Research article

Ixodes scapularis tick <u>ser</u>ine <u>p</u>roteinase <u>in</u>hibitor (serpin) gene family; annotation and transcriptional analysis

Albert Mulenga*, Rabuesak Khumthong and Katelyn C Chalaire

Address: Department of Entomology, Texas A&M University, 2475 TAMU, Minnie Belle Heep center, College Station, Texas 77843, USA

Email: Albert Mulenga* - a-mulenga@tamu.edu; Rabuesak Khumthong - rabuesak@gmail.com;
Katelyn C Chalaire - katelynchalaire@neo.tamu.edu

* Corresponding author

Abstract

Background: Serine proteinase inhibitors (Serpins) are a large superfamily of structurally related, but functionally diverse proteins that control essential proteolytic pathways in most branches of life. Given their importance in the biology of many organisms, the concept that ticks might utilize serpins to evade host defenses and immunizing against or disrupting their functions as targets for tick control is an appealing option.

Results: A sequence homology search strategy has allowed us to identify at least 45 tick serpin genes in the *Ixodes scapularis* genome that are structurally segregated into 32 intronless and 13 intron-containing genes. Nine of the intron-containing serpins occur in a cluster of 11 genes that span 170 kb of DNA sequence. Based on consensus amino acid residues in the reactive center loop (RCL) and signal peptide scanning, 93% are putatively inhibitory while 82% are putatively extracellular. Among the 11 different amino acid residues that are predicted at the P1 sites, 16 sequences possess basic amino acid (R/K) residues. Temporal and spatial expression analyses revealed that 40 of the 45 serpins are differentially expressed in salivary glands (SG) and/or midguts (MG) of unfed and partially fed ticks. Ten of the 38 serpin genes were expressed from six to 24 hrs of feeding while six and fives genes each are predominantly or exclusively expressed in either MG and SG respectively.

Conclusion: Given the diversity among tick species, sizes of tick serpin families are likely to be variable. However this study provides insight on the potential sizes of serpin protein families in ticks. Ticks must overcome inflammation, complement activation and blood coagulation to complete feeding. Since these pathways are regulated by serpins that have basic residues at their P1 sites, we speculate that *I. scapularis* may utilize some of the serpins reported in this study to manipulate host defense. We have discussed our data in the context of advances on the molecular physiology of *I. scapularis*. Although the paper is descriptive, this study provides the first step toward a comprehensive understanding of serpins in tick physiology.

Background

Ticks, segregated into two families; Ixodidae (hard ticks) and Argasidae (soft ticks) are important vectors of several pathogens with a global veterinary and public health impact [1,2]. Although research on ticks has for the most part been directed towards agricultural interests, ticks

have been recognized as the second most important vectors of human disease agents after mosquitoes [2]. Globally, the impact of tick borne disease agents on public health has been on a dramatic rise [3-6] since the discovery of *Borelia burdgoferi* as the causative agent for Lyme disease in 1982 [7,8]. Literature reviews by Parola and Roult [3] listed 15 new tick borne bacterial agents being discovered or recognized as human pathogens between 1982 and 2004. In the United States ticks transmit more causative agents of vector borne diseases than any other vector arthropod [9]. Tick borne human diseases reported in the USA include Babesiosis, ehrlichisosis, Southern Tick-Associated Rash illness (STARI), Lyme disease, tick-borne relapsing fever, anaplasmosis and Rocky Mountain spotted fever, [9]. From a human health perspective, *Ixodes scapularis* and its close relatives, *I. pacificus, I. ricinus* and *I. persulcatus* and *I. holocyclus* are the most important ticks as they transmit the majority of emerging human disease pathogens. The importance of these ticks to human health was the key justification for funding of the *I. scapularis* genome, sequencing project [10]. Key anticipated outcomes of the tick genome sequencing project will be provision of opportunities to identify unique tick genes that could be exploited for tick control and, thus the control of tick borne diseases [10]. We are interested in understanding the role of the serine proteinase inhibitors (serpin) in tick physiology and feeding.

Serpins represent the largest family of proteinase inhibitors that is widely conserved across taxa, from animals to plants, viruses and bacteria [11-18]. Of the 68 families of proteinase inhibitors listed on the MEROPS database (version 7.6, http://merops.sanger.ac.uk/, [19]), the serpin family (denoted as I4) has the largest number of entries. In humans, serpins make up 2% of total blood plasma proteins [20] and are involved in the regulation of important pathways such as blood coagulation, complement activation, inflammation, fertilization and food digestion [11-14]. The importance of serpins in humans can be attested to by more than 90 human diseases such as cirrhosis, emphysema, blood coagulation disorders and dementia that arises from serpin malfunctions due to mutation [12]. In arthropods, serpins were linked to immunity in mosquitoes [21,22], the fruit fly [23-26] and the tobacco hornworm [27], development in the fruit fly [28], control of the hemolymph coagulation cascade in the horseshoe crab [29-31]. Given the importance of serpins in the biology of multicellular organisms, it has been hypothesized that, ticks might utilize serpins to evade host defenses and immunizing against or disrupting functions of these proteins is an appealing option for designing new tick control strategies [17,32].

More than 30 serpin encoding cDNAs have now been cloned in several economically and medically important

ticks including *Amblyomma americanum* [33], *R. appendiculatus* [17] and *I. ricinus* [34] and *I. scapularis* [35]. Serpin encoding ESTs and cDNA sequences of *I. scapularis* and *R. microplus* cited in [33] are also available in GeneBank. Studies by Sugino et al., [36], Imamura et al., [37] and Prevot et al., [38] have provided encouraging evidence suggesting the potential of serpins as targets for tick control. These authors [36-38] showed that feeding of ticks on animals immunized with recombinant tick serpins caused ticks to obtain smaller blood meals and as consequence reduced tick fecundity and mortality. These findings clearly suggest that serpins play important roles in tick physiology.

Given the availability of *I. scapularis* genome sequence information in GeneBank and at VectorBase [38] we thought to get some insight on the size of the serpin protein family in ticks. We report here on the identification and characterization at least 45 *I. scapularis* serpin genes. A sequence based analyses show that ~93% of these serpins are putatively inhibitory with 88% (31/35) of full-length serpins being putatively extracellular. Our RT-PCR expression analyses data demonstrated that 84% (38/45) are differentially expressed in midguts and salivary glands of unfed and partially fed ticks. We have discussed our findings towards advances in *I. scapularis* tick molecular biology.

Methods
Discovery of Ixodes scapularis *serpin genes*
The strategy to annotate serpin genes in the *I. scapularis* genome is summarized in figure 1. In the **first step**, we downloaded supercontig DNA sequences from Vector-Base [39,40] and created a local database. In the **second step**, the supercontig database was subjected to BLASTX homology search using the FASTA3 software version 3.4 [41] to identify serpin-encoding supercontigs. Queries for this search, included α_1-antitrypsin and 30 tick serpin sequences; 17 *A. americanum* [33], one *H. longicornis* [36], four *R. appendiculatus* [17], six *I. ricinus* (DQ915842, DQ915843, DQ915844, DQ915845 and AJ269658), and two in *I. scapularis* (AAM93649 and AAV80788 [35]). In the **third step**, coding genomic DNA sequences were assembled using GenomeScan analysis [42] and the VectorNTI DNA analysis software (Invitrogen, Carlsbard, CA; version 10, academic free license). GenomeScan analysis [42] was used to align supercontig sequences with tick serpin protein sequences that were used as queries. This alignment identified exons, which were manually assembled into genomic coding sequences using the VectorNTI DNA analysis software (Invitrogen, Carlsbard, CA; version 10, academic free license). In **step four** cDNA sequences were assembled. This was done by scanning genomic DNA coding sequences from step three against the *I. scapularis* trace archive EST database. Trace archive ESTs
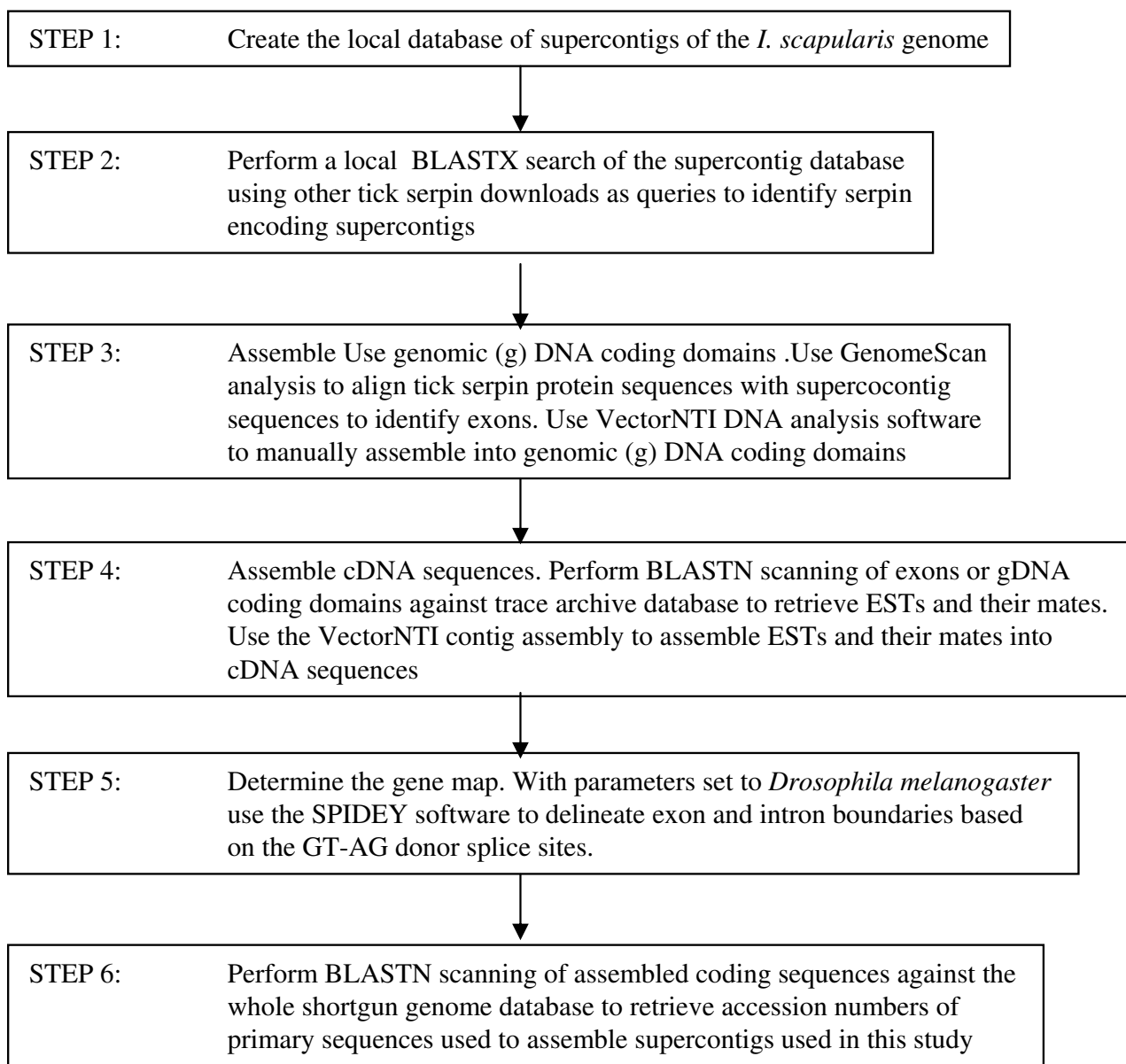
STEP 1:	Create the local database of supercontigs of the *I. scapularis* genome

STEP 2:	Perform a local BLASTX search of the supercontig database using other tick serpin downloads as queries to identify serpin encoding supercontigs

STEP 3:	Assemble Use genomic (g) DNA coding domains .Use GenomeScan analysis to align tick serpin protein sequences with supercocontig sequences to identify exons. Use VectorNTI DNA analysis software to manually assemble into genomic (g) DNA coding domains

STEP 4:	Assemble cDNA sequences. Perform BLASTN scanning of exons or gDNA coding domains against trace archive database to retrieve ESTs and their mates. Use the VectorNTI contig assembly to assemble ESTs and their mates into cDNA sequences

STEP 5:	Determine the gene map. With parameters set to *Drosophila melanogaster* use the SPIDEY software to delineate exon and intron boundaries based on the GT-AG donor splice sites.

