NIH Public Access
Author Manuscript
Gene. Author manuscript; available in PMC 2012 August 15.

Published in final edited form as:
Gene. 2011 August 15; 482(1-2): 78–93. doi:10.1016/j.gene.2011.04.008.

A snapshot of the *Ixodes scapularis* degradome

Albert Mulenga[*] and **Kelly Erikson**[+]
Texas A & M University AgriLife Research, Department of entomology, College Station, Texas 77843

Abstract

Parasitic encoded proteases are essential to regulating interactions between parasites and their hosts and thus they represent attractive anti-parasitic druggable and/or vaccine target. We have utilized annotations of *Ixodes scapularis* proteases in gene bank and version 9.3 MEROPS database to compile an index of at least 233 putatively active and 150 putatively inactive protease enzymes that are encoded by the *Ixodes scapularis* genome. The 233 putatively active protease homologs hereafter referred to as the degradome (the full repertoire of proteases encoded by the *I. scapularis* genome) represents ~1.14% of the 20485 putative *I. scapularis* protein content. Consistent with observations in other animals, the content of the *I. scapularis* degradome is ~6.0% (14/233) aspartic, ~19% (44/233) cysteine, ~40% (93/233) Metallo, ~28.3% (66/233) serine and ~6.4% (15/233) threonine proteases. When scanned against other tick sequences, ~11% (25/233) of *I. scapularis* putatively active proteases are conserved in other tick species with ≥60% amino acid identity levels. The *I. scapularis* genome does not apparently encode for putatively inactive aspartic proteases. Of the 150 putative inactive protease homologs none are from the aspartic protease class, ~8% (12/150) are cysteine, ~58.7% (88/150) metallo, 30% (45/150) serine and ~3.3% (5/150) are threonine proteases. The *I. scapularis* tick genome appears to have evolutionarily lost proteolytic activity of at least 6 protease families, C56 and C64 (cysteine), M20 and M23 (metallo), S24 and S28 (serine) as revealed by a lack of the putatively active proteases in these families. The overall protease content is comparable to other organisms. However, the paucity of the S1 chymotrypsin/trypsin-like serine protease family in the *I. scapularis* genome where it is ~12.7% (28/233) of the degradome as opposed to ~22–48% content in other blood feeding arthropods, *Pediculus humanus humanus*, *Anopheles gambiae*, *Aedes Aegypti* and *Culex pipiens quinquenfasciatus* is notable. The data is presented as a one-stop index of proteases encoded by the *I. scapularis* genome.

Keywords

Ixodes scapularis degradome; aspartic; serine; metallo; cysteine; threonine

INTRODUCTION

Ticks represent are among the most successful ectoparasites of animals and humans. Among all known arthropod vectors of disease agents, ticks transmit the most diverse animal and

© 2010 Elsevier B.V. All rights reserved.

[*]Corresponding author: a-mulenga@tamu.edu, Phone; 979 458 4300.
[+]KE was an undergraduate on NIH sponsored summer research experience

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

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

human disease agents (Sonenshine, 1993). In tropical and subtropical climates, ticks and tick borne constitute the major source of monetary losses in the livestock industry (Jongejan and Uilenberg, 2004). Some of the most important tick borne livestock diseases include the globally distributed cattle tick fever, heartwater and theileriosis on the African continent (Jongejan and Uilenberg, 2004). In terms of the impact of vector borne diseases on public health, ticks are only second to mosquitoes as vectors of important human disease agents (Sonenshine, 1993). Since 1980s when the cause agent of Lyme disease was described as the spirochete *Borrelia burgdoferri* transmitted by the blacklegged tick, *Ixodes scapularis* in the 1980s, an increased number of human tick borne disease agents have been identified (Branton and Corey, 2005). Of the 11 human tick borne diseases described in the United States, 4 (borrelisosis, anaplasmosis, babesiosis and Powassan virus) are vectored by *Ixodes* tick species (Branton and Corey, 2005, Fish and Childs, 2009). From the perspective of *Ixodes* tick species being important in public health, the USA national institute of health funded sequencing of the *I. scapularis* genome to provide resources that will lead to in depth tick molecular physiology studies (Pagel et al., 2007; Nene 2009). The rationale was that in depth tick molecular physiology studies to be facilitated by the availability of genome sequence data will open up opportunities to discover pathways that are critical to regulation of key tick pathways. These data will in turn allow for discovery of anti-tick vaccine and/or druggable targets that will lead to development of efficient ticks and tick borne disease control methods. Our long-term interest is deciphering the role(s) of protease-mediated pathways in tick physiology and/or regulation of tick parasitism.

Proteases regulate physiological roles that are essential to life of all organisms, from microbes to animals and plants (Quesada et al., 2009; Puente et al., 2003; 2005). In mammals pathways that are essential to life such as immune response, inflammation and blood coagulation are proteolytic cascades (Puente et al., 2003; 2005). In parasitic organisms, research on proteases has focused on the role of parasite-derived proteases in regulating the pathogenesis of parasitic diseases (Mckerrow et al., 2006). Parasitic proteases are now targets for anti-parasitic drug discovery (Mckerrow et al 2006; Trenholme et al., 2008).

The potential for proteases to be central to maintenance of tick homeostasis and regulation of tick and host interactions have attracted intense interest in tick protease research. At the time of this write up, search of literature in pubmed with query terms “tick proteases or tick proteinase” retrieved more than 320 research articles demonstrating a sustained research interest by the tick research community. To date studies on tick proteases have predominantly focused on profiling of proteolytic activity in tick tissues (Mulenga et al., 2001; Horn et al., 2009; Alim et al., 2008a; 2008b; Decrem et al., 2008a), cloning and characterization of tick protease encoding cDNAs (Mulenga et al., 1999; 2001; 2003; Hatta et al., 2009), validating the role(s) of tick encoded proteases in feeding regulation using RNAi mediated gene silencing (Miyoshi et al., 2004; Alim et al., 2009; Decrem et al., 2008b) and testing the anti-tick vaccine of recombinant tick proteases (Seixas et al., 2008). In this study we have utilized tick protease annotations in Gene Bank and the MEROPS database (version 9.3) (Rawlings et al., 2010) and Vectorbase to compile a snap shot of the *I. scapularis* degradome. The degradome as utilized here was defined as the full repertoire of protease genes that are expressed by an organism (Puente et al., 2003; 2005). The *I. scapularis* degradome presented in this study forms the foundation for in depth studies on the role of proteases in the physiology of *I. scapularis* and other tick species significant public health importance.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

MATERIALS AND METHODS

Sequences that were used in this study were downloaded from publicly available databases: gene bank, vectorbase (http://www.vectorbase.org/) and version 9.3 MEROPS database (http://merops.sanger.ac.uk/, Rawlings et al., 2010). To start, we downloaded 236 putatively active and 180 putatively inactive protease enzyme homologs from the MEROPS database. These protease downloads were then scanned against tick sequence entries in Gene Bank using the BLASTP homology search to assign gene bank accession numbers. To determine Gene Bank accession numbers of annotated sequences, the top best match that showed 100% amino acid residue identity with the query (MEROPS databas gene e derived download) was adopted. Sequences that produced hits to the same accession numbers were subjected to multiple amino acid sequence alignment using the web based BIOEDIT (http://www.mbio.ncsu.edu/bioedit/) DNA sequence analysis software program. Sequences that showed 100% amino acid identity were determined to be redundant annotations and culled out of the analysis. For provisional identification, we for the most part relied on annotations from the MEROPS database (Rawlings et al., 2010) with exception of instances mentioned when the putative protease was identified from the BLASTP scanning done in this study. With amino acid identity level set at ≥ 60%, *I. scapularis* proteases were scanned other tick sequences in Gene Bank to identify homologs in other tick species. The identification of putatively active or putatively inactive protease homologs was based on presence or absence of consensus active site amino acid residues as indicated on the Merops database (Rawlings et al., 2010).

RESULTS AND DISCUSSION

I. scapularis proteases annotated on the Merops database (Rawlings et al., 2010), were successfully used as query to identify redundant protease annotations in public databases. Additionally, 10 calcium activated cysteine proteases (calpain) that were not included on the list of proteases on the Merops database version 9.3 (Rawlings et al., 2010) were identified. We designated annotations that produced hits to the same gene as redundant annotations (not shown). After culling off redundant sequences, 16 putatively active and 30 inactive protease homologs, we determined that the *I. scapualaris* genome encoded for at least 233 putatively active and 150 inactive protease enzyme homologs (tables 1–12). Proteases are classified on the basis of mechanisms of cleavage and/or amino acid residues at the active site. On this basis, the MEROPS database currently classifies proteases into 9 classifications: the 6 conventional protease classes (Aspartic, Cysteine, Glutamic, Mettalo, Threonine and Serine) and others (Glutamic, Asparagine and Unknown) (Rawlings et al., 2010). Consistent with observations in mammals (Puente et al., 2003; 2005) the 233 putatively active *I. scapularis* proteases are composed of ~6.0% (14/233) aspartic, ~18% (44/233) cysteine, ~40% (93/233) Metallo, ~28.3% (66/233) serine and, ~6.3% (15/233) threonine proteases (figure 1a, table insert in figure 1). The *I. scapularis* genome does not apparently encode for putatively inactive aspartic proteases in that of the 150 putative inactive protease homologs, none are aspartic, ~8% (12/150) are cysteine, 58.8% (86/150) metallo, 30% (45/150) serine and ~3.3% (5/150) are threonine proteases (figure 1b, table insert in figure 1). The designation of putatively active and inactive protease homologs is based on whether or not the candidate protease amino acid sequence possess classical catalytic residues that characterizes a family (MEROPS database, Rawlings et al., 2010). The identification of catalytic amino acid residues has been based on studies in other organisms that are without saying biologically distinct from ticks. There is potential that ticks utilize other amino acid catalytic residues, in which case the active and inactive protease homolog count presented here may not be consistent with actual events in ticks.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

On the MEROPS database, protease classes are classified into clans on the basis of structural similarity and or amino acid sequence similarity (Rawlings et al., 2010). Clans are further divided into families based on common ancestry. In cases of unclear ancestry, families are further classified into subfamilies. At the time of drafting this manuscript, the MEROPS database version 9.3 listed 205 protease families in 45 clans (Rawlings et al., 2010). The 233 putatively active protease sequences in the *I. scapularis* genome are from 25 annotated clans and 2 non-classified clans that segregate into 54 families (figure 2a). Likewise the 150 putatively inactive proteases are from 17 annotated clans in 29 families (figure 2b). As summarized in figure 2a, the *I. scapularis* genome does not encode for putatively active proteases in at least 6 protease families: 2 cysteine protease families, C56 and C64, 2 serine protease families, S24 and S28 and 2 metalloprotease families, M20 and M23. Whether or not this is an indication of *I. scapularis* evolutionary losing proteolytic functions in these protease families is unknown.

As stated above, the purpose of this study was to provide a snapshot of proteases encoded by the *I. scapularis* genome. Thus, we have not provided a full review of each of the clans and families of proteases that are encoded by the *I. scapularis* genome. We have rather provided a one stop semi-index catalogue of putative proteases that are encoded under each clan and family with guiding descriptions of probable functions where available. Where available published data in ticks is cited. For further reading, we advise the reader to consult expert work, which we have cited through out the manuscript.

ASPARTIC PROTEASES

Aspartic proteases are so called because they require an aspartic amino acid residue at the active site (Szecsi, 1992). AP enzymes are found in many different organisms, raging from vertebrates, invertebrates including ticks, plants and microbial organisms (Szecsi, 1992; Boldbaatar et al., 2006). They regulate numerous important functions in health and disease: including food digestion, protein turn over, blood pressure regulation (Szecsi, 1992; Chappell, 2010), cancer, Alzheimer's disease and malaria pathogenesis in humans (Wang et al., 2010), blood meal digestion and embryogenesis in ticks (Abreu et al., 2004; Horn et al., 2009), hemoglobin digestion in blood feeding arthropods such as ticks (Horn et al., 2009).

Aspartic protease enzymes are currently classified into 6 clans and 19 families (Rawlings et al., 2010). The *I. scapularis* genome encodes 14 enzymes from 2 clans, clan AA, which has 4 enzymes and clan AD, which has 10 enzymes (table 1). The 4 enzymes in clan AA are from family A1, while the 10 enzymes in clan AD are from family A22. Both families A1 and A22 contains 2 subfamilies: A1A and A1B in family A1 and A22A and A22B in family A22. The 4 *I. scapularis* family A1 enzymes in are from subfamily A1A typified by pepsin. Pepsin is a digestive enzyme that is optimally functional at an acidic pH (Table 1, Szecsi, 1992). Of the 4 *I. scapularis* pepsin-like enzymes, 2 (ISCW003823 and ISCW00023880) are provisionally identified as Cathepsin D. In mammals, cathepsin D is a multifunctional enzyme that is involved in regulating numerous biological processes ranging from protein turnover, apoptosis, cell signaling regulation, cancer pathogenesis (Vashishta et al., 2009; Benes et al., 2008). In ticks, cathepsin-D like enzymes have been linked to regulating important tick physiological functions: blood meal processing (Horn et al., 2009; Sojka et al., 2008; Boldbaatar et al., 2006) and reproduction through processing of vittelogenin and embryogenesis (Abreu et al., 2004; da Silva Vaz et al., 1998). The 10 *I. scapularis* aspartic proteases enzymes in clan AD are also from a single family, A22 (table 1), which is typified by presenilin-1 and the intramembrane protease aspartic protein (IMPAS)-1. Both presenilin and IMPAS are multi-pass transmembrane aspartic proteases that perform intramembranous proteolysis of substrates such as signal peptides cleaved of extracellular proteins and other peptides that are essential in numerous cell-signaling pathways (Martoglio and Golde 2003).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Of the 10 aspartic proteases enzymes in clan AD, 2 (ISCW006186 and ISWC018713) are provisionally identified as IMPAS-1 and -2, both of which are components of the intramembranous proteolytic machinery that controls the metabolism of numerous peptides and proteins that affect several signaling cascades (Martoglio and Golde 2003).

