# University of Rhode Island DigitalCommons@URI

Past Departments Faculty Publications (CELS)

College of the Environment and Life Sciences

2002

# Exploring the sialome of the tick Ixodes scapularis

Jesus G. Valenzuela

Ivo M. B. Francischetti

Van My Pham

Mark K. Garfield

Thomas N. Mather University of Rhode Island

See next page for additional authors

Follow this and additional works at: https://digitalcommons.uri.edu/cels\_past\_depts\_facpubs

# **Citation/Publisher Attribution**

Valenzuela, J. G., Francischetti, I. M. B., Pham, V. M., Garfield, M. K., Mather, T. N., & Ribeiro, J. M. C. (2002). Exploring the sialome of the tick *Ixodes scapularis*. *Journal of Experimental Biology, 205*(18), 2843-2864. Retrieved from http://jeb.biologists.org/content/205/18/2843. Available at: http://jeb.biologists.org/content/205/18/2843

This Article is brought to you by the University of Rhode Island. It has been accepted for inclusion in Past Departments Faculty Publications (CELS) by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu. For permission to reuse copyrighted content, contact the author directly.

# Exploring the sialome of the tick Ixodes scapularis

# Authors

Jesus G. Valenzuela, Ivo M. B. Francischetti, Van My Pham, Mark K. Garfield, Thomas N. Mather, and José M. C. Ribeiro

# Terms of Use

All rights reserved under copyright.

# Exploring the sialome of the tick *Ixodes scapularis*

Jesus G. Valenzuela<sup>1</sup>, Ivo M. B. Francischetti<sup>1</sup>, Van My Pham<sup>1</sup>, Mark K. Garfield<sup>2</sup>, Thomas N. Mather<sup>3</sup> and José M. C. Ribeiro<sup>1,\*</sup>

<sup>1</sup>Medical Entomology Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda Maryland, 20892-0425, USA, <sup>2</sup>Research Technologies Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA and <sup>3</sup>Center for Vector-Borne Disease, University of Rhode Island, Kingston 02881-0804, USA

\*Author for correspondence (e-mail: jribeiro@nih.gov)

Accepted 17 June 2002

#### Summary

To attempt description of the set of mRNA and protein (sialome) expressed in the salivary glands of the tick *Ixodes scapularis*, we randomly sequenced 735 clones of a full-length salivary gland cDNA library of this arthropod and performed Edman degradation of protein bands from salivary gland homogenates (SGH) and saliva separated by SDS-PAGE. The sequences were grouped into 410 clusters, of which 383 are not associated with known *I. scapularis* sequences. 15- and 17-protein bands from PAGE yielded amino-terminal information on the saliva and salivary gland gels, respectively. We attributed 19 of these sequences to translation products of the cDNA library. Full-length sequences were obtained for 87 clones. Among these protein sequences are several protease

#### Introduction

Saliva of blood-sucking arthropods contains a large array of antihemostatic, anti-inflammatory and immunomodulatory components (Ribeiro, 1995). Tick saliva has been proposed to be important for formation and maintenance of the feeding cavity in host skin (Ribeiro, 1989; Wikel et al., 1994; Wikel, 1996). The tick Ixodes scapularis, the main vector of Lyme disease in the eastern US, has a salivary apyrase (Ribeiro et al., 1985) that destroys ADP, a main agonist of platelet aggregation. Platelet-activating factor and collagen-induced platelet aggregation inhibitors also exist in I. scapularis saliva (Ribeiro et al., 1985), as do anticlotting agents (Ribeiro et al., 1985) including Ixolaris, an inhibitor of tissue factor/FVIIa (Francischetti et al., 2002), inhibitors of neutrophil activation (Ribeiro et al., 1990) and inhibitors of T-cell activation (Ribeiro et al., 1985). The latter activity is mediated, at least in part, by an undefined protein that binds IL-2 (Gillespie et al., 2001). A salivary kininase in Ixodes (Ribeiro and Mather, 1998) destroys bradykinin, a mediator of pain and edema (Regoli and Barabe, 1980). The effects of inflammatory anaphylatoxins are also blocked, perhaps by the same kininase enzyme or other carboxypeptidases (Ribeiro and Spielman, 1986). I. scapularis saliva also has an inhibitor of the inhibitors of distinct classes, metalloproteases, novel proteins with histamine-binding domains, and several peptide families of unknown function displaying different conserved cysteine residues, many of which contain single Kunitz domains. This work provides information into the diversity of messages expressed in the salivary glands of *I. scapularis*, describes novel sequences that may be responsible for known biological activites, indicates further biological activities that may be present in *I. scapularis* saliva and identifies novel vaccine targets that may be used in Lyme disease prevention.

Key words: salivary gland, proteome, electrophoresis, hematophagy, Lyme's disease, tick, *Ixodes scapularis*.

alternative complement pathway, Isac, which was recently characterized molecularly (Ribeiro, 1987; Valenzuela et al., 2000). There is also evidence for the presence of salivary prostacyclin (Ribeiro et al., 1988) and prostaglandin E<sub>2</sub> (Ribeiro et al., 1985). Prostaglandins, in particular E<sub>2</sub> and F<sub>2</sub> $\alpha$ , have been described in saliva of other ticks (Dickinson et al., 1976; Higgs et al., 1976; Ribeiro et al., 1992); these prostaglandins are both vasodilators of skin vasculature and immunomodulators. Other than prostaglandins, Isac, the salivary anticomplement of *I. scapularis* (Valenzuela et al., 2000) and the anticlotting Ixolaris (Francischetti et al., 2002), no other pharmacologically active molecule in *I. scapularis* saliva has been molecularly characterized.

Tick saliva is also important in transmission of tick-borne pathogens for several reasons; it may enhance pathogen transmission, hypersensitivity to saliva may modify the site of inoculation of pathogens, and it may promote non-viremic transmission of viruses by cofeeding (Jones et al., 1987, 1990; Nuttall et al., 2000; Wikel et al., 1994; Wikel, 1996). A protein of unknown function (named SALP16) has been characterized by immunoscreening an expression salivary gland cDNA library obtained from *I. scapularis* nymphs (Das et al., 2000),

as have 13 other immunodominant proteins from *I. scapularis* (Das et al., 2001).

The composition of *I. scapularis* saliva is interesting in the study of the biology of parasite-host relationships, the discovery of novel biologically active components, and the identification of novel vaccine targets against I. scapularisvectored diseases. Toward these goals, we constructed a salivary gland cDNA library from blood-feeding I. scapularis and randomly sequenced 735 clones that yielded 410 cDNA clusters. Based on BLAST homology to other proteins in the non-redundant (NR) database, the presence of conserved domains of the SMART (Schultz et al., 2000) or Pfam (Bateman et al., 2000) databases, and the presence of a signal peptide indicative of secretion in these clones (Nielsen et al., 1997), we identified 100 clusters that are probably associated with secretory products. From these, we obtained full-length information on 87 different clones, herein reported, 19 of whose expression was confirmed by identification of their amino-terminal sequence in PVDF-transferred salivary proteins separated by SDS-PAGE. While descriptive in nature, this paper raises many hypotheses about the compositional diversity of blood-sucking arthropods and identifies several novel sequences that could have biological activity and possibly serve as vaccine targets.

#### Materials and methods

#### Water and organic compounds

All water used was of  $18 M\Omega$  quality and was produced by a MilliQ apparatus (Millipore, Bedford, MA, USA). Organic compounds were obtained from Sigma Chemical Corporation (St Louis, MO, USA) or as stated.

#### Ticks and tick saliva

Tick saliva was obtained by inducing partially engorged adult female *I. scapularis* to salivate (3–4 days post-attachment to a rabbit) into capillary tubes using the modified pilocarpine induction method (Valenzuela et al., 2000). Tick salivary gland extracts were prepared by collecting glands from partially engorged female *I. scapularis* as described (Valenzuela et al., 2000). Glands were stored frozen at –75°C until needed.

### Salivary gland cDNA library construction

*I. scapularis* salivary gland mRNA was isolated from 25 salivary gland pairs taken from adult females at days 3 and 4 after attachment to a rabbit host. The Micro-FastTrack mRNA isolation kit (Invitrogen, San Diego, CA, USA) was used to isolate mRNA, which was reverse transcribed to cDNA using Superscript II RNase H-reverse transcriptase (Gibco-BRL, Gaithersburg, MD, USA) and the CDS/3' primer (Clontech, Palo Alto, CA, USA). Second-strand synthesis was performed using a polymerase chain reaction (PCR)-based protocol with the SMART III primer (Clontech) as the sense primer and the CDS/3' primer as antisense primer. These two primers create *Sfi*I A and B sites at the ends of the nascent cDNA. Double-stranded cDNA was immediately treated with proteinase K

 $(0.8 \,\mu g \,\mu l^{-1})$  and washed three times with water using Amicon filters with a 100 kDa cutoff (Millipore). Double-strand cDNA was then digested with SfiI. cDNA was then fractionated using columns provided by the manufacturer (Clontech). Fractions containing cDNA of more than 400 base pairs (bp) were pooled, concentrated and washed three times with water using an Amicon filter with a 100 kDa cutoff. cDNA was concentrated and ligated into an 8-Triplex2 vector (Clontech). The resulting ligation reaction was packed using the Gigapack Gold III from Stratagene/Biocrest (Cedar Creek, TN, USA) following the manufacturer's specifications. The library thus obtained was plated by infecting log-phase XL1-blue cells (Clontech), and the amount of recombinants was determined by PCR using vector primers flanking the inserted cDNA and visualized on agarose gels with Ethidium Bromide. For more details, see Valenzuela et al. (2002).

#### Sequence of Ixodes scapularis cDNA library

The salivary gland cDNA library was plated to approximately 200 plaques per plate (150 mm diameter Petri dish). Randomly picked plaques were transferred to a 96-well polypropylene plate containing 100 µl of water per well. The bacteriophage sample  $(5 \mu l)$  was used as a template for a PCR reaction to amplify random cDNA using PT2F1 (5'-AAG TAC TCT AGC AAT TGT GAG C-3'), which is positioned upstream from the cDNA of interest (5' end), and PT2R1 (5'-CTC TTC GCT ATT ACG CCA GCT G-3'), which is positioned downstream from the cDNA of interest (3' end). Platinum Taq polymerase (Gibco-BRL) was used for these reactions. After removal of primers, the PCR product was used as a template for a cycle-sequencing reaction using the DTCS labeling kit from Beckman Coulter Inc. (Fullerton, CA, USA). The primer used for sequencing (PT2F3) is upstream from the inserted cDNA and downstream from primer PT2F1. After cycle sequencing the samples, a cleaning step was done using the multiscreen PCR 96-well plate cleaning system from Millipore. Dried samples were immediately resuspended with 25 µl of deionized ultrapure formamide (J. T. Baker, Phillipsburg, NJ, USA) and one drop of mineral oil was added to the top of each sample. Samples were sequenced immediately on a CEQ 2000 DNA sequencer (Beckman Coulter Inc.) or stored at  $-30^{\circ}$ C.

#### **Bioinformatics**

Detailed description of the bioinformatic treatment of the data can be found elsewhere (Valenzuela et al., 2002). Briefly, primer and vector sequences were removed from raw sequences, compared against the GenBank non-redundant (NR) protein database using the standalone BlastX program found in the executable package at ftp://ftp.ncbi.nlm.nih.gov/blast/executables/ (Altschul et al., 1997) and searched against the Conserved Domains Database (CDD) (found at ftp://ftp.ncbi.nlm.nih.gov/pub/mmdb/cdd/), which includes all Pfam (Bateman et al., 2000) and Smart (Schultz et al., 2000) protein domains. The predicted translated proteins were searched for a secretory signal through the SignalP server (Nielsen et al., 1997). Sequences were clustered using the BlastN program (Altschul et al., 1990) as detailed before (Valenzuela et al., 2002), and the data presented in the format of Table 1 in this paper. The electronic version of the table has additional hyperlinks to ClustalX (Jeanmougin et al., 1998) alignments as well as FASTA-formatted sequences for all clusters. The electronic table is available upon request; email: jribeiro@nih.gov.

#### Full-length sequencing of selected cDNA clones

A sample  $(4\mu)$  of the  $\lambda$ -phage containing the cDNA of interest was amplified using the PT2F1 and PT2R1 primers (same conditions as described above). The PCR samples were cleaned using the multiscreen PCR 96-well filtration system (Millipore). Cleaned samples were sequenced first with PT2F3 primer and subsequently with custom primers. Full-length sequences were again compared with databases as indicated for the nucleotide sequences above, and the data displayed as in Table 2, which has hyperlinks in its electronic version (available upon request to jribeiro@nih.gov).

#### SDS-polyacrylamide gel electrophoresis

NuPAGE 10% gels, 1 mm thick (Invitrogen), using reducing MES buffer, were electrophoresed according to the manufacturer's recommendations to resolve proteins in 60 µl of tick saliva. Salivary gland homogenates (SGH; 1.0 pairs per lane) were run in 12% gels under non-reducing conditions with Bis-Tris buffer. To estimate the molecular mass of detected proteins, SeeBlue™ markers from Invitrogen (myosin, bovine dehydrogenase, serum albumin, glutamic alcohol dehydrogenase, carbonic anhydrase, myoglobin, lysozyme, aprotinin and insulin, chain-B) were used. Samples were treated with NuPAGE LDS sample buffer (Invitrogen). For amino-terminal sequencing of the salivary proteins, the gels were transferred to PVDF membrane using 10 mmol l<sup>-1</sup> Caps, pH11.0, 10% methanol as the transfer buffer on a blot module for the XCell II Mini-Cell (Invitrogen). The membrane was stained with 0.025% Coomassie Blue in the absence of acetic acid. Stained bands were cut from the PVDF membrane and subjected to Edman degradation in a Procise sequencer (Perkin-Elmer Corporation). To find the cDNA sequences corresponding to the amino acid sequence obtained by Edman degradation, we wrote a search program that checked these amino acid sequences against the three possible protein translations of each cDNA sequence obtained in the mass sequencing project. A more detailed account of this program is found elsewhere (Valenzuela et al., 2002).

### Results

# Characterization of the library by DNA sequencing of randomly selected clones

To investigate the transcriptome of the salivary glands of feeding adult female *Ixodes scapularis* ticks, we randomly sequenced 735 clones from our unidirectionally cloned library. After clustering these sequences using BlastN with a cutoff of

10E-60, we found 410 unique clusters. All sequences within each cluster were compared with the NR protein database using the BlastX program (Altschul et al., 1997) and with the CDD database, containing all Pfam and SMART motifs (Bateman et al., 2000; Schultz et al., 2000), using the RPSblast program (Altschul et al., 1997). The three possible reading frames of each sequence were inspected for long reading frames with an initial methionine residue followed by at least 40 residues; these were submitted to the SignalP server for verification of secretory signal peptide. The results for each cluster were compiled as shown in Table 1, which displays the 30 most abundant clusters of this cDNA collection. 13 of the 30 clusters are possibly related to secretory products as they display a signal secretion peptide signature (Nielsen et al., 1997). Five clusters have indications of being related to membraneanchored or cytoplasmic proteins, while the remaining eight clusters give no conclusive indication of a leader signal peptide, probably due to diminished sequence quality at the 5'end. Notably, seven of the 30 clusters have Kunitz domains, found in many protease inhibitors such as anticlotting proteins. Of these 30 clusters, six had highly significant matches to five previously published I. scapularis salivary proteins, all of which are from clusters having a predicted signal secretory peptide sequence. When comparing all 16 known salivary protein sequences of I. scapularis (as of September 20, 2001) with the complete cDNA library described in this paper (using tBlastN), 13 were found in the library with a confidence value of 1E-30 or better, indicating they corresponded to the same or very closely related proteins. The three reported protein sequences not found in the translation of our library are: the SALP9 protein (gi|15428346), which matched the aminoterminal sequence of one of the clones and appears to be a signal sequence, yielding an E value of 1E–5 for the match; the salivary gland 16kDa protein SALP16 (gi 12002008), which identifies four cDNA with varying scores ranging from 2E-5 to 3E-15, the best alignments indicating 40% sequence identity; and finally, the 26kDa salivary protein B (gi 15428306), which has no matches to our database.

The complete Table 1 (available electronically; e-mail: jribeiro@nih.gov) containing 410 clusters was annotated to indicate whether each of the clusters is associated with a possibly secreted, probably housekeeping protein, or one of unknown function. These annotation and function assignments were based on both similarities to the NR or CDD databases and on whether the proteins indicate coding for a secretory signal peptide. We thus found 102 clusters possibly associated with secretory products. These 102 clusters account for a total of 310 sequences, or 42% of the cDNA database. Table 2 indicates the clusters possibly associated with secretory products, sorted alphabetically. The electronic version of the manuscript contains the tables for the clusters associated with probable housekeeping and unknown clusters, as well as links to all sequences, alignments and BLAST results.

Table 2 shows that, in addition to the 13 proteins indicated above, there are several clusters associated with anti-protease sequences or domains, such as  $\alpha$ -2-macroglobulin and cystatin,

Cluster	Number of						
number	seduences	Best BlastX match to NR database <sup>1</sup>	E-value <sup>2</sup>	Species of NR match	Best Rpsblast to CDD database <sup>3</sup>	E-value <sup>2</sup>	SignalP <sup>4</sup>
1	44	gi 15428308  14 kDa salivary	7E-34	Ixodes scapularis	pfam02414 Borrelia_orfA	1E-09	SIG
5	41				pfam02326 YMF19	0.0005	SIG
3	34	gi 5835682 gi 5835682 ref NP_008498.1 CO	6E-83	Ixodes hexagonus	pfam00115 COX1	1E-90	NoORF
4	20	gi 15428348  Salp10 [Ixodes s	5E-34	Ixodes scapularis	pfam00014 Kunitz_BPTI	0.001	SIG
5	18				pfam01028 Topoisomerase_I	0.0003	SIG
9	14						NoORF
7	14	gi 124810 IP52_ANESU PROTEASE INHIBITOR	0.01	Anemonia sulcata	Smart smart00131 KU	1E-06	SIG
8	11						NoORF
6	6	gi 15428310  25 kDa salivary	1E-105	Ixodes scapularis			SIG
10	8	gi 5835692 gi 5835692 ref NP_008508.1 CY	2E-73	Ixodes hexagonus	pfam00033 cytochrome_b_N	1E-38	NoORF
11	8	gi 5835683 gi 5835683 ref NP_008499.1 CO	6E-37	Ixodes hexagonus	pfam00116 COX2	5E-26	NoORF
12	8	gi 5835693 gi 5835693 ref NP_008509.1 ND	1E-50	Ixodes hexagonus	pfam00146 NADHdh	2E-52	NoORF
13	8	gi 7302028  CG1746 gene product	7E-17	Drosophila melanogaster	pfam00137 ATP-synt_C	8E-20	ANCH
14	7						ANCH
15	L						DISON
16	L				Smart smart00131 KU	4E-05	ANCH
17	9	gi 3881447  contains similarity to Pfam	0.018		Smart smart00131 KU	0.0001	NoORF
18	9				pfam02098 His_binding	0.0007	SIG
19	5	gi 3451024  cytochrome oxidase III	2E-25	Ixodes pacificus	pfam00510 COX3	1E-48	NoORF
20	5	gi 15428348  Salp10 [Ixodes s	9E-14	Ixodes scapularis			SIG
21	4	gi 7861934  Ke3 [Danio rerio]	2E-37	Danio rerio	pfam00416 Ribosomal_S13	6E-22	DISON
22	4	gi 8072217  Dscam [Drosophila	0.056	Drosophila melanogaster			SIG
23	4						SIG
24	б				Smart smart00131 KU	0.007	SIG
25	ŝ				pfam01943 Polysacc_synt	4E-08	NoORF
26	ŝ	gi 7506657 gi 7506657 pir  T32060	0.00001	Caenorhabditis elegans	Smart smart00131 KU	2E-09	NoORF
27	ŝ	gi 15428300  20 kDa salivary	3E-62	Ixodes scapularis			NoORF
28	ŝ				pfam01490 Aa_trans	0.01	NoORF
29	б				pfam01391 Collagen	0.0007	SIG
30	б				pfam00014 Kunitz_BPTI	0.009	SIG
<u>1</u>			-				
1000							

<sup>1</sup>BlastX performed with Blosum62 matrix and an E-value cutoff of 0.1 against the non-redundant (NR) protein database of NCBI. <sup>2</sup>Indicates significance of match to NR or CDD sequence of previous column.

<sup>3</sup>RPSblast performed against the conserved domains database (CDD) of NCBI containing all PFAM and Smart motifs. <sup>4</sup>Combined results from all sequences after submission of best open reading frame to the SignalP server (Ref\*\*\*). SIG, secretory signal sequence found; NoOrf, no suitable frame or start methionine was found; NoSIG, no secretory signal is found; ANCH, membrane anchor sequence found.