STEP 6:	Perform BLASTN scanning of assembled coding sequences against the whole shortgun genome database to retrieve accession numbers of primary sequences used to assemble supercontigs used in this study

Figure 1
Flow chart of the strategy used to annotate *Ixodes scapularis* serpins. The FASTA3 version 3.4 software [41] was used to perform a local BLASTX search of the local database of supercontigs of the *I. scapularis* genome sequence data using tick serpin downloads from GeneBank as queries.

showing e-values of zero (probability that the observation was not by chance) and their mates were retrieved and assembled using ContigExpress (Vector NTI package version 10). Prior to assembly any vector sequence contamination was removed using vector screen at NCBI. In **step five**, exon and intron boundaries were delineated based on donor-acceptor splice sites (GT-AG). This was accomplished by mapping cDNA sequences onto appropriate supercontig sequences using the Spidey software [43] at

NCBI with parameters set to *Drossophila melanogaster*. In **step six**, coding domains were scanned against the whole shotgun genome database to retrieve accession numbers for primary sequences that were used to assemble the supercontigs used in this study.

Bioinformatics analyses
For comparisons to known serpins, assembled coding sequences were scanned against known protein entries in

GenBank using the BLASTx or BLASTp homology search program. To validate their accuracy, inferred amino acid sequences were inspected for two amino acid motifs, "NAVYKFG" and "DVNEEG," that are conserved in most known serpins [30,44]. Full-length and partial sequences were determined on the basis that, a typical serpin molecule ranges between 350–450 amino acid residues long [11,12]. To gain insight on probable functionality, inferred amino acid sequences were scanned against amino acid motif entries in ScanProsite and signal peptides in the SignalP servers (ExPASY Proteomics Server http://ca.expasy.org/). The reactive center loop (RCL) which determines the functionality (inhibitory and non-inhibitory) of a serpin molecule was determined based on consensus 20/21 residue peptide "p17 [E]-p16 [E/K/R]-p15 [G]-p14 [T/S]-p13 [X]-p12-9 [AGS]-p8-1 [X]-p1' -4' " in carboxy-terminus region [11,45]. The numbering here is based on the convention in which residues on the amino terminal side of scissile bond (p1-p1') are labeled as p17 to p1 and those on the carboxy terminal side are labeled as p1' -p4' [46]. The putative scissile bond (p1-p1') and the p1 residue were predicted based on the conserved features that there are generally 17 amino residues (p17 to p1) between the start of the hinge region of the RCL, and scissile bond [11].

Guide phylogeny tree

In order to determine relationships among tick serpins, a guide tree of 68 tick serpins sequences, 34 full-length serpins from this study, and 34 other tick serpins that were downloaded from GeneBank, and human α_1-antitrypsin (out group, AAB59495) as an outlier were used to construct the guide tree using the neighbor joining method. Specifications were set for bootstrap values at 1000 replications, gaps proportionately distributed and correction for distance set to a Poisson distribution. To determine the homologous features that influenced the clustering patterns, sequences of each clade were subjected to multiple sequence alignment analyses.

Structural based sequence alignment and comparative modeling

To determine secondary structures, randomly selected representative sequences of each cluster in the guide phylogeny tree, serpins S1, 19, 22, 34 and 35 were subjected structural based sequence alignment using the web based STRAP (**S**tructure based **A**lignment **P**rogram) http://www.charite.de/bioinf/strap/. Subsequently, secondary structures were superimposed on the structurally aligned sequences using the human α_1-antitrypsin tertiary structure (1HP7, [47]) as template. The aligned sequences were subsequently viewed using the GeneDoc sequence analysis software for windows http://www.nrbsc.org/gfx/genedoc/index.html.

Expression analyses by RT-PCR

OligodT primed salivary gland (SG) and midgut (MG) cDNA templates of unfed ticks and ticks that were fed for 6, 24, 48, 72 and 96 hrs used in this study were a gift from Dr. Shahid Karim, formerly at the University of Rhode Island, now at University of Mississippi. The harvest of tick tissues and synthesis of cDNA were done as previously described by Kotsyfakis et al., [48]. One µl aliquots of these templates were used in a PCR reaction with specific PCR primers (Tables 1 and 2). PCR products were electrophoresed on a 2% agarose gel containing 1 µg/ml ethidium bromide. The tick actin gene (Table 1) was used for PCR template load control. To determine mRNA abundance, densitograms of amplified PCR bands were determined using the web based ImageJ software http://rsb.info.nih.gov/ij/. To correct for variations in PCR template concentrations, PCR band densities were normalized using the following formula; $Y = V + V (H-X)/X$ where Y = normalized mRNA density, V = observed Ispin PCR band density in individual tissues (MG and SG), H = highest tick actin PCR band density among tested tissues (carcass in this case 24 hrs MG), X = tick actin density in MG or SG [33].

Results
Discovery of I. scapularis *serpin genes*
We successfully used homology search engines, Fasta3 version 3.4 [41], BLASTX and BLASTN to scan supercontigs of the *I. scapularis* genome and annotated 45 serpin genes in the *I. scapularis* genome (Table 3). Of the 45 genes, serpins (S) 1 to 11 occur in a cluster spanning 170 kb of DNA sequence (figure 2) and the rest occur singly or in pairs. When scanned against the *I. scapularis* trace archive cDNA database, 72% (32/45) of the serpin coding regions showed matches in the EST database with e-values of zero (probability that this observation was not by chance) (Table 3). The delineation of introns and exons boundaries based on Spidey [43] and Genome Scan [42] programs revealed that, the 45 serpin genes are structurally segregated into 32 intronless and 13 intron-containing genes. The intron-containing genes, have four exons with the first and second exon coding for the 5' untranslated region (UTR) and the third and fourth exons coding for the open reading frame and the 3' UTR (figure 3). Of the 13 intron-containing genes, nine belong to the cluster of 11 genes shown in figure 2.

As summarized in table three, all I. scapularis serpin sequences showed best matches exclusively to other tick serpin sequences. Except for S5, 9, 17 and 40 that respectively showed 96, 97, 99, 96 and 95% amino acid identity to I. ricinus serpin (Irs) 4 (accession [acc] # DQ915844), Irs1 (acc# DQ915842), Irs8 (acc# DQ915845), the I. ricinus immunosuppressant serpin (AJ269658, [34]) and Irs2 (acs# DQ915843), the identity levels between serpins in

Table 1: Gene specific PCR primers to amplify 16 serpins that have basic residues at their P1 site

Serpin gene	Forward primer	Reverse
S1	CAAAGCTGATACAAGACAAGACCAAG	TTAAAGATTATTGACCTGTCCAGCG
S2	AGTGAAGGAACGAAGAGCTCGAC	TTGTGGAGACATCTGAGATCTTCAGA
S4	AACAGCGATGGGAATGGTTTTC	TCGTTGGGCACAGTTAGGGTC
S6	TACCAAGATGAAAGTTCTCGTGACG	CTTGGCTCAAAGGTGGTTCACTT
S7	ATGAAAGTTATCACTGCGTTCCTGT	CTTGGTTGTCTCAGAGGTGGTTCA
S8	TCAAACCTTGTTTACTTGCCCCA	TCAAACCTTGTTTACTTGCCCCA
S9	TCTTCTCGCCCTACAGCATATCC	TGTTTTGAAGCCTGGCGATTGT
S12	ATGTTCGCTCCGACTGCCCT	GTTAATTTTTCACGTTCTCAAGCCTG
S14	ATGAAGGCTTCAACGGTGTGG	GGTAATAAAAGTGGTCGCACACAATG
S16	GGTGAGTCAAGATGGCAAAATGCC	GTTTACCTGTCCTGCAAACAGCACC
S17	GAGACGAGACAGACCAATTACCTCA	GGATGACTCAGAGGGCGTTGAT
S18	TATCACAACCTCCACGATGTTCG	GTTAATTTCTCACGTTCTCAAGCCTG
S19	ATGTGGTTCCCGGCGCTCCT	CGTCCTCGATGTATTCACAAGTCCAGC
S21	CTGTTCCTCCTGATGGCGGTA	ATTTCTCACGTTCTCAAGCCTGC
S26	ATGCTTGCGGAAAAGGTTTTC	CTGGACTTCTCCCAAAAATAGCG
S27	AAGCTAGATCGACGCCATGAAG	CTAAACCCGCTTCAACATGCTG
S32	GCATCCAGAGCTTCCAAACATG	TCAATGTGATAGGAAGCACTTCGAG
Tick Actin	GGACAGCTACGTGGGCGACGAGG	CGATTTCACGCTCAGCCGTGGTGG

this study and other tick serpins ranged between 34 to 57% (Table 3). Two of the serpins genes in this study, S5 and 7 are 99 and 98% identical to previously annotated I. scapularis serpins, AAM93649 and AAV80788 respectively [35] (Table 3). At the time of preparing this manuscript, a preliminary annotation of the I. scapularis genome (version 0.5) was released http://www.vectorbase.org/index.php. When scanned against this database, 88% (40/45) of the serpin sequences produced matches with e-values of zero. The reader is notified here that because neither supercontig accession numbers nor the preliminary gene annotations are referenced in GeneBank, whole genome shotgun sequence and/or trace archive accession numbers are given in table three to provide the source for primary sequence data that was used in this study.

Inferred amino acid sequence analyses
A typical full-length serpin molecule may range between 350-to-450 amino acid residues long [11,14]. On this basis and in addition to possessing in frame start and stop codons, 35 of the 45 inferred serpin sequences were determined to be full-length (Table 3). As indicated in table three, eight of the ten partial sequences are truncated at their amino terminus, while the remaining two sequences are truncated at their c-terminal. At their c-terminus, serpins are characterized by a flexible reactive center loop (RCL), the primary determinant of functionality (inhibitory or non-inhibitory) that acts as a pseudo-substrate for the target proteinase [11,42]. Based on consensus amino acid residues [11,14,45], RCLs have been detected in all sequences (figure 4), except for serpin 25 and 28 that are truncated at their c-terminus (Table 3). On the basis that

Table 2: Gene Specific PCR primers used to amplify transcripts of *I. scapularis* serpins that have non-basic residues their P1 sites

