BMC Genomics

RESEARCH Open Access

Transcriptome profiling of intra-snail stages of the liver fluke *Fasciola hepatica* reveals key mediators underlying parasite development and interaction with the host

Mauricio Langleib[1,2], Santiago Fontenla[2], Fernanda Domínguez[2], Geetha Sankaranarayanan[3], Sabina Wlodek[2,4], Matt Berriman[3,5], Gabriel Rinaldi[3,6], Andrés Iriarte[1*] and José F. Tort[2*]

Abstract

Background While mammalian host invasion by the liver fluke *Fasciola hepatica* has been extensively studied, reports on the interaction with its intermediate host at the molecular level remain scarce. These developmental stages are particularly interesting since at least two cycles of asexual amplification occur within lymnaeid snails, leading to the production of hundreds of infective metacercariae. We analyzed transcriptomic data from miracidia and intra-molluscan stages seeking clues of these interaction and developmental processes.

Results We identified 1744 novel transcripts and several isoforms of already annotated genes. Analysis of expression across the whole life cycle resulted in five distinct gene expression clusters (egg, miracidia, intra-snail stages, invading stages and juvenile-adults). Few genes showed strict stage-specific expression, but notably most of those corresponding to miracidial and intra-snail stages were novel unannotated genes. Genes upregulated in the miracidium include enzymes involved in neurotransmitter synthesis, energy metabolism and calcium mediated signaling, consistent with the physiology of a short-lived free-living stage. Several genes associated with development and morphogenesis were characterized in early (15 days post-infection) intra-snail stages. Purine salvage pathway genes were upregulated in this time-point, consistent with a high biosynthetic demand and the absence of a complete purine synthesis pathway in *F. hepatica*. Mucins, glycan biosynthesis genes and aquaporins upregulated within late (30 dpi) intra-snail stages are interesting, considering their putative role in the next host transition.

Conclusions Different members of well-known protein families involved in host-parasite interaction such as cathepsin proteases, legumains, protease inhibitors and lipid transporters were detected with stage specific expression in early and late intra-snail stages. These findings suggest duplication and tunning during evolution of the same set of molecular mediators for the interactions with the intermediate and definitive hosts.

Keywords Fasciola, Transcriptomics, Intermediate host, Development, Invasion, Gene families

*Correspondence:
Andrés Iriarte
airiarte@higiene.edu.uy
José F. Tort
jtort@fmed.edu.uy

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

Background

Fasciolosis is the zoonotic disease with the widest distribution worldwide, affecting all continents with the only exception of Antarctica. This foodborne trematodiasis (FBT) is caused by two species of liver flukes, *Fasciola hepatica* in temperate areas and *F. gigantica* in tropical regions of Africa and Asia [1]. Fasciolosis is responsible for massive economic losses to the livestock industry, estimated globally to be US$3.2 billion annually due to reduced production yields and associated treatment costs [2]. It is also recognized as a Neglected Tropical Disease (NTD) by the World Health Organization (WHO) (https://www.who.int/news-room/questions-and-answers/item/q-a-on-fascioliasis), with at least 2.4 million people infected in more than 75 countries worldwide, with several million more at risk [3, 4]. However, the actual numbers of infections in humans and animals are likely underestimated due to the lack of comprehensive or coordinated studies, and limited availability of diagnostic tools in some developing countries [5].

The infection is acquired by ingesting plants contaminated with the parasite metacercariae, an encysted larval form that is activated by gastric and duodenal contents promoting excystation. The newly excysted juvenile (NEJ) actively migrates through the intestinal wall, breaking epithelial cells, connective tissue and muscular fibers, to the peritoneal cavity. Here the parasites penetrate the liver, feeding extensively on blood and liver parenchyma to facilitate their growth and development. This initial acute phase of the disease spontaneously resolves 5 to 6 weeks after the infection when the worms established within the biliary ducts, develop into adults. In the immunologically protected niche of the bile duct, the adults can live for years laying thousands of eggs daily, which are passed to the environment with host feces [6]. When the eggs reach a water source and find appropriate conditions of light and temperature, embryonic development is induced, leading to the hatching of the miracidium. This free-living larval stage is characterized by quick and coordinated movements facilitated by the ciliated plates in which it is coated. The miracidium actively swims in search for its intermediate host, a snail of the genus *Lymnaea*. Through coordinated mechanic action and enzymatic secretions the miracidium penetrates the snail, losing its ciliary plates and transforming into sporocyst [7, 8].

Within the intermediate snail host extensive asexual amplification occurs. A population of undifferentiated stem cells (denominated neoblasts) is thought to underlie this intra-molluscan amplification, through multiple rounds of proliferation and *de novo* embryogenesis in the absence of fertilization [9]. Five to eight germinal cells within the miracidium start proliferating when transformed into a sporocyst, each leading to the formation of a rediae, a larval form displaying a digestive system. These larvae migrate within the host feeding from snail tissues. Within the rediae, the asexual amplification cycles triggered by undifferentiated germinal cells continue, producing secondary rediae (that can perpetuate the asexual reproduction) and/or cercariae that are shed from the snail after 4 to 7 weeks post-infection. Emerging cercariae swim freely seeking vegetation (e.g., watercress) on top of which they encyst [10]. Remarkably, the asexual amplification cycles within the snail are vital for parasite propagation, since a single miracidium can produce hundreds of metacercariae [11].

The diverse developmental transitions across the liver fluke life cycle are accompanied by changes at the gene expression level. Several efforts to analyze these changes at transcriptomic and proteomic level have largely been centered in the intra-mammalian developmental stages, including NEJs, juveniles, adult worms and eggs [12–15]. More recently, spatial transcriptomic studies on adult worms have added detailed insights about this developmental stage [16]. However, large-scale changes to gene expression within the intermediate host remain poorly characterized.

In the present study, we analyzed gene expression patterns of the *F. hepatica* miracidium and three intra-snail stages collected from experimentally infected *Lymnaea viatrix* snails. We also used publicly available RNA-seq data from other stages to reveal differentially expressed genes in each case, analyzing in detail the expression profile of genes involved in development and host-parasite interaction. Additionally, we compared our findings with those from a previous study in *F. gigantica* intra-snail stages [17]. Our analyses revealed that gene expression across the life cycle can be subdivided into 5 or 6 discrete clusters of coordinated expression, mirroring the major transitions between hosts. Furthermore, diverse gene families show paralogous expansions associated with expression at different stages of the life cycle. These results suggest that the parasite employs different variants of a restricted toolkit for the invasion of both intermediate and definitive hosts.

Results and discussion

Intra-snail stage transcriptomics data assembly reveals novel genes

While there is reasonable information on transcriptomic changes associated with intra-mammalian developmental stages of *Fasciola hepatica* [12–15], there are no transcriptomics studies of intra-snail stages for this species, and only a single report in the sister species *F. gigantica* [17]. To fill this knowledge gap, here we analyzed the transcriptomic profiles of miracidia, and parasites collected from infected snails at 15-, 21- and 30-days post-infection (IS15, IS21 and IS30, respectively). Consistent

with the rather asynchronous pattern of development of liver fluke within snails [10], samples obtained at different time points were diverse, comprising late sporocysts and/or different sized rediae (primary and secondary) at the early time point (Fig. 1A-B), while at the later time point samples consisted mainly of late rediae containing developing cercariae and fully developed cercariae (Fig. 1C-D). The asynchronic development and the resulting heterogeneity of forms at intra-snail stages is an intrinsic characteristic of the parasite life cycle, and time points are traditionally considered as good proxies of the advancement of the development within the intermediate host [18].

Samples from different time points were used to prepare Illumina RNAseq libraries, barcoded, equimolarly pooled, and sequenced. We obtained more than 340 million good quality paired reads from all the libraries with mapping rates to the reference genome (PRJEB25283) ranging from 72% to 88% (Suppl. Table 1). Only 4% of the unmapped reads could be mapped to snail host genomes, suggesting low levels of contamination with host tissues. In any case, all unmapped reads where not considered in downstream analyses. Transcript assembly after mapping to the reference genome resulted in more than 10 thousand confident transcripts (over 10x coverage) in each sample (Suppl. Table 1). This count is consistent with the initial 9,732 genes described in the reference genome, and lower but in the same order of magnitude than the 58,422 assembled transcripts reported for all the life stages transcriptomics of *F. gigantica* [17]. About 50% of the "novel transcripts" reported by StringTie corresponded

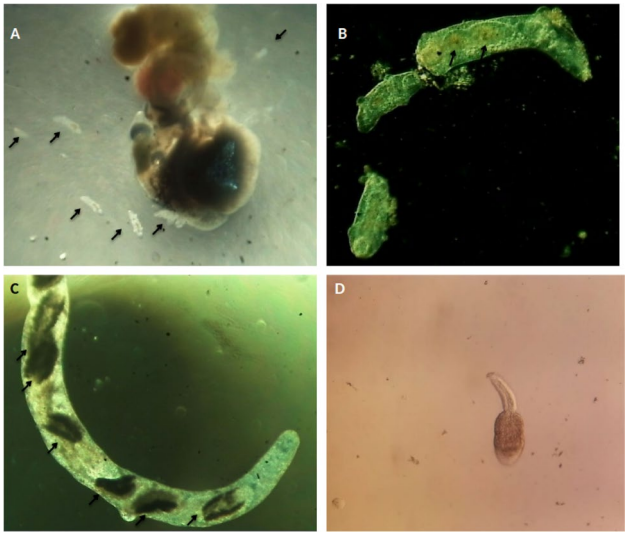

Fig. 1 Dissected snails and specimens recovered, including those sequenced in the study. **A** IS15 snail with parasites emerging (indicated with arrows). **B** Parasites representative of early stages infection. Arrows indicate proliferative cells masses within one rediae, **C** Rediae containing cercariae (arrows), representative of late (IS30) stages of infection. **D** Free cercaria observed in late stages of infection

to "splicing variants" of already annotated genes (Suppl. Figure_2A). Novel isoforms account for an additional ~ 20%, with intergenic annotations being the greater fraction within this category. These novel genes were particularly interesting, given that they could be stage-specific. We detected 1744 novel putative protein coding genes expressed in at least one of the analyzed time-points. These were added to the already annotated genes in the reference, resulting in 11,452 total protein coding genes that we used as our curated dataset from here on. Almost 70% of them have some type of functional annotation, while the rest remain as unknown genes.

