, and histidine-rich peptides possibly with antimicrobial function. Notably absent from the list are enzymes associated with purine degradation, such as apyrase and adenosine deaminase, protease inhibitors, and members of the D7 family of salivary proteins. Also absent are members of the 62-kDa and 34-kDa families, 30-kDa antigen, basic 7.6-kDa peptide, and basic 3.8-kDa peptide. Many of these absent families were abundantly transcribed in *Ædes* and *Culex* transcriptomes (Arca et al., 2002; Calvo et al., 2006a; James et al., 1991; Valenzuela et al., 2002a), and their complete absence in over 1,000 transcripts is significant, indicating that these families may play a role in blood feeding. Expression of these proteins in hemostasis and inflammation bioassays may help to uncover their function. While we take another step toward understanding the evolution of blood feeding by mosquitoes, we categorize the possible role of many novel protein families that might have an impact in human physiology or potential to act as novel antimicrobial agents.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgements

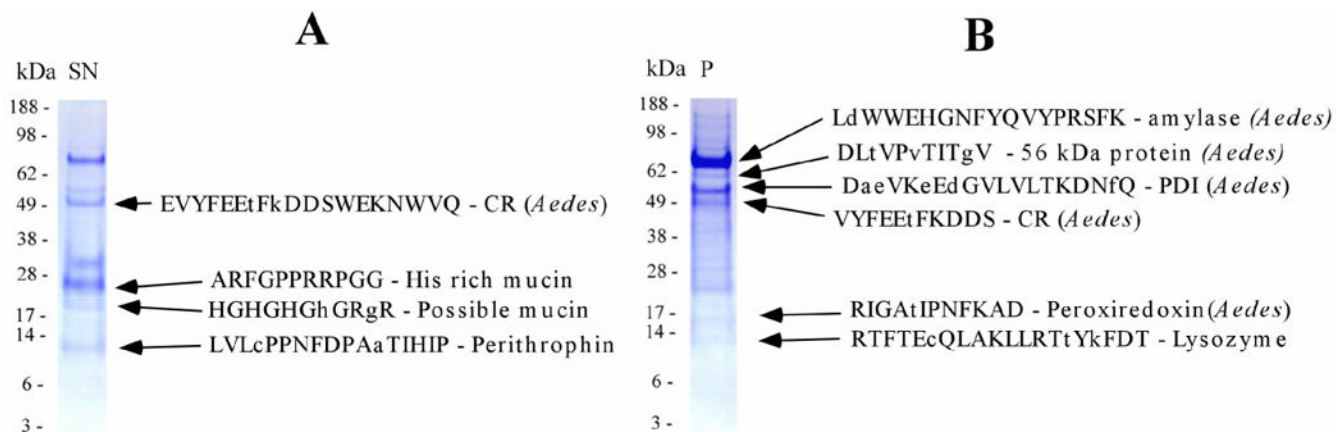
This work was supported by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health. We thank NIAID intramural editor Brenda Rae Marshall, Mark Garfield of Research Technology Branch, NIAID, for the Edman degradation reactions, and Chuong Huynh from NCBI for help with posting the databases.

