

DRAFT 29 March 2012
Molecular and Cellular Proteomics

A deep exploration of the transcriptome and ‘excretory/secretory’ proteome of adult *Fascioloides magna*

**Cinzia Cantacessi^{a,b,*}, Jason Mulvenna^{b,*,§}, Neil D. Young^{a,*}, Martin Kasny^c, Petr Horak^c,
Ammar Aziz^b, Andreas Hofmann^d, Alex Loukas^b, Robin B. Gasser^{a,§}**

From the ^aFaculty of Veterinary Science, The University of Melbourne, Parkville, Victoria 3010, Australia, and the ^bQueensland Tropical Health Alliance, James Cook University, Cairns, Queensland 4878, Australia, and the ^cFaculty of Science, Charles University in Prague, Prague, Czech Republic, and the ^dEskitis Institute for Cell and Molecular Therapies, Griffith University, Brisbane, Queensland 4111, Australia

*Equal contributions.

[§]Corresponding authors. Email: robinbg@unimelb.edu.au (R.B. Gasser) or
(Jason.mulvenna@jcu.edu.au) (J. Mulvenna).

Parasitic liver flukes of the Fasciolidae family are responsible for major socio-economic losses worldwide. However, at present, knowledge of the fundamental molecular biology of these organisms is scant. Here, we characterize, for the first time, the entire transcriptome and secreted proteome of the adult stage of 'giant liver fluke', *Fascioloides magna*, using Illumina sequencing technology and one-dimensional SDS-PAGE and OFFGEL protein electrophoresis, respectively. A total of ~54,000,000 reads were generated and assembled into ~39,000 contiguous sequences (= contigs); ~20,000 peptides were predicted and classified based on homology searches, protein motifs, gene ontology and biological pathway mapping. From the predicted proteome, 48.1% proteins could be assigned to 384 biological pathway terms, including 'spliceosome', 'RNA transport' and 'endocytosis'. Putative proteins involved in amino acid degradation were most abundant. Of the 835 secreted proteins predicted from the transcriptome of *F. magna*, 80 were identified in the excretory/secretory products from this parasite. Highly represented were antioxidant proteins, followed by peptidases (particularly cathepsins) and proteins involved in carbohydrate metabolism. The integration of transcriptomic and proteomic datasets generated herein sets the scene for future studies, aimed at exploring the potential role/s that molecules may play at the host-parasite interface and for establishing novel strategies for the treatment or control of trematodiasis.

Parasitic liver flukes (Platyhelminthes: Trematoda) of livestock, such as *Fasciola hepatica* Linnaeus, 1758 and *F. gigantica* Cobbold, 1855, are responsible for major economic losses worldwide, estimated at 3,000 million dollars, due to morbidity, mortality and decreased productivity (1-3). Of these parasites, the giant liver fluke, *Fascioloides magna*, infects a range of wild and domestic ruminants (e.g., cervids and bovids), primarily in North America and Europe (4). The life cycle of *F. magna* is indirect; eggs are released by mature flukes and excreted in the faeces by the mammalian host. In aerated water, miracidia hatch from the eggs and penetrate the body of a susceptible, aquatic intermediate snail host (e.g., *Fossaria parva* and *F. modicella* in North America; *Galba truncatula* in Europe) within ~2 hours (5). In the intermediate host, the parasite develops through the stages of sporocysts, rediae and cercariae; the latter larval stage emerges from the snail within ~40-58 days (4). The cercariae encyst as metacercariae on submerged or emergent vegetation and are then ingested by a mammalian host. There, the metacercariae excyst and the juvenile flukes penetrate the intestinal walls and migrate to the liver, where they are encapsulated (in pairs) in pseudocysts formed by the hepatic parenchyma and then develop to mature flukes (4). The migration of the immature stages through the hepatic tissues, together with the large number of pseudocysts and size of mature flukes [up to 8 cm in length; (4)] can result in liver fibrosis. Clinical signs associated with infection by *F. magna* include lethargy, anorexia, depression and weight loss, with sudden death occurring in heavily infected animals (4).

In livestock, the control of liver fluke infections has relied predominantly on the treatment with anthelmintic drugs, such as closantel, oxclozanide and triclabendazole (6). Triclabendazole is considered the drug of choice against both juvenile and adult stages of liver flukes in the definitive (mammalian) host, whereas other compounds affect only the adult stage (7, 8). Thus, triclabendazole is widely and often excessively used for the treatment of trematodiasis in livestock (6, 9), which carries a significant risk that drug resistance will develop. Indeed, there are recent published reports of triclabendazole-resistance in *F. hepatica* populations in Australia (10) and western European countries (11-15). In addition, despite major efforts in studies aimed at developing novel intervention strategies against liver flukes (16-23), there is still a paucity of information on host-parasite interactions at the molecular level. Recent studies (3, 24, 25) have provided the first insights into the molecular biology of fasciolids through exploring the transcriptomes of the adult stage of both *F. hepatica* and *F. gigantica* (3, 25) and of the composition of the excreted/secreted (ES) products of both the juvenile and adult stages of *F. hepatica* (24). In these studies, proteolytic enzymes (e.g., cathepsins, asparaginyl endopeptidase cysteine proteases, and trypsin-like serine proteases) were identified as key molecules in both the transcriptome and ES products, which are likely to play key roles in parasite migration through tissues and in the modulation of the immune responses in the mammalian hosts. Given the potential of proteolytic enzymes as vaccine candidates against trematodiasis (18-20, 26), comparative analyses of the transcriptomes and ES products of liver flukes is of critical importance for an improved understanding of their molecular biology and also for developing new treatment and control strategies against them.

Advances in ‘-omic’ and computational technologies for the pre-processing, assembly and annotation of sequence data (27-29) are substantially assisting studies of the transcriptomes and proteomes of parasitic helminths (3, 28-32). In the present study we explore, for the first time on a large scale, the transcriptome of adult *F. magna*, using Illumina-based sequencing technology and bioinformatic analyses of sequence data, and characterize the protein components of the ES products of this developmental stage. This insight into the molecular biology *F. magna* offers unprecedented opportunities for comparative investigations of various economically important liver flukes and the design of new interventions against these parasites.

EXPERIMENTAL PROCEDURES

Procurement of parasite material, RNA isolation and Illumina sequencing - Adult *F. magna* were collected from naturally infected livers of red deer (*Cervus elaphus*) (from the Brdy mountains, Czech Republic) and washed in 0.1 M phosphate-buffered saline (PBS), pH 7.2, at

37 °C. Total RNA was extracted from three individual adults of *F. magna* using the TriPure reagent (Roche) and DNase I-treated (25). RNA amounts were estimated spectrophotometrically (NanoDrop Technologies), and RNA integrity was verified using a 2100 BioAnalyser (Agilent). Polyadenylated (polyA+) RNA was purified from 10 µg of total RNA using Sera-mag oligo(dT) beads, fragmented to a length of 100-500 bp, reverse transcribed using random hexamers, end-repaired and adaptor-ligated, according to the manufacturer's protocol (Illumina). Ligated products of ~300 bp were excised from agarose and PCR-amplified (15 cycles) (25). Products were cleaned using a MinElute column (Qiagen) and paired-end sequenced on a Genome Analyzer II (Illumina) (33), according to manufacturer's protocols.