CYSTEINE PROTEASES

Cysteine proteases (CP) are so called because they utilize a cysteine amnio acid residue at the catalytic site (Dickinson, 2002). CP enzymes are conserved across taxa: from vertebrates, invertebrates, plants and viruses (Dickinson, 2002; Puente et al., 2003; 2005; Reiser et al., 2010; Mulenga et al., 2001; Reddy and Lerner, 2010). In animals, CP enzymes regulate numerous biological functions ranging from terminal protein degradation, mediating immune responses, apoptosis and inflammation regulation, pro-hormone processing, and extracellular matrix remodeling in bone development (Dickinson, 2002; Reiser et al., 2010; Leto et al., 2010). The MEROPS database (version 9.4, Rawlings et al., 2010) lists 72 CP enzyme families, 56 of which are assigned to 12 annotated CP enzyme clans, while the remaining 16 families belong non-annotated clans. The *I. scapularis* genome respectively encodes at least 44 putatively active (table 2) and 11 putatively inactive (table 3) CP enzyme homologs from 16 families clustered in 5 clans: CA, CD, CE, PB and PC (Table 2). We would like to note here that, of the 44 putatively active CP enzymes, only 31 were annotated as CP enzymes on version 9.4 MEROPS database. The additional 13 sequences include 4 sequences (marked with asterisks in table 2) ISCW000076, ISCW005981, ISCW013346 and ISCW024899, which were originally unclassified on version 9.3 MEROPS data that we have classified as under CP on the basis of their similarity to cathepsin-L enzymes in Gene Bank (not shown). The remaining 9 sequences marked with a plus sign (+) in table 2 were identified from the BLASTP data mining done in this study. Consistent with data in mammals where clan CA is the largest among CP enzyme clans (Puente et al., 2005), 33 of the 44 putatively active *I. scapularis* CP are from clan CA. The 31 CP enzymes in clan CA are from 9 families: C1, C2, C12, C19, C54, C65, C78, C85 and C86. The other 11 sequences are distributed in clan CD, which has 4 sequences in 3 families, C13, C14 and C50, clan CE, which has 5 sequences from family C48, clan PB which has one sequence each from families C44 and C89 (Table 2). Given the large number of *I. scapularis* CP enzymes, we have split discussions below based on clans for clarity and simplicity. Given that very little is known on the role of inactive CP enzyme homologs, we have not developed commentaries on putatively inactive enzymes.

Clan CA

Consistent with data in other animals where within clan CA, family C1 is the largest (Puente et al., 2005) with 11 followed by family C2, which has at least 9 enzyme sequences (Table 2). Of the remaining 13 CP enzymes in clan CA, 4 are from family C19, 2 each are from families C12, C54 and C86, and one each from families C65, C78 and C85 (Table 2). Family C1 is typified by papain in subfamily C1A or bleomycin hydrolase-like in subfamily C1B (Rawlings et al., 2010). The 11 *I. scapularis* CP enzymes in family C1 enzymes are papain-like enzymes in subfamily C1A (table 2). Data in animals show that papain-like enzymes are multifunctional regulating a broad-spectrum of biological functions ranging from protein turnover, antigen processing, and extracellular matrix modeling (Dickinson, 2002). Of the 11 *I. scapularis* papain-like enzymes (table 2), 6 have been provisionally identified. The 6 putatively identified CP enzymes include a cathepsin O-like (ISCW017550), cathepsin B-like (ISCW002679, ISCW000078, ISCW000080 and ISCW004002), enzymes which can function as digestive enzymes in the extracellular space and regulatory enzymes in the intracellular space (Reisser et al., 2010). The remaining sequence (ISCW007008) has been provisionally identified as the insect 26/29 kDa peptidase, an enzyme that appears to be widely conserved immune response molecule in

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

many arthropods (Saito et al., 1992; Fujimoto et al. 1999; Massova et al., 2009). The insect 26/29 kDa peptidase-homolog is comprised of 2 subunits, the N- and C-terminus 26 and 29 kDa fragment fused together (Fujimoto et al., 1999). Similar to data published by Fujimoto et al., (1999), the 29 kDa carboxy-terminal subunit of the *I. scapularis* 26/29 kDa peptidase is predominantly similar to cathepsin L (not shown).

At the time of preparing this manuscript a literature search revealed that the majority of tick CP enzyme encoding cDNAs cloned from ticks were provisionally identified as cathepsin L (Cruz, et al., 2010; Yamaji et al., 2009; Renard et al., 2000; 2002; Mulenga et al., 2001). Whether or not, this is an indication that cathepsin L cysteine proteases are predominantly expressed in ticks, remains to be explored. A limited number of studies have also shown that tick cathepsin L-like cysteine proteases may be involved in reproduction and embryogenesis through its role in proteolysis of vittelin and york (Renard et al., 2000; 2002), in blood meal processing through their role in proteolysis of hemoglobin (Yamaji et al., 2009), a major nutrient that ticks obtain from its blood meal.

While the lone *I. scapularis* 26/29 peptidase did not show any similarity to vertebrate sequences, it is highly conserved among invertebrates as revealed by sequence alignments that show that the *I. scapularis* 26/29 kDa peptidase is ~50–62% amino acid identical to other arthropod sequences (not shown). The high amino acid conservation of the 26/29 kDa protease in arthropods may signal the conservation of the molecular interaction cascade regulated by this enzyme among arthropods. The insect 26/29kDa protease is an extracellular immune responsive molecule in invertebrates that was shown to clear foreign proteins in the flesh fly (Saito et al., 1992; Fujimoto et al., 1999). It is interesting to note that a 26/29 kDa peptidase-like sequence directly submitted (AAO60045) to gene bank by one of the authors of this manuscript was cloned from partially fed *R. appendiculatus* midguts. While it is known that host red blood cells are internalized into tick gut epithelia cells for digestion (Sonenshine 1993), the fate of other protein components in the tick blood meal is unknown. It will be interesting to investigate if the tick 26/29 kDa peptidase is involved in clearance of these protein components of the tick blood meal.

The 9 sequences in family C2 show similarity to calcium dependent neutral cysteine proteases (calpain). Although the regulation of calpain activity remains unclear, these enzymes have been shown to be involved in a wide range of biological functions raging from signal transduction pathways and apoptosis (Sorimachi et al., 2010). Of the 13 *I. scapularis* other cysteine proteases in clan CA, 2 are from family C54 and 11 are from families C12, C19, C65, C78, C85 and C86. The 2 and 11 enzymes are respective putative house keeping enzymes, the autophagy (Kim et al., 2003) and ubiquitin (Shcherbik and Pestov, 2010) cellular homeostasis regulatory mechanisms. The autophagy and ubiquitin house keeping enzyme systems are involved in eliminating damaged, misfolded and long lived proteins, foreign proteins that are derived from pathogens and long lived organelles (Shcherbik and Pestov, 2010, Kim et al., 2003). The 2 sequences in family C54, ISCW004712 and ISCW010488 are provisionally identified as autophagy related proteins (*ATG*) 4B and 4D that are part the core molecular machinery of the autophagy system (Homma et al., 2010). It is noteworthy that, except for *P. humanus*, which encodes for 1 *ATG4* gene, genomes of other blood feeding arthropods, *Ae. aegypti* and *An. gambiae* encode for 2 *ATG4* genes similar to *I. scapularis* genome as opposed to at least 4 in homo sapiens (Mizushima and Levine, 2010). The ubiquitin system is reversible and starts with the target protein being tagged with ubiquitin by ubiquitination enzymes and ends with deubiquinating (DUB) enzymes freeing up ubiquitin to be recycled (Kim et al., 2003). Of the 11 putative molecular components of the *I. scapularis* ubiquitin (Ub) system, the lone sequence in family C78 (ISCW014202) is putatively identified Ub fold modifier protein (Sasakawa et al., 2006), which is involved in the covalent attachment of Ub to the target

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

protein while the remaining 10 sequences show similarity to DUB enzymes (Kim et al., 2003). In humans, 5 DUB families have been described of which 4 are papain-like cysteine proteases belonging to ubiquitin carboxyl-terminal hydrolases (UCH), the ubiquitin specific protease (USP/UBP), the ovarian tumor (OTU) containing domain, the Josephin containing domain UB enzymes (Kim et al., 2003). In the *I. scapularis* genome the 2 DUB enzymes sequences in family C12, ISCW022153 and ISCW000206 are respectively provisionally identified as UCH-like, while 2 of the 4 DUB enzyme sequences in family C19, ISCW010054 and ISCW019604 and one enzyme sequence in family C78, ISCW014202 are provisionally identified as USP/UBP-like. The lone enzyme sequences in families C65 (ISCW023231) and C85 (ISCW021968) are OTU-like and the 2 sequences in family C86 (ISCW006944 and ISCW017541) are Josephin-like. For in depth details, on molecular interaction cascades enzymes and biological functions of DUB enzymes, the reader is referred to an excellent review by Kim et al., (2003).

Clan CD—There are currently 6 annotated families in clan CD, families C11, C13, C14, C25, C50 and C80 that contain functionally distinct cysteine protease enzymes (Rawlings et al., 2010). The *I. scapularis* genome encodes 4 putatively active (table 2) and 3 inactive (table 3) enzymes in clan CD from families C13 (legumain-like), which has 2 active and 2 inactive enzyme sequences, C14 (caspase-like), which has one active and one inactive enzyme sequence, and C50 (separase-like), which contains one enzyme sequence. In family C13, ISCW015983 is provisionally identified as legumain-like, an enzyme, which in ticks is blood meal responsive and is potentially involved in processing of the tick blood meal (Alim et al., 2008, Sojka et al. 2008; Horn et al., 2009), while ISCW000202 is provisionally identified as glycosylphosphatidylinositol (GPI)-anchor transamidase, an enzyme that is involved in the biosynthesis of the GPI cell surface protein anchoring system (Jiang et al., 2007). In family C14, ISCW022545 is provisionally identified as caspase-8, an enzyme that plays numerous roles such as the execution-phase of cell apoptosis (Hedrick et al., 2010), and lastly, ISCW017453 in family C50 is provisionally identified as separin, an enzyme involved with segregation of DNA to opposite poles of the cell during anaphase of cell replication (Stemmann et a., 2005).

Clan CE—The MEROPS database (version 9.3) listed 6 families: C5, C48, C55, C57, C63 and C79 in clan CE (Rawlings et al., 2010). The *I. scapularis* genome encodes 5 active and one inactive clan CE enzyme sequences from family C48 (Ubiquitin-like protein [Ulp, also known as Small ubiquitin-related modifier [SUMO] or Sentrin [SENP] endopeptidases) (table 2) (Creton et al., 2010). Of the 5 active SUMO specific proteases, 4 are provisionally identified: 3 (ISCW003012, ISCW012734 and ISCW003013) as SUMO specific protease 1, and the fourth sequence ISCW014736 as SUMO specific protease 8. Like ubiquitinylation, which regulates protein turnover (Kim et al., 2003), sumoylation (the covalent attachment of the SUMO protein to target proteins) alter the function of numerous cellular proteins via post-translational modification and is involved in controlling numerous biological processes such as gene expression, genomic and chromosomal integrity, intracellular transport, and protein stability (Wilkinson and Henley, 2010; Andreou and Tavernarakis, 2009). This process is reversible via actions of the SUMO specific proteases (so called desumoylation) that allow the recycling of SUMO proteins (Kolli et al., 2010). Given the high cross-taxa conservation of the SUMO specific proteases, it is highly likely that tcik enzymes are functionally similar to their mammalian counterparts.

Families of I. scapularis CP from mixed clans—According to annotations by the MEROPS database (Rawlings et al., 2010), there are 3 mixed protease clans: clan PA containing serine and cysteine protease families, clan PB containing cysteine, serine and threonine protease families and, clan PC containing cysteine and serine protease families.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The *I. scapularis* genome encodes for one cysteine protease each from families C44 (ISCW015613) and C89 (ISCW010877) in clan PB and one potentially inactive cysteine protease sequence from family C56 in clan PC (tables 2 and 3). As shown in table 2, ISCW015613 is provisionally identified as an amidophospho-ribosyltransferase precursor, a housekeeping enzyme involved *in denovo* purine nucleotide biosynthesis (Gassmann et al., 1999). In the case of ISCW010877, it is provisionally identified as an acid ceramidase, an enzyme that catalyzes the degradation of the sphingosine and fatty acid lipid molecule (ceramide) into sphingosine and fatty acids. Actions of acid ceramidase are linked to multiple biological processes such as early embryo survival, apoptosis regulation and cancer metasis (Mao and Obeid, 2008). As both ISCW015613 and ISCW010877 are housekeeping genes conserved in most organims, they are likely to be functionally similar to mammalian enzymes.

METALLOPROTEASES

Metalloprotease (MP) enzymes contain endo- and exo-proteases that are characterized by catalytic mechanisms that require a divalent metal ion at the active site to hydrolyze the peptide bond (Ugalde et al., 2010). Based on data in animals, MP enzymes are involved in regulating a diverse of biological processes ranging from embryo development, peptide hormone processing, cytokine and growth factor secretion, nutrient absorption in the intestine, cell-cell fusion, cell adhesion and migration, structural composition of cells walls and extracellular matrix remodeling (Hatfield et al., 2010). The significance of MP enzymes in eukaryote biology can be attested to by aberrant functions of MP enzymes are implicated in numerous diseases such as arthritis, cancer, disorders of the urinary tract and central nervous system and cardiovascular diseases (Morancho et al., 2010; Klein and Bischoff, 2010). The version 9.3 MEROPS database (Rawlings et al., 2010) listed 56 metalloprotease families. Of these, 51 are from 14 clans and the rest are not assigned into any clans. The *I. scapularis* genome encodes at least 93 putatively active (tables 4 and 5) and 88 putatively inactive (table 6 and 7) MP enzyme homologs. The putatively active MP enzyme sequences are organized in 23 families (tables 4 and 5). Of the 23 MP enzyme families in the *I. scapularis* genome, 10 families (M1, M2, M3, M8, M10, M12, M13, M41, M43 and M48) are from clan MA, 2 families (M20 and M28) are from clan MH, single families each are from clans MC (family M14), ME (family M16), MF (family M17), MG (family M24), MK (family M22), MM (family M23) and MP (family M67). The remaining 3 families M49, M76 and M79 are not assigned to any clans (table 4). Please note that given the bulkiness of family M13, it is presented separately in table 5. It is interesting to note that in contrast to observations in blood feeding insects and homosapiens where serine proteases represent the majority of proteolytic enzymes followed by metalloproteases, the *I. scapularis* composition is in reverse order with the latter being more than the former. It is also interesting to not that ~67% (68/88) of non-protease MP enzyme homologs are from family M13 (table 7), which has 27 putatively active MP enzymes (table 5). The significance of the massive shrinkage of family M13 enzymes on the biology of ticks is unknown at this point. We also would like to note that we have not discussed clans and/or families that contain only non-protease homologs because very little is known on their role(s) in animal physiology.