## References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402. [PubMed: 9254694]
- Anderson JM, Oliveira F, Kamhawi S, Mans BJ, Reynoso D, Seitz AE, Lawyer P, Garfield M, Pham M, Valenzuela JG. Comparative salivary gland transcriptomics of sandfly vectors of visceral leishmaniasis. *BMC Genomics* 2006;7:52. [PubMed: 16539713]
- 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;37:107–127. [PubMed: 17244540]
- Arca B, Lombardo F, Lanfrancotti A, Spanos L, Veneri M, Louis C, Coluzzi M. A cluster of four D7-related genes is expressed in the salivary glands of the African malaria vector *Anopheles gambiae*. *Insect Mol. Biol* 2002;11:47–55. [PubMed: 11841502]
- 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]
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet* 2000;25:25–29. [PubMed: 10802651]
- Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL. The Pfam protein families database. *Nucleic Acids Res* 2000;28:263–266. [PubMed: 10592242]
- Calvo E, Andersen J, Francischetti IM, de LCM, deBianchi AG, James AA, Ribeiro JM, Marinotti O. The transcriptome of adult female *Anopheles darlingi* salivary glands. *Insect Mol. Biol* 2004;13:73–88. [PubMed: 14728669]
- 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* 2007a;37:164–175. [PubMed: 17244545]
- Calvo E, Mans BJ, Andersen JF, Ribeiro JM. Function and evolution of a mosquito salivary protein family. *J. Biol. Chem* 2006a;281:1935–1942. [PubMed: 16301315]
- Calvo E, Pham VM, Lombardo F, Arca B, Ribeiro JM. The sialotranscriptome of adult male *Anopheles gambiae* mosquitoes. *Insect Biochem. Mol. Biol* 2006b;36:570–575. [PubMed: 16835022]
- Calvo E, Ribeiro JM. A novel secreted endonuclease from *Culex quinquefasciatus* salivary glands. *J. Exp. Biol* 2006;209:2651–2659. [PubMed: 16809456]
- 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. *J Biol Chem* 2007b;282:26928–26938. [PubMed: 17650501]
- Clements, AN. *Biology of Mosquitoes: Development, nutrition and reproduction*. London: Chapman & Hall; 1992.
- 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;205:2429–2451. [PubMed: 12124367]
- Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj J* 1998;15:115–130. [PubMed: 9557871]
- Hoffman DR. Allergens in Hymenoptera venom. XXV: The amino acid sequences of antigen 5 molecules and the structural basis of antigenic cross-reactivity. *J Allergy Clin Immunol* 1993;92:707–716. [PubMed: 8227862]
- 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;44:245–254. [PubMed: 2052024]

- Jariyapan N, Choochote W, Jitpakdi A, Bates PA. Salivary gland of *Toxorhynchites splendens* Wiedemann (Diptera: Culicidae): ultrastructural morphology and electrophoretic protein profiles. *J. Med. Entomol* 2004;41:569–574. [PubMed: 15311445]
- King TP, Spangfort MD. Structure and biology of stinging insect venom allergens. *Int Arch Allergy Immunol* 2000;123:99–106. [PubMed: 11060481]
- Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 2004;5:150–163. [PubMed: 15260895]
- 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]
- Letunic I, Goodstadt L, Dickens NJ, Doerks T, Schultz J, Mott R, Ciccarelli F, Copley RR, Ponting CP, Bork P. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res* 2002;30:242–244. [PubMed: 11752305]
- Li S, Kwon J, Aksoy S. Characterization of genes expressed in the salivary glands of the tsetse fly, *Glossina morsitans morsitans*. *Insect Mol. Biol* 2001;10:69–76. [PubMed: 11240638]
- Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res* 2002;30:281–283. [PubMed: 11752315]
- Marinotti O, James A, Ribeiro JMC. Diet and salivation in female *Aedes aegypti* mosquitoes. *J. Insect Physiol* 1990;36:545–548.
- Megraw T, Kaufman TC, Kovalick GE. Sequence and expression of *Drosophila* Antigen 5-related 2, a new member of the CAP gene family. *Gene* 1998;222:297–304. [PubMed: 9831665]
- Moreira-Ferro CK, Marinotti O, Bijovsky AT. Morphological and biochemical analyses of the salivary glands of the malaria vector, *Anopheles darlingi*. *Tissue Cell* 1999;31:264–273. [PubMed: 10481298]
- Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* 1997;10:1–6. [PubMed: 9051728]
- Oliveira F, Kamhawi S, Seitz AE, Pham VM, Guigal PM, Fischer L, Ward J, Valenzuela JG. From transcriptome to immunome: identification of DTH inducing proteins from a *Phlebotomus ariasi* salivary gland cDNA library. *Vaccine* 2006;24:374–390. [PubMed: 16154670]
- Perutz MF, Finch JT, Berriman J, Lesk A. Amyloid fibers are water-filled nanotubes. *Proc Natl Acad Sci U S A* 2002;99:5591–5595. [PubMed: 11960014]
- Ribeiro JM, Andersen J, Silva-Neto MA, Pham VM, Garfield MK, Valenzuela JG. Exploring the sialome of the blood-sucking bug *Rhodnius prolixus*. *Insect Biochem. Mol. Biol* 2004a;34:61–79. [PubMed: 14976983]
- 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, 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* 2004b;34:543–563. [PubMed: 15147756]
- Ribeiro JMC, Sarkis JJJ, Rossignol PA, Spielman A. Salivary apyrase of *Aedes aegypti*: Characterization and secretory fate. *Comp. Biochem. Physiol* 1984;79B:81–86.
- Rossignol PA, Lueders AM. Bacteriolytic factor in the salivary glands of *Aedes aegypti*. *Comp. Biochem. Physiol* 1986;83B:819–822.
- Rothstein DM, Helmerhorst EJ, Spacciapoli P, Oppenheim FG, Friden P. Histatin-derived peptides: potential agents to treat localised infections. *Expert Opin. Emerg. Drugs* 2002;7:47–59. [PubMed: 15989535]
- Schaffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res* 2001;29:2994–3005. [PubMed: 11452024]
- Sharma D, Shinchuk LM, Inouye H, Wetzel R, Kirschner DA. Polyglutamine homopolymers having 8–45 residues form slablike beta-crystallite assemblies. *Proteins* 2005;61:398–411. [PubMed: 16114051]