# 2846 J. G. Valenzuela and others

Table 1. Thirty most abundant cDNA clusters from an adult female Ixodes scapularis salivary gland library

mber         Cluster         Number of sequences         Best BharX match to NR darbase <sup>1</sup> Best RFSblast to and the mumber           1         44         gil 5425301 L Mo safivary 334         1         44         gil 5425301 L Mo safivary 335         200         20         200 </th <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th>									
1         4         gil (3478308)         14. Mas atilvary         700E-34         pfm002414 Borrelia_ord/3           3         1         gil (3478308)         15. Mas atilvary         0.00E-31         pfm00207 A2M           3         1         gil (3478308)         15. Mas atilvary         0.00E-31         pfm00207 A2M           3         1         gil (34723105 25. Mas atilvary         0.00E-31         pfm00207 A2M           273         3         gil (34703) gil (372105) gil (1841612         2.00E-22         pfm00307 A2M           377         1         gil (34703) gil (3712) gil (1841612         2.00E-10         pfm0031 cytatin           377         1         gil (347703) gil (3712) gil (184011         2.00E-10         pfm0031 cytatin           377         1         gil (347703) gil (37012) gil (184012)         2.00E-10         pfm0031 cytatin           377         1         gil (347703) gil (37012) gil (184012)         2.00E-10         pfm0031 cytatin           377         1         gil (347703) gil (37012) gil (184012)         2.00E-20         pfm0031 cytatin           377         1         gil (37763) gil (37012) gil (184012)         2.00E-20         pfm0031 cytatin           378         1         gil (37763) gil (39012) gil (49012)         2.00E-20<	e nher	Cluster	Number of	Rest BlactX match to NR database <sup>1</sup>	E-value <sup>2</sup>	Best RPSblast to CDD database <sup>3</sup>	E-value <sup>2</sup>	Comments	New 94
1         44         gil15428304 [15 kDa salivary         7.00E-34         pfann02414 Boretaort7           23         1         gil15428304 [15 kDa salivary         1.00E-31         pfann0207 A2M           23         gil15428304 [56 kDa salivary         0.00E-41         pfann0207 A2M           23         gil15428304 [56 kDa salivary         0.00E-41         pfann0207 A2M           23         gil15428304 [56 kDa salivary         0.00E-41         pfann0207 A2M           23         gil15421905[gil15705]pir[ML1612         2.00E-20         pfann0021 systain           237         gil15721905[gil15705]pir[ML1612         2.00E-10         pfann0021 systain           237         gil15775[gil187763]pir[MV1612         2.00E-20         pfann0051 systain           237         gil157783[gil87763]pir[MV1612         2.00E-20         pfann0051 systain           237         gil157783[gil87763]pir[MV1612         0.00E-11         pfann01421 Reprolysin           238         gil157108         ppotterical protein         2.00E-20         pfann00576 funexin           238         gil10720060 (G-INX         2.00E-30         pfann01421 Reprolysin         gil1147791 Reprolysin           238         gil107702051 was subspotterical protein         2.00E-30         pfann01421 Reprolysin         gil1547834			en anna de la companya de la compa						
329         1         gil 15428296 [6 kDa salivary         100E-31           9         9         gil 5428309 [6 kDa salivary         100E-31           283         1         gil 5428309 [6 kDa salivary         100E-31           291         19 [5428309 [5 kDa salivary         300E-62         pfil and 200207 A2M           273         1         gil 5428300 [5 kDa salivary         300E-62         pfil and 200207 A2M           273         1         gil 5428300 [20 kDa salivary         300E-62         pfil and 20120           274         1         gil 5428300 [20 kDa salivary         300E-62         pfil and 20120           274         1         gil 5428300 [20 kDa salivary         300E-62         pfil and 20120           277         1         gil 542830 [108] hypothetical protein         100E-11         pfil 00021 hysis           277         1         gil 5763 [af 1768] hypothetical protein         1.00E-11         pfil 01421 Repolysis           278         2         2         2         2         pfil and 212 Repolysis           278         1         3         3         3         3         3         3           279         1         3         3         3         3         3         <		1	44	gi 15428308  14 kDa salivary	7.00E-34	pfam02414 Borrelia_orfA	1.00E-09	14 kDa salivary protein	
364         1         gill 5428206 [6 kDa salivary         100E-105           29         9         9         [15428310] 25 kDa salivary         100E-105           291         1         [37521905]girl(75118544         100E-105         prim00207 A2M           291         1         [37521905]girl(75118544         400E-12         LOAD_1az tax.           291         1         [313470791]girl(3470791]reftNP_1102360.11         900E-12         LOAD_1az tax.           247         1         [313470791]girl(350077 probable         900E-13         LOAD_1az tax.           247         1         [315470791]girl(350077 probable         900E-13         LOAD_1az tax.           247         1         [315470791]sirl(50077         900E-13         Dimo031 cystatin           247         1         [3154170781]sirl(50077         00E-10         Dimo035 (statin           248         1         [315911708] hypothetical protein         LOAD_1az tax.         200E-35           248         1         [31671708] hypothetical protein         200E-35         fmm0035 (statin           243         1         [31671708] hypothetical protein         200E-35         fmm01421 Repolysin           244         1         [316717002] (statan         200E-35         fmm0207 k		329	1	gi 15428294  15 kDa salivary	4.00E-35			15 kDa salivary protein	
9         9         gill 5428301 0.25 ADa salivary         100E-105           231         1         gill 5428304 0.5 ADa salivary         9.00E-41         pilm00207 A2M           231         1         gill 5428304 0.5 ADa salivary         9.00E-32         pilm00231 59/161           231         1         gill 5428304 0.50 ADa salivary         9.00E-42         pilm00450 serile.carbpe           231         1         gill 547051 pie17012         2.00E-10         pilm00450 serile.carbpe           233         1         gill 547051 pie17081 hypothetical protein         2.00E-10         pilm0031 cystatin           247         1         gill 537751 pie17081 hypothetical protein         1.00E-105         pilm01421 Reprolysin           262         2         gil59117081 hypothetical protein         2.00E-10         pilm01421 Reprolysin           133         1         gil59117081 hypothetical protein         2.00E-10         pilm01421 Reprolysin           202         1         gil59117081 hypothetical protein         2.00E-40         pilm01421 Reprolysin           203         1         gil59117081 hypothetical protein         2.00E-10         pilm01421 Reprolysin           203         1         gil59117081 hypothetical protein         2.00E-10         pilm01421 Reprolysin <tr< td=""><td></td><td>364</td><td></td><td>gi 15428296  16 kDa salivary</td><td>1.00E-31</td><td></td><td></td><td>16 kDa salivary protein</td><td></td></tr<>		364		gi 15428296  16 kDa salivary	1.00E-31			16 kDa salivary protein	
283         1         ğil (5428304) (26 kDa salivary         9,00E-41         pim0207 A2M           291         3         gil (372190) (girl)(178344         4,00E-42         pim00207 A2M           273         3         1         gil (372190) (girl)(178344         4,00E-12         pim0050 serine_carbpe           274         1         gil (37012) (gil (3012) (gil (3707)) (gil (37763) (gil		6	6	gi 15428310  25 kDa salivary	1.00E-105			25 kDa antigen	
291         1         gi[7521905[gi]7521905[pir/f18544         4.00E-41         pfan00.007 A2M           373         1         gi[1347050]         103012[ji](301031[ris]NPL         3.00E-20         pian00.450 serine_cambre           373         1         gi[134705]         103012[ji](301031]         1000131         2.00E-10           377         1         gi[134705]         gi[134705]         1000131         2.00E-10           62         2         gi[3511768]         hypothetical protein         1.00E-35         fam00131         2.04D_Jaz atz.           103         1         gi[5511768]         hypothetical protein         1.00E-35         fam01421         Repolysin           65         2         gi[5511708]         hypothetical protein         1.00E-35         fam01421         Repolysin           134         1         gi[5511708]         hypothetical protein         2.00E-35         fam01421         Repolysin           134         1         gi[5511708]         hypothetical protein         2.00E-35         fam01421         Repolysin           134         1         gi[5511708]         hypothetical protein         2.00E-45         fam01421         Repolysin           135         1         gi[5911708]         hypothetical pr		283	1	gi 15428304  26 kDa salivary	9.00E-41			26 kDa protein A (new member?)	
27         3 $gll[3428300]$ 20 kDa salivay $3.00E-62$ pfan00450 serine_carbpe           373         1         gll(3470791 [gll]447012 $2.00E-12$ DOAD_taz taz           247         1         gll(3470791 [gll]447012 $2.00E-10$ $DOAD_{122}$ taz           247         1         gll(3470791 [gll]447012 $2.00E-10$ $DOAD_{122}$ taz           247         1         gll(3470781 [hypothetical protein $2.00E-10$ $DOAD_{122}$ taz           059         1         gl[59117081 [hypothetical protein $1.00E-16$ pfan0031 cystatin           133         1         gj[59117081 [hypothetical protein $2.00E-35$ pfan01421 Reprolysin           134         1         gj[59117081 [hypothetical protein $2.00E-35$ pfan001421 Reprolysin           292         1         gj[59117081 [hypothetical protein $2.00E-35$ pfan01421 Reprolysin           293         1         gj[59117081 [hypothetical protein $7.00E-16$ pfan00142 Reprolysin           293         1         gj[5917081 [hypothetical protein $7.00E-16$ pfan01421 Reprolysin           306         1         gj[107720660] (G-1NX $2.00E-16$ pfan01421 Reprolysin		291	1	gi/7521905 gi/7521905 bir  T18544	4.00E-41	pfam00207 A2M	3.00E-25	Alpha-2-macroglobulin	Y
379         1         gi[103012]gi[103012]pir][A41612         2.00E-20         pfam00450 serine_carbpe           377         1         gi[12323279] gi[134705]ipir]50027 probable         0.00E-18         0.00E-18           62         2         gi[131233279] defensib B [Omithodoros         1.00E-05         Sinart smart00505 Knotl           63         1         gi[5911708] hypothetical protein         0.00E-11         pfam0031 cystatin           63         2         gi[5911708] hypothetical protein         2.00E-35         pfam01421 Reprofysin           63         1         gi[5911708] hypothetical protein         2.00E-35         pfam01421 Reprofysin           63         1         gi[5911708] hypothetical protein         2.00E-35         pfam01421 Reprofysin           7         1         gi[5911708] hypothetical protein         2.00E-36         pfam01421 Reprofysin           7         1         gi[5911708] hypothetical protein         2.00E-34         pfam01421 Reprofysin           7         1         gi[5911708] hypothetical protein         2.00E-46         pfam01421 Reprofysin           7         2         2         2         2.00E-41         pfam01421 Reprofysin           7         2         2         2         2         2.00E-41		27	ŝ	gi 15428300  20 kDa salivary	3.00E-62	4		Anticomplement protein	
333       1       gi[1347079][gi[1347079][ref]hyP_102360.1]       9.00E-12       LOAD_Laz taz         247       1       gi[12532479] puative [Mus musculus]       1.00E-18       Smart smart00505 Knot1         65       2       gi[911708] hypothetical protein       1.00E-05       Smart smart00505 Knot1         134       1       gi[911708] hypothetical protein       2.00E-11       pfam00031 cystatin         134       1       gi[911708] hypothetical protein       2.00E-93       pfam01421 Reprolysin         135       1       gi[911708] hypothetical protein       2.00E-93       pfam01421 Reprolysin         136       1       10770060 (G-INX       2.00E-93       pfam0168 Occludin         306       1       gi[911708] hypothetical protein       2.00E-93       pfam0164 Curversin         306       1       gi[977002] ixolaris [Ixodes       2.00E-61       pfam02168 Occludin         306       1       gi[15077002] ixolaris [Ixodes       2.00E-64       pfam0014 Kuniz_BPT1         307       1       gi[15077002] ixolaris [Ixodes       2.00E-74       pfam0014 Kuniz_BPT1         307       1       gi[15077002] ixolaris [Ixodes s       2.00E-74       pfam0014 Kuniz_BPT1         308       1       gi[15077002] ixolaris [Ixodes s       <		379	1	gi 103012 gi 103012 pir  A41612	2.00E-22	pfam00450 serine carbpept	9.00E-26	Carboxypeptidase	Y
2471 $\tilde{g}[12832479]$ putative [Mus musculus]1.00E-185771 $g[87763]g[87763]gi[87763]pir[JS0027] probable2.00E-10pfam00031 cystatin5772g[3911708] hypothetical protein2.00E-10pfam0031 cystatin1341g[3911708] hypothetical protein2.00E-93pfam01421 Reprolysin1341g[3911708] hypothetical protein2.00E-93pfam01421 Reprolysin3831g[3911708] hypothetical protein2.00E-93pfam01421 Reprolysin3661g[10720060] (G-INX4.00E-16pfam02168 Occludin3631g[1072002] isolaris [Ixodes2.00E-64pfam02168 Occludin3631g[15077002] isolaris [Ixodes2.00E-49pfam02014 Kunit2. BPT13631g[15077002] isolaris [Ixodes2.00E-49pfam00014 Kunit2. BPT13631g[15077002] isolaris [Ixodes2.00E-49pfam00014 Kunit2. BPT13631g[15077002] isolaris [Ixodes2.00E-49pfam00014 Kunit2. BPT13641g[15077002] isolaris [Ixodes2.00E-48pfam00014 Kunit2. BPT13651g[15077002] isolaris [Ixodes2.00E-49pfam00013 Kuu3661g[15077002] isolaris [Ixodes2.00E-48pfam00013 Kuu3671g[15077002] isolaris [Ixodes2.00E-48pfam0013 Kuu3681g[15077002] isolaris [Ixodes2.00E-48pfam0013 Kuu371g[15077002] isolaris [Ixodes<$		333	1	gil13470791 gi 13470791 ref NP_102360.1	9.00E-12	LOAD_taz taz	0.002	Collagen-like	Υ
377         1         gi[87763]gi[87763]gi[87763]gi[87763]gi[87763]gi[87763]gi[87763]gi[87763]gi[87053]gi[87053]gi[87053]gi[87053]gi[87053]gi[87053795] defensin B [Omithodoros         1.00E-105         Smart smart00505 Knotl           103         2         gi[5911708] hypothetical protein         1.00E-05         Smart smart00505 Knotl           134         1         gi[5911708] hypothetical protein         1.00E-05         Smart smart00505 Knotl           134         1         gi[5911708] hypothetical protein         2.00E-35         gman01421 Reprolysin           135         1         gi[5911708] hypothetical protein         2.00E-35         gman01421 Reprolysin           205         1         gi[5911708] hypothetical protein         2.00E-46         gman0131 KU           203         1         gi[10720060] (G-INX         4.00E-16         pfam02168 Occludin           203         1         gi[10720060] (G-INX         4.00E-16         pfam02138 His_binding           203         1         gi[1077021 kolaris [Loodes         1.00E-04         pfam02138 His_binding           203         1         gi[15077021 kolaris [Loodes         3.00E-04         pfam0014 Kmirz_BPT1           203         1         gi[15478348] Sub10 [Loodes s         2.00E-34         pfam0013 Kuirz_BPT1           204         2		247	1	gi[12832479] putative [Mus musculus]	1.00E-18	I		Conserved protein	Υ
62         2         plan0031 cystatin           109         1         gil 5623795] defensin B (Ornithodoros         1.00E-01         Smart smart0550 Knott           134         1         gil 5911708  hypothetical protein         1.00E-111         pfam01421 Reprolysin           135         1         gil 5911708  hypothetical protein         2.00E-93         pfam01421 Reprolysin           163         1         gil 5911708  hypothetical protein         2.00E-111         pfam01421 Reprolysin           292         1         gil 5911708  hypothetical protein         2.00E-11         pfam0132168 Occludin           306         1         gil 50770021 kubinsi [Raodes         2.00E-40         pfam02108 His_binding           202         1         gil 50770021 kubinsi [Raodes         2.00E-41         pfam02038 His_binding           203         1         gil 50770021 kubinsi [Raodes         2.00E-45         pfam0014 Kuniz_BPT1           363         1         gil 50770021 kubinsi [Raodes         2.00E-46         pfam0013 Ku           363         1         gil 50770021 kubinsi [Raodes         2.00E-45         pfam00014 Kuniz_BPT1           363         1         gil 50770021 kubinsi [Raodes         2.00E-45         pfam00014 Kuniz_BPT1           363         1		377	1	gi 87763 gi 87763 pir  JS0027 probable	2.00E-10			Conserved protein	Y
109         1         gil 36.23795         defensin B         Omithodoros         1.00E-01         Smart smart0505 Knott           134         1         gil 911708         hypothetical protein         2.00E-33         pfan01421 Reprolysin           165         1         gil 911708         hypothetical protein         2.00E-33         pfan01421 Reprolysin           163         1         gil 911708         hypothetical protein         2.00E-33         pfan01421 Reprolysin           365         1         gil 910720060         (G-INX         2.00E-93         pfan01421 Reprolysin           306         1         gil 910720060         (G-INX         4.00E-16         pfan00142168           306         1         gil 1547022         ixotaris Ixodes         2.00E-40         pfan02088 His_binding           202         1         gil 5470378         jxodes         2.00E-44         pfan00014 Kunitz_BPT1           363         1         gil 5477302         jxotaris Ixodes         3.00E-34         pfan00014 Kunitz_BPT1           363         1         gil 5477302         jxotaris Ixodes         3.00E-34         pfan00014 Kunitz_BPT1           363         1         gil 5477328         jxodes         3.00E-34         pfan00014 Kunitz_BPT1		62	7	· · · · · · · · · · · · · · · · · · ·		pfam00031 cystatin	8.00E-12	Cystatin	Y
65         2         gil5911708 hypothetical protein gil5911708 hypothetical protein gil5911708 hypothetical protein gil5911708 hypothetical protein gil5911708 hypothetical protein gil5911708 hypothetical protein gil5911708 hypothetical protein gil507102 gil5470378 hypothetical protein 306         1.00E-11         pfam01421 Reprolysin pfam02168 Occludin pfam02098 His_binding gil15077002 hybothetical protein gil5577002 hybothetical protein gil55725 hybothetical protein gil55725 hybothetical protein gil55725 hybothetical protein gil55725 hybothetical protein gil550203 hybothetical protein gil550203 hybothetical protein gil550203 hybothetical protein gil550203 hybothetical protein gil550203 hybothetical gil550203 hybothetical		109	1	gi 13623795  defensin B [Ornithodoros	1.00E-05	Smart smart00505 Knot1	6.00E-07	Defensin – antibacterial peptide	Υ
1341gi[5911708] hypothetical protein2.00E-353831gi[5911708] hypothetical protein2.00E-351631gi[5911708] hypothetical protein2.00E-353061gi[10720060] (G-NX4.00E-163061gi[154128292] histamine bindim4.00E-163061gi[1547002] ixolaris [Ixodes2.00E-613631gi[1547022] ixolaris [Ixodes2.00E-643631gi[1547022] ixolaris [Ixodes2.00E-643631gi[15077002] ixolaris [Ixodes2.00E-643631gi[15077002] ixolaris [Ixodes2.00E-643631gi[15077002] ixolaris [Ixodes2.00E-643631gi[15077002] ixolaris [Ixodes2.00E-743631gi[15077002] ixolaris [Ixodes2.00E-643631gi[15478348] Sapl0 [Ixodes s2.00E-743631gi[15478348] Sapl0 [Ixodes s2.00E-74422gi[15428348] Sapl0 [Ixodes s2.00E-18432gi[15428348] Sapl0 [Ixodes s2.00E-184420gi[15428348] Sapl0 [Ixodes s2.00E-18452gi[15428348] Sapl0 [Ixodes s2.00E-18467gi[15428348] Sapl0 [Ixodes s2.00E-18476gi[15428348] Sapl0 [Ixodes s2.00E-14476gi[15428348] Sapl0 [Ixodes s2.00E-14482gi[1542836776775]1.00E-06492gi[17		65	2	gi 5911708  hypothetical protein	1.00E-111	pfam01421 Reprolysin	0.008	Disintegrin protease	Υ
383         1         gi[5911708  hypothetical protein         2.00E-93         pfam01421 Reprolysin           163         1         gi[5911708  hypothetical protein         7.00E-11         pfam02168 Occludin           202         1         gi[10720060] (G-INX         4.00E-16         pfam02168 Occludin           306         1         gi[10720060] (G-INX         4.00E-16         pfam02168 Occludin           306         1         gi[1547037819[4470378]80] 7421 [HBP2_RHI         1.00E-61         pfam02168 Occludin           202         1         gi[15077002] ixolaris [Ixodes         2.00E-61         pfam00208 His_binding           203         1         gi[15077002] ixolaris [Ixodes         2.00E-74         pfam0014 Kunitz_BPTI           303         1         gi[15077002] ixolaris [Ixodes         2.00E-44         pfam00014 Kunitz_BPTI           42         2         gi[15470378]83 H47] contains similarity to Pfam         3.00E-45         pfam00013 Kunitz_BPTI           7         14         gi[15470378]5706571pir[T530665         0.018         pfam00013 Kunitz_BPTI           7         14         gi[15470378]57065751pir[T732060         0.018         smart smart00131 KU           7         14         gi[15470378]677021         10010         smart smart00131 KU		134	1	gi 5911708  hypothetical protein	2.00E-35			Disintegrin protease	Y
163         1         gil3911708  hypothetical protein         7.00E-16         pfam00376 Innexin           292         1         gil10720060  (G-INX         4.00E-16         pfam02168 Occludin           306         1         gil10720060  (G-INX         4.00E-16         pfam02168 Occludin           305         1         gil15428392  histamine bindim         2.00E-61         pfam02098 His_binding           363         1         gil15077002  ixolaris [Ixodes         2.00E-64         pfam02098 His_binding           363         1         gil15077002  ixolaris [Ixodes         2.00E-64         pfam00131 KU           363         1         gil15077002  ixolaris [Ixodes         2.00E-74         pfam000131 KU           363         1         gil15077002  ixolaris [Ixodes         2.00E-74         pfam000131 KU           363         1         gil15077002  ixolaris [Ixodes         2.00E-34         pfam00014 Kunitz_BPTI           42         2         gil15428348  Salp10 [Ixodes s         2.00E-18         pfam000131 KU           7         14         gil1244810         0.01         Smart smart00131 KU           7         14         gil124810         0.01         Smart smart00131 KU           7         14         gil124810<		383	1	gi 5911708  hypothetical protein	2.00E-93	pfam01421 Reprolysin	2.00E-04	Disintegrin protease	Y
292       1       gil107200601 (G-INX       4.00E-16       pfan00876 Innexin         306       1       gil107200601 (G-INX       4.00E-16       pfan002168 Occludin         306       1       gil154282921 histamine bindin       2.00E-61       pfan02098 His_binding         202       1       gil15470021 ixolaris [Ixodes       2.00E-61       pfan02098 His_binding         303       1       gil150770021 ixolaris [Ixodes       2.00E-45       pfan00014 Kuniz_BPTI         303       1       gil150770021 ixolaris [Ixodes       2.00E-45       pfan00014 Kuniz_BPTI         168       1       gil150770021 ixolaris [Ixodes       2.00E-45       pfan00014 Kuniz_BPTI         168       1       gil150770021 ixolaris [Ixodes       2.00E-45       pfan00014 Kuniz_BPTI         168       1       gil150770021 ixolaris [Ixodes s       2.00E-45       pfan00014 Kuniz_BPTI         168       1       gil150770021 ixolaris [Ixodes s       2.00E-45       pfan00014 Kuniz_BPTI         168       1       gil150770021 ixolaris [Ixodes s       2.00E-18       pfan00014 Kuniz_BPTI         168       7       14       gil150770021 ixolaris [Ixodes s       2.00E-18       pfan00131 KU         17       14       gil15428348       S		163	1	gi 5911708  hypothetical protein	7.00E-11			Disintegrin protease?	Y
306         1         5         pfan02168 Occludin pfan02168 Occludin           18         6         gill5428292] histamine bindin         pfan02168 Occludin           202         1         gill5428292] histamine bindin         2.00E-61         pfan02058 His_binding           363         1         gill577002] ixolaris [Ixodes         2.00E-61         pfan02098 His_binding           363         1         gill5077002] ixolaris [Ixodes         2.00E-44         pfan0014 Kunitz_BPT1           168         1         gill5077002] ixolaris [Ixodes         2.00E-74         pfan0014 Kunitz_BPT1           168         1         gill5077002] ixolaris [Ixodes         3.00E-45         Smart smart00131 KU           17         14         gill5428348  Salp10 [Ixodes s         2.00E-34         pfan00014 Kunitz_BPT1           7         14         gill24810]         3.00E-45         Smart smart00131 KU           7         14         gill2479700[ir71720219         0.01         Smart smart00131 KU           7         14         gill24810]         0.01         Smart smart00131 KU           7         14         gill24810]         0.01         Smart smart00131 KU           7         14         gill24810]         1.0026657         Smart sm		292	-1	gi 10720060  (G-INX	4.00E-16	pfam00876 Innexin	2.00E-16	Gap junction protein	Y
43       2       pfan02168 Occludin         18       6       pian02168 Occludin         18       6       pian02168 Occludin         202       1       gill5470378[sp]077421 [HBP2_RHI       100E-04         363       1       gill50770021 ixolaris [Ixodes       2.00E-61         389       1       gill50770021 ixolaris [Ixodes       2.00E-74       pfam02098 His_binding         133       1       gill50770021 ixolaris [Ixodes       2.00E-74       pfam0014 Kunitz_BPTI         168       1       gill50770021 ixolaris [Ixodes       3.00E-34       pfam00014 Kunitz_BPTI         168       1       gill54810       2.00E-74       pfam00014 Kunitz_BPTI         168       7       2.00E-34       pfam0014 Kunitz_BPTI         168       7       2.00E-18       pfam0014 Kunitz_BPTI         166       7       3.00E-34       pfam0014 Kunitz_BPTI         166       7       2.00E-18       pfam0014 Kunitz_BPTI         166       7       14       2.00E-18       pfam0014 Kunitz_BPTI         167       7       14       2.00E-18       pfam0014 Kunitz_BPTI         16       7       2       gill5481047       0.018       Smart smart00131 KU		306	1			pfam02168 Occludin	1.00E-06	Gap junction protein	Υ
18         6         pfan02098 His_binding           202         1         gil15428292  histamine bindim         2.00E-61         pfan02098 His_binding           363         1         gil8470378[gil8470378]gil8470378[gil8470378]gil8470378]gil8470378[gil647042]         1.00E-04         pfan02098 His_binding           363         1         gil15077002  ixolaris [Ixodes         2.00E-74         pfan00014 Kunitz_BPTI           168         1         gil1547038]spl07421 HBP2_RHI         1.00E-40         Smart smart00131 KU           168         1         gil1547002  ixolaris [Ixodes         2.00E-74         pfan00014 Kunitz_BPTI           168         1         gil1547002  ixolaris [Ixodes         3.00E-45         Smart smart00131 KU           168         7         14         20         gil15428348] Salp10 [Ixodes s         0.01         Smart smart00131 KU           16         7         14         gil124810         0.01         Smart smart00131 KU           16         7         0         01         Smart smart00131 KU           17         6         gil3581447]         contains similarity to Pfam         0.01         Smart smart00131 KU           17         6         gil7497910[gil7497910[gir[f123573         0.01         Smart smart00131 KU <td< td=""><td></td><td>43</td><td>2</td><td></td><td></td><td>pfam02168 Occludin</td><td>0.005</td><td>Gap junction protein?</td><td>Υ</td></td<>		43	2			pfam02168 Occludin	0.005	Gap junction protein?	Υ
202       1       gil15428292  histamine bindim       2.00E-61         363       1       gil8470378[sp]077421 HBP2_RHI       1.00E-04       pfam02098 His_binding         89       1       gil15077002  ixolaris [Lxodes       1.00E-04       pfam02098 His_binding         89       1       gil15077002  ixolaris [Lxodes       2.00E-74       pfam00014 Kunitz_BPT1         168       1       gil15077002  ixolaris [Lxodes       2.00E-74       pfam00014 Kunitz_BPT1         168       1       gil15473348] Salp10 [Lxodes       2.00E-74       pfam00014 Kunitz_BPT1         7       14       gil15428348] Salp10 [Lxodes s       2.00E-18       pfam00014 Kunitz_BPT1         7       14       gil124810        0.01       Smart smart00131 KU         7       14       gil124810        2.00E-16       Smart smart00131 KU         7       14       gil124810        0.01       Smart smart00131 KU         7       14       gil75005657[gil7497910][gil		18	9			pfam02098 His_binding	7.00E-04	Histamine binding domain	Y
363         1         gi[8470378]gi[8470378]sp](077421]HBP2_RHI         1.00E-40         pfann02098 His_binding           89         1         gi[15077002] ixolaris [Ixodes         1.00E-40         Smart smart00131 KU           133         1         gi[15077002] ixolaris [Ixodes         2.00E-74         pfann00014 Kunitz_BPT1           168         1         gi[15077002] ixolaris [Ixodes         2.00E-74         pfann00014 Kunitz_BPT1           168         1         gi[15428348] Salp10 [Ixodes s         2.00E-74         pfann00014 Kunitz_BPT1           42         2         gi[15428348] Salp10 [Ixodes s         2.00E-34         pfann0014 Kunitz_BPT1           7         14         gi[15428348] Salp10 [Ixodes s         0.01         Smart smart00131 KU           7         14         gi[124810]         0.01         Smart smart00131 KU           7         14         gi[7497910]gi[7497910]gir[[T32060         1.00E-05         Smart smart00131 KU           26         3         gi[770505657]gir[7505725]pir[[T23573         0.018         Smart smart00131 KU           7         148         2         gi[77020203]gir[7497910]gir[[T23573         0.016-05         Smart smart00131 KU           26         3         gir[7497910]gir[7497910]gir[[T23573         0.00E-05 <td< td=""><td></td><td>202</td><td>1</td><td>gi 15428292  histamine bindin</td><td>2.00E-61</td><td></td><td></td><td>Histamine binding protein</td><td></td></td<>		202	1	gi 15428292  histamine bindin	2.00E-61			Histamine binding protein	
89       1       gil15077002  ixolaris [Ixodes       1.00E-40       Smart smart00131 KU         133       1       gil15077002  ixolaris [Ixodes       2.00E-74       pfam00014 Kunitz_BPTI         168       1       gil15077002  ixolaris [Ixodes       3.00E-45       Smart smart00131 KU         4       20       gil15428348  Salp10 [Ixodes s       3.00E-45       Smart smart00131 KU         7       14       gil15428348  Salp10 [Ixodes s       5.00E-34       pfam00014 Kunitz_BPTI         7       14       gil124810        3.00E-45       Smart smart00131 KU         16       7       gil124810        0.01       Smart smart00131 KU         17       6       gil3881447  contains similarity to Pfam       0.01       Smart smart00131 KU         24       3       gil7497910[gil7497910]pir[[T20219       0.01       Smart smart00131 KU         26       3       gil7497910[gil7497910]pir[[T20219       7.00E-05       Smart smart00131 KU         251       2       gil7605725[gil7505725[gil7505725]pir[[T20219       7.00E-05       Smart smart00131 KU         27       2       gil779725[gil7500203]pir[[T16210       0.003       Smart smart00131 KU         28       2       gil7500203[gil7500203]pir[[T16210       0.005       pfa		363	1	gi 8470378 gi 8470378 sp 077421 HBP2_RHI	1.00E-04	pfam02098 His_binding	0.003	Histamine binding protein	Y
133       1       gi[15077002] ixolaris [Ixodes       2.00E-74       pfam00014 Kunitz_BPTI         168       1       gi[15077002] ixolaris [Ixodes       3.00E-45       Smart smart00131 KU         4       20       gi[15428348] Salp10 [Ixodes s       3.00E-45       Smart smart00131 KU         7       14       gi[15428348] Salp10 [Ixodes s       5.00E-18       pfam00014 Kunitz_BPTI         7       14       gi[124810]       0.01       Smart smart00131 KU         16       7       3.00E-45       Smart smart00131 KU         16       7       gi[3881447] contains similarity to Pfam       0.01       Smart smart00131 KU         24       3       gi[7497910]gi[7497910]piri[T20219       0.018       Smart smart00131 KU         26       3       gi[7497910]gi[7497910]piri[T20219       0.018       Smart smart00131 KU         26       3       gi[7497910]gi[7497210]       0.018       Smart smart00131 KU         26       3       gi[7497210]gi[7497219]       0.018       Smart smart00131 KU         27       28       gi[7497210]gi[7497219]       0.018       Smart smart00131 KU         26       3       gi[7497210]gi[7497219]       0.018       Smart smart00131 KU         27       28       gi[		89	1	gi 15077002  ixolaris [Ixodes	1.00E-40	Smart smart00131 KU	9.00E-13	Kunitz – Ixolaris	
168         1         gil15077002         ixolaris [Ixodes         3.00E-45         Smart smart00131 KU           4         20         gil15428348         Salp10 [Ixodes s         5.00E-34         pfam00014 Kunitz_BPTI           7         14         gil15428348         Salp10 [Ixodes s         5.00E-34         pfam00014 Kunitz_BPTI           7         14         gil15428348         Salp10 [Ixodes s         0.01         Smart smart00131 KU           16         7         gil154810          2.00E-18         pfam00014 Kunitz_BPTI           17         6         gil124810          0.01         Smart smart00131 KU           17         6         gil3881447  contains similarity to Pfam         0.018         Smart smart00131 KU           24         3         gil7497910[gil7497910]pir[[T20219         0.018         Smart smart00131 KU           26         3         gil7497910[gil7497910]pir[[T20219         7.00E-05         Smart smart00131 KU           45         2         gil7497910[gil7497910]pir[[T20219         7.00E-05         Smart smart00131 KU           47         2         gil7497910[gil7497910]pir[[T20219         7.00E-05         Smart smart00131 KU           48         2         gil770758277         G18436 gene product         0.003<		133	1	gi 15077002  ixolaris [Ixodes	2.00E-74	pfam00014 Kunitz_BPTI	4.00E-11	Kunitz – Ixolaris	
4       20       gil15428348  Salp10 [Ixodes s       5.00E-34       pfam00014 Kunitz_BPTI         7       14       gil15428348  Salp10 [Ixodes s       2.00E-18       pfam00014 Kunitz_BPTI         7       14       gil15428348  Salp10 [Ixodes s       2.00E-18       pfam00014 Kunitz_BPTI         16       7       14       gil124810        0.01       Smart smart00131 KU         17       6       gil3881447  contains similarity to Pfam       0.01       Smart smart00131 KU         17       6       gil7506657 gil750657 pir [T32060       0.018       Smart smart00131 KU         26       3       gil7497910 gil7497910 pir [T20219       0.016-05       Smart smart00131 KU         21       3       gil7497910 gil7497910 pir [T20219       1.00E-05       Smart smart00131 KU         45       2       gil7497910 gil7497910 gir [T120219       7.00E-06       Smart smart00131 KU         48       2       gil70726827  CG18436 gene product       0.003       Smart smart00131 KU         54       2       gil10726827  CG18436 gene product       0.003       Smart smart00131 KU         54       2       gil7500203 gir [T16210       0.066       pfan00014 Kunitz_BPTI         68       2       gil10726827  CG18436 gene product       3.00E-07<		168	1	gi 15077002  ixolaris [Ixodes	3.00E-45	Smart smart00131 KU	3.00E-07	Kunitz – Ixolaris	
42       2       gi[15428348] Salp10 [Ixodes s       2.00E-18       pfam00014 Kuniz_BPTI         7       14       gi[124810]       0.01       Smart smart00131 KU         16       7       8i[124810]       0.01       Smart smart00131 KU         17       6       gi[3881447] contains similarity to Pfam       0.01       Smart smart00131 KU         17       6       gi[3881447] contains similarity to Pfam       0.018       Smart smart00131 KU         24       3       gi[7506657]gi[750657]gi[750657]pir][T32060       1.00E-05       Smart smart00131 KU         26       3       gi[7497910]gi[7497910]pir][T20219       0.016       Smart smart00131 KU         26       3       gi[7505725]gi[7505725]gir][T20219       7.00E-06       Smart smart00131 KU         48       2       gi[70726827] CG18436 gene product       0.003       Smart smart00131 KU         51       2       gi[10726827] CG18436 gene product       0.003       Smart smart00131 KU         54       2       gi[10726827] CG18436 gene product       0.003       Smart smart00131 KU         54       2       gi[7500203]gir][7500203]gir][T16210       0.066       pfan00014 Kunitz_BPTI         53       1       gi[7500203]gir][760203]gir][760203]gir][716210       0.006 <t< td=""><td></td><td>4</td><td>20</td><td>gi 15428348  Salp10 [Ixodes s</td><td>5.00E-34</td><td>pfam00014 Kunitz_BPTI</td><td>0.001</td><td>Kunitz - Salp10</td><td></td></t<>		4	20	gi 15428348  Salp10 [Ixodes s	5.00E-34	pfam00014 Kunitz_BPTI	0.001	Kunitz - Salp10	
7       14       gil124810        0.01       Smart smart00131 KU         16       7       e       gil3881447  contains similarity to Pfam       0.01       Smart smart00131 KU         17       6       gil3881447  contains similarity to Pfam       0.018       Smart smart00131 KU         24       3       gil75066571gil7506577]pir[T32060       1.00E-05       Smart smart00131 KU         26       3       gil7497910[gil7497910[pir][T20219       0.01B       Smart smart00131 KU         31       3       gil7497910[gil7497910[pir][T20219       1.00E-05       Smart smart00131 KU         45       2       gil7505725[gil7505725[pir][T23573       4.00E-05       pfam00014 Kunitz_BPTI         48       2       gil107268271 CG18436 gene product       0.003       Smart smart00131 KU         54       2       gil107268271 CG18436 gene product       0.003       Smart smart00131 KU         68       2       gil107268271 CG18436 gene product       3.00E-07       Smart smart00131 KU         73       1       gil7500203[gir][T16210       9.00E-07       Smart smart00131 KU         92       1       gil7500203[gir][T16210       9.00E-07       Smart smart00131 KU         92       1       gil7500203[gir][T6210       9.00E-07       Smart s		42	7	gi 15428348  Salp10 [Ixodes s	2.00E-18	pfam00014 Kunitz_BPTI	6.00E-05	Kunitz – Salp10	
16       7         17       6       gi[3881447] contains similarity to Pfam       0.018       Smart smart00131 KU         24       3       gi[7506657]gi[7506657]pir[[T32060       1.00E-05       Smart smart00131 KU         26       3       gi[750657]gi[750657]pir[[T32060       1.00E-05       Smart smart00131 KU         26       3       gi[7497910]gi[7497910]pir[[T20219       0.016-05       Smart smart00131 KU         21       3       gi[7505725]gi[7505725]pir[[T20219       7.00E-06       Smart smart00131 KU         48       2       gi[7505725]gi[7505725]pir[[T23573       4.00E-05       pfam00014 Kunitz_BPTI         48       2       gi[10726827] CG18436 gene product       0.003       Smart smart00131 KU         54       2       gi[10726827] CG18436 gene product       0.003       Smart smart00131 KU         54       2       gi[10726827] CG18436 gene product       0.006       pfam00014 Kunitz_BPTI         68       2       gi[10726827] CG18436 gene product       3.00E-07       Smart smart00131 KU         73       1       gi[7500203]gir[[T16210       9.00E-07       Smart smart00131 KU         92       1       gi[7500203]gir[[T16210       9.00E-07       Smart smart00131 KU         92       1       gi[75002		7	14	gi 124810	0.01	Smart smart00131 KU	1.00E-06	Kunitz domain	Y
17       6       gi[3881447] contains similarity to Pfam       0.018       Smart smart00131 KU         24       3       gi[7506657]gi[7506557]pir][T32060       Smart smart00131 KU         26       3       gi[7506657]gi[7506557]pir][T32060       1.00E-05       Smart smart00131 KU         31       3       gi[7505725]gi[7505725]pir][T20219       7.00E-05       Smart smart00131 KU         45       2       gi[7505725]gi[7505725]pir][T20219       7.00E-05       Smart smart00131 KU         48       2       gi[10726827] CG18436 gene product       0.003       Smart smart00131 KU         54       2       gi[10726827] CG18436 gene product       0.003       Smart smart00131 KU         54       2       gi[10726827] CG18436 gene product       0.0066       pfam00014 Kunitz_BPTI         68       2       gi[10726827] CG18436 gene product       3.00E-07       Smart smart00131 KU         73       1       gi[7500203]gir][T16210       9.00E-05       pfam00014 Kunitz_BPTI         92       1       gi[7500203]gir][T16210       9.00E-05       pfam000131 KU         92       1       gi[7500203]gir][T16210       9.00E-05       pfam00014 Kunitz_BPTI		16	7			Smart smart00131 KU	4.00E-05	Kunitz domain	Y
24       3       gi[7506657]gi[7506657]pir][T32060       Smart smart00131 KU         26       3       gi[7506657]gi[7506657]pir][T32060       1.00E-05       Smart smart00131 KU         31       3       gi[7497910]gi[7497910]pir][T20219       7.00E-05       Smart smart00131 KU         45       2       gi[7505725]gi[7505725]gi[7505725]pir][T20219       7.00E-05       Smart smart00131 KU         48       2       gi[10726827] CG18436 gene product       0.003       Smart smart00131 KU         54       2       gi[10726827] CG18436 gene product       0.003       Smart smart00131 KU         54       2       gi[7500203]gi[7500203]pir][T16210       0.066       pfam00014 Kunitz_BPTI         68       2       gi[10726827] CG18436 gene product       3.00E-07       Smart smart00131 KU         73       1       gi[7500203]gir][T16210       9.00E-07       Smart smart00131 KU         92       1       gi[7500203]gir][T16210       9.00E-05       pfam00014 Kunitz_BPTI		17	9	gi 3881447  contains similarity to Pfam	0.018	Smart smart00131 KU	1.00E-04	Kunitz domain	Y
26       3       gi[7506657]gi[7506657]pir][T32060       1.00E-05       Smart smart00131 KU         31       3       gi[7497910]gi[7497910]pir][T20219       7.00E-06       Smart smart00131 KU         45       2       gi[7505725]gi[7505725]pir][T20219       7.00E-06       Smart smart00131 KU         48       2       gi[10726827] CG18436 gene product       0.003       Smart smart00131 KU         54       2       gi[10726827] CG18436 gene product       0.003       Smart smart00131 KU         54       2       gi[10726827] CG18436 gene product       0.006       pfam00014 Kunitz_BPTI         68       2       gi[10726827] CG18436 gene product       3.00E-07       Smart smart00131 KU         73       1       gi[7500203]gir][T16210       9.00E-07       Smart smart00131 KU         92       1       gi[7500203]gir][T16210       9.00E-05       pfam00014 Kunitz_BPTI		24	3			Smart smart00131 KU	0.007	Kunitz domain	Y
31       3       gi[7497910]gi[7497910]pir [T20219       7.00E-06       Smart smart00131 KU         45       2       gi[7505725]gi[7505725]pir [T23573       9.00E-05       pfam00014 Kunitz_BPTI         48       2       gi[10726827] CG18436 gene product       0.003       Smart smart00131 KU         51       2       gi[10726827] CG18436 gene product       0.003       Smart smart00131 KU         54       2       gi[7500203]gi[7500203]pir [T16210       0.006       pfam00014 Kunitz_BPTI         68       2       gi[10726827] CG18436 gene product       3.00E-07       Smart smart00131 KU         73       1       gi[7500203]gir][7500203]pir [T16210       9.00E-05       pfam00014 Kunitz_BPTI         92       1       gi[7500203]gir][760203]pir [T16210       9.00E-05       pfam00014 Kunitz_BPTI		26	ю	gi 7506657 gi 7506657 pir  T32060	1.00E-05	Smart smart00131 KU	2.00E-09	Kunitz domain	Y
45       2       pfam00014 Kunitz_BPTI         48       2       gi[7505725]gi[7505725]pir][T23573       4.00E-05       pfam00014 Kunitz_BPTI         51       2       gi[10726827] CG18436 gene product       0.003       Smart smart00131 KU         54       2       gi[7500203]gi[7500203]pir][T16210       0.006       pfam00014 Kunitz_BPTI         68       2       gi[10726827] CG18436 gene product       3.00E-07       Smart smart00131 KU         73       1       gi[7500203]gi[7500203]pir][T16210       9.00E-07       Smart smart00131 KU         92       1       gi[7500203]gir][T16210       9.00E-05       pfam00014 Kunitz_BPTI		31	б	gi 7497910 gi 7497910 pir  T20219	7.00E-06	Smart smart00131 KU	4.00E-08	Kunitz domain	Y
48         2         gi[7505725]gi[7505725]gi[7505725]pir][T23573         4.00E-05         pfam00014 Kunitz_BPTI           51         2         gi[10726827] CG18436 gene product         0.003         Smart smart00131 KU           54         2         gi[7500203]gi[7500203]pir][T16210         0.066         pfam00014 Kunitz_BPTI           68         2         gi[10726827] CG18436 gene product         3.00E-07         Smart smart00131 KU           73         1         gi[7500203]gi[7500203]pir][T16210         9.00E-05         pfam00014 Kunitz_BPTI           92         1         gi[7500203]gi[7500203]pir][T16210         9.00E-05         pfam00014 Kunitz_BPTI		45	2			pfam00014 Kunitz_BPTI	6.00E-06	Kunitz domain	Υ
51         2         gil10726827  CG18436 gene product         0.003         Smart smart00131 KU           54         2         gi[7500203]gi[7500203]pir [T16210         0.066         pfam00014 Kunitz_BPTI           68         2         gi[10726827] CG18436 gene product         3.00E-07         Smart smart00131 KU           73         1         gi[7500203]gi[7500203]pir [T16210         9.00E-05         pfam00014 Kunitz_BPTI           92         1         Smart smart00131 KU         Smart smart00131 KU		48	2	gi 7505725 gi 7505725 pir  T23573	4.00E-05	pfam00014 Kunitz_BPTI	2.00E-08	Kunitz domain	Υ
54         2         gi[7500203]gi[7500203]pir][T16210         0.066         pfam00014 Kunitz_BPTI           68         2         gi[10726827] CG18436 gene product         3.00E-07         Smart smart00131 KU           73         1         gi[7500203]gi[7500203]pir][T16210         9.00E-05         pfam00014 Kunitz_BPTI           92         1         Smart smart00131 KU         Smart smart00131 KU		51	2	gi 10726827  CG18436 gene product	0.003	Smart smart00131 KU	7.00E-08	Kunitz domain	Υ
68         2         gi[10726827] CG18436 gene product         3.00E-07         Smart smart00131 KU           73         1         gi[7500203]gi[7500203]pir [T16210         9.00E-05         pfam00014 Kunitz_BPTI           92         1         Smart smart00131 KU         Smart smart00131 KU		54	7	gi 7500203 gi 7500203 pir  T16210	0.066	pfam00014 Kunitz_BPTI	9.00E-05	Kunitz domain	Υ
73 1 gi 7500203 gi 7500203 pir  T16210 9.00E-05 pfam00014 Kunitz_BPTI 92 1 Smart smart00131 KU		68	7	gi 10726827  CG18436 gene product	3.00E-07	Smart smart00131 KU	3.00E-04	Kunitz domain	Y
92 1 Smart 5mart 00131 KU		73	1	gi 7500203 gi 7500203 pir  T16210	9.00E-05	pfam00014 Kunitz_BPTI	2.00E-10	Kunitz domain	Y
		92	1			Smart smart00131 KU	2.00E-08	Kunitz domain	Υ