Serpin gene	Forward primer	Reverse primer
S3	ATGAAGTACCTGGTTACGTTCCTC	TTAGAGCTTGTTGACCTGTCCCGC
S5	ATGAGGTCTCTCGCAACGTTCATG	TCAAAGATGATTGACCTGTCCCAC
S10	CGGCCTCAACCTCCTCAGGGAGC	CAGCGATTGATCTACCGCGTCCC
S11	CGACTGCAGCTGCGAGTGCCC	CTGCTCTTTCTTCTGTCACCTC
S13	CTCCGTGGAAGTCGCTATGGTC	GGTGTTGGCAAGTCTGTCCCTG
S15	ATGTCTCCCCGAGGTACCTTTCTC	TTTATGAACCGCACCAATAAAGTG
S20	ATGAAACGCTGCACCCTGGTCGC	TCACACCTCTTGAAGCCTTCCGAG
S22	ATGTACGCTCCGACTGTACTGTTCC	TCAAGCTTGCAACTGGTGCACCTC
S23	CGCCATGGTCTACGCCGGCGCC	CACACGGCCGGTTGCCTTCTCGC
S24	CAGCGAACAGCCGGAATCGAAC	GCCTATTAACGGCGAAGGATGG
S26	ATGCTTGCGGAAAAGGTCTTCTTC	CTTGCCTCTGGTCTTGTTCTTCAC
S28	GTTTGCCGGAAAGGTGTTTCTCG	GTGGCCGAGGGCAGTGGAAAG
S29	ATGATGCTTGCCGAAAAGGTTTTTC	CTAGCCAAGGGCGGGCTTAACGGC
S30	ATGGCTTCCGATTTTGGCGATTC	GATTTGGCGTACAGAGCCAAG
S31	GCCGAACTCCGGGACTCCATG	CACTTCTCCCAAAAATAGCACACG
S33	CTCGAGCTTCCTTTGCAAGCTACCC	GGATGGATCCCATGAAGAGAACAG
S34	ATGGCGATATTACCTGCTCTGAC	GCATGTCGGAGACGAGGTTCCTG
S35	ATGGACGCGGTTTCGAGATTCCTG	CTATAGGTGACGCACCGAACCC
S36	ATGATGTTGGGCCGTCTTGCAGC	TTACAGCTTGTTGACTTGACCAGC
S37	GGAGGTCAATGCCTGGGTCGAG	GCTGCAGATCAGGAACATGAACG
S38	ATGGCTATGGCCTACGCTGGAG	CTAGAGTGCGTTGATATGTCCCAC
S39	ATGGCGTCTGATTTTAGTAATTCTC	TCAGATTTGGGCTACAGAGCCAAT
S40	TCTGAAGGAATTTCTCCAGAAGTTG	TCAAAGCATGTTTACTTGCCCC
S41	ATGAAGACTCTGGCAGCATTCCTG	TTAAGCGGCCGCCTCGGTGCCTTC
S42	CTTTCCTCAAGGTTCAAGTTTGAGAC	CTTTCCTCAAGGTTCAAGTTTGAGAC
S43	GAAACCATGACGAGACGCGGTAAG	GCCGGTTGCCTTCTCGCGGATGTAG
S44	GTCGAGAGCATAGAAAAAATCCAGG	GAGACGGCCGGTAGCCTTCTCGCG
S45	CGCCTCTAGCGAGCGATTTCCTC	GACACCTTCCATTCCGGCGTCGG

inhibitory serpins have more than 50% of "A" between p12-9 and that charged amino acid residues at p14 or "P" residues at p12 or p10 in the RCL, causes loss of inhibitory function [11], 41 of the 43 serpins that have a RCL are predicted to have inhibitor functions (figure 4). In the RCL the P1 residue determines substrate specificity [11,45]. By consensus there are generally 17 amino acid residues (p17 to p1) between the start of the hinge region of the RCL, and scissile bond [p1-p1'] [11,49]. On this basis, visual inspection of putative RCL amino acid sequences revealed a diversity of 11 different residues K, R, L, G, A, S, E, C, Y, I and M at the P1 sites (figure 3). As shown in figure 2, 16 of the 43 serpin, sequences that have RCLs are predicted to possess basic charged amino acid residues, "R" or "K" at their P1 site (figure 4). Of the 16 serpin sequences, seven belong to the cluster of 11 serpin genes that are shown in figure 1. It is also interesting to note that eight other serpin genes possess basic residues at their P2 site.

Amino acid motifs scan analyses
Motif scan analyses on the ScanProsite server [50] revealed that the serpin signature motif pattern PS00284 ([LIVMFY]-G-[LIVMFYAC]-[DNQ]-[RKHOS]-[PST]-F-[LIVMFY]-[LIVMFYC]-X-[LIVMFAH] was present in all sequences (results not shown). On the SignalP3.0 server http://www.cbs.dtu.dk/services/SignalP/, 89% (31/35) of the 35 full-length serpin sequences were predicted to have leader sequences (Table 3). Other important motifs include the putative N-glycosylation (NX [S/T]) sites in 38 of the 45 serpin sequences. Except for S11, which was predicted to have eight putative N-glycosylation sites, the rest have between one to three sites. We also found the cell attachment motif "RGD" in four sequences, S1, 7,17 and 42.

Guide phylogeny analysis
In order to determine the relationship among tick serpins, we used the neighbor joining method to generate a phylogeny guide tree of 68 tick serpin sequences. From the α_1-antitrypsin outlier, 68 tick serpin sequences segregated into five main clusters, A-E supported by 93, 90, 96, 88 and 93% bootstrap values (figure 5). Except for Bmserp1, all other sequences in main branch "A" have no signal peptides. The remaining sequences, all of which have signal peptides are segregated into branches "B-E".

I. scapularis *serpins posses key residues that characterize the secondary structure the serpin superfamily*
Irving et al., [51] performed a sequence alignment of 219 serpin sequences and identified 51 core amino acid residues that underlie the functionality of inhibitory serpins. Here we adopted this analysis by aligning the 35 full serpin sequence with the human α_1-antitrypsin (accession # AAB59495) and found that the 51 core residues that underpin the efficient functioning of an inhibitory serpin [51] are 67–96% (34–49/51) conserved in *I. scapularis* serpins (figure 6). The inhibitory mechanism of the serpins relies on their structural flexibility, which is controlled by the shutter region [11,52]. Two amino acid motifs S^{53}-P^{54}-X^{55}-S^{56} and I^{157}-N^{158}-X^{159}-X^{160}-V^{161} (numbering is based on the serpin human α_1-antitrypsin) that underpin the functioning of the shutter region [11,52] are 100% conserved in 20 of the 34 *I. scapularis* serpin sequences (figure 6).

Given the high conservation of core residues that underpin the functionality of serpins, we thought to determine whether or not the serpin common fold of 7–9 α-helices and three β-sheets (A-C) and the RCL [12] were also con-
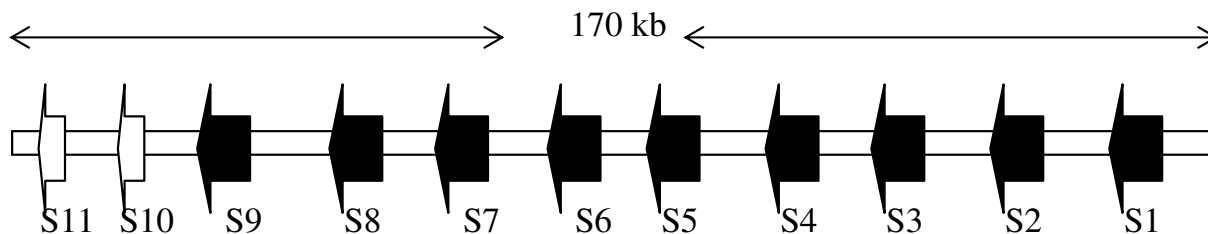
Figure 2
A cluster of 11 serpin (S) genes (S1-to-11) spanning 170 kb of DNA sequence. Structurally, nine serpin genes, S1-to-9 have introns, while S10 and 11 are intronless. When scanned against the whole genome shotgun DNA sequence database, encoding these open reading frames of these genes are ABJB010004033 for S1, ABJB010065261 for S2, ABJB010418850, ABJB011135496, ABJB0109755250 and ABJB010321224 for S3, ABJB011011198 for S4, ABJB010283366 for S5, ABJB011103238 and ABJB010459188 for S6, ABJB010459188 for S7, ABJB010459188 for S8 and ABJB010543279 for S9, 10 and 11. Based on genomic DNA sequence, complete open reading frames (ORF) were detected for all serpins except for S11 and S6, which are partial and their ORFs were assembled from EST sequences indicated in Table 3. Open arrow heads ($\Leftarrow$) denote intronless, closed arrow heads ($\Rightarrow$) denote intron-containing genes.

Exon 1   Exon 2   Exon 3   Exon 4

2.6 kb   0.8 kb   5.7 kb

53 nucleotides   53 nucleotides   881 nucleotides   410 nucleotides
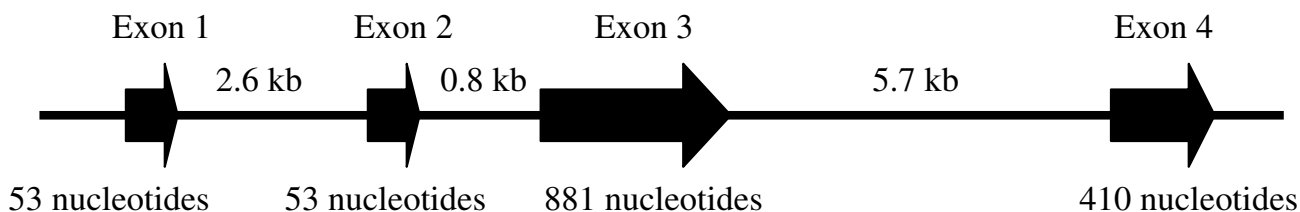
Figure 3
Typical structure of intron-containing serpin genes. The cDNA sequence that were assembled from trace archive EST sequences were aligned with supercontig DNA sequences using the Spidey mRNA and genomic DNA alignment software at NCBI [43] and identified the donor-acceptor splice sites (GT-AG). Displayed here is S3 whose cDNA was assembled from trace archive EST and its mate, accession #s 1446657576/1446661330 respectively. The cDNA was mapped onto genomic DNA sequences corresponding to whole genome shotgun sequence (ABJB010418850, ABJB011135496, ABJB0109755250 and ABJB010321224). The first and second exons encode the 5' untranslated region (UTR) and while the third and fourth exon encode the open reading frame, and the 3' UTR.

served in the deduced serpin sequences. As shown in figure 6, structural based sequence alignment of representative sequences (S1, 19, 22, 34 and 35) from each of the five main branches shown in figure 4, revealed that the serpin secondary structure is conserved in *I. scapularis* serpins (figure 7).

Transcriptional analyses
To begin determining the role of serpins in this study in tick feeding and physiology, semi-quantitative RT-PCR was used to establish temporal and spatial expression patterns in salivary glands (SG) and midguts (MG) of unfed and ticks that were partially fed for 6, 24, 48, 72 and 96 hrs (figure 8). Of the 45 serpin sequences, 40 are expressed in SG and MG; 34 being strongly expressed (figures 8A and 8B) and five weakly expressed (not shown). For clarity and limitations on space, we have presented our RT-PCR data in two panels. Panel one (figures 8A and 8B) show expression patterns of serpin sequences that possesses basic residues (R/K) at their P1 sites, while expression patterns of the rest of the genes are displayed in panel two (figures 8C and 8D). On the basis of visual intensity of PCR bands (figure 8A and 8C) and analysis of normalized PCR band densities (figure 8B and 8D), expression patterns of the 35 genes can be segregated into four profiles. In the first pattern, six genes (S1, 2, 4 and 7 in panel one, S3, 28, 38 and 44 in panel two) whose transcript increased with tick feeding were expressed from six or 24 hrs of tick feeding. The next set of genes, are constitutively expressed in both MG and SG with their transcript abundance either increasing (S17 and 32 in panel one, 23 and 37 in panel two) or decreasing (S19 in panel two) as ticks continued to feed. The third group of genes S6, 9, 12, 16, 21, 26 and 27 in panel one and S20, 25 and 40 in panel two displayed a dichotomous expression pattern. In the case of S6, it is expressed from 24 hrs in SG but constitutively expressed in MG with its transcript abundance increasing in both organs as ticks continue to feed. In the

case of serpin S16 and 26, which are induced from 24 hrs in SG and their transcript abundance increasing thereafter, are constitutively expressed in MG with transcript abundance decreasing as ticks continue to feed. For S9 and 27 in panel one, S20, 25, 30, 33, 40 and 41 in panel two, they are constitutively expressed in both organs, but their transcript abundance increases in SG and decreases in MG as ticks continue to feed. The fourth profile consist of genes that are predominantly or specifically expressed in SG (S1, 5, 16, 31, 35 and 36) and MG (S10, 14, 18, 21 and 22) with transcript abundance decreasing for the former and increasing for the later as ticks continued to feed.

