


RESEARCH ARTICLE

Open Access



Complex and dynamic transcriptional changes allow the helminth *Fasciola gigantica* to adjust to its intermediate snail and definitive mammalian hosts

Xiao-Xuan Zhang^{1,2†}, Krystyna Cwiklinski^{3*†} , Rui-Si Hu¹, Wen-Bin Zheng¹, Zhao-An Sheng⁴, Fu-Kai Zhang¹, Hany M. Elsheikha⁵, John P. Dalton^{3*} and Xing-Quan Zhu^{1*}

Abstract

Background: The tropical liver fluke, *Fasciola gigantica* causes fasciolosis, an important disease of humans and livestock. We characterized dynamic transcriptional changes associated with the development of the parasite in its two hosts, the snail intermediate host and the mammalian definitive host.

Results: Differential gene transcription analysis revealed 7445 unigenes transcribed by all *F. gigantica* lifecycle stages, while the majority ($n = 50,977$) exhibited stage-specific expression. Miracidia that hatch from eggs are highly transcriptionally active, expressing a myriad of genes involved in pheromone activity and metallopeptidase activity, consistent with snail host finding and invasion. Clonal expansion of rediae within the snail correlates with increased expression of genes associated with transcription, translation and repair. All intra-snail stages (miracidia, rediae and cercariae) require abundant cathepsin L peptidases for migration and feeding and, as indicated by their annotation, express genes putatively involved in the manipulation of snail innate immune responses. Cercariae emerge from the snail, settle on vegetation and become encysted metacercariae that are infectious to mammals; these remain metabolically active, transcribing genes involved in regulation of metabolism, synthesis of nucleotides, pH and endopeptidase activity to assure their longevity and survival on pasture. Dramatic growth and development following infection of the mammalian host are associated with high gene transcription of cell motility pathways, and transport and catabolism pathways. The intra-mammalian stages temporally regulate key families of genes including the cathepsin L and B proteases and their trans-activating peptidases, the legumains, during intense feeding and migration through the intestine, liver and bile ducts. While 70% of the *F. gigantica* transcripts share homology with genes expressed by the temperate liver fluke *Fasciola hepatica*, gene expression profiles of the most abundantly expressed transcripts within the comparable lifecycle stages implies significant species-specific gene regulation.

(Continued on next page)

* Correspondence: krystyna.cwiklinski@nuigalway.ie;

johnpius.dalton@nuigalway.ie; xingquanzhu1@hotmail.com

[†]Xiao-Xuan Zhang and Krystyna Cwiklinski contributed equally to this work.

³National Centre for Biomedical and Engineering Science (NCBES), School of Natural Sciences, National University of Ireland, Galway, Ireland

¹State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou, Gansu Province 730046, People's Republic of China

Full list of author information is available at the end of the article



(Continued from previous page)

Conclusions: Transcriptional profiling of the *F. gigantica* lifecycle identified key metabolic, growth and developmental processes the parasite undergoes as it encounters vastly different environments within two very different hosts. Comparative analysis with *F. hepatica* provides insight into the similarities and differences of these parasites that diverged > 20 million years ago, crucial for the future development of novel control strategies against both species.

Keywords: *Fasciola gigantica*, Different developmental stages, Host-pathogen interaction, RNA-sequencing, De novo Transcriptome,

Background

Fasciolosis is an economically important disease of livestock that affects more than 300 million animals (cattle, sheep, water buffalo and goats) [1, 2]. It is recognized by the WHO as an important neglected tropical disease with estimates of between 2.4 to 17 million people infected [3]. The causative agents of this disease are the helminth parasites, *Fasciola gigantica* and *Fasciola hepatica*, commonly referred to as tropical and temperate liver flukes, respectively. *F. gigantica* is widely distributed throughout Asia and Africa [4]. Global economic losses attributed to fasciolosis are estimated to be more than US \$3 billion per year [5, 6], while specific losses driven by *F. gigantica* infection are estimated to be in the region of US \$ 2.4 billion and US \$ 0.84 billion in Asia and Africa, respectively [1].

The *F. gigantica* lifecycle requires two hosts, a snail intermediate host and a mammalian definitive host. Eggs are passed in faeces from the mammalian host into the environment where embryonation is initiated and the parasites develop into miracidia. Following activation by temperature and light, miracidia hatch from the eggs and seek out and infect the snail intermediate host. *F. gigantica* infects a wide range of snail species, including the Eurasian species *Radix auricularia* and *Austropeplea viridis*, and the African species *Radix natalensis* [7]. Within the snail, the parasite progresses through several developmental stages, namely sporocysts, rediae and cercariae, by a process of clonal expansion [8, 9]. Studies by Dinnik and Dinnik [10–12] have shown that *F. gigantica* undergoes five redial generations that result in large numbers of cercariae emerging from the snail. The cercariae settle on vegetation, or remain free flowing in water, and become encysted in a double-layered weather-resistant cyst. These encysted metacercariae infect the mammalian definitive host, typically cattle and buffalo, following their ingestion with contaminated vegetation or water.

The metacercariae double-layered cyst protects the parasite on pasture and ensures its survival following ingestion. Acid proteases from the stomach/rumen remove the outer cyst wall and then following activation at neutral

pH by reducing conditions and bile salts within the duodenum the parasites excyst as newly excysted juveniles (NEJ). The NEJ rapidly migrate through the intestinal wall into the liver parenchyma, where they feed extensively on blood and liver tissue to facilitate their dramatic growth and development. The parasites move to the bile ducts, become sexually mature adults and release approximately 20,000 eggs per fluke per day [13] to continue the lifecycle.

A larger proportion of floating metacercariae cysts have been observed for *F. gigantica*, implying that water is an important source of infection in addition to encysted metacercariae on vegetation for *F. gigantica* [5, 14]. Comparative analysis of the maturation and development times of the parasites in the snail host indicates that as a consequence of higher temperatures *F. gigantica* develops more rapidly than its temperate counterpart *F. hepatica* [14]. Furthermore, studies of the infectivity of *F. gigantica* versus *F. hepatica* have shown that certain ruminant species/breeds display different susceptibilities to the two species. For example, Indonesian Thin Tailed sheep have been shown to be highly resistant to *F. gigantica* but susceptible to *F. hepatica* [15]. In contrast, Japanese Saanen goats are more susceptible to *F. gigantica* than to *F. hepatica* [16]. For ruminant species, such as Guirra sheep, that display similar susceptibilities to infection to both species, *F. gigantica* was found to be more pathogenic than *F. hepatica* [17]. The ability of *F. hepatica* to infect a vast range of mammals, including rodents, lagomorphs, ungulates, ruminants, marsupials, camelids and primates [18], suggests universal processes of invasion, virulence and immune modulation that may also play an important role for *F. gigantica*.

In the present study, we have used RNASeq to profile the complex and dynamic transcriptional changes that *F. gigantica* undergoes throughout its lifecycle, when the parasite is exposed to contrasting environmental and host-specific environments. Parsing the data by host, snail intermediate host versus mammalian host revealed that the parasite transcribes distinct subsets of gene families to interact and manipulate each host. Comparative analysis carried out using the available *F. hepatica* datasets highlighted the high level of homology between the

tropical and temperate liver flukes while, at the same time, reveals distinct patterns of gene regulation. These new insights into *Fasciola* spp. biology are critical for the future understanding of developmental biology of these major global and economically important parasites and will assist in the development of new vaccine and drug control strategies.

Results and discussion

Sequence assembly and functional annotation

Like all digenean trematodes, *F. gigantica* requires two hosts to facilitate its lifecycle, (1) a snail intermediate host where asexual reproduction results in a clonal expansion of the parasite and, (2) a mammalian definitive host where sexual reproduction occurs with the production of large numbers of eggs. Environmental free-living stages link the two hosts; miracidia that emerge from the eggs and immediately seek and penetrate the snail host, and cercariae that rupture from the snail, settle on vegetation or float on water and secrete a coating to form encysted metacercariae that are infective when ingested by the mammalian host.

To gain insight into how this parasite develops and adapts to these various conditions, we carried out transcriptome analysis of the eight lifecycle stages; eggs, miracidia, rediae, cercariae, metacercariae, liver stage juvenile parasites 42 days post infection (dpi), liver stage juvenile parasites 70 dpi and mature adult fluke taken from the bile ducts (Fig. 1). An average of 54 million high quality reads

were generated for each lifecycle stage, which were used for de novo assembly, resulting in 212,790 transcripts (Table 1). The N50 contig size was 1552 nucleotide base pairs, which is consistent with other helminth transcriptomes (*Schistosoma turkestanicum*: 1480 bp [19]; *Eurytrema pancreaticum*: 1094 bp [20]).

The sequencing reads were re-mapped to the stage-specific transcriptome datasets to determine the transcription of the clustered unigenes, and resulted in a total of 58,422 unigenes based on at least 1 fragment per kilobase per million reads (FPKM) in all three replicates across the eight transcriptomes. Analysis of differential transcription revealed that 7445 unigenes were expressed by all eight lifecycle stages (Fig. 2) while the majority ($n = 50,977$) exhibited stage-specific expression. Stage specific expression was observed for each transcriptome dataset (egg, 3372; miracidia, 5646, rediae, 1093; cercariae, 1829; metacercariae, 9605; juvenile-42dpi, 856; juvenile-70dpi, 873; adult, 2192; Fig. 2). Functional annotation of the 58,422 unigene dataset was carried out using seven in silico databases, resulting in 59.2% ($n = 34,600$) of the dataset being annotated by at least one database (Table 2).

Complexities of the *F. gigantica* lifecycle

Eggs and miracidia

Eggs passed in the faeces of the mammalian host are undeveloped and require a period of embryonation on

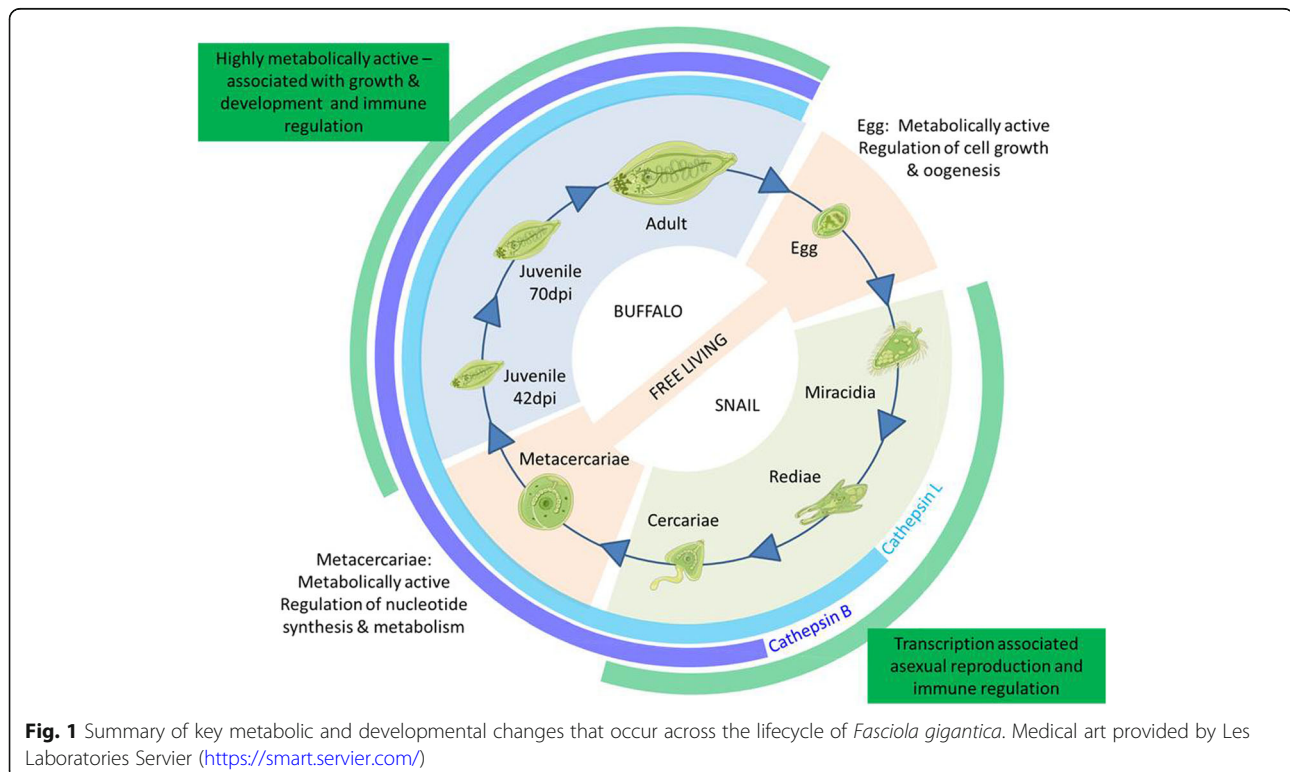


Table 1 Summary of *Fasciola gigantica* lifecycle stage-specific transcriptome sequence assembly