				Table 2. Co	ontinued			
Line number	Cluster number	Number of sequences	Best BlastX match to NR database <sup>1</sup>	E-value <sup>2</sup>	Best RPSblast to CDD database <sup>3</sup>	E-value <sup>2</sup>	Comments N	ew? <sup>4</sup>
42	108	-	gi 7498832 gi 7498832 pir  T34212	0.014	Smart smart00131 KU	7.00E-06	Kunitz domain	Y
43	124	1			Smart smart00131 KU	4.00E-05	Kunitz domain	Y
44	157	1	gi 4502167  precursor protein	2.00E-07	Smart smart00131 KU	4.00E-12	Kunitz domain	Y
45	176	1	4 4 -		Smart smart00131 KU	0.005	Kunitz domain	Y
46	264	1	gi 10726827  CG18436 gene product	6.00E-06	Smart smart00131 KU	6.00E-06	Kunitz domain	Y
47	290	1	gil6164595  lacunin [Manduca	5.00E-05	Smart smart00131 KU	0.005	Kunitz domain	Y
48	318	1	gi 400070 gi 400070 sp P31713 ISH1_STOHE	1.00E-14	Smart smart00131 KU	5.00E-22	Kunitz domain	Y
49	323	1	gil6164595  lacunin [Manduca	4.00E-12	Smart smart00131 KU	7.00E-05	Kunitz domain	Y
50	344	1	gi[7324126  Hypothetical protein	0.003	Smart smart00131 KU	7.00E-05	Kunitz domain	Y
51	30	ю			pfam00014 Kunitz_BPTI	0.009	Kunitz domain?	Y
52	272	1	gi 6651241  TAGL-beta [Mus mu	3.00E-13	Smart smart00644 Ami_2	2.00E-10	Peptidoglycan recognition protein	Y
53	248	1	gi 3282590  peritrophin 1 [Anopheles	3.00E-08	pfam01607 Chitin_bind_2	2.00E-13	Peritrophin	Y
54	145	1	gi 12002008  salivary gland 1	3.00E-21	1		Salivary 16 kD protein	
55	175	1	gi 12002008  salivary gland 1	5.00E-11	pfam02853 ACR	1.00E-04	Salivary 16 kD protein – new member	Y
56	178	1	gi 12002008  salivary gland 1	3.00E-06	pfam00335 transmembrane4	0.003	Salivary 16 kD protein – new member	Y
57	231	1	gi 15428348  Salp10 [Ixodes s	8.00E-20			Salp10	
58	20	5	gi 15428348  Salp10 [Ixodes s	9.00E-14			Salp10 – new member	
59	84	1	gi 15428348  Salp10 [Ixodes s	6.00E-11			Salp10 – new member	Y
60	274	1	gi 12018322 gi 12018322 ref NP_072152.1	2.00E-09	Smart smart00020 Tryp_SPc	7.00E-19	Serine protease	Y
61	194	1	gi 14140097  hypothetical protein	4.00E-51	Smart smart00093 SERPIN	9.00E-19	Serpin	Y
62	360	1	gi 862467  limulus intracellular	3.00E-11	pfam00079 serpin	2.00E-18	Serpin	Y
63	0	41			pfam02326 YMF19	5.00E-04	Short protein	Y
64	5	18			pfam01028 Topoisomerase_I	3.00E-04	Short protein	Y
65	14	7					Short protein	Y
99	15	7					Short protein	Y
67	44	2					Short protein	Y
68	36	7					Short protein?	Y
69	53	7	$gi 11496688 gi 11496688 ref NP_045470.1 $	0.001	pfam02118 Srg	2.00E-04	Similarity to Borrelia protein	Y
70	29	ω			pfam01391 Collagen	7.00E-04	Small collagen	Y
71	33	m	gi 15428290  thrombospondin [	3.00E-28	pfam02853 ACR	4.00E-05	Thrombospondin	
72	233	1	gi 15428290  thrombospondin [	8.00E-30	pfam02853 ACR	9.00E-04	Thrombospondin	
73	235	1	gi 15428290  thrombospondin [	1.00E-37			Thrombospondin	
74	324	1	gi 15428290  thrombospondin [	5.00E-26	pfam02853 ACR	2.00E-04	Thrombospondin	
75	392	1	gi 15428290  thrombospondin [	1.00E-10			Thrombospondin (new member?)	
76	279	1	· · ·		pfam01826 TIL	6.00E-06	Trypsin inhibitor cys rich domain	Y
LL	188	1	gi 14780055	4.00E-07	Smart smart00020 Tryp_SPc	2.00E-08	Trypsin-like protease	Y
78	22	4	gi 8072217  Dscam [Drosophila	0.056			4	Y
79	23	4						Y
80	39	7						Y
81	58	7						Y
82	61	2						Y

				Table 2. C	ontinued			
Line	Cluster	Number of			Best RPSblast to			
number	number	sednences	Best BlastX match to NR database <sup>1</sup>	E-value <sup>2</sup>	CDD database <sup>3</sup>	E-value <sup>2</sup>	Comments	New? <sup>4</sup>
83	67	2						Υ
84	95	1						Υ
85	102	1						Y
86	105	1						Y
87	122	1						Y
88	123	1						Υ
89	130	1						Υ
90	137	1						Y
91	212	1						Υ
92	229	1						Y
93	258	1						Y
94	261	1						Υ
95	277	1						Υ
96	296	1						Υ
97	353	1						Υ
98	359	1						Υ
66	375	1						Υ
100	389	1		_	pfam01028 Topoisomerase_I	8.00E-07		Υ
<sup>1</sup> Blast	X performe	vd with Blosum6	52 matrix and an E-value cutoff of 0.1 aga	inst the non-redu	ndant (NR) protein database of N	NCBI.		

