

An integrated transcriptomic and proteomic analysis of the secretome of the helminth pathogen, *Fasciola hepatica*: proteins associated with invasion and infection of the mammalian host.

Mark W. Robinson¹, Ranjeeta Menon², Sheila M. Donnelly¹, John P. Dalton^{1*} & Shoba Ranganathan^{2*}.

¹ Institute for the Biotechnology of Infectious Diseases (IBID), University of Technology Sydney (UTS), Level 6, Building 4, Corner of Thomas & Harris Street, Ultimo, Sydney, NSW 2007, Australia. ² Department of Chemistry and Biomolecular Sciences & ARC Centre of Excellence in Bioinformatics, Macquarie University, Sydney, NSW 2109, Australia.

Corresponding author: Mark W. Robinson. Institute for the Biotechnology of Infectious Diseases, University of Technology Sydney, Building 4, Harris Street, Ultimo, Sydney, New South Wales 2007, Australia. Tel: 61-2-95144127; Fax: 61-2-95148206; Email: mark.robinson@uts.edu.au

*These authors contributed equally to this work.

Running title: Developmental Regulation of the *Fasciola* secretome

Key words: Helminth, *Fasciola*, proteomics, cathepsin, asparaginyl endopeptidase

Abbreviations

NEJ – newly excysted juvenile

EST – expressed sequence tag

rEST – representative expressed sequence tag

BLAST – Basic local alignment search tool

nanoLC-ESI-MS/MS - nano liquid chromatography electrospray ionisation tandem
mass spectrometry

Summary

In order to infect their mammalian hosts, *F. hepatica* larvae must penetrate and traverse the intestinal wall of the duodenum, move through the peritoneum and penetrate the liver. After migrating through and feeding on the liver, causing extensive tissue damage, the parasites move to their final niche in the bile ducts where they mature and produce eggs. Here, we have integrated a transcriptomic and proteomic approach to profile *Fasciola* secretory proteins that are involved in host-pathogen interactions and to correlate changes in their expression with the migration of the parasite. Prediction of *F. hepatica* secretory proteins from 14,031 expressed sequence tags (ESTs) available from the Wellcome Trust Sanger Centre using the semi-automated EST2Secretome pipeline showed that the major components of adult parasite secretions are proteolytic enzymes including cathepsin L, cathepsin B and asparaginyl endopeptidase cysteine proteases as well as novel trypsin-like serine proteases and carboxypeptidases. Proteomic analysis of proteins secreted by infective larvae, immature flukes and adult *F. hepatica* showed that these proteases are developmentally regulated and correlate with the passage of the parasite through host tissues and its encounters with different host macromolecules. Proteases such as FhCL3 and cathepsin B have specific functions in larvae activation and intestinal wall penetration, while FhCL1, FhCL2 and FhCL5 are required for liver penetration, and tissue and blood feeding. Besides proteases, the parasites secrete an array of antioxidants that are also highly regulated according to their migration through host tissues. However, whereas the proteases of *F. hepatica* are secreted into the parasite gut *via* a classical ER/Golgi pathway, we speculate that the antioxidants, which all lack a signal sequence, are released *via* a non-classical trans-tegumental pathway.

Introduction

Fasciola hepatica is a helminth (worm) parasite with a world-wide distribution. Although traditionally regarded as a parasite of livestock, particularly sheep and cattle, that results in a large economic loss to the agricultural community it has recently emerged as an important human infection in many regions of the world, including South America, Iran, Egypt and mainland South-East Asia (1). Dormant larvae contained within cysts adhere to vegetation and emerge as infective juveniles (newly excysted juveniles, NEJs) in the duodenum following ingestion by animals or humans. They infect their hosts by rapidly penetrating the intestinal wall and entering the peritoneal cavity where they break through the liver capsule. After 8-12 weeks of consistent burrowing, feeding and growth within the liver parenchyma they move to their final destination within the bile ducts where they mature and produce enormous numbers of eggs (2). The two distinct clinical phases of fasciolosis are directly related to the parasites' migration: acute fasciolosis, which manifests as fever, abdominal pain, weight loss and hepatomegaly, is associated with liver tissue damage and inflammation caused by the migrating immature parasites, whereas chronic fasciolosis (usually sub-clinical) is coupled with the presence of the mature adult flukes in the bile ducts (1,3).

These invasive helminth parasites undergo complex changes as they migrate within their definitive mammalian hosts. The developing parasites encounter different host tissues and macromolecules and have to contend with a continually changing physiological microenvironment (such as pH and oxygen availability), and a mounting humoral and cellular host immune response. Morphological and ultrastructural studies clearly show major alterations of the parasite surface and gastrodermis, the two host-parasite interfaces, as they migrate and grow (4). However,

we are only beginning to understand the molecular and biochemical interactions that occur between host and parasite and how these adjust as the parasite's development progresses. A deeper knowledge of such host-pathogen interplay should provide data on novel targets for anthelmintic compounds and potential anti-parasite vaccine candidates.

Increasingly, proteomic analysis is being employed as a means to investigate the interaction of helminth parasites and their hosts, particularly in cases where obtaining pathogen material is difficult (5-9). For some helminths these studies have been facilitated by the availability of large transcriptomic datasets (10-12). Unfortunately, the *Fasciola* nucleotide sequences available in GenBank are relatively few and highly redundant (298 for *F. hepatica* and 142 for *F. gigantica* as of 15/01/2009) and the adult *F. hepatica* ESTs currently available from the Wellcome Trust Sanger Centre (14,031 reads) are unannotated (therefore the current identification of peptides of interest requires manual BLAST analysis using specific query sequences). Accordingly, in this study we employed the semi-automated EST2Secretome pipeline to analyse all available *F. hepatica* ESTs for secretory proteins potentially involved in host-pathogen interactions (13). EST2Secretome was developed in our laboratory by optimizing signal-peptide-mediated secreted protein prediction from our earlier predictions of parasitic nematode ESTs (14,15). We have integrated this transcriptomic data with a proteomic analysis of the molecules secreted by adult *F. hepatica*, with a particular emphasis on proteolytic enzymes. Furthermore, we have also analysed the somatic and secreted molecules of the infective NEJ parasites and compared these to the secretome of immature and adult parasites taken from liver tissues. In doing so we have produced a comprehensive view of how *F. hepatica* differentially and developmentally express and secrete proteolytic enzymes

and other molecules according to the specific challenges faced in the intestine, liver and bile ducts. *Fasciola hepatica* cathepsins B and cathepsin L (FhCL3) are stored as zymogens within the infective larvae ready to be *trans*-activated by specific asparaginyl endopeptidases, and released to perform the highly-specific function of host tissue invasion (intestinal epithelium and liver capsule). By contrast, cysteine proteases belonging to the phylogenetic clades FhCL1, FhCL2 and FhCL5 are expressed during the later stages within the liver and bile duct and function in tissue degradation and feeding alongside the cell-lytic protein saposin and a newly-described prolylcarboxypeptidase. Several novel developmentally-regulated cathepsin L and cathepsin B cysteine proteases, and members of two serine protease families, namely carboxypeptidase and trypsin-like serine proteases, which may also have important roles in host-parasite interplay, have also been identified. Finally, our observations have led us to propose that whereas the major proteases of *F. hepatica* are secreted into the parasite gut *via* a classical ER/Golgi pathway, an array of abundant and highly-regulated antioxidants are released *via* a non-classical trans-tegumental pathway.

Experimental Procedures

Analysis of the F. hepatica ESTs using EST2Secretome

Transcriptome analysis for secretory proteins from 14031 *Fasciola* EST sequences, available from the Sanger Centre, UK (<ftp://ftp.sanger.ac.uk/pub/pathogens/Fasciola/hepatica/ESTs/>) was performed using the semi-automated EST2Secretome pipeline recently reported by Nagaraj *et al.* (13). The EST2Secretome pipeline integrates a number of high quality programs for EST cleaning (SeqClean and RepeatMasker), assembly into contigs and singletons (CAP3),

conceptual translation (ESTSCAN) and prediction of secreted proteins, based on the presence of N-terminal secretory signal sequences (SignalP), and follows with the elimination of membrane proteins using TMHMM. This approach has proven very reliable in identifying secretory proteins from a number of helminths including the bovine lungworm, *Dictyocaulus viviparus* (16), gastrointestinal worm, *Trichostrongylus vitrinus* (15), and the hookworm, *Ancylostoma caninum* (17). The predicted secretory proteins are extensively annotated for functional protein families and motifs (InterProScan), gene ontologies (BLAST2GO) pathways (KOBAS), protein-protein interactions (comparison to IntAct), mapping to *Caenorhabditis elegans* proteins (from WormPep) and subsequent correlation to RNAi phenotype data. The SignalP threshold value for secretory signal peptide prediction was set at 0.5 as determined for the large-scale secretory protein prediction from helminth parasite ESTs (13).

Independently, un-annotated *Fasciola* ESTs that matched MS/MS data were used as queries for BLASTn (18) searches of all nucleotide sequences in the Genbank and EMBL databases. Searches were performed using the NCBI server (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Open reading frames were constructed from the *Fasciola* EST sequences (guided by their best BLASTn hits) and hypothetical proteins derived from their conceptual translation were submitted to the InterProScan algorithm (19) to detect conserved domains and motifs to help assign putative protein identifications.

Proteomic analysis of parasite somatic and secreted proteins: gel electrophoresis and mass spectrometry

The profile of proteins secreted by adult *F. hepatica* has been recently studied using two-dimensional electrophoresis (2-DE) by us and others (8,20,21). However, because of the paucity of material that can be obtained for the infective juvenile NEJ parasites and 21 day-old liver stage parasites we employed 1-DE in the present study. Nevertheless, we validated this approach by subjecting adult parasite secretory proteins to 1-DE and comparing the data with those reported by Robinson *et al.* (8). Protein samples were analysed by 1-DE using NuPage® Novex® 4-12 % Bis-Tris gels (Invitrogen). NuPage® LDS sample buffer plus Sample Reducing Agent (Invitrogen) were added to the samples and heated at 95°C for 5 min prior to electrophoresis. Gels were stained with Colloidal Coomassie Blue G250 (Sigma) and de-stained with 10 % methanol (v/v) and 7 % acetic acid (v/v). Following visualisation, gel lanes were cut into 6 sections for analysis by mass spectrometry. Briefly, the individual gel sections were cut into smaller pieces (approximately 1 mm²) and reduced and alkylated with 5 mM tributylphosphine and 20 mM acrylamide (Sigma) in 100 mM NH₄HCO₃ for 90 mins. The excised sections were then in-gel digested with trypsin (Sigma Proteomics grade) and the peptides solubilised with 2 % formic acid (Sigma) prior to analysis by nano liquid chromatography electrospray ionisation tandem mass spectrometry (nanoLC-ESI-MS/MS) using a Tempo nanoLC system (Applied Biosystems) with a C18 column (Vydac) coupled to a QSTAR Elite QqTOF mass spectrometer running in IDA mode (Applied Biosystems). Peak list files generated by the Protein Pilot v1.0 software (Applied Biosystems) using default parameters were exported to local MASCOT v2.1.0 (Matrix Science) or PEAKS (Bioinformatics Solutions Inc.) search engines for database searching.

Database searches

MS/MS data was used to search 3239079 entries in the MSDB (20060809) database using MASCOT whereas PEAKs software was used to search the 14,031 *F. hepatica* EST sequences from the Wellcome Trust Sanger Institute. The enzyme specificity was set to trypsin and propionamide (acrylamide) modification of cysteines was used as a fixed parameter and oxidation of methionines was set as a variable protein modification. The mass tolerance was set at 100 ppm for precursor ions and 0.2 Da for fragment ions. Only 1 missed cleavage was allowed. For MASCOT searches, matches achieving a molecular weight search (MOWSE) score >70 with at least two high-scoring individual peptides were considered to be significant (6,7). For PEAKs searches of the *Fasciola* EST database, at least two high-scoring (>60 %) matching peptides were required. However, other criteria were considered in assigning a positive identification including concordance between the calculated theoretical molecular mass of the protein and the observed position of the polypeptide by 1-DE. In order to account for matches to multiple members of the *Fasciola* cathepsin family, peptides specific to individual enzymes or clades were searched for (8; results section). The proteomics data and transcriptome analysis were integrated to give a more complete view of gene expression/secretion in adult *F. hepatica*. This integrated dataset provided a framework for comparison with the infective NEJs and 21 day-old immature flukes at the sub-proteome (i.e. secretome) level.