Multidimensional analysis of gene expression mirrors the major life cycle transitions in *F. hepatica*

Variations in gene expression were analyzed across the life cycle of *F. hepatica*, by integrating our intra-snail stages transcriptomics with previously published data from other stages (namely eggs, metacercariae, NEJs, juveniles and adults) [12–15]. In order to get comparable results, we reanalyzed both publicly available and novel transcriptomic data using the same framework. After quantification and quality-control, the consistency of samples was analyzed by multidimensional scaling (MDS). The first two dimensions explained ~ 50% of the variance (dimension 1, ~ 30%; dimension 2, ~ 20%) (Fig. 2A, C), with samples from the same developmental stages grouping together across five clusters: *(i)* egg and adult stages (EGG, AD), *(ii)* miracidium (MIR), *(iii)* intra-snail stages (IS15, IS21, IS30), *(iv)* metacercariae (MET) and different time points of newly excysted juveniles (NEJ1, NEJ2, NEJ24), and *(v)* juvenile (JUV). The clustering of adult and egg samples may reflect the presence of eggs within the adult reproductive organs, as expected considering the high egg production by hermaphroditic adult worms.

Consistent with this hypothesis, adding the third dimension (which explains ~ 20% of the variance) of the MDS clearly discriminates these two stages (Fig. 2B). Similarly, miracidium samples were close to the intra-snail samples, NEJ samples proximate to juvenile samples, which in turn appeared closer to the adult stage, thus recapitulating the life cycle stage progression of *F. hepatica*.

Contiguous life stages show similar profiles of highly expressed genes

More than 50% of the annotated genes showed expression across all life cycle stages, while the remaining expressed genes were generally shared by more than one stage. Stage-specific expression was detected only for a reduced set of genes in miracidium and intra-snail stages (Suppl Fig. 2B). Most of these genes corresponded to novel protein coding transcripts identified with StringTie

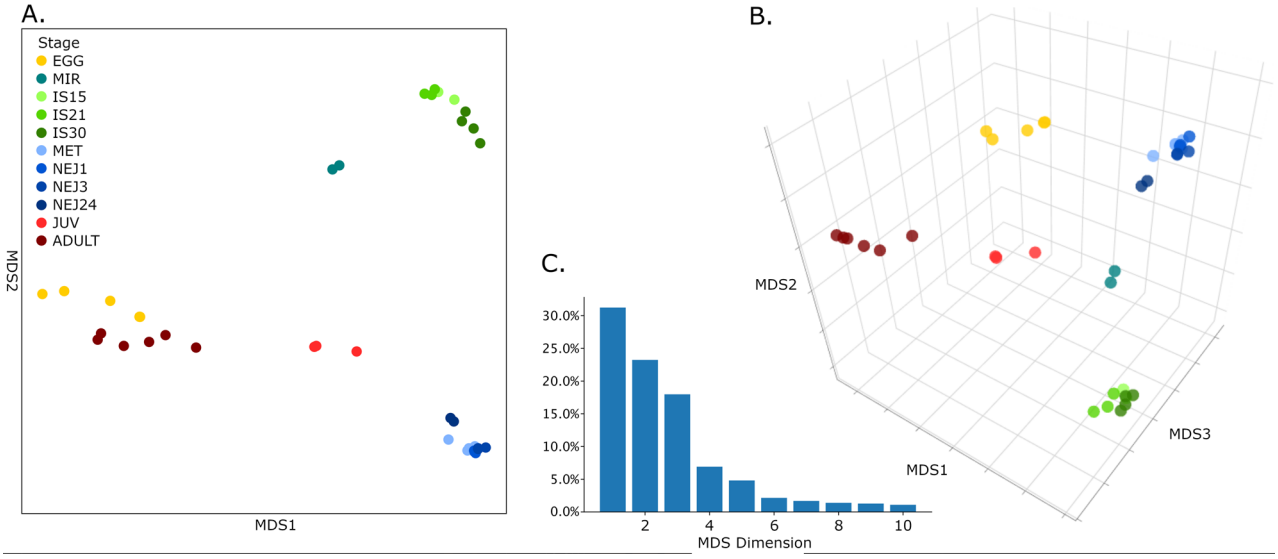

Fig. 2 Multidimensional analysis of the samples used in the study. **A** Shows the two main dimensions. **B** Adding the third-dimension splits adult and egg samples. An interactive version of this plot could be found as IP1 in the supplementary Zenodo repository (**C**) Proportion of the variance that can be explained by each dimension of the first ten dimensions

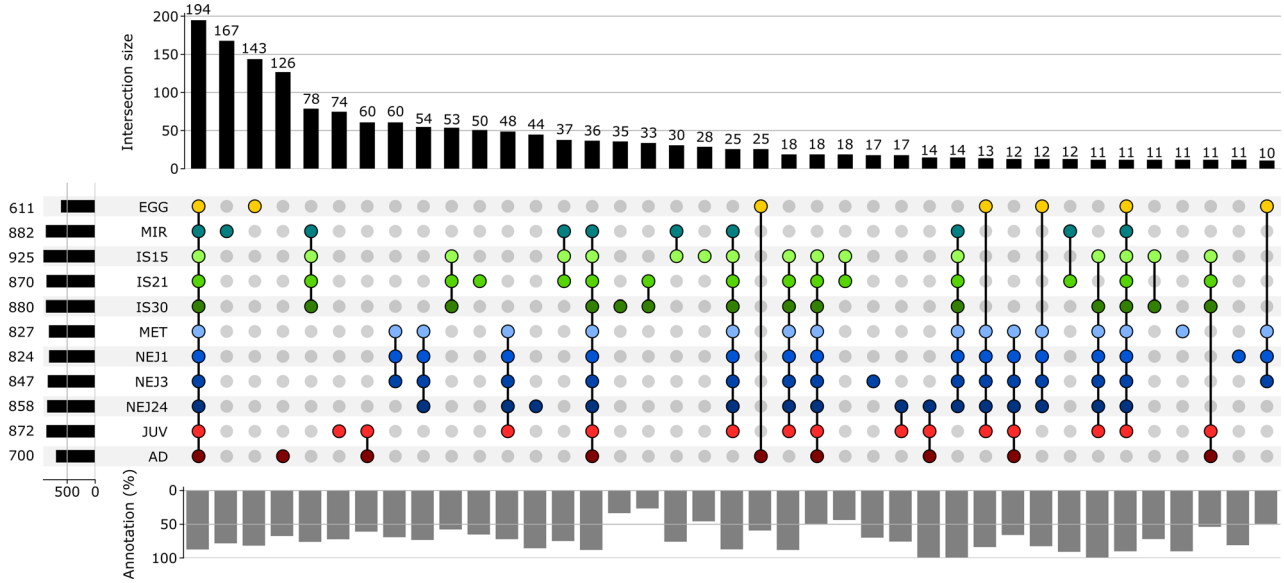

Fig. 3 Upset plot showing the intersection of the top 10% expressed genes between stages. The number of genes per stage and the size of each intersection group are indicated by left and top black bars respectively. Contiguous life stages with similar HEG patterns are indicated by different shades of color (i.e. yellow for eggs, greens for snail-related stages, blues for invasion of the mammalian host, reds for intra-liver stages). The percentage of genes with any kind of functional annotation (GO, KEGG or PFAM) within each group is shown as grey bars below the plot

in previously unannotated inter-genic regions of the reference genome, highlighting the relevance of integrating transcriptomic data from as many life cycle stages as possible [19].

To further characterize gene expression across the life cycle of the parasite, we restricted the analysis to the top 10% of expressed genes in each stage. While the larger set (~30%) still corresponds to several genes expressed in all stages that might represent housekeeping functions, other patterns could be detected (Fig. 3A).

Notably, contiguous life stages show similar profiles of highly expressed genes (HEG), particularly a set of genes highly expressed in the intra-snail stages and a group of genes highly expressed in the early invasive stages of the mammalian host. A few stage-specific HEG were also detected. In general, the sets of HEG associated with the intra-snail stages showed lower percentage of functional annotation, consistent with the limited molecular knowledge of the parasite stages associated with the intermediate host (Fig. 3B).

Relevant changes in gene expression might not be limited only to highly expressed genes; subtle changes in key regulatory genes might underly relevant developmental stage transitions. For this reason, we used *Clust* [20] to cluster genes with similar patterns of co-expression across the cycle (Fig. 4). Seven patterns of gene expression transitions were identified corresponding to genes preferentially upregulated in intra-snail stages (Fig. 4A-B), intra-mammalian stages (Fig. 4F-G), or having profiles with no clear association with specific biological conditions (Fig. 4C-E). These results may reflect the existence of differential regulatory programs being expressed, allowing the adaptation to the specific environments in each host. Our results are similar to those reported for *F. gigantica* where, rather than stage-specific, clusters of co-expression are associated with the major developmental transitions of the life cycle [17].