Bioinformatic analyses of transcriptomic sequence data - The 100 bp single-read sequences generated from the non-normalized cDNA library representing the adult stage of *F. magna* were assembled using the program Oases v0.1.22 [<http://www.ebi.ac.uk/~zerbino/oases/>; (34)]. Adapter sequences and sequences with suboptimal read quality (i.e., PHRED score of < 32.0) were eliminated. The remaining sequences (99%) were used to construct a de Bruijn-graph using a *k*-mer value of 43 bp. Since *de novo* transcriptome assemblies of short-read sequence libraries can potentially lead to an under-representation of members of large protein families, characterized by a high degree of sequence similarity, such as the papain-like cysteine peptidases (= cathepsins) (25), reads with a high degree of sequence homology (98% nucleotide identity) to the conserved cathepsin propeptide inhibitor domain (I29; 177 nucleotides, nt) of a cathepsin L sequence from *F. magna* (GenBank protein accession number: ACG50798; nucleotide: EU877764.1) were used to generate contigs from the paired-end sequence datasets employing the PRICE software (<http://derisilab.ucsf.edu>). Eight cycles of contig extension were performed, with each new cycle using contigs (>500 nt in length) generated from the previous assembly, as a template. The identity of the assembled contigs was then verified by BLASTn and BLASTx comparisons with nucleotide and amino acid sequences of trematode cathepsins, respectively, available from public databases (via NCBI, www.ncbi.nlm.nih.gov). The raw sequence reads from the cDNA library were then mapped to the non-redundant sequence data using SOAP2 (35). In brief, raw reads were aligned to the assembled, non-redundant transcriptomic data, such that each read was mapped to a unique transcript. Reads that mapped to more than one transcript (called "multi-reads") were randomly allocated to a unique transcript, such that they were recorded only once. To provide a relative assessment of transcript-abundance, the numbers of raw reads that mapped to individual contigs were normalized for sequence length [i.e., reads per kilobase per million reads, RPKM; (36)].

The non-redundant transcriptomic dataset for *F. magna* was then analyzed using an established approach (25, 37). Briefly, assembled contigs were compared (using BLASTn and BLASTx algorithms) with sequences available in public databases, including (i) NCBI (www.ncbi.nlm.nih.gov) and the EMBL-EBI Parasite Genome Blast Server (www.ebi.ac.uk), (ii) transcriptomic sequence data available for the related liver flukes *F. hepatica*, *F. gigantica* (3, 25), *Clonorchis sinensis* and *Opisthorchis viverrini* (32) (<http://www.gasserlab.org/>) and (iii) the entire genome sequences of the blood flukes *Schistosoma haematobium* (38), *S. japonicum* (39) and *S. mansoni* (40) (<http://www.genedb.org/>, October 2011; e-value cut-off: < 10⁻⁵).

Proteins were conceptually translated from the predicted coding domains of individual transcriptomic sequences and their functions predicted using InterProScan (41) employing the default search parameters. Based on their homology to conserved domains and protein families, proteins predicted for *F. magna* were assigned parental (i.e., level 2) gene ontology (GO) terms (i.e., 'biological process', 'cellular component' and 'molecular function') (<http://www.geneontology.org/>) (42) and displayed using the WEGO tool (<http://wego.genomics.org.cn/cgi-bin/wego/index.pl>) (43). Inferred proteins with homologues in other organisms were mapped to conserved biological pathways utilizing the

Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology-Based Annotation System v.2 (= KOBAS2) (44). Signal peptides were also predicted using the program SignalP 3.0, employing the neural network and hidden Markov models (45) and transmembrane domains were inferred using the program TMHMM (www.cbs.dtu.dk/services/TMHMM/) (46-48). ES peptides were predicted based on the presence of a signal peptide and sequence homology to one or more known ES proteins listed in the Secreted Protein (<http://spd.cbi.pku.edu.cn/>) (49) and the Signal Peptide (<http://proline.bic.nus.edu.sg/spdb/index.html>) (50) databases.

Isolation of excretory/secretory (ES) products - Adult *F. magna* (n = 15) were washed three times in large volumes of PBS and incubated for 6 h at 37 °C in sterile RPMI 1640 medium (Sigma) containing 100 IU/ml penicillin G sodium, 100 µg/ml streptomycin sulfate, and 250 µg/ml amphotericin B (Sigma-Aldrich). Following incubation, the supernatant was centrifuged twice at 4500 g for 20 min at 2 °C, and the ES products were concentrated using Amicon Ultra-15 filters (10 kDa cut-off; Millipore). Protein amounts were measured using the Quant-iT Protein Assay Kit (Invitrogen); 1 ml of ES products (concentration = 1 mg/ml) was lyophilized and stored at -20 °C until analysis.

Electrophoresis and in-gel digestion - Two 10 µl aliquots of *F. magna* ES products (9 mg/ml) were incubated for 2 h at 37 °C with an equal volume of sample buffer (51). The samples were then applied to a 0.75-mm-thick 4% stacking, 12.5% resolving gel (prepared using the Bio-Rad PROTEAN 3) for SDS-PAGE (51). Electrophoresis was carried out using maxima of 40 mA/gel and 300 V. Proteins were stained using Coomassie Brilliant Blue and destained in 50:8.75:41.25 methanol/acetic acid/water (v:v:v). Lanes containing ES proteins were finely sliced; the gel slices were then destained twice in 200 µl of 50% acetonitrile, 25 mM NH₄HCO₃ for 15 min at 37 °C, desiccated using a vacuum centrifuge and then resuspended in 20 mM dithiothreitol (DTT) and reduced for 1 h at 65 °C. [After removal of ... \(was it actually the DTT-containing supernatant that was removed ?\)](#) The DTT was then removed, and the samples were alkylated by adding 1 M iodoacetamide (IAA) to a final concentration of 50 mM and incubation at 22 °C in darkness for 40 min. Gel slices were washed three times for 45 min in 25 mM NH₄HCO₃ and then desiccated. Individual dried slices were then allowed to swell in 20 µl of 40 mM NH₄HCO₃, 10% acetonitrile containing 20 µg/ml trypsin (Sigma) for 1 h at 22 °C. An additional 50 µl of the same solution was added to the samples and incubated overnight at 37 °C. The supernatant was then removed from the gel slices, and residual peptides were washed from the slices by incubating them three times in 0.1% trifluoroacetic acid (TFA) for 45 min at 37 °C. The original supernatant and extracts were combined and reduced to 10 µl in a vacuum centrifuge before mass spectral analysis.

OFFGEL electrophoresis - For peptide separation, based on isoelectric point (pI), 240 µg of ES protein were reduced by adding DTT to 20 mM and 10% SDS to 2% (w/v) and an incubation at 65 °C for 1 h. Alkylation was then achieved by adding IAA to 50 mM and incubating the solution for 40 min in darkness at 22 °C. The protein sample was then co-precipitated with 1 µl of a 1 µg/µl solution of trypsin by adding 9 volumes of methanol at -20 °C. After incubation overnight at -20 °C, the sample was resuspended in 25 mM NH₄HCO₃ and incubated at 37 °C for 5 h with the addition of 1 µg of trypsin after 3 h. The 3100 OFFGEL fractionator and OFFGEL kit (pH 3-10; 24-well format) (Agilent Technologies) were prepared according to the manufacturer's protocols. The ES protein digests and undigested protein samples were diluted in the peptide- and protein-focusing buffers, respectively, to a final volume of 3.6 ml, and 150 µl of sample were loaded into each well. The samples were focused in a maximum current of 50 µA until 50 kilovolt hours (kVh) were achieved. Peptide fractions were harvested and desiccated using a vacuum centrifuge, resuspended in 25 mM NH₄HCO₃, and desiccated again before mass spectral analysis.

