Published in final edited form as:

Insect Mol Biol. 2010 April ; 19(2): 229–241. doi:10.1111/j.1365-2583.2009.00978.x.

Analysis of *Rickettsia typhi*-infected and uninfected cat flea (*Ctenocephalides felis*) midgut cDNA libraries: deciphering molecular pathways involved in host response to *R. typhi* infection

S. M. Dreher-Lesnick^{*,¶}, S. M. Ceraul^{*,¶}, S. C. Lesnick[†], J. J. Gillespie^{*,‡}, J. M. Anderson[§], R. C. Jochim[§], J. G. Valenzuela[§], and A. F. Azad^{*}

^{*}Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD, USA

[†]Department of Biology, Bioscience Research Building, University of Maryland College Park, College Park, MD, USA

[‡]Virginia Bioinformatics Institute, Bioinformatics Facility, Washington Street, Virginia Tech, Blacksburg, VA, USA

[§]Vector Molecular Biology Unit, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institute of Health, Rockville, MD, USA

Abstract

Murine typhus is a flea-borne febrile illness that is caused by the obligate intracellular bacterium, *Rickettsia typhi*. The cat flea, *Ctenocephalides felis*, acquires *R. typhi* by imbibing a bloodmeal from a rickettsemic vertebrate host. To explore which transcripts are expressed in the midgut in response to challenge with *R. typhi*, cDNA libraries of *R. typhi*-infected and uninfected midguts of *C. felis* were constructed. In this study, we examined midgut transcript levels for select *C. felis* serine proteases, GTPases and defence response genes, all thought to be involved in the fleas response to feeding or infection. An increase in gene expression was observed for the serine protease inhibitors and vesicular trafficking proteins in response to feeding. In addition, *R. typhi* infection resulted in an increase in gene expression for the chymotrypsin and rab5 that we studied. Interestingly, *R. typhi* infection had little effect on expression of any of the defence response genes that we studied. We are unsure as to the physiological significance of these gene expression profiles and are currently investigating their potential roles as it pertains to *R. typhi* infection. To our knowledge, this is the first report of differential expression of flea transcripts in response to infection with *R. typhi*.

Keywords

flea; rickettsia; midgut; *R. typhi*

© 2009 The Authors Journal compilation © 2009 The Royal Entomological Society

Correspondence: Shane M. Ceraul, Department of Microbiology and Immunology, University of Maryland School of Medicine, 685 W. Baltimore Street, Room HSFI Suite 380, Baltimore, MD 21201, USA. Tel.: +410 706 3337; fax: +410 706 0282; scera001@umaryland.edu.

[¶]Sheila M. Dreher-Lesnick and Shane M. Ceraul contributed equally.

Supporting Information

Additional Supporting Information may be found in the online version of this article under the DOI reference: DOI 10.1111/j.1365-2583.2009.00978.x

Introduction

Fleas are one of many insect vectors that transmit pathogens important to human health, including bubonic plague, cat scratch disease and flea-borne rickettsioses (Fronzes *et al.*, 2009). Fleas and flea-borne diseases continue to pose a public health threat today, with reemergence of bubonic plague and rickettsioses worldwide (Kelly *et al.*, 2005; Parola *et al.*, 2005; Eisen & Gage, 2009). Despite their continued role as vectors of disease, little is known about flea molecular responses to infection and the biology underlying their vector competency.

Murine typhus is a flea-borne febrile disease that is caused by the obligate intracellular bacterium, *Rickettsia typhi*. *R. typhi* does not adversely affect its flea host, and is maintained in a natural cycle that involves small rodents and fleas. An increase in reports of human cases in recent years is likely due to increased physician awareness, improved diagnoses, and changes in environment and human behaviour (Azad *et al.*, 1997).

Upon acquisition, *R. typhi* persists in fleas at which point they may serve as reservoirs of rickettsiae in nature (Azad *et al.*, 1997). Once a flea ingests an infected blood meal, *R. typhi* encounter the flea midgut. Many studies suggest that molecular interactions with the arthropod midgut may determine the fate of imbibed microbes (Dong *et al.*, 2006; Hu & Aksoy, 2006). For instance, knock-down of the transcriptional activator Rel in the tsetse fly, *Glossina morsitans morsitans*, results in an increase in the prevalence of *Trypanosoma brucei* in the midgut (Hu & Aksoy, 2006). Similarly, silencing of anti-Plasmodium genes in *Anopheles gambiae* resulted in an increased presence of ookinetes in the mosquito midgut (Dong *et al.*, 2006). To date, it is unclear if rickettsial infection alters flea midgut gene expression. In general, not much is known about the flea's response to microbial challenge.

For arthropod vectors, comparison of cDNA libraries can provide insights into their transcriptional response to infection with microbes. For example, comparison of a cDNA library derived from *Leishmania* spp.-infected sand flies with a cDNA library from uninfected sand flies identified several genes that may play a role in the arthropods ability to serve as a vector of disease (Ramalho-Ortigao *et al.*, 2007; Jochim *et al.*, 2008). Differences in transcript abundance were observed in genes involved in defence and antimicrobial response, as well as genes in several other functional categories (Ramalho-Ortigao *et al.*, 2007; Jochim *et al.*, 2008).

Previously, we reported on temperature-induced gene expression trends in *R. typhi* to highlight how the microbe adjusts to individual host environmental temperatures (Dreher-Lesnicks *et al.*, 2008). In the current report, we study how the vector deals with *R. typhi* infection by constructing and comparing cDNA libraries from *R. typhi*-infected and uninfected *Ctenocephalides felis* midguts. We hypothesize that *C. felis* transcript patterns in the midgut will differ between the infected and uninfected fleas, reflecting the midgut response to *R. typhi* challenge. Transcription patterns of select *C. felis* proteases, putative GTPases and defence genes were further analysed over a four-day infection time-course *in vivo* to elucidate their potential role in the flea response to *R. typhi* infection.

Results and discussion

To generate a repertoire of flea midgut transcripts that may be involved in the flea response to infection with *R. typhi*, cDNA libraries from *R. typhi*-infected (CFGF-I) and uninfected (CFGF) *C. felis* midguts were constructed. A total of 1152 transcripts from both libraries were sequenced, generating 906 high quality sequences, 472 from the uninfected and 434 from the infected midgut library. Sequences were clustered into groups based on similarity, which resulted in a total of 419 contigs derived from two or more sequences, and 334

singletons derived from only one sequence. Transcripts were assigned putative functions based on comparisons with the NR protein NCBI database, gene ontology (GO) database, as well as the NCBI conserved domain databases (CDD), which include Eukaryotic Orthologous Groups (KOG), Protein Families database (PFAM) and Simple Modular Architecture Research Tool (SMART) (Altschul *et al.*, 1997; Ashburner *et al.*, 2000; Marchler-Bauer *et al.*, 2002). The results of these analyses are summarized in Table S1. Assignment of functional class was based on matches with the lowest *E*-value from the KOG and GO databases. Flea transcripts that matched genes annotated as uncharacterized conserved function were assigned to the uncharacterized conserved function KOG. A summary of the number of transcripts per functional class is listed in Fig. 1. Functional classes are ordered by abundance of transcript, with the functional class containing the most transcripts at the top. Of the transcripts that showed similarity to known proteins, the most abundant were found in both libraries and correspond to genes involved in energy production and conversion and amino acid transport and metabolism. It is probable that these genes are involved in blood meal digestion and subsequent energy conversion. Among these, it is interesting to note that more transcripts with putative roles in energy production and conversion were identified from the library derived from uninfected fleas than *R. typhi*-infected fleas. Differences in transcript abundance between the two libraries are also found in defence mechanism, transcription, signal transduction mechanism, intracellular trafficking and secretion, and cytoskeleton functional classes (Fig. 1). These differences in trends between the two libraries may elucidate processes important in the flea's response to rickettsial infection.

