Insect Molecular Biology (2004) **13**(1), 73–88

# The transcriptome of adult female *Anopheles darlingi* salivary glands

**E. Calvo*, J. Andersen†, I. M. Francischetti†, M. deL. Capurro‡, A. G. deBianchi‡, A. A. James*, J. M. C. Ribeiro† and O. Marinotti***

**University of California, Irvine, Department of Molecular Biology and Biochemistry, Irvine, CA, USA;* †*National Institute of Allergy and Infectious Diseases, Section of Vector Biology, Laboratory of Malaria and Vector Research, Bethesda, MD, USA; and* ‡*Universidade de São Paulo, Departamento de Parasitologia – ICB, Avenue. Prof Lineu Prestes, São Paulo, Brazil*

## Abstract

***Anopheles* (*Nyssorhynchus*) *darlingi* is an important malaria vector in South and Central America; however, little is known about molecular aspects of its biology. Genomic and proteomic analyses were performed on the salivary gland products of *Anopheles darlingi*. A total of 593 randomly selected, salivary gland-derived cDNAs were sequenced and assembled based on their similarities into 288 clusters. The putative translated proteins were classified into three categories: (S) secretory products, (H) housekeeping products and (U) products with unknown cell location and function. Ninety-three clusters encode putative secreted proteins and several of them, such as an anophelin, a thrombin inhibitor, apyrases and several new members of the D7 protein family, were identified as molecules involved in haematophagy. Sugar-feeding related enzymes (α-glucosidases and α-amylase) also were found among the secreted salivary products. Ninety-nine clusters encode housekeeping proteins associated with energy metabolism, protein synthesis, signal transduction and other cellular functions. Ninety-seven clusters encode proteins with no similarity with known proteins. Comparison of the sequence divergence of the S and H categories of proteins of *An. darlingi* and *An. gambiae* revealed that the salivary proteins are less conserved than the housekeeping proteins, and therefore are changing at a faster evolutionary rate. Tabular and supplementary material containing the cDNA sequences and annotations are available at http://www.ncbi.nlm.nih.gov/projects/Mosquito/A_darlingi_sialome/**

**Keywords: *Anopheles darlingi*, salivary glands, proteome, transcriptome.**

## Introduction

The mosquito *Anopheles darlingi* (subgenus *Nyssorhynchus*) is an important vector of human malaria in South and Central America (Deane, 1986; Rubio-Palis & Zimmerman, 1997). The incidence of malaria has grown in this area during the last 30 years, attaining over a million cases annually, and in Brazil alone, four to five hundred thousand cases have been reported every year over the last decade (PAHO, 1998). The absence of an effective malaria vaccine, and the spread of drug-resistant *Plasmodium* parasites as well as insecticide resistance in vector populations (Crampton *et al.*, 1992), means there is urgent need for novel malaria control strategies. Rational approaches to new strategies are anticipated to originate from studies of insect physiology, immunology, biochemistry and molecular biology. These approaches will benefit from detailed understandings of interactions that occur between parasites and insects in all of these disciplines (Hurd, 1994).

Despite its importance as a malaria vector, little is known regarding the genome and proteome of *Anopheles darlingi*, mainly as a result of the inability to develop laboratory-adapted strains of the mosquito. Here we describe the transcriptome from the salivary glands of wild-caught female *An. darlingi*. We have selected the salivary glands as the organ to be studied because of its direct involvement in the transmission of malaria parasites to human hosts (Kappe *et al.*, 2003). Generation of a set of *An. darlingi* salivary gland cDNAs and deduced proteins provides indispensable tools for the systematic and comprehensive analysis of molecules that may play an active role in mosquito blood feeding and the pathogenesis of malaria.

Received 7 August 2003; accepted following revision 23 October 2003. Correspondence: O. Marinotti, University of California, Irvine, Department of Molecular Biology and Biochemistry, 2315 McGaugh Hall, Irvine, CA 92697–3900, USA. Tel.: +1 949 8243210; fax: +1 949 8242814; e-mail: omarinot@uci.edu

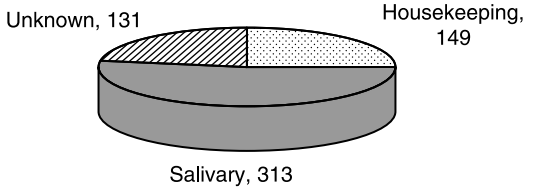
## Results and discussion

### *Organization of transcriptome information*

A total of 593 cDNA inserts were sequenced from a *An. darlingi* salivary gland cDNA library and these were assembled by the CAP program (Huang, 1992) into 289 clusters of related sequences. Using the BLAST package of programs (Altschul *et al*., 1997), we compared sequences for each cluster in the database with the nonredundant protein and nucleotide sets of the NCBI and Gene Ontology databases (Ashburner *et al*., 2000; Lewis *et al*., 2000; Hvidsten *et al*., 2001). Translated sequences also were screened with RPSBlast for protein motifs of the combined set of Pfam (Bateman *et al*., 2000) and SMART (Schultz *et al*., 2000) databases (also known as the Conserved Domains Database [CDD]). The sequences also were compared with the proteome set of the mosquito *An. gambiae* (available for FTP download at NCBI and other sites).

Finally, we submitted all translated sequences (starting with the first Met) to the Signal P server (Nielsen *et al*., 1997) to detect N-terminal amino acid sequences indicative of secretion signal peptides. With this information, the clustered database was annotated and classified into three categories of clusters: S, those associated with possible secreted products; H, those possibly associated with 'housekeeping' functions (metabolic activities anticipated in all cell types); and U, those of unknown function. Accordingly, ninety-nine cDNA clusters containing a total of 149 sequences (25.12% of the transcriptome) were classified as category H (Fig. 1). These clusters have an average of 1.5 sequences per cluster. This contrasts with the ninety-three clusters containing 313 sequences (52.78% of the transcriptome; average of 3.6 sequences per cluster) classified as category S. These average cluster sizes are statistically different from one another ($P < 0.01$, $\chi^2$ test), and consistent with what was observed in the salivary gland transcriptomes of *Ae. aegypti*, *An. gambiae*, *An. stephensi* and *Ixodes scapularis* (Francischetti *et al*., 2002b; Valenzuela *et al*., 2002b,c, 2003). Finally, ninety-seven clusters with 131 sequences (22.09% of the transcriptome; average of 1.3 sequences per cluster) were classified as category U.

A




B

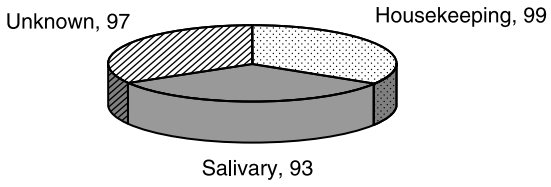



**Figure 1.** Numbers of sequences (A) or clusters (B) obtained from the 593 randomly selected clones from the adult female *Anopheles darlingi* salivary glands cDNA library. The sequences were classified as housekeeping, secretory or unknown functions. Absolute numbers of sequences or clusters are indicated on the diagrams.

### *Preliminary characterization of the salivary gland proteome of* An. darlingi

Parallel to the analysis of the salivary gland transcriptome of *An. darlingi*, sequence information was obtained for some of the most abundant proteins in the salivary glands of adult female mosquitoes. Proteins from fifteen salivary gland pairs were separated by SDS-PAGE, transferred to PVDF membranes, stained and the major polypeptides submitted to Edman degradation. Six polypeptides yielded useful information, five of which could be assigned to protein sequences predicted from our cluster database (Fig. 2). Other molecules did not yield useful amino acid sequences, either because they were blocked at their aminoterminal ends or because the Edman degradation resulted in uninformative low signals. The five clusters associated with the amino acid sequences generated from the Edman degradation reactions had between three and eleven cDNA sequences, with an average of six sequences per cluster, twice the average of the clusters in the S group. This result is consistent with the interpretation that the quantity of a particular protein correlates with its mRNA abundance. A good correlation between protein abundance and mRNA abundance was also observed in other analysed organisms (Futcher *et al*., 1999).

### *Description of secretory (S) category clusters*

Ninety-three clusters of sequences belonging to the S category were identified (Table 1). These clusters contain 1–44 cDNA sequences each and most belong to well-known families of proteins, although some do not have a known function.

### *D7-related proteins*

D7-related proteins, named after the prototype *Ae. aegypti* D7 protein (James *et al*., 1991), comprise a unique family found in the salivary glands of mosquitoes and sand flies, and are related distantly to the insect odorant-binding proteins (Hekmat-Scafe *et al*., 2000; Calvo *et al*., 2002; Valenzuela *et al*., 2002a). Two classes of D7-related proteins have been described: long (28–30 kDa), found in both mosquitoes and sand flies; and short (15–20 kDa), found so far only in mosquitoes (Arca *et al*., 1999; Calvo *et al*., 2002; Valenzuela *et al*., 2002a; Malafronte *et al*., 2003). Nine