	Egg	Miracidia	Rediae	Cercariae	Metacercariae	Juvenile-42 dpi	Juvenile-70 dpi	Adult
Raw reads	64061584	47847820	48250358	73797162	49174724	49275028	54126804	56344306
Clean reads	60943084	47847820	46199344	70605896	47239548	46714058	54126804	52842820
GC content	45.62	44.92	44.08	43.37	44.90	45.51	45.04	45.96
Q20 percentage	96.83	95.03	97.27	96.99	96.71	94.96	94.92	94.86
Total unigenes ^a	18197	28697	17854	25902	36293	17518	17187	15635

^aClustered unigenes filtered by > 1 FPKM across all three biological replicates for each lifecycle stage

pasture under suitable temperature and humidity conditions. Upon embryonation, hatching of the miracidia occurs in response to light and temperature cues [21]. Within the eggs, abundant transcription was observed for transcripts associated with calcium-dependent protein binding (GO:0048306), glycolipid transport and binding activity (GO:0017089, GO:0051861) and gluconate transmembrane transporter activity (GO:0015128). However, transcriptional analysis of the two stages reveals higher levels of transcription in the miracidial stage, with 16,202 transcripts showing stage-specific expression consistent with the high levels of parasite activity (Fig. 3); this difference was also reflected in the

higher number of unique gene ontology (GO) terms within the miracidia-specific transcripts (Fig. 3). Key GO terms associated with the miracidia including those related to pheromone activity (GO:0005186) and photosynthesis/light reaction (GO:0019684), and was reflective of the requirement of miracidia to seek out the snail host via light and chemical stimuli [8].

Invasion of the snail host involves penetration and migration through the snail tissues, using both mechanical and proteolytic means [8]. Correlating with the GO enrichment of transcripts relating to catalytic activity upregulated in the miracidia (GO:0003824; Additional file 1: Table S1), a number of transcripts were annotated with the GO term metalloproteinase activity (GO:0008237), which could play a role in the breakdown of snail tissues to facilitate parasite entry. In addition, several GO terms related to extracellular vesicles (EVs) were observed (endocytosis, GO:0006897; vesicle docking involved in exocytosis, GO:0006904) indicating that the miracidia may also use EVs to deliver a cargo of proteases and other molecules for invasion and interaction with the snail host. The GO term alpha-1,6-mannosylglycoprotein 2-beta-N-acetylglucosaminyltransferase activity (GO:0008455) was also found within the miracidia dataset, indicating an essential role in N-glycosylation of proteins in the miracidia which could suggest that parasite glycan-snail

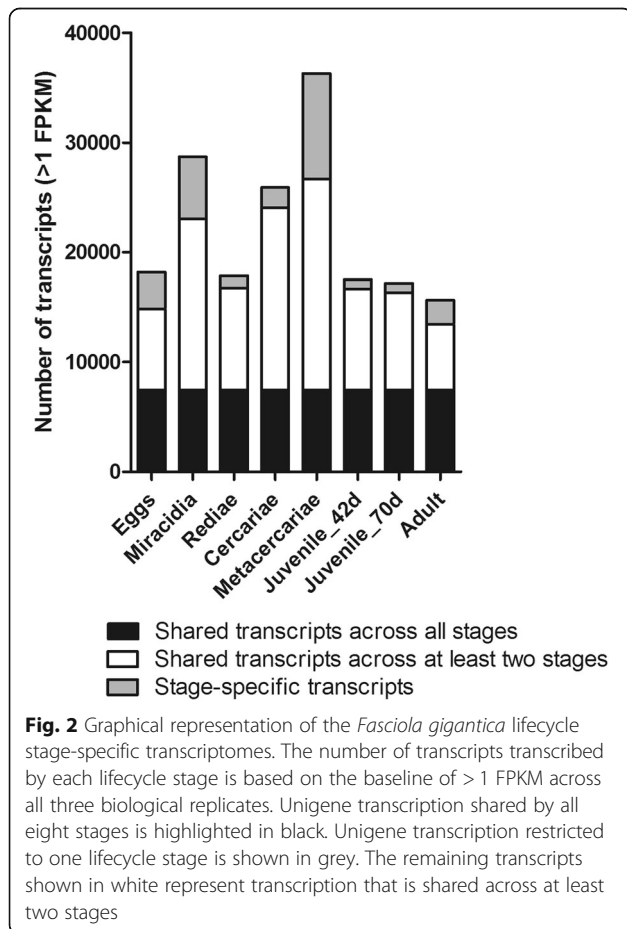
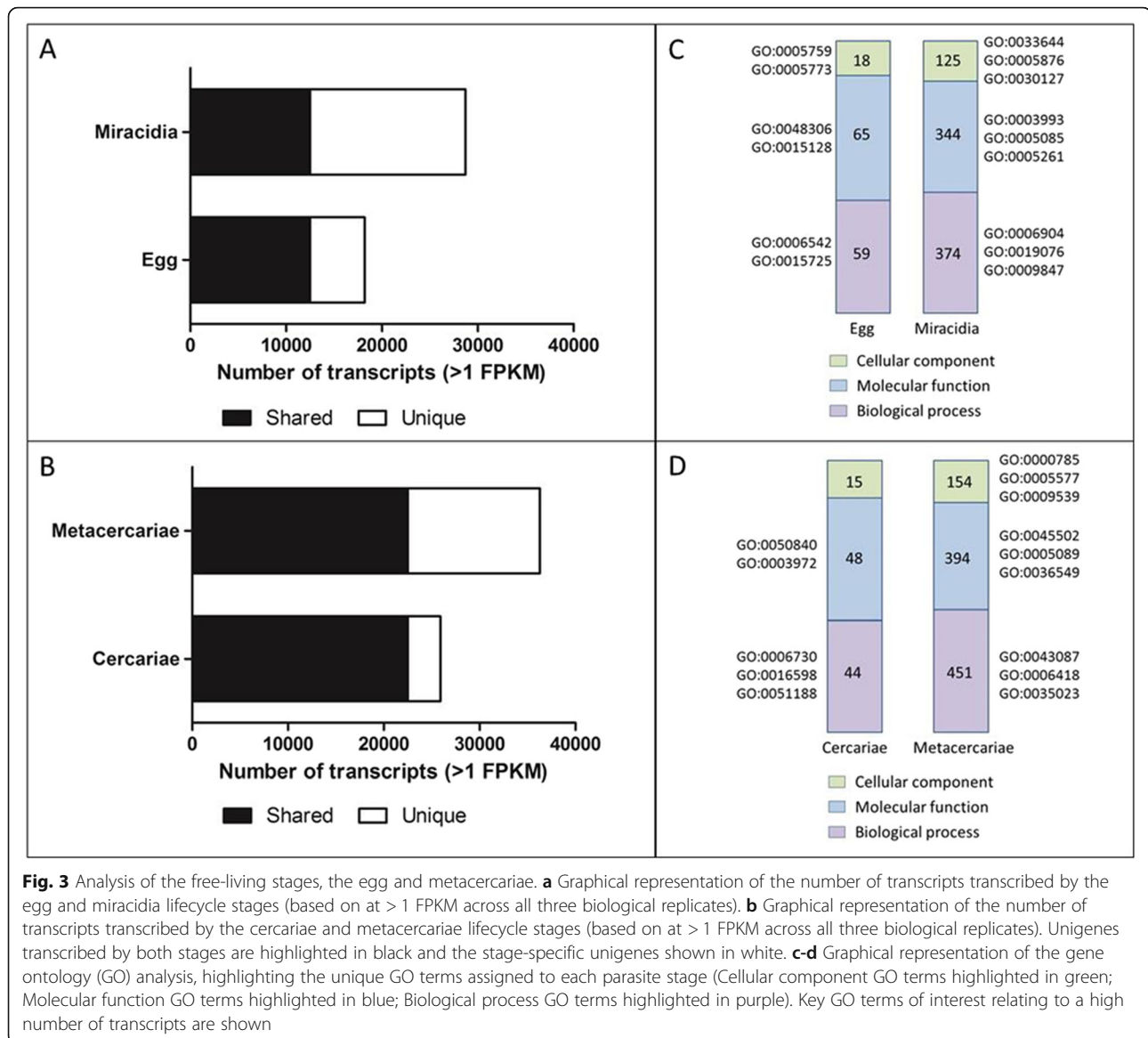


Table 2 Summary of in silico annotation of *Fasciola gigantica* unigenes (58,422; > 1 FPKM across three biological replicates)

In silico database	Number of transcripts
NCBI Non-redundant (Nr) database	30201 (51.7%)
NCBI Nucleotide (Nt) database	6673 (11.4%)
KEGG (KO) database	11165 (19.1%)
PFAM database	22591 (38.7%)
Gene Ontology	22633 (38.7%)
Biological Process	16554 (28.3%)
Cellular Component	12067 (20.7%)
Molecular Function	18658 (31.9%)
SwissProt database	18783 (32.1%)
KOG database	12316 (22.8%)
Annotated by all databases	1326 (2.3%)
Annotated by at least one database	34600 (59.2%)



lectin interactions are involved in the transformation of miracidium to sporocyst within the snail, as shown in studies of snail interactions with *F. hepatica* [22] and *Schistosoma mansoni* [23].

Cercariae and metacercariae

Comparative analysis of the cercariae removed from snails 2 days before shedding with the free-living encysted metacercariae, revealed expression of a greater number of transcripts ($n=36,293$) associated with the metacercariae and a corresponding greater number of assigned GO terms ($n = 13,797$) demonstrating dramatic changes in gene expression profile (Fig. 2). GO enrichment analysis of the shared transcripts revealed a down-regulation of GO terms associated with nucleotide synthesis and metabolism in the metacercariae consistent with a lower level of metabolic activity (Additional

file 1: Table S1). However, since several GO terms were associated with metabolism this indicates that the metacercariae are metabolically active and not completely dormant. Furthermore, 33 biological process GO terms associated with regulation were identified in metacercariae; indicative of their longevity in the external environment this stage expresses genes that regulate (a) gene transcription and translation (transcription from RNA polymerase III promoter, GO:0006359; DNA replication initiation, GO:0032297; DNA transcription, elongation, GO:0032784; translation initiation, GO:0045947), (b) protein phosphorylation (protein kinase activity, GO:0045859), and (c) signal transduction (GO:0009966; ARF protein signal transduction, GO:0032012; Rho protein signal transduction, GO:0035023; GTPase activity, GO:0043087; small GTPase mediated signal transduction, GO:0051056). The encysted parasites appear to maintain

their low metabolic state by regulating the pH (GO:0006885) as well reducing their endopeptidase activity (GO:0010951) to avoid autolytic processes.