Clan MA

Consistent with data in other animals where clan MA is the largest among all MP enzyme families (Ugalde et al., 2010), 60 of the 93 putative MP enzymes in the *I. scapularis* are from clan MA. The 60 enzyme sequences are from 10 families: M1, M2, M3, M8, M10, M12, M13, M41, M43 and M48 (table 4). We would like to notify the reader here that, we have discussed aminopeptidase (AMP)-like enzymes in family M1 together with other AMP enzymes from other families as group below. In family M2, which is typified by the blood pressure regulating enzymes, angiotensin-converting enzymes (ACE) 1 and 2 (Hanafy et al.,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

2010), the *I. scapularis* genome encodes 4 putatively active and 2 non-protease MP enzyme homologs (table 4). None of the 4 putatively active *I. scapularis* family M2 enzymes have been provisionally identified and thus their potential role(s) in tick physiology are unknown. However antibodies to a *R. microplus* midgut expressed ACE-like enzyme, *Bm*91 reduced the cattle tick's efficiency to digest its blood meal (Riding et al., 1994; Jarmey et al., 1995) suggesting the significance of these enzymes in tick physiology. In family M3, which is typified by thimet oligoprotease, an enzyme involved in antigen presentation, the *I. scapularis* genome encodes 2 putatively active enzymes (table 4) and one inactive enzyme (table 6). As shown in table 4, one of the putatively active enzymes in family M3, ISCW016804 is provisionally identified as the mitochondrial intermediate peptidase, a house keeping molecule of the mitochondrial protein import system required for maturation of nuclear proteins targeted to the mitochondrial (Gordon et al., 2000). The *I. scapularis* genome encodes a single enzyme in family M8 (table 4), which is typified by leishmanolysin, a virulence factor for *Leishmania* parasites (Pereira et al., 2010). In family M10, the *I. scapularis* genome encodes 2 putatively active enzymes, ISCW022900 and ISCW006136 that have been putatively identified as human matrix metalloprotease (MP) 2, an enzyme that regulates numerous processes including the break down of basement membrane barriers during angiogenesis (Chen et al., 2010) and *Drosophila melanogaster* MP1, which is involved in extracellular matrix remodeling during neural development, imarginal disc regeneration and larval tracheal growth and pupal head eversion (Page-McCaw et al., 2003; McClure et al., 2008). The *I. scapularis* genome encodes 12 putatively active and one inactive MP enzyme homologs in family M12. Family M12 is typified by astacin, an enzyme that regulate several functions in invertebrates such as cuticle formation and ecdysis in nematodes (Stepek et al., 2010a; 2010b) and toxin function in spiders (Trevisan et al., 2009). Of the 12 putatively active family M12 MP enzymes, 3 have been provisionally identified. The first enzyme sequence, ISCW016865 is provisionally identified as the tolkin-like enzyme that is involved in *Drosophila* larval and pupal development (Finelli et al., 1995), while the second enzyme, ISCW000910 has been identified as *Caenorhabditis elegans* ADAMT (MP enzyme containing disintegrin, MP and thrombospodin domain)-like enzyme which is involved in morphogenesis of *C. elegans* male copulatory organs (Kuno et al., 2002). The third enzyme, ISCW008310 has been identified as the *Ancylostuma caninum* matrix MP1, an enzyme involved in regulating infective larvae tissue migration (Williamson et al., 2006). There is precedence that family M12 metalloproteases may be involved in blood meal feeding regulation. Francischetti et al., (2003) showed that a recombinant M12 metalloprotease (AAP22067) had gelatinase and fibrinogenolytic activities. Fibrinogenolytic function is essential to the tick's ability to maintain host blood in a fluid state during tick feeding.

Consistent with data in other animals where the M13 family is the largest among clan MA families, nearly half; 27 of the 60 putatively active *I. scapularis* MP enzymes in clan MA are from family M13 (table 5). It is more dramatic in the case of putatively inactive homologs in that 68 of the 88 inactive MP enzyme sequences are from family M13 (table 7). In mammals, family M13 is typified by 7 membrane anchored proteins: neprilysins -1 and -2, endothelin converting enzyme -1 and -2, Kell blood group antigen, DINE and PHEX proteases all of which are involved in processing biologically active peptides that regulate of numerous biological functions such as inflammation, cardiovascular and immune systems, cell differentiation, pattern formation during development, and modulating extracellular matrix formation (Ugalde et al., 2010; Meyer et al., 2009; Isaac et al., 2009; Turner et al., 2001). Except 2 (ISCW003514 and ISCW01836) sequences that are provisionally identified as neprilysin-like (Turner et al., 2001), the rest of the 27 putatively active family M13 enzymes are not provisionally identified (table 5). The *I. scapularis* encodes 2 putatively active family M41 enzymes, which in mammals are involved in regulating mitochondrial homeostasis (Koppen et al., 2009). One of the 2 *I. scapularis* family M41 enzymes,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

ISCW014866 is putatively identified as paraplegin, a mitochondrial membrane anchored protease that degrades misfolded proteins that transport across the mitochondrial membrane (Koppen et al., 2009). Both the cytophagalysin-like family M43 and the Ste24 zinc MP-like family M48 encode single putatively active enzymes, with the latter (ISCW015561) being provisionally identified as the farnesylated-protein converting enzyme 1, an integral membrane protein that is involved in cell division (Freije et al., 1999)

It is interesting to note that, the enzymatic composition of *I. scapularis* family MA is dramatically different from other blood feeding arthropods. In *I. scapularis*, family M13 is the largest with 27 enzymes. However it is among the smaller families in mosquitoes: *Ae. aegypti* and *An. gambiae* and *Cx. pipiens quinquenfasciatus* and human body lice, *P. humanus* which respectively contain 5, 6, 9 and 7 members in family M13 (version 9.3 MEROPS database, Rawlings et al., 2010). On the other hand family M12, which in *I. scapularis* is the second largest with 14 enzymes, is the largest in *Ae. aegypti* and *An. gambiae* with 24 enzymes, *P. humanus* and *Cx. pipiens quinquenfasciatus* with 10 and 27 enzymes respectively (see the MEROPS database).

Clan MC—This clan contains enzymes that hydrolyze peptide bonds from the carboxy-terminal ends of peptides and proteins. It is composed of one family, M14 (carboxypeptidase, CP), which is further subdivided into 4 subfamilies: M14A (digestive CP enzymes), M14B (regulatory CP enzymes), M14C (bacterial peptidoglycan hydrolyzing enzymes) and M14D (cytosolic CP enzymes) (Rawlings et al., 2010). The *I. scapularis* genome encodes 3 putatively active (table 4) and 3 non-active (table 6) CP enzymes, all of which have not been provisionally identified. There is precedence that tick encoded CP enzymes may be associated with blood meal up take and/or digestion regulation. Ribeiro and Mather (1998) showed that a CP-like enzyme was present in *I. scapularis* saliva, while Motobu et al., (2007) described a blood meal responsive *Haemaphysalis longicornis* serine CP enzyme.

Clan ME—This clan contains 2 families; family M16, subdivided into pitrilysin-like enzymes in subfamily M16A and the mitochondrial processing peptidase in subfamily M16B and lastly family M44 typified by *Vaccinia virus* derived MP enzyme (MEROPs database version 9.3, Rawlings et al., 2010). The *I. scapularis* genome encodes 8 genes in clan ME from family M16, half of which are inactive MP enzyme homologs. Two of the 4 active family M16 MP enzymes, ISCW011140 and ISCW001592 have been respectively provisionally identified as nardilysin-like enzyme in subfamily M16A, which in humans is involved in regulation of axonal maturation and myelination in the central and peripheral nervous system (Ohno et al., 2009) and the mitochondrial processing peptidase beta-subunit in subfamily M16B, an enzyme involved in the processing of signal peptides of mitochondrial protein imports (Paces et al., 1993). We would like to point out one of the 4 putative active metalloprotease enzymes, which was provisionally identified as ymxg peptidase in subfamily M16A is potentially a non-tick enzyme as it did not show any similarity to any tick genes when scanned against the *I. scapularis* predicted genes. Instead, it was predominantly identical with bacteria derived sequences (not shown).

Aminopeptidase (AMP) enzymes in clans MA, MF, MG and MH—AMP enzymes are characterized by preferentially cleaving off amino acid residues at the amino-terminal end of the peptide or protein (Taylor, 1993). Because of this similarity in their proteolytic activity, we have elected to discuss AMP enzymes in the *I. scapularis* genome as a group to minimize redundancy. AMP enzymes are encoded by a broad spectrum of organisms including microbes, plants and animals (Taylor, 1993). At the time of this write up, the version 9.3 MEROPS database listed 12 AMP families in 6 MP clans (Rawlings et al., 2010). The *I. scapularis* genome encodes 17 putatively active (table 4) and 9 inactive (table

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

6) AMP enzyme homologs. Of the 17 putatively active AMP enzymes, 7 are from family M1 in clan MA, 4 each are from families M28 and M24 in clans MH and MG and the remaining 2 sequences are from family M17 in clan MF respectively (table 4).

Of the 7 AP enzymes in family M1, 3 enzymes, ISCW006685, ISCW022724 and ISCW008629 are provisionally identified as AMPN (Taylor, 1993), leukotriene A4 hydrolase (Haeggstrom et al., 2002; 2007), cytosol alanyl AP (Bukowska et al., 2003) respectively. AMPN-like enzymes are type II membrane glycoproteins that preferentially cleave neutral amino acid residues from amino-terminal ends of proteins and peptides (Taylor, 1993), while Leukotriene A4 hydrolase is a bi-functional molecule, which as an AMP enzyme has been shown to cleave derivatives of several amino acid residues especially alanine and arginine (Haeggstrom, 2002, 2007). Likewise, the cytosol alanyl AMP preferentially cleave the alanine amino acid residue from the amino-terminus region of peptides and proteins (Bukowska et al., 2003).

Although, the 4 enzyme sequences in family M28 AMP enzymes (ISCW022417, ISCW022420, ISCW022423 and ISCW002672) have not been provisionally identified on the MEROPS database, they all show similarity to subfamily M28B human glutamate carboxypeptidase. The human glutamate carboxypeptidase enzyme is involved in catalyzing the hydrolysis of the neurotransmitter N-acetyl-L-aspartyl-L-glutamate to N-acetyl-L-aspartate and L-glutamate, which is also a neurotransmitter (Mesters et al., 2006). The 4 *I. scapularis* AMP enzymes in family M24 have been provisionally identified as methionyl AMP1 (MAP, ISCW022177), Xaa-proline dipeptidases or prolidase (ISCW016014 and ISCW005124) and the mitochondrial intermediate peptidase (MIP, ISCW008932) (table 4). In mammals, the MAP-like enzymes hydrolyze the first methionine during protein synthesis (Taylor, 1993), the Xaa-proline dipeptidase cleaves dipeptides with proline or hydroxyproline at the carboxy terminus and is involved collagen turnover (Lupi et al., 2008). The MIP enzyme is involved in the maturation of nuclear pre-proteins that are trans-located to the mitochondrial. Finally, the 2 *I. scapularis* AMP enzymes in family M17 have been provisionally identified as animal (ISCW019719) and *Fasciola* spp (ISCW003100) derived Leucine AMP-like (LAMP) enzymes. LAMP enzymes preferentially catalyze the hydrolysis of leucine and other hydrophobic amino acid residues from the amino-terminus regions of peptides and proteins and are used as markers for normal kidney and liver function (Taylor, 1993).

Emerging data show that AMP enzymes may play key role(s) in regulating tick physiology (Hatta et al., 2006; 2007; 2009; 2010). For instance RNAi mediated silencing of a blood meal induced *H. longicornis* (BAF03566) LAMP-like enzyme that potentially play a role in tick vittelogenesis (Hatta et al., 2010). RNAi mediated silencing of BAF03566, which shows ~68% amino acid identity to *I. scapularis* LAMP (ISCW019719) caused *H. longicornis* fail to feed to repletion (Hatta et al., 2006). In a related study, protease inhibitor sensitivity profiling of tick gut proteolysis revealed that a LAMP-like enzyme is part of tick enzyme complex that is involved digestion of hemoglobin by *I. ricinus* ticks (Horn et al., 2010; Franta et al., 2010). Hemoglobin is the major protein nutrient that ticks obtain from host red blood cells. Thus the role of LAMP-like enzymes in degradation and assimilation of this tick nutrient source is central to tick survival. Thus, it will be interesting to clearly define the role of AMP enzymes in tick physiology.

Clan MK—This clan contains one family, M22, which is typified by O-sialoglycoprotein endopeptidases, which specifically cleave the polypeptide backbone of membrane glycoproteins that contain clusters of O-linked sialoglycans, (Nichols et al., 2006). The *I. scapularis* genome encodes 2 family M22 endopeptidases, ISCW006830 and ISCW023376, which have been respectively provisionally identified as Kinase-associated endopeptidase 1

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

and O-sialoglycoprotein endopeptidase-like 1, which are linked to genome maintenance (Hecker et al., 2009). The later is considered as the mitochondrial version of the Kae1 enzyme (Hecker et al., 2009).

Clan MM—This clan contains MP enzymes that belong to a single family, M50 that is typified by site-2 protease (S2P), which is involved in the feedback regulation of synthesis and up take of sterol and fatty acid through control of transcription factors and the sterol regulatory element binding proteins (Zelenski et al., 1999). The *I. scapularis* genome encodes one S2P-like protease (ISCW019725).

Clan MP—This clan is comprised of proteases from a single family, M67, which is part of the ubiquitin cellular regulatory system (Kim et al., 2003). As mentioned above, the ubiquitin system starts with the target protein being tagged with ubiquitin by ubiquitination enzymes and ends with deubiquinating (DUB) enzymes freeing up ubiquitin to be recycled (Kim et al., 2003). MP enzymes in family M67 are part of the DUB enzyme system. The *I. scapularis* genome encodes 5 putatively active (table 4) and 2 inactive (table 6) family M67 enzyme homologs. Of the 5 putatively active enzymes, ISCW017120 and ISCW019527 have been respectively provisionally identified as the Poh1 peptidase, an enzyme that is essential to cell viability (Gallery et al., 2007) and the Jab1/MPN domain containing metalloenzyme, which removes long ubiquitin chain from substrate proteins (Tran et al., 2003).

I. scapularis MP families not assigned into clans—The version 9.3 MEROPS database listed 5 MP families (M49, M73, M75, M76 and M45) with no assigned clan (Rawlings et al., 2010). The *I. scapularis* genome encodes one active enzyme each from families M49 and M76 in addition to 7 active enzymes in family M79 (table 4). The *I. scapularis* enzyme in family M49 respectively has been provisionally identified as a dipeptidyl-peptidase III-like, an enzyme that sequentially cleaves amino-terminal dipeptides from peptides and proteins (Baral et al., 2008) and has been shown to be involved in the metabolism of neuropeptides in insects (Isaac et al., 2009). The enzyme sequence in family M76 have the mitochondrial inner membrane protease ATP23 is involved in the assembly of mitochondrial ATPase (Zeng et al., 2006).