To further analyze gene expression across the life cycle, we classified all expressed genes based on their differential expression following a similar approach previously applied for *F. gigantica* [17]. We identified a set of 1784 Highly Variable Genes (HVGs), defined as those that differ 16 times in their expression value over the mean ($|\log_2 FC| \geq 4$ and FDR < 0.01) between any pair of

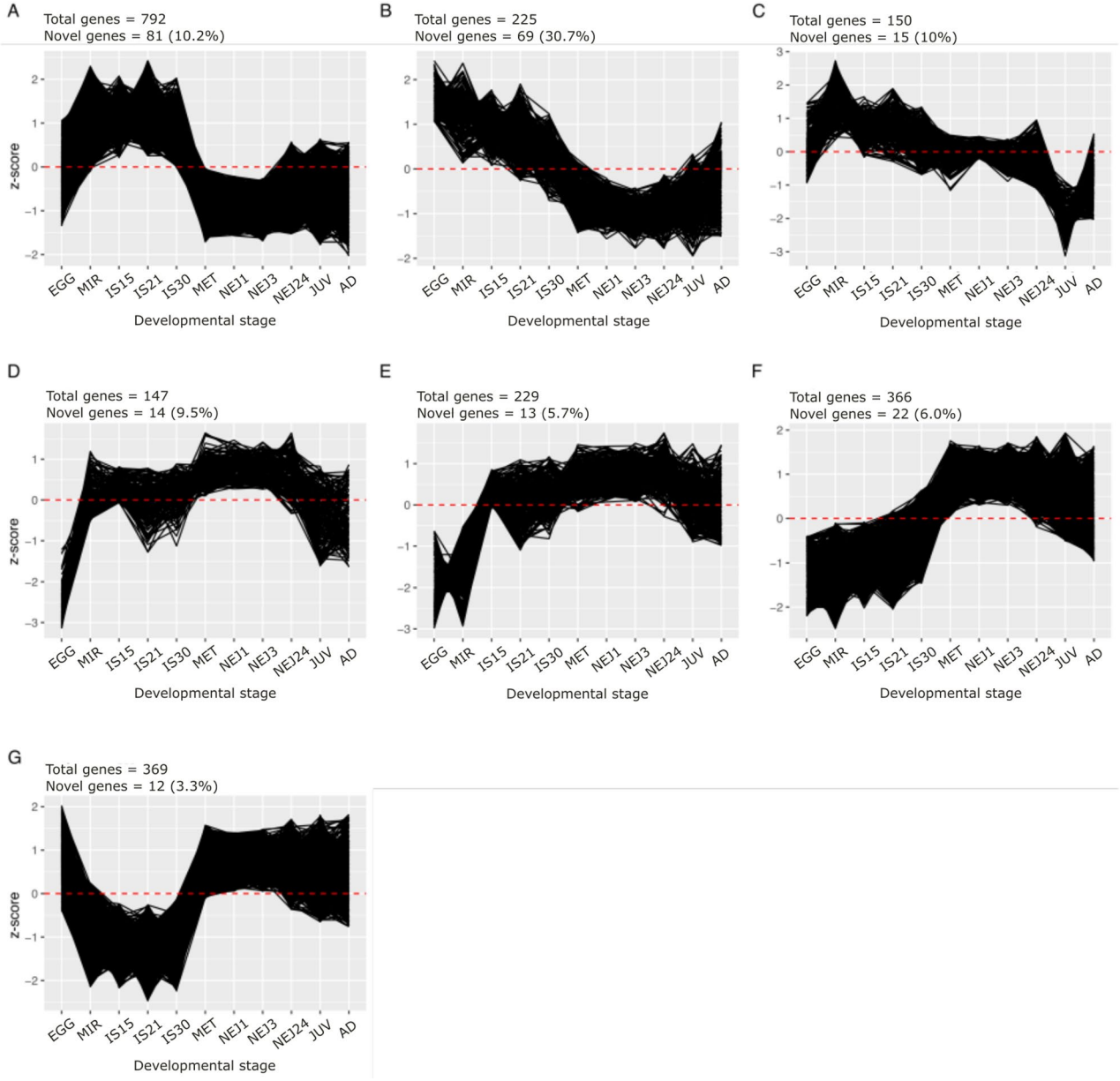

Fig. 4 Co-expression gene clusters along the life cycle of *F. hepatica*. Clusters of co-expressed genes based on the normalized expression (z-score) across all the developmental stages were generated with Clust. Genes with stable gene expression along the entire life cycle were filtered out. For each cluster the total number of genes and those that are novel are indicated over each graph

conditions. In contrast, we identified a set of 2431 genes with low variability coefficient of expression across stages (VC < 0.5). These Stable Expressed Genes (SEGs) were used as controls for further analyses (Fig. 5A).

By definition HVGs are over-expressed in certain stages, and consequently they could be classified by a simple *k*-means algorithm into groups that follow the different stage transitions (Fig. 5B). Sets of highly expressed genes in egg and miracidia are evident. Although intra-snail stages showed a similar pattern, early and late expressed genes could also be differentiated. Metacercariae and NEJ collected at different times showed a restricted and shared set of genes, while immature and adult intra-liver stages showed an overlapping set of HVGs. HVGs fell into very diverse functional annotation categories, but there was no evident enrichment of any category across the diverse clusters. A comparison of HVGs from *F. hepatica* and those detected in *F. gigantica* showed a reasonable overlap suggesting that similar mechanisms and genes are upregulated at different stages of the cycle (Suppl. Figure 3).

Differential gene expression analysis reveals motility related genes in miracidia

To analyze gene expression associated with host invasion and infection establishment in the snail we focused on the comparison of miracidia with early (15 day) intra-snail larval stages, and between early and late (15 and 30 day) intra-snail stages. Samples from 21-day intra-snail stages were excluded from the comparison since we want to focused on the time point samples that show more morphological observed differences in their contents. Furthermore, the IS21 samples were obtained and sequenced separately, so batch effects could mask differences observed.

Differential gene expression between miracidia and 15-day intra-snail stages (IS15) revealed 330 genes upregulated in miracidia, and 1215 genes in the IS15. Although no GO terms or KEGG pathways were functionally enriched in miracidia, several genes related to the infective processes, signal transduction, and movement in response to external stimuli, as well as carbohydrate metabolism, were upregulated in miracidia.

Various genes associated with motility and ciliary proteins were upregulated in miracidia (Fig. 6A, cyan and green), consistent with the active role of the miracidium ciliary plates during snail-host seeking [21]. These include structural proteins like alpha tubulin (TUBA), the ciliary rootlet coiled coil protein (rootletlin, CROCC) [22], as well as other genes associated to the ciliary function like the radial spike protein RPS4, Gas8 involved in the nexin-dynein regulatory complex (NDRC), or CDCC39 involved in regulating ciliogenesis [23]. Genes associated with ciliary function were also upregulated in *S. japonicum* miracidia [24], consistent with common mechanisms across trematodes.

Some genes involved in calcium mediated transduction like calmodulin (CaM), calcium/calmodulin kinase

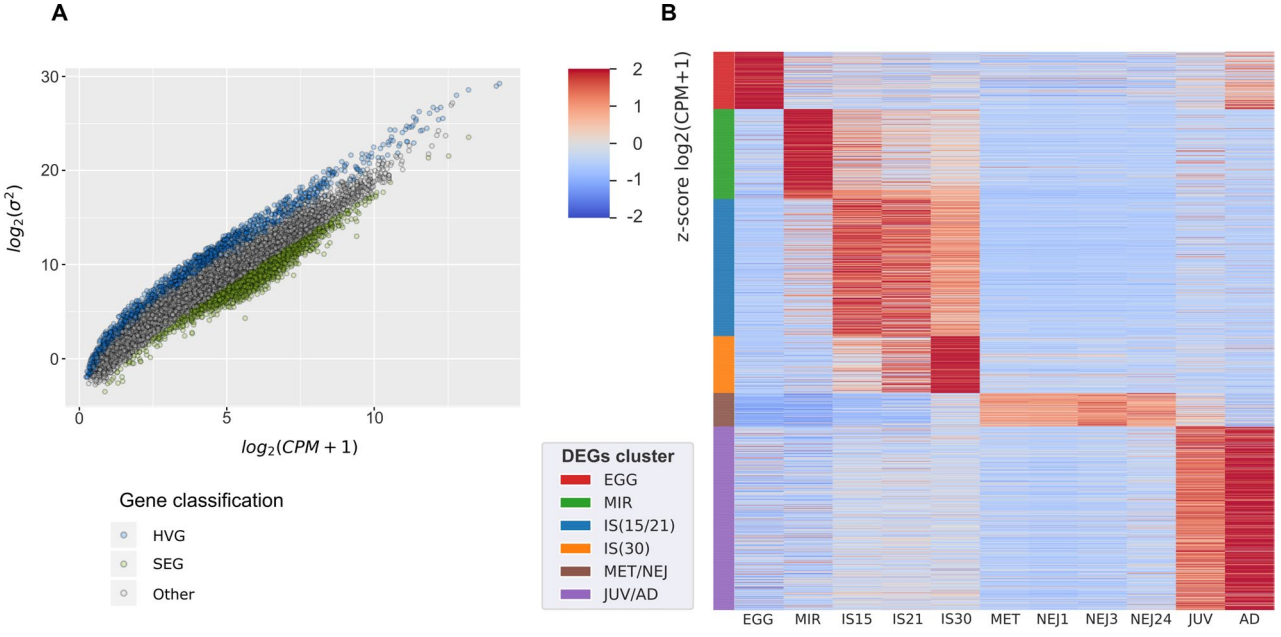

Fig. 5 Expression variance across the *F. hepatica* life cycle. **A** Ratio of average expression of a gene and its variance across the life cycle. Genes with a Highly Variable Expression (HVG, |log2FC| ≥ 4 & FDR < 0.01) are represented in blue, Stable Expression Genes (SEG) are indicated in green. **B** Expression patterns of GDAV cluster genes along the life cycle of F. hepatica. The clusters of HVG (log2FC > = 2 and FDR < 0.01) were generated by a k-means algorithm (k = 6). The six different clusters detected are indicated by color bars (left)