Protein identification using LC-MS/MS-MS = OGE (OFFGEL?) fractions and tryptic fragments from in-gel digests were separated chromatographically by HPLC (Dionex

Ultimate 3000) using a Zorbax 300SB-C18 column (3.5 μm , 150 mm \times 75 μm ; Agilent) and a linear gradient of 0–80% solvent B for 60 min. A pre-concentration step (3 min) was performed employing a Dionex μ -Precolumn cartridge (C18, 5 μm , 300 μm \times 5 mm) before commencement of the gradient. A flow rate of 300 nl/min was used for all experiments. The mobile phase consisted of solvent A (0.1% formic acid [aq]) and solvent B (80/20 acetonitrile/0.1% formic acid [aq]). Eluates from the RP-HPLC column were directly introduced into the NanoSpray II ionisation source of a QSTAR Elite Hybrid MS/MS System (Applied Biosystems), operated in positive ion electro-spray mode. All analyses were performed using Information Dependant Acquisition. Analyst 2.0 (Applied Biosystems) was used for data analysis. Briefly, the acquisition protocol consisted of the use of an Enhanced Mass Spectrum scan. The three most abundant ions detected (above the background threshold) were subjected to examination using an Enhanced Resolution scan to confirm the charge state of the multiply-charged ions. The ions with a charge state of +2, +3 or with an unknown charge were then subjected to collision-induced dissociation using a rolling collision energy, dependent upon the m/z and the charge state of the ion. Enhanced Product Ion scans were acquired, resulting in full product ion spectra for each of the selected precursors, which were then used for subsequent database searches.

Bioinformatic analyses of proteomic sequence data - Protein sequences from mass spectrometric analysis were compared with peptide sequences predicted from transcriptomic data as well to proteins available in the SwissProt database (May 2011) using Mascot v.2.3.02 (<http://www.matrixscience.com>), employing the following search parameters: enzyme = trypsin; precursor ion mass tolerance = ± 0.1 ; fixed modifications = methionine oxidation; variable modifications = carbamidomethylation; number of missed cleavages allowed = 2; charge states = +2 and +3. As required, the program MudPIT was used to compare mass spectrometric data with the theoretical parameters for known proteins, in order to assign protein scores. The results from the Mascot searches were validated using the program Scaffold v.3.00.06 (Proteome Software Inc.) (52). Peptides and proteins were identified using the Peptide Prophet algorithm (53), using a probability cut-off of 95.0% (peptides) or 99.0% probability (proteins) and contained at least two identified peptides (proteins) (54). Proteins containing similar peptides but that could not be differentiated based on MS/MS analysis were grouped to satisfy the principles of parsimony. A false discovery rate of 0.1% was calculated using protein identifications validated using the Scaffold program. Relative protein abundance was determined by normalized spectral counting (52), and proteins encoded by transcripts were annotated by BLASTx comparisons (bit-score cut-off: >30) with data available in the non-redundant (NR) protein databases (www.ncbi.nlm.nih.gov). Proteins were also classified according to GO categories using the program InterProScan (41), and putative signal peptides and transmembrane domain/s predicted using the programs SignalP (55) and TMHMM (47). Putative mannose 6-phosphate glycosylation sites were identified using NetNGlyc (56).

RESULTS

The transcriptome of F. magna - A total number of 54,004,008 Illumina reads were produced for the adult stage of *F. magna* (Table I). The assembly of the raw reads yielded 38,925 contigs (842 \pm 1,005 bases in length; range: 101-13,180), with a GC content of 44.04% (Table I). The PRICE-assembly of cathepsin-encoding contigs yielded 9 additional sequences, resulting in 38,934 assembled contigs (843 \pm 1,005 bases in length; range: 101-13,180) (Table I). A total of 26,616,266 reads (~49.3% of all raw reads) could be re-mapped to the assembled contigs, with a mean depth of coverage of 847.19 \pm 30,910.17 reads per sequence. A total of 25,405 and 23,025 of 38,934 assembled contigs (65.2% and 59.1%, respectively) matched known *F. hepatica* and *F. gigantica* homologues (e-value cut-off: $< 10^{-5}$), respectively, whereas 5,488 (14.1%) and 7,533 (19.3%) had homologues in *C. sinensis* and *O. viverrini*, respectively (Table I). A total number of 3,941 (10.1%), 3,966 (10.2%) and 4,147 (10.6%) had homologues in *S. haematobium*, *S. japonicum* and *S. mansoni*, respectively (Table I). In total, 20,140 peptides were predicted from the transcriptome of *F. magna* (Table I); 11,166 of them

(55.4%) had highest homology to proteins predicted for *F. hepatica*, followed by those from *F. gigantica* (n = 10,854; 53.9%), *O. viverrini* (n = 9,739; 48.3%) and *S. mansoni* (n = 9,376; 46.5%) (Table I).

Proteins inferred from the *F. magna* transcriptome were then categorized according to the presence of conserved (Pfam and/or InterPro) domains/signatures; 15,414 (76.5%) sequences could be mapped to known proteins, characterized by 2,965 different conserved domains (Table I). Predicted proteins were also classified according to their GO terms (i.e., molecular function, cellular localization and/or association with biological pathways). Of 20,140 proteins predicted for *F. magna*, 35.4% could be assigned to 379 'biological process', 140 'cellular component' and 533 'molecular function' parental (level 2) terms (Table I). The predominant terms were 'cellular process' and 'metabolic process' for 'biological process' (15.7% and 10.1%, respectively), 'cell' and 'cell part' for 'cellular component' (28.6% and 13.3%, respectively) and 'ATP binding' and 'catalytic binding' for 'molecular function' (36.4% and 34.8%, respectively) (Fig. 1). From the proteome predicted for *F. magna*, 9,690 (48.1%), proteins mapped to homologous proteins in the KEGG database, which were assigned to 384 biological pathway terms (Table I), including 'spliceosome' (n = 105 molecules), 'RNA transport' (86) and 'endocytosis' (76) (Table SI). Putative proteins involved in amino acid degradation (i.e., cathepsins, n = 15) were most abundant (Fig. 2), in accordance with previous studies of the transcriptomes of other parasitic trematodes, including *F. hepatica* (3, 24), *C. sinensis* (see 57-60), *O. viverrini* (Littman et al., unpublished data) and *Paragonimus westermani* (61). A total of 835 (4.1%) ES proteins were predicted from the transcriptome of *F. magna* (Table I), 4% of which were inferred to be peptidases (n = 32), including cathepsins B and L (n = 7) (Table SII).

Proteomic analysis of excretory/secretory (ES) proteins - F. magna ES proteins were subjected to SDS-PAGE and the protein constituents identified using tandem mass spectroscopy. More than 37,000 spectra from in-gel digests and OGE analysis were used for Mascot searches of the transcriptome. Using Scaffold, 80 proteins were identified by, at least, two peptides at a 99.0% probability and an estimated false discovery rate of 0.1%. The proteins identified were assigned to functional categories on the basis of InterPro domains and/or GO categories. The largest functional group represented antioxidant proteins, followed by peptidases and proteins linked to carbohydrate metabolism; approximately one quarter of the proteins identified could not be assigned to a major functional group, and they were thus grouped into a 'miscellaneous' group (Fig. 3A). In accordance with respective functional categories, GO terms highly represented among the proteins included "cysteine-type peptidase activity", "oxidoreductase activity", "serine-type endopeptidase inhibitor activity" and "carbohydrate metabolic process". Using the spectral counting method included in Scaffold, the cysteine protease inhibitor, cystatin, was the most abundant protein in the ES products, followed by two cathepsin L1 proteases, cathepsin B, calpain and ferritin. Cathepsin L was the most abundant protein identified in the ES products, with four different isoforms identified, and a cathepsin L-like protease. Similarly, multiple isoforms of cathepsin B2 and fatty acid-binding protein were identified. In all cases, isoforms were reported only if at least one unique peptide (which scored significantly) was assigned to the isoform (see Fig. 4). Using SignalP and TMHMM (47, 55), ~30% of the identified proteins were inferred to contain a classical secretory signal peptide, and represented mainly proteases and proteins associated with carbohydrate metabolism. Six proteins were predicted to contain more than two transmembrane helices, indicative of a membrane bound protein. Protein localization was reinforced using a combination of PSORT (62) and information from the literature. The majority of ES proteins were predicted to be either extracellular (= 27) or cytoplasmic (= 39), with smaller numbers inferred to be localised or co-localised to the nucleus, plasma membrane and/or mitochondria (Fig. 3B). Twelve proteins were assigned a lysosomal localization; to further validate this finding, all proteins with a predicted signal sequence were analysed for the presence of potential mannose 6-phosphate glycosylation sites. Of the 23 proteins with predicted signal peptides, ten proteins had a glycosylation probability of >0.75 and six >0.50. Proteins inferred to be glycosylated included known lysosomal proteins, such

as Pro-Xaa carboxypeptidase, alpha-mannosidase, alpha-glucosidase, and cathepsins A and B2.