**Table 1.** *An. darlingi* salivary glands cDNA clusters encoding for proteins probably secreted

| No. of sequences | Assembled contig | *E* value (best match to NR protein database) | Best match to AGPROT database | Percentage identity | Comments (similar to/putative function) |
|---|---|---|---|---|---|
| **Allergen/Antigen-5 related protein** | | | | | |
| 14 | AD-contig_214 | 1e–66 | CRA\|agCP8743 | 58 | 30 kDa allergen |
| 9 | AD-contig_66 | 3e–82 | CRA\|agCP6145 | 75 | antigen-5 |
| **D7 proteins** | | | | | |
| 44 | AD-contig_1 | 2e–67 | CRA\|agCP11198 | 61 | D7-related |
| 11 | AD-contig_230 | 5e–20 | CRA\|agCP11220 | 36 | D7-related |
| 9 | AD-contig_279 | 3e–23 | CRA\|agCP11196 | 36 | D7-related |
| 7 | AD-contig_77 | 5e–13 | CRA\|agCP11196 | 28 | D7-related |
| 5 | AD-contig_159 | 2e–82 | CRA\|agCP10845 | 50 | D7-related |
| 4 | AD-contig_229 | 4e–20 | CRA\|agCP11220 | 37 | D7-related |
| 2 | AD-contig_256 | 3e–7 | CRA\|agCP11220 | 28 | D7-related, short |
| 1 | AD-contig_228 | 6e–12 | CRA\|agCP2228 | 42 | D7clu5 short form (hamadarin) |
| 1 | AD-contig_278 | 6e–21 | CRA\|agCP11220 | 40 | short form D7clu5 |
| **Enzymes and enzymes inhibitors linked to blood and sugar meals** | | | | | |
| 5 | AD-contig_137 | 4e–33 | CRA\|agCP11208 | 43 | anophelin |
| 2 | AD-contig_253 | 4e–78 | CRA\|agCP9757 | 49 | apyrase |
| 3 | AD-contig_217 | 4e–75 | CRA\|agCP10591 | 81 | apyrase (full clone) |
| 3 | AD-contig_224 | 1e–82 | CRA\|agCP10591 | 54 | apyrase (truncated clone) |
| 1 | AD-contig_45 | 2e–32 | CRA\|agCP7190 | 50 | salivary peroxidase |
| 1 | AD-contig_115 | 1e–11 | CRA\|agCP14623 | 38 | thrombin inhibitor infestin precursor |
| 6 | AD-contig_88 | 1e–115 | CRA\|agCP12790 | 58 | maltase |
| 1 | AD-contig_265 | 1e–82 | CRA\|agCP12790 | 82 | maltase-like protein |
| 1 | AD-contig_65 | 4e–21 | EBI\|8952 | 54 | maltase-like protein |
| 1 | AD-contig_106 | 1e–24 | CRA\|agCP1208 | 36 | alpha-amylase |
| 1 | AD-contig_89 | 1e–6 | CRA\|agCP12790 | 37 | probable maltase precursor |
| 1 | AD-contig_124 | 2e–17 | CRA\|agCP12065 | 49 | salivary glucosidase |
| 15 | AD-contig_123 | 9e–32 | CRA\|agCP12065 | 49 | salivary glucosidase |
| 4 | AD-contig_125 | 2e–17 | CRA\|agCP12065 | 37 | salivary glucosidase |
| **Mucin-like proteins** | | | | | |
| 8 | AD-contig_242 | 1e–17 | CRA\|agCP1772 | 51 | mucin |
| 5 | AD-contig_244 | 5e–16 | CRA\|agCP1772 | 40 | mucin |
| 3 | AD-contig_220 | 1e–10 | CRA\|agCP7687 | 31 | mucin |
| 1 | AD-contig_243 | 1e–17 | CRA\|agCP1772 | 38 | mucin-like |
| 1 | AD-contig_51 | 2e–59 | CRA\|agCP3409 | 67 | peritrophin 1 (mucin-like peritrophin) |
| 1 | AD-contig_34 | 4e–70 | CRA\|agCP14528 | 70 | mucin? |
| **Possibly related to immunity** | | | | | |
| 10 | AD-contig_266 | 4e–15 | CRA\|agCP7505 | 51 | cecropin |
| 5 | AD-contig_181 | 9e–24 | CRA\|agCP7503 | 50 | cecropin |
| 4 | AD-contig_203 | 3e–41 | CRA\|agCP6915 | 78 | defensin |
| 2 | AD-contig_239 | 1e–22 | EBI\|7267 | 32 | putative infection responsive short peptide |
| 1 | AD-contig_129 | 3e–17 | CRA\|agCP7503 | 67 | antibiotic peptide cecropin A2 |
| 1 | AD-contig_166 | 0.012 | CRA\|agCP14093 | 85 | putative gram negative bacteria binding protein |
| 1 | AD-contig_267 | 2e–15 | CRA\|agCP7505 | 45 | cecropin CecC |
| 1 | AD-contig_180 | 2e–81 | CRA\|agCP9741 | 64 | T-cell immunomodulatory protein |
| 1 | AD-contig_59 | 2e–38 | CRA\|agCP3859 | 57 | lysozyme |
| **SG family of anopheline salivary proteins** | | | | | |
| 3 | AD-contig_199 | 2e–18 | CRA\|agCP13537 | 37 | gSG1b |
| 7 | AD-contig_28 | 0.084 | CRA\|agCP6138 | 32 | gSG2 |
| 1 | AD-contig_29 | 0.044 | CRA\|agCP6138 | 31 | gSG2 protein |
| 3 | AD-contig_207 | 1e–29 | CRA\|agCP11109 | 44 | gSG7 |
| 1 | AD-contig_206 | 1e–22 | CRA\|agCP11109 | 41 | gSG7 |
| 2 | AD-contig_247 | 1e–24 | CRA\|agCP2222 | 54 | gSG7 |
| 4 | AD-contig_202 | 3e–31 | CRA\|agCP7185 | 54 | gSG8 |
| 11 | AD-contig_255 | 1e–20 | CRA\|agCP13467 | 32 | SG1 family |
| 16 | AD-contig_201 | 2e–29 | CRA\|agCP6430 | 42 | SG3 family |
| **Chitinase** | | | | | |
| 1 | AD-contig_40 | 6e–26 | CRA\|agCP6090 | 55 | chitinase |
| **Similar to previously described salivary An. gambiae proteins of unknown function** | | | | | |
| 5 | AD-contig_16 | 1e–6 | CRA\|agCP13582 | 38 | hypothetical protein 15 |
| 5 | AD-contig_192 | 0.012 | CRA\|agCP15011 | 52 | no match in Ag (nonannotated protein) |
| 1 | AD-contig_35 | 6e–28 | CRA\|agCP8839 | 33 | putative 53.7 kDa salivary protein |
| 2 | AD-contig_15 | 1e–4 | CRA\|agCP13582 | 45 | unknown |
| 1 | AD-contig_24 | 7e–8 | EBI\|4655 | 29 | pectinesterase |
| 1 | AD-contig_78 | 5e–4 | CRA\|agCP11977 | 35 | unknown |
| 5 | AD-contig_100 | 3e–24 | CRA\|agCP8099 | 41 | unknown |
| 1 | AD-contig_189 | 7e–17 | CRA\|agCP11339 | 93 | unknown |

**Table 1.** (*Continued*)

| No. of sequences | Assembled contig | *E* value (best match to NR protein database) | Best match to AGPROT database | Percentage identity | Comments (similar to/putative function) |
|---|---|---|---|---|---|
| 1 | AD-contig_285 | 1e–4 | CRA\|agCP13137 | 56 | unknown |
| 1 | AD-contig_281 | 3e–33 | CRA\|agCP3858 | 51 | unknown |
| 1 | AD-contig_276 | 5e–28 | CRA\|agCP1678 | 53 | unknown |
| 1 | AD-contig_87 | 7e–7 | CRA\|agCP2550 | 53 | unknown |
| 1 | AD-contig_262 | 1e–37 | CRA\|agCP2969 | 73 | unknown |
| 1 | AD-contig_73 | 9e–10 | CRA\|agCP11977 | 45 | unknown |
| 1 | AD-contig_274 | 6e–27 | CRA\|agCP15376 | 52 | unknown |
| 1 | AD-contig_288 | 7e–62 | CRA\|agCP12405 | 88 | unknown |
| 1 | AD-contig_169 | 8e–4 | CRA\|agCP1607 | 78 | unknown |
| 1 | AD-contig_67 | 2e–4 | CRA\|agCP4709 | 38 | unknown |
| 1 | AD-contig_39 | 5e–8 | CRA\|agCP7243 | 90 | unknown |
| 1 | AD-contig_136 | 5e–13 | CRA\|agCP5458 | 54 | unknown |
| 1 | AD-contig_198 | 0.092 | EBI\|6164 | 46 | unknown |
| 1 | AD-contig_30 | | CRA\|agCP6138 | 31 | unknown |
| 1 | AD-contig_10 | | CRA\|agCP11711 | 30 | unknown |
| **Putative secreted proteins with unknown function** | | | | | |
| 6 | AD-contig_3 | 0.003 | No match | | hypothetical salivary protein 8.2 |
| 4 | AD-contig_2 | 0.003 | No match | | hypothetical salivary protein 8.2 |
| 1 | AD-contig_68 | 0.002 | No match | | unknown |
| 1 | AD-contig_71 | 0.035 | No match | | unknown |
| 1 | AD-contig_69 | 2e–4 | No match | | unknown |
| 1 | AD-contig_58 | | No match | | unknown |
| 1 | AD-contig_286 | | No match | | unknown |
| 1 | AD-contig_111 | | No match | | unknown |
| 1 | AD-contig_158 | | No match | | unknown |
| 4 | AD-contig_209 | | No match | | unknown |
| 1 | AD-contig_114 | | No match | | unknown |
| 1 | AD-contig_116 | | No match | | unknown |
| 1 | AD-contig_134 | | No match | | unknown |
| 1 | AD-contig_144 | | No match | | unknown |
| 1 | AD-contig_154 | | No match | | unknown |
| 1 | AD-contig_284 | | No match | | unknown |
| 1 | AD-contig_72 | | No match | | unknown |
| 1 | AD-contig_135 | | No match | | unknown |
| 1 | AD-contig_63 | | No match | | unknown |
| 1 | AD-contig_44 | | No match | | unknown |

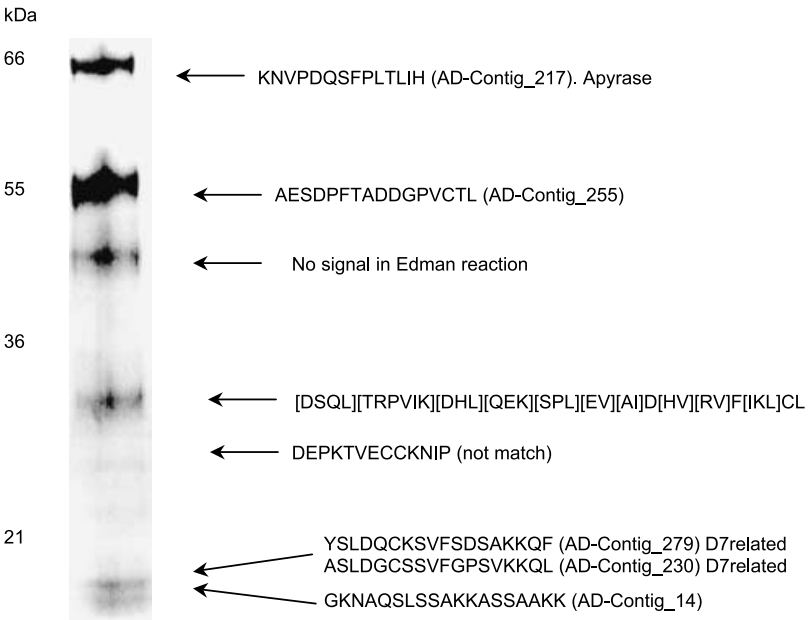



**Figure 2.** SDS-PAGE of *Anopheles darlingi* salivary gland proteins. Molecular weight markers are shown to the left and the amino acid sequences obtained by Edman degradation to the right. The corresponding cluster is indicated next to each amino acid sequence.

clusters coding for proteins of the D7 family were identified in the *An. darlingi* salivary gland transcriptome, indicating that several D7-related genes exist in the genome of this mosquito. The occurrence of multiple D7-related genes and their organization in the genome in *An. gambiae* was reported previously (Arca *et al*., 2002). The most abundant cluster (forty-four sequences) codes for the recently described protein, Anda-D7r3-2 (Calvo *et al*., 2002). Anda-D7r3-2 has 61% amino acid identity to the reported short-form D7clu2 salivary protein of *An. gambiae* (CRA|agCP11198) and 14% identity to the long D7 protein of *Ae. aegypti* (gi|159559) (James *et al*., 1991).

The D7-related proteins may inhibit activation of host plasma, as was observed for hamadarin, a D7-related protein of *An. stephensi* saliva recently characterized as an inhibitor of Factor XII (Isawa *et al*., 2002). In addition, they may have different functions, as is the case with *Rhodnius* salivary lipocalins (Montfort *et al*., 2000). Several closely related proteins of the lipocalin family exist in the saliva of *R. prolixus*, each of which has a different antihaemostatic activity (Andersen *et al*., 2002; Francischetti *et al*., 2002a). This diversity probably reflects a scenario of gene duplication and divergence of function (Sankoff, 2001).