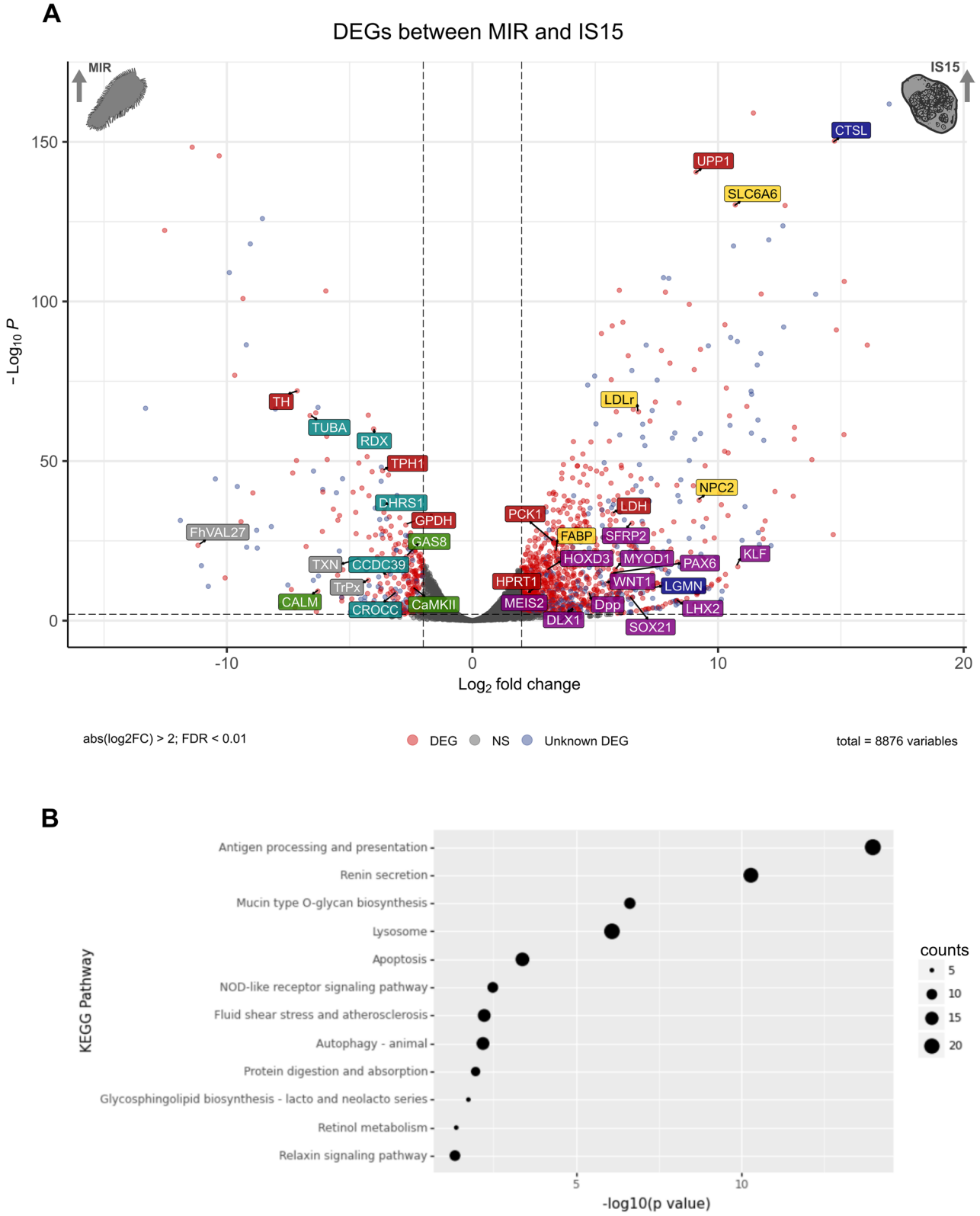

Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 Differential expressed genes (DEGs) between miracidia and IS15 stages. **A** Volcano plot of DEGs, defined by a |log2FC|>2 and FDR < 0.01 after applying the Benjamini-Hochberg correction to control false positives. DEGs with annotation are indicated by red dots, whereas those devoid of annotation are indicated by blue dots. Genes mentioned in the text with a documented role in miracidia and IS15 are colored (red, metabolic genes; blue, proteinases; green, calcium signaling; purple, development genes; grey, host-parasite interface; yellow, lipid metabolism genes). **B** Bubble plot showing the KEGG pathways enriched in genes upregulated in IS15. Bubble size depicts the gene count in each term. An interactive version of this plot could be found as IP2 in the supplementary Zenodo repository

II (CaMKII) and the kinase regulatory unit PKA-R were also upregulated in this stage. Ciliary movement is associated with calcium signaling in schistosomes mediated by cAMP [25, 26], and consistently *F. hepatica* miracidia can be immobilized by exposure to chelating agents like EDTA. Furthermore, egg hatching in schistosomes is dependent on calcium and inhibited by CaM antagonists, stressing their relevance at this stage [27, 28].

Genes involved in the biosynthesis of serotonin and dopamine neurotransmitters like tryptophan hydroxylase (TPH1) or tyrosine hydroxylase (TH) were also upregulated in miracidia. These neurotransmitters may be critical during the muscle contraction that propels this invasive larva [29]. Additionally, genes involved in energy metabolism (i.e. GPDH, highlighted in red), and the regulation of redox homeostasis (grey in Fig. 6A) including thioredoxin (TXN), peroxiredoxin (TrPx), or the oxidoreductase DHRS1 were upregulated, emphasizing the high energetic requirements of the stage, and the consequent increase in the reactive products of oxygen.

Gene expression enriched in early intra-snail stages highlights development, morphogenesis and associated resources demands

Functional enrichment of 1235 genes upregulated in IS15 stages highlighted biological processes occurring within the intermediate host (Fig. 6B). The presence of terms such as "*Protein digestion and absorption*", "*lysosome*" and "*antigen processing and presentation*" were associated with the overexpression of genes encoding proteases (highlighted in blue in Fig. 6A) mainly cathepsin and legumain families. The role of these proteins in host invasion, migration and infection establishment in intra-mammalian stages has been well characterized [30, 31]. In addition, their involvement in protein digestion, for both facilitating migration and obtaining nutrients for anabolic processes during intra-snail developmental stages, has been described for *S. mansoni* [32].

Pathways such as "*ECM-receptor interaction*", "*Neuroactive ligand-receptor interaction*" or *"Cell adhesion molecules"* could also be related to the migratory process in intra-snail stages, as interaction with the host's extracellular matrix and neuromotor control are necessary for locomotor function (Fig. 6B).

The observed enrichment of the "*signaling pathways regulating pluripotency of stem cells*" term, is consistent with the differential expression of several genes (highlighted in purple in Fig. 6A), belonging to the Wnt/beta catenin and BMP signaling pathways, known players in determining axis patterning [33–35].

Intra-snail stages are characterized by asexual proliferation and morphogenesis [9], supported by the presence of germinal cells — neoblast-like cells —described in other trematodes [36, 37]. The developmental processes of the rediae and cercariae imply the establishment of polarity along body axes. Consequently, it is not surprising to find genes associated with these developmental processes, including Wnt, Hox and Fox genes upregulated in IS15 relative to miracidia. The Wnt pathway is also relevant in development and growth of stages associated with the definitive host in *F. hepatica* (Fig. 6A).

Since proliferative cells are essential for these developmental processes, we scrutinized the expression of several genes described as neoblasts markers in other flatworms [37–40]. In general, most of the germ cell markers studied show an expected peak expression in miracidia or early intra-snail stages, due to the proliferation and differentiation of these cells in these stages of *F. hepatica* (Suppl. Figure 4). As was also expected, the observed trends were generally concordant between homologous stages of *F. hepatica* and *F. gigantica*, suggesting conserved regulatory mechanisms. This result suggests that the germ cells in these stages of *F. hepatica* may be regulated by conserved germline multipotency program (GMP) genes, similar to those of model trematodes [40, 41].

A unique aspect of trematode metabolism is their inability to perform de novo purine synthesis [42]. The clonal expansion of sporocysts and rediae entails a great investment in nucleic acid synthesis. Notably, several genes involved in purine salvage pathway (i.e. HPRT1, ADK, ADSS, ADA) were found between those highly expressed in intra-snail stages (Fig. 6A). Furthermore, genes encoding purine transporters like SLC43A4 were also among those with the highest expression in these stages (Fig. 7A). Interestingly, some genes involved in the purine salvage pathway (i.e., genes encoding the enzymes PNP, ADK, and ADSL), showed paralogs differentially expressed between intra-snail and intra-mammalian stages. This may suggest differences in purine metabolism between mollusk and mammalian hosts (Suppl Fig. 5).

The detection of differentially expressed genes encoding lipid transporters (FABP, LDLR, and NPC2), as well

as lipases and phospholipases, in the miracidia-IS15 transition (Fig. 6A), suggests a critical demand for these nutrients. Recent genomic studies have confirmed the presence of genes belonging to the anabolic and catabolic lipid pathways in *F. hepatica* [43]. In this regard, it is noteworthy that all genes encoding proteins involved in the beta-oxidation of fatty acids were expressed in the larval stages of the parasite (Suppl Fig. 6). These findings are consistent with the hypothesis that intra-snail stages of *F. hepatica* are capable of obtaining lipids from the host for both biosynthetic purposes and energy acquisition. This feature might differentiate digestive-borne trematodes to the more restricted metabolic capacities of *Schistosomatidae* species [43].

Since endogenous glycogen reserves could be rapidly depleted, it has been proposed that trematode sporocysts could parasitize host glycogen stores in vivo. Differential expression of genes coding for glycogen phosphorylase (PYG) and glycogenin (GYG) in the miracidium transition and IS15 is consistent with this notion. Also, genes associated with carbohydrate fermentation pathways and Krebs cycle show elevated expression in intra-snail stages. Differential expression of a LDH gene enables the parasite to obtain energy under hypoxic conditions through homolactic fermentation [44]. Furthermore, genes associated with malate dismutation and catabolism of protein precursors are expressed in intra-snail stages, suggesting alternative pathways of obtaining energy [30, 45]. It has been shown that *S. mansoni* sporocysts use the amino acid glutamine to provide intermediates to the Krebs cycle via the PEPCK enzyme [46], and the synthesis of this amino acid is stimulated in Bge cells of *Biomphalaria glabrata* in co-culture with *S. mansoni* sporocysts [47]. Interestingly, PEPCK is also upregulated in IS15 of *F. hepatica*, suggesting that a similar mechanism might be taking place.

Proteases and glycoproteins relevant for encystment are highly expressed in late intra-snail stages

While the transition between miracidia and IS15 highlights several changing aspects, the transition between intra-snail stages (i.e. IS15 and IS30) is less dramatic, highlighting only 44 genes upregulated in the earlier and 412 in the latter stages (Fig. 7A). This is expected, since intra-snail developmental stages show high transcriptional similarity to each other (see above). Although no functional enrichment was found within the few IS15 upregulated genes, many of them are linked to developmental pathways.