Proteases represented a significant proportion (~16%) of the proteins identified in the ES. Thirteen proteases were identified in the ES; nine cysteine proteases, including cathepsin B and L, calpain and legumain; three carboxypeptidases, including cathepsin A; and, a leucine aminopeptidase (Fig. 3C). Cathepsins were the most abundant, with examples of cathepsin A, B and L identified. Based on BLAST comparisons, all cathepsin L proteins identified in the ES products shared significant homology with *F. hepatica* or *F. magna* cathepsin L1, in accordance with the results from the analysis of the *F. magna* transcriptome (see Fig. 2). In addition to cathepsins identified by proteomic means, cathepsins D, L3 and a cathepsin B-like protease, whose corresponding transcripts were detected in the *F. magna* transcriptome, were not found in the proteome. The cathepsins identified correlated with the most abundantly expressed transcripts, suggesting that unidentified proteases were expressed at levels below the detection limit of the mass spectrometric method used (Table II). Analysis of the peptides detected in Mascot searches showed that only one peptide representing cathepsin L (ID: 39933) (see Fig. 4) originated from the pro-region of the protease, indicating that these proteases were predominantly present in an activated form. Conversely, three pro-region peptides were identified from one of the cathepsin B2 isoforms (ID: 39926) (Fig. S1), suggesting the presence of inactive forms of this protease.

The ES proteins from *F. magna* were compared with those characterized in previous studies of *F. hepatica* (24, 63-66), *C. sinensis* (67) and *O. viverrini* (68). Approximately 65% of the proteins identified here were also detected in at least one of these flukes, but only enolase, actin and triose-phosphate isomerase were identified in all three species. The ES proteomes of *F. hepatica* and *F. magna* were most similar, with 47 of 80 proteins identified here found in both species; a major difference between these proteomes was the identification of ten isoforms of cathepsin L1 in *F. hepatica* and only five in *F. magna*. The 27 proteins unique to *F. magna* were less abundant and membrane-bound and/or structural, although eight of them contained a secretory signal peptide. Apart from the difference in cathepsin L1 isoform numbers, the protease profile of *F. magna* differed from that of *F. hepatica* only by the presence of cathepsin A. However, the protease profile of both *F. hepatica* and *F. magna* (Fasciolidae) was markedly different from that of *C. sinensis* and *O. viverrini* (Opisthorchiidae) (67, 68), with only cathepsin B2 and legumain detected in the ES of both groups (Fig. 3C and 3D).

DISCUSSION

The present study provides a comprehensive snapshot of the transcriptome and the secreted proteome of the adult stage of *F. magna* and represents an invaluable resource for fundamental investigations of the molecular biology of liver flukes of both veterinary and public health importance. Almost two thirds of *F. magna* transcripts sequenced in the present study were similar to molecules identified in the transcriptomes of *F. hepatica* and *F. gigantica*, respectively (3, 25), which is indicative of the biological similarities shared among members of the family Fasciolidae. Most abundant in the transcriptome of *F. magna* were molecules containing a predicted signal peptide; this finding is in accordance with the results of a previous study of the transcriptome of *F. gigantica* (25), and is likely to reflect the crucial role/s that secreted proteins play in the biology of these organisms (69). Of the 835 predicted proteins with a signal peptide in *F. magna*, 80 were identified in the ES products from this organism at a 99% confidence, in accordance with previous proteomic analyses of the ES products of other liver flukes (i.e., 20-90 proteins identified) (24, 66-68). This finding suggests that, in addition to peptides that were undetectable because of a low concentration, the vast majority of proteins in ES products of *F. magna* was identified. In a previous study of the transcriptome and secreted proteome of *F. hepatica*, Robinson et al. (24) also observed a discrepancy between the number of secreted proteins predicted from transcriptomic data and that of proteins identified in the ES products from this species. In the present study, a total of 42 transcripts encoding papain-like cysteine peptidases, designated 'cathepsins' (e.g., cathepsins B and L), were detected in the transcriptome of *F. magna*, of which seven were

predicted to contain a signal peptide indicative of secretion (see Table SII), and eight were identified in the ES products (cf. Fig. 3C). A possible explanation for the absence of a predicted signal peptide in transcripts encoding proteins identified in the ES products is that some members of the cathepsin family could be excreted/secreted *via* a “non-classical” pathway that does not involve signal peptide cleavage (70). Conversely, the difference in number of cathepsin-encoding transcripts predicted from transcriptomic data and the number of cathepsins identified in the ES products supports the hypothesis that the transcriptome of fasciolids encodes ‘endogenous’ cathepsins which function in key biological pathways, such as egg production, protein turnover and remodelling (71), and ‘exogenous’ cathepsins whose roles appear to relate to the digestion of host molecules (72, 73).

In parasitic trematodes, cathepsins are known to be encoded by multi-gene families (74), which poses a challenge for both the *de novo* assembly of transcripts encoding different, particularly closely related, isoforms and the identification of these isoforms using mass spectrometry. The accurate characterization of isoforms by spectrometry is highly dependent on their sequence similarity and their abundance in the matrix subjected to analysis (in this case, ES). Thus, it is possible that the high sequence similarity of the cathepsin L proteins identified in the transcriptome impaired mass spectral identification of low-abundant cathepsins L, which, in some cases, relied on the specific detection of variation of only two or three peptide differences. The presence of multiple isoforms of secreted cathepsin L has been reported previously in both *F. hepatica* (24, 66) and *F. gigantica* (25). In *F. hepatica*, these proteins have been shown to be crucial for parasite survival, mediating essential processes, such as the digestion of host macromolecules and the suppression of the host immune response (20). In *F. hepatica*, 24 different cathepsin L isoforms have been identified and classified into five clades (designated ‘Clades 1-5’; ‘FhCL1-5’) (65); members of Clade 1, 2 and 5 were shown to be present among ES proteins of adult worms, whereas members of Clades 3 and 4 were detected exclusively in ES products from juvenile worms, thus suggesting that different isoforms play distinct roles in molecular mechanisms linked to the invasion of and the survival in the vertebrate host (65). The expansion of the cathepsin family of peptidases in *F. hepatica* is the result of a series of gene duplication events (65, 75). The sequence diversity displayed by members of the cathepsin protein family, together with their broad range of substrate specificities, have been hypothesized to play a role in the ability of parasitic trematodes to infect a wide range of mammalian hosts (20, 75, 76). In the present study, transcripts encoding cathepsins with significant homology to FhCL3 were identified in the transcriptome of adult *F. magna*; however, the corresponding proteins were not detected in the ES of this trematode by proteomic means. This finding contrasts a previous analysis of *F. hepatica*, in which molecules encoding FhCL3 were not detected in the transcriptome of the adult worm (65). Five distinct cathepsin L isoforms were identified in the ES proteome of *F. magna*, including four isoforms and a cathepsin L-like protease, all significant homologues of FhCL1. However, some S2 active site residues of the *F. magna* cathepsin L (see Fig. 4) peptidases were similar to those of cathepsins expressed specifically by the juvenile stages of *F. hepatica* (65). For instance, in all but one of the *F. magna* cathepsins L, the initial two residues of the active site were either Trp-Met or Phe-Met, in accordance with the residues detected in cathepsin L peptidases of the juvenile stage of *F. hepatica* (Fig. 4) (65, 71). In each case, the presence of these amino acids was confirmed by fragmentation analysis in tandem MS. Likewise, the S2 pocket of three of the five cathepsins L of *F. magna* also presented a substitution in the terminal amino acid, typical of cathepsins L specifically expressed by juvenile *F. hepatica* (65), which was also confirmed by MS (Fig. 4). The amino acid composition of the active site of cathepsin L has been shown to affect substrate specificity (65); therefore, the developmental regulation of the expression of cathepsins L is likely to represent a potential adaptation of different developmental stages of the parasite to the diverse environmental conditions encountered throughout its life cycle. Based on these observations, it is tempting to speculate that, if developmental regulation of cathepsin L peptidases occurs in *F. magna*, it will differ from that in *F. hepatica*. Further studies aimed at elucidating the expression profiles of cathepsin-encoding transcripts in different stages of *F.*