# Sialome of the tick I. scapularis 2849

and 28 clusters having the Kunitz domain found in soybean trypsin inhibitor. Two serpins were also found, one of which matches a previously reported *I. ricinus* sequence. One additional cluster has the SMART TIL signature of trypsin inhibitors. Possible inhibitors of platelet aggregation include disintegrins (four clusters) and thrombospondin (five clusters). Three clusters code for proteins having similarity to tick histamine-binding proteins, one of which has been already described in *I. scapularis*.

A sequence matching the antimicrobial defensin was found, but this clone is truncated and does not have the distal 5' end of the starting methionine. Proteins or peptides with similarity to collagen or gap junction proteins are also represented, but their function is unknown. A serine carboxypeptidase, two serine proteases and metalloproteases appear to be secreted. More than 35 clusters are associated with proteins that are possibly secreted, but their function in tick feeding is not readily apparent. Also evident from Table 2 is the existence of several related proteins. Indeed, when the clustering of the database is done with a cutoff value of 1E-20 rather than 1E-60, several of these clusters collapse (for example, those labeled short proteins or those containing Kunitz domains), although the alignments indicate that these are composed of several different, but related, gene products (results not shown; see below).

Table 2A, available on request from the author (e-mail: jribeiro@nih.gov), contains information on clusters of sequences probably associated with housekeeping function. Three of these clusters, each containing only one sequence, all code for proteins of the 5'-nucleotidase family, a family previously associated with secreted salivary apyrase of mosquitoes. Of interest were also the finding of a sulfotransferase and an alkyl hydroperoxide reductase that could be linked to synthesis of sulfated products of secretion and salivary prostanoids, respectively.

RPSblast performed against the conserved domains database (CDD) of NCBI containing all PFAM and Smart motifs

<sup>2</sup>Indicates significance of match to NR or CDD sequence of previous column.

<sup>4</sup>Y, the sequence is a newly found *Ixodes scapularis* sequence.

# Full-length sequence information on 87 clones

To obtain more information on this transcriptome collection, with emphasis on the messages possibly associated with secreted proteins (the sialome set), we obtained full-length sequence of 87 clones, the properties of which are summarized in Table 3. 62 of these sequences belong to seven distinct groups, obtained by comparing the sequences against themselves using the BlastP program with a cutoff value of 1E–20 (see Materials and methods for more detail).

glands
salivary g
pularis .
odes sca
y <i>of</i> Ixc
A librar
$\geq$
ã
rom a ci
clones fi
ıgth c
ler
-III
7 fi
ŝ
6
zation
<b>Tharacteri</b> :
Table 3

