ELSEVIER

Insect Biochemistry and Molecular Biology 34 (2004) 543–563

Insect Biochemistry and Molecular Biology

www.elsevier.com/locate/ibmb

An insight into the salivary transcriptome and proteome of the adult female mosquito *Culex pipiens quinquefasciatus*

José M.C. Ribeiro [a,*], Rosane Charlab [b], Van My Pham [a], Mark Garfield [c], Jesus G. Valenzuela [a]

[a] *Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 12735 Twimbrook Parkway, Room 2E32, Rockville, MD 20852, USA*

[b] *Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA*

[c] *Biological Resources Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA*

Received 26 December 2003; accepted 20 February 2004

Abstract

To obtain an insight into the salivary transcriptome and proteome (sialome) of the adult female mosquito *Culex quinquefasciatus*, a cDNA library was randomly sequenced, and aminoterminal information for selected proteins and peptides was obtained. cDNA sequence clusters coding for secreted proteins were further analyzed. The transcriptome revealed messages coding for several proteins of known families previously reported in the salivary glands of other blood-feeding insects as well as immune-related products such as C-type lectin, gambicin, and members of the prophenol oxidase cascade. Additionally, several transcripts coding for low-complexity proteins were found, some clearly coding for mucins. Many novel transcripts were found, including a novel endonuclease previously described in crabs and shrimps but not in insects; a hyaluronidase, not described before in mosquito salivary glands but found in venom glands and in salivary glands of sand flies and black flies; several cysteine-rich peptides with possible anticlotting function, including one similar to a previously described nematode family of anti-proteases; and a completely novel family of cysteine- and tryptophane-rich proteins (CWRC family) for which 12 full-length sequences are described. Also described are 14 additional novel proteins and peptides whose function and/or family affiliation are unknown. In total, 54 transcripts coding for full-length proteins are described. That several of these are translated into proteins was confirmed by finding the corresponding aminoterminal sequences in the SDS-PAGE/Edman degradation experiments. Electronic versions of all tables and sequences can be found at http://www.ncbi.nlm.nih.gov/projects/Mosquito/C_quinquefasciatus_sialome.
Published by Elsevier Ltd.

Keywords: Mosquito; Salivary gland; Hematophagy; Transcriptome; Sialome

1. Introduction

Culex pipiens quinquefasciatus is a cosmopolitan mosquito species found in both tropical hemispheres. It breeds in great numbers in organically polluted water, being a major nuisance and producing allergic reactions. *C. quinquefasciatus* is also an efficient vector of Bancroftian filariasis and arboviral diseases (Horsfall, 1955). It is closely related to the subtropical species *C. pipiens pipiens*, which is a relatively efficient vector of West Nile virus (Dohm et al., 2002; Turell et al., 2001).

Adult female mosquitoes inject several salivary antihemostatic substances into the host skin before taking a blood meal (Ribeiro and Francischetti, 2003). Although *C. quinquefasciatus* has comparatively fewer antihemostatic activities than other anthropophilic mosquitoes (Ribeiro, 2000), it does contain (i) salivary apyrase activity (Ribeiro, 2000) that counteracts the platelet-aggregating effect of ADP released by damaged cells and activated platelets, (ii) as yet uncharacterized anticlotting factor(s) (Ribeiro, 2000), and (iii) abundant platelet-activating factor (PAF) hydrolyzing activity (Ribeiro and Francischetti, 2001). *C. quinquefasciatus*

* Corresponding author. Tel.: +1-301-496-9389; fax: +1-301-402-2201.

E-mail address: jribeiro@nih.gov (J.M.C. Ribeiro).

0965-1748/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.ibmb.2004.02.008

also has salivary proteins of the D7 family, an ubiquitous family of proteins found in blood-sucking mosquitoes and sand flies (Valenzuela et al., 2002a). D7 family proteins found in blood-sucking Nematocera belong to the superfamily of odorant-binding proteins (OBP) (Hekmat-Scafe et al., 2000). Except for hamadarin, one of several salivary D7 proteins found in the *Anopheles stephensi* mosquito, which has anticlotting and antikinin activity by inhibiting Factor XII (Isawa et al., 2002), the function of D7 proteins remain obscure. Mosquito saliva also functions in sugar feeding, where maltases and amylases have been described in other species (James et al., 1989). Perhaps associated with protection of the sugar meal from contaminating microorganisms, mosquito saliva may also have an immune function where lysozyme (Moreira-Ferro et al., 1998; Rossignol and Lueders, 1986) and other immune proteins have been found (Dimopoulos et al., 1998; Francischetti et al., 2002b; Valenzuela et al., 2002b). The salivary cocktail of *C. quinquefasciatus* remains largely unknown.

To gain insight into the complexity of the salivary transcriptome of *C. quinquefasciatus* and to identify molecules with possible function in the process of sugar and blood feeding, we have randomly sequenced a salivary gland cDNA library from adult female mosquitoes. After clustering the resulting database, we identified transcripts possibly associated with blood and sugar feeding and herein report 54 novel full-length sequences of putative salivary proteins and peptides. The possible roles of some of these proteins are discussed, although most have unknown function.

2. Materials and methods

2.1. Mosquitoes

Adult female *C. quinquefasciatus*, Vero Beach strain, were dissected at day 0 and day 1 post emergence to remove the salivary glands, which were then used to make a PCR-based cDNA library using the Micro-FastTrack mRNA isolation kit (Invitrogen, Carlsbad, CA) and the SMART™ cDNA library construction kit (BD Biosciences-Clontech, Palo Alto, CA) exactly as described previously (Francischetti et al., 2002b). Eighty pairs of salivary glands were used for the library.

2.2. SDS-PAGE

Sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) of 20 pairs of homogenized salivary glands of *C. quinquefasciatus* adult females was done using 1-mm thickness NU-PAGE 4–12% gels run with MES buffer, or 12% gels (which better discriminate at a lower molecular weight range) run with MOPS buffer (Invitrogen), according to the manufacturer's instructions. To estimate the molecular weight of the samples, SeeBlue™ markers from Invitrogen (myosin, BSA, glutamic dehydrogenase, alcohol dehydrogenase, carbonic anhydrase, myoglobin, lysozyme, aprotinin, and insulin, chain B) were used. Salivary gland homogenates were treated with NU-PAGE LDS sample buffer (Invitrogen) with (4–12% gel) or without (12% gel) reducing reagent. The combination of these two gel types, buffer mixtures, and reducing conditions increased the probability of individualizing protein bands. Twenty pairs of homogenized salivary glands per lane (approximately 20 μg protein) were applied when visualization of the protein bands stained with Coomassie blue was required. For aminoterminal sequencing of the salivary proteins, 20 homogenized pairs of glands were electrophoresed and transferred to a 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 in the absence of acetic acid (acetic acid can acetilate amino groups and render the protein blocked for the Edman degradation reaction). Stained bands were cut from the PVDF membrane and subjected to Edman degradation using a Procise sequencer (Perkin–Elmer Corp., Foster City, CA). To find the cDNA sequences corresponding to the amino acid sequence—obtained by Edman degradation of the proteins transferred to PVDF membranes from PAGE gels—we wrote a search program (in Visual Basic) that checked these amino acid sequences against the three possible protein translations of each cDNA sequence obtained in the mass sequencing project. This program takes in account multiple amino acids eventually found in a single Edman degradation cycle and thus can identify more than one protein per Edman experiment. For details, see Valenzuela et al. (2002b).

2.3. Bioinformatic analysis

Treatment of the cDNA sequence data was as in (Francischetti et al., 2002b) and in (Valenzuela et al., 2002b), except that clustering of the cDNA sequences was accomplished using the CAP program (Huang, 1992). Accession numbers for the National Center for Biology Information (NCBI) databases are given, as recommended by NCBI, as gi|XXXX, where XXXX is the accession number. BLAST searches were done locally from executables obtained at the NCBI FTP site (ftp://ftp.ncbi.nih.gov/blast/executables/) (Altschul et al., 1997). Prediction of signal peptides indicating secretion was made through the SignalP server (Nielsen et al., 1997). Prediction of *O*-glycosylation sites was made through the NetoGly server (Hansen et al.,

1998). Sequence alignments and phylogenetic tree analysis used the ClustalW package (Thompson et al., 1997). Phylogenetic trees were constructed by the neighbor-joining method (Saitou and Nei, 1987). Bootstrapping of phylogenetic trees, corrected for multiple substitutions and excluding positions with gaps, was done with the Clustal package for 1000 trials. Phylogenetic trees were formatted with TreeView (Page, 1996) using the ClustalW output. Hidden Markov Models (HMM) of protein alignments and the search of these models against protein databases was done with the HMMER2 software package found at http://hmmer.wustl.edu/ (Bateman et al., 2000). The electronic version of the complete tables (Microsoft Excel format) with hyperlinks to web-based databases and to BLAST results are available at http://www.ncbi.nlm.nih.gov/projects/Mosquito/C_quinquefasciatus_sialome.

3. Results

3.1. SDS-PAGE electrophoreses

Using different acrylamide concentrations, two gels were used to separate homogenates of adult female *C. quinquefasciatus* salivary glands (Fig. 1). These were transferred to a PVDF membrane, stained, and the bands cut according to the numbering in Fig. 1 to obtain Edman degradation information. The results of the observed amino acid sequences will be described below in the context of the transcriptome analysis.

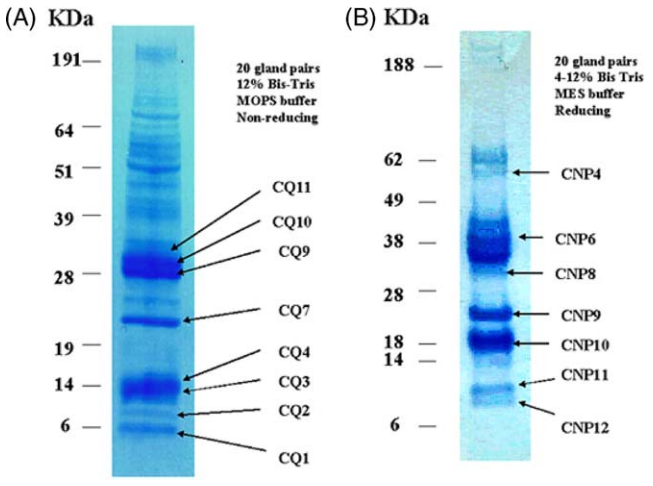

Fig. 1. Sodium dodecylsulfate–polyacrylamide gel electrophoresis (SDS-PAGE) separation of *C. quinquefasciatus* salivary gland proteins. Twenty pairs of homogenized salivary glands were previously incubated with SDS sample buffer (A) or in buffer with reducing reagent (B) before applying to a 12% gel; electrophoresis was run with Bis–Tris–MOPS (A) or to a 4–12% gel run in Bis–Tris–MES buffer (B). The gel was transferred to a PVDF membrane, the bands cut and submitted to Edman degradation. The CQ# and CNP# indicates the gel bands yielding results matching transcriptome data, as indicated in Tables 1 and 2. The numbers on the sides of the gel represent the retention position of molecular weight markers.

3.2. Transcriptome analysis

Following DNA sequencing of 503 clones of a cDNA library constructed from the salivary glands of adult female *C. quinquefasciatus* mosquitoes, these were grouped into 281 clusters and arbitrarily divided into three groups following analysis of the data: cDNA clusters probably related to secretory products (S); those probably related to housekeeping products (H); and those of unknown (U) function (Supplemental material). The S group contained 103 clusters and a total of 284 sequences, while the H group had 103 clusters and 132 sequences, and the U group had 75 clusters and 87 sequences. Accordingly, the majority of the individual cDNA clones were attributed to putative secretory products.