Quantitation of Fasciola proteins by emPAI

We analysed the exponentially modified protein abundance index (emPAI) provided in the output of the MASCOT MS/MS ion search to estimate the relative expression of proteases and antioxidants identified in the developmental stages of *F. hepatica*. The emPAI used by MASCOT is a modification of the formula developed by

Ishihama *et al.* (22) and gives a label-free relative quantification of the proteins in a mixture. The raw emPAI values obtained represent the transformed ratio of the number of experimentally observed peptides (composed of unique precursor ions, including different charge states of the same peptide, which match or exceed the threshold level for homology or identity) to the total number of peptides that can theoretically be detected within the operating mass range and retention time range of the mass spectrometer (calculated by MASCOT based on the mass of the protein, the average amino acid composition of the database searched and the enzyme specificity). For this analysis, the raw emPAI values (averaged from 3 separate gels) for all *Fasciola* proteases or antioxidants identified were added to give a figure representing total expression for each within a particular developmental stage. The raw emPAI values for each individual protease or antioxidant were then converted to a percentage of this total to estimate their relative expression levels (Figures 2A-3A). As the method used by MASCOT to calculate the number of observable peptides differs from that originally described by Ishihama *et al.* (22) and is not freely available, it was decided not to calculate emPAI values manually for those proteins identified using the PEAKS software in order to avoid introducing errors into the subsequent analysis. In order to account for redundancy, molecules potentially containing shared sequences were grouped together (e.g. all cathepsin B variants were classed as FhCB, all cathepsin L1 variants were classed as FhCL1 etc).

Excystment of F. hepatica metacercariae and preparation of somatic larvae proteins

The dormant cysts of *F. hepatica* metacercariae contain two layers, the outer of which can be contaminated with plant or other extraneous material. Here we describe a method for removing the outer cyst and adhered material so that somatic

proteins of the dormant infective larvae can be analysed. In addition, we describe a rapid method for activating the larvae and inducing them to emerge from the cysts so that *in vitro*-secreted proteins can be isolated. Thus, *F. hepatica* metacercariae (Baldwin Aquatics Inc., Monmouth, Oregon) were vortexed for 10 seconds in 0.5 % sodium hypochlorite and then incubated at room temperature for 20 min (this procedure dissolves the outer cyst layer). They were then washed three times in distilled water by centrifugation at 2000 x g for 2 min. The juvenile larvae, which were now only contained within a clear inner cyst layer, were used to prepare somatic protein extracts and secretory proteins. To prepare the somatic extracts the parasites were homogenised in RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM PMSF, 1 mM EDTA, 1 % Triton X-100, 1 % sodium deoxycholate (Sigma), 0.1% SDS, 1x Complete mini protease inhibitor cocktail (Roche)) and placed on ice for 30 min. The protein extract was centrifuged at 13,000 rpm for 10 min to remove insoluble debris and the supernatant stored at -20°C until use.

To prepare secreted proteins the washed parasites were resuspended and incubated in the following excystment medium; 0.5 % sodium bicarbonate, 0.4 % sodium chloride, 0.2 % sodium taurocholate, 0.07 % concentrated HCl and 0.006 % L-cysteine for up to 3 h at 37°C / 5 % CO₂. NEJs emerged from the cysts within two-three hours and were transferred to pre-warmed (37°C) culture medium, RPMI 1640 medium (Invitrogen) containing 2 mM L-glutamine, 30 mM HEPES, 0.1 % (w/v) glucose and 2.5 µg / ml gentamycin, and cultured for 24 hours.

To determine whether cysteine proteases were essential to cyst rupture, *F. hepatica* metacercariae (fifty per treatment in triplicate) were either excysted with medium as described above or in medium lacking the 0.006 % (w/v) L-cysteine, or supplemented with the cysteine protease inhibitors *trans*-epoxysuccinyl-L-

leucylamido(4-guanidino)butane (E-64, Sigma) or 1 mM carbobenzoxy-phenylalanyl-alanine-diazomethyl ketone (Z-Phe-Ala-CHN₂, Bachem St. Helens, UK) to a final concentration of 1 mM. The numbers of excysted parasites were counted after 3 h incubation at 37°C / 5 % CO₂ and the data was analysed using Student's *t*-test. The experiments were repeated twice.

Preparation of immature and mature adult F. hepatica secretory proteins

Immature *F. hepatica* flukes (21 day-old) were recovered from the livers of female BALB/c mice (experimentally infected with 20 metacercariae) while adult parasites were recovered from the bile ducts of Merino sheep 16 weeks after an experimental infection with 200 metacercariae. Immature and adult parasites were washed in pre-warmed (37°C), phosphate-buffered saline (PBS), pH 7.3, before transfer to culture medium (as described above) for 24 and 8 h, respectively. Secretory proteins were concentrated from the culture supernatants by precipitation with methanol / chloroform as previously described (6). Pellets were resuspended in 10 µl RIPA buffer and stored at -20°C prior to separation by electrophoresis.

Results

Transcriptomic profiling of adult F. hepatica secretory proteins

Of the 14031 adult *F. hepatica* raw EST sequences available, a total of 12954 (92.3 %) quality sequences were obtained (Table 1). Cluster analysis of the 12954 ESTs yielded 4236 representative ESTs (rESTs; 2749 contig and 1487 singleton sequences), of which 2960 (68.9 %) had open reading frames. These pre-processed ESTs ranged from 60-2093 bp, with a mean of 569 bp and a standard deviation (S.D.) of 268 bp. After clustering, the mean length of the contigs increased to 788 bp and standard

deviation (S.D) of 358 bp. The G+C content of the coding sequences was 44.5 % which is similar to the figure of 43.5 % reported for the adult bovine lungworm, *Dictyocaulus viviparus* from EST analysis (16). This value is slightly higher than that reported for the related trematode *S. mansoni* (34 %; 23), but consistent with those reported for nematodes (32-51 %; 24), *Caenorhabditis elegans* (37 %) and *C. briggsae* (38 %) (25).

All rESTs were then subjected to analysis using our recently reported semi-automated bioinformatics platform (EST2Secretome, 13) to predict secretory proteins from the *F. hepatica* EST database. Using this approach, we have identified a sub-set of parasite proteins likely to participate in the most significant interactions that occur between the adult stage of this parasite and its mammalian host. Thus, 173 *Fasciola* secretory proteins were predicted by the EST2Secretome pipeline, with 160 true positives based on the EST2Secretome annotations mapping homology to proteins in a non-redundant secreted protein database (SecProtSearch) derived from the literature, the secreted protein database, SPD (26) and the manually curated signal peptide database, SPdb (27); as well the GO annotations of subcellular localization of the top homologues identified by BLAST (92.5 % accuracy) (Table 1). Since ESTs are usually not full-length and often truncated, manual inspection of the final dataset is required as it is possible that transmembrane sequences are erroneously identified as secreted proteins and thus elude the filtration step by TMHMM.

From the detailed annotations of the 160 adult secreted proteins the predominance of cathepsin L cysteine proteases is clearly evident as these are represented by a total of 66 (41.2 %) proteins (Table 2, and Supplementary Table 1). Robinson *et al.* (8) recently showed that the adult *F. hepatica* cathepsin L proteases separated into five distinct clades. Of these 66 adult rESTs encoding cathepsin Ls, 48

(72.5 %) represented clade FhCL1 (38 sub-clade FhCL1A and 10 sub-clade FhCL1B), 11 (17 %) encoded clade FhCL2 two (3 %) encode clade FhCL5 cathepsins (8). Consistent with Robinson *et al.* (8) no cDNAs encoding clades FhCL3 or FhCL4 were detected in the adult ESTs, since these proteases have been reported as specific for the infective NEJ (see below, and 28). Interestingly, five rESTs (7.5 %) encoded cathepsin Ls that could not be placed into any of the five phylogenetic clades based on primary sequence alignment analysis.

The next most abundant secreted protein based on the number of ESTs identified was saposin-like protein 3 which has been reported as a secreted protein by Grams *et al.* (29) and suggested to play a role in red blood cell lysis (30). Other proteins of interest to our study include several novel cathepsin B cysteine endoproteases (designated cathepsin B4 – B10 in the current study), four novel asparaginyl endopeptidases or legumains (designated legumain 4 – 7 in the current study) and a cysteine protease inhibitor, cystatin. Additionally, three putative novel adult serine proteases were identified – a serine carboxypeptidase and two proteins with trypsin-like protease domains. Other ESTs encoded secreted vitelline protein B1 that is found in eggs produced by the adult parasite (31) (Table 2).

A *Fasciola* protein disulphide isomerase was also predicted that has previously been identified in the secretions of adult flukes (32). Protein disulphide isomerases have roles in protein folding and a *Fasciola* recombinant enzyme was shown to mediate the oxidative refolding of reduced RNase (32). A putative peptide with a number of cubulin domains was also predicted in the current analysis. Cubulin domains occur predominantly in extracellular proteins or plasma membrane-associated proteins with a range of functions including complement activation, tissue repair, cell signalling and inflammation (33). Although the *Fasciola* peptide contains

a predicted N-terminal transmembrane region, its molecular function remains unknown. We note that while orthologues are available for 6 proteins, including an uncharacterised secretory protein from *Clonorchis sinensis*, their function remains elusive. A total of 36 secreted proteins (21.9 %) are novel but no database matches exist at the present time (Table 2).

Proteomic profiling of adult F.hepatica secreted proteins.

We and others have previously characterised the major secretory proteins expressed by adult *F. hepatica* using 2-DE (8,20,21). To complement these earlier studies, and to validate the use of 1-DE for proteomic analysis, we analysed tryptic digests extracted from gel sections of adult *F. hepatica* secreted proteins (see Figure 1) by mass spectrometry. Twenty-two different proteins secreted by adult *F. hepatica* were identified in this analysis: 19 matched to previously identified *Fasciola* cDNAs and 3 corresponded to putative proteins encoded by novel ESTs identified by our present EST2Secretome analysis but were unidentified by Robinson *et al.* (8) (Figures 2B-3B; Supplementary tables 2-3).

In accordance with the transcriptomic predictions, cathepsin L proteases were highly represented in adult fluke secretions. Matches to 13 cathepsin L sequences were observed and included clades FhCL1, FhCL2 and FhCL5 enzymes (FhCL4 or FHCL3 enzymes were not detected) (8). These identifications were based on the presence of clade-specific peptide matches such as NSWGLSWGER (ion m/z 596.30; +2) and VTGYITVHSGSEVELK (ion m/z 590.32; +3) which are clade 1-specific peptides, three peptides DYYYVTEVK (m/z 590.26; +2), VTGYITVHSGDEIELK (m/z 604.32; +3) and LTHAVLAVGYGSQDGTDYWIVK (m/z 798.42; +3) characteristic of FhCL2 cathepsin Ls, as well as the presence of peptides such as

NSWGTWWGEDGYIR (m/z 863.90; +2) and FGLETESSYPYR (m/z 724.85; +2) that aid the identification of clade 5 cathepsin Ls (8). The three previously unidentified secreted proteins were saposin, a peptidyl-prolyl cis-trans isomerase and a protein with homology to an uncharacterised *C. sinensis* secretory protein.

Several proteins that were predicted to be secreted by EST2Secretome were not identified by mass spectrometry of the adult parasite secretome both in this study and in the studies of Jefferies *et al.* (20), Morphew *et al.* (21) and Robinson *et al.* (8); these were cathepsin B, legumain, serine carboxypeptidase and two trypsin-like enzymes, and the egg-shell vitelline protein B1. On the other hand, proteomic analysis detected four fatty acid-binding proteins (FaBP1, FaBP2, FaBP3 and Fh15) and two redox enzymes (peroxiredoxin and thioredoxin) that were not predicted by EST2Secretome analysis. A putative novel prolylcarboxypeptidase (also observed in 21-day old immature parasites, see below) was also identified by mass spectrometry and yet was not predicted to be a secretory peptide. This is surprising since at least one adult fluke EST (Fhep06a01.q1k) encoding the N-terminal end of the enzyme contains a putative signal peptide. However, this EST was not retained as a singleton, after repeat-masking and truncation, in the early stages of EST processing in EST2Secretome.

Identification of F. hepatica dormant larval somatic proteins and NEJ secretory proteins

By removing the outer cyst layer of the dormant metacercarial infective stage of the parasite we could extract somatic protein without contamination with extraneous proteins and analyse these by 1-DE. Approximately 14 protein bands could be visualised following Coomassie blue-stained SDS-PAGE (Figure 1) which on

inspection appeared very similar to the banding pattern of *Fasciola* larval somatic proteins reported by Tkalcevic *et al.* (34). Proteins in these bands were in-gel digested with trypsin and analysed by nanoLc-ESI-MS/MS and the resulting CID data was used for database searching. A total of 26 different *Fasciola* dormant larvae proteins were identified: 12 matched to previously identified *Fasciola* cDNAs (or cDNAs from related trematode species) and 14 corresponded to proteins encoded by novel ESTs (Figures 2B-3B; Supplementary tables 2-3). A further five peptides encoded by *Fasciola* ESTs were also identified but these lacked conserved protein domains and could not be assigned putative functions based on BLAST searches.

Of the 26 positively-matched *Fasciola* proteins six were proteases including, two cathepsin L3 proteases, three cathepsin B endopeptidases and an asparaginyl endopeptidase-like precursor (discussed in detail below). Others included structural proteins related to muscle function such as actin, myosin-regulatory light chain and a troponin C homologue (Supplementary tables 2-3). Four metabolic enzymes were also identified; pyruvate carboxylase (gluconeogenesis), malate dehydrogenase (TCA cycle) as well as aldolase and enolase (glycolysis). Mass spectrometry data also matched to a putative cullin protein (roles in protein degradation and ubiquitination), a cyclophilin-like peptidyl-prolyl cis-trans isomerase and a tetraspanin membrane protein. Other notable peptide matches included the *Fasciola* antioxidant enzymes peroxiredoxin, two histone proteins, a heat shock protein 70, dynein light chain, a ribosome production factor, ribosomal protein L30 and a protein bearing a conserved RNA-binding motif (Supplementary tables 2-3).