magna, as well as the presence of members of this protein family in the ES products of the juvenile stages, will be crucial to address this point.

In addition to the cathepsin L isoforms, six other proteases were detected in the ES products from *F. magna*, namely two isoforms of cathepsin B2, a cathepsin A, a legumain, two isoforms of lysosomal Pro-Xaa carboxypeptidase and a leucine aminopeptidase (cf. Fig. 3C). Like cathepsin L, cathepsin B is thought to be expressed as an inactive zymogen and is activated following the removal of a pre-pro-region by an asparaginyl endopeptidase (= legumain) (77), also present in the ES products from *F. magna*. In the present study, predominantly activated forms of cathepsin L were identified, as evidenced by the lack of peptides assigned to the pre-pro-region. In a single case, the amino acid sequence of one of the cathepsins B identified included three pre-pro-region peptides, typical of the inactive form of this protein; however, each of the peptides from this region was represented by only one spectrum, while spectral counts of peptides from the mature sequence were higher, e.g. up to 23 for one of the peptides. Therefore, it is likely that inactivated proteases constituted a small proportion of the total number of cathepsins B detected in the ES products. In addition, for the second isoform of cathepsin B2 identified herein, the Asn in the region of the amino acid sequence, which precedes the first residue of the mature protein (typical of all cathepsin L peptidases from *F. hepatica* and proposed to act as a substrate for possible activation of these proteins by legumain proteases) (77), was substituted by a Glu residue (Table SIII). The same substitution was observed in three of the cathepsins L of *F. magna* (Fig. 4). Several of the proteases identified in the *F. magna* ES products were of lysosomal origin, including cathepsin A, the two Pro-Xaa carboxypeptidases and, possibly, legumain (cf. Fig. 3C). On the basis of PSORT and the presence of mannose 6-phosphate glycosylation sites, 12 lysosomal proteins were identified as part of the *F. magna* ES, which constituted ~15% of the total identifications; after cystatin and the cathepsin proteases, lysosomal proteins were most abundant in the ES products. Recent proteomic studies of *S. mansoni* (blood fluke) and *F. hepatica* vomitus also identified a number of lysosomal proteins as enzymes putatively involved in digestive processes (66, 78). In *S. mansoni*, this observation led to the hypothesis that lysosomal proteases might be actively secreted into the gut lumen, the pH of which facilitates the activity of these proteins, in order to digest plasma components as well as haemoglobin (78). A legumain, a pro-X carboxypeptidase, a beta-glucosidase, several isoforms of ferritin and the Niemann-Pick C2 protein were also identified here as constituents of the ES products from *F. magna*. The presence of these proteins in the *F. magna* ES, as well as serpin and kunitz-type protease inhibitors, is likely a consequence of the fact that *F. magna* readily regurgitates the contents of its digestive tract into the culture media, as previously observed for *F. hepatica* (66). Indeed, the same proteins were also identified in vomitus of both the latter trematode species and *S. mansoni* (66, 78).

The vast majority of *F. magna* ES products represented a complex mixture of proteins predominantly of extracellular or cytoplasmic origin. In particular, a significant proportion of non-classically secreted ES proteins, including a number of proteins predicted to be membrane-bound, such as a tetraspanin and a glucose transporter, were identified. Unlike ES products from gastrointestinal parasitic nematodes, such as *Ancylostoma caninum* (31), in which classically secreted proteins are very abundant, non-classically secreted proteins have been consistently shown to represent a significant proportion of ES from trematode parasites, including *S. japonicum* (79), *C. sinensis* (67), *O. viverrini* (68) and *F. hepatica* (66). For the latter trematode, it has been proposed that the presence of these proteins in the ES products is a result of stress-induced shedding of the tegument during culture (64). However, morphological studies of this parasite have observed a rapid turn-over of its tegument facilitated by subtegumental cells, thus leading to the hypothesis that the shedding of the tegument and associated proteins occurs *in vivo* and may represent an immuno-defensive strategy (24, 80). Two phospholipases and an ABC transporter (11521 and 21652) were also identified among the ES proteins of *F. magna*. The presence of ABC transporters in the tegumental membrane of *F. hepatica* has prompted comparisons to the ER/Golgi-independent secretion of IL-1 β and caspase-1 in mammalian cells (24). This process involves formation of plasma membrane 'blebs', mediated by ABC transporters, which are subsequently released as

microvesicles after phospholipase-mediated fusion with the plasma membrane (81); the presence of both phospholipases and ABC transporters in the ES products from *F. magna* suggests that a similar process may occur in this trematode. While some ES proteins identified in the present study might relate to gut content regurgitated by the parasite (66), 15 antioxidant proteins were identified, none of which possessed a signal peptide indicative of secretion. A similar profile of ES antioxidant proteins, including fatty acid-binding proteins, glutathione S-transferase, peroxiredoxin and thioredoxin, has been described in *F. hepatica* (24). In the latter trematode, these proteins are proposed to play an important role in the evasion of host immune responses, which likely relates to protecting the parasite from reactive oxygen species released by host immune cells (82-85), an inhibition of the proliferative potential of spleen cells (e.g., in rats) (86) as well as the recruitment and alternative activation of macrophages (86-88). All antioxidant proteins identified here, with the exception of dihydrolipoamide dehydrogenase, were also detected in ES products from adult *F. hepatica*, suggesting that these proteins play a role in immune evasion (82-84, 86-88). In *F. hepatica*, a variable antioxidant profile was observed in ES from different developmental stages (24), which led to the hypothesis that these non-classically secreted antioxidant proteins were transported through an alternative, trans-tegumental, secretory pathway (24). However, the existence of such a pathway, in both *F. hepatica* and *F. magna*, remains to be demonstrated.

The availability of the entire genome sequences of related species of liver flukes, such as *C. sinensis* (89), and blood flukes (38-40), now provides unprecedented opportunities to (i) conduct comparative proteomic comparisons, (ii) elucidate the structures and functions of key molecules (e.g., ‘endogenous’ and ‘exogenous’ cathepsins), (iii) explore novel biological pathways (e.g., trans-tegumental, secretory pathways) and (iv) establish relationships among genes, transcripts and proteins involved specifically in the invasion of and establishment in and interactions with the host. Advancing these areas will provide a basis for studies aimed at exploring the potential of these molecules as targets for the development of novel strategies for the control of trematodiasis.