3.2.1. Housekeeping gene products expressed in the salivary glands of C. quinquefasciatus

Among the clusters of the H group are several representing energy metabolism enzymes such as cytochrome *c* oxidase subunits and ATP synthase, proteins involved in protein synthesis and protein modification, such as ribosomal proteins, initiation factors, and glycosyl transferases that might be involved in salivary protein glycosylation. Several transcription factors were also identified. Proteins associated with secretory processes, such as Golgi 4-transmembrane spanning protein, the p24 protein involved in membrane trafficking, calnexin, and the endoplasmic reticulum chaperone SIL1 were found. A cDNA coding for the vacuolar ATP synthase 16 kDa proteolipid subunit was also found. In addition, cDNA clusters coding for proteins with probable function in signal transduction were found, such as those coding for the enzyme sphingomyelin phosphodiesterase precursor, protein kinase and protein kinase inhibitors. A cluster coding for the circadian rhythm protein period was also identified. Included in the housekeeping group is a cluster of cDNA sequences coding for the previously described adenosine deaminase (ADA) from *C. quinquefasciatus*, for which no evidence of secretion in saliva was found despite a clear signal peptide indicative of secretion, whereas the salivary ADA of *Aedes aegypti* and *Lutzomyia longipalpis* gave prior evidence of being secreted enzymes (Charlab et al., 2000; Ribeiro et al., 2001). The Edman degradation sequence AKLISRLD was found in band CNP-4 of gel 2 (Fig. 1), a gel location consistent with the predicted molecular weight of the enzyme. Further, the observed aminoterminal sequence matches *Culex* ADA at position 18, where the signal peptide should have been cleaved. Other cDNA clusters

coding for proteins conserved in both *Drosophila* and *Anopheles* were included in the probable housekeeping category; although their function is unknown; their conserved sequence in Diptera suggests a conserved housekeeping function. The complete hyperlinked table for the H group, including additional products not mentioned in this paragraph, can be found electronically as indicated in the Materials and methods section.

3.2.2. *Gene products probably associated with secretory proteins expressed in the salivary glands of Culex quinquefasciatus. D7 and other odorant-binding protein families*

A novel protein named D7 was first reported in *Ae. aegypti* salivary glands (James et al., 1991). Similar proteins were later shown to occur in salivary glands of other mosquitoes and in sand flies (Arca et al., 1999; Valenzuela et al., 2002a). These proteins occur in two forms, described as long and short (Valenzuela et al., 2002a), and were classified as distant members of the odorant-binding superfamily of proteins (Hekmat-Scafe et al., 2000). Analysis of the sialotranscriptome of *C. quinquefasciatus* detected 11 cDNA clusters producing protein translations similar to members of the D7 family, and three additional clusters indicating translated proteins with weak similarity to one *An. gambiae* protein annotated as OBP (Table 1). Five of the clusters coding for D7-related proteins appear to be novel, while the remaining are either alleles, splice variants, or identical to other *Culex* D7 proteins previously described (Valenzuela et al., 2002a). Edman degradation of protein bands from SDS-PAGE transferred to PVDF matched two of these D7 proteins, as indicated in Table 1. One such match is to the previously reported long-form D7clu12 salivary protein (gi|16225986), while the other refers to a novel D7 protein sequence. Presently we report three full-length D7 protein sequences (Table 2), one of the long family, and two of the short family, all containing signal peptides indicative of secretion.

Clustal analysis of the long D7 protein CQ_LD7_3 (Table 2) with other anopheline and culicine long D7 proteins indicate a unique insert of 20 amino acid (aa) residues after the 7th conserved cysteines (Fig. 2A). This insert is flanked by Gly and Pro residues, which are often associated with beginning and end of protein loops. CQ_LD7_3 is 73% and 32% identical to the previously reported *C. quinquefasciatus* proteins D7clu12 and D7clu1, respectively (Valenzuela et al., 2002a). The bootstrapped phylogram indicates the presence of at least three robust clades (Fig. 2B). Notice that clade II contains both *Aedes* and *Anopheles* sequences, which are further separated into sub-clades, of which the anopheline group is unique for containing 8 instead of the usual 10 conserve cysteines of the long D7 family (Valenzuela et al., 2002a).

The two novel short D7 proteins are similar to the previously reported *C. quinquefasciatus* D7_clu32, and to the short D7 3 protein of *Ae. aegypti*. These four proteins are unique in having a shorter aa stretch between the 5th and 6th conserved Cys residue (Fig. 3A). The bootstrapped phylogram (Fig. 3B) indicates that *Culex* and *Aedes* short D7 proteins constitute a robust clade distinct from the anophelines, which do not group robustly, as indicated by their relatively small bootstrap values.

While one of the *An. stephensi* D7 proteins (named hamadarin) acts as an anticoagulant (Isawa et al., 2002), it is not known whether other D7 proteins function in a similar way. In parallel to the expansion of the D7 proteins in salivary glands of blood-feeding Diptera, the Hemiptera *Rhodnius prolixus* achieved a large expansion of the lipocalin family. Lipocalins, like OBPs, are specialized to bind small molecules. In *Rhodnius*, the salivary lipocalins perform various functions, from transporting nitric oxide to binding nucleotides and amines, and inhibiting blood clotting, a function unrelated to the binding of small molecules (Andersen et al., 2003; Francischetti et al., 2002a; Ribeiro et al., 1995). It is possible that, in analogy to the bug's lipocalins, the various D7 proteins in Diptera have evolved to acquire different functions.

Three clusters of the databases matched an *An. gambiae* protein annotated as odorant-binding G.21F.a (Table 1). Two full length sequences were obtained, coding for proteins containing a distinct signal peptide and predicted mature molecular weights of 13.4 and 14.6 kDa (Table 2). The alignment with *An. gambiae* G.21F.a indicates that the two *Culex* proteins are similar (26% and 30% identical) to the carboxyterminal region of the anopheline molecule, where six conserved cysteines are found, together with other conserved residues (Fig. 4). The function of these OBP-like proteins in *Culex* salivary glands is unknown.

3.2.3. *Proteins of the antigen-5 family*

Three cDNA clusters from the salivary gland library of *C. quinquefasciatus* code for proteins having similarity to proteins annotated as members of the antigen-5 family (Table 1), which are proteins found in the venom of vespids (King and Spangfort, 2000) and in the salivary glands of many blood sucking insects (Francischetti et al., 2002b; Li et al., 2001; Valenzuela et al., 2002b). These proteins belong to the larger family of cysteine-rich extracellular proteins (CRISP) ubiquitously found in animals and plants (Schreiber et al., 1997), with largely unknown function, except for one *Conus* protein that was recently shown to have proteolytic activity (Milne et al., 2003). The three cDNA clusters contain 15 ESTs, indicating that these messages

Table 1
Clusters of sequences from a cDNA library from adult female *C. quinquefasciatus* salivary glands most probably associated with secretory products

Assembled contig[a]	Sequences per contig[b]	Edman product[c]	Best match to NR protein database[d]	*E* value	Best match to CDD database[e]	*E* value	Comments
Odorant binding protein family							
Contig_142	1		Long form D7clu1	0.0	PhBP	2E−004	Long form D7clu1
Contig_201	3	CQ9 WKPFSPEETLFTYTRCMEDN CNP9 WKPFSPEETLFTY	Long form D7clu12	1E−080	PhBP	2E−004	Long form D7clu12
Contig_246	2		Long form D7clu12	1E−100			Long form D7clu12—truncated
Contig_1	25	CQ10 DEWSPMDPEEVAFEEAKCME CQ11 DEWSPMDPEEVAF CNP6 DEWSPMDPEEVAFEEAKCM	Long form D7clu12	3E−033			New long D7
Contig_2	1		D7 protein long form	3E−014	PhBP	0.011	New long D7
Contig_28	1		Short form D7clu32	5E−019	PhBP	1E−005	New short D7
Contig_78	1		Short form D7clu32	5E−045	PhBP	2E−006	New short D7
Contig_128	1		Short form D7clu32	1E−011	PhBP	0.009	New short D7
Contig_224	2		Short form D7clu32 [	1E−077	PhBP	6E−006	Short form D7clu32 salivary protein
Contig_193	3		Long form D7clu1	5E−063	PhBP	0.002	Very similar to long form D7clu1
Contig_181	1		Long form D7clu1	5E−063	PhBP	0.002	Very similar to long form D7clu1
Contig_107	1		Odorant-binding G.21F.a	2E−004			Similar to *An. gambiae* ptn
Contig_126	1		Odorant-binding G.21F.a	0.011			Similar to *An. gambiae* ptn
Contig_192	1		Odorant-binding G.21F.a	9E−005			Similar to *An. gambiae* ptn
Antigen-5 family							
Contig_36	6		Putative secreted protein	1E−042	SCP	8E−011	Antigen 5
Contig_109	1	CNP8 ADYCSDEFQKI	Putative secreted protein	1E−019			Antigen 5
Contig_226	9	CNP8 ADYCSDEFQKI	Putative secreted protein	1E−078	SCP	3E−017	Antigen 5
Enzymes involved in sugar digestion							
Contig_202	12		Probable maltase precursor	1E−137	Alpha-amylase	4E−015	Maltase
Contig_207	3		Probable maltase precursor	4E−060	Aamy	1E−039	Alpha glucosidase
Contig_33	1		Probable maltase precursor	4E−066	Aamy	4E−015	Alpha glucosidase
Contig_40	1		Probable maltase precursor	4E−030	Aamy	1E−014	Alpha glucosidase
Contig_6	1		Probable maltase precursor	4E−050	Aamy	2E−024	Maltase
Contig_267	1		Probable maltase precursor	2E−011	Aamy	8E−010	Maltase
Contig_211	3		Alpha-amylase I precursor	4E−039	Aamy	1E−011	Amylase
Contig_67	1		Alpha-amylase I precursor	3E−052	Alpha-amylase	7E−014	Amylase
Contig_145	1		Endochitinase [Encephalitozoon	3E−018	Glyco_hydro19	2E−015	Chitinase
Contig_97	1		gp8 [mycobacteriophage	9E−013	Glyco_hydro19	5E−012	Endochitinase
Other enzymes							
Contig_21	1		Chrysoptin precursor [Chryso	2E−020	Metallophos	0.009	Apyrase/5′-nucleotidase

(continued on next page)

Table 1 (*continued*)