Consistent with the abundance of cDNA sequences in the D7 family of clusters, two of eight major *An. darlingi* salivary gland proteins produced amino-terminal sequences matching D7-related proteins (Fig. 2). The sequence, ASLDGCSS, found in the 16–18 kDa region of the gel, matches the previously reported D7clu2 of *An. gambiae* (Francischetti *et al*., 2002b). The other amino-terminal sequence, YSLDQCKS, matches proteins of the previously reported *An. gambiae* D7clu5 protein (Francischetti *et al*., 2002b). No function has been discovered yet for these proteins. The abundance of D7 proteins observed by SDS-PAGE also is consistent with prior findings in which D7 was isolated as a major protein in the acidic soluble fractions of *An. darlingi* salivary glands (Calvo *et al*., 2002).

### *Mucins*

The most abundant cluster in this group (eight clones sequenced) codes for a protein with 50% identity to a putative salivary mucin of *An. stephensi* (AAO06835). This protein also is similar to the Trypanosomal mucin-like glycoproteins (Di Noia *et al*., 1998). Molecules belonging to this family of trypanosomal proteins resemble vertebrate mucins and their amino acid sequences consist of three regions. The amino and carboxyl termini are conserved among all members of the family, whereas the central region is not well conserved and contains a large number of threonine residues, some of which can be glycosylated. Indirect evidence is interpreted to suggest that these genes might encode the core protein of parasite mucins, glycoproteins that may be involved in the interaction with, and invasion of, mammalian host cells (Hansen *et al*., 1998).

Five other clusters also were identified tentatively as encoding mucin-like proteins (Table 1). These components of mosquito saliva could function as lubricants of the salivary canal and also could have other activities such as modulation of macrophages, as is the case with surface mucins of *Trypanosoma cruzi* (Acosta-Serrano *et al*., 2001; Ropert *et al*., 2002).

### *Antigen 5-related proteins*

Proteins related to this family are found in the venom glands of Hymenoptera (Henriksen *et al*., 2001) and in the salivary glands of sand flies (Charlab *et al*., 1999), tsetse flies (Li *et al*., 2001) and mosquitoes (Francischetti *et al*., 2002b; Valenzuela *et al*., 2002c). The *An. gambiae* antigen-5 protein is expressed specifically in the medial lobes of the female salivary glands of that mosquito (Arca *et al*., 1999). Antigen-5-related salivary products are members of a group of secreted proteins that belong to the CAP family (cysteine-rich secretory proteins; antigen-5 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) and to antifungal proteins in plants (Stintzi *et al*., 1993; Szyperski *et al*., 1998). One cDNA cluster, with nine clones sequenced, encodes a protein with similarity to other antigen-5-related molecules (Table 1). The *An. darlingi* antigen-5-related protein has > 75% identity to *An. gambiae* agCP6145 and is 74% identical to the salivary antigen-5-related protein 1 of the same mosquito (Francischetti *et al*., 2002b).

The general function for CAP proteins is controversial (Schreiber *et al*., 1997). Some of the CAP family proteins function as protease inhibitors (Megraw *et al*., 1998). As for the physiology of blood-sucking insects, these molecules could represent an adaptive response to inhibit coagulation, complement activation or any other component of the vertebrate host haemostasis potentially harmful to the mosquitoes. However, proteins from this family have also been associated with proteolytic activity in *Echinococcus granulosus* (Lorenzo *et al*., 2003) and *Conus textile* (Milne *et al*., 2003).

### *30 kDa antigen*

One of the most abundant clusters (fourteen sequences) encodes a product similar to the protein agCP8743 of *An. gambiae* (58% identity) and the salivary 30 kDa protein gi|18389879 (47% identity) of the same mosquito. This mosquito salivary protein was first described in *Ae. aegypti* mosquitoes (Simons & Peng, 2001) and later found to be in the salivary glands of *An. gambiae* (Francischetti *et al*., 2002b). It has a long region of low amino acid complexity, consisting mainly of Gly and Glu residues. The clone identified here may represent the homologous protein in *An. darlingi*. The function of this protein is unknown.

*The SG family*

This family of anopheline salivary gland proteins, described as SG or gSG proteins (Arca *et al*., 1999; Lanfrancotti *et al*., 2002), does not yield significant similarities (by BLASTP) to other proteins in the NCBI database, except among its own members. This family also includes the distantly related salivary *An. gambiae* TRIO protein (Francischetti *et al*., 2002b). TRIO is a multidomain protein that binds the lymphocyte-activating receptor transmembrane tyrosine phosphatase (PTPase) and contains a protein kinase domain. It was proposed that TRIO may orchestrate cell–matrix and cytoskeletal rearrangements necessary for cell migration (Lin & Greenberg, 2000).

Two other cDNA clusters (fourteen sequences) of the *An. darlingi* salivary gland library yielded similarities to *An. gambiae* proteins annotated as members of the SG1 family (Table 1). All proteins belonging to this family, including the conceptual product of the SG1-like polypeptide identified from *An. darlingi* cDNA clones, have a clear signal peptide indicative of secretion.

The *An. darlingi* transcriptome contains two clusters of cDNA (eight sequences) having sequence similarity to *An. gambiae* SG2 and gSG2 proteins (Table 1). This family, first described in *An. gambiae* as SG2, gSG2 and SG2-like protein (Arca *et al*., 1999; Lanfrancotti *et al*., 2002), consists of proteins 114–168 amino acids in length that are rich in Gly or Asn. High protein similarity matches in the NCBI database are only produced among those anopheline proteins that belong to the SG2 group. These two clusters may represent an mRNA corresponding to the *An. darlingi* gene homologous to the *An. gambiae* SG2 gene. It is interesting that *Ixodes scapularis* salivary glands also contain glycine-rich peptides of equivalent size (Valenzuela *et al*., 2002b). Their function also is unknown.

Finally, we found three clusters (twenty-four sequences) having similarity to other proteins of the SG salivary protein families (gSG3, gSG7 and gSG8) of *An. gambiae*. This subgroup of proteins is the second most abundant in the *An. darlingi* transcriptome. Proteins of the SG family, by being unique to anopheline mosquitoes, may become a useful immunological marker for exposure to this mosquito group by humans and animals.

*Enzymes and inhibitors associated with antihaemostatic activities*

Six clusters (twenty-five sequences), with an average of 2.5 clones sequenced per cluster, encode products that are similar to enzymes and enzyme inhibitors related to haematophagy. Salivary peroxidases, which act as vasodilators, were found in the salivary glands of anopheline mosquitoes (Ribeiro & Nussenzveig, 1993; Ribeiro *et al*., 1994; Ribeiro & Valenzuela, 1999). One cDNA cluster, consisting of only one sequence, matches the *An. albimanus* salivary peroxidase and may be related to the *An. gambiae* protein agCP7190.

Three different clusters (eight sequences) encode proteins having 49–89% identity to the *An. gambiae* salivary apyrase (agCP10591). One cluster, containing three sequences, represents truncated clones matching the apyrase gene product. In support of the existence of salivary apyrases in *An. darlingi*, an aminoterminal sequence (KNVPDQ) that matches the putative salivary apyrase gene product was found in the 60 kDa region of the SDS-PAGE of salivary glands (Fig. 2). Furthermore, apyrase activity was described previously in the salivary glands of this mosquito (Moreira-Ferro *et al*., 1999). Apyrases are a polyphyletic group of enzymes found ubiquitously in the salivary glands of blood-feeding insects and ticks. Apyrases degrade the neutrophil-inducing substance ATP, and the platelet-aggregating nucleotide ADP to AMP, presumably facilitating blood feeding. In *Ae. aegypti*, apyrase is a member of the 5′-nucleotidase family (Champagne *et al*., 1995). In *An. gambiae*, two such genes are expressed in the salivary glands and annotated as apyrase and 5′-nucleotidase; however, both could actually be coding for proteins with apyrase activity (Arca *et al*., 1999; Lombardo *et al*., 2000).

The *An. darlingi* salivary transcriptome has one cDNA cluster with five clones sequenced encoding a protein similar to antithrombins of the anophelin family (Valenzuela *et al*., 1999; Francischetti *et al*., 1999). The putative *An. darlingi* anophelin peptide is 43%, 41% and 86% identical to the homologues of *An. gambiae*, *An. stephensi* and *An. albimanus*, respectively.

Another cluster within the *An. darlingi* transcriptome encodes a protein with 38% identity to the *An. gambiae* agCP14623 protein and 52% identity to the thrombin inhibitor, infestin, of *Triatoma infestans* (Campos *et al*., 2002). The infestin gene encodes a protein with four nonclassical Kazal-type domains with an apparent molecular mass of 22 kDa, and belongs to a family of serine protease inhibitors. This protein, found in the *T. infestans* midgut, showed inhibitory activities towards thrombin and trypsin (Campos *et al*., 2002). Surprisingly, infestin inhibited not only thrombin and trypsin, but also factor XIIa, factor Xa and plasmin, inhibiting blood clotting during the blood meal.

*Sugar-meal digestion*

Mosquito salivary glands secrete enzymes such as maltases or α-glucosidases (*Maltase-like 1* [*Mal*l] gene, James *et al*., 1989; Marinotti *et al*., 1990) and possibly α-amylases (*Amy*l gene, Grossman & James, 1993) that help in sugar-meal digestion. The *An. darlingi* transcriptome has eight clusters, represented by thirty sequences, with similarity to enzymes linked to sugar digestion in other mosquitoes.

Four different clusters (nine sequences) showed similarity to the *Mal*l gene of *Ae. aegypti* (James *et al*., 1989) and to the *An. gambiae* protein agCP12790. This is consistent

with the finding that the α-glucosidase activity has been demonstrated in the salivary glands of *An. darlingi* (Moreira-Ferro *et al*., 1999). *Anopheles stephensi* also expresses four gene products related to maltases in its salivary glands (Valenzuela *et al*., 2003).

Additionally, we sequenced one clone with similarity to α-amylases of *Ae. aegypti* (Grossman & James, 1993) and 36% identity with the protein agCP1208 of *An. gambiae*. In *Ae. aegypti*, the *Amylase* I gene (*Amy*I) is expressed specifically in the salivary glands and its function has been proposed to be involved with carbohydrate metabolism. However, amylase activity is detected at a very low level in *Ae. aegypti* salivary gland extracts (Grossman & James, 1993).

Finally, in this group we sequenced twenty clones included in three different clusters coding for proteins with similarity to salivary glucosidases. The most abundant cluster in this group has fifteen sequences with 49% identity to the *An. gambiae* protein agCP12065 and 52% identical to the cellulose 1,4-beta-cellobiosidase from *Xylella fastidiosa* (NP_778753). This protein also could be involved in sugar digestion in *An. darlingi* mosquitoes.