Currently, the Gasser Lab is supported by funding from the Australian Research Council (ARC), Melbourne Water Corporation and the National Health and Medical Research Council (NHMRC) (RBG). Support from the Victorian Life Sciences Computation Initiative (VLSCI) and IBM Collaboratory is also acknowledged (RBG). The Hofmann and Loukas Labs are also supported by grants from the NHMRC. This study was also supported by the Linkage Infrastructure, Equipment and Facilities (LIEF) scheme from ARC (JM), and by the Czech Science Foundation (grant nos. P502/10/P248 and 206/09/H026), the Czech Ministry of Education (grant no. MSM LC06009 and MSM 0021620820) and partially by UNCE of the Charles University (grant no. 204017) (MK). JM holds a Career Development Award, (CDA1) and NDY an Early Career Researcher (ECR) Fellowship from NHMRC.

REFERENCES

1. Boray, J. C. (1997). Chemotherapy of infections with Fasciolidae. In: Boray JC (Ed.), Immunology, Pathobiology and Control of Fasciolosis. MSD AGVET, Rahway, NJ, pp. 83-97
2. Kaplan, R. M. (2001). *Fasciola hepatica*: a review of the economic impact in cattle and considerations for control. *Vet. Ther.* **2**, 40-50
3. Young, N. D., Hall, R. S., Jex, A. R., Cantacessi, C., and Gasser, R. B. (2010). Elucidating the transcriptome of *Fasciola hepatica* – a key to fundamental and biotechnological discoveries for a neglected parasite. *Biotechnol. Adv.* **28**, 222–231
4. Pybus, M. J. (2001). Liver flukes. In: Samuel, W. M., Pybus, M. J., and Kocan, A. A. (Eds.), Parasitic Diseases in Wild Mammals, Iowa State Press, Iowa City, pp. 121-149
5. Coil, W. H. (1977). The penetration of *Fascioloides magna* miracidia into the snail host *Fossaria bulimoides*. A scanning electron microscope study. *Zeitschrift für Parasitenkunde* **52**, 53-59
6. Fairweather, I. (2005). Triclabendazole: new skills to unravel an old(ish) enigma. *J. Helminthol.* **79**, 227-234
7. Maes, L., Vanparijs, O., Lauwers, H., and Deckers, W. (1990) Comparative efficacy of closantel and triclabendazole against *Fasciola hepatica* in experimentally infected sheep. *Vet. Rec.* **127**, 450-452
8. Keiser, J., and Utzinger, J. (2004). Chemotherapy for major food-borne trematodes: a review. *Expert Opin. Pharmacother.* **5**, 1711-1726
9. Flynn, R. J., Mulcahy, G., and Elsheikha, H. M. (2010). Coordinating innate and adaptive immunity in *Fasciola hepatica* infection: implications for control. *Vet. Parasitol.* **169**, 235-240
10. Overend, D. J., and Bowen, F. L. (1995). Resistance of *Fasciola hepatica* to triclabendazole. *Aust. Vet. J.* **72**, 275-276
11. Coles, G. C., Rhodes, A. C., and Stafford, K. A. (2000). Activity of closantel against adult triclabendazole-resistant *Fasciola hepatica*. *Vet. Rec.* **146**, 504
12. Moll, L., Gaasenbeek, C. P. H., Vellema, P., and Borgsteede, F. H. M. (2000). Resistance of *Fasciola hepatica* against triclabendazole in cattle and sheep in the Netherlands. *Vet. Parasitol.* **91**, 153-158
13. Thomas, I., Coles, G. C., and Duffus, K. (2000). Triclabendazole-resistant *Fasciola hepatica* in southwest Wales. *Vet. Rec.* **146**, 200
14. Fairweather, I. (2009) Triclabendazole progress report, 2005-2009: an advancement of learning? *J. Helminthol.* **83**, 139-150
15. Gaasenbeek, C. P. H., Moll, L., Cornelissen, J., Vellema, P., and Borgsteede, F. H. M. (2001). An experimental study on triclabendazole resistance of *Fasciola hepatica* in sheep. *Vet. Parasitol.* **95**, 37-43
16. Spithill, T. W., and Dalton, J. P. (1998). Progress in development of liver fluke vaccines. *Parasitol. Today* **14**, 224-228
17. Piacenza, L., Acosta, D., Basmadjian, I., Dalton, J. P., and Carmona, C. (1999). Vaccination with cathepsin L proteinases and with leucine aminopeptidase induces high levels of protection against fascioliasis in sheep. *Infect. Immun.* **67**, 1954-1961
18. Dalton, J. P., and Mulcahy, G. (2001). Parasite vaccines – a reality? *Vet. Parasitol.* **98**, 149-167
19. Dalton, J. P., Brindley, P. J., Knox, D. P., Brady, C. P., Hotez, P. J., Donnelly, S., O'Neill, S. M., Mulcahy, G., and Loukas, A. (2003). Helminth vaccines: from mining genomic information for vaccine targets to systems used for protein expression. *Int. J. Parasitol.* **33**, 621-640
20. Dalton, J. P., O'Neill, S., Stack, C., Collins, P., Walshe, A., Sekiya, M., Doyle, S., Mulcahy, G., Hoyle, D., Khaznadjji, E., Moire, N., Brennan, G., Mousley, A., Kreshchenko, N., Maule A. G., and Donnelly, S. M. (2003). *Fasciola hepatica*

- cathepsin L-like proteases: biology, function, and potential in the development of first generation liver fluke vaccines. *Int. J. Parasitol.* **33**, 1173-1181
21. McManus, D. P., and Dalton, J. P. (2006). Vaccines against the zoonotic trematodes *Schistosoma japonicum*, *Fasciola hepatica* and *Fasciola gigantica*. *Parasitology* **133**, S43-S61
 22. Acosta, D., Cancela, M., Piacenza, L., Roche, L., Carmona, C., and Tort, J. F. (2008). *Fasciola hepatica* leucine aminopeptidase, a promising candidate for vaccination against ruminant fasciolosis. *Mol. Biochem. Parasitol.* **158**, 52-64
 23. Jayaraj, R., Piedrafita, D., Dynon K., Grams, R., Spithill, T. W., and Smooker, P. M. (2009). Vaccination against fasciolosis by a multivalent vaccine of stage-specific antigens. *Vet. Parasitol.* **160**, 230-236
 24. Robinson, M. W., Menon, R., Donnelly, S. M., Dalton, J. P., and Ranganathan, S. (2009). An integrated transcriptomics and proteomics analysis of the secretome of the helminth pathogen *Fasciola hepatica*: proteins associated with invasion and infection of the mammalian host. *Mol. Cell. Proteomics* **8**, 1891-1907
 25. Young, N. D., Jex, A. R., Cantacessi, C., Hall, R. S., Campbell, B. E., Spithill, T. W., Tangkawattana, S., Tangkawattana, P., Laha, T., and Gasser, R. B. (2011). A portrait of the transcriptome of the neglected trematode, *Fasciola gigantica* – biological and biotechnological implications. *PLoS Negl. Trop. Dis.* **5**, e1004
 26. Kasny, M., Mikes, L., Hampl, V., Dvorak, J., Caffrey, C. R., Dalton, J. P., and Horak, P. (2009). Chapter 4. Peptidases of trematodes. *Adv. Parasitol.* **69**, 205-297
 27. DeMarco, R., and Verjovski-Almeida, S. (2009). Schistosomes – proteomics studies for potential novel vaccines and drug targets. *Drug Discov. Today* **14**, 472-478
 28. Chuan, J., Feng, Z., Brindley, P. J., McManus, D. P., Han, Z., Jianxin, P., and Hu, W. (2010). Our wormy world genomics, proteomics and transcriptomics in East and southeast Asia. *Adv. Parasitol.* **73**, 327-371
 29. Cantacessi, C., Campbell, B. E., and Gasser, R. B. (2012). Key strongylid nematodes of animals – impact of next-generation transcriptomics on systems biology and biotechnology. *Biotechnol. Adv.* **30**, 469-488
 30. Verjovski-Almeida, S., and DeMarco, R. (2008). Current developments in *Schistosoma* proteomics. *Acta Trop.* **108**, 183-185
 31. Mulvenna, J., Hamilton, B., Nagaraj, S. H., Smyth, D., Loukas, A., and Gorman, J. J. (2009). Proteomics analysis of the excretory/secretory component of the blood-feeding stage of the hookworm, *Ancylostoma caninum*. *Mol. Cell. Proteomics* **8**, 109-121
 32. Young, N. D., Campbell, B. E., Hall, R. S., Jex, A. R., Cantacessi, C., Laha, T., Sohn, W. M., Sripa, B., Loukas, A., Brindley, P. J., and Gasser, R. B. (2010). Unlocking the transcriptomes of two carcinogenic parasites, *Clonorchis sinensis* and *Opisthorchis viverrini*. *PLoS Negl. Trop. Dis.* **4**, e719
 33. Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-59
 34. Schulz, M. H., Zerbino, D. R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* in press
 35. Li, R., Yu, C., Li, Y., Lam, T. W., Yiu, S. M., Kristiansen, K., and Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**, 1966-1967
 36. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* **5**, 621-628
 37. Cantacessi, C., Jex, A. R., Hall, R. S., Young, N. D., Campbell, B. E., Joachim, A., Nolan, M. J., Abubucker, S., Sternberg, P. W., Ranganathan, S., Mitreva, M., and