By using an artificial medium that replicated the surfactant and reducing conditions in the duodenum we could activate the dormant infective larvae and induce them to excyst. Proteins secreted by *F. hepatica* NEJs during *in vitro* culture were

isolated from the medium, separated by 1-DE and analysed by nano-Lc-ESI-MS/MS (Figure 1). Matches to 29 different proteins, of which ten were proteases including seven cathepsin L3 proteases, one cathepsin B, and two asparaginyl endopeptidase-like proteases were obtained from MS/MS data from *Fasciola* NEJ secreted proteins.

Other protein matches to *Fasciola* cDNA sequences included enolase, three fatty acid binding proteins (Fh2, Fh3 and Fh 15) and peroxiredoxin (Figure 3B; Supplementary tables 2-3). A further 16 different putative peptides were identified following searches of the *F. hepatica* EST database including the metabolic enzymes fructose-bisphosphate aldolase, phosphoenolpyruvate carboxykinase, glyceraldehyde 3-phosphate dehydrogenase, malate dehydrogenase and an ATPase. Other matches included putative structural proteins such as calponin, spermadhesin and histones (H2A, H2B and H4), redox enzymes (thioredoxin and peptidyl-prolyl cis-trans isomerase) and an uncharacterised protein with predicted transmembrane regions. Finally, molecules with roles in protein turnover such as ubiquitin and a putative serpin were also identified.

Identification of immature F. hepatica secretory proteins

Parasites were removed from the livers of infected mice and maintained *in vitro* for collection of secreted proteins. The overall complexity of the secretory proteins of these 21 day-old immature *F. hepatica* was greater compared with that of the NEJs or adult parasites (Figure 1) and yielded a total of 45 different protein identifications (Figures 2B-3B; Supplementary tables 2-3). Of these, 34 were matched to previously identified *Fasciola* cDNAs and 11 corresponded to putative proteins encoded by novel ESTs. Mass spectrometry data also matched to peptides encoded by a further

two *F. hepatica* EST sequences that lacked conserved protein domains and could not be assigned putative functions based on BLAST searches.

Of the 45 positively-matched proteins 22 were proteases and included 14 cathepsin Ls, four cathepsin Bs, three asparaginyl endopeptidases (legumains) and a newly discovered prolylcarboxypeptidase. The remaining 23 proteins secreted by the immature liver stage parasites included a glutathione *S*-transferase (GST) sigma class enzyme and four isoforms of mu class GSTs (GST1, GST7, GST47 and GST 51), four fatty acid-binding proteins (FaBP1, FaBP2, FaBP3 and Fh15), two saposin-like proteins (SAP1 and SAP3), two enzymes of glycolysis (enolase and triosephosphate isomerase) and two enzymes involved in cell redox homeostasis (peroxiredoxin and protein disulphide isomerase). Other significant peptide matches included annexin, ferritin, ubiquitin, a 14-3-3 protein, a multi-cystatin and a putative ABC transporter protein.

Discussion

The database of *F. hepatica* ESTs available from the Wellcome Trust Sanger Centre is now sufficiently large as to allow a significant transcriptomic analysis of this helminth pathogen, which until now has been lacking in this field. We have employed our newly developed EST2Secretome pipeline to analyse these data sets with the view to identifying molecules secreted by the adult trematode, *F. hepatica*, unlike earlier applications to nematode parasites which focused on the complete transcriptome. Furthermore, we have integrated these results with data generated from proteomic analysis performed here and in previous reports (8,20,21) to build a picture of how the developing parasite sustains itself within the mammalian host, with particular emphasis on proteases as virulence and tissue-damaging factors.

EST2Secretome analysis of the adult *F. hepatica* ESTs identified 160 cDNAs encoding secreted proteins, 41% of which encoded cathepsin L cysteine proteases. The abundance of adult cathepsin L sequences was noted in a previous analysis of entries in the public databases (8,35). When these sequences were subjected to a phylogenetic investigation it was demonstrated that they could be separated into five clusters or clades, FhCL1, FhCL2, FhCL3, FhCL4 and FhCL5. Due to the high level of conservation between the clades it was not possible to study temporal patterns of specific cathepsin gene expression by conventional means (such as RT-PCR). Analysis of the >14,000 adult parasite EST sequences in the present study confirmed the expression of three of these distinct clades (FhCL1, FhCL2 and FhCL5) in this fully mature stage of the parasite. It also supported the lack of expression of clades FhCL3 and FhCL4 in adult parasites, although proteases encoding these genes are expressed by NEJs (see below and 28). We also identified five novel cathepsin L proteases that although encoded by partial nucleotide sequences (contigs 1553, 2626, 1886 and singletons Fhep55b05.q1k and Fhep18b10.q1k) may represent a new phylogenetic cluster.

The *F. hepatica* cathepsin L proteases, the largest family of proteases known in any helminth pathogen, arose by a series of gene duplications. Members underwent selective functional diversity brought about by specific alterations within the active site that ultimately produced a repertoire of proteases with overlapping but distinct substrate specificity (8). Our proteomic investigation was consistent with the EST2Secretome analysis by demonstrating the predominance of these proteases in the adult *F. hepatica* secretome (approximately 80 % of total protein secreted) and indicated a major role for the enzymes in supporting the parasites' existence within the bile duct. Here, the adult parasites are obligate blood feeders and, therefore, the

most obvious function for the cathepsin L enzymes is in the digestion of blood macromolecules. Recently, Lowther *et al.* (36) showed that members of the most predominant cathepsin L clade, FhCL1, have evolved an active site with a strong preference for hydrophobic amino acids, such as Leu, Ala, Phe and Val, that are particularly abundant in haemoglobin (42 %) indicating a specific adaptation for digestion of this protein. Haemoglobin is the prime supplier of amino acids which the parasites use in the anabolism of egg proteins, a major task since they produce 30-50,000 eggs/day/worm (37). Members of the other major clade, FhCL2, have specific amino acid substitutions in the active site cleft that gives these proteases the ability to cleave substrates with Pro residues and facilitates the degradation of proline-rich interstitial collagen during the parasites' migration through the tissues (38).

Consistent with the importance of assimilating nutrient from host red blood cells was the identification of ESTs encoding a saposin-like molecule as the next most abundant in the dataset. This protein, which was termed FhSAP-2 because of its similarity to a saponin-like protein family and the amoebapore precursors of *Entamoeba histolytica*, has been implicated in the penetration and lysis of ingested host red blood cells, not only in *Fasciola* (29,30) but also in other helminths (39). Saposin-like molecules are likely secreted from granules within cells lining the parasite oesophagus so that ingested red blood cells are lysed as soon as they enter the digestive track. Hydrolysis of released haemoglobin by the cathepsin L proteases then takes place in the acidic caecum (37).

Several *Fasciola* proteins predicted as components of the secretome by our EST2Secretome analysis were not identified by mass spectrometry in secretions of adult parasites maintained in culture (8,20) or within the bile ducts (21). The most highly represented of these was the major eggshell protein, vitelline protein B1

(vpB1), which is produced by mature vitelline cells to form the hard protective trematode eggshell (31). However, since this process occurs within the ootype and uterus of the adult fluke it is likely that vpB1 is retained within the eggshell and not secreted outside the parasite. Other examples are the transcripts encoding six novel cathepsin B cysteine proteases and five new asparaginyl endopeptidases. While some isoforms of these proteases are secreted by the infective larvae and immature liver-stage flukes (see below), they are absent from the secretome of adult *F. hepatica* or *F. gigantica* as previously reported using biochemical, proteomic and immunoblotting methods (8,40,41). Considering these observations, and the fact that cathepsin B and asparaginyl endopeptidases also function internally within cells (42,43), it is likely that these newly discovered proteases have functions within the internal tissues of this complex multicellular parasite, and are involved in generalised protein turnover, remodelling and/or catabolism. Alternatively or additionally, since immunocytochemistry and *in situ* hybridization has localised cysteine proteases and transcripts to the reproductive structures of other trematodes (44,45) these proteases may be involved in the process of egg production. Other proteins identified by EST2Secretome analysis but not by proteomics, including the previously uncharacterised proteins such as deoxyribonucleases, carboxypeptidase and trypsin-like peptidases that have homologs in *S. japonicum* and *C. sinensis*, may also be confined to internal tissues of the parasite or be expressed and secreted at levels below the detection capacity of the proteomic methods so far employed (cDNAs of these were poorly represented in the transcriptome of the adult parasite).

Conversely, a number of *F. hepatica* proteins that have previously been described as major components of the adult parasite secretions by standard protein and proteomics methods were not predicted by the EST2Secretome pipeline. These

included the fatty acid-binding proteins, FaBP1, FaBP2, FaBP3 and Fh15 (20,21), and two redox enzymes, thioredoxin and peroxiredoxin (20,46-48). Primary sequence analysis using SignalP (49) confirmed that these proteins lacked predicted N-terminal signal peptides for secretion *via* the classical ER/Golgi pathway and, hence, they must be released by alternate mechanisms. Morphey *et al.* (21) suggested that release of *Fasciola* FaBPs was due to shedding of the tegument as part of a stress response during *in vitro* culture of parasites since they were not detected *in vivo*. However, morphological studies have shown that blebbing or shedding of the *F. hepatica* surface tegument and renewal by proteins synthesised by highly metabolic sub-tegumental cells is a continuous property that may be an immuno-defensive strategy (4,50). This non-classical mechanism for protein secretion may be compared to those reported for some leaderless eukaryotic proteins, for example the ER/Golgi-independent secretion of leaderless peptides including interleukin (IL)-1 α , IL-1 β and fibroblast growth factor (FGF)-2 (51). In the case of IL-1 β , which is processed and activated by caspase-1, it has been proposed that both molecules are packaged together into plasma membrane blebs (by ATP-binding cassette (ABC) transporters) and are rapidly released as microvesicles following phospholipase-mediated fusion with the plasma membrane (52-54). The identification of a phospholipase in the transcriptome of adult flukes (Table 2) together with the presence of ABC transporters within the tegument (55) supports the possibility of a similar mechanism for ER/Golgi-independent secretion pathway in *Fasciola*.

In summary, our EST2Secretome pipeline was successful in identifying the major secreted proteins of adult *F. hepatica*. Integration of this analysis with proteomic data is important for the study of helminth host-pathogen relationships to distinguish proteins that are secreted extra-corporeally from those secreted within the

internal tissues of the parasites. Additionally, this integrated approach identified major helminth secreted proteins that may reach the exterior by novel or non-classical secretory pathways.

Proteomic analysis of the dormant and infective stage larvae

Having collated the secretome data for adult *F. hepatica* we performed a comparative analysis of this with the infective larvae that invade their host by penetrating the intestinal wall. The larvae of *F. hepatica* (<1 mm in length) are released by the intermediate snail host (*Galba truncatula*) and encyst on vegetation that is consumed by the host. For research purposes these are produced in an aquarium and the larvae are allowed to encyst on cellophane. However, because under these conditions the encysted metacercariae can become contaminated with extraneous material we used a method of washing in 2 % hypochlorite to remove this and the outer cyst wall so that a proteomic study of the dormant larvae could be performed. We also used a medium containing bile duct surfactants to activate the larvae and induce them to excyst so that their secretome could be analysed. We found that the level of *Fasciola* excystment was significantly reduced when the excystment medium excluded the reducing agent L-cysteine (91 % reduction, $p = 0.0008$, not shown) compared with medium containing L-cysteine. The requirement for reducing conditions implied a role for cysteine proteases in the excystment process since the activity of these thiol-dependent proteases are enhanced in the presence of reducing agents. We therefore proved this point by showing that the level of excystment was significantly reduced when the broad-range cysteine protease inhibitors E-64 (98 % reduction, $p = 0.0433$) or the cathepsin-specific inhibitor Z-Phe-Ala-CHN₂ (99 % reduction, $p = 0.0205$) were added to the excystment medium. Proteomic analysis of soluble extracts of

dormant larvae and the secretions of the newly excysted juveniles (NEJ) identified 26 and 29 proteins, respectively (Figures 2B-3B; Supplementary tables 2-3). Significantly, 23 % (6) of all the somatic larval proteins and 31 % of the secreted proteins were cysteine proteases consisting of cathepsin L (37 %), cathepsin Bs (45 %) and asparaginyl endopeptidases (18 %). Since the former two classes are potentially inhibited by E-64 and Z-Phe-Ala-CHN₂, but the later class not, the results implicate cathepsin Ls and/or cathepsin B in the process of cyst rupture.