*Putative immunity-related products*

Four clusters, ranging from one to ten sequences in each cluster, encode proteins with similarity to the antimicrobial peptide Cecropin family (Cec-A, -B and -C) of *An. gambiae* (Vizioli *et al*., 2000; Zheng & Zheng, 2002). Cecropins are small antibacterial peptides first isolated from the lepidopteran *Hyalophora cecropia* (Steiner *et al*., 1981), and have since been isolated from various insects, such as *Drosophila melanogaster* (Kylsten *et al*., 1990), *Ae. aegypti* (Lowenberger *et al*., 1999) and *An. gambiae* (Vizioli *et al*., 2000). Cecropins and their derivatives have a wide range of antimicrobial targets, including *Candida albicans* (Park *et al*., 1997) to *Plasmodium* parasites (Gwadz *et al*., 1989; Rodriguez *et al*., 1995), *Trypanosoma cruzi* (Barr *et al*., 1995), *Leishmania* (Akuffo *et al*., 1998) and the filarial worm, *Brugia pahangi* (Chalk *et al*., 1995). Cecropin expression is induced by bacterial challenge, but may be produced constitutively in the salivary glands to protect the sugar meal from microbial fermentation.

Another cluster, with four sequences, encodes a peptide similar to *An. gambiae* Defensin (78% identity). The expression of Defensin is predominantly induced in the mosquito fat body shortly after bacterial challenge. It also is induced locally in the midgut and salivary gland epithelia upon invasion by malaria parasites, suggesting that Defensin may have a broad role in the defence against both microbes and parasites (Richman *et al*., 1996, 1997).

Lysozyme, an antibacterial enzyme first described as a mosquito salivary activity in *Ae. aegypti* (Rossignol & Lueders, 1986), also was found in salivary glands of *An. darlingi* (Moreira-Ferro *et al*., 1998). Only one cluster with one sequence is similar to the *An. gambiae* LYC_ANOGA lysozyme precursor (66% identity). This result is in contrast with the *An. stephensi* transcriptome (Valenzuela *et al*., 2003), where an abundant cluster (fourteen sequences) was found. Salivary lysozyme may help to deter bacterial growth in sugar meals of mosquitoes, which are stored in the crop.

Three additional clusters observed in the mosquito salivary glands that encode putative secreted proteins may be involved with immunity. One cluster (one sequence) encodes a product similar to the *An. gambiae* protein (AgCP14093) and also similar to Gram-negative binding protein (AAM73871), the second cluster (two sequences) encodes a putative infection-responsive short peptide, similar to the gambicins of *An. gambiae* (Vizioli *et al*., 2001) and *Culex pipiens* (AA038515), and the third cluster (one sequence) encodes a product similar to the *An. gambiae* agCP9741 protein (64% identity) and also is similar to a T-cell immunomodulatory protein of *Homo sapiens*.

Expression of the mRNAs encoding Gram-negative bacteria binding proteins in mosquito salivary glands was described for *An. gambiae* (Dimopoulos *et al*., 1997) and *Ae. aegypti* (Valenzuela *et al*., 2002c). Gambicin was first described in *An. gambiae* by Vizioli *et al*. (2001). This molecule is an antimicrobial peptide that lacks sequence homology with other known immune-related proteins. The mature peptide can kill both Gram-positive and Gram-negative bacteria. The T-cell immunomodulatory protein function in *An. darlingi* salivary glands is unknown; however, immunomodulatory activity in the saliva of the mosquito *Ae. aegypti* has been described (Cross *et al*., 1994).

*Description of secretory products with unidentified function in* An. darlingi

The *An. darlingi* salivary gland cDNA library yielded twenty-three clusters of sequences similar to predicted proteins from the *An. gambiae* genome but not previously described in salivary gland transcriptomes. These hypothetical proteins do not yield significant matches to other proteins in the NR protein database and are yet to be characterized.

Furthermore, twenty cDNA clusters, with thirty-one clones sequenced, encode proteins predicted to have a signal peptide indicative of secretion; however, these do not match other known proteins, even when the BLAST filter to exclude low-complexity sequences was removed. When the protein sequences were compared with the *An. gambiae* genome (using tblastn), they produced no matches, indicating that these are novel proteins or truncated clones.

*Description of housekeeping (H) category clusters*

Of the ninety-nine H category clusters in the salivary transcriptome of *An. darlingi* (Table 2), forty-five clusters correspond to genes involved in protein synthesis and secretion including those that encode rRNA, ribosomal proteins, mitochondrial proteins and Golgi vesicular membrane trafficking proteins. Twenty-two clusters (thirty-six clones

**Table 2.** *An. darlingi* salivary glands cDNA clusters encoding proteins associated with housekeeping function

| No. of sequences | Assembled contig | *E* value (best match to NR protein database) | Best match to AGPROT database | Percentage identity | Comments identity (similar to/putative function) |
|---|---|---|---|---|---|
| **Heat shock-like protein** | | | | | |
| 1 | AD-contig_32 | 1e–63 | CRA\|agCP3435 | 79 | heat shock protein |
| 2 | AD-contig_246 | 3e–49 | CRA\|agCP12309 | 95 | heat shock protein cognate 4 |
| 1 | AD-contig_191 | 3e–21 | CRA\|agCP12309 | 67 | heat shock-like protein |
| 1 | AD-contig_81 | 6e–42 | CRA\|agCP11787 | 92 | chaperonin containing TCP1 |
| 1 | AD-contig_128 | 2e–27 | CRA\|agCP11981 | 91 | chaperonin-heat shock |
| **Metabolism related proteins** | | | | | |
| 1 | AD-contig_121 | 1e–15 | CRA\|agCP8069 | 74 | *N*-acetyltransferase |
| 1 | AD-contig_57 | 6e–47 | CRA\|agCP12915 | 55 | ADP-ribosylation-like factor 6 |
| 1 | AD-contig_104 | 1e–82 | CRA\|agCP8658 | 70 | dimeric dihydrodiol dehydrogenase |
| 1 | AD-contig_31 | 5e–39 | CRA\|agCP8498 | 95 | glyceraldehyde-3-phosphate dehydrogenase |
| 1 | AD-contig_287 | 4e–41 | CRA\|agCP10011 | 75 | glycosylasparaginase |
| 1 | AD-contig_119 | 5e–59 | CRA\|agCP3334 | 92 | glutamate carboxypeptidase-like |
| 1 | AD-contig_289 | 1e–63 | CRA\|agCP15559 | 96 | NADH dehydrogenase (ubiquinone) |
| 1 | AD-contig_76 | 2e–36 | CRA\|agCP3297 | 97 | NADH dehydrogenase (ubiquinone) |
| 1 | AD-contig_85 | 9e–65 | CRA\|agCP3151 | 82 | NADH dehydrogenase (ubiquinone) |
| 1 | AD-contig_82 | 6e–29 | CRA\|agCP2296 | 86 | NADH dehydrogenase (ubiquinone) |
| 1 | AD-contig_46 | 2e–70 | CRA\|agCP6292 | 84 | *N*-methyl-D-aspartate receptor-associated protein |
| 1 | AD-contig_11 | 2e–53 | CRA\|agCP1214 | 91 | peptidylprolyl cis-trans isomerase |
| 1 | AD-contig_268 | 3e–99 | CRA\|agCP3166 | 94 | peroxidase (antioxidant enzyme) |
| 1 | AD-contig_48 | 5e–42 | EBI\|9772 | 87 | peroxidase (antioxidant enzyme) |
| 1 | AD-contig_60 | 9e–73 | EBI\|2906 | 72 | phosphotidylinositol 3 kinase |
| 1 | AD-contig_26 | 2e–25 | CRA\|agCP6680 | 49 | polyhydroxyalkanoate synthesis protein |
| 1 | AD-contig_13 | 9e–24 | CRA\|agCP10881 | 52 | similar to programmed cell death |
| 1 | AD-contig_261 | 8e–92 | CRA\|agCP3730 | 89 | succinate dehydrogenase B |
| 1 | AD-contig_151 | 2e–64 | CRA\|agCP4843 | 72 | ubiquinol-cytochrome C reductase |
| 4 | AD-contig_200 | 1e–44 | CRA\|agCP13749 | 59 | ubiquitin |
| 1 | AD-contig_168 | 2e–28 | | | NADH dehydrogenase subunit |
| 1 | AD-contig_161 | 0.011 | | | NADH2 dehydrogenase (ubiquinone) |
| 2 | AD-contig_238 | 4e–31 | | | ATP-synthase |
| 1 | AD-contig_36 | 3e–79 | | | cytochrome b oxydase |
| 5 | AD-contig_148 | 4e–65 | | | cytochrome oxydase |
| 3 | AD-contig_221 | 5e–41 | | | cytochrome b oxydase |
| **Protein synthesis and secretion** | | | | | |
| 4 | AD-contig_208 | 1e–40 | CRA\|agCP1641 | 81 | ribosomal protein |
| 2 | AD-contig_235 | 1e–71 | CRA\|agCP14911 | 87 | ribosomal protein |
| 1 | AD-contig_146 | 4e–58 | CRA\|agCP11398 | 56 | ribosomal protein |
| 2 | AD-contig_231 | 5e–42 | CRA\|agCP11536 | 91 | ribosomal protein |
| 1 | AD-contig_37 | 8e–45 | CRA\|agCP8317 | 64 | ribosomal protein |
| 2 | AD-contig_237 | 6e–70 | CRA\|agCP14909 | 84 | ribosomal protein |
| 2 | AD-contig_226 | 4e–41 | CRA\|agCP3608 | 88 | ribosomal protein |
| 1 | AD-contig_280 | 7e–95 | CRA\|agCP5980 | 79 | ribosomal protein |
| 2 | AD-contig_251 | 2e–43 | CRA\|agCP6680 | 68 | ribosomal protein |
| 2 | AD-contig_257 | 3e–56 | CRA\|agCP10609 | 73 | ribosomal protein |
| 2 | AD-contig_241 | 6e–45 | CRA\|agCP1535 | 78 | ribosomal protein |
| 2 | AD-contig_234 | 6e–84 | CRA\|agCP10687 | 94 | ribosomal protein |
| 5 | AD-contig_126 | 9e–47 | CRA\|agCP14068 | 82 | ribosomal protein |
| 1 | AD-contig_64 | 3e–76 | CRA\|agCP1538 | 89 | ribosomal protein |
| 1 | AD-contig_21 | 1e–40 | CRA\|agCP14472 | 87 | ribosomal protein |
| 1 | AD-contig_132 | 1e–45 | CRA\|agCP9994 | 83 | ribosomal protein |
| 1 | AD-contig_6 | 1e–8 | CRA\|agCP11155 | 96 | ribosomal protein |
| 1 | AD-contig_122 | 1e–102 | CRA\|agCP9554 | 82 | ribosomal protein |
| 1 | AD-contig_33 | 8e–63 | CRA\|agCP12827 | 65 | ribosomal protein |
| 1 | AD-contig_92 | 4e–26 | CRA\|agCP10200 | 82 | ribosomal protein |
| 3 | AD-contig_222 | 9e–78 | CRA\|agCP12166 | 97 | ribosomal protein |
| 1 | AD-contig_113 | 1e–41 | CRA\|agCP8681 | 98 | ribosomal protein |
| 1 | AD-contig_147 | 5e–56 | CRA\|agCP13921 | 93 | ribosomal protein |
| 1 | AD-contig_9 | 1e–76 | CRA\|agCP12071 | 89 | ribosomal protein |
| 1 | AD-contig_98 | 4e–59 | CRA\|agCP1749 | 83 | ribosomal protein |
| 1 | AD-contig_97 | 9e–33 | CRA\|agCP14988 | 58 | ribosomal protein |
| 1 | AD-contig_193 | 1e–104 | CRA\|agCP7923 | 93 | ribosomal protein |
| 1 | AD-contig_4 | 1e–71 | CRA\|agCP9782 | 80 | ribosomal protein |
| 1 | AD-contig_47 | 9e–10 | CRA\|agCP9264 | 93 | ribosomal protein |
| 1 | AD-contig_149 | 2e–41 | CRA\|agCP3167 | 81 | ribosomal protein |