Parsing snail and mammalian host transcriptome data

The transcriptome data was parsed according to infected host, i.e. stages infecting the snail intermediate host

(miracidia, reidae, cercariae) were compared with those in the buffalo definitive mammalian host (juvenile 42 dpi, juvenile 70 dpi, adult) (Fig. 4). A third of the transcripts were transcribed by all the stages within the snail (33.3%; $n = 12, 850 / 38,510$). By contrast, the number of shared transcripts between the buffalo-specific stages represented approximately half of the gene transcripts

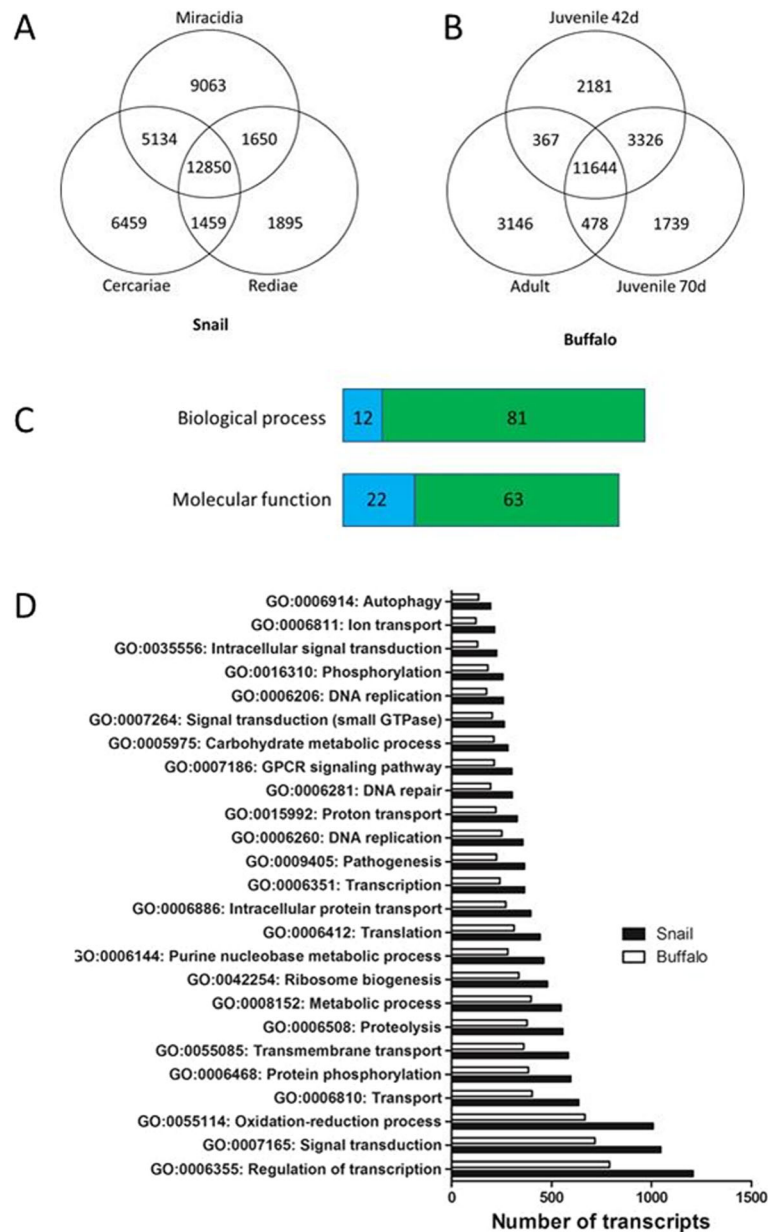


Fig. 4 Analysis of host-associated transcription of *Fasciola gigantica* transcripts. **a** Venn diagram representation of the number of transcripts transcribed by the snail intermediate host associated lifecycle stages, namely miracidia, reidae and cercariae. **b** Venn diagram representation of the number of transcripts transcribed by the buffalo mammalian definitive host associated lifecycle stages, namely juvenile 42 dpi, juvenile 70 dpi and adult parasites. **c** Graphical representation of the number of unique gene ontology (GO) terms related to biological process and molecular function transcribed by the snail associated parasite stages, highlighted in green, and the buffalo associated parasite stages, highlighted in blue. **d** Graphical representation of the number of genes annotated with biological process associated GO terms. The terms with the most transcripts assigned are shown. The snail associated parasite stages, shown in black, are compared with the buffalo associated stages, shown in white

(50.9%; $n = 11,644/22,881$), reflecting a greater similarity in gene expression between stages in the mammalian stages compared to those within the snail.

Analysis by gene ontology identified the same unique cellular component GO terms in both datasets ($n = 335$). In total 1113 unique biological processes GO terms, and 964 molecular function GO terms were assigned across both datasets. Analysis of those shared by both datasets revealed more genes assigned per GO term within the snail dataset compared with the buffalo dataset (Fig. 4). In addition, 93 biological processes GO terms (snail = 81; buffalo = 12) and 85 molecular function GO terms (snail = 63; buffalo = 22) were associated to only one of the datasets and, therefore, specific to the association with the respective host (Additional file 2: Table S2). The number of unique GO terms particular to the snail associated stages, reflects major biological changes that occur during rapid asexual development in the snail (system development, GO:0048731; tissue development, GO:0009888; regulation of sporulation, GO:0042173; actin cytoskeleton reorganization, GO:0031532; dorsal/ventral pattern formation, GO:0009953). This contrasted sharply with the buffalo associated stages, where growth is the major developmental change.

Interestingly, several GO terms associated with immune response were identified within the intra-snail stages, including activation of innate immune response (GO:0002218), B cell receptor signaling pathway (GO:0050853), and lymphocyte activation (GO:0046649). This observation is consistent with studies showing that trematode parasites

ensure their survival by manipulating the snail's innate immune system [24–26]. RNASeq analysis of the interaction between *Opisthorchis viverrini* with its snail intermediate host, revealed enrichment of T and B cell signaling pathways and down-regulation of approximately 90% of the differentially expressed genes within infected snails compared with uninfected snail [24].

Lifecycle stage-specific metabolic regulation

Analysis of the transcription of the KEGG metabolic pathways reveals that *F. gigantica* displays distinct patterns of metabolic transcription throughout its lifecycle (Fig. 5). Several pathways were highly transcribed by the eggs relative to the other lifecycle stages. In particular, KEGG pathways associated with cell growth and regulation of oogenesis, namely cell cycle (ko04110), oocyte meiosis (ko04114) and p53 signalling pathway (ko04115), which likely play a role in the embryonation process of the eggs on pasture.

Overall, the miracidia show lower relative levels of transcription of the metabolic related genes, particularly relating to pathways involved in carbohydrate metabolism, amino acid metabolism, and transport and catabolism. This is consistent with the short timeframe of this non-feeding stage, as following egg hatching the miracidia must infect a snail within 24 h, with optimal infection success occurring within 2 h of hatching [8]. The swimming rate of the miracidia, which is a determinant for infectivity, is contingent on available glycogen energy stores that are not replenished during this process [8].

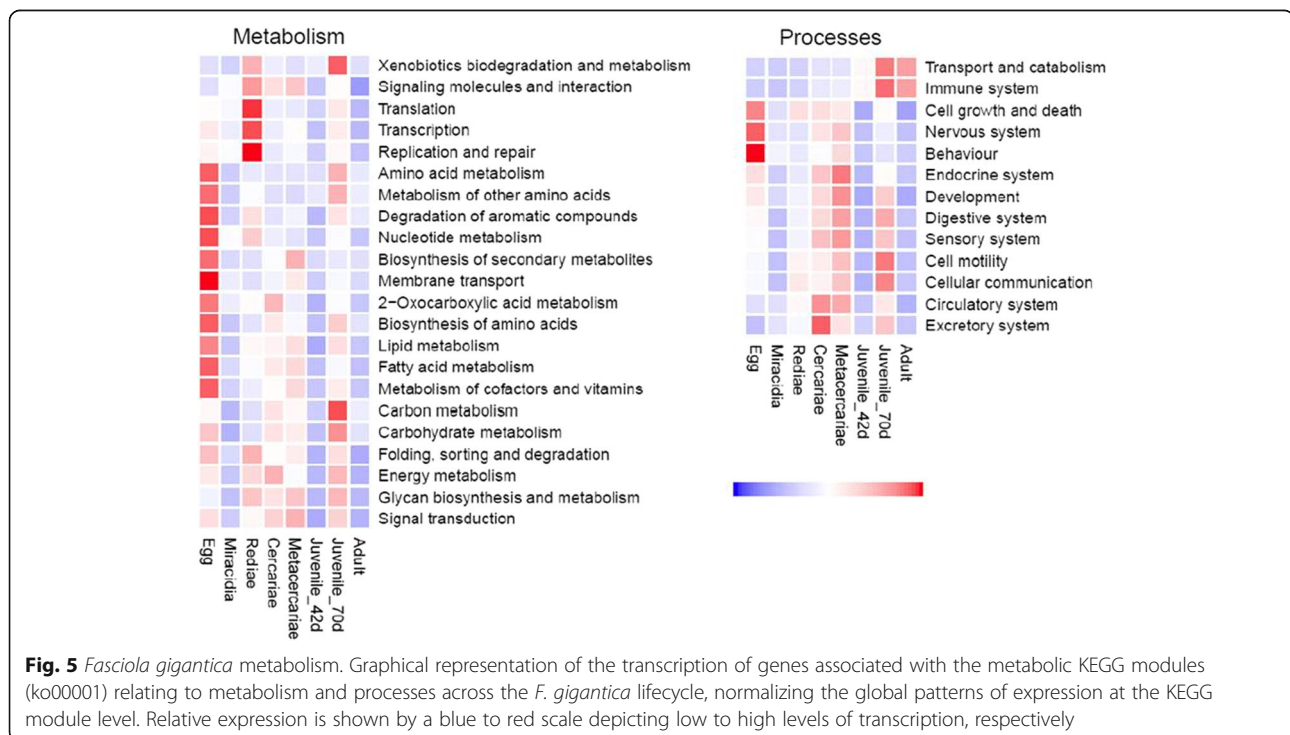


Fig. 5 *Fasciola gigantica* metabolism. Graphical representation of the transcription of genes associated with the metabolic KEGG modules (ko00001) relating to metabolism and processes across the *F. gigantica* lifecycle, normalizing the global patterns of expression at the KEGG module level. Relative expression is shown by a blue to red scale depicting low to high levels of transcription, respectively

Following invasion of the snail the miracidia transform into sporocysts from which the rediae develop. Clonal expansion from the mother redia ensues through daughter redial generations from which the cercariae finally develop [12]. Analysis of redial metabolism revealed that in comparison to the other lifecycle stages, the rediae demonstrated high levels of gene transcription, translation and replication/repair, a requirement for the development of multiple redial generations within the snail. Genes associated with the ribosome (ribosomal proteins); ribosome biogenesis (GTP binding nuclear protein Ran and H/ACA ribonucleoprotein associated protein) and the spliceosome (splicing factors and ribonucleoproteins) were amongst the most transcribed genes associated with transcription/translation. In addition, DNA-directed RNA polymerases, transcription initiation factors and translation initiation factors associated with RNA transport also contributed to the high levels of transcription.

The cercariae and metacercariae stages show similar levels of metabolic-related gene transcription. The data reveals that metacercariae regulate the rate of metabolic transcription to remain viable on pasture, which is consistent with similar observations of *F. hepatica* [27]. Overall, the cercariae and metacercariae show comparable levels of metabolic activity to the rediae based on the quantitative metabolic fingerprint values (QMF) of the KEGG metabolism and processes related pathways. However, key metabolic pathways are more abundantly transcribed by the rediae compared with cercariae and metacercariae; for example, gene transcription and translation was between 2.0–3.5 fold increased in the rediae compared with the cercariae and metacercariae. Conversely, a two-fold increase in expression of the transport and catabolism pathways was observed in the cercariae compared with the rediae. Similarly, the biosynthesis of secondary metabolites pathways and membrane transport associated pathways, were more highly expressed by the metacercariae compared to the rediae, at a six-fold and two-fold increased level of transcription, respectively.

Following ingestion by the mammalian host, *F. gigantica* parasites rapidly migrate into the liver parenchyma where feeding begins. In keeping with the rapid growth and development that takes place within the liver tissue, high levels of transcription were observed for the majority of KEGG pathways transcribed by the juvenile 70-dpi parasites. Consistent with the intense feeding and migratory activities of the juvenile parasites through the liver parenchyma, parasites at 42 dpi and 70 dpi transcribed high levels of genes associated with the lysosome and phagosome of the transport and catabolism pathways, and regulation of the actin cytoskeleton related to cell motility pathways. The genes included cathepsin L and cathepsin

B proteases, legumains and actin and tubulin genes. Similarly, these genes are also known to be involved in immune system processes, including antigen processing and presentation, platelet activation and leukocyte trans-endothelial migration, facilitating the parasite's evasion and manipulation of the host immune response. Glutathione S transferase (mu class), associated with xenobiotics biodegradation and metabolism and glutathione metabolism, was also observed as most abundantly transcribed by the juveniles at 42 dpi and 70 dpi.