SERINE PROTEASES

Serine protease (SP) enzymes are so called because their reaction mechanism requiring a catalytic serine amino acid residue (Hedstrom, 2002; Page and Di Cera, 2008; Ekici et al., 2008). In mammals, SP enzymes regulate numerous physiological processes, food digestion, development, immune response, blood coagulation, fibrinolysis, reproduction, prohormone processing and, compliment fixation (Page and Di Cera, 2008). The significance of SP enzymes in human physiology have been attested to by the many diseases such as thrombosis, emphysema and that occur in case of mutations that may lead to excessive or lack of SP enzymatic activity (Maga et al., 2010). In tick physiology the interest in SP enzyme research has been premised on the potential of tick derived SP enzymes to be involved in acquisition and digestion of the blood meal and to regulate the tick-host interface (Miyoshi et al., 2004; 2008; Mulenga et al., 2001; 2003).

At the time of this write up, the MEROPS database version 9.3 listed 45 SP enzyme families, of which 39 are from 15 defined clans and the rest from non-defined clans. The *I. scapularis* genome encodes 63 putatively active enzymes (tables 8 and 9) and 49 inactive (table 10) SP enzyme homologs from 15 SP families. Of these, 11 families are from 7 clans (SB, SC, SF, SJ, SK, SP and ST) that contain SP enzymes only. Of the other 4 families, one each is from mixed clans, PA and PC, while the remaining 2 families are from unassigned

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(U) clans. It is interesting to note that the 5 SP enzyme clans (PA, SC, SK, SJ and ST) that have been identified in nearly all forms of life (Page and Di Cera, 2008) are also present in ticks. Consistent with observations in other organisms where SP enzymes constitutes a third of the degradome (Puente et al., 2003; 2005), the 63 putatively active SP enzymes represent ~28.3% (66/233) of the *I. scapularis* degradome. Given the huge expanse of SP enzyme families, we have developed narratives on SP enzyme based on specific clans encoded by the *I. scapularis* genome.

Clan PA

Members of this clan are the mostly studied proteases and are known to mediate proteolytic cascades that are essential to life such as blood coagulation and inflammation (Hedstrom, 2002; Page and Di Cera, 2008). The MEROPS database (version 9.3) listed a mixture of 6 cysteine and 12 serine protease families in clan PA (Rawlings et al., 2010). The *I. scapularis* genome encodes 28 putatively active (table 9) and 26 inactive (table 10) SP enzyme homologs from a single subfamily, S1. All *I. scapularis* S1 enzymes belong to subfamily S1A (table 9), which is typified by chymotrypsin or trypsin (Page and Di Cera, 2008). Consistent with data in other organisms where members of S1A subfamily are the majority among SP enzymes (Puente et al., 2003; 2005; Ross et al., 2003), the 28 *I. scapularis* S1A enzymes constitute ~42% of the 66 SP enzymes encoded by the *I. scapularis* genome. However, it is interesting to note that the 28 *I. scapularis* S1A enzymes that represent ~12% (28/233) of the *I. scapularis* genome is dramatically smaller than the S1A enzyme content in other blood feeding arthropods where it is ~22% (42/188) in *P. humanus*, ~44% in *An. gambiae* (237/537) and ~48% in *Ae. Aegypti* (210/435) and *Cx pipiens quinquenfasciatus* (210/438) genomes (MEROPS database version 9.3). Although the significance of the S1A enzyme family shrinkage in *I. scapularis* may be unknown, it may reflect fundamental physiological differences between short-term blood feeders such as mosquitoes where the S1A enzyme family is large and long-term blood feeders where the S1A enzymes is smaller. Except for ISCW018541 and ISCW021255, which have been respectively provisionally identified as *Tachypleus* (horseshoe) type-clotting enzyme, which is part of the horseshoe crab hemolymph coagulation cascade (Iwanaga et al., 1992; Seki et al., 1995) and tequila peptidase, which regulates long-term memory formation in *Drosophila* (Didelot et al., 2006), the rest of the sequences have not been provisionally identified. As the majority of *I. scapularis* S1A enzymes have not been provisionally identified, we are unable to speculate on probable biological functions of S1A enzymes in tick physiology. However there is indirect evidence that show that tick S1A enzymes may be involved in tick feeding regulation in that trypsin and chymotrypsin-like enzymes are among some of the proteins that are up regulated in response to tick feeding activity (Mulenga et al, 2003; Miyoshi et al., 2004; Motobu et al., 2007).

Clan PC

Similar to clan PA above, clan PC contains cysteine protease families and one serine protease family, S51 (Rawlings et al., 2010). As shown in table 5, the *I. scapularis* genome encodes one putatively active enzyme (ISCW005900) from family S51, which has been provisionally identified as alpha-aspartyl dipeptidase, an enzyme that preferentially cleaves dipeptides of an aspartic amino acid residue and any other amino acid residue in the P1 prime position (Lessy and Miller, 2000).

Clan SB

This clan contains SP enzymes from 2 families: S8 typified by subtilisin in subfamily S8A or kexin in subfamily S8B and the S53 family typified by sedolisin (Rawlings et al., 2010). The *I. scapularis* genome encodes 3 active (table 8) and 3 inactive (table 10) enzymes from family S8. Of the 3 putatively active S8 enzymes (table 8), ISCW006099 is provisionally

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

identified as the tripeptidyl-peptidase II, a SP aminopeptidase enzyme that is involved in protein turn over (Chuang et al., 2010; Zhang et al., 2010).

Clan SC

This clan is composed of 6 families, S9, S10, S15, S28, S33 and S37 (Rawlings et al., 2010). The *I. scapularis* genome encodes 20 active (table 8) and 16 inactive (table 10) enzymes from clan SC. The clan SC enzymes are distributed in 6 families: S9 which has a 9/7 (active/inactive) enzyme composition, S10 which contains 10/2 enzymes, S28 which has 0/2 enzyme content, and S33 which contain 6/3 content (Tables 8 and 10). Three of the 9 *I. scapularis* enzymes from family S9 have been provisionally identified. The first 2 enzymes have been identified as prolyl-oligopeptidase (ISCW002432) and acylaminoacyl-peptidase (ISCW021356), enzymes that are involved in regulating bio-peptide metabolism that is essential in numerous key biological functions such as angiogenesis and learning for example (Myohanen et al., 2010). The third enzyme has been identified as olipeptidase B (ISCW016734), an enzyme, which in protozoan parasites is important for host cell invasion (Yoshida and Cortez, 2008). We would like to note that ISCW016734 might be an *I. scapularis* rickettsial endosymbiont derived enzyme as it also shows 100% amino acid identity to several rickettsia sequences (not shown). For family S10, 6 putative enzymes ISCW003663, ISCW018397, ISCW007492, ISCW010371, ISCW014727 and ISCW003059 have been provisionally identified as vitellogenic-like carboxypeptidase (table 8), an enzyme, which in humans is suspected to be involved in regulating the immune system (Harris et al., 2006).

Clan SF

The MEROPS database lists 2 families in clan SF, S24 and S26 that are respectively typified by *E. coli* LexA repressor and signal peptidase I enzymes (Rawlings et al., 2010). The *I. scapularis* genome encodes 2 putatively inactive and one putatively active enzyme in family S26 (ISCW014839). This enzyme has been provisionally identified as the invertebrate signal peptidase, an enzyme that is highly conserved across taxa and functions to remove the amino-terminal targeting signals from secreted proteins (Killic et al., 2010).

Clan SJ

This clan contains unique ATP activated enzymes from 3 families, S16, S50 and S69 (Rawlings et al., 2010). The *I. scapularis* genome encodes 2 putatively active enzymes from family S16, which have been respectively provisionally identified as Lon-A peptidase and peroxisomal Lon peptidase. However when scanned against gene bank entries, the Lon-A-peptidase enzyme may be *rickettsial* derived, as it shows amino acid identity exclusively to bacteria sequences, while the peroxisomal Lon peptidase enzyme is tick derived as it showed a 100% match to ISCW0018144. The peroxisomal Lon peptidase is a widely conserved enzyme that has been shown in yeast to be involved in degradation of unfolded and non-assembled peroxisomal matrix proteins (Aksam et al., 2007).

Clan SK

This clan contains enzymes from 3 families, S14, S41 and S49 (Rawlings et al., 2010). The *I. scapularis* genome encodes one enzyme (ISCW01678) from family S14, which has provisionally been identified as type 1 Clp peptidase. The type 1 Clp peptidase is a universally conserved enzyme that has been shown to be involved in protein turn over house keeping function in bacteria (Frees et al., 1999). We would like to caution here that ISCW01678 might be derived from the *I. scapularis* endosymbiont bacteria, in that it also shows 100% amino acid identity to several rickettsial ATP-dependent Clp protease in gene bank (not shown).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Clan SP

This clan contains a single family of enzymes, S59 typified by nucleoporin 145 (*Homo sapiens*) (Rawlings et al., 2010). The *I. scapularis* genome encodes one enzyme sequence ISCW013985 from family S59, which has been provisionally identified as nucleoporin 145, a self-processing enzyme that is part of the nuclear pore complex that forms the gateway that regulates the flow of macromolecules between the cell nucleus and the cytoplasm (Heese-Peck and Raikhel, 1998).

Clan ST

This clan contains a single family of enzymes, S54 typified by Rhomboid proteases (Rawlings et al., 2010), which are regulated intramembrane proteases that cleave the transmembrane segments of integral membrane proteins (Freeman, 2009). The *I. scapularis* genome encodes one enzyme (ISCW021997) from family S54 that has not been provisionally identified.

Families of I. scapularis SP enzymes that do not belong to annotated clans

As shown in table 8, the *I. scapularis* genome encodes 6 and 1 enzyme sequences from families S63 and S72 that are respectively typified by the EGF-like module containing mucin-like hormone receptor-like 2 and dystroglycan (*Homo sapiens*) (Rawlings et al., 2010). The 6 enzyme sequences from family S63 (ISCW012721, ISCW010660, ISCW005937, ISCW006717, ISCW000464 and ISCW001355) have not been provisionally identified. On the other hand, the single sequence from family S72 (ISCW015049) has been provisionally identified as dystroglycan, an enzyme that in mammals plays diverse roles in development and homeostasis including basement membrane formation, epithelial morphogenesis, membrane stability, cell polarization, and cell migration (Masaki and Matsumura, 2010).

THREONINE PROTEASES

Threonine enzymes, which are contained within clan PB, are so called because their proteolytic function requires a threonine amino acid residue at the active site (Rawlings et al., 2010). They are mostly known for their role as major components of the proteasome catalytic subunits, which are part of the protein turnover system (Gomes et al., 2006). The version 9.3 MEROPS database listed 4 threonine protease families, T1, T2, T3 and T6 (Rawlings et al., 2010). The *I. scapularis* genome encodes 15 putatively active (Table 12) and 5 inactive (table 13) threonine protease enzymes from 4 families (figure 2). In family T1, which is typified by beta component archaean proteasome, the *I. scapularis* genome has 3 active and 4 inactive enzymes. In family T2 typified by glycosylasparaginase precursor, the *I. scapularis* genome has 8 active and one inactive enzyme. In family T3, which is typified by gamma-glutamyltransferase 1, the *I. scapularis* genome has 8 active and one inactive enzyme. Finally in family T6 typified by polycystin-1, the *I. scapularis* genome encoded one active enzyme. The 3 putatively active enzymes in family T1, ISCW019233, ISCW006222 and ISCW05987 have been provisionally identified as proteosome catalytic subunits 1, 2 and 3, enzymes that are part of the protein turn over housekeeping mechanism in the cell (Gomes et al., 2006). In the case of the 3 enzymes in family T2, ISCW001950, ISCW000866 and ISCW001928, they have been provisionally identified as glycosylasparaginase precursor, threonine type isoaspartyl dipeptidase and taspase-1, enzymes that play important role(s) in the metabolism of β-aspartyl peptides, cell proliferation (Noronkoski et al., 1998; Cantor et al., 2009). Of the 8 enzymes in family T3, 5 sequences ISCW000530, ICSW003582, ISCW009859, ISCW010660 and ISCW012371 have been provisionally identified as gamma-glutamyltransferase homologs, an enzyme that is present in cell membranes and is involved in cross cell membrane trafficking of amino

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

acids and peptides as well as glutathione metabolism (Langford et al., 2010; Leggatt and Iwama, 2009). There are no reports on the role threonine-like enzymes in tick physiology. However given that threonine enzymes appear to be highly conserved house keeping enzymes, it is more likely that tick threonine enzymes are functionally similar to mammalian homologs.

CONCLUSIONS

The 233 putatively active proteases presented here represent ~1.14% of the 20485 putative protein sequences in the IscW1.1 release of the *I. scapularis* genome annotation (http://www.vectorbase.org/Ixodes). Comparatively, the 1.1% protease content in the *I. scapularis* genome is smaller than what has been observed in other organisms, ~1.8% in humans and *Caenorhabditis elegans*, ~1.7% in yeast and ~3.4% in *Drosophila* (Southan, 2000). Similarly, the *I. scapularis* degradome appears to be smaller when compared to short-term blood feeding arthropods, body lice, *Pediculus humanus*, mosquitoes, *Culex quinquefasciatus, Anopheles gambiae* and *Aedes aegypti* where degradomes apparently account for ~1.6, 2, 2.4 and 3.7% of the total peptide content. The reader is cautioned here that mosquito degradome sizes discussed here were calculated based annotations contained in the version 9.3 MEROPS database (Rawlings et al., 2010) and peptide content estimates contained in vectorbase release at the time of preparing this manuscript, *Agam*P3.5 (*An. gambiae*), *Aaeg*L1.2 (*Ae. aegypti*), CpiPJ1.2 (*C. quinquefasciatus*) and *Phum*U1.2 (*P. humanus*) in vectorbase. Given that these databases are subject to revision, the degradome sizes are also subject to change. The 150 non-protease homologs in *I. scapularis* that represent ~0.65% of the *I. scapularis* genome is comparable to other non-protease content in genomes of other organisms, ~0.64% in *C quinquefasciatus* and *P. humanus*, ~1.06 and 1.7% in *Ae. aegypti* and *An. gambiae* respectively. Given that we do not have gene expression data in this study, whether or not non-protease homologs are expressed is unknown at this point. At the time of writing this manuscript, the *I. scapularis* genome was not completely annotated, in that there were still substantial amount of gaps. Thus, whether or not the 233 proteases and the 150 non-protease homologgs presented here represent the full repertoire of the *I. scapularis* degradome remains to be verified in future.