An inspection of the genes upregulated in IS30 stage showed the ubiquitous presence of proteases from the cathepsin B and L families to legumains Their presence is reflected in the enrichment of terms such as "Antigen processing and presentation", "Lysosome", "Apoptosis", "Autophagy - animal", and "Protein digestion and absorption" (Fig. 7A highlighted in blue. Figure 7B). Interestingly, members of these families were also upregulated in IS15 compared to miracidia. and during the sporocyst-to-cercariae transition in trematodes such as *F. gigantica* and *S. japonicum* [17, 24]. This is expected, as proteases constitute the main component of trematode cercarial secretions [30]. Their overexpression in cercariae suggests a role in the final stages of intra-snail migration host egress. The observation of co-expression with members of the legumain family is consistent with the postulated regulatory role for members of this family in cathepsin function [48]. Furthermore, differential expression of members of protease inhibitor families such as serpins and Kunitz-type inhibitors was also observed in intra-snail stages.

An enrichment in biosynthetic pathways of carbohydrate compounds like "Mucin type O-glycan biosynthesis" or "Glycosphingolipid biosynthesis - lacto and neolacto series", was detected consistent with the expression of genes linked to mucin biosynthesis in late cercarial stages (Fig. 7A, highlighted in red and Fig. 7B). Products of mucin biosynthesis constitute the inner layer of the metacercarial cyst. In Plagiorchiidae, the glands secreting mucopolysaccharides to the cercarial tegument (future outer layer of the metacercarial cyst) and the cystogenic gland that produces the material for the inner layer of the cyst are synthesized during cercarial morphogenesis [49]. Previous genomic and transcriptomic studies indeed described a complete set of genes involved in biosynthetic pathways for these polysaccharides in the *F. hepatica* genome [50], and the presence of mucins in the liver fluke NEJs [51]. Larval stages also expressed genes required for N-glycan synthesis, although they were not differentially expressed. N-glycosylated proteins, both secreted and present in suckers and spines, and on the surface of the tegument of *F. hepatica*, has been reported, thus constituting a first line of interaction in the complex host-parasite interface [52]. The detection of differential expression of genes encoding enzymes involved in the synthesis of N- and O-glycans pathways (such as C1GALT1, CGNT1, GALNT, and B4GALT2), may be associated with parasite requirements of a rich glycoprotein covering of the tegument, which could be critical for the mammalian host invasion.

Several genes encoding structural proteins such as β- and β-actin (ACTB, ACTB/G), dynein light chains (DYNL), myophilin, actin-binding proteins (gelsolin), or microtubule dynamics regulators (RMDN1) are also highly expressed in the later intra-snail stages. These genes could be linked to the regulation of movement, necessary for cercarial emission and also for the regulation of vesicle movement. Consistent with this view is the differential expression of structural components of

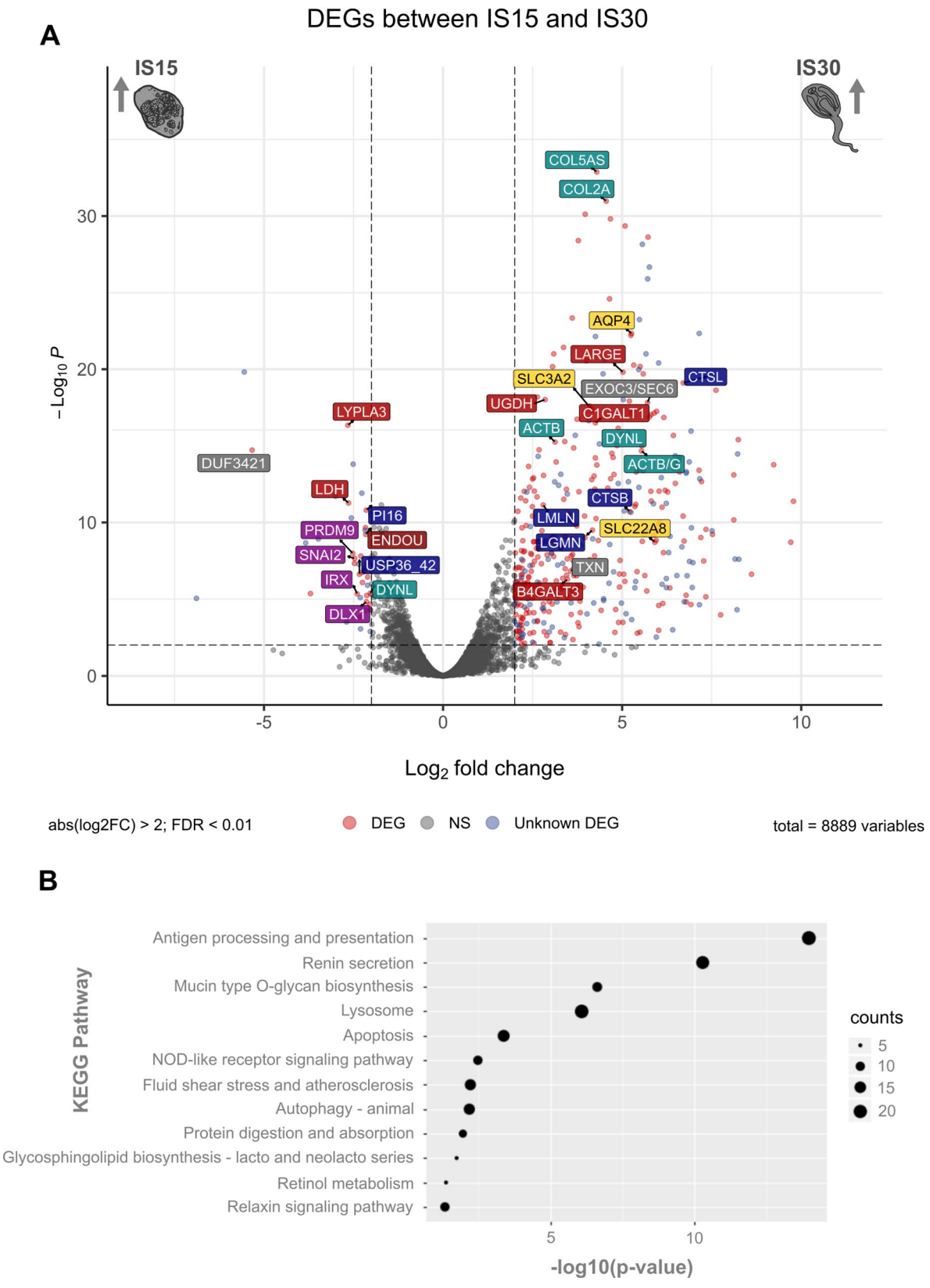

Fig. 7 (See legend on next page.)

(See figure on previous page.)

Fig. 7 Differential expressed genes (DEG) between IS15 and IS30 stages. **A** Volcano plot of DEGs, defined by |log2FC|>2 and FDR < 0.01 after applying the Benjamini-Hochberg correction to control false positives. DEGs with annotation are indicated by red dots, whereas those devoid of annotation are indicated by blue dots. Genes mentioned in the text with a documented role are highlighted with a color code (red, metabolic genes; blue, proteinases; green, calcium signaling; purple, development genes; grey, host-parasite interface; yellow, lipid metabolism genes). **B** Bubble plot showing the KEGG pathways enriched in genes upregulated in IS30. Bubble size depicts the gene count in each term. An interactive version of this plot could be found as IP3 in the supplementary Zenodo repository

extracellular vesicles (EXOC3/Sect. 6, SEC22B), which may be associated with the accumulation in vesicles of elements necessary either for cyst formation or, subsequently, for the initial interactions with the mammalian host once the parasite emerges from the cyst. Also structurally related to the tegument formation are two of the top DEGs, encoding collagen isoforms (COL2A and COL5AS), components of cercarial glycocalyx fibrils [53].

In addition, amongst genes differentially expressed we identified members of the aquaporin gene family (e.g., AQP4), key players in osmotic regulation during environmental changes associated with the life cycle [14]. Other solute transporters (SLC3A2, SLC22A8) could facilitate the exchange of small molecules such as glycerol, relevant at these transitions.

Recurrent gene families expressed across the life cycle

The terms "Protein digestion and absorption" are recurrently enriched in intra-snail stages (Figs. 6 and 7), consistent with the increased expression of numerous secreted proteases, mainly from the legumain and cathepsins B and L families. While the involvement of these gene families in definitive host infection is well known, their roles in the context of the snail infection and interaction remains poorly explored. For example, cathepsins L constitute the predominant secretion products of the intra-mammalian stages of *F. hepatica*, with proteolytic activities towards different substrates including hemoglobin, fibrinogen, and collagen in these hosts [54–57]. Several studies show a temporal progression in the expression of cathepsins L and B family members during invasion and infection establishment [14, 31, 58].

The differential expression of the cathepsin gene families during miracidium-sporocyst transition, strongly suggests that they may play a similar role in the snail. Members of these protease families during the sporocyst-to-cercaria transition have also been reported in other trematodes such as *F. gigantica* and *S. japonicum* [17, 24]. However, it was not known if the same genes were expressed in both intermediate and definitive host invasion, or if different members of the gene family were involved in similar functions. To unravel this, we generated phylogenetic trees with all the identified members of the families and analyzed their expression profile (Fig. 8). The data generated in this study showed that some family members, particularly subgroups CL8/9 and CL10, for which no transcriptomic or proteomic information was previously available, exhibited differential expression in intra-snail larval developmental stages. Interestingly, different paralogs within the same families are upregulated in early and late intra-snail stages both in *F. hepatica* and *F. gigantica*.