In keeping with the increased Krebs cycle activity observed during parasite development that has been shown to be directly correlated to the surface area of the liver fluke parasite [28], pathways involved with carbohydrate metabolism, specifically glycolysis/gluconeogenesis and the citrate cycle (TCA cycle), were highly expressed. Glycan biosynthesis and metabolism pathways, particularly relating to N-glycan biosynthesis, were also observed within the juvenile parasite datasets, consistent with the predominant mannosylated N-glycans found on the outer surface of the *F. hepatica* NEJ and adult parasites [29–31].

Once settled within their final place of residence in the bile ducts, matured adult parasites showed lower relative levels of metabolic transcription compared with the juvenile liver stage parasites. However, the pathways associated with transport and catabolism, namely the phagosome and lysosome (ko04145, ko04142), and pathways associated with the immune system, such as antigen processing and presentation (ko04612), leukocyte trans-endothelial migration (ko04670) and Fc-gamma R mediated phagocytosis (ko04666), are highly transcribed by the adult parasites. Abundant transcription of these KEGG pathways is consistent with the parasites interacting and manipulating the host immune system driving a T helper (Th) 2-type immune response whilst promoting parasite survival through the suppression of pro-inflammatory Th1-type immune responses [32]. Chronic fasciolosis associated with the blood-feeding adult parasites within the bile ducts is characterized by the infiltration of immune cells, including eosinophils, lymphocytes and plasma cells, to the portal areas [33].

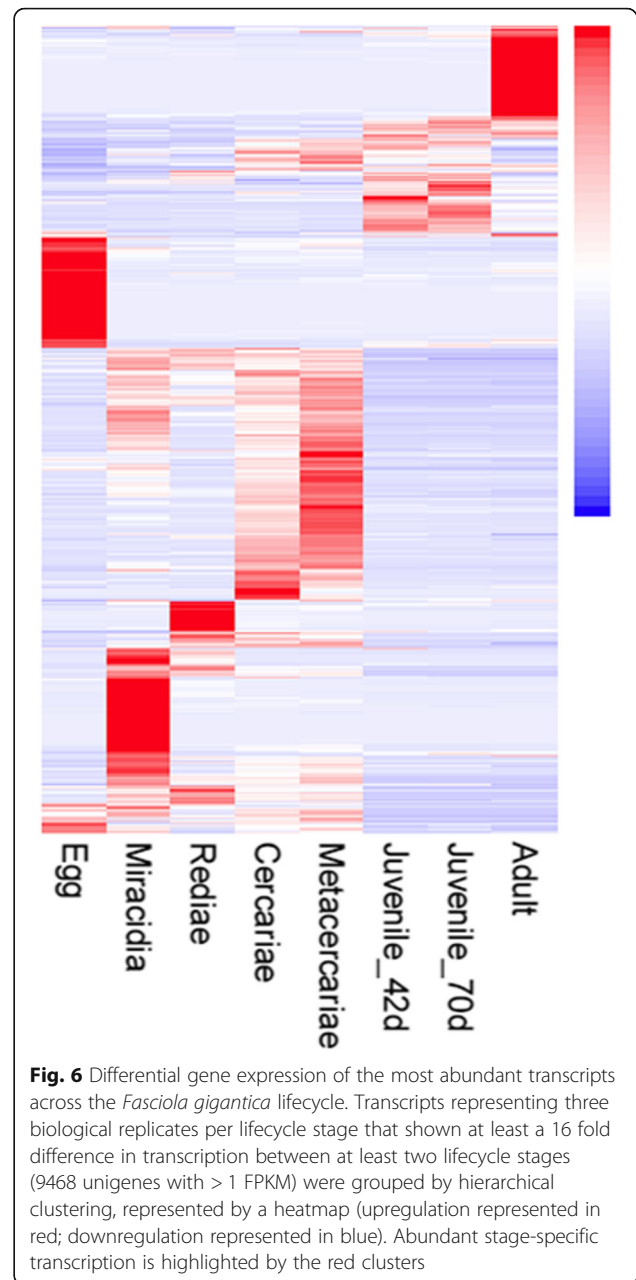
Differential gene expression during the lifecycle of *F. gigantica*

Co-expression cluster analysis of the 58,422 *F. gigantica* unigenes transcribed by the eight lifecycle stages generated 16 clusters that reflect the transitioning gene expression through the lifecycle (Additional file 3: Figure S1). The majority of clusters represented high gene expression at multiple lifecycle stages, with only two clusters (clusters 12 and 15) representing high gene transcription within one lifecycle stage, namely miracidia and adult, respectively. Functional analysis of these clusters revealed that despite different sets of genes being clustered together,

key molecular processes are enriched throughout the life-cycle and are present within multiple clusters. In particular, several GO terms were significantly enriched throughout the eight lifecycle stages associated with processes involving nucleic acids (GO:0000166;GO:0003723; GO:0003724; GO:0003676; GO:0003677; GO:0006310), transcription and translation (GO:0003743; GO:0006355; GO:0006446) and energy (ATP) regulation (GO:0005524; GO:0016887; Additional file 4: Table S3). Key GO terms associated with liver fluke biology, including catalytic activity (GO:0003824) and proteolysis (GO:0006508), were also significantly enriched in all eight lifecycle stages. Consistent with a recent study by McVeigh and colleagues of G protein-coupled receptors within the *F. hepatica* genome [34], the GO term G-protein coupled receptor signaling pathway (GO:0007186) was also highly enriched throughout the lifecycle.

In addition to the enriched processes observed throughout the lifecycle, several GO terms were uniquely enriched to particular clusters. Cluster 1 that represents high gene expression from the egg stage through to the metacercariae showed enrichment for metalloendopeptidase activity (GO:0004222) and vesicle-mediated transport (GO:0016192), consistent with our in depth analysis of the snail associated lifecycle stages. Cysteine-type endopeptidase activity (GO:0004197) and cysteine-type peptidase activity (GO:0008234) were significantly enriched within the highly expressed genes within the stages associated with the mammalian host (metacercariae through to adult fluke; cluster 5). By contrast, the only GO term significantly enriched within clusters 11 and 12 that show high levels of expression within the miracidia and cercariae is zinc ion binding (GO:0008270). Studies of the effects of heavy metals and pesticides by Falis and colleagues [35] have shown that zinc may have a protective effect on the survival of miracidia, which may reflect the high levels of genes associated with zinc ion binding. Similarly, the co-enrichment of zinc ion binding and metallo-endopeptidase activity within the eggs and miracidia is consistent with the presence of these zinc-dependent peptidases within these stages.

Deeper analysis across the lifecycle transcriptomes revealed that a total of 34,264 transcripts were significantly differentially expressed (based on at least 1 FPKM across three replicates; Additional file 5: Table S4). In particular, 9468 DEG were found to display at least a 16-fold difference in transcription between at least two of the lifecycle stages. Hierarchical clustering of these most abundant transcripts demonstrated stage-specific expression (Fig. 6), consistent with the developmental regulation of genes associated with metabolism. Analysis of 661 molecular function associated GO terms assigned to the 9468 DEG revealed high representation of GO terms associated with protein binding (GO:00005515), DNA binding (GO:0003677), ATP binding (GO:0005524) and ion binding



(GO0008270: zinc ion binding; GO:0046872: metal ion binding; GO:0005509: calcium ion binding) (Additional file 9: Table S7). In addition, key GO terms were highly represented within stage-specific clusters, e.g. GTPase activity (GO:0003924) and microtubule activity/binding (GO:0003777, GO:0008017) were highly represented within the adult parasites. Consistent with the known reliance of *Fasciola* spp. on cathepsin cysteine proteases for migration and feeding in the mammalian host, transcription for these were particularly abundant in the metacercariae and liver stages at 42dpi and 70dpi.

To further investigate how the parasites changes and develops through each lifecycle stage, we carried out

pairwise DEG analysis of the sequential lifecycle stages, as follows (Fig. 7); (1) egg versus miracidia, (2) miracidia versus rediae, (3) rediae versus cercariae, (4) cercariae versus metacercariae, (5) metacercariae versus juvenile 42 dpi, (6) juvenile 42 dpi versus juvenile 70 dpi and (7) juvenile 70 dpi versus adults. Gene ontology enrichment analysis revealed key processes were enriched within each pairwise analysis (Additional file 1: Table S1). As an example, cathepsin L and B, legumain, GST (sigma class) and protease inhibitors (cystatins, serpins and kunitz-type serine protease inhibitors) were observed to be amongst the most abundant unigenes transcribed by the rediae, cercariae, metacercariae, juvenile parasites and adults.

Transcripts upregulated in the rediae and cercariae were enriched for GO terms relating to binding, particularly protein binding, ATP binding, DNA binding, metal ion including zinc ion binding, indicative of a parasite that is metabolically highly active. Heat shock proteins, a range of protease inhibitors (serpins, cystatins and kunitz-type inhibitors) and egg related proteins (yolk ferritin and vitelline protein) were amongst the most upregulated unigenes transcribed by the eggs. Cubilin-related genes were the most represented genes within the miracidia upregulated unigenes.

Compared to the infective metacercarial stage, the juvenile 42dpi parasites displayed abundant transcription of genes relating to metabolic process (GO:0008152) and catalytic process (GO:0003824), reflecting the active feeding and growth of the juvenile parasites when they enter the liver. However, migrating juveniles at 42 dpi

and 70 dpi shared a similar transcriptional profile with no overall enrichment indicating little difference in the transcriptional activity of parasites that are slowly moving through the parenchyma mass through their tunneling activity and evading the infiltration of immune cells into damaged sites. Single genes annotated as cathepsin L, legumain, thioredoxin and glutathione S-transferase (GST) sigma class were identified, and are clearly relevant to the immunopathology associated with acute fasciolosis.

Consistent with recent analysis of the early migrating stages of *F. hepatica* [27], we observed expression and temporal regulation of a family of transmembrane water channels, the aquaporins, which facilitate the transport of water and other solutes. Gene transcripts were observed for both the aquaporins, that enable the passage of water, in addition to a subset of channels within this family that also allow the transportation of other solutes including urea and glycerol, termed the aquaglyceroporins based on the signature Froger's residues and the residues comprising the aquaporin aromatic/arginine selectivity filter [36–38]. These conserved residues were also comparable with those observed within the *F. hepatica* aquaporin sequences [27]. Differential transcription of these genes was observed across the lifecycle, with one subset of aquaporins being expressed almost exclusively by the free-living eggs and miracidia while other aquaporin genes were more highly expressed during the stages present within the mammalian host. Equally, the aquaglyceroporins genes were predominately expressed by the migrating juvenile (42dpi and 70dpi) and the adult parasites, with some genes also being transcribed by the rediae

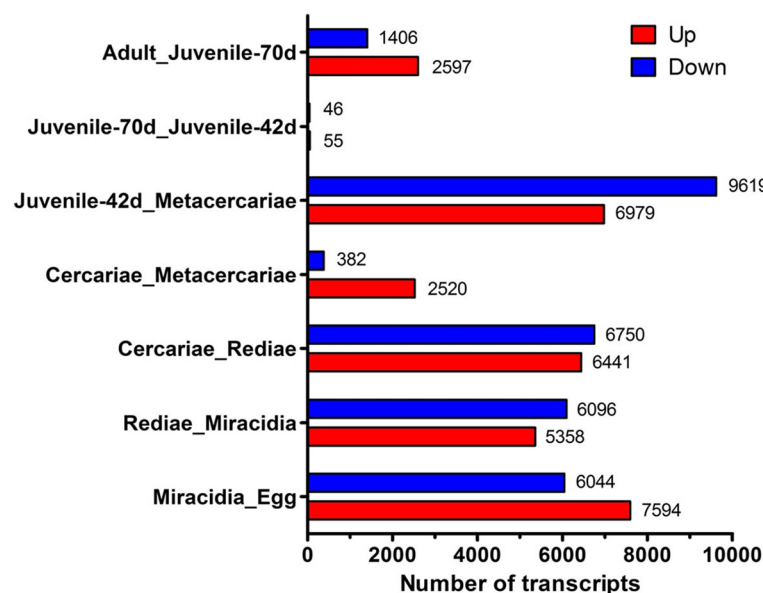


Fig. 7 Pairwise analysis of the eight *Fasciola gigantica* transcriptomes. Graphical representation of the number of statistically significant differentially expressed transcripts

stage. This regulation of surface transport molecules highlights a molecular system by which the parasites adjust to the varying osmotic conditions found between the external environments and the two different hosts the parasite infects.