Assembled contig[a]	Sequences per contig[b]	Edman product[c]	Best match to NR protein database[d]	*E* value	Best match to CDD database[e]	*E* value	Comments
Contig_17	1		agCP8269 [*Anopheles gambiae*	7E−019	IU_nuc_hydro	3E−011	Nucleoside hydrolase
Contig_59	6		agCP2051 [*Anopheles gambiae*	2E−006			Endonuclease?
Contig_216	3		agCP7510 [*Anopheles gambiae*	2E−015	Lipase	3E−007	Phospholipase A1 precursor
Contig_50	1		agCP7510 [*Anopheles gambiae* str.	4E−011	Lipase	1E−005	Lipase?
Contig_77	1		agCP3812 [*Anopheles gambiae*	3E−030	COesterase	7E−012	Esterase
Contig_95	1		Hyaluronoglucosaminidase	1E−025	Glycohydro56	2E−019	Hyaluronidase
Protease inhibitors							
Contig_74	1		Putative serpin [*Aedes aegy*	6E−012			Putative serpin
Contig_112	1		Putative serpin [*Aedes aegy*	8E−009			Putative serpin
Contig_200	3		agCP14448 [*Anopheles gambiae* E47	5E−027	WAP	8E−006	C-rich protease inhibitor?
Contig_121	1		Putative trypsin-like inhib	2E−007	TIL	3E−012	Putative trypsin-like inhibitor
Contig_229	2		Hypothetical protein Y69H2.3a	4E−004	TIL	1E−006	C-rich protease inhibitor?
Contig_134	1		Putative trypsin-like inhib	3E−008	TIL	3E−012	Trypsin inhibitor
Immunity-related proteins							
Contig_113	1		Putative secreted protein [	4E−029			Gambicin
Contig_151	1		Putative secreted protein [	5E−029			Gambicin
Contig_156	1		Putative secreted protein [	4E−052			Gram negative binding protein
Contig_84	1		CG9134-PA [*Drosophila melan*]	5E−008	Lectin_c	3E−008	Mannose binding lectin
Contig_262	1		agCP14272 [*Anopheles gambiae*	2E−020	Tryp_SPc	3E−017	Similar to Ag serine protease 14D2—PPO activator?
Mucins, other low complexity proteins, and proteins similar to uncharacterized Aedes and Anopheles proteins							
Contig_57	1		agCP13749 [*Anopheles gambiae* E61	3E−011			Similar to *A. gambiae* salivary protein
Contig_14	7		Putative 8.3 kDa secreted p	7E−006	CAP	0.084	Similar to 8.3 kDa secreted protein—mucin
Contig_47	5		HMW glutenin subunit [Triticum	9E−021	Glutenin_hmw	4E−020	Glutenin—low complexity
Contig_48	1		*Drosophila melanogaster* CG8797	1E−009	MSSP	2E−006	G-rich glutenin
Contig_70	5		Putative selenium-binding	0.001	CBM_14	0.066	Low complexity
Contig_238	9		Basic proline-rich protein [Sus4	0.004			Low complexity

Contig_277	1		Hypothetical protein [*Plasmodium*	0.001	ChtBD2	0.015	Low complexity
Contig_149	4		Conserved hypothetical protein	3E−005	Atrophin-1	0.002	Mucin
Contig_271	7		Hypothetical ORF; Ydl037cp	4E−020	Tryp_mucin	8E−013	Mucin
Contig_72	1		Proteoglycan 4 (megakaryocyte	0.004			Mucin?
Contig_217	3		Cell surface protein precursort	3E−006	TT_ORF1	0.003	Mucin?
Contig_239	2	CQ1 DQRCTYLRCRTEFRKTGAY CNP12 DQRCTYLRCRTE	glycoprotein gp2 [Equine	0.078	CBM_14	3E−004	Mucin?
Contig_245	2		Hypothetical protein; s	6E−005			Mucin?
Contig_249	8		agCP6539 [*Anopheles gambiae*	0.094			Mucin?
Contig_278	1		Hypothetical protein [*Neurospora*	1E−004			Mucin?
Contig_281	1		Conserved hypothetical protein	0.089			Mucin?
Contig_256	1		ebiP7571 [*Anopheles gambiae*	3E−028			Ser rich—mucin?
Contig_127	4		agCP6539 [*Anopheles gambiae*	8E−004			T rich mucin?
Contig_19	1		agCP6539 [*Anopheles gambiae*	0.025			T rich—mucin?
Contig_204	3		Homeotic protein Hoxd-3	7E−004			Proline/serine rich
Contig_115	17	CNP9 GKLPGMRGEA	ebiP3712 [*Anopheles gambiae*	0.043			Similar to procollagen
Contig_23	1		Putative 7.8 kDa secreted p	2E−005			Similar to *Aedes* putative 7.8 kDa peptide
Contig_182	4		Putative 30.5 kDa secreted	5E−038			Similar to *Aedes* putative 30.5 kDa salivary protein
Contig_235	2		Putative 56.5 kDa secreted	1E−035			Similar to *Aedes* salivary 56 kDa protein
Contig_179	1		Putative 56.5 kDa secreted	2E−013			Similar to *Aedes* salivary 56 kDa protein
Contig_169	1		Putative 56.5 kDa secreted	3E−043			Similar to *Aedes* salivary 56 kDa protein
Contig_26	1		agCP5466 [*Anopheles gambiae*	4E−044	Keratin_B2	0.025	Similar to *An. gambiae* protein
Unknown families							
Contig_55	1	CQ4 DVPTGCVTL					Unknown
Contig_222	2	CQ4 DVPTGCATIK					Unknown
Contig_43	1		agCP13749 [*Anopheles gambiae*	0.007			Unknown
Contig_81	5						Unknown
Contig_92	5		RIKEN cDNA	0.070			Unknown
Contig_138	4						Unknown
Contig_171	4						Unknown
Contig_185	1						Unknown

(continued on next page)

Table 1 (*continued*)

Assembled contig[a]	Sequences per contig[b]	Edman product[c]	Best match to NR protein database[d]	*E* value	Best match to CDD database[e]	*E* value	Comments
Contig_198	1	CQ4 TVPTGCVHIKN					Unknown
Contig_210	3	CQ3 DVPTGCVTLK CQ4 DVPTGCVTLK					Unknown
Contig_214	1		CG13004-PA [*Drosophila melanogaster*	4E−008			Unknown
Contig_215	10	CQ4 YVPLGCVKIKN					Unknown
Contig_218	3						Unknown
Contig_220	2	CQ4 DVPTGCVTLWH					Unknown
Contig_228	2		Cleavage and polyadenylation specif	0.080			Unknown
Contig_230	2						Unknown
Contig_243	2		gene_id:F1D9.26~unknown protein	2E−005	LGT	0.017	Unknown
Contig_165	1	CNP11 EKYCKSMKCTKVD	Putative membrane protein family	0.090			Unknown—membrane protein?
Contig_25	7						Unknown abundant cluster—signal peptide
Contig_3	7	CQ4 EVPTGCVTLK	Hypothetical protein [*Plasmodium*	0.024			Weak similarity to *Plasmodium* protein
Contig_116	4	CNP10 DVPTGCVTIKN and CQ4 DVPTGCVTIKN					Unknown
Contig_68	1						Unknown
Contig_69	1						Unknown
Contig_104	4		gene_id:F1D9.26~unknown protein	6E−012	7tm_5	1E−004	Unknown
Contig_108	1						Unknown
Contig_111	1						Unknown
Contig_123	1						Unknown
Contig_129	1						Unknown
Contig_140	1		Pfs77 protein [*Plasmodium*	0.098			Unknown
Contig_141	1						Unknown
Contig_161	1						Unknown
Contig_247	2						Unknown
Contig_41	1		Similar to RING-H2 finger protein	0.005	60KD_IMP	0.067	Unknown
Contig_58	1				COX3	0.022	Unknown
Contig_209	3		Similar to RING-H2 finger protein	2E−008	YMF19	4E−004	Unknown
Contig_254	1						Unknown
Contig_258	1		Putative transcription fact	0.010			Unknown
Contig_279	1						Unknown

[a] Contigs represent sequences derived from the CAP assembler (see Materials and methods).

[b] Indicates the number of sequences in each contig.

[c] Indicates Edman degradation amino acid (aa) sequence information derived from the experiment shown in Fig. 1, which matches translation products of the contig. Non-underlined amino-acids represents that the aa was not found in the Edman degradation product.

[d] Best protein sequence match when the contig was compared to the non-redundant protein database of NCBI using blastX with the filter option off (-FF option) (see Materials and methods).

[e] Best motif match when the contig was compared to the Conserved Domains Database using rpsblast (see Materials and methods).

Table 2
Characterization of 54 transcripts from the salivary glands of adult female *C. quinquefasciatus* mosquitoes coding for putative proteins and peptides. All translation products have a signal peptide indicative of secretion

Sequence name	Edman product[a]	Best match to NR database[b]	*E* value	Best match to CDD database[c]	*E* value	Comments
D7 and odorant binding protein family						
CQ_LD7_3	CQ9 WKPLSPEETLFTYTRCMEDI CNP8 WKPLSPEETLFTYTRCM	Long form D7clu12 salivary protein	1E−143	PhBP	2E−004	D7 long
CQ_SD7_3		Short form D7clu32 salivary protein [	7E−061	PhBP	6E−006	D7 short
CQ_SD7_4		Short form D7clu32 salivary protein [	9E−021	PhBP	2E−005	D7 short
CQ_SOBP1		Odorant-binding protein G.21F.a	8E−005	PhBP	0.045	OBP similar to Ag OBP—truncated?
CQ_SOBP2		Odorant-binding protein G.21F.a	6E−005	PhBP	0.057	OBP similar to Ag OBP
Antigen 5 family						
CQ_SAG5_1		Putative secreted protein [	2E−085	SCP	3E−021	Antigen 5 related
CQ_SAG5_2		Putative secreted protein [	8E−071	SCP	8E−019	Antigen 5 related
Enzymes						
CQ_SAPY		agCP5195 [*Anopheles gambiae*	2E−099	5_nucleotidase	3E−070	Apyrase
	CQ_PURNUC		agCP8269 [*Anopheles gambiae*	3E−089		IU_nuc_hydro
2E−033	Purine nucleosidase					
CQ_SENDONUC		agCP2051 [*Anopheles gambiae*	2E−021	NUC	8E−007	Endonuclease
CQ_SLIP		agCP7510 [*Anopheles gambiae*	6E−049	Lipase	2E−035	Lipase
CQ_SHYAL		agCP14464 [*Anopheles gambiae*	9E−075			Glyco_hydro_56
1E−075	Hyaluronidase					
Probable protease inhibitors						
CQ_SFIB		CG30197-PA [*Drosophila melanogaster*	4E−029			Cys rich
CQ_SPI_1		Hypothetical protein Y69H2.3c [*Caeno*	4E−004			Cys rich
CQ_SPI_2		Putative trypsin-like inhib	4E−007	TIL	1E−004	Cys rich
Immunity related proteins						
	CQ_SGAMBIC		Gambicin [*Culex pipiens pipiens*]	7E−044		
Gambicin						
CQ_SCLEC		CG12111-PA [*Drosophila melanor*	1E−012	CLECT	5E−013	C-type lectin
Mucins and other low complexity proteins						
CQ_SMUC_6		Putative 13.4 kDa salivary protein	1E−005			Mucin/proteoglycan 8 *O*-glyc sites
CQ_SMUC_7		Putative 8.3 kDa secreted p	2E−006			Mucin 5 *O*-Glyc sites
CQ_SMUC_1		Hypothetical protein [*Anopheles gamb*	1E−019			Mucin 17 *O*-glyc sites S/T rich
CQ_SMUC_2		Hypothetical ORF; Ydl037cp [*Saccharo*	1E−023	DYNc	0.088	Mucin Ser/Thr rich 78 *O*-glyc sites TTTDS repeats
CQ_SMUC_4				TOPEUc	0.087	Mucin Ser/Thr rich 10 *O*-glyc sites

(continued on next page)

Table 2 (*continued*)