- Gasser, R. B. (2010). A practical, bioinformatic workflow system for large data sets generated by next-generation sequencing. *Nucleic Acids Res.* **38**, e171
38. Young, N. D., Jex, A. R., Li, B., Liu, S., Yang, L., Xiong, Z., Li, Y., Cantacessi, C., Hall, R. S., Xu, X., Chen, F., Wu, X., Zerlotini, A., Oliveira, G., Hofmann, A., Zhang, G., Fang, X., Kang, Y., Campbell, B. E., Loukas, A., Ranganathan, S., Rollinson, D., Rinaldi, G., Brindley, P. J., Yang, H., Wang, J., Wang, J., and Gasser, R. B. (2012). Whole-genome sequence of *Schistosoma haematobium*. *Nat. Genet.* **44**, 221–225
 39. The *Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium. (2009). The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature* **460**, 345–351
 40. Berriman, M., Haas, B. J., LoVerde, P. T., Wilson, R. A., Dillon, G. P., et al. (2009). The genome of the blood fluke *Schistosoma mansoni*. *Nature* **460**, 352–358
 41. Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., et al. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Res.* **37**, D211–D215
 42. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29
 43. Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L., and Wang, J. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* **34**, W293–W297
 44. Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L., Gao, G., Li, C. Y., and Wei, L. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **39**, W316–W322
 45. Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795.
 46. Sonnhammer, E. L. L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology, Menlo Park, CA, AAAI; pp.175–182.
 47. [Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L.](#) (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580
 48. Moller, S., Croning, M. D. R., and Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646–653
 49. Chen, Y., Zhang, Y., Yin, Y., Gao, G., Li, S., Jiang, Y, Gu, X., and Luo, J. (2005). SPD—a web based secreted protein database. *Nucleic Acids Res.* **33**, D169–D173
 50. Choo, K. H., Tan, T. W., and Ranganathan, S. (2005). SPdb – a signal peptide database. *BMC Bioinformatics* **6**, 249
 51. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–907
 52. Searle, B. C. (2010). Scaffold: a bioinformatics tool for validating MS/MS-based proteomics studies. *Proteomics* **10**, 1265–1269
 53. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392
 54. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658
 55. 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

56. Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S., and Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**, 1633-1649
57. Cho, P. Y., Lee, M. J., Kim, T. I., Kang, S. Y., and Hong, S. J. (2006). Expressed sequence tag analysis of adult *Clonorchis sinensis*, the Chinese liver fluke. *Parasitol. Res.* **99**, 602-608
58. Cho, P. Y., Kim, T. I., Whang, S. M., and Hong, S. J. (2008). Gene expression profile of *Clonorchis sinensis* metacercariae. *Parasitol. Res.* **102**, 277-282
59. Kang, J. M., Bahk, Y. Y., Cho, P. Y., Hong, S. J., Kim, T. S., Sohn, W. M., Na, B. K. (2010). A family of cathepsin F cysteine proteases of *Clonorchis sinensis* is the major secreted proteins that are expressed in the intestine of the parasite. *Mol. Biochem. Parasitol.* **170**, 7-16
60. Yoo, W. G., Kim, D. W., Ju, J. W., Cho, P. Y., Kim, T. I., Cho, S. H., Choi, S. H., Park, H. S., Kim, T. S., and Hong, S. J. (2011). Developmental transcriptomic features of the carcinogenic liver fluke, *Clonorchis sinensis*. *PLoS Negl. Trop. Dis.* **5**, e1208
61. Kim, T. S., de Guzman, J. V., Kong, H. H., and Chung, D. I. (2006). Comparison of gene representation between diploid and triploid *Paragonimus westermani* by expressed sequence tags analyses. *J. Parasitol.* **92**, 803-816
62. Horton, P., Park, K., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* **35**, W585-W587
63. Jefferies, J. R., Campbell, A. M., vanRossum, A. J., Barrett, J., and Brophy, P. M. (2001). Proteomic analysis of *Fasciola hepatica* excretory-secretory products. *Proteomics* **1**, 1128-1132
64. Morphew, R. M., Wright, H. A., LaCourse, E. J., Porter, J., Barrett, J., Woods, D. J., and Brophy, P. M. (2011). Towards delineating functions within the *Fasciola* secreted cathepsin I protease family by integrating in vivo based sub-proteomics and phylogenetics. *PLoS Negl. Trop. Dis.* **5**, e937
65. Robinson, M. W., Tort, J. F., Lowther, J., Donnelly, S. M., Wong, E., Xu, W., Stack, C. M., Padula, M., Herbert, B., and Dalton, J. P. (2008). Proteomics 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
66. Wilson, R. A., Wright, J. M., de Castro-Borges, W., Parker-Manuel, S. J., Dowle, A. A., Ashton, P. D., Young, N. D., Gasser, R. B., and Spithill, T. W. (2011). Exploring the *Fasciola hepatica* tegument proteome. *Int. J. Parasitol.* **41**, 1347-1359
67. Zheng, M., Hu, K., Liu, W., Hu, X., Hu, F., Huang, L., Wang, P., Hu, Y., Huang, Y., Li, W., et al. (2011). Proteomic analysis of excretory secretory products from *Clonorchis sinensis* adult worms: molecular characterization and serological reactivity of excretory-secretory antigen-fructose-1, 6-bisphosphatase. *Parasitol. Res.* **1**, 1-8
68. Mulvenna, J., Sripa, B., Brindley, P. J., Gorman, J., Jones, M. K., Colgrave, M. L., Jones, A., Nawaratna, S., Laha, T., Suttiaprapa, S., Smout, M. J., and Loukas, A. (2010). The secreted and surface proteomes of the adult stage of the carcinogenic human liver fluke *Opisthorchis viverrini*. *Proteomics* **10**, 1063-1078
69. Robinson, M. W., Corvo, I., Jones, P. M., George, A. M., Padula, M. P., To, J., Cancela, M., Rinaldi, G., Tort, J. F., Roche, L., and Dalton J. P. (2011). Collagenolytic activities of the major secreted cathepsin L peptidases involved in the virulence of the helminth pathogen, *Fasciola hepatica*. *PLoS Negl. Trop. Dis.* **4**, e1012
70. Nickel, W. (2003). The mystery of nonclassical protein secretion. A current view on cargo proteins and potential export routes. *Eur. J. Biochem.* **270**, 2109-2119