Proteases are integral to bloodmeal acquisition, digestion and possibly the response to *Rickettsia typhi* infection

Closer examination of transcripts within the amino acid transport and metabolism functional group revealed a large number of transcripts encoding proteases, primarily trypsins and chymotrypsins. Analysis of both libraries identified a total of 76 transcripts in 18 contigs with similarity to trypsin and trypsin-like proteins (Table 1), and 108 transcripts in 22 contigs with similarity to chymotrypsin and chymotrypsin-like proteins (Table 2). A previous study by Gaines *et al.* (1999) identified 19 *C. felis* serine proteases and examined transcript abundance of select proteases at different life stages, in different tissues and genders, and in response to blood feeding. Of the 18 trypsin transcripts identified in this library screen, 13 are seemingly novel flea trypsin sequences (Table 1). Interestingly, 4 of the 13 novel sequences were identified from the infected library only (Contigs 92, 76, 265 and 172). Twenty-two flea chymotrypsin-like transcripts were identified from both infected and uninfected midgut libraries. Of the 22 flea chymotrypsin-like sequences identified in our libraries, 15 appear to be novel sequences (Table 2). Multiple sequence alignment of select trypsins and chymotrypsin transcripts show a moderate degree of similarity outside of conserved regions, which is consistent with a previous report of *C. felis* trypsins (Gaines *et al.*, 1999) and trypsins from invertebrates in general (Muhlia-Almazan *et al.*, 2008). Trypsins and chymotrypsins belong to the S1 family of serine proteases. All members of the S1 family of serine proteases share the conserved catalytic triad (His-Asp-Ser) and six cysteine residues that form three disulfide bonds (Fig. 2) (Muhlia-Almazan *et al.*, 2008). This alignment also shows the conserved arginine at the proteolytic cleavage site preceding the 'IVGG' motif of mature proteins (Muhlia-Almazan *et al.*, 2008).

A high abundance of serine protease transcripts have also been identified in other cDNA libraries from studies exploring the effects of blood feeding and/or infection on arthropod gene transcription. Multiple trypsins have been identified from midguts of sand flies, where differences in transcript patterns of trypsins and chymotrypsins are observed between unfed, sugar- or blood-fed sand flies (Ramalho-Ortigao *et al.*, 2003, 2007; Jochim *et al.*, 2008).

Similar differences in transcription patterns for trypsins (Table 1) and chymotrypsins (Table 2) were also found in our study, with some showing an increase and others a decrease in transcription in response to infection. Arthropod serine proteases are known to be involved in multiple processes associated with digestion, development and immune response (Gorman & Paskewitz, 2001; Muhlia-Almazan *et al.*, 2008). In the arthropod immune response, serine proteases regulate haemolymph coagulation, antimicrobial peptide synthesis and melanization of pathogen surfaces (Gorman & Paskewitz, 2001; Xu *et al.*, 2006).

In this study, a closer look at transcript number per contig for both trypsins and chymotrypsins revealed some differences between the libraries derived from infected and uninfected fleas. To further determine whether proteases play a role in the flea response to *R. typhi*, expression patterns of two serine protease transcripts, contig 15/chymotrypsin 15 (NR protein database accession number AAD21823.1) and contig 16/trypsin 16 (NR protein database accession number AAD21829.1) were examined over an infection time course (Fig. 3). Both contigs had a higher abundance of transcripts in the uninfected library compared with the infected library (Table 1 and 2).

Expression of trypsin-16 over a four-day infection time course was measured by real-time quantitative reverse-transcription (qRT)-PCR (Fig. 3A). A two-way ANOVA examining the effect of time post feeding and treatment (infected vs. uninfected blood meal) shows a significant increase in gene expression over the four-day infection time course ($P < 0.0001$; Table S2; Fig. 3A). While expression in both treatments increases over time post feeding, gene expression of trypsin-16 is significantly lower in infected than uninfected fleas throughout the experiment ($P = 0.0011$; Table S2; Fig. 3A). In summary, real time qRT-PCR analysis of trypsin-16 indicates down-regulation in *R. typhi*-infected *C. felis* midguts relative to the increase in expression attributable to feeding alone (Table S2).

Relative gene expression of chymotrypsin-15 was also measured over the four-day time course (Fig. 3B). Expression of chymotrypsin-15 increased over the four-day time course after the blood meal ($P < 0.0001$; Table S2). Furthermore, gene expression of chymotrypsin-15 was significantly higher in infected fleas than uninfected fleas ($P = 0.027$; Table S2) and there was no significant interaction between time and treatment.

Changes in trypsin expression in response to immune challenge by pathogens are also observed in other insect vectors, including sand flies (Ramalho-Ortigao *et al.*, 2003) and mosquitoes (Jahan *et al.*, 1999; Xu *et al.*, 2006). In sand flies, Ramalho-Ortigao *et al.*, 2003) observed a significant decrease in expression of the chymotrypsin PpChym2, and a significant increase of the trypsin PpTryp1 in *Leishmania major*-infected midguts. Suppression of *Aedes aegypti* midgut serine proteases decreased infectivity of Dengue virus-2 in the midgut, and decreased dissemination of the virus to the thorax and head (Molina-Cruz *et al.*, 2005; Brackney *et al.*, 2008). How differential transcription of individual *C. felis* serine proteases may be involved in either limiting or enabling infection of *R. typhi* in the midgut will be explored in future studies.

Transcriptional profiles for genes involved in intracellular trafficking and secretion may reflect rickettsial modulation or host response to infection

Upon entry into cells, rickettsiae quickly escape from the phagosome and replicate in the host cytosol (Hackstadt, 1996). To date, only a few details about the molecular processes involved in rickettsial escape from the phagosome are known. Following escape, *R. typhi* then replicates within the host cell cytosol, leading to the rupture of host cell membrane and the spread of *R. typhi* to neighbouring cells (Hackstadt, 1996). Due to its obligate intracellular lifestyle, and the need to avoid detection from host cell defences, it is likely that *R. typhi* modulates host cellular processes. Modulation of host vesicular trafficking is

documented for many intracellular bacteria, such as *Legionella pneumoniae* modulation of GTPases (Brumell & Scidmore, 2007). Analysis of cDNA libraries from both infected and uninfected fleas identified multiple transcripts that encode proteins involved in intracellular trafficking and secretion (Table 3).