**Table 2.** (*Continued*)

| No. of sequences | Assembled contig | *E* value (best match to NR protein database) | Best match to AGPROT database | Percentage identity | Comments identity (similar to/putative function) |
|---|---|---|---|---|---|
| 1 | AD-contig_96 | 3e–62 | CRA\|agCP13543 | 72 | ribosomal protein |
| 1 | AD-contig_50 | 1e–28 | CRA\|agCP5982 | 95 | ribosomal protein |
| 3 | AD-contig_219 | 1e–32 | CRA\|agCP4788 | 47 | ribosomal protein |
| 3 | AD-contig_223 | 2e–85 | CRA\|agCP7766 | 90 | ribosomal protein |
| 1 | AD-contig_179 | 1e–73 | CRA\|agCP7049 | 97 | ribosomal protein |
| 3 | AD-contig_213 | 2e–54 | CRA\|agCP11873 | 92 | ribosomal protein |
| 3 | AD-contig_225 | 5e–95 | CRA\|agCP6972 | 90 | ribosomal protein |
| 2 | AD-contig_248 | 1e–19 | CRA\|agCP6003 | 38 | ribosomal protein |
| 1 | AD-contig_12 | 8e–17 | CRA\|agCP4228 | 62 | ribosomal protein |
| 1 | AD-contig_93 | 8e–49 | CRA\|agCP1091 | 79 | RRM (RNA recognition motif) |
| 1 | AD-contig_19 | 4e–44 | CRA\|agCP9326 | 79 | Golgi vesicular membrane trafficking protein p18 |
| 1 | AD-contig_22 | 2e–73 | CRA\|agCP11339 | 93 | similar to Reticulon protein 3 |
| 1 | AD-contig_107 | 2e–68 | CRA\|agCP9383 | 81 | probable microsomal signal peptidase 25 kDa subunit |
| 8 | AD-contig_54 | | | | rRNA, mitochondrial |
| 2 | AD-contig_258 | 1e–5 | | | ribosomal protein |
| **Signalling/transport** | | | | | |
| 1 | AD-contig_56 | 2e–48 | CRA\|agCP12773 | 89 | calcyclin, signal transduction |
| 1 | AD-contig_185 | 4e–77 | CRA\|agCP10575 | 97 | calmodulin |
| 1 | AD-contig_79 | 5e–66 | CRA\|agCP8333 | 87 | translocase of inner mitochondrial membrane |
| 1 | AD-contig_283 | 4e–68 | EBI\|7451 | 76 | mitochondrial carrier protein |
| 1 | AD-contig_271 | 1e–9 | CRA\|agCP12493 | 65 | metallothionein |
| 1 | AD-contig_38 | 5e–28 | CRA\|agCP6905 | 91 | mitochondrial ATP synthase |
| 1 | AD-contig_269 | 2e–78 | CRA\|agCP4271 | 90 | myosin regulatory light chain (non-muscle) |
| 1 | AD-contig_110 | 2e–30 | EBI\|5259 | 29 | $Na^+/K^+$-exchanging ATPase |
| 1 | AD-contig_130 | 4e–63 | CRA\|agCP1147 | 95 | GABA(A) receptor-associated protein |
| 1 | AD-contig_53 | 2e–7 | | | adipokinetic hormone |
| **Structural proteins** | | | | | |
| 1 | AD-contig_263 | 3e–75 | CRA\|agCP4904 | 68 | integral membrane protein 2A |
| 1 | AD-contig_183 | 1e–102 | CRA\|agCP5029 | 94 | similar to proteasome |
| 1 | AD-contig_176 | 6e–91 | CRA\|agCP8637 | 96 | similar to transmembrane 9 superfamily |
| 1 | AD-contig_14 | 0.004 | CRA\|agCP4724 | 60 | extensin |
| 1 | AD-contig_190 | 2e–62 | CRA\|agCP4913 | 32 | membrane integral protein |
| **Transcription/translation factors** | | | | | |
| 1 | AD-contig_196 | 5e–12 | CRA\|agCP13054 | 65 | elongation factor (super cysteine rich protein) |
| 1 | AD-contig_131 | 3e–58 | CRA\|agCP13402 | 90 | histone H2A |
| 1 | AD-contig_5 | 3e–22 | CRA\|agCP6533 | 79 | nuclear regulation |
| 1 | AD-contig_282 | 6e–8 | CRA\|agCP4557 | 46 | transcription factor |
| 6 | AD-contig_112 | 7e–48 | CRA\|agCP9704 | 78 | transcription factor, MBF2 |
| 1 | AD-contig_186 | 5e–93 | CRA\|agCP7051 | 81 | translation initiation factor |
| 1 | AD-contig_184 | 1e–111 | CRA\|agCP2764 | 94 | translation initiation factor 3 |
| 1 | AD-contig_177 | 1e–118 | CRA\|agCP7851 | 89 | phenylalanine-tRNA ligase beta chain |

sequenced) are associated with energy metabolism, including several mitochondrial enzymes (cytochromes and ATP synthases) and enzymes from the glucolysis, glucose-6-P, and Krebs cycle pathways (transaldolase, transketolase, aconitase and several dehydrogenases). Eight clusters (thirteen sequences) are associated with possible transcription/translation factors, including elongation factors, transcription-initiation factors, histone H2A and nuclear regulation factor.

Ten clusters are associated with signal/transport transduction pathways including calcyclin, calmodulin, metallothionein, mitochondrial ATP synthase, GABA(A) receptor-associated protein and ATPases involved in ion transport, such as $Na^+/K^+$-ATPase. Five clusters (six sequences) are associated with products possibly involved in protein folding such as heat shock proteins and chaperonins. Five clusters are associated with cytoskeletal proteins such as integral membrane protein 2 A and extensin.

### *Description of unknown function (U) category clusters*

We sequenced 131 clones, included in ninety-seven clusters (Table 3), and annotated these as unknown sequences. These sequences did not show significant similarity with known proteins and could either represent novel proteins, unique to *An. darlingi* salivary glands, or PCR artefacts (or sequencing errors), and hence they are not described in this work.

### *Comparison of protein sequence identities between* An. darlingi *and* An. gambiae *gene products*

It has been proposed that adult female salivary gland proteins of anopheline mosquitoes and American sand flies are under strong selection owing to the deleterious effect

**Table 3.** *An. darlingi* salivary glands cDNA clusters with unknown function

| No. of sequences | Assembled contig | *E* value (best match to NR protein database) | Best match to AGPROT database | Percentage identity |
|---|---|---|---|---|
| 1 | AD-contig_75 | 2e–13 | EBI\|7994 | 85 |
| 5 | AD-contig_197 | 0.092 | EBI\|6164 | 46 |
| 1 | AD-contig_118 | 0.001 | EBI\|5715 | 31 |
| 1 | AD-contig_165 | 0.088 | EBI\|5060 | 84 |
| 2 | AD-contig_240 | 3e–13 | EBI\|4655 | 23 |
| 1 | AD-contig_195 | 0.017 | EBI\|2778 | 80 |
| 1 | AD-contig_83 | 0.086 | EBI\|1072 | 29 |
| 3 | AD-contig_211 | 0.059 | CRA\|agCP9352 | 70 |
| 1 | AD-contig_215 | 1e–9 | CRA\|agCP8743 | 51 |
| 1 | AD-contig_216 | | CRA\|agCP8743 | 70 |
| 1 | AD-contig_7 | 8e–17 | CRA\|agCP8728 | 60 |
| 5 | AD-contig_170 | 0.050 | CRA\|agCP8362 | 33 |
| 1 | AD-contig_105 | 8e–7 | CRA\|agCP8258 | 40 |
| 2 | AD-contig_254 | 0.004 | CRA\|agCP8099 | 39 |
| 1 | AD-contig_80 | 1e–62 | CRA\|agCP7842 | 72 |
| 1 | AD-contig_27 | 1e–37 | CRA\|agCP7057 | 88 |
| 1 | AD-contig_117 | 0.018 | CRA\|agCP6430 | 34 |
| 3 | AD-contig_204 | 1e–51 | CRA\|agCP6351 | 79 |
| 1 | AD-contig_194 | | CRA\|agCP6071 | 52 |
| 1 | AD-contig_20 | 9e–71 | CRA\|agCP5466 | 72 |
| 1 | AD-contig_49 | 0.003 | CRA\|agCP4972 | 60 |
| 1 | AD-contig_99 | 0.094 | CRA\|agCP4788 | 100 |
| 1 | AD-contig_182 | 6e–12 | CRA\|agCP4733 | 50 |
| 2 | AD-contig_236 | 1e–57 | CRA\|agCP4692 | 60 |
| 1 | AD-contig_61 | 0.002 | CRA\|agCP4376 | 32 |
| 1 | AD-contig_188 | 6e–65 | CRA\|agCP4171 | 93 |
| 1 | AD-contig_91 | 1e–16 | CRA\|agCP3793 | 93 |
| 1 | AD-contig_264 | 9e–4 | CRA\|agCP2750 | 66 |
| 1 | AD-contig_145 | 3e–13 | CRA\|agCP1820 | 31 |
| 1 | AD-contig_175 | | CRA\|agCP1764 | 33 |
| 1 | AD-contig_86 | | CRA\|agCP15025 | 36 |
| 2 | AD-contig_250 | $3e^{-17}$ | CRA\|agCP13443 | 24 |
| 2 | AD-contig_17 | 0.001 | CRA\|agCP13284 | 52 |
| 1 | AD-contig_205 | 0.001 | CRA\|agCP13284 | 52 |
| 1 | AD-contig_23 | 2e–5 | CRA\|agCP12804 | 32 |
| 1 | AD-contig_18 | 7e–57 | CRA\|agCP12530 | 70 |
| 2 | AD-contig_249 | 1e–8 | CRA\|agCP11977 | 46 |
| 2 | AD-contig_245 | 8e–6 | CRA\|agCP11977 | 34 |
| 1 | AD-contig_162 | | CRA\|agCP11977 | 29 |
| 1 | AD-contig_155 | 5e–22 | CRA\|agCP11977 | 56 |
| 4 | AD-contig_210 | | CRA\|agCP11543 | 42 |
| 1 | AD-contig_232 | 0.013 | CRA\|agCP11536 | 100 |
| 1 | AD-contig_142 | 7e–67 | CRA\|agCP11064 | 94 |
| 1 | AD-contig_25 | 3e–90 | CRA\|agCP10880 | 90 |
| 1 | AD-contig_260 | 0.014 | CRA\|agCP10645 | 30 |
| 6 | AD-contig_41 | 0.024 | CRA\|agCP10484 | 23 |
| 1 | AD-contig_273 | 0.008 | CRA\|agCP10434 | 32 |
| 1 | AD-contig_187 | 0.005 | no match | |
| 3 | AD-contig_212 | 0.004 | no match | |
| 3 | AD-contig_218 | 0.059 | no match | |
| 1 | AD-contig_138 | 2e−4 | no match | |
| 1 | AD-contig_277 | 0.079 | no match | |
| 1 | AD-contig_43 | 0.002 | no match | |
| 1 | AD-contig_171 | 0.009 | no match | |
| 1 | AD-contig_172 | 0.061 | no match | |
| 1 | AD-contig_173 | 0.005 | no match | |
| 1 | AD-contig_164 | 0.073 | no match | |
| 1 | AD-contig_120 | 0.071 | no match | |
| 2 | AD-contig_252 | 0.076 | no match | |
| 2 | AD-contig_42 | 0.091 | no match | |
| 1 | AD-contig_102 | 0.081 | no match | |
| 2 | AD-contig_233 | | no match | |
| 1 | AD-contig_101 | | no match | |
| 1 | AD-contig_103 | | no match | |
| 1 | AD-contig_108 | | no match | |
| 1 | AD-contig_109 | | no match | |