Phylogenetic studies showed that cDNA clones generated from transcripts of *Fasciola* larvae, designated clade FhCL3, encoded a cathepsin L protease that was specific to this infective stage (8,35); the cDNA encoding this protease, FhCL3_n122, was originally described by Harmsen *et al.* (56). Earlier studies by Tkalcovic *et al.* (34) using N-terminal sequencing to identify proteins expressed by larvae reported a sequence of a cathepsin L cysteine protease which we now know matches exactly with the predicted N-terminal of FhCL3_n122 (56). More recently, Cancela *et al.* (28) confirmed the restricted expression of FhCL3, and another protease FhCL4, to infective larvae by isolating cDNAs encoding these enzymes from a *F. hepatica* larvae-specific cDNA library and, subsequently, by employing RT-PCR expression analysis. No FhCL4 peptides were identified in the present analysis suggesting that this enzyme may be expressed at low levels, performs specific intracellular functions and is not secreted by *F. hepatica*. However, we confirmed the presence of FhCL3 in NEJ by identifying several FhCL3-specific peptides with matches to FhCL3_n122. It is noteworthy that the mass of the dormant larvae somatic cathepsin L3, which was found in a protein band migrating at ~37 kDa, is consistent with that of an inactive cathepsin L precursor, or zymogen, which is supported by the presence of mass ion m/z 644.28; +2 that matched to a peptide (SNDVSWHEWK) found only within the N-

terminal prosegment region of the protease. No peptides matching with cathepsin Ls were identified in protein bands in the region corresponding to the fully processed and active mature enzymes i.e. mass ~24 kDa (Figure 1, NEJ secretome gel section 3). By contrast, while a few cathepsin L3 peptides were present at ~37 kDa in the secretome of NEJ (including the FhCL3 prosegment peptides LGLNQFTDLTFEEFK, m/z 901.49; +2, and SNDVSWHEWK, m/z 644.28; +2), the most robust peptide matches were obtained from gel sections at the molecular mass of ~24 kDa corresponding to the size of a fully activated mature enzyme (Figure 1, NEJ secretome gel section 3). Collectively, these observations show that the dormant larvae of *F. hepatica* express a specific cathepsin L protease, FhCL3, stored as an inactive zymogen that is rapidly secreted and activated following emergence from their cysts to become infective larvae.

Fasciola cathepsin B-like cysteine proteases were also identified in somatic extracts of larvae, with peptides matching to proteins encoded by three different cDNAs that were previously designated cathepsin B1 (accession number A7UNB2, Ljunggren *et al.* unpublished), cathepsin B2 (Q8I7B2, Khaznadji *et al.* unpublished) and cathepsin B3 (A5X494, 28). All three cathepsin B proteases were identified in intense protein bands that migrated in reducing SDS-PAGE at molecular mass of ~36 – 50 kDa and ~50 – 70 kDa (Figure 1, dormant larvae gel sections 1 and 2) suggesting that like FhCL3 these may represent stored zymogens. The presence of peptides FINIEHFK (m/z 524.28, +2) and QHLGLLEETPEER (m/z 775.89, +2) confirmed the existence of cathepsin B2 as an unprocessed zymogen whereas peptides QNLGVLEETPEDR (m/z 750.36, +2) and YSVSENDLPESFDAR (m/z 864.90, +2; which spans the juncture between the prosegment and the mature domain) indicate that cathepsins B1 and B3 are also stored as zymogens by *Fasciola* NEJs.

However, unlike the FhCL3, we also found some cathepsin B at ~20 – 36 kDa (Figure 1, NEJ secretome gel section 3) which is consistent with the occurrence of processed active enzymes within the dormant larvae (which have a predicted molecular mass of 29.6 kDa, 57). While each of the three cathepsin B were also found as fully active enzymes in the secretions of larvae, in the region of ~20 – 36 kDa, we also discovered a novel family member, encoded by a *Fasciola* EST HAN4015b05.q1kT3 (here designated cathepsin B4) (Figure 2B; Table 3).

Cysteine proteases are produced as inactive zymogens consisting of a prosegment and mature domain (58). The prosegment lies along the active site groove (in reverse to the direction of protein substrates) preventing unwanted hydrolysis during trafficking and storage of the protease within the cell. Removal of the prosegment from the mature domain exposes the active site of the hydrolase to entry of macromolecular protein substrates. Dalton *et al.* (59) proposed that this activation step is mediated by proteolytic clipping at a ‘protease-susceptible’ region between the prosegment and mature enzyme, and that this event may be performed by the same or another protease molecule. Dalton and Brindley (60) and Dalton *et al.* (59) proposed that in *F. hepatica* and *S. mansoni* an asparaginyl endopeptidase (otherwise known as legumain), which cleaves peptide bonds C-terminal to Asn residues, are involved in the *trans*-processing and activation of cathepsin L and B cysteine proteases. In support of this argument all members of these classes of cysteine protease in both helminth parasites possess an Asn residue in the vicinity of the prosegment/mature juncture (8,60). Furthermore, Sajid *et al.* (62) and Beckham *et al.* (57) provided experimental evidence by showing that a recombinant asparaginyl endopeptidase could *trans*-process and activate recombinant cathepsin B of *F. hepatica* and *S. mansoni*. In the present study, we identified two asparaginyl endopeptidases in the

somatic proteins and secretome of *Fasciola* NEJs. One of the identified enzymes (legumain-like precursor TrEMBL accession number Q711M2, termed legumain 1) matched exactly with an N-terminal sequence for asparaginyl endopeptidase obtained from NEJ somatic proteins by Tkalcevic *et al* (34). Although this enzyme retains the His₁₅₈ of the conserved catalytic dyad residues, the Cys₂₀₀ is replaced by a serine residue and, therefore, this enzyme may display an altered substrate-specificity compared to the normal asparaginyl endopeptidases. However, the enzyme does contain a conserved asparagine residue (Asn₃₂₃) that is required for C-terminal processing to an active mature enzyme (63). The second legumain-like protease identified in the NEJ somatic protein extracts is encoded by a *Fasciola* EST (identifier Fhep29h09.q1k, termed legumain 2) but as this EST is incomplete at the 5' end, it is not possible to determine whether the enzyme encoded by this sequence retains the conserved active site His₁₅₈/Cys₂₀₀ dyad. Interestingly, the enzyme encoded by the legumain 2 EST lacks the conserved Asn₃₂₃ required for auto-activation. The overlapping primary sequence of the two enzymes exhibit 51 % identity and, therefore, individual enzymes were easily differentiated on the basis of their tryptic peptides. Notwithstanding the anomalies within the sequence of these two enzymes, the data suggest that the NEJ possess the enzymatic machinery to rapidly process and activate the stored zymogens of cathepsin L and B. Recently Morita *et al.* (64) reported that purified bovine asparaginyl endopeptidase can degrade fibronectin, a major component of the extracellular matrix. If they possess similar biochemical properties, a second tangible role for the asparaginyl endopeptidases secreted by the infective larvae could be to facilitate penetration of the host intestine.

Proteomic analysis of the migrating liver-stage parasite

The developmental stage of *F. hepatica* that migrates through the liver tissue is responsible for the clinical manifestations associated with acute animal and human fasciolosis. The migrating parasite causes extensive physical tissue destruction, tunnelling and haemorrhaging, and induces immunologically-related inflammatory damage (65). To gain insight on the variety of molecules produced by this parasite stage that may induce this pathology 21 day-old immature parasites were removed from the liver of infected mice and maintained in culture. A proteomics analysis of the medium revealed that these parasites secreted a greater range of proteins than the infective and adult stage flukes. In particular, they secreted the greatest range of cathepsin L cysteine proteases, including the FhCL3 that we identified in the NEJ (this was secreted solely as a mature active form of ~24 kDa). Members of cathepsin L clades FhCL1, FhCL2 and FhCL5 originally identified in the adult parasite were also secreted by the immature flukes (these could be distinguished on the basis of the clade-/sub-clade-specific peptide matches described above and reported previously by Robinson *et al.* (8)).

Fasciola cathepsins B1 and B2 were also secreted by 21 day-old immature flukes, while cathepsin B3 was not detected. However, mass spectrometry data matched with two novel putative cathepsin B enzymes that were discovered in the EST2Secretome of adult ESTs. The enzyme encoded by EST Fhep45b05.q1k (designated cathepsin B5) was identified based on matches with the sequence-specific peptides NIMYEIMK (m/z 521.25; +2), LLGWGVEDGEK (m/z 601.79; +2) and FYAISSYNVYGGEK (m/z 799.36; +2) while the other, EST HAN4006g01.q1kT3 (designated cathepsin B6), was differentiated due to specific matches with the peptides FSTPK (m/z 579.29; +1), HTTGALLGGHAIR (m/z 652.35; +2) and TSYNLLHNEETIMK (m/z 846.90; +2).

Three asparaginyl endopeptidases were identified within the secretory proteins of the immature 21 day-old flukes. These included the NEJ legumain-like precursor, legumain 1, with the altered Cys₂₀₀/Ser₂₀₀ active site residue, but did not include the NEJ legumain 2. However, two asparaginyl endopeptidases (designated legumains 3 and 4) appeared to be specific to the liver migrating stage and, like their homologues that were originally identified in *F. gigantica* (41), retained the conserved catalytic His₁₅₈/Cys₂₀₀ dyad and Asn₃₂₃ that is required for C-terminal processing (63). Unlike the asparaginyl endopeptidases of NEJs, those secreted by the immature flukes all migrated at ~36 kDa suggesting that they had undergone some processing events to lead to fully active enzymes. It is possible that besides functioning in the *trans*-processing of cathepsin L and B, the asparaginyl endopeptidases of the immature parasites participate in host tissue degradation.

Mass spectrometry data from the immature 21 day-old *F. hepatica* (and subsequently found in adult parasite secretome) matched with several ESTs encoding a putative prolylcarboxypeptidase (Figure 2B; Table 3). The exopeptidase contains a predicted N-terminal signal peptide and a conserved catalytic triad (Ser₁₆₁, Asp₄₁₃, His₄₃₈) that is characteristic of the s28 family of serine peptidases including prolylcarboxypeptidase. Although this is the first report of a putative prolylcarboxypeptidase from a trematode pathogen, orthologues have been identified in parasitic nematodes including the gastrointestinal worm *Haemonchus contortus* where the enzymes act as anti-coagulants to facilitate a blood-feeding lifestyle (66). It is likely that the *Fasciola* prolylcarboxypeptidase performs a similar role, acting alongside the saposin-like proteins to prevent blood coagulation to ensure effective lysis of ingested host red blood cells. A role for the putative prolylcarboxypeptidase in

Fasciola nutrition is supported by the absence of the enzyme in the NEJ which does not possess a functional gut.

Relative expression levels of Fasciola proteases and relationship to virulence and tissue invasion.

For *F. hepatica* NEJs, the emPAI data indicated that cathepsin L3 and cathepsin B enzymes were expressed at similar high levels (accounting for 37 % and 45 % of total proteases, respectively) whereas asparaginyl endopeptidases represented 18 % of total proteases detected. By using inhibitors of cathepsin-like proteases we have shown that excystment of the infective larvae are dependant on the cathepsin L3 and cathepsin B proteases. Asparaginyl endopeptidases are likely to be essential for the rapid *trans*-processing and activation of the stored zymogens of cathepsin L and B and, hence, it is probable that one of the first steps in activation of the dormant larvae is the switching on of the genes encoding these enzymes. A role for cathepsin L and B proteases in the penetration of the host by NEJ was recently demonstrated using RNA interference methods by McGonigle *et al.* (67). Knockdown of both cathepsin L and cathepsin B transcripts in NEJ parasites blocked the ability of the larvae to penetrate and traverse the intestinal wall of a rodent host. Therefore, the two types of cysteine proteases, working together, must be responsible for degrading the intestinal tissue macromolecules to enable the rapid penetration of the host intestine. Cysteine proteases are also essential for infectivity of the closely-related schistosome parasites (*Schistosoma mansoni* and *Trichobilharzia regentii*) which enter their hosts *via* the skin (68), and for the cyst emergence and infectivity of larvae of the trematode *Paragonimus westermani* (69-71).

Correlating with the parasites' migration into the liver tissue, and the development of the gut apparatus (beginning at ~10 days, 36) the expression pattern of proteases becomes more complex. For the migrating parasite, the secretion of FhCL3 and cathepsin B becomes less important as these proteases account for only 3 % and 2 % of total proteases, respectively. In contrast, members of the other cathepsin L clades become more highly represented in the secretome, FhCL1 (50 %), FhCL2 (27 %) and FhCL5 (7 %) (Figure 2A). This striking shift in protease expression indicates a requirement for a new set of proteases for liver migration and for the degradation of tissue and blood macromolecules. As discussed above FhCL1 proteases have a particular adaptation to haemoglobin digestion while FhCL2 proteases exhibit unique collagenase-like activities (FhCL5 appears to be more closely related to FhCL1, 72). Asparaginyl endopeptidase secretion remains relatively high, accounting for 11 % of total proteases secreted by the immature flukes, which may be required for activation the cathepsin L proteases and possibly for directly engaging in tissue degradation. The expression of the prolylcarboxypeptidase and saposin at this stage signals the beginning of blood feeding. Working together this profile of molecules is an effective tissue-destroying machinery essential to the parasites migration and growth but, at the same time, responsible for the pathogenesis associated with acute animal and human fasciolosis.