To determine if *R. typhi* infection affects transcription of *C. felis* genes involved in intracellular trafficking, we examined transcription of two GTPases identified in our cDNA screen of the flea midgut over a four-day time course as described above. One GTPase is a putative *C. felis* signal recognition particle (Cf-srp), and the second, a putative rab5 GTPase (Cf-rab5) (Fig. 4). Examination of the results of the four-day time course experiment suggests expression of Cf-srp increases in uninfected flea midgut post-blood meal, while expression in *R. typhi*-infected flea midgut does not (Fig. 4A). Consistent with this observation, statistical analysis shows a significant interaction between time and treatment on Cf-srp gene expression ($P = 0.024$; Table S2) as well as significant time ($P = 0.0063$) and treatment effects ($P = 0.0012$; Table S2). This pattern of down-regulation is particularly intriguing as SRPs direct nascent secretory or membrane proteins to the protein translocation machinery of cells, the endoplasmic reticulum in eukaryotes, for subsequent protein sorting (Keenan *et al.*, 2001). How and whether *R. typhi* modulates host protein sorting is yet to be determined.

Another GTPase identified in the *C. felis* cDNA library is a rab5-like GTPase. Rab5 is recruited to early endosomes in eukaryotic cells and is involved in phagosome maturation (Brumell & Scidmore, 2007). In the four-day time course experiment, expression of Cf-rab5 increased after blood meal ($P < 0.0001$; Table S2). Furthermore, transcription of rab5 is significantly higher in infected than uninfected *C. felis* midguts ($P = 0.0096$) and there is no evidence of significant interaction between time and treatment ($P = 0.9981$; Table S2). These results suggest that infection with *R. typhi* increases expression of rab5-like GTPase during four days after infection above expression caused by feeding alone.

Manipulation of Rab GTPases by intracellular bacteria is well documented in multiple organisms such as *Mycobacterium tuberculosis*, *Coxiella burnetii*, *Chlamydia* spp., *Legionella pneumophila* and *Listeria monocytogenes* (Brumell & Scidmore, 2007). Many intracellular bacteria replicate inside membrane-bound organelles within the host, thus, modulation of host Rab5 to either delay or avoid phagosome maturation is a crucial step in early stages of infection (Brumell & Scidmore, 2007). Unlike the aforementioned intracellular bacteria, *Listeria* and *Rickettsia* spp. escape from the phagosome soon after cell entry and replicate within the host cytosol (Whitworth *et al.*, 2005; Pizarro-Cerda & Cossart, 2006). Recent evidence suggests that *L. monocytogenes* modulate Rab's early in the infectious cycle, and exclude Rab5 in early endosomes (Henry *et al.*, 2006). Absence of Rab5 may be beneficial to *L. monocytogenes* survival within the macrophage (Henry *et al.*, 2006), as over-expression of Rab5 enhanced intracellular killing of *L. monocytogenes* (Prada-Delgado *et al.*, 2005). In contrast, little is known about the interaction between rickettsiae and the vesicular trafficking system. More specifically, because Rab5 is found in the early endosome, a compartment that rickettsiae spend little time in, we decided to assess the expression profile in response to infection. In this study, overall expression of *C. felis* rab5 increased in response to infection with *R. typhi* (Fig. 4B). We are unsure at this time if *R. typhi* is exploiting Rab5 or if the increase in *rab5* expression stabilizes the early endosome, thereby preventing escape by *R. typhi*. The latter would represent a host mechanism to limit *R. typhi* infection and needs to be further explored. Regardless of its physiological significance, the increase in Rab5 expression as a result of infection and its localization in the early endosome make this GTPase a protein of interest.

Feeding-induced expression of defence response genes may prepare the midgut environment for microbial challenge

For blood-feeding arthropods, the initial interaction between a microbe and the insect occurs in the midgut. To this end, a heightened level of immunological activity in response to feeding prepares the midgut for any microbe that is imbibed with a blood meal. To combat infiltrating microbes, soluble effectors or by-products of digestion are secreted into the lumen of the midgut (Fogaca *et al.*, 1999; Nakajima *et al.*, 2003; Sonenshine *et al.*, 2005). Soluble effectors may include opsinizing lectins, lytic cationic peptides such as defensin and lysozyme and Serine Protease Inhibitors that regulate the melanization cascade (Gadjeva *et al.*, 2001; Philpott *et al.*, 2001; Abraham *et al.*, 2005). Our screen identified flea genes with innate immune functions, including a flea defensin and previously published *C. felis* serpins (Brandt *et al.*, 2004) (Table 4). These soluble effectors will be most effective before *R. typhi* enters host cells and during spread from one host cell to the next. To this end, we examined the effect of *R. typhi* infection on gene expression patterns for defensin, a lectin, a SERPIN as well as a peptidoglycan receptor (PGRP-LB), which is required for immune deficiency pathway signalling and homeostasis (Girardin & Philpott, 2006; Zaidman-Remy *et al.*, 2006; Wang *et al.*, 2009).

We examined expression of these genes in infected and uninfected fleas in a four-day time course as described above. Analysis of these experiments suggest that transcript abundance for defensin, lectin and SERPIN increase in response to feeding (defensin, $P < 0.0001$; lectin, $P < 0.0001$ and $P < 0.001$; serpin, $P < 0.018$, Table S2), which is consistent with observations for defensin gene expression in the stable fly, *Stomoxys calcitrans*, and the sheep tick, *Ixodes ricinus* (Lehane *et al.*, 1997; Munks *et al.*, 2001; Boulanger *et al.*, 2002; Hamilton *et al.*, 2002; Rudenko *et al.*, 2005) (Table S2). Interestingly, none of these genes show a significant treatment effect on gene expression due to infection (Table S2).

Rickettsia spp. appears to vary in their ability to elicit antimicrobial gene expression. Previously, we reported an increase for defensin and lysozyme gene expression, over time, in the tick midgut in response to a challenge with the non-pathogenic rickettsia, *Rickettsia montanensis* (Ceraul *et al.*, 2007). Most recently, *R. montanensis* challenge of the American dog tick, *Dermacentor variabilis*, resulted in an increase in gene expression for a kunitz-type serine protease inhibitor that was also shown to possess rickettsiostatic properties (Ceraul *et al.*, 2008). On the other hand, *Rickettsia peacockii* failed to elicit or suppress antimicrobial gene expression in an *I. scapularis* and *Dermacentor andersoni* cell lines (Mattila *et al.*, 2007). In contrast to the observations for *R. peacockii*, infection of THP-1 cells (acute monocytic leukemia cell line) with *Anaplasma phagocytophilum*, an Alphaproteobacterium in the order Rickettsiales, results in a down-regulation of host defence response genes (Garcia-Garcia *et al.*, 2009).

The variability in immune stimulation of arthropods by *Rickettsia* spp. may be due to the biological systems in which the phenomena are studied. Although, gene expression for trypsin-16 and rab5 in our study increased as a result of challenge with *R. typhi* (Fig. 3 and 4, Table S2), it is possible that the rickettsial dose used in our experiments was too low to elicit expression of the standard immune response genes beyond that observed for blood-feeding. We also realize that transcription patterns are not definitive due to the complexities of protein expression. To address these concerns, we are continuing our study of immune stimulation in fleas at the levels of signalling and hope to provide more definitive conclusions in the future.