- Shepard JJ, Andreadis TG, Vossbrinck CR. Molecular phylogeny and evolutionary relationships among mosquitoes (Diptera: Culicidae) from the northeastern United States based on small subunit ribosomal DNA (18S rDNA) sequences. *J Med Entomol* 2006;43:443–454. [PubMed: 16739399]
- Stintzi A, Heitz T, Prasad V, Wiedemann-Merdinoglu S, Kauffmann S, Geoffroy P, Legrand M, Fritig B. Plant 'pathogenesis-related' proteins and their role in defense against pathogens. *Biochimie* 1993;75:687–706. [PubMed: 8286442]
- Szyperski T, Fernandez C, Mumenthaler C, Wuthrich K. Structure comparison of human glioma pathogenesis-related protein GliPR and the plant pathogenesis-related protein P14a indicates a functional link between the human immune system and a plant defense system. *Proc. Natl. Acad. Sci. U S A* 95:2262–2266.
- Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, et al. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 2003;4:41. [PubMed: 12969510]
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997;25:4876–4882. [PubMed: 9396791]
- Valenzuela JG, Charlab R, Gonzalez EC, Miranda-Santos IKF, Marinotti O, Francischetti IM, Ribeiro JMC. The D7 family of salivary proteins in blood sucking Diptera. *Insect Mol. Biol* 2002a;11:149–155. [PubMed: 11966880]
- Valenzuela JG, Francischetti IM, Pham VM, Garfield MK, Ribeiro JM. Exploring the salivary gland transcriptome and proteome of the *Anopheles stephensi* mosquito. *Insect Biochem. Mol. Biol* 2003;33:717–732. [PubMed: 12826099]
- Valenzuela JG, Pham VM, Garfield MK, Francischetti IM, Ribeiro JMC. Toward a description of the sialome of the adult female mosquito *Aedes aegypti*. *Insect Biochem. Mol. Biol* 2002b;32:1101–1122. [PubMed: 12213246]



**Figure 1.**

Coomassie blue-stained PVDF membrane resulting from transfer of SDS-PAGE of 50 pairs of homogenized glands of adult *T. amboinensis*. Sample was heated for 10 min at 85°C and spun down before loading onto a NuPAGE gel (A) the supernatant and (B) the pelleted material. Molecular weight markers are shown on the left of each gel. Indicated to the right of each gel are the Edman degradation product obtained from the band and the best match to proteins deduced either from this work or from *Aedes aegypti* (indicated by *Aedes*). Lower cases in the peptide sequence indicate a mismatch in the sequence. PDI, protein disulfide isomerase; CR, calreticulin.