An adaptive role has been attributed to the lineage specific expansions of cathepsins observed in *Schistosomatidae*, *Opisthorchiidae*, and *Fasciolidae* (Fig. 8) [43, 57, 60]. The integration of transcriptomic and proteomic data into phylogenetic analysis supports the idea of functional divergence, with differential expression at different developmental stages in both *F. hepatica* and *F. gigantica*. The grouping of different family members according to their expression pattern in both *Fasciola* species suggests, following the parsimony criterion, the existence of at least four different genes in the common ancestor of the species of the suborder Echinostomata (*F. hepatica*, *F. gigantica*, *F. buski*, *and E. caproni*), with different expression patterns (egg, intra-snail stages, and invasive and adult stages in the mammalian host). The expansion of these ancestral genes and the conservation of their expression pattern could broadly explain the observed pattern in our dataset. Functional duplication and divergence is an established evolutionary mechanism and has been studied in multiple families of secreted proteins [61–64]. Some authors propose the evolutionary process of these families in two phases: a first stage, the retention of gene copies that provides some type of selective advantage (e.g., increased dosage of the encoded product [62]), and a second stage where functional redundancy between copies allows functional divergence over time without influencing the fitness of the carrier organisms [65, 66]. Studies of the active sites of genes expressed in mammalian stages show functional differences associated with changes at a few positions that result in different substrate preferences [55, 56, 67, 68]. Preliminary observations showed changes at conserved sites in the genes expressed in intra-snail stages, suggesting that different specificities may also exist here.

Cathepsins are synthesized as inactive proenzymes, and may be activated by aspartyl-endopeptidases, known as legumains, that also constitute a multigene family [48]. Interestingly, the integration of transcriptomic data from larval stages with the phylogenetic pattern inferred for legumains also showed diverse clades with differential expression across stages, similar to that observed for cathepsins. Likewise, the phylogenetic analysis suggests

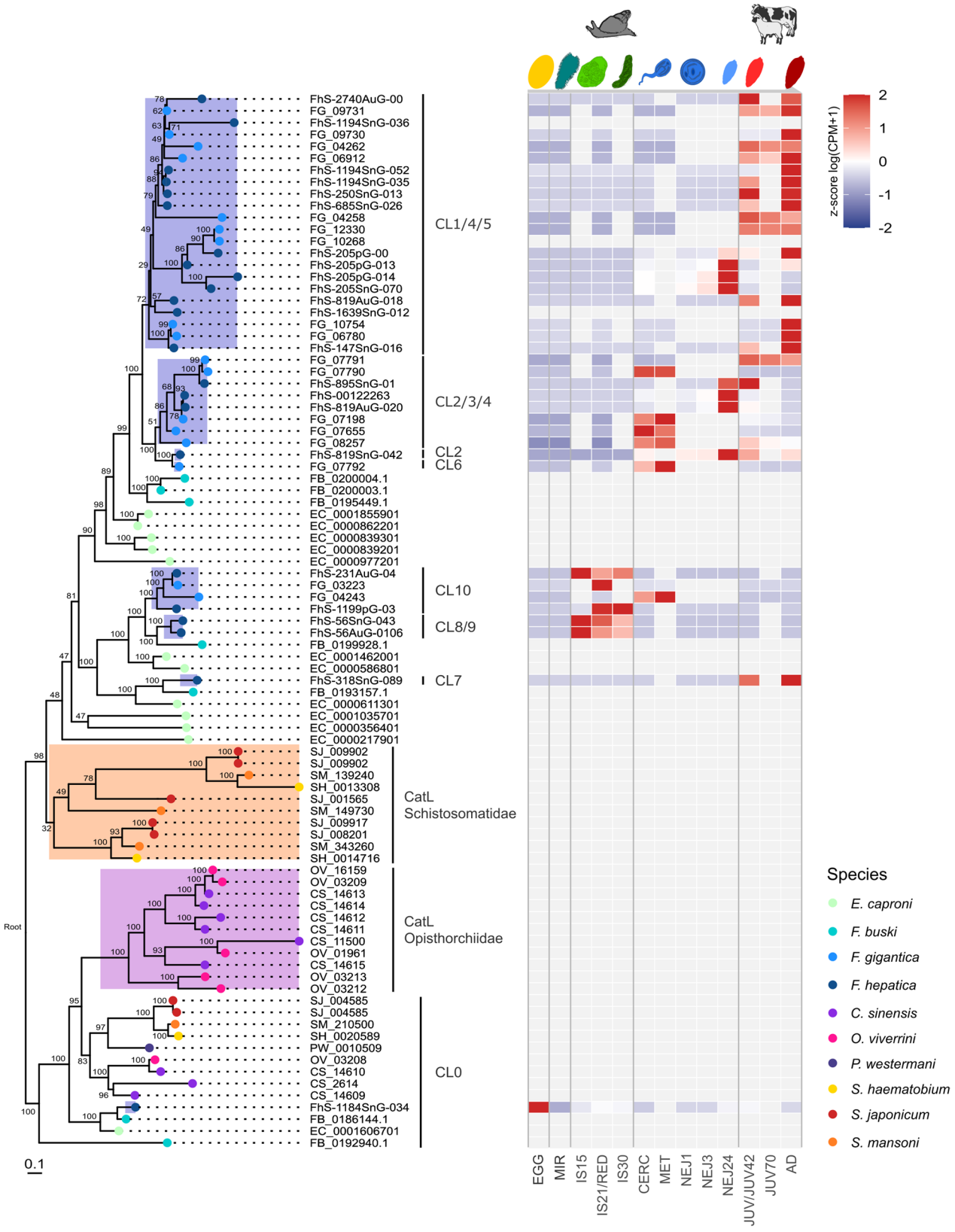

Fig. 8 Phylogeny of the cathepsin L multigene family. Inferred phylogenetic tree of diverse members of the cathepsin L multigene family within selected trematodes. Diamonds represent lineage specific collapsed branches (number of genes indicated). The tree was midpoint-rooted using the ape library (R; [59]). Expression trends throughout the life cycle for genes encoded in the genomes of *F. hepatica* (blue dots) and *F. gigantica* (sky blue) are indicated

that two legumain copies were present in the common ancestor of the Echinostomata suborder (Fig. 9).

The divergence observed in both multigene families at the level of gene expression, with some members expressed in intra-snail developmental stages and others in intra-mammalian developmental stages, is compatible with a scenario of coevolution between both families. The intra-snail specific proteases might be playing

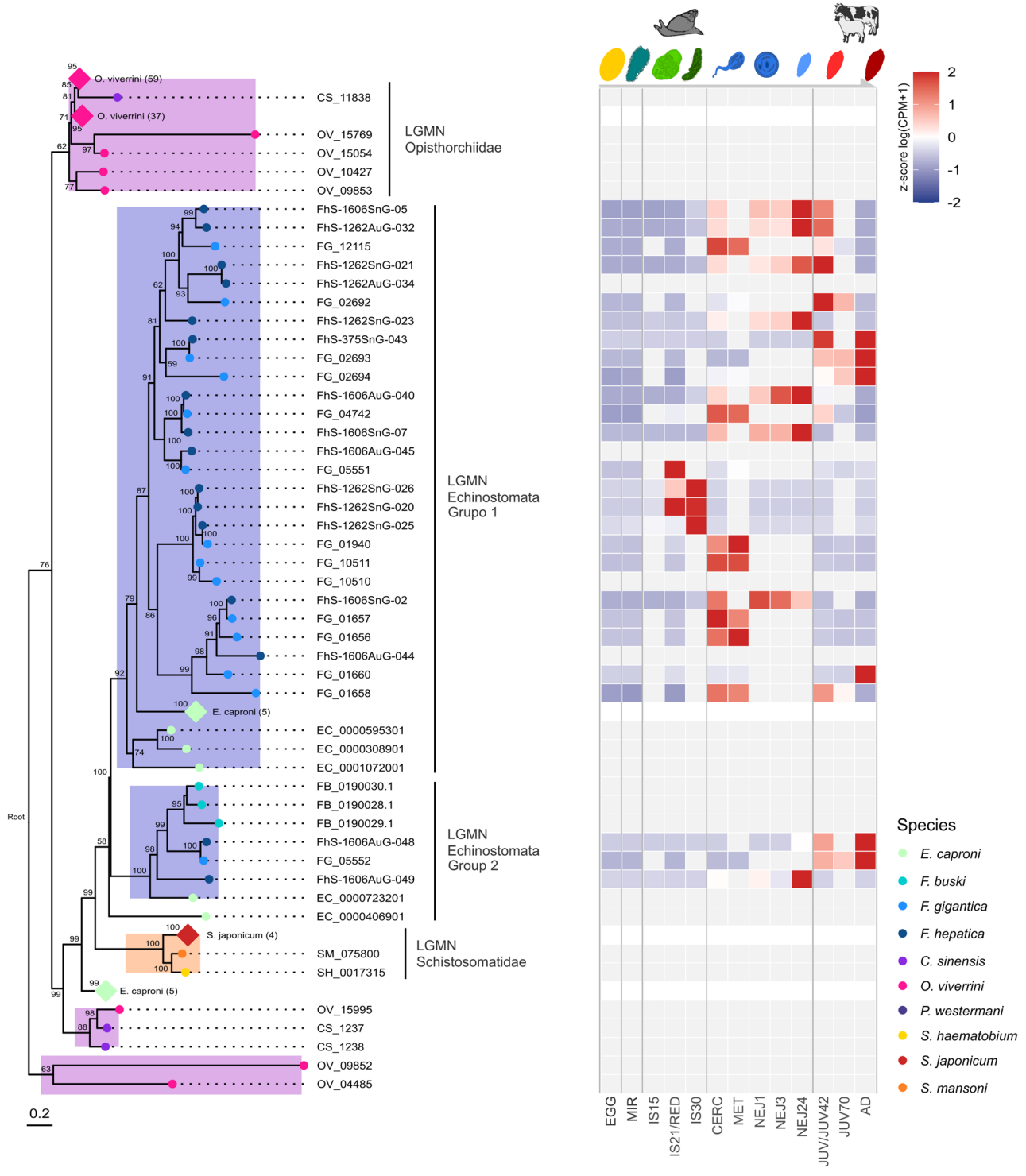

Fig. 9 Phylogeny of the legumain multigene family. Inferred phylogenetic tree of legumain multigene family members across selected trematode species. Diamonds represent lineage specific collapsed branches (number of genes indicated). The tree was midpoint-rooted using the ape library (R; [59]). The expression trends throughout the life cycle for genes encoded in the genomes of *F. hepatica* (blue dots) and *F. gigantica* (sky blue) are represented

similar roles in parasite host invasion as those described for intra-mammalian stages, that is, establishment and nutrient uptake. *F. hepatica* infected snails tend to be larger and have a reduced capacity to produce eggs than uninfected counterparts. This has been associated with the migration of redia and castration of the snail by direct ingestion of the ovotestis [69], a phenomenon where these proteases might play a leading role.