Discussion and conclusion
Analysis of genome sequence data has led to the discovery of large families of serpins in multicellular organisms, including 36 in humans, nine in *Caenorhabditis elegans* [16], 29 in *Drosophila* [26] and plants [53]. On the Merops database, 17 and 18 serpin sequences are listed for *Aedes aegypti* and *Anopheles gambiae*. In ticks, documentation of multiple serpin encoding cDNAs has provided indirect evidence suggesting that ticks do encode large serpin families. For instance, we recently described 17 serpin cDNAs that are expressed by *A. americanum* during feeding [33]. Here we describe the annotation and characterization of 45 serpin genes in the *I. scapularis* genome. The observed high amino acid sequence identity between *I. scapularis* and *I. ricinus* serpins was not surprising as these two ticks belong to the same genus. It is important to point out that eight of the 45 annotated were not represented in the preliminary peptide build at VectorBase. This finding may raise the prospect of error in annotations reported here. Interestingly, this possibility is ruled out, as ESTs of six of the eight serpins in this study (S1, 13, 23, 24, 33 and 39) were present in the trace archive database, while coding regions of serpins S26 and 32 were amplified from unfed and partially fed ticks (see figure 7).

Table 3: Summary, characterization of serpin inferred amino acid sequences

Serpin # ID	Signal peptide	Putative Function	Trace archive EST/mate Accession #s	Whole genome shotgun accession #s	Best match (% identity)
1	+	Inhibitory	1731419061/1731424162	ABJB010004033	IriS-4, DQ915844 (56)
2	+	Inhibitory	1446657576/1446661330	ABJB010418850 and ABJB011135496	IriS-4, DQ915844 51)
3	+	Inhibitory	1446665747/1446661466	ABJB010065261	IriS-4, DQ915844 (56)
4	+	Inhibitory	1446883315/1446875792	ABJB011011198	IriS-4, DQ915844 (57)
5	+	Inhibitory	1446915260/1446919205	ABJB010283366	IscAAM93649 (97)[REF1]
6	+	Inhibitory	1446658181/1446658460	ABJB011103238	IscAAV80788 (73)
7	+	Inhibitory	1446654646/1446653845	ABJB011103238 and ABJB010459188	IscAAV80788 (97) [REF1]
8	+	Inhibitory	NEF	ABJB010459188	IscAAV80788 (73)
9	+	Inhibitory	1798691964/1798683282	ABJB010543279	IscAAV80788 (73)
10	P	Inhibitory	NEF	ABJB010543279	Iri-1 DQ915842 (95)
11	+	Inhibitory	NEF	ABJB010543279	Lospin17 [REF2](46)
12	+	Inhibitory	1798658594/no mate	ABJB010080534	Lospin17 [REF2](29)
13	+	Non-Inhibitory	1798680529/1798669906	ABJB010568406	HloS-2, AB162827 (50)
14	+	Inhibitory	1446647997/1446656254	ABJB010572532	Lospin17 [REF2](38)
15	+	Inhibitory	NEF	ABJB010589966	Lospin17 [REF2](48)
16	+	Inhibitory	NEF	ABJB010189666	Lospin17 [REF2](44)
17	+	Inhibitory	1446918763/1446915174	ABJB010803180	IriS-1 DQ915842 (30)
18	+	Inhibitory	1446686436/1446684945	ABJB010296835	IriS-8 DQ915845 (98)
19	+	Inhibitory	1446935509/1446936853	ABJB010938637 and ABJB010859665	HloS-2, AB162827 (50)
20	+	Inhibitory	1446937497/1446940930	ABJB010863599	Lospin10 [REF2](34)
21	-	Inhibitory	799333999/1799375613	ABJB011069597 and ABJB010801481	Lospin17 [REF2](52)
22	+	Inhibitory	1798749904/1798771804	ABJB010325534 and ABJB010620410	HloS-2, AB162827 (51)
23	+	Inhibitory	1702098781/1702102057	ABJB010111671 and ABJB010862034	HloS-2, AB162827 (50)
24	+	Inhibitory	NEF	ABJB010229324	Lospin4 [REF2](46)
25	+	NR	NEF	ABJB010327382	Lospin17 [REF2](41)

Table 3: Summary, characterization of serpin inferred amino acid sequences *(Continued)*

26	+	Inhibitory	1702098781/1702098781	ABJB011108013	HloS-2, AB162827 (44)
27	+	Inhibitory	1446877265/1446875646	ABJB010216105	HloS-2, AB162827 (48)
28	+	NR	NEF	ABJB011030283	Lospin4 [REF2](34
29	+	Inhibitory	1702098781/1702102057	ABJB010216676	Lospin17 [REF2](49)
30	-	Inhibitory	1798755892/1798766236	ABJB011013454	Lospin12 [REF2](43)
31	P	Inhibitory	NEF	ABJB010583840	Lospin17 [REF2](47)
32	-	Inhibitory	1799314980/1799282524	ABJB010837586	Lospin7 [REF2](61)
33	P	Inhibitory	1446912351/1446909212	ABJB010588517	Lospin17 [REF2](44)
34	+	Inhibitory	NEF	ABJB010122781	Iris [REF2](93)
35	-	Non-inhibitory	1701896038/1701906143	ABJB010889434	IriS-2, DQ915843 (46)
36	+	Inhibitory	1446674847/1446673722	ABJB010262605	Iris [REF3](93)
37	-	Inhibitory	1446912351/1446909212	ABJB010112590	IriS-2, DQ915843 (60)
38	-	Inhibitory	1446875683/1446883661	ABJB011132035	Iris [REF3](60)
39	P	Inhibitory	1799178683/1799156584	ABJB010146202	Iris [REF3](55)
40	P	Inhibitory	NEF	ABJB011073580	Lospin17 [REF2](39)
41	+	Inhibitory	1799307827/1799285191	ABJB010040374	IriS-2, DQ915843 (93)
42	+	Inhibitory	1798653329/1798670720	ABJB010592280	IscAAV80788 (66)
43	P	Inhibitory	NEF	ABJB010519540	IriS-1 DQ915842 (61)
44	P	Inhibitory	1701931459/1701895980	ABJB010496686	RapS-3 [REF4](44)
45	P	Inhibitory	NEF	ABJB010314107	Lospin7 [REF2](41)

[REF1][35], [REF2][33], [REF3][34], [REF4] [17], NR, denotes partial in the carboxy-terminal region and no reactive center loop found, thus functionality not determined. Plus sign (+) denotes signal peptide is present, P = partial at the amino terminus region and presence of SP not determined. NEF denotes, not EST found. ²Because of space these two (ABJB0109755250 and ABJB010321224) additional accession numbers not listed in table above

The adoption of the consensus serpin secondary structures [11,12] and the high conservation of the core amino acid residues [51] that underpin the structure and functionality of serpins strongly suggested that, *I. scapularis* serpins are functional. The majority of known serpins function as inhibitors of serine proteinases and hence the name [54]. However others with activity against cysteine proteinases and those with no inhibitor functions have also been described [12]. Although on the basis of sequence analysis [11], we are able to distinguish between inhibitory and non-inhibitory serpins, available data in this study is insufficient to specify their preferred proteinase sub-

strates. Putative RCLs and scissile bonds of serpins in this study were predicted based on consensus that there are 17 amino acid residues between the start of the RCL hinge region and the scissile bond [11,45,51]. As some of the characterized serpins such as α_2-antiplasmin [11] or serpin1k from *Manducca sexta* [55], utilize shorter RCLs, we are interpreting our predictions of RCLs and scissile bonds with caution.

Our finding that 82% of the full-length serpins in this study have signal peptides is consistent with observations in humans where the majority of known serpins exist in

A   p17[E]-p16[E/K/R]-p15[G]-p14[T/S]-p13[X]-p12-9[AGS]-p8-1[X]-p8-1[X]-p1'-4'

				P												P'				
Gene ID	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	1'	2'	3'	4'
Serpin1	E	E	G	T	E	A	A	A	V	S	G	A	V	F	V	T	K	T	L	G	L
Serpin2	E	E	G	T	E	A	A	A	A	S	G	V	V	A	V	N	R	L	I	G	V
Serpin3	E	E	G	S	E	A	A	G	A	T	G	A	G	L	M	P	L	S	A	A	V
Serpin4	E	E	G	T	E	A	A	G	V	S	G	V	V	V	N	T	R	L	V	E	V
Serpin5	E	A	G	T	E	A	A	G	A	T	G	V	I	I	V	A	E	S	L	V	E
Serpin6	E	E	G	T	E	A	A	V	V	S	G	V	I	G	A	T	R	V	A	L	L
Serpin7	E	E	G	T	E	A	A	V	V	S	A	V	I	G	G	L	R	S	G	S	F
Serpin8	E	E	G	T	E	A	A	V	V	S	G	V	V	G	A	I	R	A	S	A	L
Serpin9	E	E	G	S	E	A	A	A	V	S	S	V	V	A	V	T	R	I	G	T	Q
Serpin10	D	G	G	F	V	A	A	A	V	T	T	V	R	N	V	N	G	G	R	G	R
Serpin11	E	E	R	A	V	S	P	G	G	G	A	I	A	T	P	K	A	V	P	V	F
Serpin12	E	K	G	T	V	A	A	A	V	T	S	I	S	M	G	M	R	S	S	P	G
Serpin13	E	E	G	A	G	F	I	N	P	T	P	G	W	T	P	N	A	E	R	A	M
Serpin14	E	E	G	T	A	A	A	S	A	T	S	V	G	I	V	A	K	S	I	P	Q
Serpin15	E	E	G	T	V	A	T	A	A	T	E	V	K	F	L	V	S	S	G	G	Y
Serpin16	E	A	G	S	E	A	E	S	M	T	V	V	K	V	S	G	K	S	A	E	T
Serpin17	E	E	G	S	E	A	A	A	V	T	G	F	V	I	Q	L	R	T	A	A	F
Serpin18	E	K	G	T	V	A	A	A	V	T	S	I	S	M	A	A	R	T	S	Q	I
Serpin19	E	K	G	T	Q	A	S	A	S	T	V	A	V	I	V	S	R	I	G	T	P
Serpin20	E	Q	G	T	E	A	V	A	L	S	S	G	I	V	R	H	S	R	P	P	E
Serpin21	E	K	G	T	V	A	A	A	V	T	S	I	S	V	A	V	R	T	S	L	I
Serpin22	E	K	G	T	V	A	A	A	V	T	S	I	S	M	R	V	G	S	S	P	T
Serpin23	E	E	G	T	V	A	T	G	V	T	V	V	L	I	V	R	S	V	P	R	S
Serpin24	E	E	G	T	V	A	T	G	V	T	V	V	A	T	A	M	S	R	P	P	S
Serpin26	E	E	G	T	V	G	T	G	A	T	V	V	V	M	E	K	R	A	P	H	S
Serpin27	E	E	G	T	R	A	V	A	A	T	Q	A	Q	F	V	S	K	S	L	V	Q
Serpin28	E	E	G	T	V	A	T	A	A	T	V	V	L	S	I	K	S	G	P	P	L
Serpin30	E	E	G	T	E	A	A	A	A	T	A	V	M	M	V	K	C	C	L	S	F
Serpin31	E	E	G	T	V	A	G	A	A	T	D	V	V	L	G	R	I	L	P	T	R
Serpin32	E	E	G	T	E	A	A	A	A	S	A	F	A	V	N	A	R	C	A	V	Y
Serpin33	E	E	G	T	E	A	A	A	A	T	A	I	P	M	M	L	M	S	A	R	F
Serpin34	E	A	G	S	E	A	S	P	S	T	S	G	L	V	S	K	S	V	E	T	L
Serpin35	E	E	G	A	E	G	V	F	A	T	P	V	I	M	M	C	Y	G	G	A	S
Serpin36	E	E	G	S	E	A	S	S	A	T	G	V	L	V	E	P	G	S	G	M	P
Serpin37	E	E	G	T	E	A	A	A	A	T	A	I	L	M	D	C	I	P	Q	V	V
Serpin38	E	E	G	S	V	A	A	A	T	T	G	I	G	I	V	A	I	S	L	P	P
Serpin39	E	E	G	T	E	A	A	A	A	T	A	M	V	M	L	C	C	M	S	F	P
Serpin40	E	E	G	T	E	A	A	A	A	T	V	V	R	I	G	K	S	L	P	R	S
Serpin41	E	E	G	T	E	A	A	A	A	T	G	V	V	I	V	P	Y	S	L	G	P
Serpin42	E	K	Y	S	E	A	A	A	V	T	A	V	A	V	P	L	I	V	G	G	D
Serpin43	E	E	G	T	V	A	G	A	A	T	E	V	K	E	G	R	I	L	I	S	F
Serpin44	E	E	G	T	E	A	S	A	A	T	G	I	L	A	P	L	G	S	G	M	P
Serpin45	E	E	G	D	E	T	A	A	D	A	G	M	E	G	V	S	L	T	L	D	H