Sequence name	Edman product[a]	Best match to NR database[b]	*E* value	Best match to CDD database[c]	*E* value	Comments
CQ_SMUC_5		Surface protein PspC [*Strept*	8E−004			Pro rich—muicn 9 *O*-Glyc sites
CQ_S56.6PTN		Putative 56.5 kDa secreted	1E−134			Similar to *Aedes* putative 56.5 kDa protein—5 *O*-Glyc sites
CQ_S30K_2		30 kDa salivary gland allergen	1E−016			Similar to *Aedes* 30 kDa salivary gland allergen 11-*O*-glyc sites
CQ_S30K_1	CNP9 SGKLPGMRGEA	Putative 30 kDa allergen-li	0.006			Similar to 30 kDa salivary gland allergen—GDE rich
CQ_QQQ_1		Putative selenium-binding protein	3E−004			Novel QQQ repeats
CQ_GQP_1		High molecular weight glute	2E−033			GQ-rich glutenin like
Unknown families						
CQ_MYS_15		Granulin-A precursor [Danio	0.083	ChtBD2	0.031	Cys rich
CQ_DVP_1	CNP10 DVPTGCATIKS					CWRC family
CQ_DVP_2	CQ3 DVPTGCVTLK					CWRC family
CQ_DVP_3	CQ4 DVPTGCVTL					CWRC family
CQ_DVP_4	CQ13 DVPTGCVTIKN CNP10DVPTGCVTIKN					CWRC family
CQ_MYS_10						CWRC family
CQ_MYS_11						CWRC family
CQ_MYS_12						CWRC family
CQ_MYS_13						CWRC family
CQ_MYS_14						CWRC family
CQ_MYS_18						CWRC family
CQ_MYS_6						CWRC family
CQ_MYS_9		Similar to *Homo sapiens* (Human)	0.092			CWRC family
CQ_MYS_16		agCP15395 [*Anopheles gambiae*	0.033			Novel
CQ_MYS_17		Hypothetical protein [*Escherichia coli*]	7E−004			Novel
CQ_MYS_20						Novel
CQ_MYS_4		agCP12530 [*Anopheles gambiae*	4E−040			Novel
CQ_MYS_5	CQ1 DQRCTYLRCRTEFRKTGAY			ChtBD2	0.071	Novel
CQ_MYS_7						Novel
CQ_MYS_8						Novel
CQ_MYS_1						Peptide
CQ_MYS_19						Peptide
CQ_MYS_2		Erythroid-specific transcription facto	0.024			Peptide
CQ_MYS_21						Peptide
CQ_MYS_22						Peptide
CQ_MYS_3		Putative transcription fact	0.008			Peptide

[a] Indicates Edman degradation amino acid (aa) sequence information derived from the experiment shown in Fig. 1, which matches translation products of the indicated sequence. Non-underlined amino acids represents that the aa was not found in the Edman degradation product.

[b] Best protein sequence match when the protein sequence was compared to the non-redundant protein database of NCBI using blastp with the filter option off (-FF option) (see Materials and methods).

[c] Best motif match when the protein sequence was compared to the Conserved Domains Database using rpsblast (see Materials and methods).

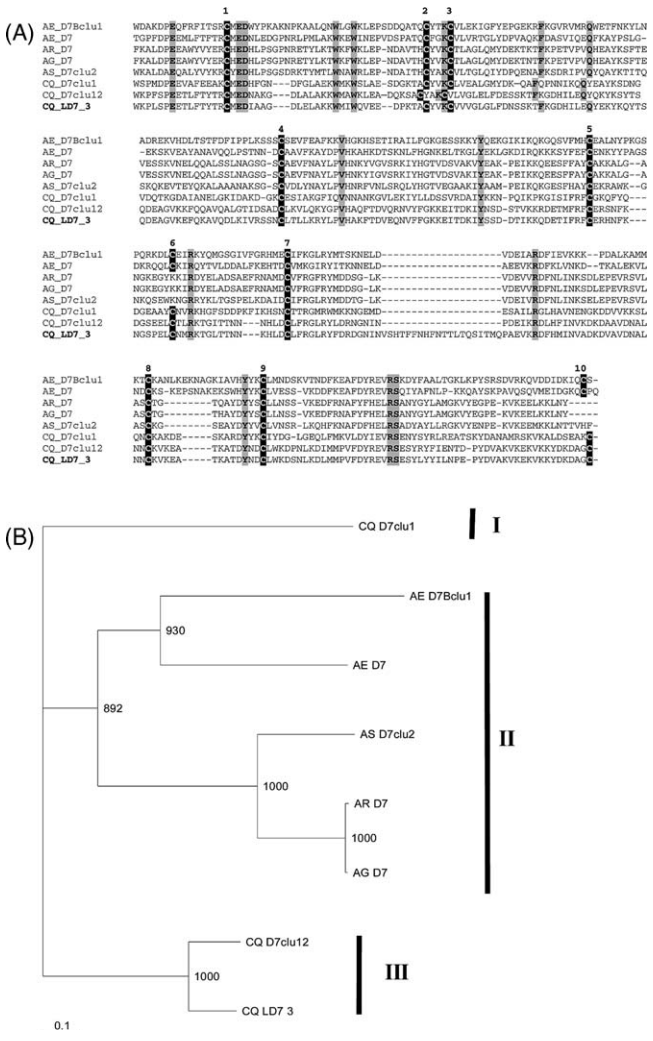

Fig. 2. Comparison of CQ_LD7_3, a novel long D7 protein from *Culex quinquefasciatus*, with other long D7 proteins from mosquitoes. (A) Clustal alignment of CQ_LD7_3 with CQ_D7clu12 (gi|16225986), CQ_D7clu1 (gi|16225983), AR_D7 (gi|16225958), AE_D7Bclu1 (gi|16225992), AS_D7clu2 (gi|16225974), and AE_D7 (gi|159559). CQ stands for *C. quinquefasciatus*, AR for *Anopheles arabiensis*, AS for *An. stephensi* and AE for *Aedes aegypti*. Cysteines are shown in reverse background and are numbered. Conserved residues are shown in gray background. The aminoterminal region of the proteins containing the signal peptide is not shown. The symbol (∗) under the alignments indicates amino acid identity between the sequences, whereas the symbol (:) indicates conserved amino acid substitution, and the (.) indicates partially conserved residue. (B) Unrooted phylogram indicating the three clades of sequences. The numbers at the nodes are the bootstrap values for 1000 trials. The bar under 0.1 indicates 10% amino acid divergence distance.

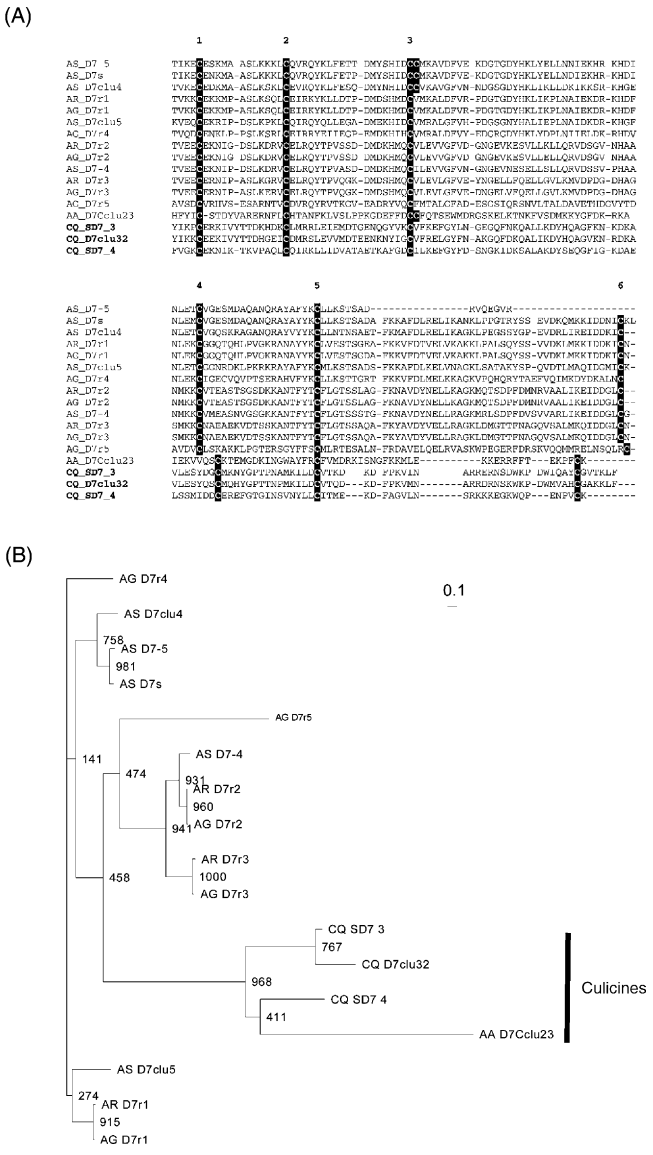

Fig. 3. Comparison of CQ_SD7_3 and CQ_SD7_4, two novel short D7 proteins from *Culex quinquefasciatus*, with other short D7 proteins from mosquitoes. (A) Clustal alignment with AS_D7-5 (gi|29501378), AS_D7-4 (gi|29501376), AA_D7Cclu23 (gi|16225995), CQ_D7clu32 (gi|16225989), AS_D7clu5 (gi|16225980), AS_D7clu4 (gi|16225977),AS_D7s (gi|16225971), AR_D7r3 (gi|16225968), AR_D7r2 (gi|16225965), AR_D7r1 (gi|16225961), AG_D7r3 (gi|4538891), AG_D7r2 (gi|4538889), AG_D7r1 (gi|4538887), AG_D7r4 (gi|17016228) and AG_D7r5 (gi|18378603). AG refers to *Anopheles gambiae*. The aminoterminal region of the proteins, containing the signal peptide is not shown. (B) Phylogram of the resulting alignment indicating the robust clade of the culicine short D7 proteins. See legend of Fig. 1A for other details.

are abundantly transcribed or stable. In support of the abundant translation of one or more of these messages was the finding of the aminoterminal sequence ADYXSDEFQKI (where X represents lack of signal or lack of AA match in a complex sequence having one or more AA per Edman degradation cycle) in the gel band CNP8 (Fig. 1), which matches both contigs 109 and 226 (Table 1).

Two clones coding for products with similarity to antigen-5 proteins were fully sequenced, producing the predicted protein sequences CQ_SAG5_1 and CQ_SAG5_2 (Table 2). The alignments with other Dipteran salivary antigen-5 family members, and with the salivary allergen 2 of the cat flea *Ctenocephalides felis* are shown in Fig. 5. Note that most Diptera have

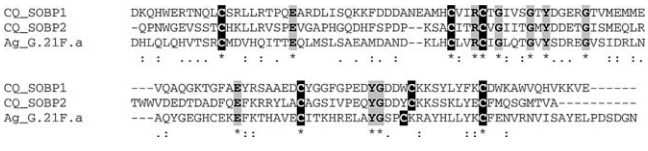

Fig. 4. Comparison of CQ_SOBP1 and CQ_OBP2 with the *Anopheles gambiae* protein G.21F.a (gi|27414083). See legend of Fig. 1A for other details.

12 conserved cysteine residues (Fig. 5A), while the flea has eight. The flea sequence is also uniquely recognizable for containing the sequence His–Tyr–Thr rather than the dipteran consensus His–Phe–Thr before Cys 7. The two higher Diptera (*Glossina* and *Stomoxys*) sequences are uniquely characterized by not having the 9th and 11th conserved cysteine, but rather having two other cysteine sets (Fig. 5A). A subgroup of mosquito proteins from both *Aedes* and *Anopheles* lack the 9th and 11th conserved cysteine residues. The phylogram shows trichotomy with clades containing different species, indicative of ancient origin of these gene duplications (Fig. 5B). Notice that all mosquito sequences lacking the 9/11 cysteine residues, except for one single sequence containing them, cluster within one clade (shown as III in Fig. 5B). Another clade (II) contains the higher Diptera, the flea and two culicine sequences, and the remaining sequences of mosquito and *Lutzomyia* sand fly (Fig. 5B), all having the 9/11 Cys residues, are clustered into clade I.