**Table 3.** (*Continued*)

| No. of sequences | Assembled contig | *E* value (best match to NR protein database) | Best match to AGPROT database | Percentage identity |
|---|---|---|---|---|
| 1 | AD-contig_127 | | no match | |
| 1 | AD-contig_133 | | no match | |
| 1 | AD-contig_139 | | no match | |
| 1 | AD-contig_140 | | no match | |
| 1 | AD-contig_141 | | no match | |
| 1 | AD-contig_143 | | no match | |
| 1 | AD-contig_163 | | no match | |
| 1 | AD-contig_167 | | no match | |
| 1 | AD-contig_174 | | no match | |
| 1 | AD-contig_178 | | no match | |
| 1 | AD-contig_52 | | no match | |
| 1 | AD-contig_62 | | no match | |
| 1 | AD-contig_227 | | no match | |
| 1 | AD-contig_259 | | no match | |
| 1 | AD-contig_270 | | no match | |
| 1 | AD-contig_272 | | no match | |
| 1 | AD-contig_275 | | no match | |
| 1 | AD-contig_70 | | no match | |
| 1 | AD-contig_74 | | no match | |
| 1 | AD-contig_8 | | no match | |
| 1 | AD-contig_84 | | no match | |
| 1 | AD-contig_90 | | no match | |
| 1 | AD-contig_94 | | no match | |
| 1 | AD-contig_95 | | no match | |
| 1 | AD-contig_55 | | no match | |
| 1 | AD-contig_150 | | no match | |
| 1 | AD-contig_152 | | no match | |
| 1 | AD-contig_153 | | no match | |
| 1 | AD-contig_156 | | no match | |
| 1 | AD-contig_157 | | no match | |
| 1 | AD-contig_160 | | no match | |

that vertebrate host immunity has on feeding (Lanzaro *et al*., 1999; Valenzuela *et al*., 2003). However, salivary gland genes involved in blood feeding also may be rapidly evolving to adapt to a different repertoire of hosts (Valenzuela *et al*., 2003). Because the primary host of both *An. darlingi* and *An. gambiae* is human (Deane, 1986; Constantini *et al*., 1998), and these two anophelines belong to different subgenera, we compared their genes belonging to the secreted (S) and housekeeping (H) categories. For this comparison, we used the *An. gambiae* (subgenus *Celia*) protein data set recently submitted to NCBI and *An. darlingi* (subgenus *Nyssorhynchus*) sequences that originated from two or more cDNA sequences and that gave > 100 amino acid residues of match to the *An. gambiae* sequences when these were compared by blastp with the filter removed.

Both the average and the variance (% identity) of the two data sets were significantly different ($P < 0.0001$). The H genes had an average of $80.11 \pm 16.3\%$ identity, whereas the S genes had $47 \pm 14.1\%$ (average $\pm$ SD; averages tested by *t*-test with nonequal variances; variances tested by the *F*-test) (Table 4). We conclude that the salivary gland genes encoding secreted products are rapidly evolving in comparison with the housekeeping genes of these species. Valenzuela *et al*. (2003) found similar results when the salivary glands transcriptomes of *An. stephensi* and *An. gambiae* were compared. These two species belong to the same subgenus (*Celia*) and when compared showed 93% of identity for gene products of the housekeeping group whereas the salivary proteins are only 62% identical. These results support the idea that S genes may be good markers for assessing phylogeny among closely related species, as has been demonstrated with triatomine bugs using the salivary hemeproteins (Soares *et al*., 1998, 2000). Manguin *et al*. (1999) showed weak differentiation among *An. darlingi* populations ranging from Mexico to Argentina. However, previous studies based on behavioural (patterns of biting activity), morphological (body size and polytene chromosome patterns) and molecular (allozymes and ITS2 sequences) differences among geographically distinct populations have indicated the possibility that *An. darlingi* is a complex of closely related species (Lounibos & Conn, 2000). The analysis of salivary gland genes may be a useful tool for further analysis of the *An. darlingi* taxonomic status.

## Final remarks

This is the first extensive work of DNA sequence and analysis conducted with a neotropical anopheline mosquito.

**Table 4.** Identity of amino acids between housekeeping and putative secreted proteins of *An. darlingi* and *An.gambiae* salivary glands

| Assembled contig | Comments | Match | % identity |
|---|---|---|---|
| **Housekeeping** | | | |
| AD-contig_246 | heat shock protein cognate 4 | CRA\|agCP12309 | 95 |
| AD-contig_200 | ubiquitin | CRA\|agCP13749 | 59 |
| AD-contig_208 | ribosomal | CRA\|agCP1641 | 81 |
| AD-contig_235 | ribosomal protein | CRA\|agCP14911 | 87 |
| AD-contig_231 | ribosomal protein | CRA\|agCP11536 | 91 |
| AD-contig_237 | ribosomal protein | CRA\|agCP14909 | 84 |
| AD-contig_226 | ribosomal protein | CRA\|agCP3608 | 88 |
| AD-contig_251 | ribosomal protein | CRA\|agCP6680 | 68 |
| AD-contig_257 | ribosomal protein | CRA\|agCP10609 | 73 |
| AD-contig_241 | ribosomal protein | CRA\|agCP1535 | 78 |
| AD-contig_234 | ribosomal protein | CRA\|agCP10687 | 94 |
| AD-contig_126 | ribosomal protein | CRA\|agCP14068 | 82 |
| AD-contig_222 | ribosomal protein | CRA\|agCP12166 | 97 |
| AD-contig_219 | ribosomal protein | CRA\|agCP4788 | 47 |
| AD-contig_223 | ribosomal protein | CRA\|agCP7766 | 90 |
| AD-contig_213 | ribosomal protein | CRA\|agCP11873 | 92 |
| AD-contig_225 | ribosomal protein | CRA\|agCP6972 | 90 |
| AD-contig_248 | ribosomal protein | CRA\|agCP6003 | 38 |
| AD-contig_112 | transcription factor, MBF2 | CRA\|agCP9704 | 78 |
| Average ± SD | | | 80 ± 16.3 |
| **Salivary** | | | |
| AD-contig_214 | 30 kDa allergen | CRA\|agCP8743 | 58 |
| AD-contig_66 | antigen-5 | CRA\|agCP6145 | 75 |
| AD-contig_1 | D7 | CRA\|agCP11198 | 61 |
| AD-contig_230 | D7 | CRA\|agCP11220 | 36 |
| AD-contig_279 | D7 | CRA\|agCP11196 | 36 |
| AD-contig_77 | D7 | CRA\|agCP11196 | 28 |
| AD-contig_159 | D7 | CRA\|agCP10845 | 50 |
| AD-contig_229 | D7 | CRA\|agCP11220 | 37 |
| AD-contig_256 | D7, short | CRA\|agCP11220 | 28 |
| AD-contig_137 | anophelin | CRA\|agCP11208 | 43 |
| AD-contig_253 | apyrase | CRA\|agCP9757 | 49 |
| AD-contig_217 | apyrase (full clone) | CRA\|agCP10591 | 81 |
| AD-contig_224 | apyrase | CRA\|agCP10591 | 54 |
| AD-contig_88 | maltase | CRA\|agCP12790 | 58 |
| AD-contig_123 | salivary glucosidase | CRA\|agCP12065 | 49 |
| AD-contig_125 | salivary glucosidase | CRA\|agCP12065 | 37 |
| AD-contig_242 | mucin | CRA\|agCP1772 | 51 |
| AD-contig_244 | mucin | CRA\|agCP1772 | 40 |
| AD-contig_220 | mucin | CRA\|agCP7687 | 31 |
| AD-contig_266 | cecropin | CRA\|agCP7505 | 51 |
| AD-contig_181 | cecropin | CRA\|agCP7503 | 50 |
| AD-contig_203 | defensin | CRA\|agCP6915 | 78 |
| AD-contig_239 | putative infection responsive short peptide | EBI\|7267 | 32 |
| AD-contig_199 | gSG1b | CRA\|agCP13537 | 37 |
| AD-contig_28 | gSG2 | CRA\|agCP6138 | 32 |
| AD-contig_207 | gSG7 | CRA\|agCP11109 | 44 |
| AD-contig_247 | gSG7 | CRA\|agCP2222 | 54 |
| AD-contig_202 | gSG8 | CRA\|agCP7185 | 54 |
| AD-contig_255 | SG1 family | CRA\|agCP13467 | 32 |
| AD-contig_201 | SG3 family | CRA\|agCP6430 | 42 |
| Average ± SD | | | 47 ± 14.1 |

Furthermore, this study was conducted with *An. darlingi*, an important malaria vector and the most anthropophilic and endophilic species among the Amazonian anophelines (Tadei *et al.*, 1998). Despite its importance as a malaria vector, only twenty-nine nucleotide sequence and thirteen protein sequence entries were available in the NCBI database prior to this study. The description of the salivary transcriptome of *An. darlingi* (subgenus *Nyssorhynchus*) and its comparison to the information available from previously studies anophelines (*An. gambiae* and *An. stephensi*, subgenus *Celia*) represent an advance in the understanding of the mosquito salivary gland functioning and salivary constitution. The comparative analysis of the transcriptomes of several anopheline mosquitoes, belonging to different subgenera and having distinct primary hosts, may supply better tools for the determination of phylogeny of closely related species, population structure and speciation processes, and ultimately identify genes related to vectorial

capacity and host preference. All of this information is likely to be useful for the improvement of existing and development of novel transmission-reduction malaria control strategies.

## Experimental procedures

### *Mosquitoes and cDNA library construction*

Adult female *An. darlingi* were caught in Porto Velho, Rondonia, Brazil, and sent to the Institute of Biomedical Sciences, University of São Paulo, Brazil. PolyA$^+$ RNA was extracted from sixty dissected pairs of salivary glands using the Micro-FastTrack mRNA isolation kit (Invitrogen, Carlsbad, CA, USA), which was then used to make a PCR-based cDNA library using the SMART™ cDNA library construction kit (BD Biosciences-Clontech, Palo Alto, CA, USA) as described by Francischetti *et al*. (2002b).