Hemoglobin constitutes the main protein component of the hemolymph of trematode snail hosts (i.e., ~ 70–80% in *B. glabrata*) [70], and consequently would be one of the main substrates for intra-snail proteases, facilitating the amino acids necessary for the biosynthetic processes of larval development, in a scenario similar to that proposed for *S. mansoni* [71]. Host protein digestion is mainly extracellular, although it has been suggested a final phase of endo-lysosomal digestion is facilitated by leucine aminopeptidases (LAPs) [30]. In addition, we have previously revealed a tentative role for LAPs during the process of schistosome egg hatching [72]. Several of these proteases were differentially expressed in the later intra-snail stages, consistent with proteases being the main component of trematode cercarial secretions [30]. Their overexpression in cercariae suggests a role for members of these families in the final stages of intra-snail migration and host egression.

In the later stages differential expression of members of protease inhibitor families such as serpins and Kunitz-type inhibitors was also observed (Fig. 10). Different clades within this inhibitory families showed differential expression profiles among developmental stages. Two independent clades showed preferential expression in intra-snail stages, significantly higher in later stages (IS30) of *F. hepatica* or in cercariae and metacercariae of *F. gigantica*. Further experimental studies, including RNAi-based experiments are needed to dissect in more detail the role of the diverse proteases and inhibitors in the intermediate host.

Within the upregulated genes in miracidia, several members of the SCP/TAPS (also known as CAP) protein superfamily were found. These proteins have been described across all classes of flatworms and other helminths, particularly in parasitic species [73], where the VAL ("Venom Allergen Like") family from *Schistosoma mansoni* has been extensively studied [74]. Proteins of these family have been previously detected in the mammalian invading stages of *F. hepatica* [13]. Our inferred phylogenetic tree of the VAL family suggested independent gene amplifications in trematodes (e.g., in the families *Schistosomatidae* and *Fasciolidae*) (Fig. 11), which may suggest adaptive value for host-parasite survival.

Expansions/duplications before and after speciation could indicate long term and ongoing diversifying selection [75]. Notably, similar SCP/TAPS gene family expansions have been observed in the cestode *M. corti* [76]. Furthermore, the phylogenetic tree suggested that at least four paralogous sequences were present in the common ancestor of the Echinostomata suborder (*E. caproni*, *F. hepatica*, *F. gigantica*, and *F. buski*). Three of the clades were represented by single or low copy family members from both *F. hepatica* and *F. gigantica* with concordant expression patterns in miracidia, the cercarial to metacerarial transition or adults. The remaining clade represents a large gene expansion with predominant expression in miracidia, and minor isoforms expressed in other developmental stages. The increased copy number driven by gene duplication may have led to functional redundancy, easing selective constraints and enabling functional divergence [62, 65].

Consistently, it has been shown in the family Schistosomatidae both temporal and spatial regulation of SmVAL gene expression, with specific members expressed in the tegument, esophagus, oral and ventral suckers, and acetabular glands of cercariae [77–81]. Particularly, SmVAL9 has being detected by proteomic analyses as secreted during the miracidia-to-sporocyst transition, as well as being characterized as expressed in the parenchymal, tegumental and germinal masses of miracidia [82]. More recently, single cell atlas approaches of *S. mansoni* miracidia [83] and sporocysts [84] consistently revealed the expression of several VAL genes in more than 60% of larval gland cells. Unfortunately, there is scarce information of this gene family within *Fasciolidae*, with the detection of VAL-encoding genes in the mammalian infective juveniles (NEJ) [13] and the detection of tegumental localization of FhVAL11 by lectin array analysis [52]. The differential expression of SCP/TAPS family members in miracidia and intra-snail stages of *F. hepatica* here warrants more detailed studies of their roles at the host–parasite interface or in intra-snail stages developmental regulation.

Conclusions

Digenetic trematodes complete their life cycles within an obligatory molluscan intermediate host and a vertebrate definitive host. Despite their biological singularities and relevance, the intra-molluscan stages have historically received less research attention [49]. Elucidating mechanisms of asexual proliferation within the snail is critical, as this process directly amplifies the number of infective larvae. Likewise, studies of the molecular basis underpinning the maintenance and development of pluripotent stem cells—present throughout the life cycle—may point to an attractive Achilles heel of this pathogen. Recent studies have started to disentangle molecular aspects underlying the biology of the intra-snail stages, mainly in the model species *S. mansoni* [37, 83–87].

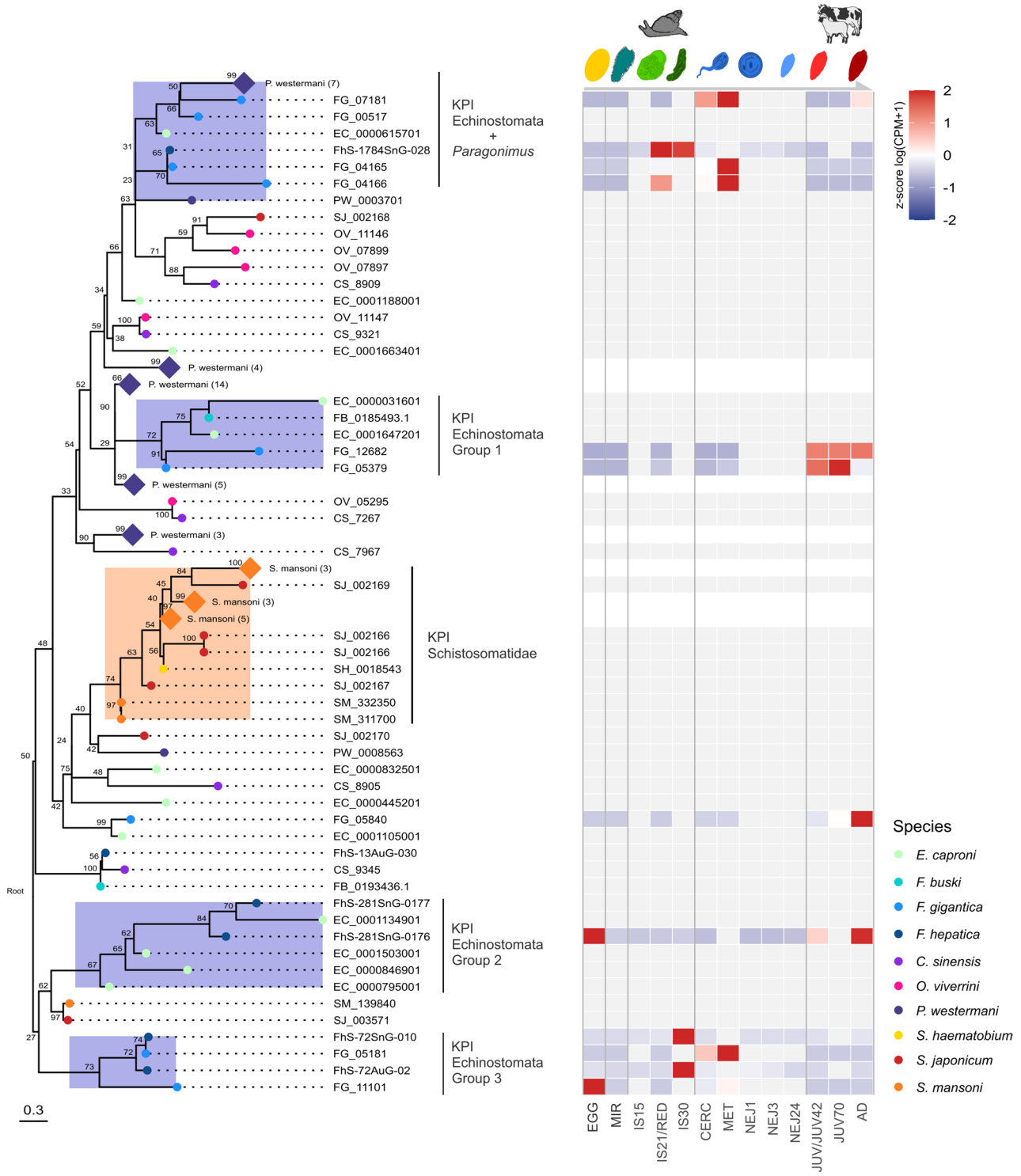

Fig. 10 Inferred phylogenetic tree for members of serine-protease inhibitors of the Kunitz type. The tree was midpoint-rooted using the ape library in R [59]. The expression trends throughout the life cycle for genes encoded in the genomes of *F. hepatica* and *F. gigantica* are represented

Here we present transcriptomic data from the invasive and intra-snail developmental stages of *F. hepatica*, compared these data to those publicly available from the sister species *F.gigantica* [17], and performed several comparative transcriptomic analyses. The integration of data from all available developmental stages of *F. hepatica* led to a complete transcriptomic view of the life cycle, with multivariate analyses showing differences between