3.2.4. Enzymes with sugar digestion function

Several clusters of transcripts gave similarity to enzymes associated to sugar meal digestion, including an abundant cluster with 13 transcripts coding for a protein very similar to the salivary putative maltase precursor of *A. aegypti* (James et al., 1989), and other clusters coding for a protein similar to the previously described salivary amylase precursor of *Ae. aegypti* (Grossman et al., 1997; Grossman and James, 1993)

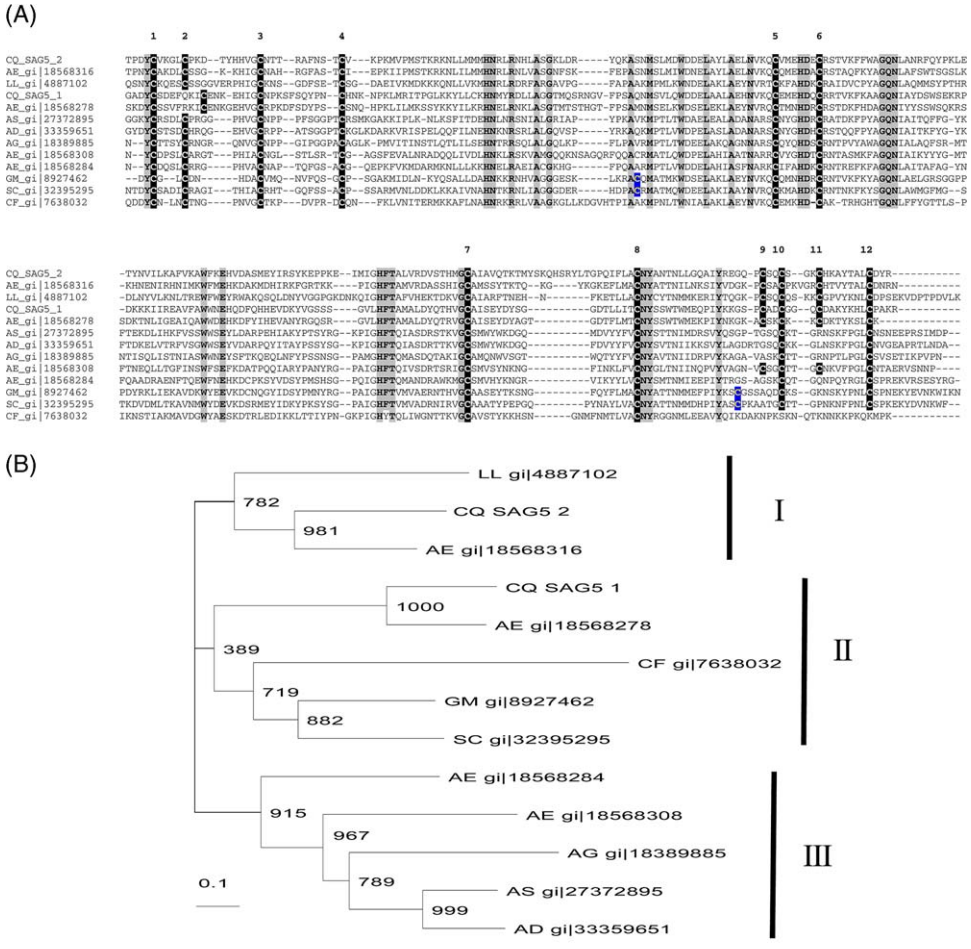

Fig. 5. Comparison of CQ_SAG5_1 and CQ_SAG5_2, 2 novel members of the antigen- 5 family of proteins, with other salivary proteins of blood-sucking Diptera and fleas. The NCBI accession number for the proteins is found following the gi| indicator. AE = *Aedes aegypti*, AS = *Anopheles stephensi*, AD = *Anopheles darlingi*, AG = *Anopheles gambiae*, SC = *Stomoxys calcitrans*, GM = *Glossina morsitans*, CF = *Ctenocephalides felix*. (A) Clustal alignment of the sequences. (B) Unrooted phylogram. For other details, see legend of Fig. 2.

(Table 1). Some of these clusters coding for maltase and amylase-like proteins appear to be truncated and may represent the same gene product. Maltase, but not amylase, gene products have also been described in salivary transcriptomes of anopheline mosquitoes (Francischetti et al., 2002b; Valenzuela et al., 2003), while amylase-coding transcripts and enzyme activity have been described in adult female *Lutzomyia* salivary glands (Charlab et al., 1999; Ribeiro et al., 1999).

In addition to maltase and amylase enzymes, two clusters, each with one transcript, were found to code for putative proteins with similarity to viral, microbial, protozoan, and plant, but only distantly to metazoan, endochitinases. It is possible that these endochitinases transcripts derive from a *Culex* parasite symbiont genome.

3.2.5. *Enzymes possibly associated with blood feeding*

Several transcripts in the salivary cDNA library code for enzymes that may play a role in the blood meal of *C. quinquefasciatus*: The singleton sequence of cluster 21 codes for a product with similarity to proteins annotated as apyrase and 5′ nucleotidases, including *An. gambiae* and *Ae. aegypti* apyrases and the chrysoptin precursor, an anti-platelet salivary protein of a *Chrysops* spp. tabanid fly (Reddy et al., 2000), which was reported as a novel anti-platelet protein but has substantial similarities to 5′-nucleotidases and apyrases. Mosquito salivary apyrases are members of the ubiquitous 5′-nucleotidase family that evolved to display hydrolytic activity to di- and tri- instead of mononucleotides (Champagne et al., 1995; Smartt et al., 1995). While most 5′-nucleotidases are extracellular membrane-bound enzymes by virtue of a glycoinositol anchor that esterifies to a carboxyterminally conserved serine residue, these salivary apyrases have become secreted enzymes by losing the motif leading to the attachment of the lipidic anchor (Champagne et al., 1995). The full-length sequence of *Culex* apyrase was obtained (Table 2), displaying 39–41% identity with *An. gambiae*, *Ae. aegypti* apyrases, and Chrysoptin (Table 2, electronic version). *Culex* apyrase lacks the conserved serine surrounded by aromatic or aliphatic amino acids in its carboxyterminal region (the site of esterification of inositol phospholipid membrane anchors) when compared to the equivalent region of membrane bound ecto-5′-nucleotidases of vertebrate and invertebrate origins (Fig. 6). The lack of an anchor site indicates that this enzyme, which contains a signal peptide, is secreted and not held to the cellular membrane as an ecto-enzyme. Although *C. quinquefasciatus* has low salivary apyrase activity when compared with other mosquitoes, the activity is clearly present (Ribeiro, 2000), and may help to prevent platelet aggregation by ADP released from injured cells and platelets during the probing phase of the blood meal. It is possible that the reported *Culex* salivary apyrase activity (Ribeiro, 2000) derives from this gene product.

One transcript from the *C. quinquefasciatus* transcriptome coded for a protein with similarity to *An. gambiae* and *Drosophila melanogaster* proteins of unknown function, and also with *Aedes* putative salivary purine hydrolase, an enzyme converting inosine into hypoxanthine plus ribose (Table 1). The transcript for such enzyme was previously found in *Aedes* sialotranscriptomes (Ribeiro and Valenzuela, 2003; Valenzuela et al., 2002b), and later, the enzymatic activity was found in *Ae. aegypti* salivary glands, the richest known source of this enzyme (Ribeiro and Valenzuela, 2003). It was proposed that this enzyme confers selective advantage to the mosquito by destroying purines that may lead to mast cell degranulation during the feeding process. Interestingly, in the same work, this enzyme activity was not found in the salivary gland homogenates of *Culex* or *Anopheles*. It may be possible that the *Culex* gene product, which is only 38% identical to the *Aedes* enzyme, has different nucleotide specificity.

Also associated with nucleotide hydrolysis is one relatively abundant cluster (with six ESTs) coding for enzymes annotated as endonucleases (Table 1), and also giving similarity to two tsetse salivary proteins named Tsal1 and Tsal2 (Li et al., 2001), of unknown function. The full-length clone of the *Culex* protein (CQ_SENDONUC; Table 2) has the NUC Smart motif, indicative of DNA/RNA non-specific endonucleases and phosphodiesterases (Table 2), and matches shrimp and crab endonucleases as well (Shagin et al., 2002; Wang et al., 2000). These arthropod enzymes have activity similar to vertebrate pancreatic DNAse I;

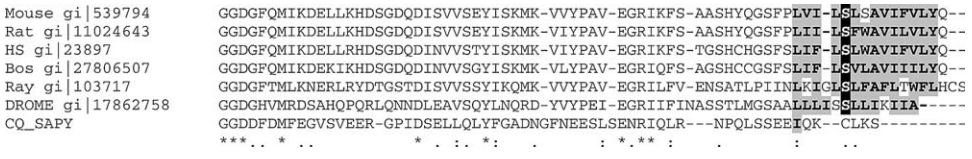

Fig. 6. Lack of inositol phosphate anchoring domain in *Culex quinquefasciatus* putative salivary apyrase (CQ_SAPY). Alignment of the carboxyterminal region of *C. quinquefasciatus* salivary putative apyrase with 5′-nucleotidases from rat, mouse, human (HS), bovine (*Bos*), electric ray, and *Drosophila melanogaster* (DROME). The conserved serine where the inositol phosphate esterifies is marked in black background. Hydrophobic and aromatic amino acids are marked in gray background. For other details, see legend of Fig. 2A.

however, they belong to a distinct protein family, despite sharing similarities near the active site (Shagin et al., 2002). This protein family, or any other endonucleases, was not previously identified in mosquitoes or sand fly salivary glands. Alignment of the arthropod proteins shows the presence of 10 conserved cysteines residues, of which *Culex* does not have the 10th, but has four additional Cys residues (Fig. 7). The shrimp enzyme, for which more detailed structural studies have been done (Shagin et al., 2002), has an extra Cys that appears to be linked to an unknown ~500 Da residue. The putative active center of the shrimp enzyme has been deduced from conserved amino acid residues found by aligning the active sites of different DNAses (Shagin et al., 2002). Most of these conserved residues (marked in blue in Fig. 7) are present in the *Culex* enzyme, as follows: The KGH triad on the shrimp enzyme (marked A in Fig. 7) is actually represented as RGH in the mosquito enzyme. This RGH triad was found in most DNAses used in the Shagin et al. (2002) alignments, and thus represents a conserved element. Except for the last conserved Arg residue (under the F mark in Fig. 7) all other six regions are conserved in the mosquito sequence. A Lys exists in the mosquito sequence within one residue of distance of the missing Arg. The mosquito putative enzyme thus has eight of