### *SDS-PAGE*

Sodium dodecyl sulphate polyacrylamide electrophoresis (SDS-PAGE) of salivary gland proteins of *An. darlingi* adult females was performed using 1 mm thick, gradient (4–12%), NU-PAGE gels (Invitrogen). Gels were run with MES buffer according to the manufacturer's instructions. To estimate the molecular weight of the salivary gland proteins, SeeBlue™ markers from Invitrogen (myosin, BSA, glutamic dehydrogenase, alcohol dehydrogenase, carbonic anhydrase, myoglobin, lysozyme, aprotinin, insulin, chain B) were used.

Salivary glands were treated with NU-PAGE LDS sample buffer (Invitrogen) and fifteen pairs of homogenized salivary glands (approximately 15 μg of protein) were applied per lane. Proteins were stained with Coomassie blue G when required. For amino-terminal sequencing, fifteen pairs of salivary glands were electrophoresed and transferred to a polyvinylidene difluoride (PVDF) membrane using 10 mM CAPS, pH 11.0, and 10% methanol as the transfer buffer on a blotmodule for the XCell II Mini-Cell (Invitrogen). The membrane was stained with Coomassie blue G 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., Foster City, CA, USA). To identify the cDNAs encoding the amino acid sequences obtained by Edman degradation, a search program written in Visual Basic (Valenzuela *et al*., 2002c) was used, which checked the amino acid sequences against the three possible protein translations of each cDNA sequence obtained in the *An. darlingi* mass-sequencing project.

### *cDNA sequence clustering*

Randomly selected cDNA clones obtained from the salivary glands cDNA library were sequenced and analysed as in Francischetti *et al*. (2002b) and in Valenzuela *et al*. (2002c), except that clustering of the cDNA sequences was accomplished using the CAP program (Huang, 1992). Accession numbers for sequences originating from the *An. gambiae* proteome (Holt *et al*., 2002) are given as agCP ####, where #### corresponds to the referenced gene product. BLAST searches were done locally from programs obtained at the NCBI FTP site (ftp://ftp.ncbi.nih.gov/blast/executables/) (Altschul *et al*., 1997). The electronic versions of the complete tables (Microsoft Excel format) with hyperlinks to web-based databases and to BLAST results (full versions of the tables presented here) are available at: http://www.ncbi.nlm.nih.gov/projects/Mosquito/A_darlingi_sialome/.

## Acknowledgements

We thank Dr Nirmala Xavier for helpful comments and suggestions, and Lynn Olson for help in preparing the manuscript. This work was funded by grants from the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), UNDP/World Bank/WHO Special programme for research and training in tropical diseases, and the National Institutes of Health (AI29746).

## References

Acosta-Serrano, A., Almeida, I.C., Freitas-Junior, L.H., Yoshida, N. and Schenkman, S. (2001) The mucin-like glycoprotein superfamily of *Trypanosoma cruzi*: structure and biological roles. *Mol Biochem Parasitol* **114**: 143–150.

Akuffo, H., Hultmark, D., Engstom, A., Frohlich, D. and Kimbrell, D. (1998) *Drosophila* antibacterial protein, cecropin A, differentially affects non-bacterial organisms such as Leishmania in a manner different from other amphipathic peptides. *Int J Mol Med* **1**: 77–82.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* **25**: 3389–3402.

Andersen, J.F., Francischetti, I.M., Valenzuela, J.G., Schuck, P. and Ribeiro, J.M. (2002) Inhibition of hemostasis by a high-affinity biogenic amine-binding protein from the saliva of a blood-feeding insect. *J Biol Chem* **278**: 4611–4617.

Arca, B., Lombardo, F., de Lara Capurro, M., della Torre, A., Dimopoulos, G., James, A.A. and Coluzzi, M. (1999) Trapping cDNAs encoding secreted proteins from the salivary glands of the malaria vector *Anopheles gambiae*. *Proc Natl Acad Sci USA* **96**: 1516–1521.

Arca, B., Lombardo, F., Lanfrancotti, A., Spanos, L., Veneri, M., Louis, C. and Coluzzi, M. (2002) A cluster of four D7-related genes is expressed in the salivary glands of the African malaria vector *Anopheles gambiae*. *Insect Mol Biol* **11**: 47–55.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S. and Eppig, J.T. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–29.

Barr, S.C., Rose, D. and Jaynes, J.M. (1995) Activity of lytic peptides against intracellular *Trypanosoma cruzi* amastigotes *in vitro* and parasitemias in mice. *J Parasitol* **81**: 974–978.

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L. and Sonnhammer, E.L. (2000) The Pfam protein families database. *Nucl Acids Res* **28**: 263–266.

Calvo, E., deBianchi, A.G., James, A.A. and Marinotti, O. (2002) The major acid soluble proteins of adult female *Anopheles darlingi* salivary glands include a member of the D7-related family of proteins. *Insect Biochem Mol Biol* **32**: 1419–1427.

Campos, I.T., Amino, R., Sampaio, C.A., Auerswald, E.A., Friedrich, T., Lemaire, H.G., Schenkman, S. and Tanaka, A.S. (2002) Infestin, a thrombin inhibitor presents in *Triatoma infestans* midgut, a Chagas' disease vector: gene cloning, expression and characterization of the inhibitor. *Insect Biochem Mol Biol* **32**: 991–997.

Chalk, R., Townson, H. and Ham, P.J. (1995) *Brugia pahangi*: the effects of cecropins on microfilariae *in vitro* and in *Aedes aegypti*. *Exp Parasitol* **80**: 401–406.

Champagne, D.E., Smartt, C.T., Ribeiro, J.M. and James, A.A. (1995) The salivary gland-specific apyrase of the mosquito *Aedes aegypti* is a member of the 5′-nucleotidase family. *Proc Natl Acad Sci USA* **92**: 694–698.

Charlab, R., Valenzuela, J.G., Rowton, E.D. and Ribeiro, J.M. (1999) Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly *Lutzomyia longipalpis*. *Proc Natl Acad Sci USA* **96**: 15155–15160.

Constantini, C., Sagnon, N., Della Torre, A., Diallo, M., Brady, J., Gibson, G. and Coluzzi, M. (1998) Odor-mediated host preferences of West African mosquitoes, with particular reference to malaria vectors. *Am J Trop Med Hyg* **58**: 56–63.

Crampton, J.M., Comley, I., Eggleston, P., Hill, S., Hughes, M., Knapp, T., Lycett, G., Urwin, R. and Warren, A. (1992) Molecular biological approaches to the study of vectors in relation to malaria control. *Mem Inst Oswaldo Cruz* **87** (Suppl. 3): 43–49.

Cross, M.L., Cupp, E.W. and Enriquez, F.J. (1994) Differential modulation of murine cellular immune responses by salivary gland extract of *Aedes aegypti*. *Am J Trop Med Hyg* **51**: 690–696.

Deane, L.M. (1986) Malaria vectors in Brazil. *Mem Inst Oswaldo Cruz* **81**: 5–14.

Di Noia, J.M., D'Orso, I., Aslund, L., Sanchez, D.O. and Frasch, A.C. (1998) The *Trypanosoma cruzi* mucin family is transcribed from hundreds of genes having hypervariable regions. *J Biol Chem* **273**: 10843–10850.

Dimopoulos, G., Richman, A., Muller, H.M. and Kafatos, F.C. (1997) Molecular immune responses of the mosquito *Anopheles gambiae* to bacteria and malaria parasites. *Proc Natl Acad Sci USA* **94**: 11508–11513.

Francischetti, I.M., Andersen, J.F. and Ribeiro, J.M. (2002a) Biochemical and functional characterization of recombinant *Rhodnius prolixus* platelet aggregation inhibitor 1 as a novel lipocalin with high affinity for adenosine diphosphate and other adenine nucleotides. *Biochemistry* **41**: 3810–3818.

Francischetti, I.M., Valenzuela, J.G., Pham, V.M., Garfield, M.K. and Ribeiro, J.M. (2002b) Toward a catalog for the transcripts and proteins (sialome) from the salivary gland of the malaria vector *Anopheles gambiae*. *J Exp Biol* **205**: 2429–2451.

Francischetti, I.M., Valenzuela, J.G. and Ribeiro, J.M. (1999) Anophelin: kinetics and mechanism of thrombin inhibition. *Biochemistry* **38**: 16678–16685.

Futcher, B., Latter, G.I., Monardo, P., McLoughlin, C.S. and Garrels, J.I. (1999) A sampling of the yeast proteome. *Mol Cell Biol* **19**: 7357–7368.

Grossman, G.L. and James, A.A. (1993) The salivary glands of the vector mosquito, *Aedes aegypti*, express a novel member of the amylase gene family. *Insect Mol Biol* **1**: 223–232.

Gwadz, R.W., Kaslow, D., Lee, J.Y., Maloy, W.L., Zasloff, M. and Miller, L.H. (1989) Effects of magainins and cecropins on the sporogonic development of malaria parasites in mosquitoes. *Infect Immun* **57**: 2628–2633.

Hansen, J.E., Lund, O., Tolstrup, N., Gooley, A.A., Williams, K.L. and Brunak, S. (1998) NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj J* **15**: 115–130.

Hekmat-Scafe, D.S., Dorit, R.L. and Carlson, J.R. (2000) Molecular evolution of odorant-binding protein genes OS-E and OS-F in *Drosophila*. *Genetics* **155**: 117–127.

Henriksen, A., King, T.P., Mirza, O., Monsalve, R.I., Meno, K., Ipsen, H., Larsen, J.N., Gajhede, M. and Spangfort, M.D. (2001) Major venom allergen of yellow jackets, Ves v 5: structural characterization of a pathogenesis-related protein superfamily. *Proteins* **45**: 438–448.

Hoffman, D.R. (1993) Allergens in Hymenoptera venom XXIV: the amino acid sequences of imported fire ant venom allergens Sol i, II, Sol i, III, and Sol i, IV. *J Allergy Clin Immunol* **91**: 71–78.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, J.M. and Wides, R. (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**: 129–149.

Huang, X. (1992) A contig assembly program based on sensitive detection of fragment overlaps. *Genomics* **14**: 18–25.

Hurd, H. (1994) Interactions between parasites and insects vectors. *Mem Inst Oswaldo Cruz* **89** (Suppl. 2): 27–30.

Hvidsten, T.R., Komorowski, J., Sandvik, A.K. and Laegreid, A. (2001) Predicting gene function from gene expressions and ontologies. *Pac Symposium Biocomput*: 299–310.

Isawa, H., Yuda, M., Orito, Y. and Chinzei, Y. (2002) A mosquito salivary protein inhibits activation of the plasma contact system by binding to factor XII and high molecular weight kininogen. *J Biol Chem* **277**: 27651–27658.

James, A.A., Blackmer, K., Marinotti, O., Ghosn, C.R. and Racioppi, J.V. (1991) Isolation and characterization of the gene expressing the major salivary gland protein of the female mosquito, *Aedes aegypti*. *Mol Biochem Parasitol* **44**: 245–253.

James, A.A., Blackmer, K. and Racioppi, J.V. (1989) A salivary gland-specific, maltase-like gene of the vector mosquito, *Aedes aegypti*. *Gene* **75**: 73–83.

Kappe, S.H., Kaiser, K. and Matuschewski, K. (2003) The *Plasmodium* sporozoite journey: a rite of passage. *Trends Parasitol* **19**: 135–143.

Kylsten, P., Samakovlis, C. and Hultmark, D. (1990) The cecropin locus in *Drosophila*: a compact gene cluster involved in the response to infection. *EMBO J* **9**: 217–224.