Group 1       gil15428308 AF209921_1 (A         TB103       gil15428308 AF209921_1 (A         ISL1129       gil15428308 AF209921_1 (A         ISL1129       gil15428308 AF209921_1 (A         ISL1342       gil15428308 AF209921_1 (A         ISL1342       gil15428308 AF209921_1 (A         TA07       gil15428308 AF209921_1 (A         TA170       gil15428308 AF209921_1 (A         TA170       gil15428308 AF209921_1 (A         TA260       gil15428308 AF209921_1 (A         TA260       gil15428308 AF209921_1 (A         TA260       gil15428308 AF209921_1 (A         TB131       gil15428308 AF209921_1 (A         TB133       gil15428308 AF209921_1 (A         TB131       gil15428308 AF209921_1 (A         TB133       gil15428308 AF209921_1 (A         TB149       gil15428308 AF209921_1 (A         TB149       gil15428308 AF209921_1 (A         TB205       TB205         T	(AF209921) 1 2 (AF209921) 1 2 (AF209921) 1 2 (AF209921) 1 4 (AF209921) 1 5 (AF209921) 1 5 (AF209921) 1 3 (AF209921) 1 1 (AF209921) 1 1 (AF209921) 1 1 (AF209921) 1 1 (AF20921) 1 3 (AF20921) 1 3 (AF20921) 1 4 (AF20921) 1 3 (AF20921) 1 3 (AF2012) 1 3 (AF2012) 1 3 (AF2012) 1 3 (AF2012) 1 3 (AF2012)	.00E-41 .00E-34 .00E-34 .00E-40 .00E-40 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34			Similar to 14 kDa salivary gland protein				
TB103       gil15428308 AF209921_1 (A         ISL1025       gil15428308 AF209921_1 (A         ISL1129       gil15428308 AF209921_1 (A         ISL932       gil15428308 AF209921_1 (A         ISL335       gil15428308 AF209921_1 (A         TA07       gil15428308 AF209921_1 (A         TA170       gil15428308 AF209921_1 (A         TA170       gil15428308 AF209921_1 (A         TA204       gil15428308 AF209921_1 (A         TA204       gil15428308 AF209921_1 (A         TA204       gil15428308 AF209921_1 (A         TA204       gil15428308 AF209921_1 (A         TB131       gil15428308 AF209921_1 (A         TB133       gil15428308 AF209921_1 (A         TB133       gil15428308 AF209921_1 (A         TB133       gil15428308 AF209921_1 (A         TB133       gil15428308 AF209921_1 (A         TB149       gil15428308 AF209921_1 (A         TB205       gil15428308 AF209921_1 (A         TB205       gil15428308 AF209921_1 (A         TB205       gil15428308 AF209921_1 (A	I (AF20921) 1 2. I (AF20921) 1 2. I (AF20921) 1 2. I (AF20921) 1 4. I (AF20921) 1 5. I (AF20921) 1 5. I (AF20921) 1 5. I (AF20921) 1 7. I (AF20921) 1 1. I (AF20921) 1 1. I (AF20921) 1 1. I (AF20921) 1 4. I (AF20921) 1 3. I (AF20921) 1 5. I (AF20921)	.00E-41 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34			Similar to 14 kDa salivary gland protein				
<ul> <li>ISL1025 gil15428308 AF209921_1 (A</li> <li>ISL1129 gil15428308 AF209921_1 (A</li> <li>ISL1342 gil15428308 AF209921_1 (A</li> <li>TA07 gil15428308 AF209921_1 (A</li> <li>TA170 gil15428308 AF209921_1 (A</li> <li>TA204 gil15428308 AF209921_1 (A</li> <li>TA204 gil15428308 AF209921_1 (A</li> <li>TA266 gil15428308 AF209921_1 (A</li> <li>TB131 gil15428308 AF209921_1 (A</li> <li>TB132 gil15428308 AF209921_1 (A</li> <li>TB133 gil15428308 AF209921_1 (A</li> <li>TB149 gil15428308 AF209921_1 (A</li> <li>TB179 gil15428308 AF209921_1 (A</li> <li>TB179 gil15428308 AF209921_1 (A</li> <li>TB123</li> <li>TB239 cluster30 gil15428308 AF209921_1 (A</li> <li>TB376 gil15428308 AF209921_1 (A</li> <li>TB376 gil15428308 AF209921_1 (A</li> <li>TB239 cluster30 gil15428308 AF209921_1 (A</li> <li>TB330 cluster134 gil15428308 AF209921_1 (A</li> <li>TB192_cluster54 gil15428308 AF209921_1 (A</li> <li>TB239_cluster51 gil15428308 AF209921_1 (A</li> <li>TB30_4_cluster54 gil15428308 AF209921_1 (A</li> <li>TA304_cluster51 gil15428308 AF209921_1 (A</li> <li>TA304_cluster51 gil15428308 AF209921_1 (A</li> <li>TA304_cluster51 gil15428308 AF209921_1 (A</li> <li>TA304_cluster51 gil15409 A9500 A9500 A9500 A9500 A9500 A9500 A9500 A9500 A9500 A950</li></ul>	I (AF20921) 1 1. (AF209221) 1 2. [1 (AF209221) 1 2. [1 (AF209221) 1 4. [1 (AF209221) 1 5. [1 (AF209221) 1 5. [1 (AF209221) 1 7. [1 (AF209221) 1 1. [1 (AF209221) 1 1. [1 (AF209221) 1 1. [1 (AF209221) 1 1. [1 (AF209221) 1 2. [1 (AF20921) 1 2. [1 [AF20921] 1 2. [1 [AF20421] 1 [AF204	.00E-34 .00E-40 .00E-40 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34				11.4	21	9.3	4.1
<ul> <li>ISL1129</li> <li>ISL1129</li> <li>ISL1342</li> <li>ISL1342</li> <li>ISL332</li> <li>ISL332</li> <li>ISL332</li> <li>ISL332</li> <li>ISL332</li> <li>ISL332</li> <li>ISL332</li> <li>ISL332</li> <li>ISL332</li> <li>ISL333</li> <li>ISL3428308</li> <li>IS20921_1 (A</li> <li>IA204</li> <li>IS15428308</li> <li>IA720921_1 (A</li> <li>IA204</li> <li>IS15428308</li> <li>IA720921_1 (A</li> <li>IB131</li> <li>IS15428308</li> <li>IA720921_1 (A</li> <li>IB131</li> <li>IS15428308</li> <li>IA720921_1 (A</li> <li>IB131</li> <li>IS15428308</li> <li>IA720921_1 (A</li> <li>IB131</li> <li>IB149</li> <li>IS15428308</li> <li>IA720921_1 (A</li> <li>IB133</li> <li>IB154</li> <li>IB154</li> <li>IB154</li> <li>IB154</li> <li>IS15428308</li> <li>IA720921_1 (A</li> <li>IB154</li> <li>IB154</li> <li>IB154</li> <li>IB154</li> <li>IS15428308</li> <li>IA720921_1 (A</li> <li>IB154</li> <li>IB154</li> <li>IA149</li> <li>II5428308</li> <li>IA720921_1 (A</li> <li>IB376</li> <li>IB154</li> <li>IS15428308</li> <li>IA720921_1 (A</li> <li>IB239</li> <li>IA149</li> <li>II5428308</li> <li>IA700921_1 (A</li> <li>IB239_cluster30</li> <li>IB192_cluster30</li> <li>IS15428308</li> <li>IA710</li> <li>IS124810</li> <li>II54381447</li> <li>IS1444_cluster7</li> <li>IS124810</li> <li>IS124810</li> <li>IS124810</li> <li>IS124810</li> <li>IS1444_cluster7</li> <li>IS124810</li> <li>IS1449_cluster64</li> <li>IS1444_cluster70</li> <li>IS124810</li> <li>IS1444_cluster70</li> <li>IS144810</li> <li>IS144810</li> <li>IS1444_cluster70</li> <li>IS144810</li> <li>IS1444_cluster70</li> <li>IS1444_cluster108</li> <li>IS1440</li> <li>IS1440</li> <li>IS1444_cluster108</li> <li>IS1440</li> <li>IS1440</li> <li>IA140_cluster108</li> <li>IA140</li> <li>IA140_cluster70</li> <li>IA140_cluster70</li> <li>IA140_cluster108</li> <li>IA140</li> <li>IA140_cluster70</li> <li>IA140_cluster70<td>I (AF20921) 1 2. (AF20921) 1 4. I (AF20921) 1 5. (AF20921) 1 5. I (AF20921) 1 5. I (AF20921) 1 7. I (AF20921) 1 1. I (AF20921) 1 1. I (AF20921) 1 1. I (AF20921) 1 1. I (AF20921) 1 4. I (AF20921) 1 4. I (AF20921) 1 4. I (AF20921) 1 3. I (AF20921) 1 4. I (AF20921) 1 3. I (AF2052021) 1 3. I (AF20520221) 1 3. I (AF20520221) 1 3. I (AF20520221) 1 3. I (AF</td><td>.00E-40 .00E-40 .00E-34 .00E-34 .00E-36 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34</td><td></td><td></td><td>Similar to 14 kDa salivary gland protein</td><td>13.6</td><td>21</td><td>11.5</td><td>9.19</td></li></ul>	I (AF20921) 1 2. (AF20921) 1 4. I (AF20921) 1 5. (AF20921) 1 5. I (AF20921) 1 5. I (AF20921) 1 7. I (AF20921) 1 1. I (AF20921) 1 1. I (AF20921) 1 1. I (AF20921) 1 1. I (AF20921) 1 4. I (AF20921) 1 4. I (AF20921) 1 4. I (AF20921) 1 3. I (AF20921) 1 4. I (AF20921) 1 3. I (AF2052021) 1 3. I (AF20520221) 1 3. I (AF20520221) 1 3. I (AF20520221) 1 3. I (AF	.00E-40 .00E-40 .00E-34 .00E-34 .00E-36 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34			Similar to 14 kDa salivary gland protein	13.6	21	11.5	9.19
<ul> <li>ISL1342</li> <li>ISL1342</li> <li>ISL1342</li> <li>ISL932</li> <li>ISL932</li> <li>ISL932</li> <li>ISL932</li> <li>ISL932</li> <li>ISL932</li> <li>ISL932</li> <li>IS15428308 AF209921_1 (A</li> <li>TA170</li> <li>TA204</li> <li>TA204</li> <li>TA266</li> <li>IS15428308 AF209921_1 (A</li> <li>TA266</li> <li>IS15428308 AF209921_1 (A</li> <li>TA266</li> <li>IS15428308 AF209921_1 (A</li> <li>IB131</li> <li>IS1812</li> <li>IS15428308 AF209921_1 (A</li> <li>IB131</li> <li>IS15428308 AF209921_1 (A</li> <li>IB131</li> <li>IS164</li> <li>IS17812</li> <li>IS15428308 AF209921_1 (A</li> <li>IB149</li> <li>IS15428308 AF209921_1 (A</li> <li>IB152</li> <li>IB154</li> <li>IB154</li> <li>IB154</li> <li>IB154</li> <li>IS15428308 AF209921_1 (A</li> <li>IB152</li> <li>IB154</li> <li>IB192_cluster30</li> <li>IS15458308 AF209921_1 (A</li> <li>IB192_cluster30</li> <li>IS15452573 hypotheti</li> <li>IA340_cluster134</li> <li>IS1552</li> <li>IS1545573 hypotheti</li> <li>IA340_cluster134</li> <li>IS1552573 hypotheti</li> <li>IA340_cluster134</li> <li>IS17464_cluster7</li> <li>IS17464_cluster7</li> <li>IS17464_cluster7</li> <li>IS17464_cluster7</li> <li>IS17464_cluster7</li> <li>IS17464_cluster64</li> <li>IS14464_cluster7</li> <li>IS17499202031716210 hypotheti</li> <li>IA340_cluster166</li> <li>IS174992011P52_ANESU PRO</li> <li>IS14992001P52_Nypotheti</li> <li>IA340_cluster164</li> <li>IS174992031716210 hypotheti</li> <li>IA344_cluster76</li> <li>IS174979101P52_Nypotheti</li> <li>IA344_cluster76</li> <li>IS1749702031716210 hypotheti</li> <li>IA344_cluster76</li> <li>IA444_cluster76</li> <li>IA444_cluster76</li> <li>IA444_clu</li></ul>	(AF20921) 1 4. (AF20921) 1 5. (AF20921) 1 5. (AF20921) 1 5. (AF20921) 1 7. (AF20921) 1 7. (AF20921) 1 1. (AF20921) 1 1. (AF20921) 1 1. (AF20921) 1 4. (AF20921) 1 4. (AF20921) 1 4. (AF20921) 1 3.	.00E-40 .00E-34 .00E-34 .00E-36 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34			Similar to 14 kDa salivary gland protein	13.6	21	11.5	9.14
<ul> <li>ISL932 gil15428308 AF209921_1 (A TA07 gil15428308 AF209921_1 (A TA170 gil15428308 AF209921_1 (A TA204 gil15428308 AF209921_1 (A TA266 gil15428308 AF209921_1 (A TA266 gil15428308 AF209921_1 (A ISTB12 gil15428308 AF209921_1 (A IB179 gil15428308 AF209921_1 (A TB179 gil15428308 AF209921_1 (A TB179 gil15428308 AF209921_1 (A TB179 gil15428308 AF209921_1 (A TB205 gil15428308 AF209921_1 (A TB205 gil15428308 AF209921_1 (A TB223 gil15428308 AF209921_1 (A TB223 gil15428308 AF209921_1 (A TB223 gil15428308 AF209921_1 (A TB234 gil15428308 AF209921_1 (A TB235 gil15428308 AF209921_1 (A TB239_cluster30 gil15428308 AF209921_1 (A TB376 gil15428308 AF209921_1 (A TB330_cluster30 gil7505725 T23573 hypotheti TA340_cluster10 gil1542_ANESU PRC ISTB31_cluster108 gil7505725 T23573 hypotheti TA244_cluster16 gil3881447  (Z19157) contain ITB11_cluster16 gil750657 T3060 hypotheti TA244_cluster54 gil760657 T3060 hypotheti TA244_cluster56 gil760557 T3060 hypotheti TB11_cluster108 gil760552 T3050060 hypotheti TB11_cluster76 gil7606557 T3060 hypotheti</li> </ul>	I (AF20921) 1 1. [(AF20921) 1 5. [(AF20921) 1 5. [(AF20921) 1 3. ](AF20921) 1 7. [(AF20921) 1 1. [(AF20921) 1 1. [(AF20921) 1 1. [(AF209221) 1 1. [(AF209221) 1 3. ](AF209221) 1 4. [(AF209221) 1 4. [(AF209221) 1 2. ](AF209221) 1 2. ](AF209921) 1 2. ](AF209221) 1 2. ](AF209241) 1 2. ](AF2041) [(AF2041) [(AF204	.00E-34 .00E-36 .00E-36 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34 .00E-44			Similar to 14 kDa salivary gland protein	11.4	21	9.3	4.13
TA07       gil15428308 AF209921_1 (A         TA170       gil15428308 AF209921_1 (A         TA204       gil15428308 AF209921_1 (A         TA256       gil15428308 AF209921_1 (A         TA260       gil15428308 AF209921_1 (A         TA260       gil15428308 AF209921_1 (A         TA260       gil15428308 AF209921_1 (A         TB131       gil15428308 AF209921_1 (A         TB131       gil15428308 AF209921_1 (A         TB179       gil15428308 AF209921_1 (A         TB179       gil15428308 AF209921_1 (A         TB179       gil15428308 AF209921_1 (A         TB205       gil15428308 AF20921_1 (A         TB206       gil15428308 AF20921_1 (A	<ul> <li>I. (AF20921) 1</li> <li>A. (AF20921) 1</li> </ul>	.00E-34 .00E-36 .00E-34 .00E-34 .00E-34 .00E-34 .00E-34 .00E-32			Similar to 14 kDa salivary gland protein	13.7	21	11.6	9.25
TA135       gil15428308 AF209921_1 (A         TA170       gil15428308 AF209921_1 (A         TA266       gil15428308 AF209921_1 (A         TA260       gil15428308 AF209921_1 (A         TA260       gil15428308 AF209921_1 (A         TB131       gil15428308 AF209921_1 (A         TB131       gil15428308 AF209921_1 (A         TB131       gil15428308 AF209921_1 (A         TB179       gil15428308 AF209921_1 (A         TB179       gil15428308 AF209921_1 (A         TB179       gil15428308 AF209921_1 (A         TB205       gil15428308 AF20921_1 (A         TB210       gil15428308 AF20921_1 (A	<ul> <li>I. (AF20921) 1</li> <li>3.</li> <li>I. (AF20921) 1</li> <li>3.</li> <li>I. (AF20921) 1</li> <li>1.</li> <li>I. (AF20921) 1</li> <li>1.</li> <li>I. (AF20921) 1</li> <li>1.</li> <li>I. (AF20921) 1</li> <li>1.</li> <li>I. (AF20921) 1</li> <li>3.</li> <li>3.</li> <li>4.</li> <li>4.</li> <li>(AF20921) 1</li> <li>4.</li> <li>4.</li> <li>4.</li> <li>(AF20921) 1</li> <li>4.</li> <li>4.</li> <li>(AF20921) 1</li> <li>4.</li> <li>4.</li> <li>(AF20921) 1</li> <li>3.</li> <li>(AF20921) 1</li> <li>3.</li> </ul>	.00E-36 .00E-34 .00E-34 .00E-34 .00E-34 .00E-32 .00E-44			Similar to 14 kDa salivary gland protein	13.7	21	11.6	8.87
TA170       gil15428308 AF209921_1 (A         TA266       gil15428308 AF209921_1 (A         TA260       gil15428308 AF209921_1 (A         TB131       gil15428308 AF209921_1 (A         TB131       gil15428308 AF209921_1 (A         TB131       gil15428308 AF209921_1 (A         TB131       gil15428308 AF209921_1 (A         TB152       gil15428308 AF209921_1 (A         TB179       gil15428308 AF209921_1 (A         TB205       gil15428308 AF209921_1 (A         TB376       gil15428308 AF209921_1 (A         TB376       gil15428308 AF209921_1 (A         TB47       gil15428308 AF209921_1 (A         TB239_cduster30       gil15428308 AF209921_1 (A         TB47       gil15428308 AF209921_1 (A         TB47       gil15428308 AF209921_1 (A         TB47       gil15428308 AF209921_1 (A         TB47       gil15428308 AF209921_1 (A	I (AF20921) 1 3. (AF20921) 1 7. I (AF20921) 1 1. I (AF20921) 1 1. I (AF20921) 1 1. I (AF20921) 1 3. I (AF20921) 1 4. I (AF20921) 1 4. I (AF20921) 1 4. I (AF20921) 1 3. I (AF20921) 1 3.	.00E-34 .00E-34 .00E-34 .00E-34 .00E-32 .00E-44			Similar to 14 kDa salivary gland protein	13.4	21	11.4	8.92
TA204       gi 15428308 AF209921_1 (A         TA256       gi 15428308 AF209921_1 (A         TB131       gi 15428308 AF209921_1 (A         TB131       gi 15428308 AF209921_1 (A         TB131       gi 15428308 AF209921_1 (A         TB152       gi 15428308 AF209921_1 (A         TB152       gi 15428308 AF209921_1 (A         TB179       gi 15428308 AF209921_1 (A         TB205       gi 15428308 AF209921_1 (A         TB205       gi 15428308 AF209921_1 (A         TB213       gi 15428308 AF209921_1 (A         TB223       gi 15428308 AF209921_1 (A         TB224       gi 15428308 AF209921_1 (A         TB225       gi 15428308 AF209921_1 (A         TB236       gi 15428308 AF209921_1 (A         TB376       gi 15428308 AF209921_1 (A         TB376       gi 15428308 AF209921_1 (A         TB376       gi 15428308 AF209921_1 (A         TB47       gi 15428308 AF209921_1 (A         <	(AF20921) 1 7. (AF20921) 1 1. (AF20921) 1 1. (AF20921) 1 1. (AF20921) 1 3. (AF20921) 1 3. (AF20921) 1 4. (AF20921) 1 4. (AF20921) 1 3.	.00E-34 .00E-34 .00E-34 .00E-32 .00E-44			Similar to 14 kDa salivary gland protein	13.6	21	11.6	9.19
TA256       gil15428308 AF209921_1 (A         TA260       gil15428308 AF209921_1 (A         TB131       gil15428308 AF209921_1 (A         TB131       gil15428308 AF209921_1 (A         TB152       gil15428308 AF209921_1 (A         TB152       gil15428308 AF209921_1 (A         TB179       gil15428308 AF209921_1 (A         TB179       gil15428308 AF209921_1 (A         TB205       gil15428308 AF209921_1 (A         TB223       gil15428308 AF209921_1 (A         TB223       gil15428308 AF209921_1 (A         TB376       gil15428308 AF209921_1 (A         TB376       gil15428308 AF209921_1 (A         TB47       gil15428308 AF209921_1 (A         TB539_coluster3       gil15428308 AF20950_1_1 (A	I (AF20921) 1 1. (AF20921) 1 1. I (AF20921) 1 1. I (AF20921) 1 3. I (AF20921) 1 3. I (AF20921) 1 4. I (AF20921) 1 4. I (AF20921) 1 3. I (AF20921) 1 3.	.00E-34 .00E-34 .00E-32 .00E-44			Similar to 14 kDa salivary gland protein	13.6	21	11.5	9.19
TA260       gil15428308 AF209921_1 (A         ISTB12       gil15428308 AF209921_1 (A         TB131       gil15428308 AF209921_1 (A         TB152       gil15428308 AF209921_1 (A         TB179       gil15428308 AF209921_1 (A         TB179       gil15428308 AF209921_1 (A         TB179       gil15428308 AF209921_1 (A         TB205       gil15428308 AF209921_1 (A         TB205       gil15428308 AF209921_1 (A         TB21       gil15428308 AF209921_1 (A         TB21       gil15428308 AF209921_1 (A         TB23       gil15428308 AF209921_1 (A         TB21       gil15428308 AF209921_1 (A         TB23       gil15428308 AF209921_1 (A         TB23       gil15428308 AF209921_1 (A         TB47       gil15428308 AF20921_1 (A         TB51       gil15428308 AF20921_1 (A         TB53       gil15428308 AF20921_1 (A         TB239_cl	I (AF20921) 1 1. I (AF20921) 1 1. I (AF20921) 1 3. I (AF20921) 1 4. I (AF20921) 1 4. I (AF20921) 1 4. I (AF20921) 1 3. I (AF20921) 1 2.	.00E-34 .00E-32 .00E-44			Similar to 14 kDa salivary gland protein	13.6	21	11.5	9.19
ISTB12 gi 15428308 AF209921_1 (A TB131 gi 15428308 AF209921_1 (A TB152 gi 15428308 AF209921_1 (A TB179 gi 15428308 AF209921_1 (A TB205 gi 15428308 AF209921_1 (A TB205 gi 15428308 AF209921_1 (A TB223 gi 15428308 AF209921_1 (A TB276 gi 15428308 AF209921_1 (A TB376 gi 15428308 AF209921_1 (A gi 15428308 AF209921_1 (A TB376 gi 15428308 AF209921_1 (A gi 15428308 AF209921_1 (A TB376 gi 15428308 AF209921_1 (A TB47 gi 15428308 AF209921_1 (A TB47 gi 15428308 AF209921_1 (A TB47 gi 15428308 AF209921_1 (A TB376 gi 15428308 AF209921_1 (A TB376 gi 15428308 AF209921_1 (A TB376 gi 15428308 AF209921_1 (A TB47 gi 15428308 AF209921_1 (A TB376 gi 15428308 AF209921_1 (A TB376 gi 15428308 AF209921_1 (A TA340_cluster134 gi 15428308 AF209921_1 (A TA340_cluster134 gi 15428308 AF2093765) CC TB192_cluster51 gi 124810 IF52_ANESU PRC ISTB31_cluster108 gi 7500203 T16210 hypotheti TA244_cluster16 gi 2881447  (Z19157) contain ISTB31_cluster16 gi 7497910 T20219 hypotheti TA244_cluster54 gi 7506657/T73060 hypotheti TB11_chuster56 gi 7506657/T73060 hypotheti	I (AF20921) 1 1. (AF20921) 1 3. (AF20921) 1 4. (AF20921) 1 4. (AF20921) 1 4. (AF20921) 1 3.	.00E-32 .00E-44			Similar to 14 kDa salivary gland protein	13.7	21	11.6	9.28
TB131       gi 15428308 AF209921_1 (A         TB152       gi 15428308 AF209921_1 (A         TB179       gi 15428308 AF209921_1 (A         TB205       gi 15428308 AF209921_1 (A         TB376       gi 15428308 AF209921_1 (A         TB376       gi 15428308 AF209921_1 (A         TB47       gi 15428308 AF209921_1 (A         TB51       gi 15428308 AF209921_1 (A         TB47       gi 15428308 AF209921_1 (A         TB51       gi 15428308 AF20951_1 (A         TB1339_cluster30       gi 15428308 AF20951_1 (A         TA340_cluster31       gi 1544810 IF52_ANESU PRO         TA340_cluster54       gi 124810 IF52_ANESU PRO         TB11_cluster104       gi 124810 IF52_ANESU P	I (AF20921) 1 3. (AF20921) 1 4. (AF20921) 1 4. (AF20921) 1 4. (AF20921) 1 3.	.00E-44			Similar to 14 kDa salivary gland protein	13.9	21	11.8	9.41
TB149       gi 15428308 AF209921_1 (A         TB152       gi 15428308 AF209921_1 (A         TB179       gi 15428308 AF209921_1 (A         TB205       gi 15428308 AF209921_1 (A         TB223       gi 15428308 AF209921_1 (A         TB223       gi 15428308 AF209921_1 (A         TB223       gi 15428308 AF209921_1 (A         TB376       gi 15428308 AF209921_1 (A         TB376       gi 15428308 AF209921_1 (A         TB47       gi 15428308 AF209921_1 (A         TB51       gi 15428308 AF209921_1 (A         TB47       gi 15428308 AF209921_1 (A         TB51       gi 15428308 AF209921_1 (A         TB47       gi 15428308 AF209921_1 (A         TB51       gi 15428308 AF209921_1 (A         TB40_cluster30       gi 15428308 AF209921_1 (A         TA340_cluster30       gi 15428308 AF20950_ANESU PRC         TB192_cluster54       gi 154810 IF52_ANESU PRC         TB192_cluster54       gi 7500203 T16210 hypotheti         TA330_cluster166       gi 7500203 T16210 hypotheti         TA330_cluster124       gi 7500203 T16210 hypotheti         TA330_clus	( (AF209921) 1 4. ( (AF209921) 1 4. ( (AF209921) 1 3. ( (AF209921) 1 3.				Similar to 14 kDa salivary gland protein	13.4	21	11.4	7.52
TB152       gi 15428308 AF209921_1 (A         TB179       gi 15428308 AF209921_1 (A         TB205       gi 15428308 AF209921_1 (A         TB222       gi 15428308 AF209921_1 (A         TB376       gi 15428308 AF209921_1 (A         TB47       gi 15428308 AF209921_1 (A         TB51       gi 15428308 AF209921_1 (A         TB53       gi 15428308 AF209921_1 (A         TB44       gi 15428308 AF20951_1 (A         TB33       gi 15428308 AF20951_1 (A         TA340_cluster51       gi 174810 IF52_ANESU PRC         TB192_cluster54       gi 17502037716210 hypotheti         TA230_cluster51       gi 7502037716210 hypotheti         TA304_cluster51       gi 7497983277347212 hypotheti         TA244_cluster51       gi 74979101720219 hypotheti         TA311_chuster54 <t< td=""><td>l (AF20921) 1 4. l (AF20921) 1 3. l (AF20921) 1 2.</td><td>.00E-33</td><td></td><td></td><td>Similar to 14 kDa salivary gland protein</td><td>13.6</td><td>21</td><td>11.5</td><td>9.34</td></t<>	l (AF20921) 1 4. l (AF20921) 1 3. l (AF20921) 1 2.	.00E-33			Similar to 14 kDa salivary gland protein	13.6	21	11.5	9.34
TB179       gi 15428308 AF209921_1 (A         TB205       gi 15428308 AF209921_1 (A         TB222       gi 15428308 AF209921_1 (A         TB376       gi 15428308 AF209921_1 (A         TB51       gi 15428308 AF209921_1 (A         TB53       gi 15428308 AF209921_1 (A         TB33       gi 15428308 AF20951_1 (A         TA340_cluster134       gi 174810 IF52_ANESU PRC         ISTB342_cluster54       gi 17502037716210 hypotheti         TA230_cluster54       gi 75002037716210 hypotheti         TA230_cluster124       gi 75020337716210 hypotheti         TA244_cluster51       gi 749798337734212 hypotheti         TA244_cluster51       gi 74979910/1720219 hypotheti         TA244_cluster54       gi 74979910/1720219 hypotheti	I (AF209921) 1 3. I (AF209921) 1 2.	.00E-34			Similar to 14 kDa salivary gland protein	13.7	21	11.5	9.08
TB205       gi 15428308 AF209921_1 (A         TB222       gi 15428308 AF209921_1 (A         TB223       gi 15428308 AF209921_1 (A         TB376       gi 15428308 AF209921_1 (A         TB51       gi 15428308 AF209921_1 (A         Group 2       gi 15428308 AF209921_1 (A         TB539_cluster30       gi 15428308 AF209921_1 (A         TB339_cluster30       gi 15428308 AF209921_1 (A         TB339_cluster30       gi 15428308 AF209921_1 (A         TA340_cluster134       gi 15428308 AF209951_0 PRC         TA3464_cluster7       gi 174810 IP52_ANESU PRC         TB192_cluster54       gi 1726827  (AE003765) CC         TB192_cluster54       gi 75002037 T16210 hypotheti         TA230_cluster54       gi 75002037 T16210 hypotheti         TA244_cluster54       gi 7499910 T52212 hypotheti         TA244_cluster51       gi 74979100 T720219 hypotheti         TA11_cluster108       gi 74979101 T20219 hypotheti         TA244_cluster51       gi 74979101 T20219 hypotheti	(AF209921)1 2	.00E-34			Similar to 14 kDa salivary gland protein	13.7	21	11.5	9.25
TB222       gi 15428308 AF209921_1 (A         TB223       gi 15428308 AF209921_1 (A         TB376       gi 15428308 AF209921_1 (A         TB51       gi 15428308 AF209921_1 (A         Group 2       gi 15428308 AF209921_1 (A         TB39_cluster30       gi 15428308 AF209921_1 (A         TB330_cluster30       gi 15428308 AF209921_1 (A         TB330_cluster30       gi 15428308 AF209921_1 (A         TA340_cluster134       gi 15428308 AF209921_1 (A         TA340_cluster51       gi 124810 IF52_ANESU PRC         TB192_cluster51       gi 124810 IF52_ANESU PRC         TB192_cluster51       gi 10726827  (AE003765) CC         TB192_cluster54       gi 7500203 T16210 hypotheti         TA230_cluster54       gi 7500203 T16210 hypotheti         TA244_cluster54       gi 7507203 T16210 hypotheti         TA244_cluster51       gi 7507203 T16210 hypotheti         TA244_cluster51       gi 7497910 T720219 hypotheti         TA11_cluster76       gi 750650 hypotheti         TB11_cluster71       gi 750650 hypotheti		.00E-37			Similar to 14 kDa salivary gland protein	13.5	21	11.5	9.28
TB223       gi 15428308 AF209921_1 (A         TB376       gi 15428308 AF209921_1 (A         TB47       gi 15428308 AF209921_1 (A         TB51       gi 15428308 AF209921_1 (A         Group 2       gi 15428308 AF209921_1 (A         Group 2       gi 15428308 AF209921_1 (A         TB39_cluster30       gi 15428308 AF209921_1 (A         Group 2       gi 15428308 AF209921_1 (A         TB39_cluster30       gi 15428308 AF209921_1 (A         TB339_cluster30       gi 15428308 AF209921_1 (A         T340_cluster134       gi 15428308 AF209921_1 (A         ISTB464_cluster7       gi 124810 IF52_ANESU PRC         ISTB382_cluster51       gi 10726827  (AE003765) CC         TB192_cluster16       gi 17500203 T16210 hypotheti         TA230_cluster16       gi 7500203 T16210 hypotheti         TA244_cluster16       gi 750203 T16210 hypotheti         TA244_cluster16       gi 750203 T16210 hypotheti         TA244_cluster13       gi 7497910 T20219 hypotheti         TB11_cluster108       gi 7506060 hymotheti         T811_cluster75       gi 760657 T73060 hymotheti	l (AF209921) 1 3.	.00E-42			Similar to 14 kDa salivary gland protein	13.5	21	11.5	9.32
TB376       gi 15428308 AF209921_1 (A         TB47       gi 15428308 AF209921_1 (A         TB51       gi 15428308 AF209921_1 (A         Group 2       gi 15428308 AF209921_1 (A         Group 2       gi 15428308 AF209921_1 (A         Group 2       gi 15428308 AF209921_1 (A         TB239_cluster30       gi 15428308 AF209921_1 (A         TB239_cluster30       gi 15428308 AF209921_1 (A         TA149_cluster48       gi 15428308 AF209921_1 (A         TA340_cluster134       gi 124810 IP52_ANESU PRC         ISTB464_cluster7       gi 124810 IP52_ANESU PRC         ISTB382_cluster51       gi 10726827  (AE003765) CC         TB192_cluster164       gi 7500203 T16210 hypotheti         TA304_cluster164       gi 7502033 T16210 hypotheti         TA230_cluster108       gi 74978121 hypotheti         TA244_cluster168       gi 7497910 T20219 hypotheti         TA244_cluster31       gi 7497910 T20219 hypotheti         TB11_cluster764       gi 750650 hypotheti	l (AF209921) 1 3.	.00E-45			Similar to 14 kDa salivary gland protein	13.3	21	11.3	9.41
TB47       gi 15428308 AF209921_1 (A         TB51       gi 15428308 AF209921_1 (A         Group 2       TB239_cluster30         TB239_cluster30       gi 15428308 AF209921_1 (A         Group 2       TB239_cluster30         TB239_cluster30       gi 15428308 AF209921_1 (A         TB239_cluster30       gi 15428308 AF209921_1 (A         TB239_cluster30       gi 124810 IF52_ANESU PRC         ISTB464_cluster7       gi 124810 IF52_ANESU PRC         ISTB382_cluster51       gi 10726827  (AE003765) CC         TB192_cluster51       gi 10726827  (AE003765) CC         TA30_cluster124       gi 7500203 T16210 hypotheti         TA230_cluster108       gi 7500203 T16210 hypotheti         TA244_cluster108       gi 7497910 T20219 hypotheti         TA11_cluster108       gi 7497910 T20219 hypotheti         TB11_cluster754       gi 756657/T73060 hypotheti	l (AF209921) 1 3.	.00E-39			Similar to 14 kDa salivary gland protein	11.2	21	9.1	4.67
TB51       gi 15428308 AF209921_1 (A         Group 2       TB239_cluster30         TB239_cluster30       gi 7505725 T23573 hypotheti         TB239_cluster48       gi 124810 IP52_ANESU PRC         ISTA464_cluster7       gi 124810 IP52_ANESU PRC         ISTB382_cluster51       gi 10726827  (AE003765) CC         TB192_cluster54       gi 7500203 T16210 hypotheti         TA30_cluster124       gi 7500203 T16210 hypotheti         TA30_cluster166       gi 7500203 T16210 hypotheti         TA230_cluster108       gi 750203 T16210 hypotheti         TA244_cluster108       gi 750203 T16210 hypotheti         TA244_cluster108       gi 750203 T16210 hypotheti         TA11_cluster108       gi 75060 hypotheti         TB11_cluster756       gi 75060 hypotheti	l (AF209921) 1 2.	.00E-50			Similar to 14 kDa salivary gland protein	14	22	11.9	9.65
Group 2         TB239_cluster30         TA149_cluster30         TA149_cluster30         gi 7505725 T23573 hypotheti         TA340_cluster134         gi 124810 IP52_ANESU PRC         ISTB464_cluster51         gi 124810 IP52_ANESU PRC         ISTB382_cluster51         gi 10726827  (AE003765) CC         TB192_cluster54         gi 7500203 T16210 hypotheti         TA30_cluster124         TA30_cluster166         gi 7497910 T6212 hypotheti         TA304_cluster16         gi 7497827 T32212 hypotheti         TA244_cluster168         gi 7497910 T20219 hypotheti         TA11_cluster76         oi1766657/T73060 hypotheti	l (AF209921) 1 1.	.00E-42			Similar to 14 kDa salivary gland protein	13.3	21	11.4	5.23
1B2.59_cluster50         TA149_cluster48       gi[7505725]T23573 hypotheti         TA340_cluster134       gi[124810]IP52_ANESU PRC         ISTA464_cluster7       gi[124810]IP52_ANESU PRC         ISTB382_cluster51       gi[10726827] (AE003765) CC         TB192_cluster54       gi[7500203]T16210 hypotheti         TA30_cluster124       gi[7500203]T16210 hypotheti         TA30_cluster124       gi[7500203]T16210 hypotheti         TA30_cluster124       gi[7500203]T16210 hypotheti         TA230_cluster124       gi[7500203]T16210 hypotheti         TA244_cluster108       gi[7500203]T1234212 hypotheti         TA244_cluster108       gi[7497910]T20219 hypotheti         TB11_cluster756       gi[766657]T73060 hypotheti						ц С	9	L N	00 0
TA149_cluster48         gil7505725[T23573 hypotheti           TA340_cluster134         gil124810[IP52_ANESU PRC           ISTA464_cluster7         gil124810[IP52_ANESU PRC           ISTB382_cluster51         gil10726827] (AE003765) CC           TB192_cluster54         gil7500203[T16210 hypotheti           TA30_cluster124         gil7500203[T16210 hypotheti           TA304_cluster124         gil7500203[T16210 hypotheti           TA230_cluster124         gil7497917[C219157] contain           ISTB31_cluster108         gil7497910[T20219 hypotheti           TA244_cluster31         gil7497910[T20219 hypotheti           TB11_cluster764         gil76657[T172060 hypotheti					Unknown function	C.8	19	6.0	9.89
TA340_cluster134       gi 124810]IP52_ANESU PRC         ISTA464_cluster7       gi 124810]IP52_ANESU PRC         ISTB382_cluster51       gi 10726827  (AE003765) CC         TB192_cluster54       gi 7500203 T16210 hypotheti         TA304_cluster124       gi 7500203 T16210 hypotheti         TA304_cluster108       gi 7500203 T16210 hypotheti         TA230_cluster124       gi 7500203 T16210 hypotheti         TA234_cluster108       gi 7497832[T34212 hypotheti         TA244_cluster31       gi 7497910[T20219 hypotheti         TB11_cluster108       gi 756657](T3060 hypotheti	hetical prot 5.	.00E-05	ofam00014 Kunitz_BPTI	6.00E-05	Protease inhibitor – Kunitz domain	8.7	27	5.9	9.61
ISTA464_cluster7 gi 124810]IP52_ANESU PRC ISTB382_cluster51 gi 10726827  (AE003765) CC TB192_cluster54 gi 7500203 T16210 hypotheti TA230_cluster124 gi 3881447  (Z19157) contain ISTB31_cluster108 gi 7497910[T20219 hypotheti TA244_cluster31 gi 7497910[T20219 hypotheti TB11_cluster76 oi175066 hypotheti	PROTEASE INHIB	0.004			Protease inhibitor?	8.8	20	6.9	4.68
ISTB382_cluster51 gi 10726827  (AE003765) CC TB192_cluster54 gi 7500203 T16210 hypotheti TA230_cluster124 gi 3881447  (Z19157) contain ISTB31_cluster108 gi 7497832[T34212 hypotheti TA244_cluster31 gi 7497910[T20219 hypotheti TB11_cluster76 oi175066 hypotheti	PROTEASE INHIB	0.004			Protease inhibitor?	8.9	19	7	4.28
TB192_cluster54         gi/7500203[T16210 hypotheti           TA230_cluster124         TA304_cluster16         gi/3881447] (Z19157) contain           ISTB31_cluster108         gi/7498327[T34212 hypotheti         TA244_cluster26         gi/7497910[T20219 hypotheti           TB11_cluster76         gi/7506557](T37060 hypotheti         TB11_cluster76         gi/7506557](T37060 hypotheti	) CG18436 gen 8.	.00E-04	smart smart00131 KU	3.00E-04	Protease inhibitor - Kunitz domain	6	22		5.53
TA230_cluster124 TA304_cluster16 gi 3881447  (Z19157) contain ISTB31_cluster108 gi 7497832[T34212 hypotheti TA244_cluster31 gi 7497910[T20219 hypotheti TB11_cluster26 oi 7566657][T37060 hypotheti	hetical prot	0.047			Unknown function	6	17	7.3	9.33
TA304_cluster16         gi 3881447  (Z19157) contain           ISTB31_cluster108         gi 7497832[T34212 hypotheti           TA244_cluster31         gi 7497910[T20219 hypotheti           TB11_cluster36         oi 75666577[T3060 hypotheti					Unknown function	9.1	20	7.1	4.63
ISTB31_cluster108 gi/749832[T34212 hypotheti TA244_cluster31 gi/7497910[T20219 hypotheti TB11_cluster26 oi/766657[T32060 hypotheti	itains simil	0.079			Unknown function	9.2	22	٢	8.83
TA244_cluster31 gi 7497910 T20219 hypotheti TB11_cluster26 oil7606657/T32060 hypotheti	hetical prot	0.00			Unknown function	9.2	21	7.1	4.29
TB11_cluster26	hetical prot 9.	.00E-06	fam00014 Kunitz_BPTI	1.00E-04	Protease inhibitor - Kunitz domain	9.9	27	7.2	6.48
	hetical prot 5.	.00E-06	mart smart00131 KU	2.00E-05	Protease inhibitor - Kunitz domain	10.2	27	7.5	9.3
ISTA12_cluster45 gi 6164595 AF078161_1 (AF	(AF078161) la	0.001			Protease inhibitor?	10.3	19	8.2	5.84
TB76B gi 6164595 AF078161_1 (AF	(AF078161) la	0.058			Unknown function	10.4	19	8.4	10
Group 3									
TB135B					Unknown function	5.4	19	3.5	3.9
TA78_cluster2					Unknown function	6.3	19	4.4	6.03