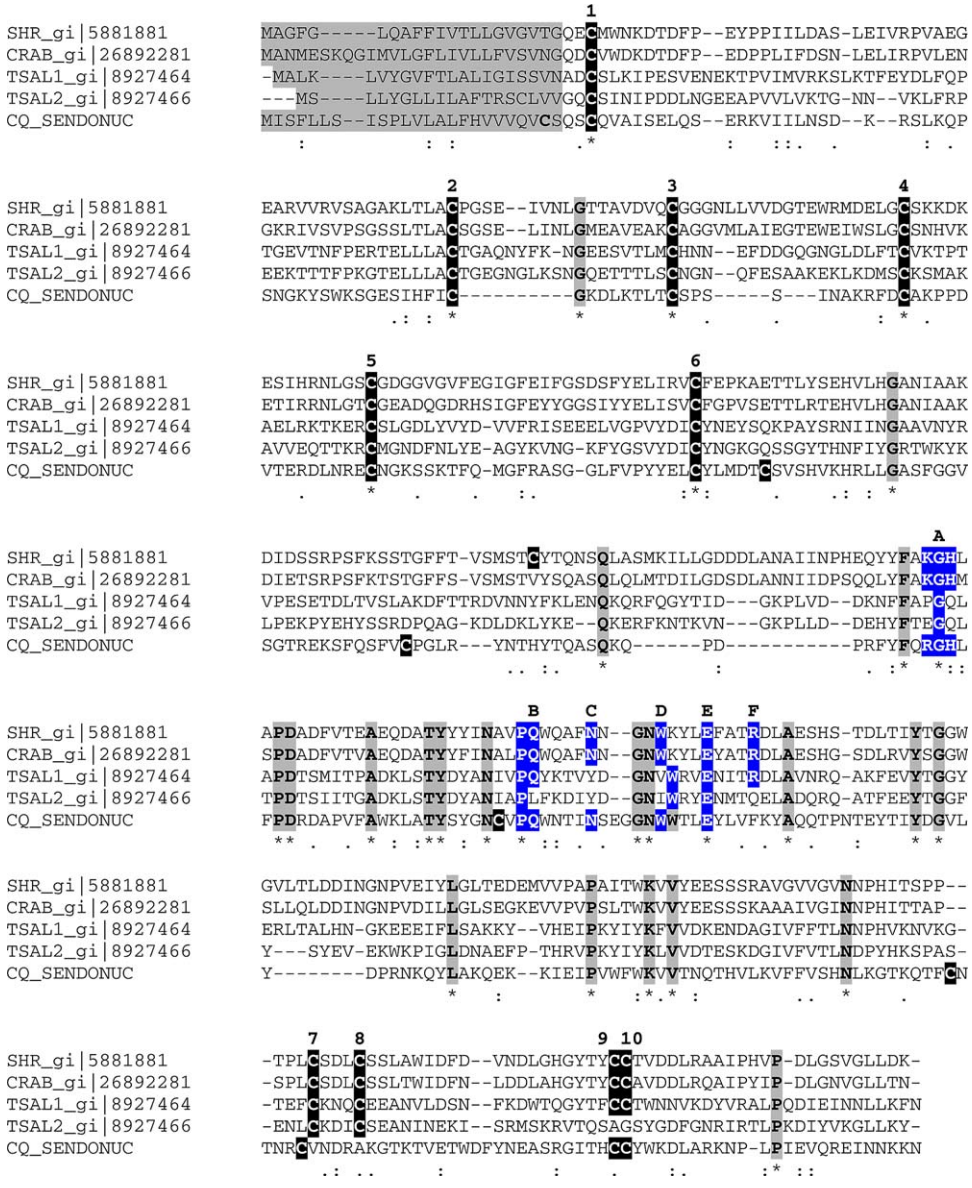

Fig. 7. Alignment of *Culex quinquefasciatus* putative salivary endonuclease with shrimp (SHR) and crab endonucleases, and two salivary proteins of the tsetse *Glossina morsitans* (Tsal1 and Tsal2). The 10 conserved cysteines are shown in reversed background and are numbered. Amino acids composing the active center are shown in blue background and are marked by letters above the alignments. Other identically conserved residues are shown in gray background. The signal peptide is shown in gray background in the aminoterminal region of the protein sequences.

the nine conserved active site amino acid residues found in this arthropod family of DNAses. The tsetse putative active site residues are less conserved, displaying five (Tsal2) or six (Tsal1) conserved residues out of nine. It is highly suggestive that CQ_SENDONUC, which has a clear signal peptide indicative of secretion, codes for a secreted endonuclease or a phosphodiesterase.

Three cDNA clusters code for proteins with similarity to esterases, phospholipases and lipases (Table 1), one of which was fully sequenced (Table 2). CQ_SLIP (Table 2) codes for a salivary lipase containing a signal peptide indicative of secretion sharing 39% identity to an *An. gambiae* putative protein, and other proteins annotated as lipase or phospholipases. CQ_SLIP has a Pfam lipase motif suggesting a triglyceride lipase activity. The role of these lipid-hydrolyzing enzymes in blood feeding is unknown, but could be responsible for the reported salivary PAF-hydrolase activity from *Culex* (Shagin et al., 2002) or, speculatively, with formation of the vasodilatory and anesthetic cannabinoids 2-arachidonoylglycerol or arachidonoylethanolamide (Howlett, 2002).

Hyaluronidase, an enzyme common in arthropod venom where it serves to spread the poison into the target tissues, has been previously reported in the salivary glands of sand flies, black flies, and ticks (Cerna et al., 2002; Charlab et al., 1999; Neitz et al., 1978; Ribeiro et al., 2000), but not in mosquitoes. Interestingly, one cDNA sequence from the *Culex* sialotranscriptome codes for a protein with similarity to the honeybee venom hyaluronidase (Table 1). Full-length sequence of this clone (Table 2) indicates a protein containing a signal peptide indicative of secretion and having 46% identity to an *An. gambiae* putative protein, and 45% identity to bee venom hyaluronidase. It is possible that *Culex*, similar to sand flies, has co-opted this enzyme into its salivary armamentarium, to help the spread of antihemostatic agents in the host skin during probing and blood feeding.

3.2.6. *Putative protease inhibitors*

Six clusters of cDNA transcripts representing nine sequences, code for putative protease inhibitors of different families (Table 1). Two are similar to the putative salivary serpin of *Ae. aegypti* (Valenzuela et al., 2002b), one of which appears to be a truncated clone. The four remaining clones code for Cys-rich peptides with a Pfam TIL domain indicative of a trypsin inhibitor-like cysteine-rich domain. Three of these clones were fully sequenced, indicating they code for Cys-rich peptides, all containing a signal peptide indicative of secretion. CQ_SFIB gives similarity to fibroin and to the human major epididymis-specific protein E4 precursor, which is an endopeptidase inhibitor. Alignments of matching peptides of similar size from *Drosophila*, *An. gambiae*, human, and pig origins indicate a number of conserved cysteines and other residues (Fig. 8). The mammalian peptides are clearly distinguished by having additional Cys residues. An HMM was created from the alignment shown in Fig. 8 and used to search the NR protein database. This resulted in 68 sequences with a statistical significance to the HMM, from which 26 have an E value smaller than 1E−6. Fourteen of these proteins are annotated as having WAP (whey acidic protein), four-disulfide core domain, and include epididymal secretory proteins of several mammals. This signature, consisting of two domains, has been proposed to be associated with protease inhibitors and growth factors (Ranganathan et al., 1999). Domain 1 of WAP is similar to the region marked as A in Fig. 8, but domain 2 of WAP is only partially similar to the B region shown in Fig. 8. We conclude that this protein family belongs to a subset of the WAP family of proteins.

CQ_SPI_1 putative translation product, which has a signal peptide indicative of secretion, matches a trypsin inhibitor-like family member of the worm *Caenorhabditis elegans* (Table 2), and has a distinct cluster of amino acids matching an anticlotting peptide of the worm *Ancylostoma caninum* (Stanssens et al., 1996) (Fig. 9). These worm anticlotting peptides have 10 Cys residues (Stanssens et al., 1996), eight of which are conserved with CQ_SPI_1 (Fig. 9). Interestingly, no similarities to CQ_SPI-_1 were found with other known mosquito salivary proteins. This is not due to its rarity, at least in the *Culex* transcriptome, because two clones in the cDNA library coded for CQ_SP_1 (Table 1).

CQ_SPI_2 codes for a putative secreted Cys-rich protein having similarities to other peptides from parasitic worms (gi|23451019), from a putative salivary protein from *A. stephensi* (gi|27372905) (Valenzuela et al., 2003), and from bee hemolymph (Bania et al., 1999)

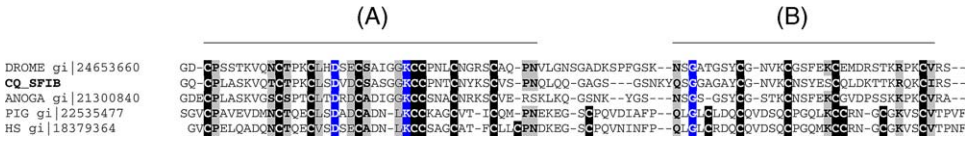

Fig. 8. Alignment of the salivary cysteine-rich peptide similar to fibroin (CQ_SFIB) with similar peptides from *Anopheles gambiae* (ANOGA), *Drosophila melanogaster* (DROME), pig, and human (HS) origins. Conserved cysteines are marked in black background; other identically conserved residues are marked in blue background, while functionally conserved residues are in gray background. Region A is similar to domain 1 of WAP (whey acidic protein) four-disulfide core domain and region B. Region marked B is similar to domain 2 of WAP.

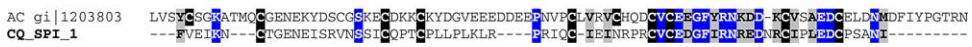

Fig. 9. Alignment of the *Culex quinquefasciatus* putative salivary protease inhibitor CQ_SPI_1 with the anticoagulant protein C2 precursor from *Ancylostoma caninum*. Conserved cysteines are marked in black background; other identically conserved residues are marked in blue background, while functionally conserved residues are in gray background. The bar indicates the region of greater conservation between these two peptides.

annotated as serine protease inhibitors (Table 2). The alignments do not show remarkable conservation beyond 10 common Cys residues, and a few other residues (Fig. 10). The role of these Cys-rich peptides as protease inhibitors, and their specificity, can only be speculated upon.

3.2.7. *Immunity-related proteins*

Lysozyme activity has been previously described in the salivary glands of both adult male and female mosquitoes, where it may play a role in preventing bacterial growth in sugar meals stored in the mosquito crop (Moreira-Ferro et al., 1998; Pimentel and Rossignol, 1990; Rossignol and Lueders, 1986). More recently, other immune-related products, such as antimicrobial peptides and lectins, have been found to be expressed in the salivary gland of infected mosquitoes (Dimopoulos et al., 1998), and in the sialotrancriptomes of mosquitoes (Valenzuela et al., 2002b, 2003). The *C. quinquefasciatus* sialotranscriptome produced two clones matching *Culex pipiens* gambicin (Bartholomay et al., 2003), an antibacterial peptide first characterized in *An. gambiae* (Vizioli et al., 2001), and designated putative infection-responsive short peptide. CQ_SGAMBIC, which was fully sequenced several times, is 97% identical at the amino acid level to *C. pipiens* gambicin, and 69% and 59% identical to the *Ae. aegypti* and *An. gambiae* homologues, respectively (Table 2, electronic version). It is interesting that when the *Ae. aegypti* putative protein was described, no function for that protein could be assigned.

CQ-contig_82 (Table 1) codes for a product similar to proteins annotated as C-type lectins, and also producing a Pfam lectin_c match (Table 1). Full-length sequence of this clone, named CQ_SCLEC (Table 2), indicates similarity to several insect proteins, including the putative salivary C-type lectin of *Ae. aegypti* (Valenzuela et al., 2002b), with which it shares 30% identity and 44% similarity at the AA level over the 154 residue length of the *Aedes* protein. The best match to the *An. gambiae* proteome, a non-annotated protein, produces 30% and 51% identity and similarity, respectively, over the 155 residue length of the protein. Although C-type lectins are implicated in animal immunity reactions (Dimopoulos et al., 2000; Vasta et al., 1999), it is also associated with anti-clotting activity in snake venom (Koo et al., 2002; Monteiro and Zingali, 2000), and in the salivary glands of the sand fly *Lutzomyia longipalpis* (Charlab et al., 1999), in addition to other functions (Loukas and Maizels, 2000). The high divergence of CQ_SCLEC when compared with the *Aedes* and *Anopheles* sequences, which are the probable orthologs, suggests that this lectin is evolving at a fast pace.