After the parasite has entered the bile duct it completes its maturation and becomes an obligate blood feeder. Blood provides all the necessary nutrients (macro- and micro-molecules) needed for production of enormous numbers of eggs, the principle function of the adult parasite (37). The total protease secretion is assumed completely by the various cathepsin L proteases, clade FhCL1 (69 %), FhCL2 (22 %) and FhCL5 (9 %), which can degrade macromolecules such a haemoglobin to small

peptides which are absorbed by the parasite for further digestion to free amino acids within its tissues (36,37). The emPAI values determined for adult parasite proteases are similar to expression data obtained using by 2-DE and densitometry (67 %, 27 % and 5 % respectively, 8) and validates the use of emPAI for estimating *Fasciola* secretory protein levels in the other developmental stages. Within the bile duct the parasite can exist for years (cattle) or even decades (sheep) causing relatively little pathology, apart from anaemia and bile duct hyperplasia.

F. hepatica secretes a battery of antioxidant molecules

It is worth highlighting that for all three developmental stages of *F. hepatica* investigated in this study an array of antioxidant molecules were secreted in an abundance second only to the proteases. These included peroxiredoxin, thioredoxin, protein disulphide isomerase, four fatty acid-binding proteins (FaBP 1-3 and Fh15) and five glutathione *S*-transferases (GST sigma and mu class) (20,21,32,46-48,73). These molecules have been implicated in fluke immune avoidance mechanisms; for example, secreted isoforms of *Fasciola* GSTs were shown to decrease the proliferative response of rat spleen cells and diminished nitric oxide production by macrophages (73) whereas peroxiredoxin plays a key role driving host Th2 immune responses via the recruitment and alternative activation of peritoneal macrophages (74,75). *Fasciola* peroxiredoxin, thioredoxin and protein disulphide isomerase may also protect the parasites from harmful reactive oxygen species released by host immune cells (32,47,48,76).

Analysis of the emPAI values for these various anti-oxidants indicates that, like the proteases, their secretion is highly regulated during the migration of the parasite (Figure 3A). The profile of secretory antioxidants is similar for the NEJs and

adult flukes; both secrete several FaBPs together with peroxiredoxin (although thioredoxin is also secreted by the adult worms). However, the diversity of the antioxidant molecules secreted by the immature liver-stage parasites is notably different to these two stages with the dramatic production of glutathione-S-transferases (GST), including sigma class GST and four isoforms of mu class GSTs (GST1, GST7, GST47 and GST 51); together these account for 50 % of total antioxidants expressed at this stage. GST may provide a particular defence for the fluke against a mounting host cellular immune response as the immature parasite is in direct contact with the immune systems as it migrates through the liver parenchyma. At this stage also the parasite has moved from an aerobic to an anaerobic environment and therefore significant changes take place in its metabolism, particularly in the tegument which is the primary host-parasite interface (77). Clearly, the combined action of FaBPs, peroxiredoxin and thioredoxin offer sufficient protection for the adult flukes residing within the immunologically-privileged bile ducts.

It was also noted that each of the aforementioned antioxidant classes found in the secretome of *F. hepatica* do not possess a signal sequence for secretion (with the exception of protein disulphide isomerase that also has a putative ER-retention signal and may act as a chaperone to ensure proper folding of classically secreted *Fasciola* peptides). This suggests that they are secreted *via* non-classical mechanisms, possibly shedding or blebbing of the tegument (see above). Accordingly, we propose that two distinct mechanisms for protein export operate for the most abundant proteins in *Fasciola*: (a) a classical secretion of proteases and other molecules via the ER/Golgi from specialised gastrodermal cells of the gut as previously reported (78), and (b) a non-classical secretion of leaderless *Fasciola* antioxidants *via* a trans-tegumental route. This has clear implications for the development of novel anti-*F. hepatica*

chemotherapy and could explain the additive effects of two of the most potent anti-flukicides triclabendazole and clorsulon which act upon the tegument and gastrodermal cells, respectively (79-81).

Conclusion

Our integrated transcriptomics and proteomics analysis has provided an in-depth overview of protein secretion by the animal and human pathogen, *F. hepatica* that represents a significant step towards a comprehensive understanding of the host-parasite interactions in fasciolosis. Of vital importance now is the completion of transcriptome datasets for both the infective NEJ larvae and immature liver-stage parasites so that a comparative analysis of secretory proteins can be performed. Nevertheless, we have described the secretion of the major molecules of this parasite, and correlated their expression with critical steps in its migration and development within the mammalian host. This will allow a more strategic and rational approach to future anthelmintic- or vaccine-development programs. It is interesting that several of the major components of *Fasciola* secretions identified here are currently leading candidates for development as first generation anti-fluke vaccines (including cathepsins, peroxiredoxin, glutathione *S*-transferase and fatty acid-binding proteins; 82) or are promising targets for novel flukicidal drugs (cathepsins; 83) demonstrating the value of our integrated approach for the future identification of new targets for therapeutic intervention.

Acknowledgements

MWR is supported by a UTS Chancellor's Postdoctoral Fellowship. SMD and JPD were supported by the National Health and Medical Research Council of Australia

Project, Grant 352912. JPD is a recipient of a NSW Government BioFirst Award 2004. The bioinformatics analysis was partially supported by a grant from the Australian Research Council (ARC) (LP0667795). We thank Matt Padula, Proteomics Technology Centre of Expertise (UTS), for assistance with the mass spectrometry and Matt Berriman of the Wellcome Trust Sanger Centre.

References

1. Garcia, H.H., Moro, P.L. and Schantz, P.M. (2007) Zoonotic helminth infections of humans: echinococcosis, cysticercosis and fascioliasis. *Curr. Opin. Infect. Dis.* **20**, 489-494.
2. Andrews, S.J. (1999) The life-cycle of *Fasciola hepatica*. In *Fasciolosis* (Dalton, J.P., ed) pp 1-29, CABI, Oxford, United Kingdom.
3. Mas-Coma, S., Bargues, M.D. and Esteban, J.G. (1999) Human Fasciolosis. In *Fasciolosis* (Dalton, J.P., ed) pp 411-434 CABI, Oxford, United Kingdom.
4. Fairweather, I., Threadgold, L. T. and Hanna, R. E. B. (1999). Development of *Fasciola hepatica* in the mammalian host. In *Fasciolosis* (Dalton, J.P., ed), pp 47-111, CABI, Oxford, United Kingdom.
5. Hewitson, J.P., Harcus, Y.M., Curwen, R.S., Dowle, A.A., Atmadja, A.K., Ashton, P.D., Wilson, A. and Maizels, R.M. (2008) The secretome of the filarial parasite, *Brugia malayi*: proteomic profile of adult excretory-secretory products. *Mol. Biochem. Parasitol.* **160**, 8-21.
6. Robinson, M.W. and Connolly, B. (2005) Proteomic analysis of the excretory-secretory proteins of the *Trichinella spiralis* L1 larva, a nematode parasite of skeletal muscle. *Proteomics* **5**, 4525-4532.
7. Robinson, M.W., Greig, R., Beattie, K., Lamont, D. and Connolly, B. (2007) Comparative analysis of the excretory-secretory proteome of the muscle larva of *Trichinella pseudospiralis* and *Trichinella spiralis*. *Int. J. Parasitol.* **37**, 139-148.
8. Robinson, M.W., Tort, J.F., Wong, E., Donnelly, S.M., Lowther, J., Xu, W., Stack, C.M., Padula, M., Herbert, B. and Dalton, J.P. (2008) Proteomic and phylogenetic analysis of the cathepsin L protease family of the helminth pathogen, *Fasciola hepatica*: expansion of a repertoire of virulence-associated factors. *Mol. Cell. Proteomics* **7**, 1111-1123.
9. Knudsen, G.M., Medzihradsky, K.F., Lim, K.C., Hansell, E. and McKerrow, J.H. (2005) Proteomic analysis of *Schistosoma mansoni* cercarial secretions. *Mol. Cell. Proteomics* **4**, 1862-1875.
10. Ojopi, E.P., Oliveira, P.S., Nunes, D.N., Paquola, A., DeMarco, R., Gregório, S.P., Aires, K.A., Menck, C.F., Leite, L.C., Verjovski-Almeida, S. and Dias-Neto, E. (2007) A quantitative view of the transcriptome of *Schistosoma mansoni* adult-worms using SAGE. *BMC Genomics* **8**, 186.
11. Oliveira, G. (2007) The *Schistosoma mansoni* transcriptome: an update. *Exp. Parasitol.* **117**, 229-235.

12. Laha, T., Pinlaor, P., Mulvenna, J., Sripa, B., Sripa, M., Smout, M.J., Gasser, R.B., Brindley, P.J. and Loukas, A. (2007) Gene discovery for the carcinogenic human liver fluke, *Opisthorchis viverrini*. *BMC Genomics* **8**, 189.
13. Nagaraj, S.H., Gasser, R.B. and Ranganathan, S. (2008) Needles in the EST Haystack: Large-Scale Identification and Analysis of Excretory-Secretory (ES) Proteins in Parasitic Nematodes Using Expressed Sequence Tags (ESTs). *PLoS. Negl. Trop. Dis.* **2**, e301.
14. Nagaraj, S.H., Deshpande, N., Gasser, R.B. and Ranganathan, S. (2007) ESTExplorer: an expressed sequence tag (EST) assembly and annotation platform. *Nucleic Acids Res.* **35**, W143-147.
15. Nagaraj, S.H., Gasser, R.B., Nisbet, A.J. and Ranganathan, S. (2008) *In silico* analysis of expressed sequence tags from *Trichostrongylus vitrinus* (Nematoda): comparison of the automated ESTExplorer workflow platform with conventional database searches. *BMC Bioinformatics* **9 Suppl 1**:S10.
16. Ranganathan, S., Nagaraj, S.H., Hu, M., Strube, C., Schnieder, T. and Gasser, R.B. (2007) A transcriptomic analysis of the adult stage of the bovine lungworm, *Dictyocaulus viviparus*. *BMC Genomics.* **8**, 311.
17. Datu, B.J., Gasser, R.B., Nagaraj, S.H., Ong, E.K., O'Donoghue, P., McInnes, R., Ranganathan, S. and Loukas, A. (2008) Transcriptional Changes in the Hookworm, *Ancylostoma caninum*, during the Transition from a Free-Living to a Parasitic Larva. *PLoS Negl Trop Dis.* **2**, e130.
18. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.
19. Zdobnov, E.M. and Apweiler, R. (2001) InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847-848.
20. Jefferies, J.R., Campbell, A.M., van Rossum, A.J., Barrett, J. and Brophy, P.M. (2001) Proteomic analysis of *Fasciola hepatica* excretory-secretory products. *Proteomics* **1**, 1128-1132.
21. Morphew, R.M., Wright, H.A., LaCourse, E.J., Woods, D.J. and Brophy, P.M. (2007) Comparative proteomics of excretory-secretory proteins released by the liver fluke *Fasciola hepatica* in sheep host bile and during *in vitro* culture ex host. *Mol. Cell. Proteomics* **6**, 963-972.
22. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J. and Mann, M. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics.* **4**, 1265-1272.
23. El-Sayed, N.M., Bartholomeu, D., Ivens, A., Johnston, D.A. and LoVerde, P.T. (2004) Advances in schistosome genomics. *Trends Parasitol.* **20**, 154-157.
24. Mitreva, M., Zarlenga, D.S., McCarter, J.P. and Jasmer, D.P. (2007) Parasitic nematodes - from genomes to control. *Vet. Parasitol.* **148**, 31-42.
25. Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., Coulson, A., D'Eustachio, P., Fitch, D.H., Fulton, L.A., Fulton, R.E., Griffiths-Jones, S., Harris, T.W., Hillier, L.W., Kamath, R., Kuwabara, P.E., Mardis, E.R., Marra, M.A., Miner, T.L., Minx, P., Mullikin, J.C., Plumb, R.W., Rogers, J., Schein, J.E., Sohrmann, M., Spieth, J., Stajich, J.E., Wei, C., Willey, D., Wilson, R.K., Durbin, R. and Waterston, R.H. (2003) The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol.* **1**, E45.