Except for *I. scapularis*, genomes of the more than 860 tick species that have been described this far have not been sequenced. Thus, the comparative relationship of degradomes between *I. scapularis* and other tick species will remain unknown. Results summarized in table 13 provide some insight on probable similarity levels between *I. scapularis* and other tick species degradomes. Data in table 13 show that ~12% (25/233) putatively active *I. scapularis* proteases have homologs in other tick species with amino acid identity levels of ≥60%. As is to be expected putative proteases of *I. ricinus*, which like *I. scapularis* belongs prostriata ticks show 95–99% amino acid identity levels (table 13). Based on these amino acid identity levels, we can confidently speculate that the *I. scapularis* degradome size described in this study informs on the probable size of *I. ricinus* degradome.

Acknowledgments

This research was supported by grants to “AM” from the Texas AgriLife Research Vector-Borne Disease Program and the NIH summer employment supplement to NIHAI074789 supported by the American Recovery and Reinvestment Act of 2009.

References

Abreu LA, Valle D, Manso PP, Façanha AR, Pelajo-Machado M, Masuda H, Masuda A, Vaz I Jr, Lenzi H, Oliveira PL, Logullo C. Proteolytic activity of *Boophilus microplus* Yolk pro-Cathepsin D

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(BYC) is coincident with cortical acidification during embryogenesis. Insect Biochem Mol Biol. 2004; 34:443–449. [PubMed: 15110865]

Aksam EB, Koek A, Kiel JA, Jourdan S, Veenhuis M, van der Klei IJ. A peroxisomal lon protease and peroxisome degradation by autophagy play key roles in vitality of Hansenula polymorpha cells. Autophagy. 2007; 3:96–105. [PubMed: 17172804]

Alim AM, Tsuji N, Miyoshi T, Islam MK, Huang X, Motobu M, Fujisaki K. Characterization of asparaginyl endopeptidase, legumain induced by blood feeding in the ixodid tick *Haemaphysalis longicornis*. Insect Biochem Mol Biol. 2007; 37:911–922. [PubMed: 17681230]

Alim MA, Tsuji N, Miyoshi T, Islam MK, Hatta T, Fujisaki K. Legumains from the hard tick *Haemaphysalis longicornis* play modulatory roles in blood feeding and gut cellular remodelling and impact on embryogenesis. Int J Parasitol. 2009; 39:97–107. [PubMed: 18718474]

Alim MA, Tsuji N, Miyoshi T, Islam MK, Hatta T, Yamaji K, Fujisaki K. Developmental stage- and organ-specific expression profiles of asparaginyl endopeptidases/legumains in the ixodid tick *Haemaphysalis longicornis*. J Vet Med Sci. 2008a; 70:1363–1366. [PubMed: 19122407]

Alim MA, Tsuji N, Miyoshi T, Islam MK, Huang X, Hatta T, Fujisaki K. HlLgm2, a member of asparaginyl endopeptidases/legumains in the midgut of the ixodid tick Haemaphysalis longicornis, is involved in blood-meal digestion. J Insect Physiol. 2008b; 54:573–585. [PubMed: 18222467]

Andreou AM, Tavernarakis N. SUMOylation and cell signalling. Biotechnol, J. 2009; 4:1740–1752. [PubMed: 19946876]

Baral PK, Jajcanin-Jozić N, Deller S, Macheroux P, Abramić M, Gruber K. The first structure of dipeptidyl-peptidase III provides insight into the catalytic mechanism and mode of substrate binding. J Biol Chem. 2008; 283:22316–22324. [PubMed: 18550518]

Benes P, Vetvicka V, Fusek M. Cathepsin D--many functions of one aspartic protease. Crit Rev Oncol Hematol. 2008; 68:12–28. [PubMed: 18396408]

Boldbaatar D, Sikalizyo SC, Battsetseg B, Xuan X, Fujisaki K. Molecular cloning and functional characterization of an aspartic protease from the hard tick Haemaphysalis longicornis. Insect Biochem Mol Biol. 2006; 36:25–36. [PubMed: 16360947]

Boldbaatar D, Sikalizyo SC, Battsetseg B, Xuan X, Fujisaki K. Molecular cloning and functional characterization of an aspartic protease from the hard tick Haemaphysalis longicornis. Insect Biochem Mol Biol. 2006; 36:25–36. [PubMed: 16360947]

Bos DH, Mayfield C, Minchella DJ. Analysis of regulatory protease sequences identified through bioinformatic data mining of the *Schistosoma mansoni* genome. BMC Genomics. 2009; 10:488. [PubMed: 19845954]

Bratton RL, Corey RG. Tick Borne disease. American family physician. 2005; 71:2323–2330. [PubMed: 15999870]

Bukowska A, Tadje J, Arndt M, Wolke C, Kähne T, Bartsch J, Faust J, Neubert K, Hashimoto Y, Lendeckel U. Transcriptional regulation of cytosol and membrane alanyl-aminopeptidase in human T cell subsets. Biol Chem. 2003; 384:657–665. [PubMed: 12751795]

Cantor JR, Stone EM, Chantranupong L, Georgiou G. The human asparaginase-like protein 1 hASRGL1 is an Ntn hydrolase with beta-aspartyl peptidase activity. Biochemistry. 2009; 48:11026–11031. [PubMed: 19839645]

Chappell KJ, Stoermer MJ, Fairlie DP, Young PR. West Nile Virus NS2B/NS3 protease as an antiviral target. Curr Med Chem. 2008; 15Chapter 21(Unit 21.10):2771–284. [PubMed: 18991636]

Chen L, DID, Luo G, Zheng L, Tan Y, Zhang X, Xu N. Immunochemical staining of MT2-MMP correlates positively to angiogenesis of human esophageal cancer. Anticancer Res. 2010; 30:4363–4368. [PubMed: 21036765]

Chuang CK, Rockel B, Seyit G, Walian PJ, Schönegge AM, Peters J, Zwart PH, Baumeister W, Jap BK. Hybrid molecular structure of the giant protease tripeptidyl peptidase II. Nat Struct Mol Biol. 2010; 17:990–996. [PubMed: 20676100]

Creton S, Jentsch S. SnapShot: The SUMO system. Cell. 2010; 143:848–848.e1. [PubMed: 21111242]

Cruz CE, Fogaça AC, Nakayasu ES, Angeli CB, Belmonte R, Almeida IC, Miranda A, Miranda MT, Tanaka AS, Braz GR, Craik CS, Schneider E, Caffrey CR, Daffre S. Characterization of proteinases from the midgut of *Rhipicephalus (Boophilus) microplus* involved in the generation of antimicrobial peptides. Parasit Vectors. 2010; 27:63. [PubMed: 20663211]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

da Silva Vaz I Jr, Logullo C, Sorgine M, Velloso FF, Rosa de Lima MF, Gonzales JC, Masuda H, Oliveira PL, Masuda A. Immunization of bovines with an aspartic proteinase precursor isolated from *Boophilus microplus* eggs. Vet Immunol Immunopathol. 1998; 66(3–4):331–341. [PubMed: 9880109]

Decrem Y, Beaufays J, Blasioli V, Lahaye K, Brossard M, Vanhamme L, Godfroid E. A family of putative metalloproteases in the salivary glands of the tick *Ixodes ricinus*. FEBS J. 2008; 275:1485–1499. [PubMed: 18279375]

Decrem Y, Mariller M, Lahaye K, Blasioli V, Beaufays J, Zouaoui BK, Vanhaeverbeek M, Cérutti M, Brossard M, Vanhamme L, Godfroid E. The impact of gene knock-down and vaccination against salivary metalloproteases on blood feeding and egg laying by *Ixodes ricinus*. Int J Parasitol. 2008; 38:549–60. [PubMed: 17959179]

Dickinson DP. Cysteine peptidases of mammals: their biological roles and potential effects in the oral cavity and other tissues in health and disease. Crit Rev Oral Biol Med. 2002; 13:238–275. [PubMed: 12090464]

Didelot G, Molinari F, Tchénio P, Comas D, Milhiet E, Munnich A, Colleaux L, Preat T. Tequila, a neurotrypsin ortholog, regulates long-term memory formation in *Drosophila*. Science. 2006; 313:851–853. [PubMed: 16902143]

Ekici OD, Paetzel M, Dalbey RE. Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. Protein Sci. 2008; 17:2023–2037. [PubMed: 18824507]

Finelli AL, Xie T, Bossie CA, Blackman RK, Padgett RW. The tolkin gene is a tolloid/BMP-1 homologue that is essential for Drosophila development. Genetics. 1995; 141:271–281. [PubMed: 8536976]

Fish D, Childs JE. Community-based prevention of Lyme disease and other tick-borne diseases through topical application of acaricide to white-tailed deer: background and rationale. Vector Borne Zoonotic Dis. 2009; 9:357–364. [PubMed: 19650729]

Francischetti IM, Mather TN, Ribeiro JM. Cloning of a salivary gland metalloprotease and characterization of gelatinase and fibrin(ogen)lytic activities in the saliva of the Lyme disease tick vector Ixodes scapularis. Biochem Biophys Res Commun. 2003; 305(4):869–875. [PubMed: 12767911]

Franta Z, Frantova H, Konvickova J, Horn M, Sojka D, Mares M, Kopacek P. Dynamics of digestive proteolytic system during blood feeding of the hard tick *Ixodes ricinus*. Parasit Vectors. 2010; 3:119. [PubMed: 21156061]

Freeman M. Rhomboids: 7 years of a new protease family. Semin Cell Dev Biol. 2009; 20:231–239. [PubMed: 19022390]

Frees D, Ingmer H. ClpP participates in the degradation of misfolded protein in Lactococcus lactis. Mol Microbiol. 1999; 31:79–87. [PubMed: 9987112]

Freije JM, Blay P, Pendás AM, Cadiñanos J, Crespo P, López-Otín C. Identification and chromosomal location of two human genes encoding enzymes potentially involved in proteolytic maturation of farnesylated proteins. Genomics. 1999; 58:270–280. [PubMed: 10373325]

Gallery M, Blank JL, Lin Y, Gutierrez JA, Pulido JC, Rappoli D, Badola S, Rolfe M, Macbeth KJ. The JAMM motif of human deubiquitinase Poh1 is essential for cell viability. Mol Cancer Ther. 2007; 6:262–268. [PubMed: 17237285]

Gassmann MG, Stanzel A, Werner S. Growth factor-regulated expression of enzymes involved in nucleotide biosynthesis: a novel mechanism of growth factor action. Oncogene. 1999; 18:6667–6676. [PubMed: 10597272]

Gomes AV, Zong C, Ping P. Protein degradation by the 26S proteasome system in the normal and stressed myocardium. Antioxid Redox Signal. 2006; 8:1677–1691. [PubMed: 16987021]

Gordon DM, Dancis A, Pain D. Mechanisms of mitochondrial protein import. Essays Biochem. 2000; 36:61–73. [PubMed: 12471903]

Haeggström JZ, Tholander F, Wetterholm A. Structure and catalytic mechanisms of leukotriene A4 hydrolase. Prostaglandins Other Lipid Mediat. 2007; 83:198–202. [PubMed: 17481555]

Haeggström JZ, Kull F, Rudberg PC, Tholander F, Thunnissen MM. Leukotriene A4 hydrolase. Prostaglandins Other Lipid Mediat. 2002; 68–69:495–510.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Hanafy S, Tavasoli M, Jamali F. Inflammation Alters Angiotensin Converting Enzymes (ACE and ACE-2) Balance in Rat Heart. Inflammation. 2010 [Epub ahead of print].

Harris J, Schwinn N, Mahoney JA, Lin HH, Shaw M, Howard CJ, da Silva RP, Gordon S. A vitellogenic-like carboxypeptidase expressed by human macrophages is localized in endoplasmic reticulum and membrane ruffles. Int J Exp Pathol. 2006; 87:29–39. [PubMed: 16436111]

Hatfield KJ, Reikvam H, Bruserud O. The crosstalk between the matrix metalloprotease system and the chemokine network in acute myeloid leukemia. Curr Med Chem. 2010; 17:4448–4461. [PubMed: 21062258]

Hatta T, Tsuji N, Miyoshi T, Alim MA, Islam MK, Fujisaki K. Leucine aminopeptidase in the ixodid tick *Haemaphysalis longicornis*: endogenous expression profiles in midgut. J Vet Med Sci. 2009; 71:589–594. [PubMed: 19498284]

Hatta T, Kazama K, Miyoshi T, Umemiya R, Liao M, Inoue N, Xuan X, Tsuji N, Fujisaki K. Identification and characterisation of a leucine aminopeptidase from the hard tick *Haemaphysalis longicornis*. Int J Parasitol. 2006; 36:1123–1132. [PubMed: 16814790]

Hatta T, Tsuji N, Miyoshi T, Islam MK, Alim MA, Yamaji K, Fujisaki K. Leucine aminopeptidase, HlLAP, from the ixodid tick Haemaphysalis longicornis, plays vital roles in the development of oocytes. Parasitol Int. 2010; 59:286–9. [PubMed: 20230906]

Hatta T, Umemiya R, Liao M, Gong H, Harnnoi T, Tanaka M, Miyoshi T, Boldbaatar D, Battsetseg B, Zhou J, Xuan X, Tsuji N, Taylor D, Fujisaki K. RNA interference of cytosolic leucine aminopeptidase reduces fecundity in the hard tick, *Haemaphysalis longicornis*. Parasitol Res. 2007; 100:847–854. [PubMed: 17136388]

Hecker A, Graille M, Madec E, Gadelle D, Le Cam E, van Tilbergh H, Forterre P. The universal Kae1 protein and the associated Bud32 kinase (PRPK), a mysterious protein couple probably essential for genome maintenance in Archaea and Eukarya. Biochem Soc Trans. 2009; 37:29–35. [PubMed: 19143597]

Hedrick SM, Ch'en IL, Alves BN. Intertwined pathways of programmed cell death in immunity. Immunol Rev. 2010; 236:41–53. [PubMed: 20636807]

Hedstrom L. An overview of serine proteases. Curr Protoc Protein Sci. 2002 Feb.2002

Heese-Peck A, Raikhel NV. The nuclear pore complex. Plant Mol Biol. 1998; 38:145–162. [PubMed: 9738965]

Horn M, Nussbaumerová M, Sanda M, Kovárová Z, Srba J, Franta Z, Sojka D, Bogyo M, Caffrey CR, Kopácek P, Mares M. Hemoglobin digestion in blood-feeding ticks: mapping a multipeptidase pathway by functional proteomics. Chem Biol. 2009a; 16:1053–1063. [PubMed: 19875079]

Horn M, Nussbaumerová M, Sanda M, Kovárová Z, Srba J, Franta Z, Sojka D, Bogyo M, Caffrey CR, Kopácek P, Mares M. Hemoglobin digestion in blood-feeding ticks: mapping a multipeptidase pathway by functional proteomics. Chem Biol. 2009b; 16:1053–1063. [PubMed: 19875079]

Isaacb RE, Blandb ND, Shirras AD. Neuropeptidases and the metabolic inactivation of insect neuropeptides. Gen Comp Endocrinol. 2009; 162:8–17. [PubMed: 19135055]

Iwanaga S, Miyata T, Tokunaga F, Muta T. Molecular mechanism of hemolymph clotting system in Limulus. Thromb Res. 1992; 68:1–32. [PubMed: 1448796]

Jarmey JM, Riding GA, Pearson RD, McKenna RV, Willadsen P. Carboxydipeptidase from *Boophilus microplus*: a "concealed" antigen with similarity to angiotensin-converting enzyme. Insect Biochem Mol Biol. 1995; 25:969–974. [PubMed: 8541886]

Jiang WW, Zahurak M, Zhou ZT, Park HL, Guo ZM, Wu GJ, Sidransky D, Trink B, Califano JA. Alterations of GPI transamidase subunits in head and neck squamous carcinoma. Mol Cancer. 2007; 6:74. [PubMed: 18028549]

Jongejan F, Uilenberg G. The global importance of ticks Parasitology. 2004; 129(Suppl):S3–14.