B

Figure 4
Predicted *I. scapularis* serpin reactive center loops (RCL) (B). RCLs were determined based on the eight, residue pattern (A) that characterizes inhibitory serpins [45]. The residues in the RCL are numbered according to the standard nomenclature developed by Schechter and Berger [46], where residues on the amino-terminal side of the scissile (P1-P1') are not primed and those on the carboxy terminal side primed. Assuming that there are 17 residues between the base of RCL hinge and the scissile bond [45], boldfaced residues are predicted P1 residues with the broken line arrowhead indicating the position of the scissile bond.

SERPIN-ID / HAT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	29	30	32	34	35	36	38	41	42	45
F^{33}	-	-	-	-	-	-	L	L	-	-	-	-	-	-	-	-	L	-	-	L	-	-	-	-	S	-	-	-	L	-	-	-	-	-	-
N^{49}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
S^{53}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P^{54}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
S^{56}	-	-	-	-	-	-	-	-	-	-	G	-	-	G	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-
L^{61}	M	M	M	M	-	M	M	M	-	-	-	-	V	-	-	-	M	-	-	-	-	-	-	-	-	-	-	-	V	-	M	-	M	M	M
G^{67}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T^{72}	-	-	-	-	-	-	-	-	-	-	-	S	C	S	S	-	-	S	-	-	S	S	S	S	-	S	-	-	S	-	-	-	-	-	-
L^{80}	-	F	-	F	F	-	-	-	-	-	-	-	-	-	-	-	F	-	-	-	-	-	-	-	-	-	-	-	M	-	-	-	-	-	-
F^{130}	Y	-	Y	Y	-	Y	Y	Y	-	Y	Y	-	Y	Y	Y	-	Y	-	Y	L	-	-	Y	Y	-	-	-	Y	Y	-	-	-	Y	Y	Y
F^{147}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	M	-	-	-	-	-	-	V	-	-	-	-	L	-	-	-	-	-	-
I^{157}	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	V	V	V	V	V	V	V	V	V	-	V	L	L	V	-	V	-	-
N^{158}	-	-	-	-	-	-	-	-	-	D	-	-	-	-	-	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
V^{161}	-	-	-	-	-	-	-	-	-	-	-	M	-	F	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-
T^{165}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
G^{167}	E	D	-	-	-	N	N	D	-	-	-	-	-	-	-	K	-	-	-	D	-	-	-	-	-	S	S	Y	F	D	S	-	-	-	Q
I^{169}	-	-	-	-	-	-	-	-	-	-	-	-	V	-	-	-	V	-	-	-	-	-	V	-	-	-	-	-	V	-	-	-	-	-	-
T^{180}	-	-	L	-	-	-	-	-	-	-	D	-	-	G	-	-	-	S	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L^{184}	-	-	-	-	-	-	-	-	-	N	-	-	V	-	-	A	-	-	-	V	-	-	-	-	V	-	I	I	-	-	-	I	-	-	-
N^{186}	S	-	-	-	-	-	-	-	-	-	-	-	S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F^{190}	-	-	-	-	-	-	-	-	-	-	-	-	L	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	L	-	-	-	-
K^{191}	-	-	-	-	-	-	-	-	-	-	N	N	A	-	-	-	-	N	-	R	N	N	N	-	-	N	-	-	-	-	-	-	-	-	-
G^{192}	-	-	-	-	-	-	-	-	-	D	-	-	T	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-
W^{194}	-	-	-	-	-	-	-	-	-	-	-	-	L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F^{198}	-	-	-	-	-	-	-	-	-	-	-	-	S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T^{203}	-	-	-	-	-	-	-	-	-	-	-	-	S	-	-	-	-	N	-	-	-	-	-	-	-	-	-	S	-	-	-	-	-	-	-
F^{208}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
V^{128}	-	-	-	-	-	-	-	-	-	-	R	-	A	-	R	K	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
M^{220}	T	T	T	T	T	T	T	T	T	T	T	T	-	V	-	-	F	-	T	-	-	T	T	T	-	T	-	-	T	-	T	-	T	T	T
M^{221}	-	-	-	-	-	-	-	-	-	-	I	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Y^{244}	-	-	-	E	-	-	-	-	-	-	H	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L^{254}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-	-	-	-	-	-	-
P^{255}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P^{289}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
K^{290}	-	-	-	-	-	R	-	-	-	R	-	-	-	-	-	-	-	-	-	-	Q	-	-	-	-	-	-	-	R	-	-	-	R	-	-
L^{299}	-	-	I	-	-	-	-	-	-	F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L^{303}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
G^{307}	-	-	-	-	-	-	-	-	-	-	-	-	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	R	-	-	-	-	-	-	-
F^{312}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A^{316}	-	-	T	-	-	-	-	-	-	S	-	-	S	P	-	-	-	-	-	S	-	-	S	S	S	-	-	S	-	C	V	-	-	C	-
L^{327}	I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-	-	-	-	-	-
H^{334}	-	-	Q	-	-	Q	Q	Q	-	-	-	-	-	-	-	-	Q	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E^{342}	-	-	-	-	-	-	-	-	-	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
G^{344}	-	-	-	-	-	-	-	-	-	-	-	-	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y
A^{347}	-	-	-	-	-	-	-	-	-	-	-	-	S	-	F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-
P^{369}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F^{370}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L^{383}	M	-	-	-	-	-	-	-	-	-	I	-	-	H	-	M	-	-	-	-	-	-	-	-	-	-	-	V	-	-	-	-	I	-	-
F^{384}	-	-	-	-	-	-	-	-	-	-	L	-	L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
G^{386}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-
P^{391}	L	L	L	L	L	L	L	V	L	L	L	L	L	L	L	L	I	L	L	L	V	L	L	L	L	M	L	I	V	L	L	L	I	L	L