26. Chen, Y., Zhang, Y., Yin, Y., Gao, G., Li, S., Jiang, Y. and Luo, J. (2005) SPD—a web-based secreted protein database. *Nucleic Acids Res.* **33**, D169–D173.
27. Choo, K.H., Tan, T.W. and Ranganathan, S. (2005) SPdb—a signal peptide database. *BMC Bioinformatics* **6**, 249.
28. Cancela, M., Acosta, D., Rinaldi, G., Silva, E., Durán, R., Roche, L., Zaha, A., Carmona, C. and Tort, J.F. (2008) A distinctive repertoire of cathepsins is expressed by juvenile invasive *Fasciola hepatica*. *Biochimie* **90**, 1461-1475.
29. Grams, R., Adisakwattana, P., Ritthisunthorn, N., Eursitthichai, V., Vichasri-Grams, S. and Viyanant, V. (2006) The saposin-like proteins 1, 2, and 3 of *Fasciola gigantica*. *Mol. Biochem. Parasitol.* **148**, 133-143.
30. Espino, A.M. and Hillyer, G.V. (2003) Molecular cloning of a member of the *Fasciola hepatica* saposin-like protein family. *J. Parasitol.* **89**, 545-552.
31. Robinson, M.W., Colhoun, L.M., Fairweather, I., Brennan, G.P. and Waite, J.H. (2001) Development of the vitellaria of the liver fluke, *Fasciola hepatica* in the rat host. *Parasitology* **123**, 509-518.
32. Salazar-Calderón, M., Martín-Alonso, J.M., Castro, A.M. and Parra, F. (2003) Cloning, heterologous expression in *Escherichia coli* and characterization of a protein disulfide isomerase from *Fasciola hepatica*. *Mol. Biochem. Parasitol.* **126**, 15-23.
33. Bork, P. and Beckmann, G. (1993) The CUB domain. A widespread module in developmentally regulated proteins. *J. Mol. Biol.* **231**, 539-545.
34. Tkalcovic, J., Ashman, K. and Meeusen, E. (1995) *Fasciola hepatica*: rapid identification of newly excysted juvenile proteins. *Biochem. Biophys. Res. Commun.* **213**, 169-174.
35. Irving, J.A., Spithill, T.W., Pike, R.N., Whisstock, J.C. and Smooker, P.M. (2003) The evolution of enzyme specificity in *Fasciola* spp. *J. Mol. Evol.* **57**, 1-15.
36. Lowther, J., Robinson, M.W., Donnelly, S.M., Xu, W., Stack, C.M., Matthews, J.M. and Dalton, J.P. (2009) The importance of pH in regulating the function of *Fasciola hepatica* cathepsin L1 cysteine protease. *PLoS. Negl. Trop. Dis.* (In Press).
37. Dalton, J.P., Caffrey, C.R., Sajid, M., Stack, C., Donnelly, S., Loukas, A., Don, T., McKerrow, J., Halton, D.W. and Brindley, P.J. (2006) Proteases in Trematode Biology. In *Parasitic Flatworms: Molecular biology, Biochemistry, Immunology and Physiology*, A.G. Maule and N.J. Marks, eds. Pp 348-368. Oxford, United Kingdom: CAB International.
38. Stack, C.M., Caffrey, C.R., Donnelly, S.M., Seshadri, A., Lowther, J., Tort, J.F., Collins, P.R., Robinson, M.W., Xu, W., McKerrow, J.H., Geiger, S., Marion, R., Brinen, L.S. and Dalton, J.P. (2008). Structural and functional relationships in the virulence-associated cathepsin L proteases of the parasitic liver fluke, *Fasciola hepatica*. *J. Biol. Chem.* **283**, 9896-9908.
39. Don, T.A., Oksov, Y., Lustigman, S. and Loukas, A. (2007) Saposin-like proteins from the intestine of the blood-feeding hookworm, *Ancylostoma caninum*. *Parasitology* **134**, 427-436.
40. Law, R.H., Smooker, P.M., Irving, J.A., Piedrafita, D., Ponting, R., Kennedy, N.J., Whisstock, J.C., Pike, R.N. and Spithill, T.W. (2003) Cloning and expression of the major secreted cathepsin B-like protein from juvenile *Fasciola hepatica* and analysis of immunogenicity following liver fluke infection. *Infect Immun.* **71**, 6921-6932.
41. Adisakwattana, P., Viyanant, V., Chaicumpa, W., Vichasri-Grams, S., Hofmann, A., Korge, G., Sobhon, P. and Grams, R. (2007) Comparative molecular analysis of two asparaginyl

- endopeptidases and encoding genes from *Fasciola gigantica*. *Mol. Biochem. Parasitol.* **156**, 102-116.
42. Turk, D. and Guncar, G. (2003) Lysosomal cysteine proteases (cathepsins): promising drug targets. *Acta. Crystallogr. D. Biol. Crystallogr.* **59**, 203-213.
 43. Chen, J.M., Dando, P.M., Stevens, R.A., Fortunato, M. and Barrett, A.J. (1998) Cloning and expression of mouse legumain, a lysosomal endopeptidase. *Biochem J.* **335**, 111-117.
 44. Michel, A., Ghoneim, H., Resto, M., Klinkert, M.Q. and Kunz, W. (1995) Sequence, characterization and localization of a cysteine proteinase cathepsin L in *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* **73**, 7-18.
 45. Meemon, K., Grams, R., Vichasri-Grams, S., Hofmann, A., Korge, G., Viyanant, V., Upatham, E.S., Habe, S. and Sobhon, P. (2004) Molecular cloning and analysis of stage and tissue-specific expression of cathepsin B encoding genes from *Fasciola gigantica*. *Mol. Biochem. Parasitol.* **136**, 1-10.
 46. Gourbal, B.E., Guillou, F., Mitta, G., Sibille, P., Thèron, A., Pointier, J.P. and Coustau, C. (2008) Excretory-secretory products of larval *Fasciola hepatica* investigated using a two-dimensional proteomic approach. *Mol. Biochem. Parasitol.* **161**, 63-66.
 47. Salazar-Calderón, M., Martín-Alonso, J.M., Ruiz de Eguino, A.D. and Parra, F. (2001) Heterologous expression and functional characterization of thioredoxin from *Fasciola hepatica*. *Parasitol Res.* **87**, 390-395.
 48. Sekiya, M., Mulcahy, G., Irwin, J.A., Stack, C.M., Donnelly, S.M., Xu, W., Collins, P. and Dalton, J.P. (2006) Biochemical characterisation of the recombinant peroxiredoxin (FhePrx) of the liver fluke, *Fasciola hepatica*. *FEBS Lett.* **580**, 5016-5022.
 49. Emanuelsson, O., Brunak, S., von Heijne, G. and Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP, and related tools. *Nat. Protoc.* **2**, 953-971.
 50. Dalton, J.P., Skelly, P. and Halton, D.W. (2004) Role of the tegument and gut in nutrient uptake by parasitic plathyhelminths. *Can. J. Zool.* **82**, 211-232.
 51. Keller, M., Rüegg, A., Werner, S. and Beer, H.D. (2008) Active caspase-1 is a regulator of unconventional protein secretion. *Cell* **132**, 818-831.
 52. MacKenzie, A., Wilson, H.L., Kiss-Toth, E., Dower, S.K., North, R.A. and Surprenant, A. (2001) Rapid secretion of interleukin-1beta by microvesicle shedding. *Immunity* **15**, 825-835.
 53. Andrei, C., Margiocco, P., Poggi, A., Lotti, L.V., Torrisi, M.R. and Rubartelli, A. (2004) Phospholipases C and A2 control lysosome-mediated IL-1 beta secretion: Implications for inflammatory processes. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9745-9750.
 54. Mambula, S.S., Stevenson, M.A., Ogawa, K. and Calderwood, S.K. (2007) Mechanisms for Hsp70 secretion: crossing membranes without a leader. *Methods* **43**, 168-175.
 55. Kumkate, S., Chunchob, S. and Janvilisri, T. (2008) Expression of ATP-binding cassette multidrug transporters in the giant liver fluke *Fasciola gigantica* and their possible involvement in the transport of bile salts and anthelmintics. *Mol. Cell. Biochem.* **317**, 77-84.
 56. Harmsen, M.M., Cornelissen, J.B., Buijs, H.E., Boersma, W.J., Jeurissen, S.H. and van Milligen, F.J. (2004) Identification of a novel *Fasciola hepatica* cathepsin L protease containing protective epitopes within the propeptide. *Int. J. Parasitol.* **34**, 675-682.
 57. Beckham, S.A., Law, R.H., Smooker, P.M., Quinsey, N.S., Caffrey, C.R., McKerrow, J.H., Pike, R.N. and Spithill, T.W. (2006) Production and processing of a recombinant *Fasciola*

- hepatica* cathepsin B-like enzyme (FhcatB1) reveals potential processing mechanisms in the parasite. *Biol. Chem.* **387**, 1053-1061.
58. Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J.S. and Cygler, M. (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO. J.* **15**, 5492-5503.
 59. Dalton, J.P., Brindley, P.J., Donnelly, S. and Robinson, M.W. (2009) The enigmatic asparaginyl endopeptidase of helminth parasites. *Trends Parasitol.* (In Press).
 60. Dalton, J.P. and Brindley, P.J. (1996) Schistosome asparaginyl endopeptidase Sm 32 in hemoglobin digestion. *Parasitol. Today* **12**, 125.
 61. Robinson, M.W., Dalton, J.P. & Donnelly, S. (2008) Helminth pathogen cathepsin proteases: it's a family affair. *Trends Biochem. Sci.* **33**, 601-608.
 62. Sajid, M., McKerrow, J.H., Hansell, E., Mathieu, M.A., Lucas, K.D., Hsieh, I., Greenbaum, D., Bogyo, M., Salter, J.P., Lim, K.C., Franklin, C., Kim, J.H. and Caffrey, C.R. (2003) Functional expression and characterization of *Schistosoma mansoni* cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. *Mol. Biochem. Parasitol.* **131**, 65-75.
 63. Chen, J., Fortunato, M. and Barrett, A.J. (2000) Activation of human prolegumain by cleavage at a C-terminal asparagine residue. *Biochem. J.* **352**, 327-334.
 64. Morita, Y., Araki, H., Sugimoto, T., Takeuchi, K., Yamane, T., Maeda, T., Yamamoto, Y., Nishi, K., Asano, M., Shirahama-Noda, K., Nishimura, M., Uzu, T., Hara-Nishimura, I., Koya, D., Kashiwagi, A., Ohkubo, I. (2007) Legumain/asparaginyl endopeptidase controls extracellular matrix remodeling through the degradation of fibronectin in mouse renal proximal tubular cells. *FEBS Lett.* **581**, 1417-1424.
 65. Behm, C.A. and Sangster, N.C. (1999) Pathology, Pathophysiology and Clinical Aspects. In *Fasciolosis* (Dalton, J.P., ed), pp 185-224, CABI, Oxford, United Kingdom.
 66. Geldhof, P. and Knox, D. (2008) The intestinal contortin structure in *Haemonchus contortus*: An immobilised anticoagulant? *Int. J. Parasitol.* **38**, 1579-1588.
 67. McGonigle, L., Mousley, A., Marks, N.J., Brennan, G.P., Dalton, J.P., Spithill, T.W., Day, T.A. and Maule, A.G. (2008) The silencing of cysteine proteases in *Fasciola hepatica* newly excysted juveniles using RNA interference reduces gut penetration. *Int. J. Parasitol.* **38**, 149-155.
 68. Kašný, M., Dalton, J.P., Mikeš, L. and Horák, P. (2007). Comparison of cysteine and serine peptidase activities in *Trichobilharzia regenti* and *Schistosoma mansoni* cercariae. *Parasitology.* **22**, 1-11.
 69. Ikeda, T. (2003) Involvement of cysteine proteinases in excystment of *Paragonimus ohirai* metacercariae induced by sodium cholate and A23187. *J. Helminthol.* **77**, 21-26.
 70. Chung, Y.B., Kim, T.S. and Yang, H.J. (2005) Early cysteine protease activity in excretory bladder triggers metacercaria excystment of *Paragonimus westermani*. *J. Parasitol.* **91**, 953-954.
 71. Na, B.K., Kim, S.H., Lee, E.G., Kim, T.S., Bae, Y.A., Kang, I., Yu, J.R., Sohn, W.M., Cho, S.Y. and Kong, Y. (2006) Critical roles for excretory-secretory cysteine proteases during tissue invasion of *Paragonimus westermani* newly excysted metacercariae. *Cell. Microbiol.* **8**, 1034-1046.
 72. Smooker, P.M., Whisstock, J.C., Irving, J.A., Siyaguna, S., Spithill, T.W. and Pike, R.N. (2000) A single amino acid substitution affects substrate specificity in cysteine proteinases from *Fasciola hepatica*. *Protein Sci.* **9**, 2567-2572.

73. Cervi, L., Rossi, G. and Masih, D.T. (1999) Potential role for excretory-secretory forms of glutathione-S-transferase (GST) in *Fasciola hepatica*. *Parasitology*. **119**, 627-633.
74. Donnelly, S., O'Neill, S.M., Sekiya, M., Mulcahy, G. and Dalton, J.P. (2005) Thioredoxin peroxidase secreted by *Fasciola hepatica* induces the alternative activation of macrophages. *Infect. Immun.* **73**, 166-173.
75. Donnelly S, Stack CM, O'Neill SM, Sayed AA, Williams DL, Dalton JP. (2008) Helminth 2-Cys peroxiredoxin drives Th2 responses through a mechanism involving alternatively activated macrophages. *FASEB J.* **22**, 4022-4032.
76. McGonigle, S., Curley, G.P. and Dalton, J.P. (1997) Cloning of peroxiredoxin, a novel antioxidant enzyme, from the helminth parasite *Fasciola hepatica*. *Parasitology* **115**, 101-104.
77. Tielens, G.M. Metabolism. In *Fasciolosis* (Dalton, J.P., ed) pp 277-306 CABI, Oxford, United Kingdom.
78. Collins, P.R., Stack, C.M., O'Neill, S.M., Doyle, S., Ryan, T., Brennan, G.P., Mousley, A., Stewart, M., Maule, A.G., Dalton, J.P. and Donnelly, S. (2004) Cathepsin L1, the major protease involved in liver fluke (*Fasciola hepatica*) virulence: propeptide cleavage sites and autoactivation of the zymogen secreted from gastrodermal cells. *J. Biol. Chem.* **279**, 17038-17046.
79. Robinson, M.W., Trudgett, A., Hoey, E.M. And Fairweather, I. (2002) Triclabendazole-resistant *Fasciola hepatica*: β -tubulin and response to *in vitro* treatment with triclabendazole. *Parasitology* **124**, 325-338.
80. Meaney, M., Allister, J., McKinsty, B., McLaughlin, K., Brennan, G.P., Forbes, A.B. and Fairweather, I. (2007) *Fasciola hepatica*: ultrastructural effects of a combination of triclabendazole and clorsulon against mature fluke. *Parasitol Res.* **100**, 1091-1104.
81. Meaney, M., Haughey, S., Brennan, G.P. and Fairweather, I. (2005) Ultrastructural observations on oral ingestion and trans-tegumental uptake of clorsulon by the liver fluke, *Fasciola hepatica*. *Parasitol Res.* **95**, 201-212.
82. MacManus, D.P. and Dalton, J.D. (2006) Vaccines against the zoonotic trematodes *Schistosoma japonicum*, *Fasciola hepatica* and *Fasciola gigantica*. *Parasitology* **133**, S43-61.
83. Alcala-canto, Y.A., Ibarra-Velarde, F., Sumano-Lopez, H., Gracia-Mora, J. and Alberti-Navarro, A. (2007) Effect of a cysteine protease inhibitor on *Fasciola hepatica* (liver fluke) fecundity, egg viability, parasite burden, and size in experimentally infected sheep. *Parasitol. Res.* **100**, 461-465.