Regulation of the complement membrane attack complex (MAC) CD59-like genes across the snail- and mammalian-associated stages was also observed, implying that the parasite uses host mimicry to evade and manipulate the immune systems of its hosts. Nine CD59-like genes have been identified within the *F. hepatica* genome [39] indicating that both *Fasciola* species may use similar methods to inhibit host complement.

Comparative analyses between *F. gigantica* and *F. hepatica*

Levels of homology between the available sequences for *F. gigantica* and *F. hepatica* were determined by reciprocal nucleotide BLAST analysis (*F. gigantica*, $n = 58,422$, present study; *F. hepatica*, $n = 22,676$, derived from the *F. hepatica* genome [40]). Homologous sequences were observed for approximately 71% of the *F. gigantica* transcripts ($n = 41,355$), which included 34,868 transcripts (~60% transcripts) that shared nucleotide homology at BLAST e value $< e^{-30}$. Our results are consistent with the analysis by Young and colleagues who compared the transcriptomes of adult *F. gigantica* and *F. hepatica* [41]. However, it is important to note that the gene models from the *F. hepatica* genome were determined based on transcriptome data from stages from the mammalian host and did not include potential snail-associated life-cycle stage-specific genes [40].

Further analysis was carried out between the lifecycle stages for which transcriptome data was available for both *Fasciola* species, namely egg, metacercariae and adult flukes. Gene ontology analysis of the respective transcriptomes revealed that approximately 900 GO terms were shared by the two *Fasciola* species (egg: 862; metacercariae: 931; adult: 903; Additional file 6: Table S5). However, the majority of significantly enriched GO terms showed species-specific GO enrichment namely within the *F. gigantica* datasets (Additional file 6: Table S5). Key GO terms identified include cysteine-type endopeptidase inhibitor activity (GO:0004869), threonine-type endopeptidase activity (GO:0004298) and positive regulation of catalytic activity (GO:0043085) that were enriched within the *F. gigantica* egg transcriptome and mannose metabolic process (GO:0006013), glycolytic process (GO:0006096) and ferric iron binding (GO:0008199) for the *F. gigantica* adult transcriptome.

Similarly, an abundance of GO terms were identified that were unique to only one of *Fasciola* species, with the majority of significantly enriched GO terms only being identified within the *F. gigantica* egg and adult datasets

(Additional file 6: Table S5). Consistent with *F. gigantica* using blood as a food source to facilitate the production of thousands of eggs daily [13], significant GO term enrichment was observed within the *F. gigantica* egg and adult transcriptomes, particularly pertaining to iron and heme transport and processes (GO:0015684: iron ion transport; GO:0015093: ferrous iron transmembrane transporter activity; GO:0006783: heme biosynthetic process). In addition, consistent with the adult parasites being sexually mature, significant enrichment was observed for GO terms relating to processes involving hormones (GO:0008207: C21-steroid hormone metabolic process; GO:0008209: androgen metabolic process; GO:0008210: estrogen metabolic process).

Double clustering analysis was carried out on the genes transcribed by *F. gigantica* and *F. hepatica* eggs, metacercariae and adults to investigate whether these three lifecycle stages show comparable levels of gene transcription (Additional file 7: Figure S2). Twelve clusters were generated, representing a total 36,990 genes (*F. hepatica*: 8511; *F. gigantica*: 28,479) based on gene transcription by the three lifecycle stages. GO enrichment analysis revealed that across all the clusters the majority of the significantly enriched terms were associated with gene transcription and translation (Additional file 8: Table S6). Reflecting the distinctly different stages of the parasite, the clustering analysis revealed that the parasites transcribe different genes at the different time-points of the lifecycle. Specifically, clusters 6 and 7 represents genes transcribed by the adult parasites, with enrichment for unique GO terms associated with peptidase inhibitor activity (GO:0030414), aspartic-type endopeptidase activity (GO:0004190) and G protein-coupled receptor signaling pathway (GO:0007186). High levels of gene transcription within the metacercariae are represented by clusters 9–11, which display enrichment for unique GO terms associated with small GTPase mediated signal transduction (GO:0007264), regulation of transcription (GO:0006355) and translation initiation factor activity (GO:0003743).

Only three clusters contained comparable numbers of *F. hepatica* and *F. gigantica* genes (cluster 4, cluster 5, cluster 9), indicating that despite high levels of homology at the gene sequence level (~70%; as above), this is not reflected in the transcription of these genes. Our analysis revealed low numbers of homologous genes within each of the clusters, ranging from 38 transcripts (cluster 10) to 228 transcripts (cluster 8) (Additional file 7: Figure S2), further highlighting the differences in levels of gene transcription between the two *Fasciola* species.

To investigate whether the differences in gene transcription was reflected in the most abundantly expressed genes, in depth reciprocal BLAST analysis was carried out on 250 transcripts from each *F. gigantica* lifecycle stages to identify the homologous sequences within *F.*

hepatica (Additional file 10: Table S8). Although these genes represent between 1 and 2% of the total number of genes transcribed by each of the *F. gigantica* lifecycle stages, based on the level of the transcription (FPKM) they represent 74.4, 49.8 and 81.4% of the total transcription of the egg, metacercariae and adult transcriptomes, respectively. With the exception of 22 genes, homologous sequences were identified for all the *F. gigantica* transcripts across the three life-cycle stages. Moreover, these sequences show a high level of nucleotide sequence identity with their *F. hepatica*-specific counterpart (> 80% nucleotide sequence identity; e value $< e^{-20}$; Fig. 8). Consistent with the cluster analysis (Additional file 7: Figure S2), these homologous sequences did not show comparable levels of expression, with only 75 sequences showing similar levels of expression within both the *F. hepatica* and *F. gigantica* datasets. This comparative data suggests species-specific differential gene regulation that warrants further investigation to determine if this is related to intermediate/definitive host specificity.

Analysis of the 250 genes abundantly expressed by both species highlighted key molecules of biological interest. Eggs were found to abundantly express yolk ferritin, aspartic protease and heat shock proteins, as well as redox based anti-oxidants such as peroxiredoxin (Prx) and glutathione S transferase (GST)-sigma class typically found in dormant/encysted parasite metacercarial stages [27, 42–45]. Consistent with studies of the mammalian infective stage [27, 46], transcripts annotated as cathepsin L and B, and legumain, in addition to redox based anti-oxidants Prx, fatty acid binding protein (FABP) and superoxide dismutase (SOD) were highly expressed within the metacercariae. While similar genes were identified with the adult-specific transcripts, this stage also highly expressed cysteine peptidase inhibitors (cystatin) and the potent immunomodulatory protein termed helminth defense molecule (FhHDM [47]).

Regulated transcription of the *F. gigantica* cathepsin L and B gene families

Consistent with our analysis of the *F. hepatica* genome and developmental stage-specific transcriptomes, we identified many *F. gigantica* transcripts annotated as cathepsin L and B peptidases, in addition to their *trans*-activators, the asparaginyl endopeptidases (legumains) [48]. Unlike other helminths, *Fasciola* spp. rely almost exclusively on cathepsin L and cathepsin B like cysteine proteases to facilitate tissue migration, feeding, and immune evasion/modulation [49, 50]. Analysis of the recently sequenced *F. hepatica* genome has confirmed that the large cathepsin L gene family has arisen following gene duplication, which based on current genome assemblies is comprised of 23 genes, followed by functional divergence [51]. In addition, a family of 11 cathepsin B gene

sequences was identified within the *F. hepatica* genome that arose in similar fashion [51]. Our knowledge of *F. gigantica* cathepsin gene family structure has largely been inferred from studies of *F. hepatica*. However, 16 *F. gigantica* cathepsin L sequences, which have been predominantly identified as cathepsin L1, and seven *F. gigantica* cathepsin B sequences are publically available via GenBank. In addition, further sequences, identified as part of an RNAseq project of adult *F. gigantica* parasites isolated from water buffalo at an abattoir in Thailand, include three cathepsin L-like sequences and 17 cathepsin B-like sequences [41].

Interrogation of the eight *F. gigantica* transcriptomes presented in this study using the available *F. hepatica* cathepsin and legumain sequences identified 37 transcript clusters as cathepsin L sequences, 24 transcript clusters as cathepsin B sequences and 23 transcript clusters as legumain sequences (Fig. 9). Hierarchical clustering analysis has shown that similar to *F. hepatica*, the *F. gigantica* cathepsins are also strictly developmentally regulated (Fig. 9). In addition to a group of cathepsin L genes abundantly expressed by the liver migratory stages and the adult parasites (likely homologous to *F. hepatica* cathepsins L1, L2, L3 and L5), several cathepsin L sequences were abundantly expressed by the snail rediae stage. By contrast, we observed that eggs, miracidia and rediae do not abundantly express cathepsin B-like peptidases, indicating that this peptidase class plays a less central role in migration in the snail despite being critical for the mammalian host life-cycle stages.

Of further interest was the observation that the legumain genes are abundantly transcribed, from the intra-snail redial stages through to the adult life-cycle stages (Fig. 9). This is consistent with the abundant expression of the various cathepsin genes, but also implies that these *trans*-activating peptidases are fundamental in regulating the proteolytic activity of the parasite throughout its lifecycle.

Conclusions

In this study, we have traced the dynamic transcriptional changes that occur throughout the lifecycle of the major helminth *F. gigantica* that allow the parasite to adjust to changing external environments and to both the intermediate invertebrate snail host and the definitive vertebrate mammalian host. In-depth analysis of the eight transcriptomes has elucidated key molecules and pathways important for each life-cycle stage, providing insight into *F. gigantica*-specific pathogenicity (Fig. 1). This data allowed the first opportunity for in-depth comparative analysis with *F. hepatica*, facilitating the investigation of transcription of stage-specific genes. This study and the availability of a unique and comprehensive transcriptomic data-set will lay the foundation for future studies to enhance our understanding of the similarities/