71. 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
72. 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
73. Turk, D., and Guncar, G. (2003). Lysosomal cysteine proteases (cathepsins): promising drug targets. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 203-213
74. Robinson, M. W., Dalton, J. P., Donnelly, S. (2008). Helminth pathogen cathepsin proteases: it's a family affair. *Trends Biochem.* **33**, 601-608
75. 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
76. Tort, J., Brindley, P. J., Knox, D., Wolfe, K. H., and Dalton, J. P. (1999). Proteinases and associated genes of parasitic helminths. *Adv. Parasitol.* **43**, 161-266
77. Dalton, J. P., Brindley, P. J., Donnelly, S., and Robinson, M. W. (2009). The enigmatic asparaginyl endopeptidase of helminth parasites. *Trends Parasitol.* **25**, 59-61
78. Hall, S. L., Braschi, S., Truscott, M., Mathieson, W., Cesari, I. M., and Wilson R. A. (2011). Insights into blood feeding by schistosomes from a proteomic analysis of worm vomitus. *Mol. Biochem. Parasitol.* **179**, 18-29
79. Liu, F., Cui, S. J., Hu, W., Feng, Z., Wang, Z. Q., and Han, Z. G. (2008). Excretory/secretory proteome of the adult developmental stage of human blood fluke, *Schistosoma japonicum*. *Mol. Cell. Proteomics* **8**, 1236-1251
80. Dalton, J. P., Skelly, P., and Halton, D. W. (2004). Role of the tegument and gut in nutrient uptake by parasitic platyhelminths. *Can. J. Zool.* **24**, 211-232
81. MacKenzie, A., Wilson, H. L., Kiss-Toth, E., Dower, S. K., North, R. A., and Surprenant, A. (2001). Rapid secretion of interleukin-1 β by microvesicle shedding. *Immunity* **15**, 825-835
82. 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
83. Salazar-Calderón, M., Martin-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
84. Salazar-Calderón, M., Martin 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
85. 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
86. 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
87. 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
88. Donnelly, S., Stack, C. M., O'Neill, S. M., Sayed, A. A., Williams, D. L., and Dalton, J. P. (2008). Helminth 2-Cys peroxiredoxin drives Th2 responses through a mechanism involving alternatively activated macrophages. *FASEB J.* **22**, 4022-4032
89. Wang, X., Chen, W., Huang, Y., Sun, J., Men, J., Liu, H., Luo, F., Guo, L., Lv, Z., Deng, C., Zhou, C., Fan, Y., Li, X., Huang, L., Hu, Y., Liang, C., Hu, X.,

Xum, J., and Yum, X. (2011). The draft genome of the carcinogenic human liver fluke *Clonorchis sinensis*. *Genome Biol.* **12**, R107

TABLE I.- *Summary of the nucleotide sequence data for adult Fascioloides magna prior and following assembly, and detailed bioinformatic annotation and analyses.*

<i>Raw reads (paired-end)</i>	Contigs (average length \pm SD; min–max length)
	GC content (%)
	Raw reads mapped to contigs (%)
Containing an Open Reading frame (%)	
With homologues in <i>Fasciola hepatica</i> (%)	
	<i>Fasciola gigantica</i>
	<i>Clonorchis sinensis</i>
	<i>Opisthorchis viverrini</i>
	<i>Schistosoma mansoni</i>
	<i>Schistosoma japonicum</i>
	<i>Schistosoma haematobium</i>
Returning InterProScan results (%)	
	<i>Number of InterPro terms</i>
Gene Ontology (%)	
	<i>Number of Biological process terms (level 2)</i>
	<i>Cellular component</i>
	<i>Molecular function</i>
Returning a KOBAS result (%)	
Number of predicted biological pathways (KEGG)	
Predicted proteins with signal peptides (%)	
	with transmembrane domains (%)
Homologous to proteins in the SPD database (%)	

TABLE II - *Numbers of cathepsin-encoding transcripts and corresponding proteins identified in the transcriptome and excretory/secretory products of Fascioloides magna, respectively.*

	No. of transcripts	No. of proteins
Cathepsin L1	18	4
Cathepsin L-like protease	2	1
Cathepsin L3	3	0
Cathepsin B2	18	2
Cathepsin B-like protease	1	0
Cathepsin A	9	1
Cathepsin D	1	0
Totals	70	8

FIGURE LEGENDS

FIG 1. *Gene Ontology*. Bar graph displaying the Gene Ontology (GO) terms (according to the categories ‘cellular component’, ‘molecular function’ and ‘biological process’) assigned to peptides inferred from transcriptomic data for *Fascioloides magna* and plotted using the WEGO tool (43).

FIG. 2. *Transcription profiling*. Bar graph showing the 25 most abundant proteins encoded in the transcriptome of *Fascioloides magna*.

FIG. 3. *Overview of proteins identified in the excretory/secretory (ES) products of Fascioloides magna*. *A*, Functional categories of the 80 proteins identified in the ES products of *F. magna*. Anti-oxidant proteins represented the largest functional group identified, followed by proteases, proteins involved in carbohydrate metabolism and protease inhibitors. Four proteins could not be assigned to any functional category, whereas ~25% were assigned to miscellaneous functional categories. *B*, Putative subcellular localizations of the identified proteins. Locations were predicted using PSORT (62) and based on evidence from the peer-reviewed literature. *C*, Proteases identified in the ES of *F. magna*. Abbreviations: SC - number of unique peptides assigned to the protein; CO - percent cover; emPAI - the exponentially modified protein abundance index; LO - the putative subcellular localization of the protein; the presence of a signal peptide (SP) or an N-linked glycosylation site (N) is denoted by a ‘+’; the presence of the protein in the ES of adult *F. hepatica* (*Fh*), *C. sinensis* (*Cs*) or *O. viverrini* (*Ov*) is also denoted by a ‘+’. *D*. Venn diagram depicting the number of similar proteins identified in the ES of adult *F. magna* (*Fm*), *F. hepatica* (*Fh*), *O. viverrini* (*Ov*) and *C. sinensis* (*Cs*).

FIG. 4. *Cathepsin L1*. Sequence alignment of four cathepsin L1 isoforms and a cathepsin L-like protease (39930) identified in the excretory/secretory (ES) products of *Fascioloides magna*. Peptides (assigned using Mascot; <http://www.matrixscience.com>) representing each isoform or multiple isoforms are highlighted in green and red, respectively. Each amino acid sequence corresponding to the mature protein is delimited by a solid black line.

LEGENDS TO SUPPLEMENTARY MATERIAL

FIG. S1 - *Cathepsins B*. Sequence alignment of two cathepsin B isoforms identified in the excretory/secretory (ES) products of *Fascioloides magna*. Peptides identified using Mascot (<http://www.matrixscience.com>) are highlighted in green; the arrow indicates the site of pro-protein cleavage of cathepsin B.

TABLE SI - *Biological pathways*. Predicted biological pathways linked to molecules encoded in the transcriptome of the adult stage of *Fascioloides magna*.

TABLE SII - *Signal peptides*. List of transcripts in the adult stage of *Fascioloides magna* predicted to encode a signal peptide.

TABLE SIII - *Secreted proteins*. Excretory/secretory (ES) proteins of *Fascioloides magna*, identified using LC-MS/MS.