Group and									
sequence name	Best match to NR protein database	E-value	Best CDD match	E-value	Comment	$MW^1$	$SP^2$	MW <sup>3</sup>	Iq
TA37B					Unknown function	6.3	19	4.4	4.83
TB184B					Unknown function	6.3	19	4.4	4.83
TB194B					Unknown function	6.4	19	4.4	4.83
TB07B	gi 114248 ASF1_HELAN ANTHER-SPECIFI	5.00E-04			Proline rich peptide	6.5	19	4.6	7.61
TA58_cluster29					Unknown function	6.5	19	4.6	5.51
TB73_cluster15	gi 114248 ASF1_HELAN ANTHER-SPECIFI	0.003			Proline rich peptide	9.9	19	4.6	9.78
TA20B					Unknown function	9.9	19	4.6	4.97
TB138B					Unknown function	6.6	19	4.5	5.33
TB111B					Unknown function	6.7	19	4.7	7.61
TB313B	gi 13751843  (AL590734) hypothetica	0.029			Unknown function	6.8	19	4.8	9.3
Group 4 TB396B	eil15428348 AF278575 1 (AF278575) S	6.00E-37			Similar to Salp 10	9.6	17	8.1	5.75
TB50B	eil15428348 AF278575 1 (AF278575) S	2.00E-35			Similar to Salp 10	6.6	17	8.1	4.9
ISTA15_cluster42	gil15428348 AF278575_1 (AF278575) S	5.00E-24			Similar to Salp 10	9.9	19	7.9	5.25
TB386B	gi 15428348 AF278575_1 (AF278575) S	4.00E-37			Similar to Salp 10	10.2	22	7.9	7.58
TA13	gi 15428348 AF278575_1 (AF278575) S	4.00E-27			Similar to Salp 10	10.2	17	8.5	7.47
TB42_cluster20	gi 15428348 AF278575_1 (AF278575) S	8.00E-25			Similar to Salp 10	10.4	22	8.1	7.48
TB144B	gi 15428348 AF278575_1 (AF278575) S	2.00E-34			Similar to Salp 10	10.4	17	8.7	9.27
TB15B	gi 15428348 AF278575_1 (AF278575) S	6.00E-34			Similar to Salp 10	10.4	17	8.7	9.27
Group 5									
ISL1022_cluster264	gi 10726827  (AE003765) CG18436 gen	3.00E-18			Protease inhibitor	35.7	19	33.7	8.24
ISL1352_cluster68	gi 10726827  (AE003765) CG18436 gen	2.00E-14			Protease inhibitor?	36.3	19	34.2	9.29
ISL1228_cluster344	gi 10726827  (AE003765) CG18436 gen	3.00E-13	Smart smart00131 KU	6.00E-04	Protease inhibitor -Kunitz domain	38	22	35.5	9.46
Group 6 ISL1040_cluster18			pfam02098 His_binding	3.00E-04	Histamine binding domain	34.2	17	32.6	5.73
ISL1276_cluster363	gi 8470378 HBP2_RHIAP FEMALE-SPECIF	2.00E-04			Similar to Rhipicephalushistamine binding ptn	36.6	18	34.6	5.15
Group 7							č		
ISL929_cluster233	gil15428290 AF209912_1 (AF209912) t	2.00E-31			Similar to L. scapularis thrombospondin	12.8	57	10.2	9.24
INL13/3_cluster33	g1 2428290 AF209912_1 (AF209912) t	/.00E-26	pramu 2853 ACK	5.00E-04	Similar to I. scapularis thrombospondin	C.51	18	0.11	C4.C
Remaining sequences									
Similar to other salivary	proteins								
ISL1083_cluster9 TB103_cluster1	gi 15428310 AF209922_1 (AF209922) 2 gi 15428308 AF209921_1 (AF209921) 1	1.00E-117 2.00E-39			95% identical to 25 kda salivary antigen Similar to 14 kDa salivary gland protein	25.3 10.7	20 15	23.3 9.3	9.66 4.14
Histamine Binding, not	Group 6								
ISL868_cluster49	pfam(	02098 His_bin	ding	3.00E-04	Histamine binding domain	25.4	21	23.3	6.41

Table 3. Continued

Sialome of the tick I. scapularis 2851

Group and sequence name	Best match to NR protein database	E-value	Best CDD match	E-value	Comment	$MW^{1}$	$SP^2$	MW <sup>3</sup>	Id
Protease inhibitors ISL1095_cluster291 ISL888_cluster62 ISTA397_cluster168 ISL1156_cluster318 ISL1268_cluster3360	gi 7521905 T18544 alpha-2-macroglob gi 15077002 AF286029_1 (AF286029) i gi 400070 ISH1_STOHE KUNITZ-TYPE PR gi 862467  (D32211) limulus intrace	1.00E-87 5.00E-59 6.00E-15 9.00E-26	pfam00207 A2M pfam00031 cystatin Smart smart00131 KU Smart smart00131 KU pfam00079 serpin	5.00E-50 3.00E-08 1.00E-11 3.00E-15 3.00E-15	Protease inhibitor – A2M partial sequence Protease inhibitor – Cystatin Protease inhibitor – Ixolaris like Protease inhibitor – Kunitz domain Protease inhibitor – Serpin type	42.3 14.3 18.7 9 23.7	24 25 23	11.9 16 6.6 20.8	5.67 4.93 5.58 10.5 5.81
Enzymes ISL1194_5nuc ISL1316_cluster379 ISL316_cluster188 ISL1033_cluster188 ISL1033_cluster583 ISL1324_cluster238	gi 12644305 5NTD_BOOMI C95 gi 12060148 AF106704_1 (AF106704) p gi 3970893  (AB020544) serin protei gi 5911708  (A1269650) hypothetical gi 5911708  (A1269650) hypothetical gi 2644364 NUDM_DROME NADH-UBIQUIN	1.00E-107 1.00E-34 1.00E-14 0 1.00E-158 1.00E-82	pfam01009 5_nucleotidase pfam00450 serine_carbpept Smart smart00020 Tryp_SPc pfam01421 Reprolysin pfam01712 dNK	1.00E-70 3.00E-31 3.00E-23 4.00E-13 1.00E-14 2.00E-14	5'-nucleotidase/apyrase secreted Carboxypeptidase – secreted Serine protease – signal anchor Zn Metalloprotease Secreted Zn Metalloprotease Secreted NADH-Ubiquinone oxireductase	64.6 38.2 41.4 44.1 46.1 47.8	21 29 849* 19*	62.5 34.9 36.7 38.2 45.8	5.6 6.8 4.73 8.94 9.0 6.11
Unknown function ISTB418_cluster179 ISI818 ISL1188_cluster39 ISL914_cluster14 ISTA482_cluster5 TA310_cluster133 ISL1068_cluster53 TB277_cluster160 ISL922_cluster53 ISL1270_cluster53 ISL1270_cluster22	gi 14733532  DKFZP564N1362 protein gi 11496688  B. burgdorferi predict gi 8072217 AF260530_1 (AF260530) Ds	6.00E-06 7.00E-04 0.043			Unknown function Secreted in saliva Unknown function Unknown function Unknown function Unknown function Unknown function Unknown function Unknown function	7.2 7.8 8.3 8.9 8.9 10.3 11.1 11.1 11.1 12.2 24.7 24.7 24.5	23 23 19 22 20 22 20 19 19	4.3 5.3 6.4 8.4 9.4 8.9 110.2 118.7 222.5	10.1 8.68 4.34 6.34 10.5 9.17 9.17 9.07 4.98 5.38
<sup>1</sup> Molecular mass c <sup>2</sup> Site predicted to c <sup>3</sup> Molecular mass o Other headings as	f the putative protein. ccur signal peptide cleavage, according to the 5 f the putative mature protein. in Table 1.	signalP prog	şram (*).						

Table 3. Continued

Δ	gi 15428308	MGLTGTMLVL-VSLAFFGSAAAHNC	QNGTRPASEQDREG	DYYCW	NAETKSWD	QFF <b>FG</b> NG <b>E</b> K	<b>CFYNSG</b> DHG	TCONGE	CHLTNNSG	G <b>P</b> NETD
/ \	тв47	MGLTGTTLVLAVSLAFFGSAAAHNC	ONGTRPASEONREG	DYYCW	NAETKSWD		CFYNSGDRG	TCONGE	CHLTTTSG	G <b>P</b> NETD
	TB223	MGLTGTTLVL-VSLAFFGSAAAHNC	ONGTRPASEKNREG	DYYCW	NAETKLWD	- Off <b>fg</b> ng <b>e</b> k	CFYNTGDRG		CHLTTSSG	<b>P</b> NEAD
	TA135	MGLTGTTLVL-VSLVFFGSAAAHN <mark>C</mark>	KNGTRPASEENREG	DYYC	NDGTNSWD	OFF <b>FG</b> NG <b>E</b> I	CFYNSGEKG	ICONGE	CHLTNNSG	G <b>P</b> NETD
	TB131	MGLTGTTLVL-VSLAFFGSAAAHN <mark>C</mark>	KNGTRPASEENREG	DYYC	NDGTNSWD	) FF <b>FG</b> NG <b>E</b> I	<b>CFYN</b> SGEKG	ICONGE	CHLTNNSG	<b>P</b> NETD
	ISL1129	MGLTGTTLVL-VCVAFFGSAAAHN <mark>C</mark>	ONGTRPASEENREG	DYYC	NEVTNSWD	- OFF <b>FG</b> NG <b>E</b> R	CFYNTGENG	KCONGE	CHLTTNSD	<b>P</b> NETD
	TB376	MGLTGTTLML-VCVAFFGTAAAHNC	- KNGTRPASEENREG	DYYCW	NEVTNSWD		CFYNTGENG	KCONGE	CHLTTNSD	G <b>P</b> NETD
	TB205	MGLTGTTLVL-VCVAFFGTAAAHNC	KNGTRPASEENREG	DYYC	NEVTNSWD	) FF <b>FG</b> NG <b>E</b> R	CFYNTGENG	KCONGE	CHLTTNSG	<b>P</b> DDTD
	TB222	MGLTGTTLVL-VCVAFFGSAAAHN <mark>C</mark>	ONGTRPASEKNREG	DYYC	NAETKSWD	- DFF <b>FG</b> DG <b>E</b> R	CFYNTGENG	TCRNGE	CHLTTSSG	<b>P</b> NETD
	ISTB12	MGLTGTTLVL-VSLAFFGSAAAHNC	ONGTRPTSEKNREG	DFYCW	NADTNLWD	~ KFF <b>FG</b> NG <b>E</b> K	<b>CFYNTG</b> EK <b>G</b>	TCLNGE	CHLTTSSG	<b>P</b> DDTG
	TB51	MGLTGTALVL-VSLAFFGSAAAHNC	ONGTRPASEENREG	DYYCW	NSETOSWD	)YF <b>FG</b> DG <b>E</b> R	CFYNSGDRG	ICONGE	CHLTTSSG	<b>PDDTD</b>
	TA170	MELTGITLVL-VSLAFFGSAAAETC	RNGTRPATOTDREG	DYYC	NTLTSSWD	~ KYF <b>FG</b> DE <b>E</b> F	<b>CFYNTG</b> LKG	TCKNGE	CHLTSEGGY	/ <b>P</b> TDPH
	TA204	MELTGITLVL-VSLALFGSAAAETC	RNGTRPASOTDREG	DYYC	NTLTSSWD	(YF <b>FG</b> DE <b>E</b> F	<b>CFYNTG</b> LKG	TCKNGE	CHLTSEGG	/ <b>P</b> TDPH
	ISL1025	MELTGITLVL-VSLAFFGSAAAETC	RNGTRPASOTDREG	DYYCW	NTLTSSWD	(YF <b>FG</b> DE <b>E</b> P	CFYNTGLKG	TCKNGE	CHLTSEGG	/ <b>P</b> TDPH
	TA256	MELTGITLVL-VSLAFFGSAAAETC	RNGTRPASOTDREG	DYYC	NSLTSSWD	(YF <b>FG</b> DE <b>E</b> F	<b>CFYNTG</b> LKG	TCKNGE	CHLTSEGGY	/ <b>P</b> TDPH
	TA07	MEFTGITLVL-VSVAFFGSAAAETC	RNGTRPASOTDREG	DYYC	NTLTSSWD	(YF <b>FG</b> DE <b>E</b> F	CFYNTGLRG	TCKNGE	CHLTSEGN	/ <b>P</b> TDPD
	TB149	MEFTGITLVL-VSVAFFGSAAAETC	RNGTRPASOTDREG	DYYCW	NTLTSSWD	(YF <b>FG</b> DE <b>E</b> F	CFYNTGLRG	T <b>C</b> K <b>NG</b> G	CHLTSEGN	/ <b>P</b> TDPD
	TB152	MEFTGITLVL-VSVAFFGSAAAETC	RNGTRPASOTDREG	DYYC	NTLTSSWD	(YF <b>FG</b> DE <b>E</b> F	CFYNTGLRG	TCKNGE	CHLTSEGGY	/ <b>P</b> TDPN
	TA260	MEFTGITLVL-VSLTFFGSAAAETC	RNGTRPASOTOREG	DYYCW	NSOTSSWD	(YF <b>FG</b> DN <b>E</b> F	CFYNTGLRG	TCONGE	CHLTSEGG	/ <b>P</b> TDPN
	TB179	MEFTGITLVL-VSLTFFGSAAAETC	RNGTRPGSOTOREG	DYYC	NSOTSSWD	(YF <b>FG</b> DN <b>E</b> F	CFYNTGLRG	TCONGE	CHLTSEGG	/ <b>P</b> TDPN
	ISL932	MEFTGITLLL-VSLAFFGSAAAETC	RNGTRPASOTOREG	DYYC	NLOTSSWD	(YF <b>FG</b> DN <b>E</b> P	<b>CFYNTG</b> LR <b>G</b>	TCONGE	CHLTSEGGY	/ <b>P</b> TDPN
	TB103	MGLTEIMLVL-VSLAFVATAAAHDC	ONGTRPASEEKREG	DYYCW	NTETKSWD	kff <b>fg</b> ng <b>e</b> r	CFYNNGDEG	LCONGE	CHLT TDSG	<b>PNDTD</b>
	ISL1342	MGLTEIMLVL-VSLAFVATAAAHDC	ONGTRPASEEKREG	YDYYCW	NTETKSWD	kff <b>fg</b> ng <b>e</b> r	CFYNNGDEG	LCONGE	CHLTTDSG	/ <b>P</b> NDTD
		* :* *:* *.:.::***. *	:**** :: .***	*:***	* *. **	::***: *	****.* .*	* **	****	* :.
	gi 15428308	DYTPAPT <b>E</b> KPKQKKKKTKKTKKPKR	KSKKDQEKNL	B s	L1342					
	TB47	DYTAAPT <b>E</b> KPKQKKKKTKKTKKPKR	KSKKDQEKNF	1	B103		0.1		TA260 TB17	9
	TB223	DNTPAPT <b>E</b> KPKQKKKKPKKTKKPKR	KSKKD				0.1		V <sup>ISI</sup>	.932
	TA135	ENTPATTEKPKQKKKKTKKPKKPKR	KSKKDQ							FA07
	TB131	DNTPAPT <b>E</b> KPKQKKKKPKKPKKPKR	KSKKDH		\					<b>T</b> B14
	ISL1129	DNTPPPT <b>E</b> KPKQKKKKPKKPKKPKR	KSKKDQ		```	<b>\</b>			$\sim$	TB152
	TB376	DNTPPPTEKPK				$\mathbf{N}$			TA204	TA256
	TB205	DNTPPPT <b>E</b> KPKQKKKKPKKPKKPKR	KSKKDQ			$\mathbf{X}$			ISL1025	
	TB222	DNTPPPT <b>E</b> KPKQKKKKPKKTKKPKR	KSRKDQ						TA170	
	ISTB12	DNTPPPT <b>E</b> KPKQKKKKPKKTKKPKR	KSKKDQKENF		HD	с \				
	TB51	ENTPPPTEKPKQKKKKPKKTKEPKR	KSKKD				E	TC		
	TA170	QYPSEPTEKPKKNKKKSKKTKKPK-	KSKKPKNN							
	TA204	QYPSEPT <mark>E</mark> KPKKNKKKSKKTKKPK-	KSKKPKNN			)				
	ISL1025	QYPSEPTEKPKKNKKKSKKTKKPK-	KSKKPKNN			1				
	TA256	QYPSEPTEKPKKNKKKSKKTKKPK-	KSKKPKNN		TB51	- 1	IDIC			
	TA07	QYPSEPTEKPKKSKKKSKKTKKPK-	KTKKPKDN				HNC			
	TB149	QYPSEPTEKPKKSKKKSKKTKKPK-	KTKKPKDN			1	oi 1	5428308 ø]	AAK97824 1	1
	TB152	QYPSEPT <mark>E</mark> KPKKNKKKSKKTKKPK-	KSKKPKDN			$\sim$	TD47	20500 5	,	1
	TA260	QYPSEPT <mark>E</mark> KPRKNKKKSKKTKKPK-	KSKKPKDN		ISTB12		TD222			
	TB179	QYPSEPT <b>E</b> KPKKNKKKSKKTKKPK-	KTKKPKDN				18223			
	ISL932	QYPSEPT <b>E</b> KPKKNKKKSKKTKKPK-	KTKKPKDN		трэээ	/	\	_		
	TB103	AKIEETEEELEA			10222		TP121	>		
	ISL1342	AKIEETE <mark>E</mark> ELEA			TI	B205	11129			
		. *:				TB376	-			

Fig. 1. (A) Alignment of related peptides, group 1 (Table 3) deduced from an *Ixodes scapularis* salivary gland cDNA library. Gray background, predicted signal secretory peptide as indicated by the SignalP program. Cysteine residues are shown white over a black background. Other conserved amino acid residues are shown in bold over a gray background. Asterisks, colons or stops below the sequences indicate identity, high conservation and conservation of the amino acids, respectively. (B) Unrooted cladogram indicating the three families HDC, ETC and HNC, named for their mature peptide amino-terminal sequence. The bar represents the degree of divergence among sequences.

### Peptide group 1

Peptides from group 1 consist of 22 sequences (Table 3) representing the most abundant family of messages in the salivary gland library (cluster 1; Table 1). These sequences have high similarity to the 14 kDa protein of *I. scapularis* (gi|15428308), but have no other significant matches to the NR database. No motifs were found when compared to CDD database. All sequences have a putative signal peptide indicative of secretion, which end in the tripeptide Ala-Ala-Ala (Fig. 1). Alignment of these 22 novel sequences with the 14 kDa protein (gi|15428308) (Fig. 1) indicates these proteins belong to three closely related families, ETC, HNC and HDC, for the amino-terminal sequence of the predicted mature peptides (see cladogram in Fig. 1). These proteins have a mature molecular mass predicted to vary from 9.1 to 11.5 kDa; most are basic in nature due to a lysine-rich carboxy-terminal

region. They all possess a conserved sequence Asn-Gly-Thr-Arg-Pro, starting at position 5 of the putative mature protein, which was detected twice by Edman degradation of protein bands excised from gels subjected to SDS-PAGE from separated tick salivary proteins (see below). Except for one sequence (ISL1342), all have six conserved cysteine residues. The function of these proteins remains elusive.

#### Peptide group 2

Peptides from group 2 (Table 3) are putative mature proteins varying in molecular mass from 6.5 to 8.4 kDa, of both basic and acidic pI. Four of the 13 proteins gave significant matches to Kunitz domains, indicating they may be protease inhibitors or otherwise interact with other protein domains. Most of the peptides gave BlastP matches to the NR protein database, indicating similarity to proteins annotated as protease

A	ISTB31_CLUSTER108 ISTB382_CLUSTER51 TA304_CLUSTER16 ISTA464_CLUSTER7 TA340_CLUSTER134 TA230_CLUSTER124 TB11_CLUSTER26 TA149_CLUSTER48 TA244_CLUSTER31 TB76B TB239_CLUSTER30 TB192_CLUSTER54 ISTA12_CLUSTER45	MKAILAVT CVFSAVV-LI MKAILAVT CIFSAVV-LI MKAILAVT CIFSAVV-LI MKAILAVT CIVSAVV-LI MKAILAVT CIVSAVV-LL MKAILAVT CIFSAVV-LL MKATLAAV FLATVV TI MKATIAAL CILAAVV VI MKAALVAIFFLAAVAYSM MKAALVAIFFLAAVAYSM MKAALVAIFFLAAVAYSM MKATLVAI CFLAAVAHSM MKDSIAVL FLVALSYVL	SALSKDV EAPHPTF SALSREV EAPHAMF SALSKQK EAPHATF SALPQEV EGPHAIS SALPQEV EGPHAIS SALPREV EAPYASA ALLPEDI RAPHAVA ALLPENI RAPHPVS GRLTEQQ RTPVPSS GRLTEQQ RTPVPSS GRLSEEQ RRPVPST ATLTEDE RAQLAFS	SCDGQAP COCAPGVT SCDPNAP SCDPNAP SCDPNAP SCASDAE SCASDAE SCASGTP MCAEDAK SCASGTP SCASGT-S'	LRTSYYFNNGTGK ESEF PRVTYYFNNGTGK ESEF LGNFYYYNNGTQR EKEF LANWYYFNNGTAR ESLF LGDFYFNHGTGR VSEF PKLLFYFNNGTNR ESYF VKETWYFNNRTNK EKYS ITTMYYFDNHTDR QNYL ITTMYFFNNTNK EKYC TRTIYSFNNNTNK EKYC TRTIYSFNNNTNK EKYC TRTIYSFNNNTNK EKYC TRTIYSFNNTNK EKYC TRUFYSFFNNTNC ESYT	'-G '-G '-G '-G !-G !DS !-G !DS !-G
	ISTB31_CLUSTER108 ISTB382_CLUSTER51 TA304_CLUSTER16 ISTA464_CLUSTER7 TA340_CLUSTER134 TA230_CLUSTER124 TB11_CLUSTER26 TA149_CLUSTER48 TA244_CLUSTER31 TB76B.SEQ TB239_CLUSTER30 TB192_CLUSTER54 ISTA12_CLUSTER45	CSNGRTDFSSEEDCRQAW CGNGRNDFPTAEKCRDAC CGGPNN-FPSKGQCRKEC CGGPNE-FSTEDECKKKN CGGPNE-FSTEDECKK-N CHGPMN-YPTEDECRQC CGGGLNDFGSKACCKDSC CGGGYNDFGSKACCKDSC CGGGYNDFGSKGCGISEC CGEGINQFEKKCCCISEC CGEGINQFEKKCCCSEC CGEGINQFEKKCCCSEC CDTGKNRFPSLGKCINEC	PYGIYASND PYGIYANNG VRSEYTR PYGIYASSS PYGRNKPNSWKPTKV PYG-NK PYGRHHPPGKRGKGR PYGRHHPPGKRGKGR RLLH PYGRHHPPGMRVRGT	ARQ  KL- F Y		
В	Sequence gi  13124581  gi  4102791  gi  7497910  gi  1083171  gi  14735875  gi  14735889  gi  14735889  gi  142848	Description TFPI TFPIbeta Hypothetical protein TFPI TFPI Hypothetical protein Salb10	Species Mouse C. elegans Rabbit Human J. scapularis	Score 29.6 26.1 23.1 22.5 20.8 20.8 19.2	E-value 5.3e-008 7.9e-007 9e-006 1.4e-005 5.4e-005 5.4e-005 0.00019	

Fig. 2. (A) Alignment of group 2 peptides (Table 3). See Fig. 1A legend for an explanation of the layout. (B) The alignment in A was used to build a hidden Markov model that retrieved the shown sequences when searched against the non-redundant NCBI protein database.

inhibitors. Cysteine residues are conserved in most peptides of this group, as well as a N-X-T preceding the third conserved cysteine of the mature peptides (Fig. 2). Remarkably, there is significant conservation of the predicted signal peptide. In the first 24 amino acid positions, there are 12 positions that are identical or conserved (excluding the initial methionine), whereas for the remaining 63 ungapped positions there are 13 conserved positions. The  $\chi^2$ -test indicates these ratios to be significant at *P*=0.0223. This conservation of the signal peptide was observed earlier in a family of antimicrobial peptides of frog skin skin (Charpentier et al., 1998), and in semenogelins, a family of mammalian semen proteins (Lundwall and Lazure, 1995).

To further investigate the nature of the peptide group 2, we built a hidden Markov model based on the alignment shown in Fig. 2, using the –f switch to allow for the presence of multiple domains in the resulting model. Search of the NR database produced six matches with an E value of 5.4E–005 or lower, three of which are the mouse, the rabbit and the human anticlotting protein, tissue factor pathway inhibitor (TFPI). TFPI is a blood coagulation inhibitor containing three tandem Kunitz domains; two of these domains have been demonstrated to interact with Factor VIIa or Factor Xa (Girard et al., 1989). Single Kunitz molecules with specificity for Factor VIIa or elastase have also been characterized in libraries from phage

display (Dennis and Lazarus, 1994) and from extracts of the parasitic worm *Ancylostoma ceylanicum ceylanicum* (Milstone et al., 2000), respectively. The model also recognized another *I. scapularis* salivary protein, SALP10, but with a higher (less significant) E value of 1.9E–4.