Figure 5
Conservation of core amino acid residues adopted from [51]. 35 full-length *I. scapularis* serpin sequences were aligned with the human, α_1-antitrypsin (α_1-HAT, accession # AAB59495) using structural based sequence alignment program (STRAP, http://www.charite.de/bioinf/strap/). The aligned sequences were visually inspected for conservation of core amino acid residues. The minus sign (-) denotes conservation of the specific amino acid residues. Replaced residues are indicated.
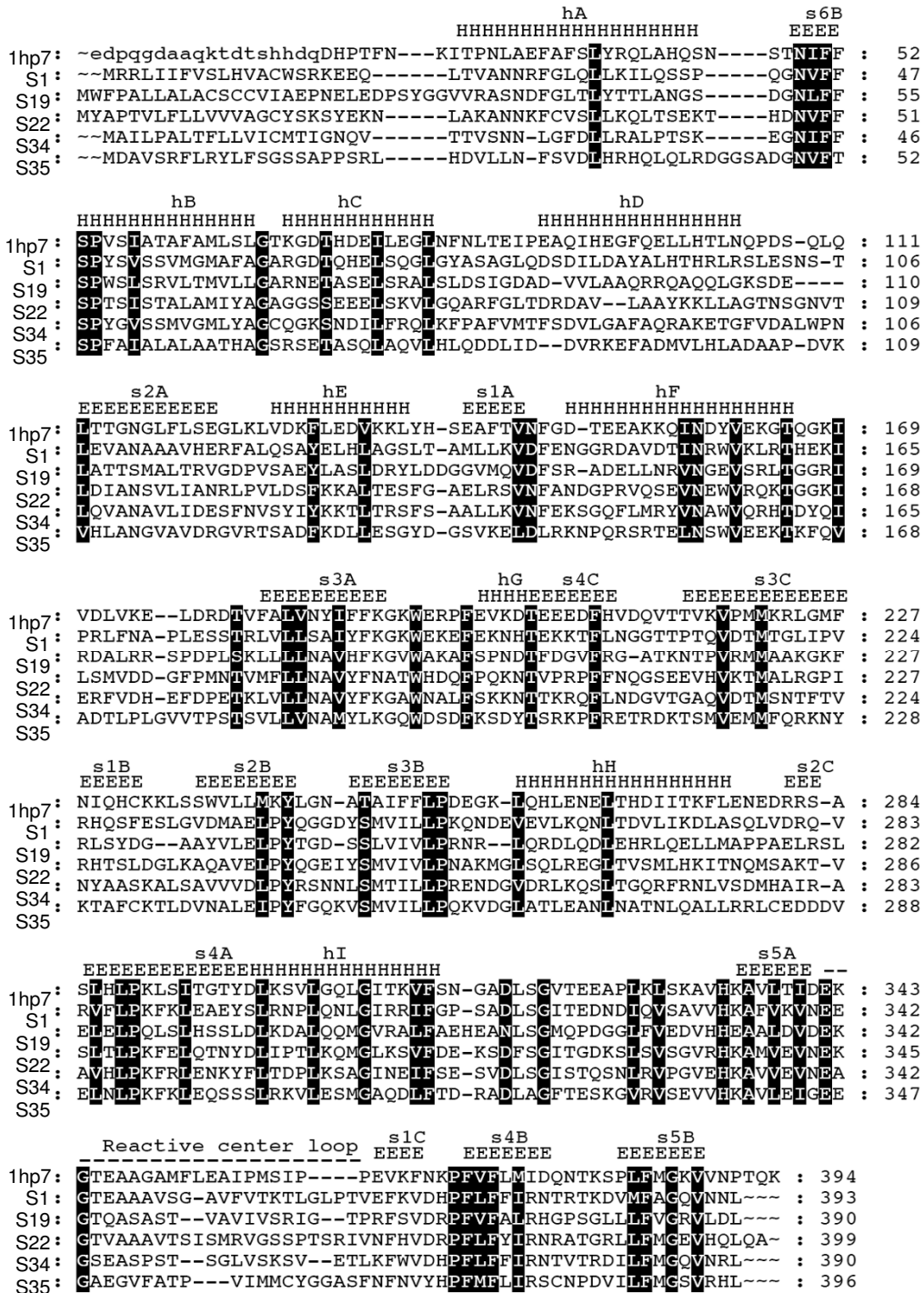
Figure 6
Neighbor joining guide phylogeny tree. Deduced amino acid sequences of 35 full-length serpins were aligned with 17 serpins from *A. americanum* [33], *R. appendiculatus* serpin (Ras) -1-to-4 [17], *H. longicornis* (Hl) serpin, *I. ricinus* (Irserp) serpin 1, 2, 4, 8 and Iris = immunosuppressive *Ixodes* serpin, *I. scapularis* (Is) serpin and *B. microplus* (Bm) serpin 1-to-6. Note that, except for Bmserpin-5, which, is annotated in GenBank, the other Bm serpins were obtained as ESTs from the TIGR database and translated in Mulenga et al., [33]. Cluster labels "A-to-E", represents branching from the outlier, the serpin superfamily archetype, α_1-antitrypsin. Accession numbers of whole genome shotgun genomic DNA sequences corresponding to supercontig sequences used in the annotation of *I. scapularis* serpins (marked with an asterisk sign) are given in Table 3.

```
                                                                 hA                    s6B
                                                 HHHHHHHHHHHHHHHHHHHHH                  EEEE
1hp7: ~edpqgdaaqktdtshhdqDHPTFN---KITPNLAEFAFSLYRQLAHQSN----STNIFF : 52
  S1: ~~MRRLIIFVSLHVACWSRKEEQ------LTVANNRFGLQLLKILQSSP-----QGNVFF : 47
 S19: MWFPALLALACSCCVIAEPNELEDPSYGGVVRASNDFGLTLYTTLANGS-----DGNLFF : 55
 S22: MYAPTVLFLLVVVAGCYSKSYEKN-----LAKANNKFCVSLLKQLTSEKT-----HDNVFF : 51
 S34: ~~MAILPALTFLLVICMTIGNQV------TTVSNN-LGFDLLRALPTSK-----EGNIFF : 46
 S35: ~~MDAVSRFLRYLFSGSSAPPSRL-----HDVLLN-FSVDLHRHQLQLRDGGSADGNVFT : 52

        hB                hC                        hD
      HHHHHHHHHHHH     HHHHHHHHHHH          HHHHHHHHHHHHHHHHHH
1hp7: SPVSIATAFAMLSLGTKGDTHDETLEGLNFNLTEIPEAQIHEGFQELLHTLNQPDS-QLQ : 111
  S1: SPYSVSSVMGMAFAGARGDTQHELSQGLGYASAGLQDSDILDAYALHTHRLRSLESNS-T : 106
 S19: SPWSLSRVLTMVLLGARNETASELSRALSLDSIGDAD-VVLAAQRRQAQQLGKSDE---- : 110
 S22: SPTSISTALAMIYACGAGGSSEEELSKVLGQARFGLTDRDAV--LAAYKKLLAGTNSGNVT : 109
 S34: SPYGVSSMVGMLYACCQGKSNDILFRQLKFPAFVMTFSDVLGAFAQRAKETGFVDALWPN : 106
 S35: SPFAIALALAATHACSRSETASQLAQVLHLQDDLID--DVRKEFADMVLHLADAAP-DVK : 109

      s2A              hE           s1A            hF
    EEEEEEEEEEE     HHHHHHHHHH     EEEEE      HHHHHHHHHHHHHHHHHHHH
1hp7: LTTGNGLFLSEGLKLVDKFLEDVKKLYH-SEAFTVNFGD-TEEAKKQINDYVEKGTQGKI : 169
  S1: LEVANAAAVHERFALQSAYELHLAGSLT-AMLLKVDFENGGRDAVDTINRWVKLRTHEKI : 165
 S19: LATTSMALTRVGDPVSAEYLASLDRYLDDGGVMQVDFSR-ADELLNRVNGEVSRLTGGRI : 169
 S22: LDIANSVLIANRLPVLDSFKKALTESFG-AELRSVNFANDGPRVQSEVNEWVRQKTGGKI : 168
 S34: LQVANAVLIDESFNVSYIYKKTLTRSFS-AALLKVNFEKSGQFLMRYVNAWVQRHTDYQI : 165
 S35: VHLANGVAVDRGVRTSADFKDLLESGYD-GSVKELDLRKNPQRSRTELNSWVEEKTKFQV : 168

        s3A             hG    s4C         s3C
      EEEEEEEEEE     HHHHEEEEEEE    EEEEEEEEEEEEEE
1hp7: VDLVKE--LDRDTVFALVNYIFFKGKWERPFEVKDTEEEDFHVDQVTTVKVPMMKRLGMF : 227
  S1: PRLFNA-PLESSTRLVLLSAIYFKGKWEKEFEKNHTEKKTFLNGGTTPTQVDTMTGLIPV : 224
 S19: RDALRR-SPDPLSKLLLLNAVHFKGVWAKAFSPNDTFDGVFRG-ATKNTPVRMMAAKGKF : 227
 S22: LSMVDD-GFPMNTVMFLINAVYFNATWHDQFPQKNLVPRPFFNQGSEEVHVKTMALRGPI : 227
 S34: ERFVDH-EFDPETKLVLINAVYFKGAWNALFSKKNTTKRQFLNDGVTGAQVDTMSNTFTV : 224
 S35: ADTLPLGVVTPSTSVLLVNAMYLKGQWDSDFKSDYTSRKPFRETRDKTSMVEMMFQRKNY : 228

      s1B      s2B          s3B            hH            s2C
     EEEEE     EEEEEEEE    EEEEEEEE    HHHHHHHHHHHHHHHH        EEE
1hp7: NIQHCKKLSSWVLLMKYLGN-ATAIFFLPDEGK-LQHLENELTHDIITKFLENEDRRS-A : 284
  S1: RHQSFESLGVDMAELPYQGGDYSMVILLPKQNDEVEVLKQNLTDVLIKDLASQLVDRQ-V : 283
 S19: RLSYDG--AAYVLELPYTGD-SSLVIVLPRNR--LQRDLQDLEHRLQELLMAPPAELRSL : 282
 S22: RHTSLDGLKAQAVELPYQGEIYSMVIVLPNAKMGLSQLREGLTVSMLHKITNQMSAKT-V : 286
 S34: NYAASKALSAVVVDLPYRSNNLSMTILLPRENDGVDRLKQSLTGQRFRNLVSDMHAIR-A : 283
 S35: KTAFCKTLDVNALEIPYFGQKVSMVILLPQKVDGLATLEANLNATNLQALLRRLCEDDDV : 288

      s4A           hI                                     s5A
    EEEEEEEEEEEEEEHHHHHHHHHHHHHHHH                       EEEEEE --
1hp7: SLHLPKLSITGTYDLKSVLGQLGITKVFSN-GADLSGVTEEAPLKLSKAVHKAVLTIDEK : 343
  S1: RVFLPKFKLEAEYSLRNPLQNLGIRRIFGP-SADLSGVSAVVHKAFVKVNEE : 342
 S19: ELELPQLSLHSSLDLKDALQQMGVRALFAEHEANLSCMQPDGGLFVEDVHHEAALDVDEK : 342
 S22: SLTLPKFELQTNYDLIPTLKQMGLKSVFDE-KSDFSGITGDKSLSVSGVRHKAMVEVNEK : 345
 S34: AVHLPKFRIENKYFLTDPLKSAGINEIFSE-SVDLSGISTQSNLRVPGVEHKAVVEVNEA : 342
 S35: ELNLPKFKIEQSSSLRKVLESMGAQDLFTD-RADLACFTESKGVRVSEVVHKAVLEIGEE : 347

      Reactive center loop    s1C      s4B       s5B
      _____      EEEE   EEEEEEE    EEEEEEE
1hp7: GTEAAGAMFLEAIPMSIP----PEVKFNKPFVFLMIDQNTKSPLFMGKVVNPTQK : 394
  S1: GTEAAAVSG-AVFVTKTLGLPTVEFKVDHPFLFFIRNTRTKDVMFACQVNNL~~~ : 393
 S19: GTQASAST--VAVIVSRIG--TPRFSVDRPFVFAIRHGPSGLLLFVGRVLDL~~~ : 390
 S22: GTVAAAVTSISMRVGSSPTSRIVNVDRPFLFYIRNRATGRLLFMGEVHQLQA~ : 399
 S34: GSEASPST--SGLVSKSV--ETLKFWVDHPFLFFIRNTVTRDILFMGQVNRL~~~ : 390
 S35: GAEGVFATP---VIMMCYGGASFNFNVYHPFMFLIRSCNPDVILFMGSVRHL~~~ : 396
```

Figure 7

Structural-based multiple sequence alignment. Using human α1-antitrypsin (1hbp7) as modeling template, five serpin sequences (S1, S19, S22, S34 and S35) representing each of the five main branches in figure 4 were aligned using the structural based sequence alignment program (STRAP, http://www.charite.de/bioinf/strap). Secondary structures assigned based on α1-antitrypsin tertiary structure (PDB ID, 1hbp7), are labeled as "H" for α-helix and "E" for beta-strand. Helices are labeled from "hA" to "hI", β-strands that constitutes β-sheet A-to-C are labeled as "sA", "sB" and "sC" respectively. The serpin superfamily core residues [52] are shaded in black. The reactive center loop domain is noted.
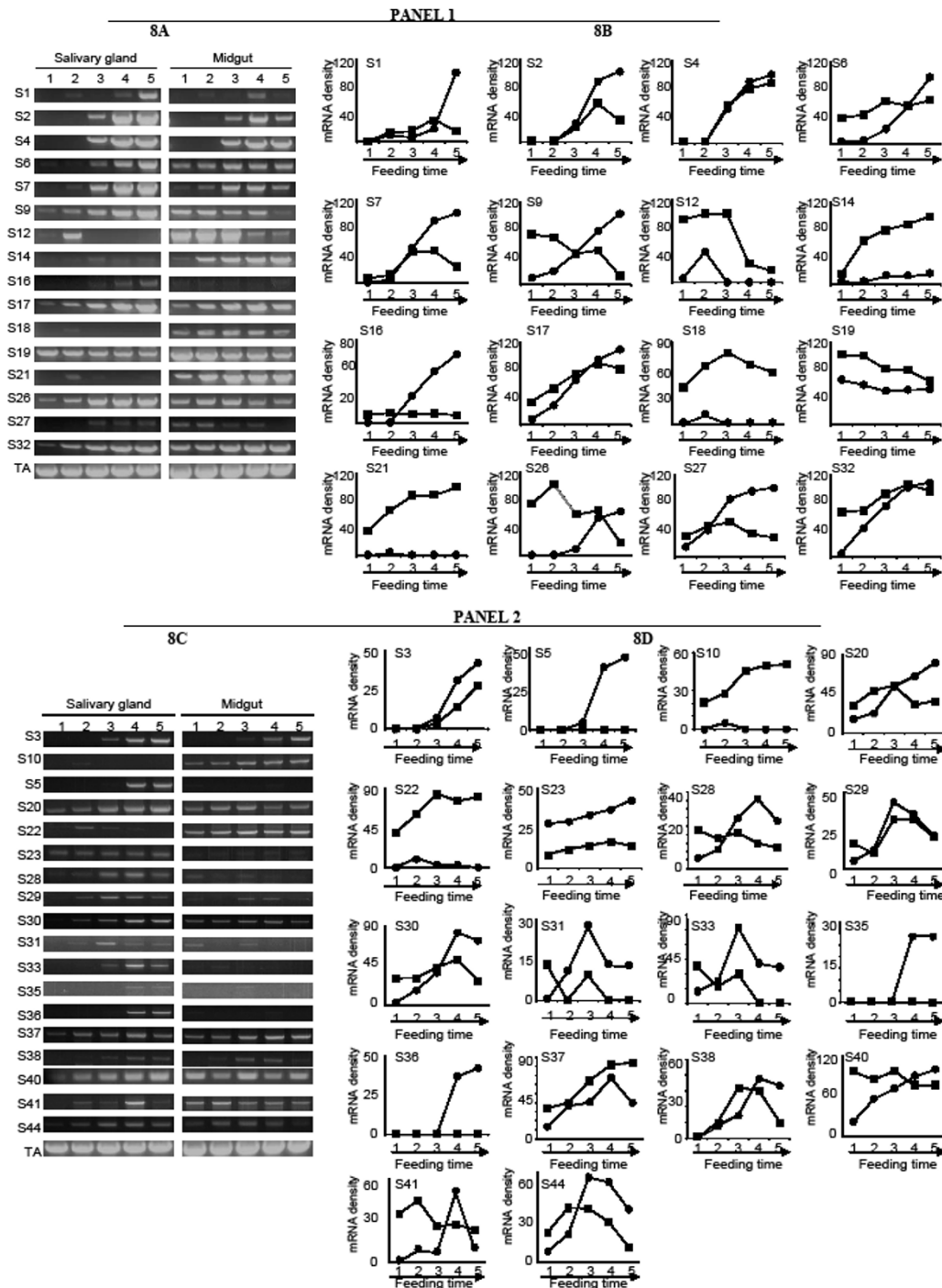
Figure 8
Transcription profiles (8A and 8C) and normalized PCR band densities representing relative mRNA abundance (8B and 8D). cDNA synthesized from total RNA of salivary glands (SG) and midgut (MG) of unfed and ticks that were partially fed for 6 to 96 hrs was subjected to semi-quantitative RT-PCR using gene specific primers shown in tables 1 and 2. PCR band densities were determined using the ImageJ software http://rsb.info.nih.gov/ij. Determined densities were normalized using the following formula, $Y = V + V(H-X)/X$ where Y = normalized mRNA density, V = observed Ispin PCR band density in individual tissues (MG and SG), H = highest actin PCR band density, X = tissue (MG and SG) actin band density.