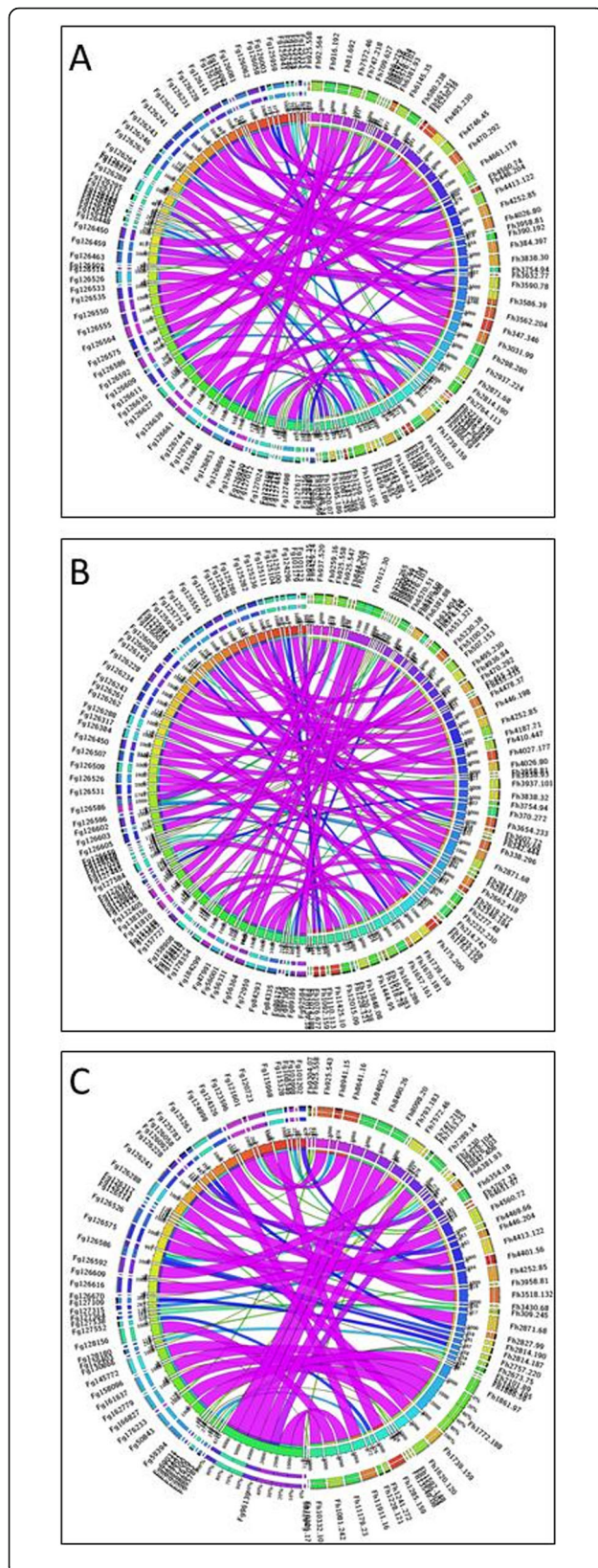


Fig. 8 Comparative analysis between *Fasciola hepatica* and *Fasciola gigantica*. Circos diagrams illustrating the best reciprocal BLAST hits of the most abundantly expressed transcripts between *F. hepatica* and *F. gigantica* with the transcriptomes of (a) egg, (b) metacercariae and (c) adult. The width and colour of the ribbons (genes) is determined by the BLAST e-value score, (Additional file 10: Table S8); narrow to wide ribbon, and a colour change through red-green-blue-purple, indicating a low to high scoring match. Full transcripts names are included in Additional file 10: Table S8

differences between the *Fasciola* species, which may account for their host specificities.

Studies have shown that in areas where both *F. gigantica* and *F. hepatica* species are present such as China, India, Korea, Japan and Vietnam, these parasites have the ability to hybridize/introgress [52–57]. *F. gigantica* and *F. hepatica* parasites are hermaphrodites and have the ability to reproduce by self-fertilization as well as cross-fertilization. However, hybrid forms are more likely to be aspermic and multiply by parthenogenesis [58]. Current methods that differentiate between these two *Fasciola* species and the hybrid forms rely on analysis of mitochondrial markers (with limited nuclear markers) and do not provide an accurate reflection of the level of hybridization/introgression occurring within the hybrid forms [57, 59–61]. The availability of extensive *F. hepatica* genomic and transcriptomic sequencing coupled with proteomic studies [46] has improved our understanding of liver fluke biology and has driven the development of new panels of markers including SNPs [40] and microsatellites [62] for population genetic studies. Coupled with the present advances in our knowledge of *F. gigantica* transcriptomics described in this paper we can now progress investigations into the biology and pathogenicity of emerging *Fasciola* hybrids.

Methods

Parasite materials

Eggs of *F. gigantica* were collected from the gall-bladders of naturally infected buffaloes from the Guangxi Zhuang Autonomous Region (GZAR), PR China. The newly-hatched miracidia were obtained from eggs incubated at 29 °C for 11 days. To collect the snail-specific stages, *Galba perversa* snails were infected with *F. gigantica* miracidia (3–5 miracidia/snail). Rediae and cercariae were collected after 15 and 40 days post-infection (dpi), respectively. After 42 days of snail infection, the snails were shed to recover the metacercariae, the infective stage of the mammalian host. A total of 15 water buffaloes were infected orally with 500 viable metacercariae per animal in GZAR, PR China. At 42, 70 and 98 dpi, five infected buffaloes were slaughtered and the parasites recovered from the liver tissue and gall bladders (15 juvenile flukes at 42 dpi, 15 juvenile flukes

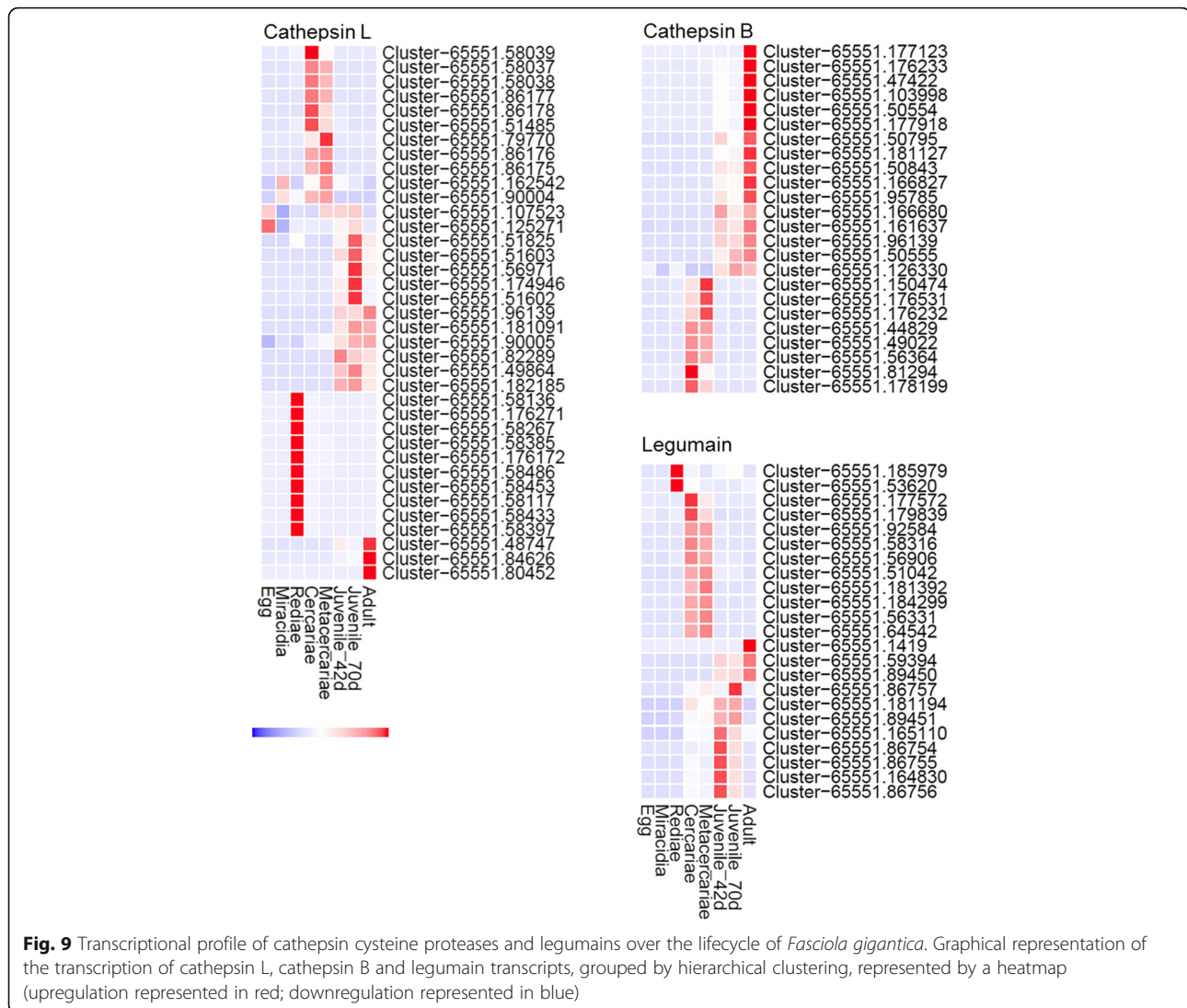


Fig. 9 Transcriptional profile of cathepsin cysteine proteases and legumains over the lifecycle of *Fasciola gigantica*. Graphical representation of the transcription of cathepsin L, cathepsin B and legumain transcripts, grouped by hierarchical clustering, represented by a heatmap (upregulation represented in red; downregulation represented in blue)

at 70 dpi and 6 adult flukes at 98 dpi [63]. The identification of *F. gigantica* was based on characteristic morphological features, which was further confirmed by amplification and sequencing of the internal transcribed spacer (ITS) region as previously described [60]. Each collected sample was washed > 10 times in physiological saline in order to remove any bacterial or host DNA contaminants and was snap frozen in liquid nitrogen followed by storage at -80°C for later use.

RNA isolation, library preparation and transcriptome sequencing

Total RNA was extracted using TRIzol according to the manufacturer's instructions (Invitrogen, USA) from three biological replicates of each *F. gigantica* lifecycle stage (pool of eggs comprising of $n > 10,000$; pool of miracidia comprising $n > 10,000$; pool of rediae comprising $n > 200$; pool of cercariae $n > 200$; pool of metacercariae comprising $n >$

5000; five juvenile flukes at 42 dpi; five juvenile flukes at 70 dpi and two adult *F. gigantica*). All RNA samples were treated with 20 units of RQ1 RNase-Free DNase (Promega, Madison, USA) to remove any residual genomic DNA according to the manufacturer's instructions. RNA and sample library integrity and concentration were confirmed using the Bioanalyzer 2100 (Agilent Technologies) and the Qubit[®] RNA Assay Kit by a Qubit[®] 2.0 Fluorometer (Life Technologies, CA, USA). Library construction was performed according to the Illumina sample preparation for RNA-seq protocol. The poly (A) mRNA was isolated from total RNA using magnetic beads with Oligo (dT), followed by fragmentation and cDNA synthesis. The resulting short mRNA fragments were purified and resolved with ethidium bromide (EB) buffer for end reparation and single nucleotide A (adenine) and subsequent adapter addition. Sequencing of the library preparations was performed by Novogene Bioinformatics Technology Company Ltd., Beijing, China (www.novogene.com).

novogene.cn) on an Illumina HiSeq X Ten platform to obtain paired-end reads.

Assembly, annotation and gene expression analysis

Illumina HiSeq reads were trimmed to $Q \geq 30$ and adapters removed using in house Python scripts. De novo reconstruction of transcriptomes from RNA-Seq data was performed by constructing a k-mer dictionary ($k = 25$), and running the Trinity software according to the software instructions [64]. Briefly, the sequence data were partitioned into many individual de Bruijn graphs, each representing the transcriptional complexity at a given gene or locus. Each graph was processed independently to extract full-length splicing isoforms and to tease apart transcripts derived from paralogous genes. The final full-length transcripts were analysed by Corset to hierarchically cluster the contigs and compute the transcript length and distribution of sequences [65]. The sequences were aligned to the available *F. gigantica* genome sequences (Genbank MKHB00000000.1 [66];) to remove any sequences that originated from contaminating snail or buffalo tissue ingested by the parasites and/or extracted during parasite sample preparation (parameter settings: total score > 100, query cover > 95, e-value < $1 \times e^{-10}$).

The transcriptional expression of the unigenes within the lifecycle stage-specific transcriptomes was calculated using the RSEM program (v1.1.17 [67];). The alignments produced by Expectation-Maximization analysis resulted in digital expression levels for each unigene that were normalized to produce fragments per kb per million reads (FPKM [68];). The resulting *P* values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted *P*-value < 0.05 as determined using DESeq were assigned as differentially expressed. Co-expression cluster analysis was carried out using Clust (v1.0.0 beta [69];), using default parameters of automatic normalization, cluster tightness of 1.0 and filtering of flat and low expression genes. Gene ontology enrichment analysis was carried out using hypergeometric tests using R. Further hierarchical clustering was carried out on those transcripts that displayed at least a 16-fold difference in expression between any of the lifecycle stages, represented by 9467 transcript clusters with a baseline cut-off of 1 FPKM in all three replicates, graphically represented using heatmaps generated using the R program, pheatmap.