Kilic A, Klose S, Dobberstein B, Knust E, Kapp K. The Drosophila Crumbs signal peptide is unusually long and is a substrate for signal peptide peptidase. Eur J Cell Biol. 2010; 89:449–461. [PubMed: 20189678]

Kim JH, Park KC, Chung SS, Bang O, Chung CH. Deubiquitinating enzymes as cellular regulators. J Biochem. 2003; 134:9–18. [PubMed: 12944365]

Klein T, Bischoff R. Physiology and pathophysiology of matrix metalloproteases. Amino Acids. 2010 [Epub ahead of print].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Kolli N, Mikolajczyk J, Drag M, Mukhopadhyay D, Moffatt N, Dasso M, Salvesen G, Wilkinson KD. Distribution and paralogue specificity of mammalian deSUMOylating enzymes. Biochem J. 2010; 430:335–44. [PubMed: 20590526]

Koppen M, Bonn F, Ehses S, Langer T. Autocatalytic processing of m-AAA protease subunits in mitochondria. Mol Biol Cell. 2009; 20:4216–4224. [PubMed: 19656850]

Kuno K, Baba C, Asaka A, Matsushima C, Matsushima K, Hosono R. The Caenorhabditis elegans ADAMTS family gene adt-1 is necessary for morphogenesis of the male copulatory organs. J Biol Chem. 2002; 277:12228–12236. [PubMed: 11796737]

Langford MP, Redmond P, Chanis R, Misra RP, Redens TB. Glutamate, excitatory amino acid transporters, Xc- antiporter, glutamine synthetase, and gamma-glutamyltranspeptidase in human corneal epithelium. Curr Eye Res. 2010; 35:202–211. [PubMed: 20373878]

Lassy RA, Miller CG. Peptidase E, a peptidase specific for N-terminal aspartic dipeptides, is a serine hydrolase. J Bacteriol. 2000; 182:2536–2543. [PubMed: 10762256]

Leggatt RA, Iwama GK. Exogenous glutathione can increase glutathione levels in tissues of rainbow trout (Oncorhynchus mykiss) through extracellular breakdown and intracellular synthesis. Comp Biochem Physiol C Toxicol Pharmacol. 2009; 150:322–328. [PubMed: 19477299]

Leto G, Sepporta MV, Crescimanno M, Flandina C, Tumminello FM. Cathepsin L in metastatic bone disease: therapeutic implications. Biol Chem. 2010; 391:655–664. [PubMed: 20370324]

Lupi A, Tenni R, Rossi A, Cetta G, Forlino A. Human prolidase and prolidase deficiency: an overview on the characterization of the enzyme involved in proline recycling and on the effects of its mutations. Amino Acids. 2008; 35:739–752. [PubMed: 18340504]

Maga TK, Nishimura CJ, Weaver AE, Frees KL, Smith RJ. Mutations in alternative pathway complement proteins in American patients with atypical hemolytic uremic syndrome. Hum Mutat. 2010; 31:E1445–1460. [PubMed: 20513133]

Mao C, Obeid LM. Ceramidases: regulators of cellular responses mediated by ceramide, sphingosine, and sphingosine-1-phosphate. Biochim Biophys Acta. 2008; 1781:424–434. [PubMed: 18619555]

Martoglio B, Golde TE. Intramembrane-cleaving aspartic proteases and disease: presenilins, signal peptide peptidase and their homologs. Hum Mol Genet. 2003; 2:R201–20. 1. [PubMed: 12966028]

Masaki T, Matsumura K. Biological role of dystroglycan in Schwann cell function and its implications in peripheral nervous system diseases. J Biomed Biotechnol. 2010; 2010:article# 740403.

Másová A, Sindelka R, Kubista M, Kindl J, Jirácek J. Gene expression responses in larvae of the fleshfly Sarcophaga bullata after immune stimulation. Folia Biol (Praha). 2009; 55:98–106. [PubMed: 19545489]

McClure KD, Sustar A, Schubiger G. Three genes control the timing, the site and the size of blastema formation in *Drosophila*. Dev Biol. 2008; 319:68–77. [PubMed: 18485344]

Meyer H, Panz M, Zmojdzian M, Jagla K, Paululat A. Neprilysin 4, a novel endopeptidase from Drosophila melanogaster, displays distinct substrate specificities and exceptional solubility states. J Exp Biol. 2009; 212(Pt 22):3673–3683. [PubMed: 19880729]

Miyoshi T, Tsuji N, Islam MK, Kamio T, Fujisaki K. Cloning and molecular characterization of a cubilin-related serine proteinase from the hard tick Haemaphysalis longicornis. Insect Biochem Mol Biol. 2004; 34:799–808. [PubMed: 15262284]

Miyoshi T, Tsuji N, Islam MK, Alim MA, Hatta T, Huang X, Fujisaki K. A set of serine proteinase paralogs are required for blood-digestion in the ixodid tick *Haemaphysalis longicornis*. Parasitol Int. 2008; 57:499–505. [PubMed: 18775510]

Miyoshi T, Tsuji N, Islam MK, Kamio T, Fujisaki K. Gene silencing of a cubilin-related serine proteinase from the hard tick Haemaphysalis longicornis by RNA interference. J Vet Med Sci. 2004; 66:1471–1473. [PubMed: 15585970]

Mizushima N, Levine B. Autophagy in mammalian development and differentiation. Nat Cell Biol. 2010; 12:823–830. [PubMed: 20811354]

Morancho A, Rosell A, García-Bonilla L, Montaner J. Metalloproteinase and stroke infarct size: role for anti-inflammatory treatment? Ann NY Acad Sci. 2010; 1207:123–133. [PubMed: 20955435]

Motobu M, Tsuji N, Miyoshi T, Huang X, Islam MK, Alim MA, Fujisaki K. Molecular characterization of a blood-induced serine carboxypeptidase from the ixodid tick *Haemaphysalis longicornis*. FEBS J. 2007; 274:3299–3312. [PubMed: 17542992]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Motobu M, Tsuji N, Miyoshi T, Huang X, Islam MK, Alim MA, Fujisaki K. Molecular characterization of a blood-induced serine carboxypeptidase from the ixodid tick Haemaphysalis longicornis. FEBS J. 2007; 274:3299–3312. [PubMed: 17542992]

Mulenga A, Tsuda A, Sugimoto C, Onuma M. Blood meal acquisition by ticks; molecular advances and implications for vaccine development. Jpn J Vet Res. 2002; 49:261–272. [PubMed: 11949474]

Mulenga A, Sugimoto C, Onuma M. Characterization of proteolytic enzymes expressed in the midgut of *Haemaphysalis longicornis*. Jpn J Vet Res. 1999; 46:179–184. [PubMed: 10414026]

Mulenga A, Misao O, Sugimoto C. Three serine proteinases from midguts of the hard tick *Rhipicephalus appendiculatus*; cDNA cloning and preliminary characterization. Exp Appl Acarol. 2003; 29(1–2):151–164. [PubMed: 14580067]

Mulenga A, Sugimoto C, Ingram G, Ohashi K, Misao O. Characterization of two cDNAs encoding serine proteinases from the hard tick *Haemaphysalis longicornis*. Insect Biochem Mol Biol. 2001; 31:817–285. [PubMed: 11378417]

Mulenga A, Sugimoto C, Ingram G, Ohashi K, Onuma M. Molecular cloning of two *Haemaphysalis longicornis* cathepsin L-like cysteine proteinase genes. J Vet Med Sci. 1999; 61:497–502. [PubMed: 10379941]

Myöhänen T, Tenorio-Laranga J, Jokinen B, Vázquez-Sánchez R, Moreno-Baylach M, García-Horsman J, Männistö P. Prolyl oligopeptidase induces angiogenesis both in vitro and in vivo in a novel regulatory manner. Br J Pharmacol. 2010 [Epub ahead of print].

Nene V. Tick genomics--coming of age. Front Biosci. 2009; 14:2666–2673. [PubMed: 19273226]

Nichols CE, Johnson C, Lockyer M, Charles IG, Lamb HK, Hawkins AR, Stammers DK. Structural characterization of *Salmonella typhimurium* YeaZ, an M22 O-sialoglycoprotein endopeptidase homolog. Proteins. 2006; 64:111–123. [PubMed: 16617437]

Noronkoski T, Stoineva IB, Ivanov IP, Petkov DD, Mononen I. Glycosylasparaginase-catalyzed synthesis and hydrolysis of beta-aspartyl peptides. J Biol Chem. 1998; 273:26295–26297. [PubMed: 9756857]

Ohno M, Hiraoka Y, Matsuoka T, Tomimoto H, Takao K, Miyakawa T, Oshima N, Kiyonari H, Kimura T, Kita T, Nishi E. Nardilysin regulates axonal maturation and myelination in the central and peripheral nervous system. Nat Neurosci. 2009; 12:1506–1513. [PubMed: 19935654]

Paces V, Rosenberg LE, Fenton WA, Kalousek F. The beta subunit of the mitochondrial processing peptidase from rat liver: cloning and sequencing of a cDNA and comparison with a proposed family of metallopeptidases. Proc Natl Acad Sci USA. 1993; 90:5355–5358. [PubMed: 8506385]

Page MJ, Di Cera E. Serine peptidases: classification, structure and function. Cell Mol Sci. 2008; 65:1220–1236.

Pagel VZJ, 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(12):1297–1305. [PubMed: 17624352]

Page-McCaw A, Serano J, Santé JM, Rubin GM. *Drosophila* matrix metalloproteinases are required for tissue remodeling, but not embryonic development. Dev Cell. 2003; 4:95–106. [PubMed: 12530966]

Pereira FM, Santos-Mallet JR, Branquinha MH, d'Avila-Levy CM, Santos AL. Influence of leishmanolysin-like molecules of Herpetomonas samuelpessoai on the interaction with macrophages. Microbes Infect. 2010; 12(12–13):1061–1070. [PubMed: 20670690]

proteinase are probably derived from a common precursor protein. J Biochem. 125:566–573. [PubMed: 10050046]

Puente XS, Gutiérrez-Fernández A, Ordóñez GR, Hillier LW, López-Otín C. Comparative genomic analysis of human and chimpanzee proteases. Genomics. 2005; 86:638–647. [PubMed: 16162398]

Puente XS, López-Otín C. A genomic analysis of rat proteases and protease inhibitors. Genome Res. 2004; 14:609–622. [PubMed: 15060002]

Puente XS, Sánchez LM, Overall CM, López-Otín C. Human and mouse proteases: a comparative genomic approach. Nat Rev Genet. 2003; 4:544–558. [PubMed: 12838346]

Quesada V, Ordóñez GR, Sánchez LM, Puente XS, López-Otín C. The Degradome database: mammalian proteases and diseases of proteolysis. Nucleic Acids Res. 2009; 37(Database issue):D239–243. [PubMed: 18776217]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Rawlings ND, Barrett AJ, Bateman A. MEROPS: the peptidase database. Nucleic Acids Res. 2010; 38(Database issue):D227–233. [PubMed: 19892822]

Reddy VB, Lerner EA. Plant cysteine proteases that evoke itch activate protease-activated receptors. Br J Dermatol. 2010; 163:532–535. [PubMed: 20491769]

Reiser J, Adair B, Reinheckel T. Specialized roles for cysteine cathepsins in health and disease. J Clin Invest. 2010; 120:3421–3431. [PubMed: 20921628]

Renard G, Garcia JF, Cardoso FC, Richter MF, Sakanari JA, Ozaki LS, Termignoni C, Masuda A. Cloning and functional expression of a *Boophilus microplus* cathepsin L-like enzyme. Insect Biochem Mol Biol. 2000; 30:1017–1026. [PubMed: 10989288]

Renard G, Lara FA, de Cardoso FC, Miguens FC, Dansa-Petretski M, Termignoni C, Masuda A. Expression and immunolocalization of a *Boophilus microplus* cathepsin L-like enzyme. Insect Mol Biol. 2002; 11:325–328. [PubMed: 12144697]

Ribeiro JM, Mather TN. *Ixodes scapularis*: salivary kininase activity is a metallo dipeptidyl carboxypeptidase. Exp Parasitol. 1998; 89:213–221. [PubMed: 9635445]

Riding GA, Jarmey J, McKenna RV, Pearson R, Cobon GS, Willadsen P. A protective “concealed” antigen from Boophilus microplus. Purification, localization, and possible function. J Immunol. 1994; 153:5158–5166. [PubMed: 7963573]

Ross J, Jiang H, Kanost MR, Wang Y. Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationships. Gene. 2003; 304:117–131. [PubMed: 12568721]

Saito H, Kurata S, Natori S. Purification and characterization of a hemocyte proteinase of Sarcophaga, possibly participating in elimination of foreign substances. Eur J Biochem. 1992; 209:939–944. [PubMed: 1425700]

Sasakawa H, Sakata E, Yamaguchij Y, Komatsuj M, Tatsumij K, Kominami E, Tanaka K, Kato K. Solution structure and dynamics of Ufm1, a ubiquitin-fold modifier 1. Biochem Biophys Res Commun. 2006; 343:21–26. [PubMed: 16527251]

Seixas A, Leal AT, Nascimento-Silva MC, Masuda A, Termignoni C, da Silva Vaz I Jr. Vaccine potential of a tick vitellin-degrading enzyme (VTDCE). Vet Immunol Immunopathol. 2008; 124(3–4):332–340. [PubMed: 18490061]

Seki N, Muta T, Oda T, Iwaki D, Kuma K, Miyata T, Iwanaga S. Horseshoe crab (1,3)-beta-D-glucan-sensitive coagulation factor G. A serine protease zymogen heterodimer with similarities to beta-glucan-binding proteins. J Biol Chem. 1994

Shcherbik N, Pestov DG. Ubiquitin and ubiquitin-like proteins in the nucleolus: multitasking tools for a ribosome factory. Genes Cancer. 2010; 1:681–689. [PubMed: 21113400]

Sojka D, Franta Z, Horn M, Hajdusek O, Caffrey CR, Mares M, Kopácek P. Profiling of proteolytic enzymes in the gut of the tick *Ixodes ricinus* reveals an evolutionarily conserved network of aspartic and cysteine peptidases. Parasit Vectors. 2008; 1:7. [PubMed: 18348719]

Sonenshine, DE. Biology of ticks. Oxford University Press; Oxford: 1993.