the extracellular form [14]. Findings in this study are not unique, in that similar results were reported in *A. americanum* where 13 of the 17 putatively inhibitory serpins in *A. americanum* were predicted to be extracellular [33]. From the perspective of finding target antigens for vaccine development, it is encouraging to note that the majority of *I. scapularis* serpins are putatively extracellular as they will be accessible to host immune response factors. Predictions based on sequence analysis, may not be consistent with the situation *in vivo*. However it is interesting to note that the four serpin sequences (S30, 32, 35 and 38) that were predicted to be intracellular sequences, based on lack of a signal peptide also posses "C" residues in the exposed regions of their RCL, a feature that has been observed in many intracellular serpins [14].

The use of alternatively spliced RCLs appears to be a wide spread strategy in insects to diversify the range of target proteinases that can be controlled by a single serpin gene [56-58]. A classic observation of this phenomenon is the serpin gene-1 of the tobacco hornworm, *M. sexta*, which has 12 different alternatively, spliced RCLs [56]. Effectively this gives rise to 12 serpins regulating 12 different proteolytic pathways. An interesting structural feature among the 12 variants of *M. sexta* serpin gene 1 is that the first 336 amino acids are exactly identical with difference restricted to the RCL [56]. The identity patterns among *I. scapularis* serpins sequences of where, stretches of identical and variable domains were spread across the entire sequence suggest that the diversity among serpins in this study may have arisen by duplication other than alternative splicing of RCL encoding exons.

From the perspective of understanding how the tick manipulates the mammalian host's defense against tick feeding, the finding of 18 serpins with basic residues at their P1 sites was exciting. In humans, this feature is associated with key serpins such as α_1-antichymotrypsin, α_1-antiplasmin, antithrombin III, protein C inhibitor and C1 inhibitor [11], which regulate important pathways such as inflammation, blood coagulation and complement activation. As these pathways are thought to represent the mammalian host's defense against tick feeding [59,60], it is tempting to imagine that ticks could utilize some of these serpins to manipulate host defense to facilitate tick feeding and disease transmission. It is also possible that these serpins may not be directly be involved in facilitation of feeding. However, like their mammalian counterparts, they may be involved in regulation of important pathways in the tick, which if disrupted can affect the capacity of ticks as vectors.

Although the biological significance of gene expression data will be strengthened if correlated with protein production, our RT-PCR data provide some useful insights on

probable biological roles of serpin genes in this study in the physiology of *I. scapularis*. Speculatively *I. scapularis* genes that were induced or up regulated after ticks had penetrated their host skin may signal their involvement in facilitation of blood meal up take. This is particularly true for the 11 genes that were induced in both SG and MG (S1, 2, 3, 4 and 7) or SG alone (S5, 6, 16, 26, 35 and 36) in ticks that were fed for 6–24 hrs. This period corresponds to the preparatory feeding phase when tick attaches onto host skin and creates its feeding lesion [2]. During this period the tick must overcome inflammation and blood coagulation for it to successfully start the feeding process. Similarly, the group of serpin genes (S17, 23, 25, 32, 37, 38 and 40) that were constitutively expressed but their transcript abundance increased with tick feeding may also play a role in facilitating blood meal up take. For those genes that were constitutively expressed, S9, 12 and 27 in the MG, and S19 in both SG and MG, but were progressively down regulated as ticks continued to feed, could be involved in regulating physiological processes at the front end of tick feeding process. The expression of S10, 14, 18, 21 and 22, specifically or predominantly in the MG is interesting as it signals the potential role for these genes in facilitation of not only blood meal processing, but also in the crossing of the gut barrier by pathogens. From the perspective of our long-term interest to find tick proteins that are used by ticks to evade host immunity, it was exciting to note that some serpins were specifically expressed in SG. It will be exciting to investigate whether or not any of these genes are injected into the host during tick feeding. It is possible that the genes analyzed here could be expressed in multiple tick organs besides the SG and MG. However, from the perspective of our long-term interests to understand molecular mechanisms that underlie tick-host interactions, our analysis here is biased to biological functions of serpins at the SG and MG levels. The SG is critical for feeding and disease transmission while the MG is important for blood meal processing and the passage of pathogens from the blood meal into the tick hemolymph [2]. Our future questions will thus address the role of the serpins in facilitation of tick feeding and blood meal processing.

Most known serpins are glycosylated [12,16] and thus it is not surprising that 40 of the 43 serpin sequences that were tested are predicted to possess putative N-glycosyslation sites. As pool feeders, ticks accomplish feeding by lacerating small blood vessels and then sucking blood from the hematoma that forms in the feeding site [2]. In order to complete feeding, ticks must prevent host blood from coagulating to ensure continued blood flow into the hematoma for the entire tick feeding period, which may last for over the 10–14 and 4–7 day feeding periods for adults and immature ticks respectively [2]. From the perspective of solving the paradox of how the tick interferes

with the coagulation cascade of its mammalian host, it was interesting to note that ∼9% (4/45) of the sequences contain the RGD motif. Previous studies have shown that tick proteins containing the RGD motif such as variabilin [61] and savignygrin [62] were potent inhibitors of platelet aggregation. Platelet aggregation is critical to stopping bleeding of injured small blood vessels such as occurs at the tick bite [59]. Thus, if functional, the RGD motif containing serpins could represent important molecular targets aimed at countering the ability of ticks to suppress the mammalian's host's blood clotting system. In addition to the anti-platelet aggregation function, RGD motif containing proteins are also involved in regulating cell-cell interactions in mammals [63], immunity in arthropods [64], and plants [65]. From the foregoing, it is clear that the RGD motif containing serpins could also be involved in regulation of multiple other functions in the tick, besides platelet function at the tick feeding site. It is interesting to note that our RT-PCR data suggested that the expression of three (S1, 7 and 16)) of the four serpin RGD motif-containing genes was responsive to tick feeding activity (see figure 8).

When compared to the 3100 Mbp human genome that encodes at least 36 serpin genes [16] the 45 serpin genes identified in *I. scapularis*, which has a 2100 Mbp genome is considerably high. While the biological significance of the high number of serpin genes in the *I. scapularis* biology cannot be ascertained at present, we speculate that this may signal the significance of serpins in tick physiology. In light of lack of genome sequence data of many tick species, the sizes of tick serpin families will remain unknown. Ticks are diverse, both in terms of their biology [2] and their genome sizes [66-68]. Thus it is likely that the sizes of serpin families in ticks are going to vary. However, this study provides some insight on the probable sizes of serpin families in ticks.

Authors' contributions

AM was responsible for conception, design, data interpretation, discussion and presentation of the work. RK equally participated in the conception, was responsible for data mining and assembly and performed part of the transcription analysis. KCC was responsible for completing all transcription analysis. All authors participated in the write up of the manuscript.

Acknowledgements

References

1. Jonjean F, Uilenberg G: **The global importance of ticks.** *Parasitology* 2004, **129**:S3-S14.
2. Sonenshine DE: **Biology of ticks.** Oxford University Press, Oxford; 1993.
3. Parola P, Raoult D: **Tropical rickettsioses.** *Clinics in dermatology* 2006, **24**:191-200.
4. Parola P, Raoult D: **Ticks and tickborne bacterial diseases in humans: an emerging infectious threat.** *Clin Infect Dis* 2001, **32**:897-928.
5. Parola P, Raoult D: **Tick-borne bacterial diseases Emerging in Europe.** *Clin Microbiol Infect* 2001, **7**:80-83.
6. Parola P, Paddock CD, Raoult D: **Tick-Borne Rickettsioses around the World: Emerging Diseases Challenging Old Concepts.** *Clinical Microbiol Rev* 2005, **18**:719-756.
7. Burgdorfer W: **Discovery of the Lyme disease spirochaete and its relation to tick vectors.** *Yale J Biol Med* 1984, **57**:515-520.
8. Burgdorfer W, Barbour AG, Hayes SF, Banach JL, Grunwaldt E, Davis JP: **Lyme disease-a tick borne spirochaetosis.** *Science* 1982, **216**:1317-1319.
9. Schwartz SN, Liles WC, Spach DH: **Tick-Borne Diseases.** *The New England Journal of Medicine* 1994, **29**:936-947.
10. Hill CA, Wikel SK: **The *Ixodes scapularis* genome project: an opportunity for advancing tick research.** *Trends in Parasitol* 2005, **21**:151-153.
11. Gettins PWG: **Serpin structure, mechanism, and function.** *Chem Rev* 2002, **102**:4751-4803.
12. Huntington JA: **Shape-shifting serpins-advantages of a mobile mechanism.** *Trends Biochem Sci* 2006, **31**:427-435.
13. Silverman GA, Bird PI, Carrell RW, Church FC, Caughlin PB, Gettins PGW, Moyer RW, Pemberton PA, Remold-O'Donnell E, Salvesen GS, Travis J, Whisstock JC: **The serpins are expanding superfamily of structurally similar but functionally diverse proteins: evolution, mechanism of inhibition, novel functions, and a revised nomenclature.** *J Biol Chem* 2001, **276**:33293-33296.
14. Silverman GA, Whisstock JC, Askew DJ, Pak SC, Luke CJ, Cataltepe S, Irving JA, Bird PI: **Human clade B serpins (ov-serpins) belong to a cohort of evolutionarily dispersed intracellular proteinase inhibitor clades that protect cells from promiscuous proteolysis.** *Cell Mol Life Sci* 2004, **61**:301-325.
15. Whisstock JC, Bottomley SP, Bird PI, Pike RN, Coughlin P: **Serpins 2005-fun between the β-sheets, Meeting report based upon presentations made ath the 4th International Symposium on Serpin Structure, function and biology (Cairns, Australia).** *FEBS J* 2005, **272**:4868-4873.
16. Law HPR, Zhang Q, McCowan S, Buckle AM, Silverman GA, Wong W, Rosado CJ, Langendorf CG, Pike RN, Bird PI, Whisstock JC: **An overview of the serpin superfamily.** *Genome Bilogy* 2006, **7**:216-226.
17. Mulenga A, Tsuda A, Onuma M, Sugimoto C: **Four serine proteinase inhibitors (serpin) from the brown ear tick, Rhiphicephalus appendiculatus; cDNA cloning and preliminary characterization.** *Insect Biochem Mol Biol* 2003, **33**:267-276.
18. Irving JA, Steenbakkers PJM, Lsk AM, Op den Camp HJM, Pike RN, Wisstock JC: **Serpins in prokaryotes.** *Mol Biol Evol* 2002, **19**:1881-1890.
19. Rawlings ND, Morton FR, Kok CY, Kong J, Barrett AJ: **MEROPS: the peptidase database.** *Nucleic Acids Res* 2008:D320-325.
20. van Gent Diana, Sharp Paul, Morgan Kevin, Kalsheker Noor: **Serpins: structure, function and molecular evolution.** *The Int J Biochem Cell Biol* 2003, **35**:1536-1547.
21. Michel K, Suwanchaichinda C, Morlais I, Lambrechts L, Cohuet A, Awono-Ambene PH, Simard F, Fontenille D, Kanost MR, Kafatos FC: **Increased melanizing activity in *Anopheles gambiae* does not affect development of Plasmodium falciparum.** *Proc Natl Acad Sci USA* 2006, **103**:16858-16863.
22. Abraham EG, Pinto SB, Ghosh A, Vanlandingham DL, Budd A, Higgs S, Kafatos FC, Jacobs-Lorena M, Michel K: **An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites.** *Proc Natl Acad Sci* 2005, **102**:16327-16332.
23. Brüning M, Lummer M, Bentele C, Smolenaars MM, Rodenburg KW, Ragg H: **The Spn4 gene from Drosophila melanogaster is a multipurpose defense tool directed against proteases from three different peptidase families.** *Biochem J* 2007, **401**:325-331.
24. Nappi AJ, Frey F, Carton Y: *Drosophila* serpin 27A is a likely target for immune suppression of the blood cell-mediated

melanotic encapsulation response. *J Insect Physiol* 2005, **51**:197-205.