Annotation of the *F. gigantica* transcripts was carried using the following in silico tools: (1) NCBI BLAST (v2.2.28; e-value $\leq 1 \times e^{-5}$) was used to interrogate the NCBI non-redundant (Nr) and nucleotide (Nt) databases, the KOG and COG databases and the protein sequence database, SwissProt; (2) Gene Ontology (GO) analysis using custom Perl Scripts (e-value $\leq 1 \times e^{-6}$) and Blast2GO v2.5 software [70]; (3) Analysis of protein conserved

domain was carried using the Pfam database, using HMMER 3.0 (e-value ≤ 0.01). Any unigenes that did not align to any database were predicted by ESTScan software [71]. Stage-specific transcription was analysed using GO and KEGG databases, based on the stringent criteria of transcripts with < 1 FPKM in all three replicates to focus on the abundantly transcribed genes. Metabolic pathway analysis was carried out by normalizing the global patterns of expression at the KEGG module level by grouping the transcript clusters annotated by KEGG together per KEGG module/pathway, graphically represented using heatmaps generated using the R program, pheatmap.

Comparative analyses with *Fasciola hepatica*

Reciprocal BLAST analyses were carried out using the *F. gigantica* DEG unigene dataset (58,422 transcript clusters) and the gene model dataset derived from the *F. hepatica* genome (22,676 gene models [40];), using a BLASTn (v2.2.29; e-value $\leq 1 \times e^{-5}$). Comparative gene ontology analysis was carried out on egg, metacercariae and adult transcriptomes of *F. gigantica* and *F. hepatica*. Enrichment of GO terms was determined using hypergeometric tests in R. Co-expression cluster analysis was carried out on the genes transcribed by the egg, metacercariae and adult stages of *F. gigantica* and *F. hepatica* using Clust (v1.0.0 beta [69];), using default parameters of automatic normalization, cluster tightness of 1.0 and filtering of flat and low expression genes. Gene ontology enrichment analysis was carried out using hypergeometric tests using R. The most abundantly expressed genes were determined from the transcriptional profile with the egg, metacercariae and adult parasites from *F. gigantica* (present study) and *F. hepatica* [40]. Reciprocal BLAST analyses were carried out using the most abundantly transcribed 250 transcripts from each transcriptome (BLASTn, v2.2.29; e-value $\leq 1 \times e^{-5}$). Circos diagrams were generated using Circos Table Viewer online [72], using the exponential value of the e-value comparison from BLASTn analysis.

F. gigantica cathepsin cysteine proteases and asparaginyl endopeptidases (legumains)

Transcript clusters representing cathepsin L and B proteases and legumains were identified based on BLAST analyses (v2.2.29) using publically available sequences from NCBI (Additional file 11: Table S9), comparative analyses with the sequences annotated within the *F. hepatica* genome [40, 51] and parsing by putative annotation. The transcriptional profile of these sequences across the *F. gigantica* lifecycle was determined and analysed by hierarchical clustering graphically represented using heatmaps generated using the R program, pheatmap.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-019-6103-5>.

Additional file 1: Table S1. Enriched GO terms of the DEG from pairwise comparisons between the *Fasciola gigantica* lifecycle stages.

Additional file 2: Table S2. Unique GO terms associated with snail and buffalo-associated lifecycle stages.

Additional file 3: Figure S1. Cluster analysis of the unigenes transcribed across the *Fasciola gigantica* lifecycle based on gene expression. Sixteen clusters were generated by cluster analysis using Clust, shown by the individual graphs (C0-C15). The lifecycle stages are represented as follows: egg, mir: miracidia, red: rediae, cer: cercariae, met: metacercariae, 42d: juvenile fluke 42dpi, 70d: juvenile fluke 70dpi and ad: adult.

Additional file 4: Table S3. Gene ontology enrichment cluster analysis of the unigenes transcribed by eight *Fasciola gigantica* lifecycle stages.

Additional file 5: Table S4. List of the 34,264 *Fasciola gigantica* differentially expressed genes.

Additional file 6: Table S5. Enriched GO terms within egg, metacercariae and adult transcriptomes from comparative analysis between *Fasciola gigantica* and *Fasciola hepatica*.

Additional file 7: Figure S2. Comparative cluster analysis of gene transcription by *Fasciola gigantica* and *Fasciola hepatica* egg, metacercariae and adult lifecycle stages. Twelve clusters were generated by cluster analysis using Clust, shown by the individual graphs (C0-C11). The lifecycle stages are represented as follows: egg, met: metacercariae and ad: adult. The number of species-specific transcripts are shown, with the numbers in brackets representing the number of homologous sequences observed within each cluster.

Additional file 8: Table S6. Gene ontology enrichment cluster analysis of the co-expressed genes transcribed by *Fasciola gigantica* and *Fasciola hepatica* egg, metacercariae and adult lifecycle stages.

Additional file 9: Table S7. Number of the 9468 DEG displaying at least a 16-fold difference annotated with gene ontology terms relating to Molecular Function.

Additional file 10: Table S8. Comparative analysis of the top 250 most abundant *Fasciola gigantica* transcripts compared with homologous sequences in *Fasciola hepatica*.

Additional file 11: Table S9. GenBank accession numbers of the publically available *Fasciola gigantica* cathepsin L, cathepsin B and legumain genes.

Abbreviations

BLAST: Basic Local Alignment Search Tool; DEG: Differentially expressed genes; dpi: Days post infection; EB: Ethidium bromide; EVs: Extracellular vesicles; FABP: Fatty acid binding protein; FPKM: Fragment per kilobase per million reads; GO: Gene ontology; GST: Glutathione S-transferase; GZAR: Guangxi Zhuang Autonomous Region; HDM: Helminth defense molecule; ITS: Internal transcribed spacer; MAC: Membrane attack complex; NEJ: Newly excysted juveniles; PRX: Peroxiredoxin; QMF: Quantitative metabolic fingerprint; SNP: Single nucleotide polymorphisms; SOD: Superoxide dismutase; WHO: World Health Organisation

Acknowledgements

We thank Mr. Jian-Gang Ma, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, for assistance during sample collection.

Authors' contributions

XQZ and JPD conceived the study and contributed funding and resources. XXZ, KC, RSH, WBZ, FKZ and ZAS prepared parasite material for sequencing and performed the research. KC, XXZ and RSH carried out the analysis. XXZ, KC, HME, JPD and XQZ interpreted the data and prepared the data for publication. XXZ, KC and HME wrote the manuscript. KC, JPD, XXZ and XQZ revised the manuscript for final publication. All authors have read and approved the manuscript.

Funding

This work was supported by the National Key Basic Research Program (973 Program) of China (Grant No. 2015CB150300) and the Elite Program of Chinese Academy of Agricultural Sciences. JPD is a recipient of a Science Foundation Ireland (SFI, Republic of Ireland) Professorship in Molecular Parasitology.

Availability of data and materials

The *F. gigantica* transcriptome datasets supporting the conclusions of this article are available via NCBI BioProject: PRJNA350370. The *F. hepatica* datasets [40] used for the homology studies are freely available from European Nucleotide Archive under accessions LN627018-LN647175 (assembly data), PRJEB6687 (genomic read data) and PRJEB6904 (transcriptomic read data).

Ethics approval and consent to participate

The water buffalos, from which *F. gigantica* were collected, were handled in accordance with good animal practice as required by the Animal Ethics Procedures and Guidelines of the People's Republic of China, and the study was approved by the Animal Administration and Ethics Committee of Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences (Permit number: LVRIAE-2015-08).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou, Gansu Province 730046, People's Republic of China. ²College of Veterinary Medicine, Qingdao Agricultural University, Qingdao, Shandong Province 266109, People's Republic of China. ³National Centre for Biomedical and Engineering Science (NCBES), School of Natural Sciences, National University of Ireland, Galway, Ireland. ⁴College of Animal Science and Technology, Guangxi University, Nanning, Guangxi Zhuang Autonomous Region 530005, People's Republic of China. ⁵Faculty of Medicine and Health Sciences, School of Veterinary Medicine and Science, University of Nottingham, Nottingham, UK.

Received: 25 May 2019 Accepted: 13 September 2019

Published online: 12 October 2019

References

- Copeman DB, Copland RS. Importance and potential impact of liver fluke in cattle and buffalo. In: Gray GD, Copland RS, Copeman DB, editors. Overcoming liver fluke as a constraint to ruminant production in South-East Asia. Canberra: Australian Centre for International Agricultural Research; 2008. p. 1–34.
- Isah UM. Studies on the prevalence of fascioliasis among ruminant animals in northern Bauchi state, North-Eastern Nigeria. *Parasite Epidemiol Control*. 2019;5:e00090.
- Mas-Coma S, Valero MA, Bargues MD. Fascioliasis. *Adv Exp Med Biol*. 2019; 1154:71–103.
- Mas-Coma S. Human Fascioliasis: epidemiological pattern in human endemic areas of South America, Africa and Asia. *Southeast Asian J Trop Med Public Health*. 2014;35:1–11.
- Spithill TW, Smooker PM, Coperman DB. *Fasciola gigantica*: epidemiology, control, immunology and molecular biology. In: Dalton JP, editor. Fasciolosis: CABI Publishing; 1999. p. 465–527.
- Mehmood K, Zhang H, Sabir AJ, Abbas RZ, Ijaz M, Durrani AZ, Saleem MH, Ur Rehman M, Iqbal MK, Wang Y, Ahmad HI, Abbas T, Hussain R, Ghori MT, Ali S, Khan AU, Li J. A review on epidemiology, global prevalence and economical losses of fasciolosis in ruminants. *Microb Pathog*. 2017;109:253–62.
- Mas-Coma S, Valero MA, Bargues MD. Chapter 2. *Fasciola*, lymnaeids and human fascioliasis, with a global overview on disease transmission, epidemiology, evolutionary genetics, molecular epidemiology and control. *Adv Parasitol*. 2009;69:41–146.