Experimental procedures

Rickettsia typhi growth and isolation for flea infections

Rickettsia typhi were grown in L929 mouse fibroblast (MOI 10) and Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 5% FBS in 75 cm² or 150 cm² filter top flasks. The infected cells were maintained at 37 °C for 4 days before harvest of the rickettsiae. To harvest rickettsiae, the infected L929 cells were scraped and resuspended in 3 ml of fresh media. The lysates were then sonicated at setting 6.5 twice for 30 s using a Sonic Dismembrator (Fischer Scientific, Pittsburg, PA, USA). To remove large host cell material, lysates were centrifuged at 1000× g for 5 min. The supernatants (containing rickettsiae) were then placed over a 25% renografin/SPG (218 mM sucrose, 3.76 mM K₂HPO₄, 7.1 mM K₂HPO₄, 4.9 mM Potassium glutamate) bed and centrifuged at 14 000× g for 10 min. *R. typhi* were routinely quantified using the BacLight Live/dead assay (Molecular probes, Eugene, OR, USA). Bacterial pellets were either frozen at −80 °C or immediately used for the flea infections.

Flea infection

Ctenocephalides felis fleas were purchased from HESKA Corporation (Loveland, CO, USA) and maintained in our BSL-3 insectary. An artificial feeder (artificial dog) was used to infect fleas with *R. typhi*. Briefly, samples of 50–100 fleas were placed into special membrane-feeding capsules and provided either 4 ml of infected or uninfected (negative control) whole sheep's blood. For the infection, we added 1×10^6 renografin purified *R. typhi* to the blood. All blood reservoirs were maintained at 37 °C within the flea-feeding apparatus. Blood was replenished on day 3 for all feeds. For all flea infections, we verified absence/presence of *R. typhi* from RNA isolated from control and infected flea batches (see below) with primers directed at the 17 kDa gene using standard RT-PCR protocols. Midguts used to construct the cDNA libraries were dissected from fleas that were fed for five days. Fleas were immobilized and dissected in RNase-free 1XPBS. We dissected midguts from 20 fully engorged fleas from both the infected and control groups, rinsed the midguts in three additional changes of RNase-free 1XPBS to remove host blood, and placed them in 20 µl of RNeasy lysis buffer (Qiagen, Valencia, CA, USA). Samples were stored at 4 °C overnight and subsequently placed at −80 °C until use for cDNA library construction (outlined below). For our flea infection time course, we harvested one capsule each of infected and uninfected fleas at 6 h, 18 h, 24 h, 48 h, 72 h and 96 h post infection. Flea midguts were dissected as described above, but instead of being placed in RNeasy lysis buffer, were directly placed in RLT buffer from the ALLPREP DNA/RNA isolation kit by Qiagen (Valencia, CA, USA). Midguts were homogenized by passing them five times through a 1-ml syringe fitted with a 27-gauge needle. Total RNA was isolated per manufacturer's protocols.

cDNA library construction

Both infected and uninfected *C. felis* midgut libraries were constructed concurrently. *C. felis* midgut mRNA was isolated using the Micro-fast Track mRNA isolation kit (Invitrogen, Carlsbad, CA, USA). Both cDNA libraries were constructed using the SMART cDNA library construction kit (Clontech, Mountain View, CA, USA) with minor modifications as previously described (Anderson *et al.*, 2008; Valenzuela *et al.*, 2001a,b). Sequencing of both libraries was done as described previously (Anderson *et al.*, 2008; Jochim *et al.*, 2008; Valenzuela *et al.*, 2001a,b).

Bioinformatic analysis

Data analysis followed standard protocols outlined previously (Anderson *et al.*, 2008; Jochim *et al.*, 2008; Valenzuela *et al.*, 2001a,b). The analysis output is found in Table S1.

Select transcript sequences from the cDNA library were aligned to reference sequences with MUSCLE (Edgar, 2004a, 2004b) using default parameters. Alignments were color-coded and annotated using Adobe Illustrator®CS2.

Endpoint reverse-transcription PCR and quantitative real-time reverse-transcription-PCR

Primers were designed using MacVector (Table 5) and the guidelines set forth in Stratagene's Guide to Quantitative PCR (QPCR Guide SMO90106). Flea midgut RNA was isolated as described above. After DNase treatment, RNA was checked for genomic DNA (gDNA) contamination by PCR using standard conditions. To detect the presence or absence of *R. typhi* in fleas, we performed a one-step RT-PCR reaction with *R. typhi*-specific 17 kDa primers and 1 µl *C. felis* RNA following manufacturer's protocols (Qiagen). For real time analysis, cDNA was synthesized with random hexamers from gDNA-free RNA using Superscript III First Strand Synthesis Supermix Kit for qRT-PCR (Invitrogen) according to the manufacturer. qPCR reactions were set up according to manufacturer's protocol for the SYBRGreener Supermix for qPCR (Invitrogen) using *C. felis* cDNA from our time point samples as template. Reactions were run on a Stratagene Mx3000p thermocycler (Stratagene, La Jolla, CA, USA). PCR efficiencies for each primer pair were determined using LinReg PCR (Ramakers *et al.*, 2003), and were then incorporated into the normalization calculations using the Qgene software program (Muller *et al.*, 2002). A *C. felis* actin gene was used as the endogenous housekeeping gene for normalization.

Gene expression analysis

We analysed gene expression in infected and uninfected fleas over a four-day time course using JMP8 statistical software (SAS, Cary, NC, USA). Relative expression data for all genes were not normally distributed (Shapiro-Wilk, $P < 0.05$) and were therefore log transformed prior to analysis. We used a two-way ANOVA to examine the effect that feeding and treatment (infected vs. uninfected blood meal), as a function of time, has on gene expression. We also examined the interaction of feeding and treatment and how it affects gene expression. Results of statistical analyses are presented in Table S2.

GenBank submission

All sequences presented in Table S1 were submitted as expressed sequence tags to GenBank. All trypsin, chymotrypsin, vesicular trafficking protein and defence response sequences were submitted as truncated or full-length sequences to GenBank. Accession numbers appear in the appropriate table.

Conclusion

Unlike most vector-borne pathogens, rickettsiae have the unique ability to successfully infect the host cell and persist in their arthropod. While *R. typhi* can infect and persist in multiple flea species (Farhang-Azad *et al.*, 1983, 1984), transmission via the common cat flea, *C. felis*, has contributed to the re-emergence of murine typhus in the USA, and perhaps worldwide (Azad *et al.*, 1997; Kelly *et al.*, 2005; Parola *et al.*, 2005). Since rickettsiae are acquired by the flea through an infected blood meal, the flea midgut and midgut-associated proteins are the first hurdle *R. typhi* encounter. This initial encounter likely plays an important role in early stages of rickettsial infection and subsequent persistence within the flea host.