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

Lanzaro, G.C., Lopes, A.H., Ribeiro, J.M., Shoemaker, C.B., Warburg, A., Soares, M. and Titus, R.G. (1999) Variation in the salivary peptide, maxadilan, from species in the *Lutzomyia longipalpis* complex. *Insect Mol Biol* **8**: 267–275.

Lewis, S., Ashburner, M. and Reese, M.G. (2000) Annotating eukaryote genomes. *Curr Opin Struct Biol* **10**: 349–354.

Li, S., Kwon, J. and Aksoy, S. (2001) Characterization of genes expressed in the salivary glands of the tsetse fly, *Glossina morsitans morsitans*. *Insect Mol Biol* **10**: 69–76.

Lin, M.Z. and Greenberg, M.E. (2000) Orchestral maneuvers in the axon: TRIO and the control of axon guidance. *Cell* **101**: 239–242.

Lombardo, F., Di Cristina, M., Spanos, L., Louis, C., Coluzzi, M. and Arca, B. (2000) Promoter sequences of the putative *Anopheles gambiae* apyrase confer salivary gland expression in *Drosophila melanogaster*. *J Biol Chem* **275**: 23861–23868.

Lorenzo, C., Salinas, G., Brugnini, A., Wernstedt, C., Hellman, U. and Gonzalez-Sapienza, G. (2003) *Echinococcus granulosus* antigen 5 is closely related to proteases of the trypsin family. *Biochem J* **369**: 191–198.

Lounibos, L.P. and Conn, J.E. (2000) Malaria vectors heterogeneity in South America. *Am Entomol* **46**: 238–249.

Lowenberger, C., Charlet, M., Vizioli, J., Kamal, S., Richman, A., Christensen, B.M. and Bulet, P. (1999) Antimicrobial activity

spectrum, cDNA cloning, and mRNA expression of a newly isolated member of the cecropin family from the mosquito vector *Aedes aegypti*. *J Biol Chem* **274**: 20092–20097.

Malafronte, R., dos, S., Calvo, E., James, A.A. and Marinotti, O. (2003) The major salivary gland antigens of *Culex quinquefasciatus* are D7-related proteins. *Insect Biochem Mol Biol* **33**: 63–71.

Manguin, S., Wilkerson, R.C., Conn, J.E., Rubio-Palis, Y., Danoff-Burg, J.A. and Roberts, D.R. (1999) Population structure of the primary malaria vector in South America, *Anopheles darlingi*, using isozyme, random-amplified polymorphic DNA, internal transcribed spacer 2, and morphologic markers. *Am J Trop Med Hyg* **60**: 364–376.

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

Megraw, T., Kaufman, T.C. and Kovalick, G.E. (1998) Sequence and expression of *Drosophila* Antigen 5-related 2, a new member of the CAP gene family. *Gene* **222**: 297–304.

Milne, T.J., Abbenante, G., Tyndall, J.D., Halliday, J. and Lewis, R.J. (2003) Isolation and characterization of a cone snail protease with homology to CRISP proteins of the pathogenesis-related protein superfamily. *J Biol Chem* **278**: 31105–31110.

Montfort, W.R., Weichsel, A. and Andersen, J.F. (2000) Nitrophorins and related antihemostatic lipocalins from *Rhodnius prolixus* and other blood-sucking arthropods. *Biochim Biophys Acta* **1482**: 110–118.

Moreira-Ferro, C.K., Daffre, S., James, A.A. and Marinotti, O. (1998) A lysozyme in the salivary glands of the malaria vector *Anopheles darlingi*. *Insect Mol Biol* **7**: 257–264.

Moreira-Ferro, C.K., Marinotti, O. and Bijovsky, A.T. (1999) Morphological and biochemical analyses of the salivary glands of the malaria vector, *Anopheles darlingi*. *Tissue Cell* **31**: 264–273.

Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**: 1–6.

PAHO. (1998) In *Health in the Americas*, pp. 123–145. Pan-American Health Organization (PAHO), Brazil.

Park, S.S., Shin, S.W., Park, D.S., Oh, H.W., Boo, K.S. and Park, H.Y. (1997) Protein purification and cDNA cloning of a cecropin-like peptide from the larvae of fall webworm (*Hyphantria cunea*). *Insect Biochem Mol Biol* **27**: 711–720.

Ribeiro, J.M.C. and Nussenzveig, R.H. (1993) The salivary catechol oxidase/peroxidase activities of the mosquito, *Anopheles albimanus*. *J Exp Biol* **179**: 273–287.

Ribeiro, J.M.C., Nussenzveig, R.H. and Tortorella, G. (1994) Salivary vasodilators of *Aedes triseriatus* and *Anopheles gambiae* (Diptera: Culicidae). *J Med Entomol* **31**: 747–753.

Ribeiro, J.M. and Valenzuela, J.G. (1999) Purification and cloning of the salivary peroxidase/catechol oxidase of the mosquito *Anopheles albimanus*. *J Exp Biol* **202**: 809–816.

Richman, A.M., Bulet, P., Hetru, C., Barillas-Mury, C., Hoffmann, J.A. and Kafatos, F.C. (1996) Inducible immune factors of the vector mosquito *Anopheles gambiae*: biochemical purification of a defensin antibacterial peptide and molecular cloning of preprodefensin cDNA. *Insect Mol Biol* **5**: 203–210.

Richman, A.M., Dimopoulos, G., Seeley, D. and Kafatos, F.C. (1997) *Plasmodium* activates the innate immune response of *Anopheles gambiae* mosquitoes. *EMBO J* **16**: 6114–6119.

Rodriguez, M.C., Zamudio, F., Torres, J.A., Gonzalez-Ceron, L., Possani, L.D. and Rodriguez, M.H. (1995) Effect of a cecropin-like synthetic peptide (Shiva-3) on the sporogonic development of *Plasmodium berghei*. *Exp Parasitol* **80**: 596–604.

Ropert, C., Ferreira, L.R., Campos, M.A., Procopio, D.O., Travassos, L.R., Ferguson, M.A., Reis, L.F., Teixeira, M.M., Almeida, I.C. and Gazzinelli, R.T. (2002) Macrophage signaling by glycosylphosphatidylinositol-anchored mucin-like glycoproteins derived from *Trypanosoma cruzi* trypomastigotes. *Microbes Infect* **4**: 1015–1025.

Rossignol, P.A. and Lueders, A.M. (1986) Bacteriolytic factor in the salivary glands of *Aedes aegypti*. *Comp Biochem Physiol* **83B**: 819–822.

Rubio-Palis, Y. and Zimmerman, R.H. (1997) Ecoregional classification of malaria vectors in the neotropics. *J Med Entomol* **34**: 499–510.

Sankoff, D. (2001) Gene and genome duplication. *Curr Opin Genet Dev* **11**: 681–684.

Schreiber, M.C., Karlo, J.C. and Kovalick, G.E. (1997) A novel cDNA from *Drosophila* encoding a protein with similarity to mammalian cysteine-rich secretory proteins, wasp venom antigen 5, and plant group 1 pathogenesis-related proteins. *Gene* **191**: 135–141.

Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P. and Bork, P. (2000) SMART: a web-based tool for the study of genetically mobile domains. *Nucl Acids Res* **28**: 231–234.

Simons, F.E. and Peng, Z. (2001) Mosquito allergy: recombinant mosquito salivary antigens for new diagnostic tests. *Int Arch Allergy Immunol* **124**: 403–405.

Soares, R.P., Gontijo, N.F., Romanha, A.J., Diotaiuti, L. and Pereira, M.H. (1998) Salivary heme proteins distinguish *Rhodnius prolixus* from *Rhodnius robustus* (Hemiptera: Reduviidae: Triatominae). *Acta Trop* **71**: 285–291.

Soares, R.P., Sant'Anna, M.R., Gontijo, N.F., Romanha, A.J., Diotaiuti, L. and Pereira, M.H. (2000) Identification of morphologically similar *Rhodnius* species (Hemiptera: Reduviidae: Triatominae) by electrophoresis of salivary heme proteins. *Am J Trop Med Hyg* **62**: 157–161.

Steiner, H., Hultmark, D., Engstrom, A., Bennich, H. and Boman, H.G. (1981) Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature* **292**: 246–248.

Stintzi, A., Heitz, T., Prasad, V., Wiedemann-Merdinoglu, S., Kauffmann, S., Geoffroy, P., Legrand, M. and Fritig, B. (1993) Plant 'pathogenesis-related' proteins and their role in defense against pathogens. *Biochimie* **75**: 687–706.

Szyperski, T., Fernandez, C., Mumenthaler, C. and Wuthrich, K. (1998) 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 USA* **95**: 2262–2266.

Tadei, W.P., Thatcher, B.D., Santos, J.M., Scarpassa, V.M., Rodrigues, I.B. and Rafael, M.S. (1998) Ecologic observations on *Anopheline* vectors of malaria in the Brazilian Amazon. *Am J Trop Med Hyg* **59**: 325–335.

Valenzuela, J.G., Charlab, R., Gonzalez, E.C., Miranda-Santos, I.K.F., Marinotti, O., Francischetti, I.M. and Ribeiro, J.M.C. (2002a) The D7 family of salivary proteins in blood sucking Diptera. *Insect Mol Biol* **11**: 149–155.

Valenzuela, J.G., Francischetti, I.M.B., Pham, V.M., Garfield, M.K., Mather, T.N. and Ribeiro, J.M.C. (2002b) Exploring the sialome of the tick, *Ixodes scapularis*. *J Exp Biol* **205**: 2843–2864.

Valenzuela, J.G., Francischetti, I.M., Pham, V.M., Garfield. M.K. and Ribeiro, J.M. (2003) Exploring the salivary gland transcriptome

and proteome of the *Anopheles stephensi* mosquito. *Insect Biochem Mol Biol* **33**: 717–732.

Valenzuela, J.G., Francischetti, I.M. and Ribeiro, J.M. (1999) Purification, cloning, and synthesis of a novel salivary anti-thrombin from the mosquito *Anopheles albimanus*. *Biochemistry* **38**: 11209–11215.

Valenzuela, J.G., Pham, V.M., Garfield, M.K., Francischetti, I.M. and Ribeiro, J.M.C. (2002c) Toward a description of the sialome of the adult female mosquito *Aedes aegypti*. *Insect Biochem Mol Biol* **32**: 1101–1122.

Vizioli, J., Bulet, P., Charlet, M., Lowenberger, C., Blass, C., Muller, H.M., Dimopoulos, G., Hoffmann, J., Kafatos, F.C. and Richman, A. (2000) Cloning and analysis of a cecropin gene from the malaria vector mosquito, *Anopheles gambiae*. *Insect Mol Biol* **9**: 75–84.

Vizioli, J., Bulet, P., Hoffmann, J.A., Kafatos, F.C., Muller, H.M. and Dimopoulos, G. (2001) Gambicin: a novel immune responsive antimicrobial peptide from the malaria vector *Anopheles gambiae*. *Proc Natl Acad Sci USA* **98**: 12630–12635.

Zheng, X.L. and Zheng, A.L. (2002) Genomic organization and regulation of three cecropin genes in *Anopheles gambiae*. *Insect Mol Biol* **6**: 517–525.