8. Graczyk TK, Fried B. Development of *Fasciola hepatica* in the intermediate host. In: Dalton JP, editor. Fasciolosis: CABI Publishing; 1999. p. 31–46.
9. Hodgkinson JE, Cwiklinski K, Beesley N, Hartley C, Allen K, Williams DJL. Clonal amplification of *Fasciola hepatica* in *Galba truncatula*: within and between isolate variation of triclabendazole-susceptible and -resistant clones. Parasit Vectors. 2018;11:363.
10. Dinnik JA, Dinnik NN. Observations on the succession of redial generations of *Fasciola gigantica* Cobbold in a snail host. Z Tropenmed Parasitol. 1956;7:397–419.
11. Dinnik JA, Dinnik NN. The influence of temperature on the succession of redial and cercarial generations of *Fasciola gigantica* in a snail host. Parasitol. 1964;54:59–65.
12. Rondelaud D, Belfaiza M, Vignoles P, Moncef M, Dreyfuss G. Redial generations of *Fasciola hepatica*: a review. J Helminthol. 2009;83:245–54.
13. Boray JC. Experimental fascioliasis in Australia. Adv Parasitol. 1969;7:95–210.
14. Phalee A, Wongsawad C, Rojanapaibul A, Chai JY. Experimental life history and biological characteristics of *Fasciola gigantica* (Digenea: Fasciolidae). Korean J Parasitol. 2015;53:59–64.
15. Pleasance J, Raadsma HW, Estuningsih SE, Widjajanti S, Meeusen E, Piedrafita D. Innate and adaptive resistance of Indonesian thin tail sheep to liver fluke: a comparative analysis of *Fasciola gigantica* and *Fasciola hepatica* infection. Vet Parasitol. 2011;178:264–72.
16. Itagaki T, Ichinomiya M, Fukuda K, Fusyuku S, Carmona C. Hybridization experiments indicate incomplete reproductive isolating mechanism between *Fasciola hepatica* and *Fasciola gigantica*. Parasitol. 2011;138:1278–84.
17. Valero MA, Bargues MD, Khoubbane M, Artigas P, Quesada C, Berinde L, Ubeira FM, Mezo M, Hernandez JL, Agramunt VH, Mas-Coma S. Higher physiopathogenicity by *Fasciola gigantica* than by the genetically close *F. hepatica*: experimental long-term follow-up of biochemical markers. Trans R Soc Trop Med Hyg. 2016;110:55–66.
18. Robinson MW, Dalton JP. Zoonotic helminth infections with particular emphasis on fasciolosis and other trematodiasis. Philos Trans R Soc Lond Ser B Biol Sci. 2009;364:2763–76.
19. Liu GH, Xu MJ, Chang QC, Gao JF, Wang CR, Zhu XQ. *De novo* transcriptomic analysis of the female and male adults of the blood fluke *Schistosoma turkestanicum*. Parasit Vectors. 2016;9:143.
20. Liu GH, Xu MJ, Song HQ, Wang CR, Zhu XQ. *De novo* assembly and characterization of the transcriptome of the pancreatic fluke *Eurytrema pancreaticum* (trematoda: Dicrocoeliidae) using Illumina paired-end sequencing. Gene. 2016;576:333–8.
21. Andrews S. The Life cycle of *Fasciola hepatica*. In: Dalton JP, editor. Fasciolosis: CABI Publishing; 1999. p. 1–29.
22. Georgieva K, Georgieva S, Mizinska Y, Stoitsova SR. *Fasciola hepatica* miracidia: lectin binding and stimulation of *in vitro* miracidium-to-sporocyst transformation. Acta Parasitol. 2012;57:46–52.
23. Peterson NA, Hokke CH, Deelder AM, Yoshino TP. Glycotype analysis in miracidia and primary sporocysts of *Schistosoma mansoni*: differential expression during the miracidium to sporocyst transformation. Int J Parasitol. 2009;39:1331–44.
24. Prasopdee S, Sotillo J, Tesana S, Laha T, Kulsantiwong J, Nolan MJ, Loukas A, Cantacessi C. RNA-Seq reveals infection-induced gene expression changes in the snail intermediate host of the carcinogenic liver fluke, *Opisthorchis viverrini*. PLoS Negl Trop Dis. 2014;8:e2765.
25. Knight M, Arican-Goktas HD, Ittiprasert W, Odoemelum EC, Miller AN, Bridger JM. Schistosomes and snails: a molecular encounter. Front Genet. 2014;5:230.
26. Pila EA, Li H, Hambrook JR, Wu X, Hanington PC. Schistosomiasis from a snail's perspective: advances in snail immunity. Trends Parasitol. 2017;33:845–57.
27. Cwiklinski K, Jewhurst H, McVeigh P, Barbour T, Maule AG, Tort J, O'Neill SM, Robinson MW, Donnelly S, Dalton JP. Infection by the helminth parasite *Fasciola hepatica* requires rapid regulation of metabolic, virulence, and invasive factors to adjust to its mammalian host. Mol Cell Proteomics. 2018;17:792–809.
28. Tielens AG, van den Heuvel JM, van den Bergh SG. The energy metabolism of *Fasciola hepatica* during its development in the final host. Mol Biochem Parasitol. 1984;13:301–7.
29. Ravida A, Aldridge AM, Driessen NN, Heus FA, Hokke CH, O'Neill SM. *Fasciola hepatica* surface coat glycoproteins contain mannoseylated and phosphorylated N-glycans and exhibit immune modulatory properties independent of the mannose receptor. PLoS Negl Trop Dis. 2016;10:e0004601.
30. Ravida A, Cwiklinski K, Aldridge AM, Clarke P, Thompson R, Gerlach JQ, Kilcoyne M, Hokke CH, Dalton JP, O'Neill SM. *Fasciola hepatica* surface tegument: glycoproteins at the interface of parasite and host. Mol Cell Proteomics. 2016;15:3139–53.
31. Garcia-Campos A, Ravida A, Nguyen DL, Cwiklinski K, Dalton JP, Hokke CH, O'Neill S, Mulcahy G. Tegument glycoproteins and cathepsins of newly excysted juvenile *Fasciola hepatica* carry mannosidic and paucimannosidic N-glycans. PLoS Negl Trop Dis. 2016;10:e0004688.
32. Cwiklinski K, O'Neill SM, Donnelly S, Dalton JP. A prospective view of animal and human Fasciolosis. Parasite Immunol. 2016;38:558–68.
33. Molina-Hernandez V, Mulcahy G, Perez J, Martinez-Moreno A, Donnelly S, O'Neill SM, Dalton JP, Cwiklinski K. *Fasciola hepatica* vaccine: we may not be there yet but we're on the right road. Vet Parasitol. 2015;208:101–11.
34. McVeigh P, McCammick E, McCusker P, Wells D, Hodgkinson J, Paterson S, Mousley A, Marks NJ, Maule AG. Profiling G protein-coupled receptors of *Fasciola hepatica* identifies orphan rhodopsins unique to phylum Platyhelminthes. Int J Parasitol Drugs Drug Resist. 2018;8:87–103.
35. Falis M, Krupicer I, Legáth J, Pitoňák V, Špalek M. The effects of heavy metals and pesticides on survival of miracidia stage of *Fasciola hepatica*. Helminthol. 2013;4:287–90.
36. Song J, Mak E, Wu B, Beitz E. Parasite aquaporins: current developments in drug facilitation and resistance. Biochim Biophys Acta. 1840;2014:1566–73.
37. Kreida S, Tomroth-Horsefield S. Structural insights into aquaporin selectivity and regulation. Curr Opin Struct Biol. 2015;33:126–34.
38. Froger A, Tallur B, Thomas D, Delamarque C. Prediction of functional residues in water channels and related proteins. Protein Sci. 1998;7:1458–68.
39. Shi Y, Toet H, Rathinasamy V, Young ND, Gasser RB, Beddoe T, Huang W, Spithill TW. First insight into CD59-like molecules of adult *Fasciola hepatica*. Exp Parasitol. 2014;144:57–64.
40. Cwiklinski K, Dalton JP, Dufresne PJ, La Course J, Williams DJ, Hodgkinson J, Paterson S. The *Fasciola hepatica* genome: gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution. Genome Biol. 2015;16:71.
41. Young ND, Jex AR, Cantacessi C, Hall RS, Campbell BE, Spithill TW, Tangkawattana S, Tangkawattana P, Laha T, Gasser RB. A portrait of the transcriptome of the neglected trematode, *Fasciola gigantica*-biological and biotechnological implications. PLoS Negl Trop Dis. 2011;5:e1004.
42. Larsen PL. Aging and resistance to oxidative damage in *Caenorhabditis elegans*. Proc Natl Acad Sci U S A. 1993;90:8905–9.
43. Devaney E. Thermoregulation in the life cycle of nematodes. Int J Parasitol. 2006;36:641–9.
44. Wang Y, Ezemaduka AN, Tang Y, Chang Z. Understanding the mechanism of the dormant dauer formation of *C. elegans*: from genetics to biochemistry. IUBMB Life. 2009;61:607–12.
45. Parkinson J, Wasmuth JD, Salinas G, Bizarro CV, Sanford C, Berriman M, Ferreira HB, Zaha A, Blaxter ML, Maizels RM, Fernandez C. A transcriptomic analysis of *Echinococcus granulosus* larval stages: implications for parasite biology and host adaptation. PLoS Negl Trop Dis. 2012;6:e1897.
46. Cwiklinski K, Dalton JP. Advances in *Fasciola hepatica* research using 'omics' technologies. Int J Parasitol. 2018;48:321–31.
47. Robinson MW, Donnelly S, Hutchinson AT, To J, Taylor NL, Norton RS, Perugini MA, Dalton JP. A family of helminth molecules that modulate innate cell responses via molecular mimicry of host antimicrobial peptides. PLoS Pathog. 2011;7:e1002042.
48. Dalton JP, Brindley PJ, Donnelly S, Robinson MW. The enigmatic asparaginyl endopeptidase of helminth parasites. Trends Parasitol. 2009;25:59–61.
49. Tort J, Brindley PJ, Knox D, Wolfe KH, Dalton JP. Proteinases and associated genes of parasitic helminths. Adv Parasitol. 1999;43:161–266.
50. Kašný M, Mikeš L, Hampl V, Dvořák J, Caffrey CR, Dalton JP, Horák P. Peptidases of Trematodes. Adv Parasitol. 2009;69:205–97.
51. Cwiklinski K, Donnelly S, Drysdale O, Jewhurst H, Smith D, De Marco VC, Pritsch IC, O'Neill S, Dalton JP, Robinson MW. The cathepsin-like cysteine peptidases of trematodes of the genus *Fasciola*. Adv Parasitol. 2019;104:113–64.
52. Rhee JK, Eun GS, Lee SB. Karyotype of *Fasciola* sp. obtained from Korean cattle. Kisaengchunghak Chapchi. 1987;25:37–44.
53. Itagaki T, Tsutsumi K. Triploid form of *Fasciola* in Japan: genetic relationships between *Fasciola hepatica* and *Fasciola gigantica* determined by ITS-2 sequence of nuclear rDNA. Int J Parasitol. 1998;28:777–81.
54. Itagaki T, Tsutsumi KI, Ito K, Tsutsumi Y. Taxonomic status of the Japanese triploid forms of *Fasciola*: comparison of mitochondrial ND1 and COI sequences with *F. hepatica* and *F. gigantica*. J Parasitol. 1998;84:445–8.
55. Terasaki K, Noda Y, Shibahara T, Itagaki T. Morphological comparisons and hypotheses on the origin of polyploids in parthenogenetic *Fasciola* sp. J Parasitol. 2000;86:724–9.

56. Itagaki T, Sakaguchi K, Terasaki K, Sasaki O, Yoshihara S, Van Dung T. Occurrence of spermic diploid and aspermic triploid forms of *Fasciola* in Vietnam and their molecular characterization based on nuclear and mitochondrial DNA. *Parasitol Int.* 2009;58:81–5.
57. Peng M, Ichinomiya M, Ohtori M, Ichikawa M, Shibahara T, Itagaki T. Molecular characterization of *Fasciola hepatica*, *Fasciola gigantica*, and aspermic *Fasciola* sp. in China based on nuclear and mitochondrial DNA. *Parasitol Res.* 2009;105:809–15.
58. Ichikawa-Seki M, Peng M, Hayashi K, Shoriki T, Mohanta UK, Shibahara T, Itagaki T. Nuclear and mitochondrial DNA analysis reveals that hybridization between *Fasciola hepatica* and *Fasciola gigantica* occurred in China. *Parasitol.* 2017;144:206–13.
59. Le TH, Blair D, McManus DP. Complete DNA sequence and gene organization of the mitochondrial genome of the liver fluke, *Fasciola hepatica* L. (Platyhelminthes; Trematoda). *Parasitol.* 2001;123:609–21.
60. Liu GH, Gasser RB, Young ND, Song HQ, Ai L, Zhu XQ. Complete mitochondrial genomes of the 'intermediate form' of *Fasciola* and *Fasciola gigantica*, and their comparison with *F. hepatica*. *Parasit Vectors.* 2014;7:150.
61. Harrison RG, Larson EL. Hybridization, introgression, and the nature of species boundaries. *J Hered.* 2014;105(Suppl 1):795–809.
62. Cwiklinski K, Allen K, LaCourse J, Williams DJ, Paterson S, Hodgkinson JE. Characterisation of a novel panel of polymorphic microsatellite loci for the liver fluke, *Fasciola hepatica*, using a next generation sequencing approach. *Infect Genet Evol.* 2015;32:298–304.
63. Zhang XX, Cong W, Elsheikha HM, Liu GH, Ma JG, Huang WY, Zhao Q, Zhu XQ. *De novo* transcriptome sequencing and analysis of the juvenile and adult stages of *Fasciola gigantica*. *Infect Genet Evol.* 2017;51:33–40.
64. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29:644–52.
65. Davidson NM, Oshlack A. Corset: enabling differential gene expression analysis for *de novo* assembled transcriptomes. *Genome Biol.* 2014;15:410.
66. Tripathi T, Ghosh A, Todur VN, Kalita P, Vijayakumar R, Kalita J, Shukla R, Chetri PB, Shukla H, Sonkar A, Lyngdoh DL, Singh R, Chikara SK, Tripathi T. Draft genome of the liver fluke *Fasciola gigantica*. *bioRxiv.* 2018:451476.
67. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
68. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008;5:621–8.
69. Abu-Jamous B, Kelly S. Clust: automatic extraction of optimal co-expressed gene clusters from gene expression data. *Genome Biol.* 2018;19:172.
70. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 2008;36:3420–35.
71. Iseli C, Jongeneel CV, Bucher P. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc Int Conf Intell Syst Mol Biol.* 1999:138–48.
72. Circos Table Viewer online: <http://mkweb.bcgsc.ca/tableviewer/visualize/>. Accessed 25 Jan 2019.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