Understanding what mechanisms contribute to balancing the flea response to *R. typhi* challenge to establish a persistent infection remain largely unknown, in part due to the lack of genomic information of flea midgut genes. Larger scale cDNA library construction and subsequent high throughput sequencing are powerful tools to generate a repertoire of flea

genes for further characterization, and provide insight into molecular responses of the flea midgut to invading pathogens.

In this study, we have identified 472 *C. felis* midgut transcripts from both an uninfected and *R. typhi*-infected *C. felis* cDNA library. Bioinformatic analysis of both libraries has identified an abundance of flea genes with function in energy transport and conversion, as well as amino acid transport and metabolism. Most notably, a large number of serine proteases were discovered, in addition to those already published for *C. felis*. Transcripts involved in intracellular trafficking and secretion, as well as host defence, are also highlighted.

Follow up studies examining differences in transcription of select genes were performed using standard real time RT-PCR methods. Here, transcription patterns of two serine proteases and two GTPases were monitored over time in response to infection with *R. typhi*. Infection with *R. typhi* significantly altered expression of all four flea genes examined. It is not yet clear how or why these specific genes change in expression. Additionally, we found that *R. typhi* infection had little to no effect on the transcription patterns of the selected defence response genes. Further targeted characterization of these flea genes to determine their role during the response to challenge with *R. typhi* will help our understanding of the complex molecular interactions between vectors and the pathogens they harbour.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Jose Ribero for providing the bioinformatics analyses software and for review of the manuscript. AFA acknowledges support from NIAID grants, AI017828, AI59118 and AI043006. JJG acknowledges support from NIAID contract HHSN266200400035C awarded to Bruno W. Sobral (Virginia Bioinformatics Institute at Virginia Tech). Because RCJ, JMA and JGV are government employees and part of this is a government work, the work is in the public domain in the United States. This work was supported in part by the Intramural Research Program of the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIAID or the National Institutes of Health.

References

- Abraham EG, Pinto SB, Ghosh A, Vanlandingham DL, Budd A, Higgs S, et al. An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites. *Proc Natl Acad Sci U S A*. 2005; 102:16327–16332. [PubMed: 16260729]
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res*. 1997; 25:3389–3402. [PubMed: 9254694]
- Anderson JM, Sonenshine DE, Valenzuela JG. Exploring the mialome of ticks: an annotated catalogue of midgut transcripts from the hard tick, *Dermacentor variabilis* (Acari: Ixodidae). *BMC Genomics*. 2008; 9:552. [PubMed: 19021911]
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat Genet*. 2000; 25:25–29. [PubMed: 10802651]
- Azad AF, Radulovic S, Higgins JA, Noden BH, Troyer JM. Flea-borne rickettsioses: ecologic considerations. *Emerg Infect Dis*. 1997; 3:319–327. [PubMed: 9284376]
- Boulanger N, Munks RJ, Hamilton JV, Vovelle F, Brun R, Lehane MJ, et al. Epithelial innate immunity. A novel antimicrobial peptide with antiparasitic activity in the bloodsucking insect *Stomoxys calcitrans*. *J Biol Chem*. 2002; 277:49921–49926. [PubMed: 12372834]

- Brackney DE, Foy BD, Olson KE. The effects of midgut serine proteases on dengue virus type 2 infectivity of *Aedes aegypti*. *Am J Trop Med Hyg.* 2008; 79:267–274. [PubMed: 18689635]
- Brandt KS, Silver GM, Becher AM, Gaines PJ, Maddux JD, Jarvis EE, et al. Isolation, characterization, and recombinant expression of multiple serpins from the cat flea, *Ctenocephalides felis*. *Arch Insect Biochem Physiol.* 2004; 55:200–214. [PubMed: 15027074]
- Brumell JH, Scidmore MA. Manipulation of rab GTPase function by intracellular bacterial pathogens. *Microbiol Mol Biol Rev.* 2007; 71:636–652. [PubMed: 18063721]
- Ceraul SM, Dreher-Lesnicks SM, Gillespie JJ, Rahman MS, Azad AF. A New Tick Defensin Isoform and Antimicrobial Gene Expression in Response to *Rickettsia montanensis* Challenge. *Infect Immun.* 2007; 75:1973–1983. [PubMed: 17261604]
- Ceraul SM, Dreher-Lesnicks SM, Mulenga A, Rahman MS, Azad AF. Functional characterization and novel rickettsiostatic effects of a Kunitz-type serine protease inhibitor from the tick *Dermacentor variabilis*. *Infect Immun.* 2008; 76:5429–5435. [PubMed: 18779339]
- Dong Y, Aguilar R, Xi Z, Warr E, Mongin E, Dimopoulos G. *Anopheles gambiae* immune responses to human and rodent Plasmodium parasite species. *Plos Pathog.* 2006; 2:52.
- Dreher-Lesnicks SM, Ceraul SM, Rahman MS, Azad AF. Genome-wide screen for temperature-regulated genes of the obligate intracellular bacterium, *Rickettsia typhi*. *BMC Microbiol.* 2008; 8:61. [PubMed: 18412961]
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acids Res.* 2004a; 32:1792–1797. [PubMed: 15034147]
- Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 2004b; 5:113. [PubMed: 15318951]
- Eisen RJ, Gage KL. Adaptive strategies of *Yersinia pestis* to persist during inter-epizootic and epizootic periods. *Vet Res.* 2009; 40:1. [PubMed: 18803931]
- Farhang-Azad A, Traub R, Wisseman CL Jr. *Rickettsia mooseri* infection in the fleas *Leptopsylla segnis* and *Xenopsylla cheopis*. *Am J Trop Med Hyg.* 1983; 32:1392–1400. [PubMed: 6418017]
- Farhang-Azad A, Traub R, Sofi M, Wisseman CL Jr. Experimental murine typhus infection in the cat flea, *Ctenocephalides felis* (Siphonaptera: Pulicidae). *J Med Entomol.* 1984; 21:675–680. [PubMed: 6502625]
- Fogaca AC, da Silva PI Jr, Miranda MT, Bianchi AG, Miranda A, Ribolla PE, et al. Antimicrobial activity of a bovine hemoglobin fragment in the tick *Boophilus microplus*. *J Biol Chem.* 1999; 274:25330–25334. [PubMed: 10464258]
- Fronzes R, Schafer E, Wang L, Saibil HR, Orlova EV, Waksman G. Structure of a type IV secretion system core complex. *Science.* 2009; 323:266–268. [PubMed: 19131631]
- Gadjeva M, Thiel S, Jensenius JC. The mannan-binding-lectin pathway of the innate immune response. *Curr Opin Immunol.* 2001; 13:74–78. [PubMed: 11154921]
- Gaines PJ, Sampson CM, Rushlow KE, Stiegler GL. Cloning of a family of serine protease genes from the cat flea *Ctenocephalides felis*. *Insect Mol Biol.* 1999; 8:11–22. [PubMed: 9927170]
- Garcia-Garcia JC, Barat NC, Trembley SJ, Dumler JS. Epigenetic silencing of host cell defense genes enhances intracellular survival of the rickettsial pathogen *Anaplasma phagocytophilum*. *Plos Pathog.* 2009; 5:e1000488. [PubMed: 19543390]
- Girardin SE, Philpott DJ. PGRP-LB minds the fort. *Immunity.* 2006; 24:363–366. [PubMed: 16618594]
- Gorman MJ, Paskewitz SM. Serine proteases as mediators of mosquito immune responses. *Insect Biochem Mol Biol.* 2001; 31:257–262. [PubMed: 11167095]
- Hackstadt T. The biology of rickettsiae. *Infect Agents Dis.* 1996; 5:127–143. [PubMed: 8805076]
- Hamilton JV, Munks RJ, Lehane SM, Lehane MJ. Association of midgut defensin with a novel serine protease in the blood-sucking fly *Stomoxys calcitrans*. *Insect Mol Biol.* 2002; 11:197–205. [PubMed: 12000638]
- Henry R, Shaughnessy L, Loessner MJ, Alberti-Segui C, Higgins DE, Swanson JA. Cytolysin-dependent delay of vacuole maturation in macrophages infected with *Listeria monocytogenes*. *Cell Microbiol.* 2006; 8:107–119. [PubMed: 16367870]