25. Pelte N, Robertson AS, Zou Z, Belorgey D, Dafforn TR, Jiang H, Lomas D, Reichhart JM, Gubb D: **Immune challenge induces N-terminal cleavage of the Drosophila serpin Necrotic.** *Insect Biochem Mol Biol* 2006, **36**:37-46.

26. Reichhart JM: **Tip of another iceberg: Drosophila serpins.** *Trends Cell Biol* 2005, **15**:659-665.

27. Zou Z, Jiang H: **Manduca sexta serpin-6 regulates immune serine proteinases PAP-3 and HP8. cDNA cloning, protein expression, inhibition kinetics, and function elucidation.** *J Biol Chem* 2005, **280**:14341-14348.

28. Ligoxygakis P, Roth S, Reichhart JM: **A Serpin Regulates Dorsal-Ventral Axis Formation in the Drosophila Embryo.** *Current Biology* 2003, **13**:2097-2102.

29. Agarwala KL, Kawabata S, Miura Y, Kuroki Y, Iwanaga S: **Limulus intracellular coagulation inhibitor type 3. Purification, characterization, cDNA cloning, and tissue localization.** *J Biol Chem* 1996, **271**:23768-23774.

30. Miura Y, Kawabata S, Wakamiya Y, Nakamura T, Iwanaga S: **A limulus intracellular coagulation inhibitor type 2. Purification, characterization, cDNA cloning, and tissue localization.** *J Biol Chem* 1995, **270**:558-565.

31. Miura Y, Kawabata S, Iwanaga S: **A Limulus intracellular coagulation inhibitor with characteristics of the serpin superfamily. Purification, characterization, and cDNA cloning.** *J Biol Chem* 1994, **269**:542-547.

32. Mulenga A, Sugino M, Nakajima M, Sugimoto C, Onuma M: **Tick-Encoded serine proteinase inhibitors (serpins); potential target antigens for tick vaccine development.** *J Vet Med Sci* 2001, **63**:1063-1069.

33. Mulenga A, Khumthong R, Blandon MA: **Molecular and expression analysis of a family of the Amblyomma americanum tick Lospins.** *J Exp Biol* 2007, **210(Pt 18)**:3188-3198.

34. Leboulle G, Rochez C, Louahed J, Rutti B, Brossard M: **Isolation of Ixodes ricinus salivary gland mRNA encoding factors induced during blood feeding.** *Amer J Trop Med Hygiene* 2002, **66**:225-233.

35. Ribeiro JM, Alarcon-Chaidez F, Francischetti IM, Mans BJ, Mather TN, Valenzuela JG, Wikel SK: **An annotated catalog of salivary gland transcripts from Ixodes scapularis ticks.** *Insect Biochem Mol Biol* 2006, **36**:111-129.

36. Sugino M, Imamura S, Mulenga A, Nakajima M, Tsuda A, Ohashi K, Unuma M: **A serine proteinase inhibitor (serpin) from ixodid tick Haemaphysalis longicornis; cloning and preliminary assessment of its suitability as a candidate for a tick vaccine.** *Vaccine* 2003, **21**:2844-2851.

37. Imamura S, da Silva VJ, Sugino M, Ohashi K, Onuma M: **A serine proteinase inhibitor (serpin) from Haemaphysalis longicornis as an anti-tick vaccine.** *Vaccine* 2005, **23**:1301-1311.

38. Prevot PP, Couvreur B, Denis V, Brossard M, Vanhamme L, Godfroid E: **Protective immunity against Ixodes ricinus induced by a salivary serpin.** *Vaccine* 2007, **25**:3284-3292.

39. Lawson D, Arensburger P, Atkinson P, Besansky NJ, Bruggner RV, Butler R, Campbell KS, Christophides GK, Christley S, Dialynas E, Emmert D, Hammond M, Hill CA, Kennedy RC, Lobo NF, MacCallum MR, Madey G, Megy K, Redmond S, Russo S, Severson DW, Stinson EO, Topalis P, Zdobnov EM, Birney E, Gelbart WM, Kafatos FC, Louis C, Collins FH: **VectorBase: a home for invertebrate vectors of human pathogens.** *Nucleic Acids Res* 2007:D503-505.

40. Megy K, Hammond M, Lawson D, Bruggner RV, Birney E, Collins FH: **Genomic resources for invertebrate vectors of human pathogens, and the role of VectorBase.** *Infect Genet Evol* 2008, **9(3)**:308-13.

41. Pearson WR, Lipman DJ: **Improved tools for biological sequence comparison.** *Proc Natl Acad Sci USA* 1988, **85**:2444-2448.

42. Burge CB, Karlin S: **Finding the genes in genomic DNA.** *Curr Opin Struct Biol* 1998, **8**:346-354.

43. Wheelan SJ, Church DM, Ostell JM: **Spidey: A Tool for mRNA-to-Genomic alignments.** *Genome Research* 2001, **11**:1952-1957.

44. Han J, Zhang H, Min G, Kemler D, Hashimoto C: **A novel Drosophila serpin that inhibits serine proteases.** *FEBS Letters* 2000, **468**:194-198.

45. Hopkins PC, Carrell RW, Stone SR: **Effects of mutations in the hinge region of serpins.** *Biochemistry* 1993, **32**:7650-7657.

46. Schechter I, Berger A: **On the size of the active site in the proteases i.e. papain.** *Biochem Biophys Res Commun* 1967, **27**:157-162.

47. Kim S, Woo J, Seo EJ, Yu M, Ryu S: **A resolution structure of an uncleaved alpha (1)-antitrypsin shows variability of the reactive center and other loops.** *J Mol Biol* 2001, **306**:109-119.

48. Kotsyfakis M, Karim S, Andersen JF, Mather TN, Ribeiro JMC: **Selective Cysteine Protease Inhibition Contributes to Blood-feeding Success of the Tick Ixodes scapularis.** *J Biol Chem* 2007, **282**:29256-29263.

49. Whisstock JC, Bottomley SP: **Molecular gymnastics: serpin structure, folding and misfolding.** *Curr Opin Struct Biol* 2006, **16**:761-768.

50. de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N: **ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins.** *Nucleic Acids Res* 2006:W362-365.

51. Irving JA, Pike RN, Lesk AM, Whisstock JC: **Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function.** *Genome Res* 2000, **10**:1845-1864.

52. Krem MM, Di Cera E: **Conserved Ser residues, the shutter region, and speciation in serpin evolution.** *J Biol Chem* 2003, **278**:37810-37814.

53. Roberts TH, Hejgaard J: **Serpins in plants and green algae.** *Funct Integr Genomics* 2008, **8**:1-27.

54. Hunt LT, Dayhoff MO: **A surprising new protein superfamily containing ovalbumin, antithrombin-III, and alpha 1-proteinase inhibitor.** *Biochem Biophys Res Commun* 1980, **95**:864-8671.

55. Li J, Wang Z, Canagarajah B, Jiang H, Kanost M, Goldsmith EJ: **The structure of active serpin 1K from Manduca sexta.** *Structure* 1999, **7**:103-109.

56. Jiang H, Wang Y, Huang Y, Mulnix AB, Kadel J, Cole K, Kanost MR: **Evolution of a family of alternate exons encoding the reactive site loop.** *J Biol Chem* 1996, **271**:28017-28023.

57. Kruger O, Ladewi J, Koster K, Ragg H: **Widespread occurrence of serpin genes with multiple reactive centre-containing exon cassettes in insects and nematodes.** *Gene* 2002, **293**:97-105.

58. Pak SC, Kumar V, Tsu C, Luke CJ, Askew YS, Mills DR, Bromme D, Silverman GA: **SRP-2 is a cross-class inhibitor that participates in postembryonic development of the nematode Caenorhabditis elegans: initial characterization of the clade L serpins.** *J Biol Chem* 2004, **279**:15448-15459.

59. Ribeiro JMC: **Role of saliva in blodd-feeding by arthropods.** *Ann Rev Entomol* 1987, **32**:463-478.

60. Valenzuela JG, Charlab R, Mather TM, Ribeiro JMC: **Purification, Cloning, and Expression of a Novel Salivary Anti-Complement Protein from the Tick, Ixodies Scapularis.** *J Biol Chem* 2000, **32**:747-754.

61. Wang X, Coons LB, Taylor DB, Stevens SE Jr, Gartner TK: **Variabilin, a novel RGD-containing antagonist of glycoprotein IIb-IIIa and platelet aggregation inhibitor from the hard tick Dermacentor variabilis.** *J Biol Chem* 1996, **271**:17785-90.

62. Mans BJ, Louw AI, Neitz AW: **Savignygrin, a platelet aggregation inhibitor from the soft tick Ornithodoros savignyi, presents the RGD integrin recognition motif on the Kunitz-BPTI fold.** *J Biol Chem* 2002, **277**:21371-2178.

63. Lechner AM, Assfalg-Machleidt I, Zahler S, Stoeckelhuber M, Machleidt W, Jochum M, Nägler DK: **RGD-dependent binding of procathepsin X to integrin alphavbeta3 mediates cell-adhesive properties.** *J Biol Chem* 2006, **281**:39588-39597.

64. Johansson MW: **Cell adhesion molecules in invertebrate immunity.** *Dev Comp Immunol* 1999, **23**:303-315.

65. Mellersh DG, Heath MC: **Plasma membrane-cell wall adhesion is required for expression of plant defense responses during fungal penetration.** *Plant Cell* 2001, **13**:413-424.

66. Ullmann AJ, Lima CM, Guerrero FD, Piesman J, Black WC 4th: **Genome size and organization in the blacklegged tick, Ixodes scapularis and the Southern cattle tick, Boophilus microplus.** *Insect Mol Biol* 2005, **14**:217-222.

67. Pagel Van Zee J, Geraci NS, Guerrero FD, Wikel SK, Stuart JJ, Nene VM, Hill CA: **Tick genomics: the Ixodes genome project and beyond.** *Int J Parasitol* 2007, **37**:1297-1305.

68. Geraci NS, Spencer JJ, Paul RJ, Wikel SK, Hill CA: **Variation in genome size of argasid and ixodid ticks.** *Insect Biochem Mol Biol* 2007, **37**:399-408.