Sorimachi H, Hata S, Ono Y. Expanding members and roles of the calpain superfamily and their genetically modified animals. Exp Anim. 2010; 59:549–566. [PubMed: 21030783]

Southan C. Assessing the protease and protease inhibitor content of the human genome. J Pept Sci. 2000; 6:453–458. [PubMed: 11016882]

Stemmann O, Boos D, Gorr IH. Rephrasing anaphase: separase FEARs shugoshin. Chromosoma. 2005; 113:409–417. [PubMed: 15703941]

Stepek G, McCormack G, Birnie AJ, Page AP. The astacin metalloprotease moulting enzyme NAS-36 is required for normal cuticle ecdysis in free-living and parasitic nematodes. Parasitology. 2010a; 27:1–12.

Stepek G, McCormack G, Page AP. Collagen processing and cuticle formation is catalysed by the astacin metalloprotease DPY-31 in free-living and parasitic nematodes. Int J Parasitol. 2010b; 40:533–542. [PubMed: 19883650]

Szecsi PB. The aspartic proteases. Scand J Clin Lab Invest Suppl. 1992; 210:5–22. [PubMed: 1455179]

Taylor A. Aminopeptidases: Structure and function. FASEB J. 1993; 7:290–298. [PubMed: 8440407]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Tran HJ, Allen MD, Löwe J, Bycroft M. Structure of the Jab1/MPN domain and its implications for proteasome function. Biochemistry. 2003; 42:11460–11465. [PubMed: 14516197]

Trenholme KR, Brown CL, Skinner-Adams TS, Stack C, Lowther J, To J, Robinson MW, Donnelly SM, Dalton JP, Gardiner DL. Aminopeptidases of malaria parasites: new targets for chemotherapy. Infect Disord Drug Targets. 2010; 10:217–225. [PubMed: 20334618]

Ugalde AP, Ordóñez GR, Quirós PM, Puente XS, López-Otín C. Metalloproteases and the degradome. Methods Mol Biol. 2010; 622:3–29. [PubMed: 20135273]

Vashishta A, Ohri SS, Vetvicka V. Pleiotropic effects of cathepsin D. Endocr. Metab Immune Disord Drug Targets. 2009; 9:385–391.

Wang S, Wang R, Chen L, Bennett DA, Dickson DW, Wang DS. Expression and functional profiling of neprilysin, insulin-degrading enzyme, and endothelin-converting enzyme in prospectively studied elderly and Alzheimer's brain. J Neurochem. 2010; 115:47–57. [PubMed: 20663017]

Wilkinson KA, Henley JM. Mechanisms, regulation and consequences of protein SUMOylation. Biochem J. 2010; 428:133–145. [PubMed: 20462400]

Williamson AL, Brindley PJ, Loukas A. Hookworm cathepsin D aspartic proteases: contributing roles in the host-specific degradation of serum proteins and skin macromolecules. Parasitology. 2003; 126(Pt 2):179–185. [PubMed: 12636356]

Yamaji K, Tsuji N, Miyoshi T, Islam MK, Hatta T, Alim MA, Takenaka A, Fujisaki K. Hemoglobinase activity of a cysteine protease from the ixodid tick *Haemaphysalis longicornis*. Parasitol Int. 2009; 58:232–237. [PubMed: 19446040]

Yoshida N, Cortez M. Trypanosoma cruzi: parasite and host cell signaling during the invasion process. Subcell Biochem. 2008; 47:82–91. [PubMed: 18512343]

Zelenski NG, Rawson RB, Brown MS, Goldstein JL. Membrane topology of S2P, a protein required for intramembranous cleavage of sterol regulatory element-binding proteins. J Biol Chem. 1999; 274:21973–21980. [PubMed: 10419520]

Zeng X, Neupert W, Tzagoloff A. The metalloprotease encoded by ATP23 has a dual function in processing and assembly of subunit 6 of mitochondrial ATPase. Mol Biol Cell. 2006; 18:617–626. [PubMed: 17135290]

Zhang J, Wong J, Gao G, Luo H. Tripeptidyl peptidase II serves as an alternative to impaired proteasome to maintain viral growth in the host cells. FEBS Lett. 2010 [Epub ahead of print].

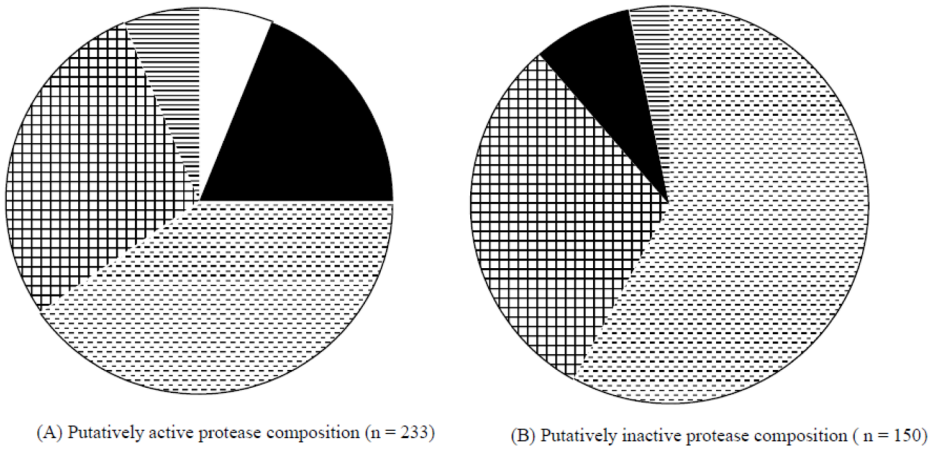

Table insert: Composition of *Ixodes scapularis* degradome

Protease class	Active (%)	Inactive (%)
Aspartic	6.0	0
Cysteine	19	8.1
Metallo	40	58.1
Serine	28.3	30.4
Threonine	6.4	3.4

Figure 1.

Composition of putatively active (A) and putatively inactive (B) proteases encoded by the *Ixodes scapularis* genome. Percent compositions of active and inactive proteases are noted in the table insert in figure 1.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

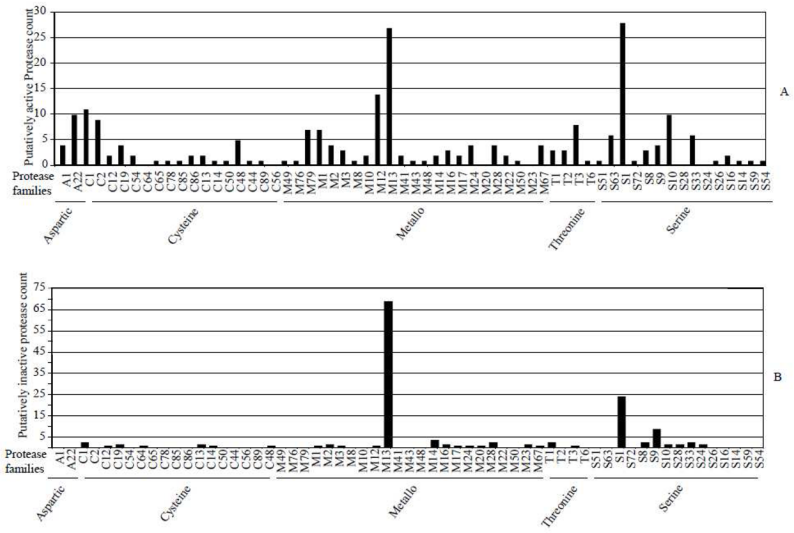

Figure 2.

Protease counts per family of proteases encoded by the *Ixodes scapularis* genome. After culling off redundant annotations, raw counts of putatively active (A) and putatively inactive (B) protease counts in each clan and family were determined.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Putatively active aspartic proteases encoded by the *I. scapularis* genome

Clan	Family	Archetype	Provisional ID	Gene ID
AA	A1	Pepsin A	Cathepsin D	ISCW003823
			Cathepsin D	ISCW023880
			-	ISCW007278
			-	ISCW013185
AD	A22	Presenilin 1	* Impas 1 peptidase	ISCW006186
			* Impas 2 peptidase	ISCW018713
			-	ISCW020583
			-	ISCW022390
			-	ISCW003378
			-	ISCW021952
			-	ISCW013270
			-	ISCW014541
			-	ISCW014542
			-	ISCW013270

* Impas = intramembrane protease aspartic protein,

- = Not provisionally identified

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Putatively active cysteine proteases encoded by the *I. scapularis* genome

Clan	Family	Archetype	Provisional ID	Gene ID
CA	C1	Papain	Cathepsin O	ISCW017550
			Cathepsin B	ISCW002679
				ISCW000078
				ISCW000080
				ISCW004002
			Insect 26/29 kDa peptidase	ISCW007008
			-	ISCW011264
			-	ISCW000076*
			-	ISCW005981*
			-	ISCW013346*
			-	ISCW024899*
	C2	Calcium activated neutral cysteine protease	-	+ ISCW012923
			-	+ ISCW009596
			-	+ ISCW002386
			-	+ ISCW006015
			-	+ ISCW017819
			-	+ ISCW024359
			-	+ ISCW004616
			-	+ ISCW017525
			-	+ ISCW012923
	C12	Ubiquitinyl hydrolase L1	Ubiquitinyl hydrolase-BAP1	ISCW022153
			Ubiquitinyl hydrolase-UCH37	ISCW000206
	C19	Ubiquitin-specific peptidase 14	Ubiquitin-specific peptidase 2	ISCW010054
			Ubiquitin-specific peptidase 14	ISCW019604
			-	ISCW003597
			-	ISCW022612
	C54	Autophagin 1	-	ISCW004712
			-	ISCW010488
	C65	Otubain 1	Otubain 1	ISCW023231
	C78	UfSP1 peptidase	UfSP1 peptidase	ISCW014202
	C85	DUBA	DUBA	ISCW021968
	C86	Ataxin 3	-	ISCW017541
			-	ISCW006944
CD	C13	Legumain	Legumain	ISCW015983
			Glycosylphosphatidylinositol	ISCW000202

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Clan	Family	Archetype	Provisional ID	Gene ID
	C14	Caspase 1	-	ISCW022545
	C50	Separase	Separase (*Homo sapiens*-type)	ISCW017453
CE	C48	[e]Ulp1 peptidase	[f]SENP1 peptidase	ISCW003012
				ISCW012734
				ISCW003013
			SENP8 peptidase	ISCW014736
			-	ISCW018648
PB	C44	Amidophospho-ribosyltransferase precursor	Amidophospho-ribosyltransferase precursor	ISCW015613
	C89	Acid ceramidase precursor	Acid ceramidase precursor	ISCW010877

[*] Originally classified as unclassified on version 9.3 MEROPS database (Rawlings et al., 2010).

- = Not provisionally identified

[+] Not listed on MEROPS database, Identified BLASTP scanning done in this study

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Putatively inactive cysteine proteases encoded by the *I. scapularis* genome

Clan	Family	Archetype	Provisional ID	Gene ID
CA	C1	Papain	Subfamily C1A non-peptidase homologues	ISCW009204
				ISCW011263
				ISCW024213
	C12	Ubiquitinyl hydrolase L1	Family C12 non-peptidase homologues	ISCW023394
	C19	Ubiquitin-specific peptidase 14	Family C19 non-peptidase homologues	ISCW008894
				ISCW009539
	C64	Cezanne deubiquitinylating peptidase	Family C64 non-peptidase homologues	ISCW014708
CD	C13	Legumain	Family C13 non-peptidase homologues	ISCW024575
				ISCW002213
	C14	Caspase 1	Subfamily C14A non-peptidase homologues	ISCW013172
CE	C48	Ulp1 peptidase	Family C48 non-peptidase homologues	ISCW016451
PC	C56	Pfp1 peptidase	Family C56 non-peptidase homologues	ISCW005013

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 4

Putatatively active metalloproteases encoded by the *I. scpularis* genome excluding family M13

Clan	Family	Archetype	Provisional ID	Gene ID
M-	M49	Dipeptidyl-peptidase III	Dipeptidyl peptidase III	ISCW013099
	M76	Atp23 peptidase	Atp23 peptidase	ISCW023892
	M79	Prenyl protease 2	-	ISCW007768
			-	ISCW024488
			-	ISCW011978
			-	ISCW012546
			-	ISCW015066
			-	ISCW015256
			-	ISCW003688
MA	M1	Aminopeptidase N	-	ISCW000925
			-	ISCW006685
			-	ISCW005385
			-	ISCW005386
			-	ISCW008629
			-	ISCW013459
			-	ISCW022724
	M2	Angiotensin-converting enzyme peptidase	-	ISCW016862
			-	ISCW005573
			-	ISCW010296
			-	ISCW003356
	M3	Thimet oligopeptidase	Mitochondrial intermediate peptidase	ISCW016804
			-	ISCW012570
			-	ISCW017437
	M8	Leishmanolysin	-	ISCW016195
	M10	Matrix metallopeptidase 1	Matrix metallopeptidase-2	ISCW022900
			Dm1 matrix metallopeptidase	ISCW006136
	M12	Astacin	Tolkin	ISCW016865
			ADAMTS adt-1 (*Caenorhabditis elegans*)-type peptidase	ISCW000910
			Metallopeptidase MTP-1	ISCW008310
			-	ISCW024332
			-	ISCW002794
			-	ISCW016405
			-	AF483730_1
			-	ISCW023638
			-	AF483731_1
			-	ISCW023637
			-	ISCW001423