The sialotranscriptome of *C. quinquefasciatus* also produced a match to the putative gram negative bacteria-binding protein (GNBP) of *An. gambiae* (Dimopoulos et al., 1997); however, this clone is truncated (Table 1). GNBPs show similarities to the β-1,3 glucan-binding region of glucanases and are likely components of the PPO-activation cascade. This same clone produced a better match to a putative secreted salivary protein of *Ae. aegypti*, most likely the homologue of *An. gambiae* GNBP. In physiologic agreement with the presence of GNBP is the finding that CQ-contig_262 (also a truncated clone) (Table 1) codes for the carboxyterminal region of serine proteases including *An. gambiae* serine protease 14D2, a hemolymph protease that has a CLIP domain and changes in transcript abundance in response to bacteria injections (Gorman et al., 2000). The common finding of these immune-related transcripts associated with the salivary glands is indicative that this organ may produce more than lysozyme to control bacterial infections in the crop-stored sugar meals.

3.2.8. *Mucins and other low-complexity proteins*

Twenty-one clusters representing a total of 84 cDNA clone sequences code for proteins of low complexity, some of which are similar to mucins (Table 1). The abundant cluster 115, with 17 sequences, codes for pro-

Fig. 10. Alignment of CQ_SP_2, a putative salivary serine protease inhibitor from *Culex quinquefasciatus*, with the putative trypsin-like inhibitor protein precursor from the worm *Oesophagostomum dentatum* (OD), the putative salivary secreted serine protease inhibitor from *Anopheles stephensi* (ANST) and the bee AMCI_APIME Chymotrypsin inhibitor.

teins with the sequence GKLPGMRGEA predicted following the cleavage of a signal peptide for which the sequence GKLPGMRXEA was found by Edman degradation in the gel band CNP-9 (Table 1 and Fig. 1). Similarly, the sequence DQRCTYLRCRTEFRKTGAY is coded by sequences in cluster 239 (with two sequences) after a predicted signal peptide from which the sequence DQXXTYLRXXTEFRKTGAY was found by Edman degradation in bands CQ-1 and CNP-12 (Table 1 and Fig. 1), suggesting these two clusters code for abundantly expressed proteins.

Full-length sequence information for eight different clones coding for low complexity proteins was obtained (Table 2). Six code for probable mucins, containing from 5 to as many as 78 glycosylation sites as predicted by the program NetoGly (Hansen et al., 1998). Two of these predicted proteins are similar to previously described salivary putative proteins of *Ae. aegypti* and *Ae. stephensi* (CQ_SMUC6 and CQ_SMUC7), and one is similar to a putative *An. gambiae* hypothetical protein (CQ_SMUC1). CQ_SMUC2 is extremely Ser–Thr rich, these two amino acids constituting 45% of the protein total residues. Two other potentially *O*-galactosylated mucins (CQ_S56.6PTN and CQ_S30K_2) are similar to previously described salivary proteins of *Ae. aegypti*, named putative 56.5 kDa secreted protein and 30 kDa salivary gland allergen, which are proteins of unknown function. An additional putative protein (CQ_S30K_1), containing only one predicted *O*-galactosylation site, is weakly similar to the *Aedes* 30 kDa salivary gland allergen. The amino acid sequence XGKLPGMRXEA obtained by Edman degradation matches the predicted mature aminoterminal after signal peptide cleavage (Tables 1 and 2).Two other low-complexity putative proteins of unknown function have Gln (CQ_QQQ_1) or Gly–Gln–Gln (CQ_GQQ_1) repeats (Table 2). These proteins may be involved in extracellular matrix adhesion phenomena.

3.2.9. *Unknown protein families*

Thirty-eight clusters representing 91 sequences code for putative secreted proteins of unknown function and with no significant matches to known proteins, even when the low-complexity filter of the BLAST program is turned off (Table 1). Eight of these clusters were matched by Edman degradation products of protein bands depicted in the Fig. 1 experiment, indicating these messages are expressed into relatively abundant proteins. From these clusters, 27 clones were sequenced from the starting Met to the stop codon, yielding information on putative secreted proteins of mature molecular weights varying from a relatively small peptide of 1.7 kDa to a 42 kDa protein.

To further characterize these novel proteins, these protein sequences of unknown families were compared among themselves with the program BlastP (low-complexity filter off) and the proteins arbitrarily clustered for those yielding 40% amino acid similarity over at least 70% of the length of the larger protein pair. It was thus found that five of these proteins constitute a novel family containing four conserved Cys residues in addition to other residues, including four tryptophanes, a relatively rare amino acid (Fig. 11A). All have signal peptide indicative of secretion, and mature molecular weights of 15–17 kDa.

A second protein family was detected by clustering the salivary *Culex* protein as indicated in the previous paragraph. This family also contains four conserved cysteines and three to four conserved tryptophanes (Fig. 11B). These also code for secreted proteins with mature molecular weight ranges of 15–17 kDa.

We found it interesting that these two novel families have the same number of conserved cysteines, have conserved Trp residues, and are of similar molecular mass. Additional inspection of other protein sequences in the unknown category of Table 2 having molecular masses of 15–17 kDa identified two additional proteins

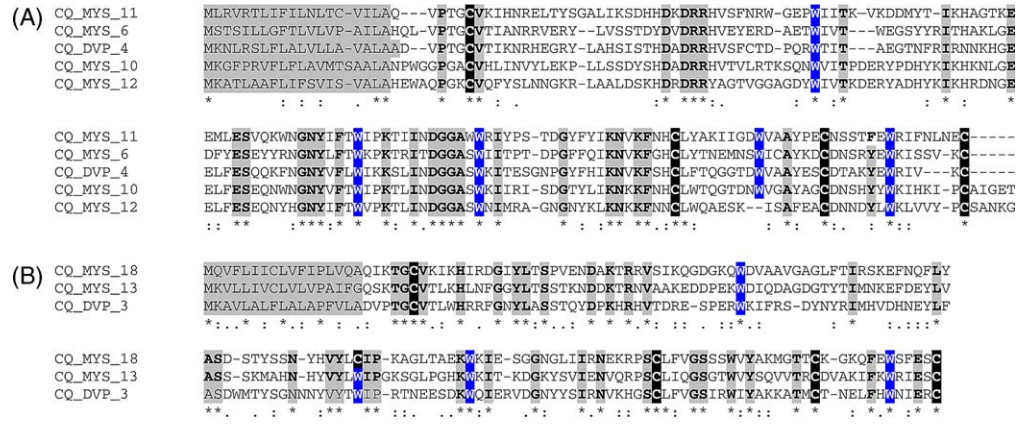

Fig. 11. Two novel families (A and B) of cysteine- and tryptophane-rich putative salivary proteins from *Culex quinquefasciatus*.

that are possible members of this superfamily. The ClustalW alignment of these 10 proteins indicates a superfamily having four conserved cysteines and three conserved tryptophane residues (Fig. 12A). Four members of this family have the sequence DVPG... following cleavage of the predicted signal peptide (Marked in gray background in Fig. 12A). Aminoterminal sequences containing DVPG... were found by Edman degradation of protein bands from the gel experiments shown in Fig. 1 (Tables 1 and 2) indicating that at least some members of this protein family are abundantly expressed in the salivary glands of *Culex*. The bootstrapped phenogram (Fig. 12B) resulting from the alignment shown in Fig. 12A, indicates that some of these family members share clear relationships; however, overall the family has evolved, most likely by gene duplications, beyond recognition of a common ancestor. This indicates a very ancient origin for this unique superfamily, or, alternatively, a very fast pace of evolution.

An HMM made with the alignments in Fig. 12A was used to search the NR database (containing 1,529,764 sequences on 10-28-2003) to which all salivary protein sequences reported in Table 2 were also added. Significant matches were found only to the *Culex* protein reported in Fig. 12A and, additionally, CQ_MYS_14 and CQ_MYS_9. We conclude that *C. quinquefasciatus* expresses a novel family of proteins in their salivary glands. These proteins are indicated in Table 2 as CWRC family, for cysteine–tryptophane-rich-proteins of *Culex*. The function of this protein family is unknown, but its members could be used in immunoassays to detect human exposure to this mosquito genus, assuming they are antigenic. Table 2 also reports for 14 additional full-length putative protein or peptide sequences for which no clue to their function or family membership is known.

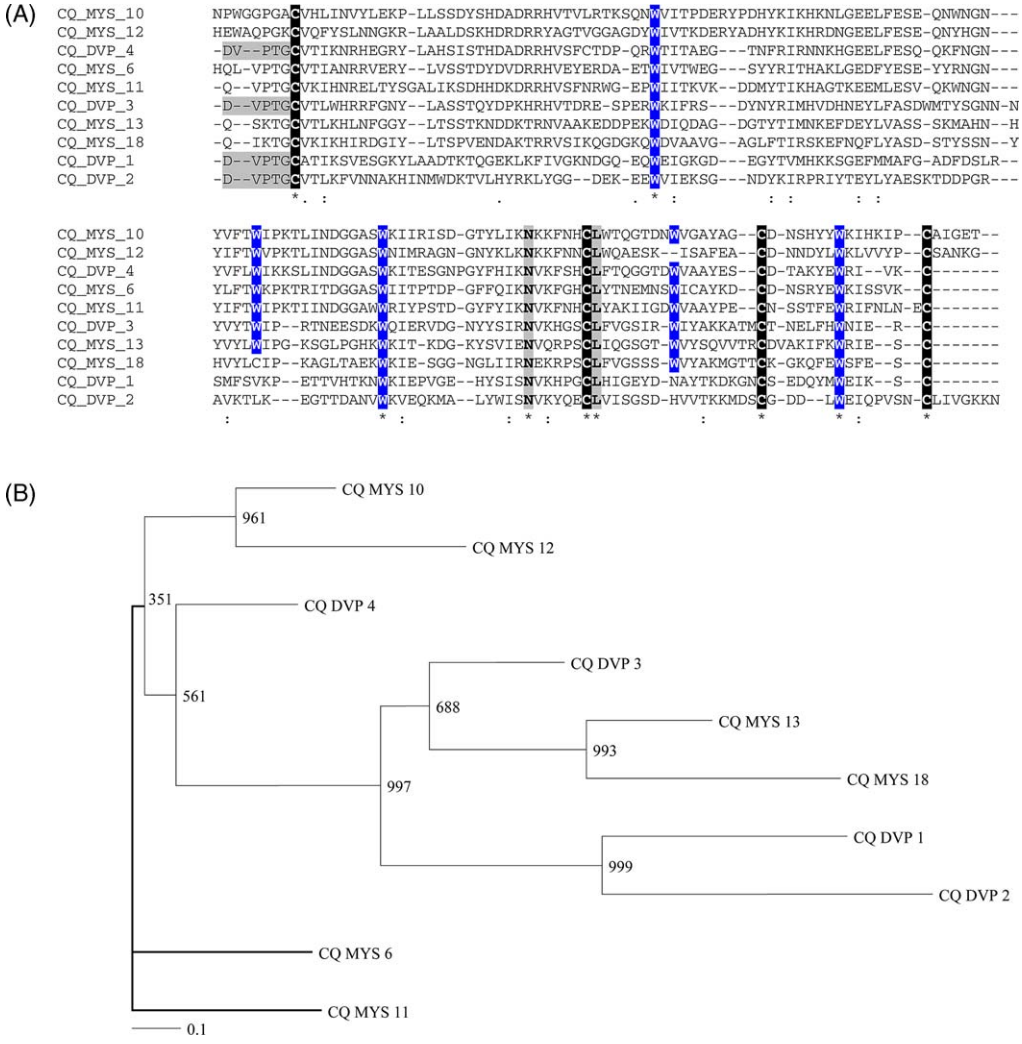

Fig. 12. The CWRC superfamily of *Culex quinquefasciatus* salivary proteins. (A) Alignment of 10 protein sequences rich in cysteine and tryptophane. (B) Unrooted phylogram based on the alignment in A. For other details, see legend to Fig. 2.