Figure 1. Analysis of *F. hepatica* somatic and secretory proteins by 1-DE.

Typical 1-D profile of somatic proteins expressed by dormant *F. hepatica* larvae as well as proteins secreted by *F. hepatica* NEJs (NEJ), 21 day-old immature flukes (Immature) and adult parasites (Adult). Proteins (10 µg) were separated using NuPage® Novex® 4-12 % Bis-Tris gels (Invitrogen) and stained with colloidal Coomassie blue G250 prior to analysis by mass spectrometry.

Figure 2. Expression of *F. hepatica* proteases.

A. The expression levels of the major proteolytic enzymes used by *F. hepatica* during its mammalian life-cycle. The label-free exponentially modified protein abundance index (emPAI, 22) derived from *Fasciola* MS/MS CID data was used to determine the abundance of the various cathepsin Ls (FhCL), cathepsin Bs (FhCB) and asparaginyl endopeptidases (FhAE) that are expressed by juvenile flukes (dormant larvae) and secreted by the immature (liver stage) flukes and adult (bile duct stage) parasites. B. Identification of *F. hepatica* proteases by nano-Lc-ESI-MS/MS. Dormant larvae somatic proteins and those secreted by NEJs, 21-day old immature flukes and adult *F. hepatica* were separated by 1-DE and analysed by nano-Lc-ESI-MS/MS. Mass spectrometry data was submitted to either the MSDB sequence database or a custom-made database composed of all *F. hepatica* ESTs (14,031 reads) currently available from the Wellcome Trust Sanger Centre using MASCOT and PEAKS software respectively. Adult *F. hepatica* ESTs that matched with MS/MS data were submitted as queries to BLASTn (18) or conceptually translated and submitted to InterProScan (19) to detect conserved domains and motifs. ¹ MASCOT scores are given for matches to *F. hepatica* cDNAs; ² number of unique peptide matches; ³ raw emPAI values provided by the MASCOT search engine give an approximate quantification of

the protein (22). These values were converted to a percentage of total protease expression within each *Fasciola* developmental stage (see experimental procedures section) and were used to generate the pie charts shown in A; ⁴ section of the 1-D gel from which the peptides were identified; ⁵ PEAKS scores (expressed as a percentage) are given for matches to peptides encoded by *F. hepatica* ESTs; ⁶ single peptide-based identification – see supplementary figure 4.

Figure 3. Expression of *F. hepatica* antioxidants.

A. The expression levels of the major antioxidant molecules used by *F. hepatica* during its mammalian life-cycle. The emPAI values (22) derived from *Fasciola* CID data were used to determine the abundance of the various antioxidants that are expressed and/or secreted by juvenile flukes (dormant larvae and NEJs) and secreted by the immature (liver stage) flukes and adult (bile duct stage) parasites. B. Identification of *F. hepatica* antioxidants by nano-Lc-ESI-MS/MS. Dormant larvae somatic proteins and those secreted by NEJs, 21-day old immature flukes and adult *F. hepatica* were separated by 1-DE and analysed by nano-Lc-ESI-MS/MS. MASCOT searches were performed against the MSDB sequence database. ¹ number of unique peptide matches; ² raw emPAI values provided by the MASCOT search engine give an approximate quantification of the protein (22). These values were converted to a percentage of total antioxidant expression within each *Fasciola* developmental stage (see experimental procedures section) and were used to generate the pie charts shown in A; ³ section of the 1-D gel from which the peptides were identified.

Table 1. Summary analysis of 14031 adult *F. hepatica* ESTs using EST2Secretome.

The contigs and singletons generated by pre-processing, overall representative ESTs (rESTs), peptides from conceptual translation and putative secretory proteins identified are shown.

<i>F. hepatica</i> ESTs	Numbers
Raw sequences obtained	14031
Cleaned sequences	12954 (92.3 %)
Clusters of multiple sequences (contigs)	2749 (19.6 %)
Clusters of singletons	1487 (10.6 %)
Total rESTs	4236 (30.2 %)
Putative peptides	2960 (68.9 % rESTs)
Secreted proteins (SignalP cut-off: 0.5)	160 (5.4 % peptides)

Table 2. Secretory proteins predicted from adult *F. hepatica* rESTs.

Adult *F. hepatica* secretory proteins were predicted using the EST2Secretome pipeline (13) using the default SignalP threshold value of 0.5. Their putative functionality based on BLAST analysis and the presence of InterPro domains is shown. Supporting proteomic data from the three *Fasciola* life-cycle stages and the RNAi phenotypes for their *C. elegans* homologs are also shown. ¹ Identified in adult *F. hepatica* secretions following gel filtration (data not shown). ² New protein family assigned to this protein: PF11703. *NEJ*, newly excysted juvenile; *Imm*, 21 day-old immature fluke; *Ad*, adult fluke; *Ste*, sterile; *Lvl*, larval lethal; *Lva*, larval arrest; *Emb*, embryonic lethal; *Gro*, slow growth; *Unc*, locomotion abnormal; *Daf*, dauer formation abnormal; *Pvl*, protruding vulva; *Dpy*, dumpy; *Slu*, sluggish; *Clr*, clear; *age*, life span abnormal; *Age*, extended life span; *Stp*, sterile progeny; *Rup*, exploded through vulva; *Ric*, Aldicarb resistant; *Cons*, constipated; *Mig*, cell migration abnormal; *Gom*, gonad migration abnormal; *Bmd*, organism morphology abnormal.

Description (Top BLAST hit)	Organism	InterPro domains	rESTs	Proteomic data	RNAi phenotypes of <i>C. elegans</i> homologs
Cathepsin L	<i>F. hepatica</i>	Peptidase C1A	66	NEJ, Imm, Ad	Emb, Gro, Unc
Novel proteins (no significant hits)	-	-	36	-	-
Saposin-like protein 3	<i>F. hepatica</i>	Saposin B	13	Imm, Ad	Emb, Gro, Unc, Ste, Pvl, Lva
Vitelline protein B1	<i>F. hepatica</i>	Trematode eggshell synthesis	9	-	Daf
Cathepsin B endoprotease	<i>F. hepatica</i>	Peptidase C1A	6	NEJ, Imm	-
Legumain	<i>O. viverrini</i>	Peptidase C13, legumain	5	NEJ, Imm	Emb
Cystatin	<i>F. hepatica</i>	Proteinase inhibitor I25, cystatin	3	-	Gro
Unknown proteins	<i>C. sinensis</i>	-	3	NEJ, Ad ¹	Ste, Lvl, Lva, Emb
Uncharacterised proteins ²	<i>S. japonicum</i>	New Pfam domain	2	-	-
Protein disulphide isomerase	<i>F. hepatica</i>	Disulphide isomerase	2	Imm	Gro, Lva, Bmd, Dpy, Emb, Unc, Slu, Clr
SJCHGC01895 protein	<i>S. japonicum</i>	Peptidase S1 and S6	2	-	Emb
Cubulin	<i>C. familiaris</i>	CUB	2	NEJ	-
Unnamed protein	<i>T. nigroviridis</i>	Deoxyribonuclease I	1	-	Emb, Lva
SJCHGC00967 protein	<i>S. japonicum</i>	Deoxyribonuclease II	1	-	age
Apoferitin-2	<i>S. japonicum</i>	Ferritin	1	Imm	bar-1(ga80)
Hypothetical protein	<i>A. aegypti</i>	Phospholipase D	1	-	-
SJCHGC06223 protein	<i>S. japonicum</i>	Peptidase S10, serine carboxypeptidase	1	-	-
Cyclophilin B	<i>X. tropicalis</i>	Peptidyl-prolyl cis-trans isomerase	1	NEJ, Ad	-
Unnamed protein	<i>K. lactis</i>	-	1	-	-
Peptidoglycan recognition protein	<i>A. irradians</i>	Amidase 2	1	-	Stp
Myoglobin 2	<i>P. westermani</i>	Globin, globin-like	1	-	Cons, Gom, Emb, Rup, Bmd, Mig, Gro
SJCHGC09717 protein	<i>S. japonicum</i>	-	1	-	-
Gag-pol polyprotein	<i>S. purpuratus</i>	-	1	-	Age, Ric

Table 3. Database of the repertoire of proteases expressed by *F. hepatica* in the mammalian host. Somatic proteases expressed by dormant larvae and those secreted by NEJs, immature flukes and adult *F. hepatica* identified by mass spectrometry. ¹ expression data only available at the transcript level.

Accession	<i>Fasciola</i> protein	Dormant larvae	NEJ secretome	Immature secretome	Adult secretome
FhCL1A					
Q8T5Z9	Cathepsin L	-	-	+	+
Q7JNQ9	Cathepsin L1	-	-	+	+
Q6R018	Cathepsin L protein	-	-	+	+
Q2HPD3	Cathepsin L1 proteinase	-	-	+	+
Q24940	Cathepsin L-like proteinase	-	-	+	+
FhCL1B					
Q9GRW5	Cathepsin L1	-	-	+	+
FgCL1C					
Q8MUT6	Cathepsin L	-	-	-	-
FhCL2					
Q7JNQ8	Secreted cathepsin L2	-	-	+	+
A5Z1V3	Secreted cathepsin L2	-	-	+	+
A5X483	Cathepsin L2	-	-	+	+
A3FMG6	Cathepsin L	-	-	+	-
FhCL3					
Q95VA7	Cathepsin L	+	+	+	-
A8W638	Cathepsin L	+	+	+	-
A8W7J0	Metacercarial procathepsin L	-	+	-	-
Q9GRW4	Partial procathepsin L3	-	+	-	-
Q9GRW6	Partial procathepsin L3	-	+	-	-
B3TM67	Cathepsin L3	-	+	-	-
B3TM68	Cathepsin L3	-	+	-	-
FhCL4					
Fhep55b05.q1k	Putative cathepsin L	-	-	-	-
FhCL5					
Q9NGW3	Cathepsin L	-	-	+	+
Q9NB30	Cathepsin L	-	-	+	+
Cathepsin B					
A7UNB2	Cathepsin B1	+	-	+	-
Q8I7B2	Cathepsin B2	+	-	+	-
A5X494	Cathepsin B3	+	-	-	-
HAN4015b05.q1kT3	Putative cathepsin B4	-	+	-	-
Fhep45b05.q1k	Putative cathepsin B5	-	-	+	-
HAN4006g01.q1kT3	Putative cathepsin B6	-	-	+	+ ¹
Fhep44e10.q1k	Putative cathepsin B7	-	-	-	+ ¹
Fhep11b02.q1k	Putative cathepsin B8	-	-	-	+ ¹
FhContig1639	Putative cathepsin B9	-	-	-	+ ¹
FhContig2164	Putative cathepsin B10	-	-	-	+ ¹
Asparaginyl endopeptidases					
Q711M2	Legumain-like precursor 1	+	+	+	-
Fhep29h09.q1k	Putative legumain 2	+	+	-	-
A6Y9U8	Legumain 3	-	-	+	+ ¹
A6Y9U9	Legumain 4	-	-	+	+ ¹
Fhep21f02.q1k	Putative legumain 5	-	-	-	+ ¹
FhContig1272	Putative legumain 6	-	-	-	+ ¹
FhContig2292	Putative legumain 7	-	-	-	+ ¹
Prolylcarboxypeptidase (s28)					
Fhep30b01.q1k	Putative prolylcarboxypeptidase	-	-	+	+
Serine carboxypeptidase (s10)					
FhContig542	Putative serine carboxypeptidase	-	-	-	+ ¹
Trypsin-like serine proteases					
FhContig492	Putative peptidase S1/S6	-	-	-	+ ¹
FhContig2453	Putative trypsin-like Ser/Cys	-	-	-	+ ¹