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Clan	Family	Archetype	Provisional ID	Gene ID
			-	ISCW021991
			-	ISCW013359
			-	AAP22067.1|
	M41	FtsH peptidase	Paraplegin	ISCW014866
			-	ISCW009180
	M43	Cytophagalysin	-	ISCW002454
	M48	Ste24 peptidase	Farnesylated-protein converting enzyme 1	ISCW015561
MC	M14	Carboxypeptidase A1	-	ISCW015949
			-	ISCW023731
ME	M16	Pitrilysin	Mitochondrial processing peptidase beta-subunit	ISCW001592
			Nardilysin	ISCW011140
			YmxG peptidase	ISCW016724
MF	M17	Leucyl aminopeptidase	Leucyl aminopeptidase (animal)	ISCW019719
			Leucine aminopeptidase (*Fasciola*-type)	ISCW003100
MG	M24	Methionyl aminopeptidase 1	Methionyl aminopeptidase 1	ISCW022177
			Xaa-Pro dipeptidase (bacteria-type)	ISCW016014
				ISCW005124
			Mitochondrial intermediate cleaving peptidase 55 kDa	ISCW008932
MH	M28	AminopeptidaseS	Plasma glutamate carboxypeptidase	ISCW011412
			-	ISCW022417
			-	ISCW022420
			-	ISCW022423
MK	M22	O-sialoglycoprotein peptidase	Kae1 putative peptidase	ISCW006830
			OSGEPL1-like protein	ISCW023376
MM	M50	S2P protease	S2P peptidase	ISCW019725
MP	M67	Poh1 peptidase	Poh1 peptidase	ISCW017120
			Jab1/MPN domain metalloenzyme	ISCW019527
			-	ISCW007692
			-	ISCW008644

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 5

Putatively active family M13 metalloproteases encoded by the *I. scapularis* genome

Clan	Family	Archetype	Provisional ID	Gene ID
MA	M13	Neprilysin	-	ISCW003514
			-	ISCW018366
			-	ISCW004387
			-	ISCW018482
			-	ISCW024406
			-	ISCW006162
			-	ISCW006696
			-	ISCW007959
			-	ISCW006925
			-	ISCW020572
			-	ISCW008746
			-	ISCW009172
			-	ISCW010709
			-	ISCW022314
			-	ISCW022338
			-	ISCW012276
			-	ISCW013494
			-	ISCW011968
			-	ISCW024902
			-	ISCW014776
			-	ISCW014688
			-	ISCW013664
			-	ISCW023669
			-	ISCW024930
			-	ISCW024196
			-	ISCW002659
			-	ISCW004216

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 6

Putatively inactive metalloproteases encoded by the *I. scapularis* genome excluding members of family M13

Clan	Family	Archetype	Provisional ID	Gene ID
MA	M2	Angiotensin-converting enzyme peptidase unit 1	Family M2 non-peptidase homologues	ISCW014169
				ISCW016487
	M3	Thimet oligopeptidase	Subfamily M3A non-peptidase homologues	ISCW017206
	M12	Astacin	Subfamily M12A non-peptidase homologues	ISCW007161
MC	M14	Carboxypeptidase A1	Subfamily M14A non-peptidase homologues	ISCW007964
			Subfamily M14b non-peptidase homologues	ISCW004264
				ISCW019087
ME	M16	Pitrilysin	Subfamily M16c non-peptidase homologues	ISCW012694
MF	M17	Leucyl aminopeptidase	Family M17 non-peptidase homologues	ISCW023735
MG	M24	Methionyl aminopeptidase 1	Subfamily M24A non-peptidase homologues	ISCW019205
MH	M20	Glutamate carboxypeptidase	Subfamily M20D non-peptidase homologues	ISCW016472
	M28	Aminopeptidase Y	Subfamily M28A non-peptidase homologues	ISCW002672
			Subfamily M28b non-peptidase homologues	ISCW007773
				ISCW005069
			Family M28 non-peptidase homologues	ISCW022413
MO	M23	Beta-lytic metallopeptidase	Subfamily M23B non-peptidase homologues	ISCW000049
			Subfamily M23B non-peptidase homologues	ISCW003494
MP	M67	Poh1 peptidase	Subfamily M67A non-peptidase homologues	ISCW018927

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 7

Putatively inactive family M13 metalloproteases encoded by the *I. scapularis* genome

Clan	Family	Archetype	Provisional ID	Gene ID
MA	M13	Neprilysin	Family M13 inactive homolog	ISCW007711
				ISCW023236
				ISCW000249
				ISCW000073
				ISCW000558
				ISCW017380
				ISCW017096
				ISCW024191
				ISCW002982
				ISCW002041
				ISCW018087
				ISCW004642
				ISCW003638
				ISCW004533
				ISCW024243
				ISCW005180
				ISCW024417
				ISCW005350
				ISCW005864
				ISCW006336
				ISCW006308
				ISCW005859
				ISCW008290
				ISCW024419
				ISCW006979
				ISCW007149
				ISCW007020
				ISCW007468
				ISCW007705
				ISCW019277
				ISCW008136
				ISCW007447
				ISCW024562
				ISCW009160
				ISCW009889
				ISCW009040
				ISCW020266

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Clan	Family	Archetype	Provisional ID	Gene ID
				ISCW010099
				ISCW008656
				ISCW008658
				ISCW020857
				ISCW010145
				ISCW010006
				ISCW024592
				ISCW010295
				ISCW021722
				ISCW011462
				ISCW010591
				ISCW012489
				ISCW012068
				ISCW013095
				ISCW012849
				ISCW012747
				ISCW013945
				ISCW024871
				ISCW014215
				ISCW024865
				ISCW023305
				ISCW014706
				ISCW014756
				ISCW023963
				ISCW017354
				ISCW003353
				ISCW007468
				ISCW003121
				ISCW024865
				ISCW005798
				ISCW006308

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 8

Putatively active serine proteases encoded by the *I. scapularis* genome excluding family S1

Clan	Family	Archetype	Provisional ID	Gene ID
PC	S51	Dipeptidase E	Alpha-aspartyl dipeptidase (eukaryote)	ISCW005900
S-	S63	EGF-like module containing mucin-like hormone receptor-like 2	-	ISCW001355
			-	ISCW000464
			-	ISCW006717
			-	ISCW005937
			-	ISCW019673
			-	ISCW012721
	S72	Dystroglycan	Dystroglycan	ISCW015049
SB	S8	Subtilisin Carlsberg	Tripeptidyl-peptidase II	ISCW006099
				ISCW010275
				ISCW006048
SC	S9	Prolyl oligopeptidase	Prolyl oligopeptidase	ISCW002432
			Acylaminoacyl-peptidase	ISCW021356
			Oligopeptidase B	ISCW016734
			-	ISCW023976
	S10	Carboxypeptidase Y	Vitellogenic carboxypeptidase- like protein	ISCW003663
				ISCW018397
				ISCW007492
				ISCW010371
				ISCW014727
				ISCW003059
			-	AF483728_1
			-	ISCW012161
			-	ISCW022356
			-	ISCW022360
	S33	Prolyl aminopeptidase	-	ISCW002326
			-	ISCW002327
			-	ISCW017323
			-	ISCW004791
			-	ISCW019273
			-	ISCW021363
SF	S26	Signal peptidase I	Signal peptidase (invertebrate)	ISCW014839
SJ	S16	Lon A protease	Lon-A peptidase	ISCW007434
			Peroxisomal Lon peptidase	ISCW008144
SK	S14	Peptidase Clp	Peptidase Clp (type 1)	ISCW016708
SP	S59	Nucleoporin 145	Nucleoporin 145	ISCW013985
ST	S54	Rhomboid 1	-	ISCW021997

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 9

Putatively active family S1 serine proteases encoded by the *I. scapularis* genome

Clan	Family	Archetype	Provisional ID	Gene ID
PA	S1	chymotrypsin A	-	AF483729_1
			-	ISCW000320
			-	ISCW001310
			-	ISCW001619
			-	ISCW001322
			-	ISCW016684
			-	ISCW002695
			-	ISCW016985
			-	ISCW002489
			-	ISCW017115
			-	ISCW004276
			-	ISCW004835
			-	ISCW006069
			-	ISCW018539
			-	ISCW011174
			-	ISCW012149
			-	ISCW011961
			-	ISCW021893
			-	ISCW012429
			-	ISCW013112
			-	ISCW014005
			-	ISCW023349
			-	ISCW014551
			-	ISCW015186
			-	ISCW002979
			-	ISCW002999
			-	ISCW002674
			-	ISCW015172

- = Not provisionally identified

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 10

Putatively inactive serine proteases encoded by the *I. scapularis* genome

Clan	Family	Archetype	Provisional ID	Gene ID
PA	S1	Chymotrypsin A	Subfamily S1A non-peptidase homologues	ISCW015718
				ISCW015778
				ISCW001620
				ISCW002331
				ISCW017114
				ISCW003779
				ISCW004734
				ISCW006333
				ISCW005538
				ISCW006166
				ISCW006167
				ISCW006169
				ISCW007529
				ISCW021255
				ISCW011978
				ISCW012738
				ISCW014680
				ISCW016466
				ISCW002093
				ISCW000362
				ISCW000366
				ISCW002675
				ISCW012701
				ISCW004277
SB	S8	Subtilisin Carslberg	Subfamily S8A non-peptidase homologues	ISCW010275
				ISCW006048
				ISCW021721
SC	S9	Prolyl oligopeptidase	Subfamily S9 non-peptidase homologues	ISCW003960
				ISCW001875
				ISCW019824
				ISCW010323
				ISCW007846
				ISCW022036
				ISCW018728
				ISCW018730
				ISCW023102
	S10	Carboxypeptidase Y	Family S10 non-peptidase homologues	ISCW021184

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Clan	Family	Archetype	Provisional ID	Gene ID
				ISCW024936
	S28	Lysosomal Pro- Xaa carboxypeptidase	Family S28 non-peptidase homologues	ISCW003037
			Family S28 non-peptidase homologues	ISCW022743
	S33	Prolyl aminopeptidase	Family S33 non-peptidase homologues	ISCW005779
				ISCW002782
				ISCW012952
SF	S24	Repressor LexA	Family S24 non-peptidase homologues	ISCW011531
				ISCW013904

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 11

Putatively active threonine proteases encoded by the *I. scapularis* genome

Clan	Family	Archetype	Provisional ID	Gene ID
PB	T1	Archaean proteasome, beta component	Proteasome catalytic subunit 1	ISCW019233
			Proteasome catalytic subunit 2	ISCW006222
			Proteasome catalytic subunit	ISCW005987
	T2	Glycosyl-asparaginase precursor	Glycosyl-asparaginase precursor	ISCW001950
			Isoaspartyl dipeptidase (threonine type)	ISCW000866
			Taspase-1	ISCW001928
	T3	Gamma-glutamyl-transferase	Gamma-glutamyl-transferase 2 (bacterial)	ISCW000530
				ISCW003582
				ISCW009859
				ISCW010660
				ISCW012371
			-	ISCW005288
			-	ISCW012872
			-	ISCW013659
	T6	Polycystin 1	-	ISCW004388

- = Not provisionally identified

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 12

Putatively inactive threonine proteases by the *I. scapularis* genome

Clan	Family	Archetype	Provisional ID	Gene ID
PB	T1	Archaean proteasome, beta component	Subfamily T1A non-peptidase homologues	ISCW005463
				ISCW007139
				ISCW007217
				ISCW010978
	T3	Gamma-glutamyl-transferase family	Family T3 non-peptidase homologues	ISCW007979

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 13

Putatively active *Ixodes scapularis* proteases that are conserved in other tick species at ≥60% amino acid identity level

Clan	Family	*I. scapularis* protease accession #	Other tick species homolog accession #	Amino acid identity (%)
AA	A1	ISCW003823	*Ixodes ricinus* (ADU03674)	98
			Haemaphysalis longicornis (BAE53722)	72
			Rhipicephalus microplus (ACP21315)	71
		ISCW023880	*I. ricinus* (ADU03675)	99
			Dermancentor variabilis (ACF35520)	71
		ISCW013185	*I. ricinus* (ABO26561)	96
CA	C1	ISCW009204	*I. ricinus* (ABO26562)	71
			R. haemaphysaloides (AAQ16118)	67
			H. longicornis (BAH86062)	64
			D. variabilis (ABS70713)	64
			R. microplus (AAF61565)	60
			R. appendiculatus (AAO60046)	62
		ISCW002679	*I. ricinus* (ABO26563)	83
			D. variabilis (ACF35521)	71
			D. variabilis (ACF35525)	67
		ISCW000078	*I. ricinus* (ABO26563)	95
			D. variabilis (ACF35526)	69
			D. variabilis (ACF35521)	78
			Argas monolakensis (ABI52787)	66
		ISCW000080	*I. ricinus* (ABO26563)	93
			D. variabilis (ACF35526)	68
			D. variabilis (ACF35521)	77
			D. variabilis (ACF35525)	76
		ISCW004002	*H. longicornis* (BAF43801)	73
			D. variabilis (ACF35521)	63
		ISCW007008	*D. variabilis* (ACF35529)	83
			R. appendiculatus (AAO60045)	63
		ISCW000076	*I. ricinus* (ABO26563)	88
			D. variabilis (ACF35526)	73
			R. appendiculatus (AAO60044)	73
			D. variabilis (ACF35521)	78
		ISCW005981	*I. ricinus* (ABO26563)	74
			D. variabilis (ACF35521)	66
			D. variabilis (ACF35526)	63
			D. variabilis (ACF35525)	65
		ISCW013346	*I. ricinus* (ABO26563)	69

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Clan	Family	*I. scapularis* protease accession #	Other tick species homolog accession #	Amino acid identity (%)
			D. variabilis (ACF35521)	62
			D. variabilis (ACF35525)	62
			R. appendiculatus (AAO6004)	61
		ISCW024899	*I. ricinus* (ABV29335)	96
		ISCW003494	*I. ricinus* (ABV29335)	97
	C54	ISCW004712	*H. longicornis* (BAI82576)	76
CD	C13	ISCW015983	*I. ricinus* (AAS94231)	66
			H. longicornis (BAF51711)	63
			H. longicornis (BAF95090)	60
PB	C44	ISCW009866	*H. longicornis* (BAF46861)	84
MA	M1	ISCW006685	*D. variabilis* (ACQ90243)	68
	M12	ISCW023638	*I. ricinus* (CAO00625)	90
			I. ricinus (CAO00626)	85
			I. ricinus (CAB55817)	90
			I. pacificus (AAT9220)	90
		ISCW023637	*I. ricinus* (CAO00626)	65
			I. ricinus (CAO00625)	64
			I. pacificus (AAT92201)	61
		AAP22067	*I. ricinus* (CAO00626)	92
			I. ricinus (CAO00625)	84
			I. ricinus (CAB55817)	84
			I. pacificus (AAT92201)	80
MF	M17	ISCW019719	*H. longicornis* (BAF03566)	68
MP	M67	ISCW019527	*Amblyomma americanum* (ACG76242)	95
PA	S1	ISCW006167	*D. andersoni* (AAO12856)	90
			D. variabilis (AAN78224)	90
SC	S9	ISCW024395	*R. microplus* (AAF00496)	64