#### Peptide group 3

Group 3 cDNA sequences code for short peptides of mature molecular mass ranging from 3.5–4.8 kDa of both basic and acidic nature (Table 3). All sequences are relatively glycineand proline-rich. Some sequences give weak matches to proteins in the NR database annotated as collagen; these possess two conserved cysteine residues in the mature peptide and remarkable conservation of the secretory signal peptide (Fig. 3). All amino acid sites of the predicted signal secretory peptide are conserved, against 18 of 35 sites on the mature peptide. A  $\chi^2$ -test is significant at *P*=0.0422. It is possible some of these sequences are alleles of an extremely polymorphic locus or, alternatively, that they represent different conserved loci. The possible function of these peptides remains elusive.

#### Peptide group 4

Group 4 sequences code for putative mature peptides having four conserved cysteine residues, molecular mass 7.9–8.7 kDa, of both basic and acidic nature. All display strong similarity





Fig. 3. (A) Alignment of group 3 peptides (Table 3). See Fig. 1A legend for an explanation of the layout. (B) The unrooted cladogram of all sequences. The bar represents the degree of divergence among sequences.

(BlastP against NR database) to a protein from *I. scapularis* named SALP10 (gi|15428348), and weak similarities to mammalian tissue pathway inhibitor (TFPI) and bungarotoxin. 19 of 21 first amino acids (excluding initial methionine) are conserved (Fig. 4), as compared to 33 of the 69 amino acids of the mature peptide. This difference is highly significant ( $\chi^2$ -test, *P*<0.001), indicating higher conservation of the signal peptide rather than the mature protein. An HMM model made from the alignment shown in Fig. 4 retrieved only SALP10 from the NR database, with an E value of 1.9E–070 but no other significant matches.

#### Peptide group 5

Three sequences in group 5 (Table 3) code for proteins of mature molecular mass ranging from 33.7 to 35.5 kDa of a basic nature, and having 24 conserved cysteine residues (Fig. 5). Comparisons with the NR protein database using BlastP indicate similarities to proteins annotated as protease inhibitors, including TFPI and the protein Ixolaris, an inhibitor of Factor VIIa (Francischetti et al., 2002). ISL228\_Cluster344 has a Kunitz domain, as indicated by the SMART database. These proteins probably code for anticlotting compounds.

#### Peptide group 6

Group 6 represents sequences giving similarities histamine-binding proteins (Table 3, Fig. 6). to ISL1040\_cluster233 has no matches to the NR protein database but has a significant match by RPSBlast to the Pfam histaminebinding domain, whereas ISL1276\_cluster 363 has no such match but instead has similarity to the tick Rhipicephalus apendiculatus histamine-binding protein found in the NR protein database. These two proteins are mildly acidic and have a mature molecular mass of 32.6 kDa and 34.6 kDa, respectively. It is probable that these proteins function by binding histamine or other small ligands.

#### Peptide group 7

The two sequences in group 7 match a sequence deposited in the NR database from I. scapularis and annotated as thrombospondin. The two predicted mature sequences, with eight conserved cysteine residues, code for two peptides of molecular mass 10.2 and 11.6 kDa, one basic and the other acidic in nature. Their similarities to thrombospondin proteins are not apparent. Both sequences have weak similarities to disintegrin metalloproteases, and ISL373\_cluster33 has the cysteine-rich domain of ADAM proteases as predicted by the Pfam database. No RGD domains found in disintegrins are observed in these sequences, nor in any of the other sequences reported in Table 3. Fig. 7 shows the alignment of the two proteins with the Ixodes thrombospondin found in the NR database. The role of the cysteine-rich domain of ADAM proteases is not known but it is postulated to interact with integrins and/or other attachment motifs of cells and matrix proteins (Hooper, 1994). Accordingly, these peptides could be involved in disruption of platelet aggregation, cell-matrix interactions and/or inhibition of angiogenesis (Roberts, 1996).

The remaining 24 novel sequences presented in this paper can be grouped as: (i) similar to previously reported *I. scapularis* salivary proteins; (ii) a novel, shorter, protein with a Pfam histamine-binding motif, but not similar to other HBP found in the NR database (when compared by a BlastP search); (iii) five novel proteins coding for different inhibitors of proteolytic activity; (iv) six enzymes; and (v) ten proteins probably secreted and with unknown function.

TB50B TB396B TA13 TB386B cil5428348 Salp10	MKATIAVLCFLVAVAYAIVVEARMASQPIDNDALNGROVKP-KECPGNSKTVYYYDPKSG MKATIAVLCFLVAVAYAIVVEARMASQPIDNGALNPKCEKP-KECPGNSKTVYYYDPKSG MKATIAVLCVLVAVAYAIVVEARMKSHPIDNGPLDPKCVKP-RECPGTSQTVSYYNPEAG MKATIAVLCVLVAVAYTIVVEARMKSHPIDNGPLDPKCVKP-RECPGTSQTVSYYNPEAG MKAAIAVUCCLVAVAYTIVVEARMANODIDDCDLNBKCVKP-KCCPGDEKTYSYYDPAKG
TB144B	MKATMAVLCFLAAVAYAIVVDARIASQPIDDGALNPKCEKP-KECPGSDRTVYYYNRKEG
TB15_cluster4 TB42 cluster20	MKATMAVLCFLAAVVYAIVVDARIASQP <b>ID</b> NGALNPKCEKP-KECPGSDRTVYYYNRKEG MKATIAVLCFLVALAYAIVVEAOMG-OPIDNGPLNPLCEKPPSNCLGNASYAYVYNRTKG
ISTA15_cluster42	MKATIP-LCFIVAVAYAIAAK-PLRATIIDNGELNPRCEKP-NDCPGNQRTVFYYNRTAG
	***::. **.:.*:*:*: : **:. *: * ** * * . *: *
TB50B	CQHIQLGADCTDNGNYRTLAECNQYCLTPPGK
TB396B	CQHIQLGADCTDNGNYRTLAECNQYCLTPPGK
TA13	CQLIQLGANCTDNGNYKTLGECNKHCLPAPGTPSRLG-
TB386B	CQLIQLGANCTDNGNYKTLGECNKHCLPAPGTPSRLG-
gi 15428348 Salp10	CQLIKLGENCTDNGNYPTLEDCNRHCLPPPGKQTRLS-
TB144B	CKGVQLGKNCTDNGNYDSLELCNTYCLPAPCKQPRLA-
TB15_cluster4	CKGVQLGKNCTDNGNYDSLELCNTYCLPAPGKQPRLA-
TB42_cluster20	CYSVGLSEGCGNNGYYTTPQDCHQYCLPPPGRQGGRRI
ISTA15_cluster42	
	GQQIRLGAGCSDNGNIQILDEGLIKGAPPPGKQV



Fig. 4. (A) Alignment of group 4 peptides (Table 3). See Fig. 1A legend for an explanation of the layout. The SALP10 peptide from the NR protein database is added for comparison. (B) The unrooted cladogram of all sequences. The bar represents the degree of divergence among sequences.

ISL1352_CLUSTER68 ISL1022_CLUSTER264 ISL1228_CLUSTER344	MHKKIWWTLIAAAFGICSGQN MHKQTVWTLVAAALGICSGQN MQRNILWISVVAAFGVFHFGECTYQDSGEDSSS *::: * :.**::	ISEEENVGLDEPQLGQGRNVVKGWTYKS IFQEGNVGYYEPDLGRGRDIIKGWTYDA MTEESRGDGPPHASRGLMNIPGWFYDG * * * * :* :* **
ISL1352_CLUSTER68 ISL1022_CLUSTER264 ISL1228_CLUSTER344	DLDKCYMFYHVKRDDYRNENIFLTETACNKKG YLDKCYVFHHAKRSHYGNENIFQSESLCNQRG SRDQCRRYHFPDQQFDMAKNKFKTVTECRKSC *:* ::: :* *: : *:: *:	NHVPVV     YAKRPPSKGTSDHPVATYDP       NPTVPAK     YAKPPSKGKSDLPVVTYDP       STVPLQ     FKKPPQTIRTMGLPVSTYNS       **     *:     ** **:
ISL1352_CLUSTER68 ISL1022_CLUSTER264 ISL1228_CLUSTER344	NTGRCINIRATKGRGVENVENNGVS©TKKCRDA NNGRCLDIRATKQGGAENVENNRAS©KKQCLDA TQGECVTIAVRPGQTGPNIFRREVECNETCRDE . *.*: * *:** . * .	ADLRLCLNATEADCEHMENPSTSYRYDN ADLRLCLNATKADCEYIGDPSYRYNA PEYGKCAPLHIVDCGGSTGNHYNL : * ** :*:
ISL1352_CLUSTER68 ISL1022_CLUSTER264 ISL1228_CLUSTER344	VSQTCKKSADGSCGGFQSAEKCFQRCAVLVEN ETETCEAAPAGTCGGFQSAEKCFQRCGILVEN DKQTCEKTTKNKCGPFATLEDCFKRCARYIQR :**::** : * ***** ::.*	CTLPIQNITTCEKPTKRYGYNKEKSOC CTLPIQNITTCENPTKRYGYNKEKDOC CNIPLLKSEYCDIVDLRYWYNSESKOC *.:*:: *: ** ** **
ISL1352_CLUSTER68 ISL1022_CLUSTER264 ISL1228_CLUSTER344	EEFLGCADGGNSFEEAKECWSLCAPKHRCNM EELFGCADGGNSFQCAKECWSVCAPKHRCNM EEIMGCADDVYNFPTAKECWETCSSKEESRCLC **::**** * ***** *: * **	MSPDTGHFPKLGLYRRHYFDVTTNDGRS MSPDTGRFAKLGLYTRYYFDVTTNECTS 2PPDLGKLGIGRTRYYYNITSNRGLT ** *:: *:*:::*:* *:
ISL1352_CLUSTER68 ISL1022_CLUSTER264 ISL1228_CLUSTER344	ARKLKPRVPG-NTNLFATPEECEQICKPQYQGT ARKLKRTVPG-NTNLFVTADECKQICKPQYQGT TTHVAFWQNTEKKNNFKRRSDCENTCRPKHKDV ::::::::::::::::::::::::::::::::::::	PPEH PPEH /KKL :



# Messages coding for proteins similar, but not identical, to previously reported I. scapularis sequences

ISL1083\_cluster9 is 95% identical to the previously reported 25 kDa proteins of *I. scapularis* (not shown) and may represent an allele of a highly polymorphic gene or another closely related gene. ISL1083\_cluster9, which does not display

a histamine-binding motif, is highly similar to two other proteins found in the NR database that are also from *Ixodes scapularis* salivary gland cDNA libraries and annotated as histamine-binding, 17 kDa proteins. The alignment of the four sequences shows highly conserved areas, including the putative secretory signal peptide (Fig. 8). The mature 17 kDa



protein is a truncated version of the other three proteins, containing two conserved cysteine residues in the mature form, while the remaining proteins have an additional four cysteine residues. These proteins may have a function in blood feeding by binding small mediator molecules involved in hemostasis or inflammation.

# Novel putative protein containing the histamine-binding domain

ISL868\_cluster49 has no similarities to proteins in the NR database but has a histamine-binding motif, a predicted signal peptide, and the molecular mass of the mature protein is 23.3 kDa. This molecular mass is similar to ISL1083\_cluster9



analyzed above, *I. scapularis* 25 kDa protein A, and *I. scapularis* histamine-binding protein, to which ISL868 may be distantly related.

#### Sequences coding for different protease inhibitors

Five predicted proteins appear to function as protease inhibitors. ISL1095\_cluster291, an  $\alpha$ -2-macroglobulin truncated clone with highest similarity to the *Limulus* protein, also demonstrates very high similarity to vertebrate proteins. These protein inhibitors are very large and entrap the proteases that they inhibit; they may also bind to cytokines (Armstrong and Quigley, 1999; Borth, 1992). Because the clone we describe in this paper is the truncated carboxyterminal region, we do not know whether there is a signal peptide indicative of secretion coded in this message.

ISL888\_cluster62 codes for a secreted peptide with mature molecular mass of 11.9 kDa containing the cystatin domain of cysteine protease inhibitors; 15 kDa cystatin has been described previously in several nematodes (Dainichi et al., 2001; Hartmann et al., 1997; Manoury et al., 2001). These nematode cystatins inhibit the lymphocyte asparaginyl endopeptidase involved in class II antigen processing in human B cells and inhibit T-cell proliferation. A similar function may be served by ISL888\_cluster62.

ISTA397\_cluster68 is similar to the *I. scapularis* TFPI-like molecule Ixolaris (alignment in Fig. 9), a molecule containing one complete and one incomplete Kunitz domain (Francischetti et al., 2002). ISTA397\_cluster68 has the same number of cysteine residues in the first and second Kunitz domains as does Ixolaris. ISTA397\_cluster68 may accordingly work also as a TFPI, or inhibit some other proteases such as chymotrypsin or trypsin (Petersen et al., 1996).

ISL1156\_cluster318 codes for a 10kDa peptide with a

Fig. 10. Lack of GPI anchor motif in ISL1194\_5nuc: alignments of Ι. scapularis 5'-nucleotidase with rat, mouse (Mus), human (Homo), bovine and **Boophilus** tick 5'-(bos). nucleotidase. Identical residues in five or more sequences are bold with gray background. Similar polar or gi|11024643|Rat SEYISKMKVIYPAVEGRIKFS---AASHYQGSFPLIILSFWAVILVLYQ gi|6754900|Mus SEYISKMKVVYPAVEGRIKFS---AASHYQGSFPLVILSFWAMILILYQ gi|4505467|Homo STXISKMKVIYPAVEGRIKFS---TESHCHGSFSLIFLSLWAVIFVLYQ gi|404502|Bos SCYISKMKVLYPAVEGRIQFS---AGSHCCGSFSLIFLSVLAVIILYQ gi|12644305|Boophilus MKYMNSTSPITTALDGRVTFLKTNQASDACLNLASPFLVLLVLVVFYHL ISL1194\_5nuc IEYARKMSPIKEPEEGRVIMYDNPRPANSTAGLPIDANKTSPKFSAAKP

hydrophobic amino acids are shown with gray background. The arrow indicates the conserved serine residue involved in the GPI anchor of 5'nucleotidases. The box shows the hydrophobic carboxyterminal region, which penetrates the cell membrane. Green background in the box indicates hydrophobic amino acid residues.

Kunitz domain, having considerable similarity to other proteins from the NR database annotated as protease inhibitors of both vertebrate and invertebrate origins.

Finally, ISL1268\_cluster360 codes for a mature protein of 20.8 kDa with a serpin motif, highly similar to *Limulus* coagulation inhibitor and to other serine protease inhibitors of both vertebrate and invertebrate origins. Interestingly, the mRNA has two open reading frames, both of which code for serpins, one with a typical secretory peptide, the other apparently leading to an intracellular protein. The specificity and activity of these putative protease inhibitors remain to be determined.

#### Sequences coding for different enzymes

Six clones are reported to code for enzymes. ISL1194 5nuc codes for a protein with high similarity to invertebrate and vertebrate 5'-nucleotidases and apyrases. 5'-nucleotidases have a signal peptide indicative of secretion, which causes the protein to be expressed extracellularly, and a carboxy terminus in which a GPI anchor fixes the protein to the extracellular side of the membrane (Ogata et al., 1990). The GPI anchor is attached to a conserved serine residue, followed by a stretch containing 15 or 16 hydrophobic amino acid residues. Neither mosquito salivary apyrase, a secreted enzyme, nor a 5'nucleotidase of sand fly saliva, has this conserved serine. These enzymes also lack the hydrophobic carboxy terminus, allowing the enzyme to be secreted (Champagne et al., 1995; Charlab et al., 1999). Analysis of the carboxy terminus of ISL1194\_5nuc (Fig. 10) shows that it does not have the conserved serine found in mammalian and constitutive tick 5'-nucleotidases. Instead of 15-16 hydrophobic residues, it contains only eight such residues. Furthermore, it contains four charged (K+E) and three polar (T+S) residues, making the carboxy terminus unlikely to be intramembranous. ISL1194\_5nuc is thus possibly responsible for the previously described salivary apyrase of *I. scapularis* (Ribeiro et al., 1985), or may code for a secreted 5'-nucleotidase.

ISL1316\_cluster379 codes for a serine carboxypeptidase containing a signal peptide indicative of secretion. The specificity of this putative carboxypeptidase is unknown. It probably does not code for the previously described kininase activity of *I. scapularis* saliva, which has kinetic characteristics of another family of peptidases, the angiotensin converting enzymes (ACE) (Ribeiro and Mather, 1998). ISL1316\_cluster379 carboxypeptidase could, however, be the salivary enzyme described previously to inactivate the serum anaphylatoxins C3a and C5a (Ribeiro and Spielman, 1986).

ISL812\_cluster188 codes for a protein with high similarity to proteins from the NR database annotated as chymotrypsin, elastase, enterokinase and enteropeptidase. The best protein match is from a protease from the tick *Haemaphysalis longicornis* (Mulenga et al., 1999). ISL812\_cluster188 putative protein has a strong signal anchor as determined by the SignalP program. It probably is not secreted and serves a housekeeping function.

ISL1033\_cluster65 and ISL1324\_cluster383 have very high similarity to a hypothetical protein from the tick *I. ricinus* and to other proteins in the NR database annotated as disintegrins and metalloproteases. Both have the Pfam reprolysin motif indicative of a zinc metalloprotease family, most commonly found in snake venoms (Hooper, 1994). Neither has a signal sequence indicative of secretion; however, the amino-terminal sequences for both were found in protein bands of one-dimensional electrophoresis of saliva samples (see below).

# Sialome of the tick I. scapularis 2859

Finally, ISL939\_cluster238 has very high similarity Drosophila NADH-ubiquinone to melanogaster oxidoreductase, a typical mitochondrial enzyme ranging in molecular mass from 69 to 75 kDa, and to other proteins annotated as deoxyguanosine/deoxyadenosine kinases, consistent with the finding of a deoxynucleoside kinase (DNK) motif from the Pfam database. DNK are 44 to 56 kDa enzymes described on both mitochondria and cytosol (http://brenda.bc.uni-koeln.de). ISL939\_cluster238 codes for a putative protein containing a signal peptide indicative of secretion, with a mature molecular mass of 45.8 kDa. It is thus possible that ISL939\_cluster238 codes for a secreted DNK in saliva with an unknown function in the tick feeding process.

#### Sequence coding for proteins of unknown function

Eleven additional clones were fully sequenced, either because they represented abundant clones or because their partial sequence contained a signal peptide indicative of secretion. Although all of these full-length clones code for putative proteins displaying a signal peptide indicative of secretion, no function was indicated when their sequences were compared to the NR or CDD database. ISTB418\_cluster179 codes for a 4.3 kDa basic peptide with similarity to human and murine proteins of unknown function. ISL942 cluster53 has similarity to a Borrelia burgdorferi protein (E value 1E-4) and weak similarity to a tick histamine-binding protein (E value 0.006). This putative protein, and that coded by ISL1270\_cluster22, has a predicted mature molecular mass of 22.5-22.6 kDa, similar to the protein described in Table 3 as histamine-binding, not group 5 (ISL868\_cluster49). Alignments of these three putative proteins reveal no obvious similarities (not shown).



# N-terminal sequence

Fig. 11. Edman degradation of protein bands from Ixodes scapularis saliva and their corresponding products. gene Numbers at left indicate the position of the molecular mass markers (kDa).

# Initial characterization of the proteome set of Ixodes scapularis

To obtain further information on the salivary proteome set of *I. scapularis*, electrophoresis of saliva and SGH were performed by one-dimensional SDS-PAGE followed by transference of the proteins to PVDF membranes, staining with Coomassie Blue, and submission of the cut bands to Edman degradation. 15 and 19 bands yielded useful sequence information from saliva and salivary gland gels, respectively (Figs 11, 12). With the exception of one larger molecular mass band in the saliva gel (FEVGKDYYY...), and three sequences on the SGH gel, we tentatively assigned all other sequences to a gene product, as follows.

#### Host proteins

Sequences originating from proteins in saliva included two matching rabbit albumin and one matching the  $\alpha$ -chain of rabbit hemoglobin. Similarly, the SGH-derived sequences included both the  $\alpha$ - and  $\beta$ -chains of rabbit hemoglobin as well as a sequence with high similarity to Ig- $\kappa$  light chain.

# Amino-terminal sequences matching putative proteins coded by cDNA sequences from cluster 1

Two sequences in each of the two gels fractionating saliva and SGH matched putative proteins belonging to the most abundant cluster of cDNA sequences. The observed aminoterminal sequences matched those predicted by the SignalP program. Mature sequences from group 1 peptides start with either HX or ET, followed by C-[QKRQ]-NGTRPAS (see above and Fig. 1). Accordingly, the sequences HNXQNG-

#### TRPASEENREGXDY and HKXQNGTRPASEKNREGXDY

were obtained from protein bands of saliva separated by SDS-PAGE and corresponding to the sequences of clones ISL1129 and TB222. Gels from SGH yielded the Edman degradation products HNXQDGTRPASE and HNXKNGTRPASE, matching clones ISTA48 and TA379 for which we do not have full-length sequences. Notably, although proteins from group 1 (Table 3) vary in molecular mass from 9.3 to 11.5 kDa, they all are located in the 20–24 kDa region in both gels. It is thus possible that the proteins of this cluster make dimers through disulfide bridges even when the samples are run under reducing conditions or, alternatively, they may be modified by posttranslation mechanisms such as glycosylation.

## Amino-terminal sequences matching putative proteins coded by cDNA sequences from cluster 14

Two proteins belonging to cluster 14 were also represented in both gels and, in both cases, represented by the pair of sequences from clones ISTB346 and ISL914. The observed amino-terminal sequences are in agreement with the mature peptide sequence predicted by the SignalP program. Although the mature peptide predicted by ISL914\_cluster14 is 7 kDa, it was found in the 10–12 kDa regions of the reduced saliva gel and in the 30 kDa region of the non-reduced SGH gel, indicating that these molecules may form multimers through disulfide bridges. Alternatively, this peptide may have a compact structure in its oxidized state that precludes sufficient binding of SDS, leading to less charge and apparently higher molecular mass in the gel experiment (Pitt-Rivers and Impiombato, 1968). No Asn



Fig. 12. Edman degradation of protein bands from Ixodes scapularis salivary gland homogenates, their and corresponding gene products. Numbers at left indicate the position of the molecular mass markers (kDa).

# Amino-terminal sequence matching the tick anticomplement protein, Isac

The sequence SEDGLE... obtained from saliva run in the SDS-PAGE gel and the tripeptide SED on the SGH gel were found in a location with an apparent molecular mass of 48 kDa (Figs 11, 12), matching the previously reported inhibitor of the C3 convertase, Isac (Valenzuela et al., 2000). Isac has a molecular mass of 18.5 kDa but behaves in gel chromatography as though it has a larger molecular mass than predicted (Valenzuela et al., 2000).

## Amino-terminal sequences from salivary proteins matching putative proteins within the metalloprotease reprolysin domain

Two amino-terminal sequences were obtained from the gel used to separate tick saliva that match metalloproteases having the reprolysin domain. These two clones (ISL1324 and ISL1033) were fully sequenced as described above. ISL1033\_cluster65 codes for a 44.1 kDa protein, while ISL1324\_cluster383 codes for a 46.1 kDa protein. The SignalP program does not predict these protein to be secreted. The observed amino-terminal sequences represent unusually distant sites from the starting methionine residue, at positions 49 and 72, predicting mature proteins of 36.7 and 38.2 kDa and compatible with their migration on gels (Fig. 11, Table 3). These proteins may be secreted by a different pathway from the other proteins, perhaps a product of apocrine secretion (Aumuller et al., 1999). They may also be the result of proteolytic processing of a pro-enzyme. It is also possible that both clones are truncated at their 5'-end, where a conserved stretch of 169 residues is sandwiched between the pre- and proproteinase in snake venom metalloproteases (Jia et al., 1996). Indeed, ISL033\_cluster65 is very similar to a hypothetical protein of I. ricinus (gi|5911708), which contains a longer predicted amino-terminal. These metalloproteases may be involved in digestion of skin matrix constituents or fibrinogen, like the hemorrhagic metalloproteases of snake venoms (Leonardi et al., 1999; Tortorella et al., 1998).