Figure 1

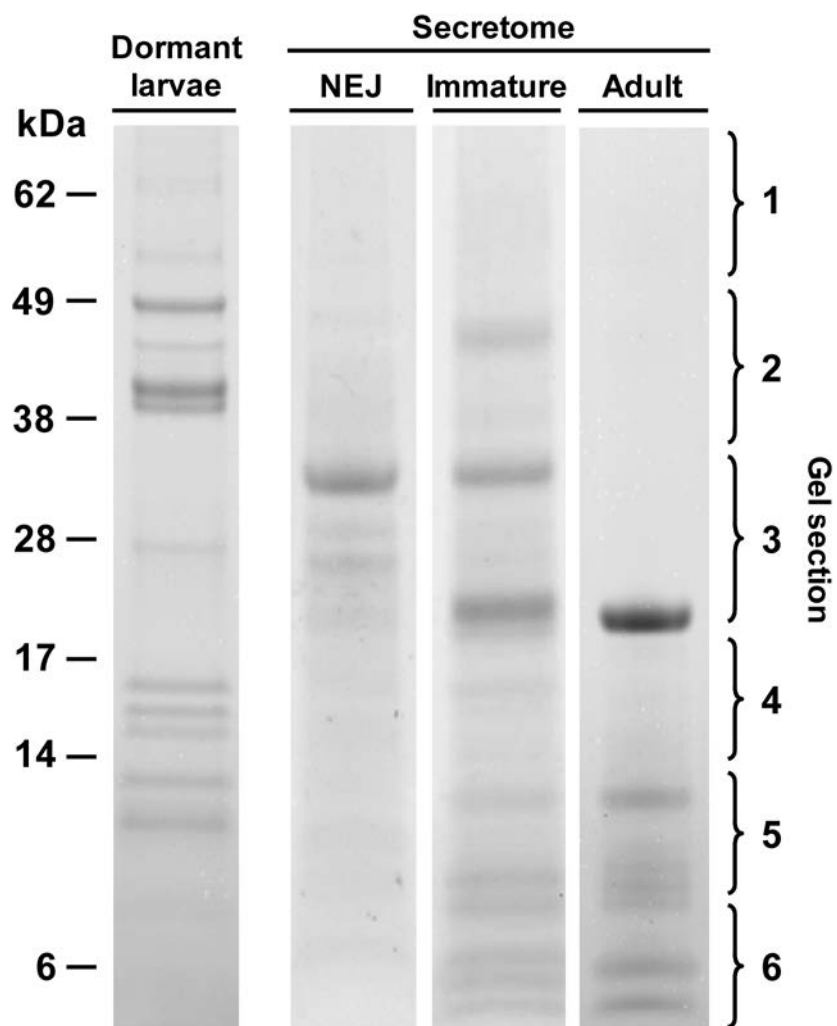
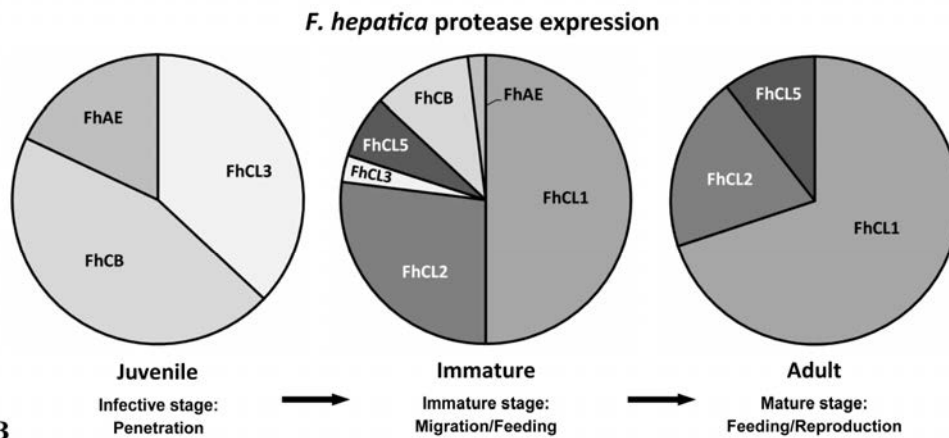


Figure 2

A

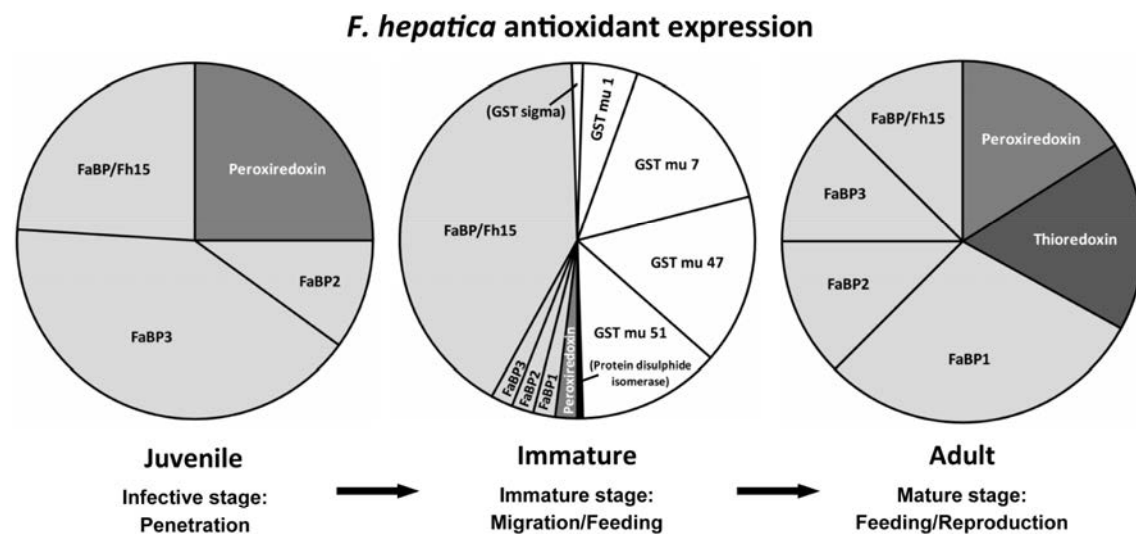


B

Accession	<i>Fasciola</i> protein	Score ¹	Unique peptides ²	% cover	Mr	emPAI ³	Section ⁴
Dormant larvae somatic							
A7UNB2	Cathepsin B	294	5	24	37.6	0.29	2
Q8I7B2	Procathepsin B2 precursor	111	3	9	38.0	0.09	2
A5X494	Cathepsin B3	223	1 ⁶	22	31.0	0.36	1-3
A8W7J0	Procathepsin L	235	7	21	35.6	0.30	2
Q9GRW6	Procathepsin L3	194	7	21	35.3	0.31	2
Q711M2	Legumain-like precursor 1	266	7	21	47.8	0.30	1-3
Fhep29h09.q1k	Putative legumain 2	76 ⁵	5	12	-	-	1
NEJ secretome							
P80527	Hemoglobinase-like protein 1	117	2	85	2.2	2.17	2,3
Fhep29h09.q1k	Putative legumain 2	99	12	12	-	-	1-5
Q9GRW4	Partial procathepsin L3	99 ⁵	11	23	35.1	-	2-5
Q9GRW6	Partial procathepsin L3	99 ⁵	8	23	35.4	-	2-5
Q95VA7	Cathepsin L	99 ⁵	5	27	37.4	-	2-5
B3TM67	Cathepsin L3	99 ⁵	9	26	35.1	-	2-5
B3TM68	Cathepsin L3	99 ⁵	9	39	35.3	-	2-5
A8W638	Metacercarial cathepsin L	99 ⁵	12	34	37.4	-	2-5
A8W7J0	Metacercarial procathepsin L	99 ⁵	13	38	35.7	-	2-5
HAN4015b05.q1kT3	Putative cathepsin B4	88 ⁵	3	11	-	-	3
Immature fluke secretome							
A7UNB2	Cathepsin B	113	2	7	37.6	0.09	2
Q8I7B2	Pro-cathepsin B2 precursor	99	2	6	38.0	0.18	5
Fhep45b05.q1k	Putative cathepsin B5	97 ⁵	6	13	-	-	2
HAN4006g01.q1kT3	Putative cathepsin B6	99 ⁵	6	9	-	-	2
Q711M2	Legumain-like precursor	412	31	16	47.8	0.70	2
A6Y9U8	Legumain-1	328	15	20	47.9	0.59	2
A6Y9U9	Legumain-2	176	3	10	48.3	0.22	2
Q8T5Z9	Cathepsin L	334	12	32	35.1	1.06	3
Q7JNQ9	Secreted cathepsin L1	447	16	37	36.7	1.82	3
Q6R018	Cathepsin L protein	292	13	24	36.6	1.00	3
Q24940	Cathepsin L-like proteinase	356	15	32	36.8	1.36	3
Q9GRW5	Cathepsin L1	282	12	27	35.1	0.88	3
Q7JNQ8	Secreted cathepsin L2	363	12	37	37.0	0.67	3
A5Z1V3	Secreted cathepsin L2	470	14	39	37.0	0.98	3
A5X483	Cathepsin L2	367	14	39	24.5	1.78	3
A3FMG6	Cathepsin L	228	9	30	36.9	0.19	3
Q95VA7	Cathepsin L	235	7	15	37.4	0.18	3
A8W638	Cathepsin L	274	6	25	37.3	0.29	3
Q9NGW3	Cathepsin L	204	7	11	36.9	0.54	3
Q9NB30	Cathepsin L	257	5	15	37.1	0.41	3
Q2HPD3	Cathepsin L1 proteinase	220	12	19	36.5	0.63	6
HAN4005b02.q1kT3	Putative prolylcarboxypeptidase	99 ⁵	10	33	-	-	
Adult secretome							
Q7JNQ9	Secreted cathepsin L1	552	20	39	36.7	1.82	1
A5Z1V3	Secreted cathepsin L2	444	12	34	37.0	1.16	1
Q9NB30	Cathepsin L	372	9	18	37.1	0.41	1
Q8T5Z9	Cathepsin L	481	17	40	35.1	1.25	1
Q9GRW5	Cathepsin L1	476	21	32	35.1	1.06	1
Q6R018	Cathepsin L protein	451	22	38	36.6	1.37	1
Q24940	Cathepsin L-like proteinase	444	19	33	36.8	1.58	1
A5X483	Cathepsin L2	401	16	45	24.5	2.16	1
Q24945	Cathepsin L-like protease	348	7	35	18.5	0.95	1
P81222	Cathepsin L-like proteinase	297	12	57	12.3	4.58	1
Q9NGW3	Cathepsin L	254	7	12	36.9	0.41	1
A5X484	Cathepsin L1	114	4	20	24.3	0.14	1
A8I598	Cathepsin L	77	2	7	28.1	-	1
Fhep30b01.q1k	Putative prolylcarboxypeptidase	92 ⁵	8	13	-	-	1

Figure 3

A



B

Accession	<i>Fasciola</i> Protein	Score	Unique peptides ¹	% cover	Mr	emPAI ²	Section ³
Dormant larvae somatic							
B0LT92	Peroxiredoxin	163	3	16	25.4	0.29	3
NEJ secretome							
Q7M4G1	Fatty acid-binding protein type 2	90	3	27	14.9	0.26	5
Q9U1G6	Fatty acid-binding protein type 3	217	3	31	14.6	1.03	5
Q7M4G0	Fatty acid-binding protein Fh15	104	2	20	14.7	0.61	5
Immature fluke secretome							
Q06A71	GST sigma class	138	4	23	24.5	0.14	3
P56598	GST mu class 26 kDa isozyme 1 (GST1)	303	9	38	25.7	0.63	3
P31671	GST mu class 26 kDa isozyme 7 (GST7)	361	14	34	25.3	2.05	3
P31670	GST mu class 26 kDa isozyme 47 (GST47)	381	15	44	25.3	2.04	3
P30112	GST mu class 26 kDa isozyme 51 (GST51)	390	16	39	25.3	1.70	3
B0LT92	Peroxiredoxin	237	5	21	24.5	0.29	3
O76945	Protein disulphide isomerase	199	6	14	55.1	0.06	3
Q9UAS2	Fatty acid-binding protein 1	77	4	16	14.6	0.23	5
Q7M4G1	Fatty acid-binding protein type 2	141	6	50	14.9	0.23	6
Q9U1G6	Fatty acid-binding protein type 3	141	3	28	14.6	0.23	5
Q7M4G0	Fatty acid-binding protein Fh15	349	14	59	14.7	5.55	5
Adult secretome							
B0LT92	Peroxiredoxin	211	5	21	25.4	0.29	1-3
Q9U1G7	Thioredoxin	83	2	22	11.6	0.30	2
Q9UAS2	Fatty acid-binding protein 1	107	3	16	14.6	0.52	2,3
Q7M4G1	Fatty acid-binding protein type 2	153	5	44	14.9	0.23	2,3
Q9U1G6	Fatty acid-binding protein type 3	81	2	18	14.6	0.23	2
Q7M4G0	Fatty acid-binding protein Fh15	208	7	51	14.7	0.23	2,3