4. Concluding remarks

The adult female *C. quinquefasciatus* sialotranscriptome, when compared with the equivalent sets of *Aedes* and *Anopheles* mosquitoes, displays some remarkable differences. The messages for a completely novel family of proteins containing at least 12 members were discovered in the present study (Table 2 and Fig. 12). At least some members of the family are confirmed to be expressed, as indicated by the presence of the predicted aminoterminal sequences in regions of the gel coinciding with the expected size of the proteins (Tables 1 and 2, Fig. 1). Additionally, several putative protease inhibitors were found, including a member of a family previously thought to be nematode-specific (Fig. 9), and a family previously found in *An. stephensi* salivary transcriptome.

From the point of view of transcripts coding for putative secreted enzymes, the *Culex* sialotranscriptome is also remarkable in that it contains not only apyrase (found in *Anopheles* and *Aedes*), adenosine deaminase (found otherwise only in *Aedes*), purine nucleosidase (previously *Aedes* only), but it also contains an endonuclease described before only in shrimps and crabs, and hyaluronidase, thus far found only in the salivary glands of sand flies and black flies, although this enzyme is a common occurrence in venoms of vertebrate and invertebrate origins. Although adenosine deaminase and apyrase activities were demonstrated before in *Culex* salivary glands, the presence of these other activities remains to be demonstrated. Several lipases and esterases are also encoded, and one of these could account for the previously described PAF-hydrolase of *Culex* salivary glands.

Most of the transcripts, as in other sialotranscriptomes, belong to either known families of unknown function, or are entirely of an unknown function and families. Some of these may code for the still molecularly uncharacterized anticoagulants and vasodilators of *Culex quinquefasciatus*. The publicly available transcriptome of this mosquito may advance the effort to characterize these pharmacologically interesting molecules.

Acknowledgements

We are grateful to Drs Robert Gwadz, Thomas Wellems, and Thomas Kindt for support and encouragement and to Nancy Shulman for editorial assistance.

References

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

Andersen, J.F., Francischetti, I.M., Valenzuela, J.G., Schuck, P., Ribeiro, J.M., 2003. 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., 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.

Bania, J., Stachowiak, D., Polanowski, A., 1999. Primary structure and properties of the cathepsin G/chymotrypsin inhibitor from the larval hemolymph of *Apis mellifera*. Eur. J. Biochem. 262, 680–687.

Bartholomay, L.C., Farid, H.A., Ramzy, R.M., Christensen, B.M., 2003. *Culex pipiens pipiens*: characterization of immune peptides and the influence of immune activation on development of Wuchereria bancrofti. Mol. Biochem. Parasitol. 130, 43–50.

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L., Sonnhammer, E.L., 2000. The Pfam protein families database. Nucleic Acids Res. 28, 263–266.

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

Champagne, D.E., Smartt, C.T., Ribeiro, J.M., 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., Rowton, E.D., Ribeiro, J.M., 2000. The salivary adenosine deaminase from the sand fly *Lutzomyia longipalpis*. Exp. Parasitol. 95, 45–53.

Charlab, R., Valenzuela, J.G., Rowton, E.D., 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. U S A 96, 15155–15160.

Dimopoulos, G., Casavant, T.L., Chang, S., Scheetz, T., Roberts, C., Donohue, M., Schultz, J., Benes, V., Bork, P., Ansorge, W., et al., 2000. *Anopheles gambiae* pilot gene discovery project: identification of mosquito innate immunity genes from expressed sequence tags generated from immune-competent cell lines. Proc. Natl. Acad. Sci. USA 97, 6619–6624.

Dimopoulos, G., Richman, A., Muller, H.M., 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.

Dimopoulos, G., Seeley, D., Wolf, A., Kafatos, F.C., 1998. Malaria infection of the mosquito *Anopheles gambiae* activates immune-responsive genes during critical transition stages of the parasite life cycle. EMBO J 17, 6115–6123.

Dohm, D.J., O'Guinn, M.L., Turell, M.J., 2002. Effect of environmental temperature on the ability of *Culex pipiens* (Diptera: Culicidae) to transmit West Nile virus. J. Med. Entomol. 39, 221–225.

Francischetti, I.M., Andersen, J.F., 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., 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.

Gorman, M.J., Andreeva, O.V., Paskewitz, S.M., 2000. Molecular characterization of five serine protease genes cloned from *Anopheles gambiae* hemolymph. Insect Biochem. Mol. Biol. 30, 35–46.

Grossman, G.L., Campos, Y., Severson, D.W., James, A.A., 1997. Evidence for two distinct members of the amylase gene family in

the yellow fever mosquito, *Aedes* aegypti. Insect Biochem. Mol. Biol. 27, 769–781.

Grossman, G.L., 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.

Hansen, J.E., Lund, O., Tolstrup, N., Gooley, A.A., Williams, K.L., 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., Carlson, J.R., 2000. Molecular evolution of odorant-binding protein genes OS-E and OS-F in *Drosophila*. Genetics 155, 117–127.

Horsfall, W.R., 1955. Mosquitoes: their Bionomics and Relation to Disease Transmission. The Ronald Press Co, New York.

Howlett, A.C., 2002. The cannabinoid receptors. Prostaglandins Other Lipid Mediat. 68-69, 619–631.

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

Isawa, H., Yuda, M., Orito, Y., Chinzei, Y., 2002a. A mosquito salivary protein inhibits activation of the plasma contact system by binding to factor XII and high molecular weight kininogen. J. Biol. Chem. 13, 13.

James, A.A., Blackmer, K., Marinotti, O., Ghosn, C.R., 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., Racioppi, J.V., 1989. A salivary gland-specific, maltase-like gene of the vector mosquito, *Aedes aegypti*. Gene 75, 73–83.

King, T.P., Spangfort, M.D., 2000. Structure and biology of stinging insect venom allergens. Int. Arch. Allergy Immunol. 123, 99–106.

Koo, B.H., Sohn, Y.D., Hwang, K.C., Jang, Y., Kim, D.S., Chung, K.H., 2002. Characterization and cDNA cloning of halyxin, a heterogeneous three-chain anticoagulant protein from the venom of *Agkistrodon halys brevicaudus*. Toxicon 40, 947–957.

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

Loukas, A., Maizels, R.M., 2000. Helminth C-type lectins and host-parasite interactions. Parasitol. Today 16, 333–339.

Milne, T.J., Abbenante, G., Tyndall, J.D., Halliday, J., 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.

Monteiro, R.Q., Zingali, R.B., 2000. Inhibition of prothrombin activation by bothrojaracin, a C-type lectin from *Bothrops jararaca* venom. Arch. Biochem. Biophys. 382, 123–128.

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

Neitz, A.W., Howell, C.J., Potgieter, D.J., Bezuidenhout, J.D., 1978. Proteins and free amino acids in the salivary secretion and haemolymph of the tick *Amblyomma hebraeum*. Onderstepoort J. Vet. Res. 45, 235–240.

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

Page, R.D., 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.

Pimentel, G.E., Rossignol, P.A., 1990. Age dependence of salivary bacteriolytic activity in adult mosquitoes. Comp. Biochem. Physiol. 96B, 549–551.

Ranganathan, S., Simpson, K.J., Shaw, D.C., Nicholas, K.R., 1999. The whey acidic protein family: a new signature motif and three-dimensional structure by comparative modeling. J. Mol. Graph Model 17, 106–113, (see also pp. 134-136).

Reddy, V.B., Kounga, K., Mariano, F., Lerner, E.A., 2000. Chrysoptin is a potent glycoprotein IIb/IIIa fibrinogen receptor antagonist present in salivary gland extracts of the deerfly. J. Biol. Chem. 275, 15861–15867.

Ribeiro, J.M., 2000. Blood-feeding in mosquitoes: probing time and salivary gland anti- haemostatic activities in representatives of three genera (*Aedes*, *Anopheles*, *Culex*). Med. Vet. Entomol. 14, 142–148.

Ribeiro, J.M., Charlab, R., Rowton, E.D., Cupp, E.W., 2000. Simulium vittatum (Diptera: Simuliidae) and Lutzomyia longipalpis (Diptera: Psychodidae) salivary gland hyaluronidase activity. J. Med. Entomol. 37, 743–747.

Ribeiro, J.M., Charlab, R., Valenzuela, J.G., 2001. The salivary adenosine deaminase activity of the mosquitoes *Culex quinquefasciatus* and *Aedes aegypti*. J. Exp. Biol. 204, 2001–2010.

Ribeiro, J.M., Francischetti, I.M., 2001. Platelet-activating-factor-hydrolyzing phospholipase C in the salivary glands and saliva of the mosquito *Culex quinquefasciatus*. J. Exp. Biol. 204, 3887–3894.

Ribeiro, J.M., Francischetti, I.M., 2003. Role of arthropod saliva in blood feeding: Sialome and post-sialome perspectives. Annu. Rev. Entomol. 48, 73–88.

Ribeiro, J.M., Valenzuela, J.G., 2003. The salivary purine nucleosidase of the mosquito, *Aedes aegypti*. Insect Biochem. Mol. Biol. 33, 13–22.

Ribeiro, J.M.C., Rowton, E.D., Charlab, R., 1999. Salivary amylase activity of the phlebotomine sand fly, *Lutzomyia longipalpis*. Insect Biochem. Mol. Biol. 30, 271–277.

Ribeiro, J.M.C., Schneider, M., Guimaraes, J.A., 1995. Purification and characterization of Prolixin S (Nitrophorin 2), the salivary anticoagulant of the blood sucking bug, *Rhodnius prolixus*. Biochem. J. 308, 243–249.

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

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Schreiber, M.C., Karlo, J.C., 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.

Shagin, D.A., Rebrikov, D.V., Kozhemyako, V.B., Altshuler, I.M., Shcheglov, A.S., Zhulidov, P.A., Bogdanova, E.A., Staroverov, D.B., Rasskazov, V.A., Lukyanov, S., 2002. A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. Genome Res. 12, 1935–1942.

Smartt, C.T., Kim, A.P., Grossman, G.L., James, A.A., 1995. The Apyrase gene of the vector mosquito, *Aedes aegypti*, is expressed specifically in the adult female salivary glands. Exp. Parasitol. 81, 239–248.

Stanssens, P.B.P.W., Gansemans, Y.J.L.L.Y., Huang, S.M.S.M.J., Lauwereys, M.C.M.H.P.J., Lasters, I.V.G.P., 1996. Anticoagulant repertoire of the hookworm *Ancylostoma caninum*. Proc. Nat Acad. Sci. USA 93, 2149–2154.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Turell, M.J., O'Guinn, M.L., Dohm, D.J., Jones, J.W., 2001. Vector competence of North American mosquitoes (Diptera: Culicidae) for West Nile virus. J. Med. Entomol. 38, 130–134.

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

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

Valenzuela, J.G., Francischetti, I.M., Pham, V.M., Garfield, M.K., Ribeiro, J.M., 2003. Exploring the salivary gland transcriptome and proteome of the *Anopheles stephensi* mosquito. Insect Biochem. Mol. Biol. 33, 717–732.

Vasta, G.R., Quesenberry, M., Ahmed, H., O'Leary, N., 1999. C-type lectins and galectins mediate innate and adaptive immune functions: their roles in the complement activation pathway. Dev. Comp. Immuno. 23, 401–420.

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

Wang, W.Y., Liaw, S.H., Liao, T.H., 2000. Cloning and characterization of a novel nuclease from shrimp hepatopancreas, and prediction of its active site. Biochem. J. 346 (Pt 3), 799–804.