- Hu C, Aksoy S. Innate immune responses regulate trypanosome parasite infection of the tsetse fly *Glossina morsitans morsitans*. *Mol Microbiol*. 2006; 60:1194–1204. [PubMed: 16689795]
- Jahan N, Docherty PT, Billingsley PF, Hurd H. Blood digestion in the mosquito, *Anopheles stephensi*: the effects of *Plasmodium yoelii nigeriensis* on midgut enzyme activities. *Parasitology*. 1999; 119(Pt 6):535–541. [PubMed: 10633914]
- Jochim RC, Teixeira CR, Laughinghouse A, Mu J, Oliveira F, Gomes RB, et al. The midgut transcriptome of *Lutzomyia longipalpis*: comparative analysis of cDNA libraries from sugar-fed, blood-fed, post-digested and *Leishmania infantum chagasi*-infected sand flies. *BMC Genomics*. 2008; 9:15. [PubMed: 18194529]
- Keenan RJ, Freymann DM, Stroud RM, Walter P. The signal recognition particle. *Annu Rev Biochem*. 2001; 70:755–775. [PubMed: 11395422]
- Kelly P, Roberts S, Fournier PE. A review of emerging flea-borne bacterial pathogens in New Zealand. *N Z Med J*. 2005; 118:U1257. [PubMed: 15682209]
- Lehane MJ, Wu D, Lehane SM. Midgut-specific immune molecules are produced by the blood-sucking insect *Stomoxys calcitrans*. *Proc Natl Acad Sci U S A*. 1997; 94:11502–11507. [PubMed: 9326639]
- Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucl Acids Res*. 2002; 30:281–283. [PubMed: 11752315]
- Mattila JT, Munderloh UG, Kurtti TJ. *Rickettsia peacockii*, an endosymbiont of *Dermacentor andersoni*, does not elicit or inhibit humoral immune responses from immunocompetent *D. andersoni* or *Ixodes scapularis* cell lines. *Dev Comp Immunol*. 2007; 31:1095–1106. [PubMed: 17428539]
- Molina-Cruz A, Gupta L, Richardson J, Bennett K, Black WT, Barillas-Mury C. Effect of mosquito midgut trypsin activity on dengue-2 virus infection and dissemination in *Aedes aegypti*. *Am J Trop Med Hyg*. 2005; 72:631–637. [PubMed: 15891140]
- Muhlia-Almazan A, Sanchez-Paz A, Garcia-Carreno FL. Invertebrate trypsins: a review. *J Comp Physiol [B]*. 2008; 178:655–672.
- Muller PY, Janovjak H, Miserez AR, Dobbie Z. Processing of gene expression data generated by quantitative real-time RT-PCR. *Biotechniques*. 2002; 32:1372–1374. 1376, 1378–1379. [PubMed: 12074169]
- Munks RJ, Hamilton JV, Lehane SM, Lehane MJ. Regulation of midgut defensin production in the bloodsucking insect *Stomoxys calcitrans*. *Insect Mol Biol*. 2001; 10:561–571. [PubMed: 11903625]
- Nakajima Y, Ogihara K, Taylor D, Yamakawa M. Antibacterial hemoglobin fragments from the midgut of the soft tick, *Ornithodoros moubata* (Acari: Argasidae). *J Med Entomol*. 2003; 40:78–81. [PubMed: 12597657]
- Parola P, Davoust B, Raoult D. Tick- and flea-borne rickettsial emerging zoonoses. *Vet Res*. 2005; 36:469–492. [PubMed: 15845235]
- Philpott DJ, Girardin SE, Sansonetti PJ. Innate immune responses of epithelial cells following infection with bacterial pathogens. *Curr Opin Immunol*. 2001; 13:410–416. [PubMed: 11498296]
- Pizarro-Cerda J, Cossart P. Subversion of cellular functions by *Listeria monocytogenes*. *J Pathol*. 2006; 208:215–223. [PubMed: 16362984]
- Prada-Delgado A, Carrasco-Marin E, Pena-Macarro C, Del Cerro-Vadillo E, Fresno-Escudero M, Leyva-Cobian F, et al. Inhibition of Rab5a exchange activity is a key step for *Listeria monocytogenes* survival. *Traffic*. 2005; 6:252–265. [PubMed: 15702993]
- Ramakers C, Ruijter JM, Deprez RH, Moorman AF. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett*. 2003; 339:62–66. [PubMed: 12618301]
- Ramvalho-Ortigao JM, Kamhawi S, Rowton ED, Ribeiro JM, Valenzuela JG. Cloning and characterization of trypsin- and chymotrypsin-like proteases from the midgut of the sand fly vector *Phlebotomus papatasi*. *Insect Biochem Mol Biol*. 2003; 33:163–171. [PubMed: 12535675]
- Ramvalho-Ortigao M, Jochim RC, Anderson JM, Lawyer PG, Pham VM, Kamhawi S, et al. Exploring the midgut transcriptome of *Phlebotomus papatasi*: comparative analysis of expression profiles of