# Presence in saliva of the peptide coded by clones TA242, ISL1014 and ISL818

These clones were classified as being of unknown function because they did not produce any significant matches when compared with protein NR or CDD databases. Their aminoterminal sequences, as predicted by the SignalP program, were found in protein bands of saliva separated by SDS-PAGE.

# Calreticulin sequences of SGH proteins

The sequences DPTVYFK... and DPAIYFK..., found in protein bands from SDS-PAGE-separated SGH, match the secreted calreticulin of the tick *Amblyomma americanum* (gi| 3924593) and rat calreticulin (gi| 11693172), respectively (Fig. 11). We have not found any sequence matching

calreticulin in our own library, which appears to be underrepresented for cDNA sequences coding for proteins of molecular mass greater than 50 kDa. These two amino-terminal sequences indicate that calreticulins, abundant intracellular proteins (Nash et al., 1994), are probably produced in *I. scapularis* salivary glands, although their secretory nature is not obvious.

# Housekeeping and other protein sequences found in SGH proteins

The sequence AKDFIAGGVA matches those from cluster 64 with very high similarity to the mitochondrial carrier enzyme ATP/ADP translocase. The sequence MQIFV..., matching the cDNA clone ISL844 from cluster 201, has very high similarity to ubiquitin. The amino-terminal sequence DPIMGYT... was not found in the possible translations of our cDNA library but does match putative oxidoreductases found in the NR protein database. Finally, the sequence NEDLIL... does not match any possible translation product of our cDNA library but does match the SALP17 protein from *I. scapularis* (gi| 15428298) at position 112. The protein sequence ARXDAYDNXSGIRARLH matched clone TB210.

### Discussion

We constructed a PCR-based cDNA library from the salivary glands of the tick I. scapularis, sequenced 735 random clones, clustered the cDNA sequences based on a BLAST algorithm, and obtained full-length information on 87 novel proteins and peptides, most of which appear to be secreted in saliva. Further, we collected information on amino-terminal sequences from proteins from saliva and SGH by SDS-PAGE. We confirmed expression for 19 proteins, including four members of the most abundant cDNA population (cluster 1), two members of another abundant cDNA cluster (cluster 14), two secreted zinc metalloproteases of the reprolysin family (the previously identified anticomplement peptide), and three proteins of unknown function. Several tick-host proteins were found in both saliva and SGH. While the possible function and structure of the sequences obtained are described in Results, two additional items remain to be discussed: (i) observation of a large redundancy of related sequences and (ii) origin of host proteins in saliva and SGH.

Our library contains a remarkably large degree of redundancy, as shown by the many related mRNAs, most of which are too different to be alleles from polymorphic loci. In addition to those shown in Figs 1–10, the previously reported salivary anticomplement protein (gi|8896135) is 82% identical to SALP20 (gi|5428300) (Das et al., 2001). The long evolutionary history of ticks may be responsible for this complex plethora of related proteins. Indeed, when we sequenced similar salivary cDNA libraries from sand flies (Charlab et al., 1999; Valenzuela et al., 2001), and mosquitoes (Valenzuela et al., 2002), we found far less diversity of related molecules. This variability in the tick salivary cDNA library is consistent with the reported high polymorphism of salivary

proteins among individual ticks analyzed by SDS-PAGE (Wang et al., 1999). The adaptive role of this gene-duplication phenomenon may derive from divergence of functions in duplicated genes. For example, a Kunitz-containing protease inhibitor might evolve into another protease inhibitor of different specificity, thus targeting another protease of the host blood-clotting pathway. Another possible adaptive role for gene duplication is the generation of different antigenicity epitopes within molecules of the same function, allowing the tick to better evade host immune responses. It is interesting to speculate whether each of these protein variants would have a differential temporal expression. Because our cDNA library was made from 25 adult female tick salivary glands removed from the tick 3-4 days after host attachment, and because ticks vary up to 2 days in their total feeding time (5-7 days from attachment to a rabbit), it is likely that our library represents an average of messages translated within a broad range of physiologic ages. A microarray experiment with messages obtained from ticks at different times post-attachment could be used to detect individual messages produced at unique times by individual ticks, thus testing the hypothesis of temporal switching of similar salivary proteins in I. scapularis.

With regard to the related messages found in the salivary gland cDNA library of *I. scapularis*, the higher conservation of signal peptides found in peptide groups 2-4, compared with the remaining protein sequences, is remarkable. This pattern was also found in secreted peptide families of vertebrates (Charpentier et al., 1998; Lundwall and Lazure, 1995). Increased evolution of secreted rather than signal peptides indicates possible conservation of a 'secretion signal cassette' or strong evolutionary pressure for variation of the secreted moiety, consistent with an antigenic variation scenario.

This diversity of related salivary proteins, whether they vary from tick to tick or temporally within individual ticks, will certainly pose an additional burden in the attempts to develop a vaccine against tick salivary antigens that may protect against

tick-borne pathogens (Valenzuela et al., 2001). Defining invariant antigens, and/or using a cocktail vaccine approach will be important for a successful vaccine development strategy.

With regard to the finding of host proteins in tick saliva and SGH, we cannot rule out contamination by host blood trapped in the tick mouthparts by tick regurgitation during saliva collection, or by tick-gut contents during salivary gland dissection. Although our cDNA library did not contain a single rabbit sequence match, and the tick mouthparts were thoroughly washed before saliva collection, this does not eliminate the possibility of regurgitation. Host Ig secretion in tick saliva has been reported before in other ticks with Igbinding proteins (IGBP) (Wang and Nuttall, 1995a,b, 1999), and is postulated to be the carrier for this host protein through the tick midgut and salivary gland epithelia. The biological reason for tick IGBP may be related to counteracting the possible noxious effects of host Ig against midgut or hemocoel targets; any other explanation for this seemingly wasteful secretion of host albumin and hemoglobin is not immediately apparent. It is interesting to speculate whether these host proteins are modified by the tick by glycosylation or by other additions. Incorporation of such antigenic epitopes into self molecules may be a strategy for tick suppression of host immunity against potentially antigenic carbohydrate determinants. Further, hemoglobin degradation leads to formation of hemorphins, opioid peptides active in the immune system and in pain reception (Nyberg et al., 1997). Hemoglobin-derived peptides may also have antimicrobial activities (Fogaca et al., 1999).

The functions of most tick sequences described in this paper are unknown. Some, such as group 2, are relatively short peptides with single Kunitz domains (Fig. 2, Table 3). When compared with snake dendrotoxins, which are also small peptides containing a single Kunitz domain (Harvey, 2001), similarities are apparent (Fig. 13) not only in the typical

SGHLLLLLGLLTLWAELTPVSGAAKYCKLPLRIGPCKRK---IPSFY-YK

INECPYGNHHPPGMRVRGTY--

	gi  1097974 prf  2114418A gi  125035 sp P00980 IVBI_DENAN gi  266399 sp P00979 IVBI_DENPO gi  125044 sp P00983 IVBB_DENPO gi  125046 sp P00984 IVBE_DENPO TA149 CLUSTER48	SGHLLLLGLLTLWAELTPVSGAAKYCKLPLRIGPCKRKIPSFY-YK QPRRKLCILHRNPGRCYDKIPAFY-YN QPLRKLCILHRNPGRCYQKIPAFY-YN CPLRKLCILHRNPGRCYQKIPAFY-YS LQHRTFCKLPAEPGPCKASIPAFY-YN IALLPENICRAPHPISSCAPG-AVKETWY-FN
	TB11_CLUSTER26	CIICRAPHAVASCAADIKPKLLFY-FN
	TB192_CLUSTER54	MGRLSEEQ <mark>C</mark> RRPVPSTS <mark>C</mark> ASGVRTI <b>YY</b> - <b>F</b> S
of		Pos
des	gi 385318 gb AAB26998.1	WKAKQCLPFDYSCCGGNANREKTIEECRRTC-VG
des	gi 385318 gb AAB26998.1  gi 1097974 prf  2114418A	WKAKQCLPFDYSGCGGNANRFKTIEECRRTC-VG
des 2 vary	gi 385318 gb AAB26998.1  gi 1097974 prf  2114418A gi 125035 sp P00980 IVBI_DENAN	WKAKQCLPFDYSGCGGNANRFKTIEECRRTC-VG
des 2 vary	gi 385318 gb AAB26998.1  gi 1097974 prf  2114418A gi 125035 sp P00980 IVBI_DENAN gi 266399 sp P00979 IVBI_DENPO	WKAKQCLPFDYSGCGGNANRFKTIEECRRTC-VG WKAKQCLPFDYSGCGGNANRFKTIEECRRTC-VG
des 2 vary of by	gi   385318  gb   AAB26998.1   gi   1097974  prf   2114418A gi   125035  sp   P00980   IVBI_DENAN gi   266399  sp   P00979   IVBI_DENPO gi   125044  sp   P00983   IVBE_DENPO gi   125044  sp   P00983   IVBE_DENPO	WKAKQCLPFDYSGCGGNANRFKTIEECRRTC-VG WKAKQCLPFDYSGCGGNANRFKTIEECRRTC-VG
des 2 2 2 ary of by	gi  385318  gb AAB26998.1   gi  1097974  prf   2114418A gi  125035  sp P00980   IVBI_DENAN gi  266399  sp P00979  IVBI_DENPO gi  125044  sp P00984   IVBE_DENPO gi  125046  sp P00984   IVBE_DENPO Tal49 CLUSTEP48	WKAKQCLPFDYSCCGGNANRFKTIEECRRTC-VG
des 2 2 ary of by ion.	gi  385318  gb  AAB26998.1   gi  1097974  prf   2114418A gi  125035  sp  P00980  IVBI_DENAN gi  266399  sp  P00979  IVBI_DENPO gi  125044  sp  P00983  IVBB_DENPO gi  125046  sp  P00984  IVBE_DENPO TA149_CLUSTER48 TB11_CLUSTER26	WKAKQCLPFDYSCCGGNANRFKTIEECRRTC-VG

dendrotoxins and short pepti from Ixodes scapularis group peptides deduced from a saliv gland cDNA library. Regions positive charge (Pos) are shown by bars. Gly, glycine-rich region. Conserved cysteine residues are shown in black background; other conserved residues are in gray background.

Fig. 13. Clustal alignments

gi 125046 sp P00984 IVBE DENPO TA149 CLUSTER48 TB11 CLUSTER26 TB192 CLUSTER54 ISTA12\_CLUSTER45

gi|385318|gb|AAB26998.1|

Gly Pos

PSLGKC

NNTNOCE -- SYTGCDTGKNRF

conservation of the Kunitz cysteine residues but also in conserved glycine-rich and basic amino acid-rich regions. These peptides may function as dendrotoxins that variously affect membrane functions. These and other peptides are of a size amenable to either direct synthesis or production by recombinant methods, and will eventually be tested for their biological activities in various bioassays. Other biological activities, such as the several antiproteases and metalloproteases, can be identified with different enzyme assays. Our ongoing studies should increase our understanding of how ticks successfully evade the hemostatic and immune responses of their hosts.

The authors thank Drs Robert Gwadz, Louis Miller and Thomas Kindt for encouragement and support, and Nancy Schulman for editorial assistance.

#### References

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410.
- 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. *Nucleic Acids Res.* 25, 3389-3402.
- Armstrong, P. B. and Quigley, J. P. (1999). Alpha2-macroglobulin: an evolutionarily conserved arm of the innate immune system. *Dev. Comp. Immunol.* 23, 375-390.
- Aumuller, G., Wilhelm, B. and Seitz, J. (1999). Apocrine secretion fact or artifact? Anat. Anz. 181, 437-446.
- Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. and Sonnhammer, E. L. (2000). The Pfam protein families database. *Nucleic Acids Res.* 28, 263-266.
- Borth, W. (1992). Alpha 2-macroglobulin, a multifunctional binding protein with targeting characteristics. *FASEB J.* 6, 3345-3353.
- 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.
- Charpentier, S., Amiche, M., Mester, J., Vouille, V., Le Caer, J. P., Nicolas, P. and Delfour, A. (1998). Structure, synthesis, and molecular cloning of dermaseptins B, a family of skin peptide antibiotics. J. Biol. Chem. 273, 14690-14697.
- Dainichi, T., Maekawa, Y., Ishii, K., Zhang, T., Nashed, B. F., Sakai, T., Takashima, M. and Himeno, K. (2001). Nippocystatin, a cysteine protease inhibitor from *Nippostrongylus brasiliensis*, inhibits antigen processing and modulates antigen-specific immune response. *Infect. Immun.* 69, 7380-7386.
- Das, S., Banerjee, G., DePonte, K., Marcantonio, N., Kantor, F. S. and Fikrig, E. (2001). Salp25D, an *Ixodes scapularis* antioxidant, is 1 of 14 immunodominant antigens in engorged tick salivary glands. J. Infect. Dis. 184, 1056-1064.
- Das, S., Marcantonio, N., Deponte, K., Telford, S. R., 3rd, Anderson, J. F., Kantor, F. S. and Fikrig, E. (2000). SALP16, a gene induced in *Ixodes* scapularis salivary glands during tick feeding. Am. J. Trop. Med. Hyg. 62, 99-105.
- Dennis, M. S. and Lazarus, R. A. (1994). Kunitz domain inhibitors of tissue factor-factor VIIa. I. Potent inhibitors selected from libraries by phage display. J. Biol. Chem. 269, 22129-22136.
- Dickinson, R. G., O'Hagan, J. E., Shotz, M., Binnington, K. C. and Hegarty, M. P. (1976). Prostaglandin in the saliva of the cattle tick Boophilus microplus. Aust. J. Exp. Biol. Med. Sci. 54, 475-486.
- Fogaca, A. C., da Silva, P. I., Jr, Miranda, M. T., Bianchi, A. G., Miranda, A., Ribolla, P. E. and Daffre, S. (1999). Antimicrobial activity of a bovine

hemoglobin fragment in the tick *Boophilus microplus*. J. Biol. Chem. 274, 25330-25334.

- Francischetti, I. M., Valenzuela, J. G., Andersen, J. F., Mather, T. N. and Ribeiro, J. M. C. (2002). Ixolaris, a novel recombinant tissue factor pathway inhibitor (TFPI) from the salivary glands of the tick, *Ixodes scapularis*: identification of factor X and factor Xa as scaffolds for the inhibition of factor VIIa/tissue factor complex. *Blood* **99**, 3602-3612.
- Gillespie, R. D., Dolan, M. C., Piesman, J. and Titus, R. G. (2001). Identification of an IL-2 binding protein in the saliva of the Lyme disease vector tick, *Ixodes scapularis. J. Immunol.* **166**, 4319-4326.
- Girard, T. J., Warren, L. A., Novotny, W. F., Likert, K. M., Brown, S. G., Miletich, J. P. and Broze, G. J., Jr (1989). Functional significance of the Kunitz-type inhibitory domains of lipoprotein-associated coagulation inhibitor. *Nature* 338, 518-520.
- Hartmann, S., Kyewski, B., Sonnenburg, B. and Lucius, R. (1997). A filarial cysteine protease inhibitor down-regulates T cell proliferation and enhances interleukin-10 production. *Eur. J. Immunol.* 27, 2253-2260.
- Harvey, A. L. (2001). Twenty years of dendrotoxins. Toxicon 39, 15-26.
- Higgs, G. A., Vane, J. R., Hart, R. J., Porter, C. and Wilson, R. G. (1976). Prostaglandins in the saliva of the cattle tick, *Boophilus microplus* (Canestrini) (Acarina, Ixodidae). *Bull. Ent. Res.* 66, 665-670.
- Hooper, N. M. (1994). Families of zinc metalloproteases. *FEBS Lett.* 354, 1-6.
- Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. and Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* 23, 403-405.
- Jia, L. G., Shimokawa, K., Bjarnason, J. B. and Fox, J. W. (1996). Snake venom metalloproteinases: structure, function and relationship to the ADAMs family of proteins. *Toxicon* 34, 1269-1276.
- Jones, L. D., Davies, C. R., Steele, G. M. and Nuttall, P. A. (1987). A novel mode of arbovirus transmission involving a nonviremic host. *Science* 237, 775-777.
- Jones, L. D., Davies, C. R., Williams, T., Cory, J. and Nuttall, P. A. (1990). Non-viraemic transmission of Thogoto virus: vector efficiency of *Rhipicephalus appendiculatus* and *Amblyomma variegatum. Trans. R. Soc. Trop. Med. Hyg.* 84, 846-948.
- Leonardi, A., Aragon-Ortiz, F., Gubensek, F. and Krizaj, I. (1999). Partial primary structure of a fibrinogenase from the venom of the snake Lachesis stenophrys. *J. Chromatogr. A* **852**, 237-243.
- Lundwall, A. and Lazure, C. (1995). A novel gene family encoding proteins with highly differing structure because of a rapidly evolving exon. *FEBS Lett.* **374**, 53-56.
- Manoury, B., Gregory, W. F., Maizels, R. M. and Watts, C. (2001). Bm-CPI-2, a cystatin homolog secreted by the filarial parasite *Brugia malayi*, inhibits class II MHC-restricted antigen processing. *Curr. Biol.* 11, 447-451.
- Milstone, A. M., Harrison, L. M., Bungiro, R. D., Kuzmic, P. and Cappello, M. (2000). A broad spectrum Kunitz type serine protease inhibitor secreted by the hookworm *Ancylostoma ceylanicum*. J. Biol. Chem. 275, 29391-29399.
- Mulenga, A., Sugimoto, C. and Onuma, M. (1999). Characterization of proteolytic enzymes expressed in the midgut of *Haemaphysalis longicornis*. *Jpn. J. Vet. Res.* 46, 179-184.
- Nash, P. D., Opas, M. and Michalak, M. (1994). Calreticulin: not just another calcium-binding protein. *Mol. Cell. Biochem.* 135, 71-78.
- 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.
- Nuttall, P. A., Paesen, G. C., Lawrie, C. H. and Wang, H. (2000). Vectorhost interactions in disease transmission. J. Mol. Microbiol. Biotechnol. 2, 381-386.
- Nyberg, F., Sanderson, K. and Glamsta, E. L. (1997). The hemorphins: a new class of opioid peptides derived from the blood protein hemoglobin. *Biopolymers* **43**, 147-156.
- Ogata, S., Hayashi, Y., Misumi, Y. and Ikehara, Y. (1990). Membraneanchoring domain of rat liver 5'-nucleotidase: identification of the COOHterminal serine-523 covalently attached with a glycolipid. *Biochemistry* 29, 7923-7927.
- Pitt-Rivers, R. and Impiombato, F. S. A. (1968). The binding of sodium dodecyl sulphate to various proteins. *Biochem. J.* 109, 825-830.
- Petersen, L. C., Bjorn, S. E., Olsen, O. H., Nordfang, O., Norris, F. and Norris, K. (1996). Inhibitory properties of separate recombinant Kunitztype-protease-inhibitor domains from tissue-factor-pathway inhibitor. *Eur. J. Biochem.* 235, 310-316.

- Regoli, D. A. and Barabe, J. (1980). Pharmacology of bradykinin and related kinins. *Pharmacol. Rev.* **32**, 1-46.
- Ribeiro, J. M. and Mather, T. N. (1998). *Ixodes scapularis*: salivary kininase activity is a metallo dipeptidyl carboxypeptidase. *Exp. Parasitol.* 89, 213-221.
- Ribeiro, J. M. C. (1987). Ixodes dammini: Salivary anticomplement activity. Exp. Parasitol. 64, 347-353.
- Ribeiro, J. M. C. (1989). Role of saliva in tick/host associations. *Exp. Appl. Acarol.* 7, 15-20.
- Ribeiro, J. M. C. (1995). Blood-feeding arthropods: Live syringes or invertebrate pharmacologists? *Infect. Agents Dis.* 4, 143-152.
- Ribeiro, J. M. C., Evans, P. M., MacSwain, J. L. and Sauer, J. (1992). *Amblyomma americanum*: Characterization of salivary prostaglandins E2 and F2alpha by RP-HPLC/bioassay and gas chromatography-mass spectrometry. *Exp. Parasitol.* **74**, 112-116.
- Ribeiro, J. M. C., Makoul, G., Levine, J., Robinson, D. and Spielman, A. (1985). Antihemostatic, antiinflammatory and immunosuppressive properties of the saliva of a tick, *Ixodes dammini. J. Exp. Med.* 161, 332-344.
- Ribeiro, J. M. C., Makoul, G. and Robinson, D. (1988). *Ixodes dammini*: Evidence for salivary prostacyclin secretion. J. Parasitol. 74, 1068-1069.
- Ribeiro, J. M. C. and Spielman, A. (1986). Ixodes dammini: Salivary anaphylatoxin-inactivating activity. Exp. Parasitol. 62, 292-297.
- Ribeiro, J. M. C., Weis, J. J. and Telford, S. R., III. (1990). Saliva of the tick *Ixodes dammini* inhibits neutrophil function. *Exp. Parasitol.* 70, 382-388.
- Roberts, D. D. (1996). Regulation of tumor growth and metastasis by thrombospondin-1. FASEB J. 10, 1183-1191.
- 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. *Nucleic Acids Res.* 28, 231-234.

- Tortorella, M. D., Pratta, M. A., Fox, J. W. and Arner, E. C. (1998). The interglobular domain of cartilage aggrecan is cleaved by hemorrhagic metalloproteinase HT-d (atrolysin C) at the matrix metalloproteinase and aggrecanase sites. *J. Biol. Chem.* **273**, 5846-5850.
- Valenzuela, J. G., Belkaid, Y., Garfield, M. K., Mendez, S., Kamhawi, S., Rowton, E. D., Sacks, D. L. and Ribeiro, J. M. (2001). Toward a defined anti-*Leishmania* vaccine targeting vector antigens: characterization of a protective salivary protein. J. Exp. Med. 194, 331-342.
- Valenzuela, J. G., Charlab, R., Mather, T. N. and Ribeiro, J. M. (2000). Purification, cloning, and expression of a novel salivary anticomplement protein from the tick, *Ixodes scapularis*. J. Biol. Chem. 275, 18717-18723.
- Valenzuela, J. G., Pham, V. M., Garfield, M. K., Francischetti, I. M. and Ribeiro, J. M. C. (2002). Toward a description of the sialome of the adult female mosquito Aedes aegypti. Insect Biochem. Mol. Biol., in press.
- Wang, H. and Nuttall, P. A. (1995a). Immunoglobulin G binding proteins in male *Rhipicephalus appendiculatus* ticks. *Parasite Immunol.* 17, 517-524.
- Wang, H. and Nuttall, P. A. (1995b). Immunoglobulin-G binding proteins in the ixodid ticks, *Rhipicephalus appendiculatus, Amblyomma variegatum* and *Ixodes hexagonus. Parasitology* 111, 161-165.
- Wang, H. and Nuttall, P. A. (1999). Immunoglobulin-binding proteins in ticks: new target for vaccine development against a blood-feeding parasite. *Cell. Mol. Life Sci.* 56, 286-295.
- Wang, H., Kaufman, W. R. and Nuttall, P. A. (1999). Molecular individuality: polumorphism of salivary gland proteins in three species of ixodid tick. *Exp. Appl. Acarol.* 23, 969-975.
- Wikel, S., Ramachandra, R. N. and Bergman, D. K. (1994). Tick-induced modulation of the host immune response. *Int. J. Parasitol.* 24, 59-66.
- Wikel, S. K. (1996). Host immunity to ticks. Ann. Rev. Entomol. 41, 1-22.