- sugar-fed, blood-fed and *Leishmania major*-infected sandflies. BMC Genomics. 2007; 8:300. [PubMed: 17760985]
- Rudenko N, Golovchenko M, Edwards MJ, Grubhoffer L. Differential expression of *Ixodes ricinus* tick genes induced by blood feeding or *Borrelia burgdorferi* infection. J Med Entomol. 2005; 42:36–41. [PubMed: 15691006]
- Sonenshine DE, Hynes WL, Ceraul SM, Mitchell R, Benzine T. Host blood proteins and peptides in the midgut of the tick *Dermacentor variabilis* contribute to bacterial control. Exp Appl Acarol. 2005; 36:207–223. [PubMed: 16132735]
- Valenzuela JG, Belkaid Y, Garfield MK, Mendez S, Kamhawi S, Rowton ED, et al. Toward a defined anti-Leishmania vaccine targeting vector antigens: characterization of a protective salivary protein. J Exp Med. 2001a; 194:331–342. [PubMed: 11489952]
- Valenzuela JG, Belkaid Y, Rowton E, Ribeiro JM. The salivary apyrase of the blood-sucking sand fly *Phlebotomus papatasi* belongs to the novel Cimex family of apyrases. J Exp Biol. 2001b; 204:229–237. [PubMed: 11136609]
- Wang J, Wu Y, Yang G, Aksoy S. Interactions between mutualist Wigglesworthia and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission. Proc Natl Acad Sci U S A. 2009; 106:12133–12138. [PubMed: 19587241]
- Whitworth T, Popov VL, Yu XJ, Walker DH, Bouyer DH. Expression of the *Rickettsia prowazekii* pld or tlyC gene in *Salmonella enterica* serovar Typhimurium mediates phagosomal escape. Infect Immun. 2005; 73:6668–6673. [PubMed: 16177343]
- Xu W, Huang FS, Hao HX, Duan JH, Qiu ZW. Two serine proteases from *Anopheles dirus* haemocytes exhibit changes in transcript abundance after infection of an incompatible rodent malaria parasite, *Plasmodium yoelii*. Vet Parasitol. 2006; 139:93–101. [PubMed: 16567047]
- Zaidman-Remy A, Herve M, Poidevin M, Pili-Floury S, Kim MS, Blanot D, et al. The Drosophila amidase PGRP-LB modulates the immune response to bacterial infection. Immunity. 2006; 24:463–473. [PubMed: 16618604]

Figure 1. Distribution of transcripts per functional group for uninfected (CFGF) and infected (CFGF-infected) *Ctenocephalides felis* midgut cDNA libraries. Transcripts were clustered and subsequently compared with multiple databases (see Experimental procedures). Functional groups were assigned a function based on homology to known proteins in the gene ontology and KOG databases.

Figure 2.

Alignments of select trypsin (A) and chymotrypsin (B) transcripts identified from the *Ctenocephalides felis* midgut libraries. The catalytic triad residues are highlighted in grey and marked with an asterisk (*). Conserved cysteine residues are in highlighted in black and disulfide bonds are marked with brackets. The active 'IVGG' motif of the mature protease is boxed, and the proteolytic cleavage site (R residue) is marked with an arrow.

Figure 3.

Gene expression analysis of (A) trypsin-like serine protease and (B) chymotrypsin-like serine protease over a four day infection time course. (A) Whole model analysis of gene expression of trypsin 16 shows a significant effect of time and treatment (infection) on gene expression ($R^2 = 0.518$; 'time' $P < 0.0001$; 'treatment' $P = 0.0011$; Table S2), with a significant difference in gene expression of trypsin-16 between infected and uninfected fleas post feeding. (B) Gene expression of chymotrypsin-15 also showed significant increase in gene expression over time post feeding, with a significant difference in gene expression between infected and uninfected fleas ($R^2 = 0.473$; 'time' $P < 0.0001$; 'treatment' $P = 0.027$; Table S2).

Figure 4.

Gene expression analysis of (A) A putative signal recognition particle (SRP), over a 4-day infection timecourse and (B) A putative rab5 GTPase. (A) Whole model analysis of gene expression of contig-220 (SRP) shows an effect of time and treatment on gene expression over the four day infection time course ($R^2 = 0.46$; 'time' $P = 0.006$; 'treatment' $P = 0.0012$). Additionally, whole model analysis indicates an interaction between the effect of time and treatment on gene expression ($P = 0.024$; Table S2) (B) Gene expression of contig-137, a putative rab5-like gene also showed significant increase in gene expression over time post feeding, with a significant difference in gene expression between infected and uninfected fleas ($R^2 = 0.54$; 'time' $P < 0.0001$; 'treatment' $P = 0.0096$; Table S2).

Table 1

Summary trypsin-like transcripts identified from *Ctenocephalides felis* midguts

Contig number	Best match to NR protein database	E value	Novel flea sequence	Localization *	CFGF	CFGF-I	Total number of sequences	Name of clone (and EST number)	Accession number
10	Trypsin-like serine protease	1E-127	y	SIG	13	11	24	CFGFIM1_G02	GU017990
34	PREDICTED: similar to CG16996-PA	1E-065	y	SIG	2	4	6	CFGFIS2_12	GU017991
41	ENSANGP000000021346 [<i>Anopheles gambiae</i> str. PEST]	1E-060	y	SIG	1	4	5	CFGFIS2_G03	GU017992
42	ENSANGP000000010972 [<i>Anopheles gambiae</i> str. PEST]	1E-018	y	Ind	1	0	1	CFGFS1_A06	GU017993
62	MPA3 allergen [<i>Periplaneta americana</i>]	2E-061	y	Ind	1	2	3	CFGFIM1_F09	GU017994
76	trypsin-like serine protease [<i>Ctenocephalideia felis</i>]	2E-067	y	Ind	0	2	2	CFGFIL1_F02	GU017995
92	trypsin-like serine protease [<i>Ctenocephalideia felis</i>]	4E-097	y	Ind	0	2	2	CFGFIM2_B01	GU017996
172	Predicted: similar to CG10472-PA [<i>Tribolium castaneum</i>]	6E-005	y	Ind	0	1	1	CFGFIM1_A08	GU017997
381	Serine protease [<i>Aedes aegypti</i>]	2.6	y	Ind	1	0	1	CFGFS1_E08	GU017998
35	ENSANGP000000010625 [<i>Anopheles gambiae</i> str. PEST]	1E-036	y	Cyt	1	0	1	CFGFL2_E10	GU017999
56	Serine protease [<i>Culicoides sonorensis</i>]	2E-034	y	Cyt	1	2	3	CFGFIM1_F05	GU018000
85	Trypsin-3 precursor	2E-026	y	Cyt	1	1	2	CFGFIM1_B03	GU018001
265	trypsin-like serine protease [<i>Ctenocephalideia felis</i>]	8E-021	y	Cyt	0	1	1	CFGFIS2_F08	GU18002
16	trypsin-like serine protease [<i>Ctenocephalideia felis</i>]	1E-148	n	SIG	7	5	12	CFGFIM1_D11	AAD21829.1
131	trypsin-like serine protease [<i>Ctenocephalideia felis</i>]	1E-103	n	Ind	0	1	1	CFGFIL1_D02	AAD21838.1
40	trypsin-like serine protease [<i>Ctenocephalideia felis</i>]	8E-088	n	Cyt	3	3	6	CFGFIM2_H03	AAD21830.1
55	trypsin-like serine protease [<i>Ctenocephalideia felis</i>]	1E-088	n	Cyt	4	0	4	CFGFS2_G08	AAD21839.1
296	trypsin-like serine protease [<i>Ctenocephalideia felis</i>]	6E-059	n	Cyt	1	0	1	CFGFL2_A03	AAD21837.1

* Prediction results from SignalP server.
Cyt, cytoplasmic; Ind, determinate; SIG, secreted.

Table 2
Summary of chymotrypsin-like transcripts identified from *Ctenocephalides felis* midguts

Contig number	Best match to NR protein database	E value	Novel flea sequence	Localization *	CFGF	CFGF-I	Total number of sequences	Name of clone (and EST number)	Accession number
14	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	5E-051	y	SIG	9	6	15	CFGF1L1_E03	GQ851941
17	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	7E-082	y	SIG	4	7	11	CFGFIM2_D02	GQ851942
31	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	1E-056	y	SIG	4	3	7	CFGF1L1_A05	GU018007
38	chymotrypsin precursor MDPIE	7E-041	y	SIG	2	1	3	CFGFIM2_D04	GU018008
65	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	1E-124	y	SIG	1	4	5	CFGFIM1_G11	GU018009
66	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	1E-136	y	SIG	2	4	6	CFGFIM2_C07	GU018010
187	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	5E-077	y	SIG	0	1	1	CFGFIM1_E01	GU018011
107	PREDICTED: similar to chymotrypsin-like	7E-022	y	Ind	1	0	1	CFGFL1_H05	GU018012
108	chymotrypsin precursor MDPIE	2E-038	y	Ind	1	0	1	CFGFM2_H09	GU018013
18	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	1E-016	y	Cyt	1	0	1	CFGFS1_F10	GU018014
36	PREDICTED: similar to chymotrypsin-like	1E-047	y	Cyt	1	2	3	CFGFIM1_E04	GU018015
37	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	3E-026	y	Cyt	4	0	4	CFGFS1_C11	GU018016
39	chymotrypsin-like protease II	7E-029	y	Cyt	4	0	4	CFGFS1_H07	GU018017
114	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	1E-028	y	Cyt	2	0	2	CFGFS2_E07	GU018018
161	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	1E-015	y	Cyt	0	1	1	CFGF1L2_D08	GU018019
15	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	1E-133	n	SIG	11	3	14	CFGFL2_D08	AAD21823.1
47	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	1E-129	n	SIG	10	2	12	CFGF1L1_D11	AAD21827.1
46	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	1E-140	n	SIG	1	5	6	CFGFIS2_D04	AAD21840.1
69	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	2E-091	n	SIG	2	1	3	CFGFS1_B12	AAD21824.1
67	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	2E-024	n	Ind	2	1	3	CFGFS1_D11	AAD21825.1
43	chymotrypsin-like serine protease [<i>Ctenocephalides felis</i>]	1E-118	n	Cyt	2	1	3	CFGFIM2_A08	AAD21831.1

* Prediction results from SignalP server.

Cyt, cytoplasmic; Ind, determinate; SIG, secreted.

Table 3

List of *Ctenocephalides felis* transcripts encoding genes with putative functions in intracellular trafficking and secretion

Contig number	Best match to NR protein or KOG ^a database	E value	Novel flea sequence	Localization *	CFGF	CFGF-I	Total number of sequences	Name of clone (and EST number)	Accession number
220	Signal recognition particle (GTPase)	3E-059	y	SIG	0	1	1	CFGFIS1_A12	GU018003
137	Rab5-like protein (GTPase)	4E-064	y	Ind	0	1	1	CFGFIL1_E11	GU018004
293	Vesicle coat complex CopII, subunit SFB2 ^a	3E-049	y	Ind	1	0	1	CFGFL1_H07	GU018005
303	Vacuolar sorting protein (VSP24) ^b	1E-081	y	Cyt	1	0	1	CFGFL2_C10	GU018006

* Prediction results from SignalP server. EST, expressed sequence tag; cyt, cytoplasmic; IND, indeterminate; SIG, secreted.

Table 4

List of *Ctenocephalides felis* transcripts involved with antimicrobial defense and host interaction

Contig number	Best match to NR protein database	E value	Novel flea sequence	Localization *	CFGF	CFGF-I	Total number of sequences	Name of clone (and EST number)	Accession number
11	Defensin A [<i>Anomala cuprea</i>]	3E-004	y	SIG	15	8	23	CFGFIM1_C01	GU059273
147	Serpin 7 [<i>Ctenocephalides felis</i>]	1E-029	n	Cyt	0	1	1	CFGFIL1_H05	AAN73317.1
162	Serpin 5 [<i>Ctenocephalides felis</i>]	4E-035	n	Cyt	0	1	1	CFGFIL2_D10	AAN73321.1
361	Serpin 4 [<i>Ctenocephalides felis</i>]	1E-009	n	Ind	1	0	1	CFGFM2_E03	AAN73322.1
320	Complement component 1q binding protein-like protein [<i>Aedes aegypti</i>]	6E-032	y	Cyt	1	0	1	CFGFM1_A03	GU059274
291	Peptidoglycan recognition protein LB	1E-049	y	SIG	1	0	1	CFGFL1_G07	GU059275
28	Galactose-specific C-type lectin	2E-014	y	SIG	3	3	6	CFGFM1_F07	GU059276
30	Galactose-specific C-type lectin	1E-010	y	Ind	1	0	1	CFGFM1_E01	GU059277
341	tetraspanin	2E-054	y	Ind	1	0	1	CFGFM1_G12	GU059278

* Prediction results from SignalP server.

EST, expressed sequence tag; Cyt, cytoplasmic; Ind, determinate; SIG, secreted.

Table 5

List of primers used in gene expression studies

Cluster	Annotation	Forward Primer	Reverse primer
16	Trypsin-like serine protease	ACATTCACCCAGACTACGACACG	TCCTTGCCACCTCTTCTAACTCCG
15	Chymotrypsin-like serine protease	GCTGCTCACTGTTTGGGTTATGAC	TGTTGTCTTACACATTGCTGGTGG
220	Signal recognition	GGTATTGTGTTTGTGTTGATTCC	GATTTCCTCCGTGCTCAC
137	RAB5B-like protein	TCTACCAGACAGAACGCTGC	GGAGAAGCCTGCCTTTACAAC
130	Glutathione-S-Transferase (sigma)	GGAAAGGAAGTGCACCAGAGTTTG	GCCCAAGTTAGTCTTCCAAGAGC
235	peritrophin-like protein	GATAAGTGCCCGAATATGCTGAC	CCATTCTACATTTGTGCGTAATC
11	Defensin	CAATGGCAGAAGACACACCAGATG	TCCTCCAAGTTGGCATACTACACG
341	Tetraspanin	GGTGAGCTTGCTGAAACCTTGGA	ACTTTGGAGCTCATCACAGCAGGA
28–29	Galactose-specific C-type lectin	ACGTAGTTGGAGGCAAATTGGCAG	AGCAGGACACATCATTCCATCCCT
147	Serpin	TGCCACAGGAGGTTTCATAATGGC	GGATTAACGTGTTGACCATGGAAC
291	PGRP-LB	ATAGGCTACAGCTTGCAGTGGGA	ATCGCTCCAGTTCCGATTAGGCA
30	Galactose-specific C-type lectin	AGGGAATTTGCTCGTGAGACCTAC	ATTATCTGGTTGACCAGGAGCCCA
79	muscle-specific (actin)	GGTCGGTATGGGACAAAAGGAC	GTAGATTGGGACGGTGTGAGAGAC