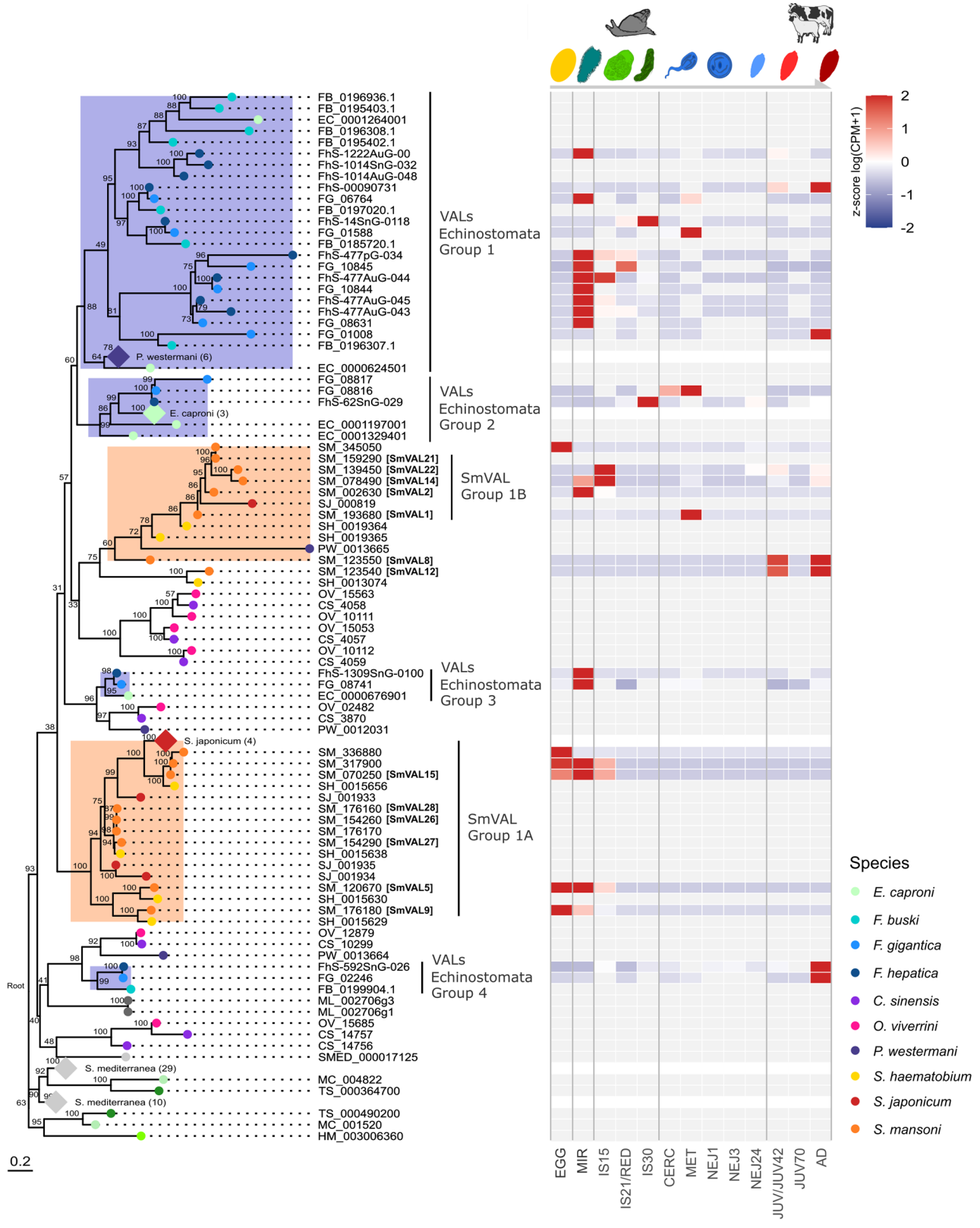

Fig. 11 Inferred phylogenetic tree for members of the SCA/TAPS (VAL) proteins across different platyhelminths. The tree was midpoint-rooted using the ape library R; [59]. Expression trends throughout the life cycle for genes encoded in the genomes of *F. hepatica*, *F. gigantica* and *S. mansoni* are indicated

free living and parasitic stages. The clustering of biologically similar developmental stages in five distinct groups is repeatedly observed either when analyzing the most highly expressed genes in each stage, when inferring gene clusters with similar expression patterns, or when grouping differentially expressed genes with the highest proportion of change between stages.

The study of differentially expressed genes among intra-snail stages highlighted several functional groups already reported as relevant in trematode biology, as proteins involved in the execution and coordination of the miracidium motor complex developmental genes, proteases with tentative roles in snail infection, migration, and protein digestion within the snail, or proteins with antioxidant roles that may interact with the host immune system. Interestingly, in intra-snail developmental stages we also identified the overexpression of group of genes linked to metabolism, morphological development, and the multipotency regulatory program and the purine salvage pathways highlighting key processes for survival and asexual proliferation.

Combining expression data with phylogenetic analysis we observed that within expanded multigene families such as cathepsins L, legumains or VAL proteins, different clades were expressed in intra-snail and intra-mammalian stages, presumably conferring selective advantages within each environment. Similar patterns were observed in multigene families for which little functional characterization is available. Further characterization of these families will shine light into parasite's adaptation to its intermediate host.

Our findings provided an initial transcriptomic picture of intra-snail stages, offering a framework to compare these processes across different trematodes. A more detailed picture of these processes is being obtained in Schistosomes with transcriptomic studies at cellular resolution [41, 83, 88, 89]. Applying these approaches to liver fluke will inform on the similarities and differences in the early development across trematode lineages. These single cells approaches can be combined with spatial transcriptomics, recently made available for the adult stage of *F.hepatica* [16] to offer a fine grained picture of parasite development and adaptation. The application of functional approaches such as RNA interference and genome editing [90] will provide valuable information for some of these genes in *F. hepatica*, revealing tentative novel targets for control strategies.

Methods

Parasite collection, RNA isolation and sequencing

Fasciola hepatica eggs obtained from gallbladders of naturally infected cattle collected at a local abattoir were embryonated in the dark in fresh sterile water for 2 weeks at 20 °C and activated with light. Freshly emerged miracidia were collected and frozen for RNA extraction, or used for in vitro infection of Lymnaea viatrix snails from a snail colony maintained at the lab. Eighty individual snails (3–4 mm length) were exposed to 5 miracidia, and maintained at 25 °C [91]. Infected snails were dissected at 15-, 21- and 30-days post-infection (named IS15, IS21 and IS30) and all the available parasite material were recovered and stored in Trizol until extraction. It worths mentioning that redial and cercarial development within snails is asynchronic, with primary mother rediae and daughter rediae coexisting early in infection, and cercariae being produced later in infection [10]. Consistent with this, we observed under the stereoscope that the early samples (IS15) contained rediae (either primary or secondary), with no cercariae detected, while the late sample (IS30) were enriched in cercarial containing rediae, and cercariae (Fig. 1). Consequently, the samples here analyzed does not represent strictly pure stages but biological proxies of the main representative forms found at different time points of infection.

Samples from the same time point were pooled to generate triplicates and total RNA was extracted using the DirectZol RNA kit. Replicates from each condition were sent for sequencing. A total of 11 samples were sequenced, either with standard Illumina PE75 kits at the Welcome Sanger Institute, or using the TruSeq Stranded mRNA LT Sample Prep Kit (*PE100*) at Macrogen according to the scheme indicated in Suppl. Table 1. Sequencing reads were deposited at SRA, accession PRJNA1039822.

Transcriptome assembly

Reads were processed with Trimmomatic v0.38 [92] removing adapters, sequences shorter than 75 bp and those with PHRED score < 30 at the beginning and end of sequences. Quality control was performed on raw and trimmed reads with FastQC v0.11.8 [93]. A single report compiling quality controls was obtained with MultiQC v1.9 [94], available in Zenodo.

Guided transcriptome assemblies were inferred for each condition, which were integrated to obtain a comprehensive genome annotation for subsequent analyses (see Suppl. Figure 1 for a pipeline graphical depiction). Briefly, trimmed reads were mapped against the reference genome assembly (PRJEB25283 at WormBase ParaSite 15 [95] with HiSat2 v2.0.5 [96], specifying read orientation (`--rna-strandness RF` option, all other parameters running by default), SAM alignments were sorted and converted to BAM files with samtools v1.9 [97].

Transcriptomic assemblies were obtained with StringTie v1.3.6 [98] specifying read orientation (`--rf` option, all other parameters running by default). An integrated annotation was obtained with gffcompare v0.11.2 [99], running under default parameters. Transcripts encoded in intergenic regions (i.e., reported under category "u" by

gffcompare) include potential genes not previously predicted and/or identified with prior transcriptomic information. Coding DNA sequences (CDSs) were annotated for these transcripts using TransDecoder v5.5.0 (github.com/TransDecoder/TransDecoder), running under default parameters. The protein sequences encoded in these transcripts were included in subsequent protein sequence analyses, using the longest ORF in all cases. The resulting genome annotation was employed for subsequent analyses. Nevertheless, an alternative genome assembly (PRJNA179522; [43] was also employed for the study of specific multigene families.

Gene expression analyses

Transcriptomic data for all currently available life cycle stages of *F. hepatica* were mapped against the integrated transcript annotation described above using *kallisto* v0.46.2 [100]. All raw reads were processed as described for miracidia and intra-snail stages samples. The employed transcriptomic set comprises miracidial and intra-snail stages (IS15, IS21 and IS30) generated in this study, and published data from eggs, metacercariae, newly excysted juveniles (1, 3 and 24 h post-excystment), juveniles and adults [14, 15, 43, 101] (Suppl. Table 1).

Gene expression levels were obtained from estimated counts with the tximport v4.0.3 R library [102], and normalized into *transcripts per million* (TPM). For differential expression analysis, expression levels were normalized into *counts per million* (CPM) and corrected by the TMM method [103] with edgeR v3.24.3 library [104, 105]. In order to take into account the heterogeneity between life cycle stages, genes with at least 1 CPM in all replicates of at least one life cycle stage were considered for further analyses following [106]. We checked if different normalization criteria affect expression trends and compared estimated values in CPM or TPM adjusted by trimmed mean values. We found that there is a strong correlation, with minor discrepancies in genes with few counts, suggesting that expression values do not change significantly with normalization criteria. This does not affect significantly the expression patterns seen within the diverse gene families analyzed or clustering patterns observed by Clust analysis (Supplementary Data 1).

In order to assess intra- and inter-sample dispersion, multidimensional scaling (MDS) analyses were carried out on the TMM-normalized, log-normalized CPM expression matrix, as implemented in scikit-learn v1.4.2 [107]. The first three dimensions were considered for visualization. Two-dimensional scatterplots (MDS1 vs. MDS2) were generated with seaborn v0.13.2 and matplotlib v3.8.4, while interactive three-dimensional plots (MDS1–MDS3) were produced using plotly v5.24.1.

Co-expression clusters were inferred with *clust* v1.12.0 [20], using normalized expression values (TPM) and specifying a tightness parameter of t = 4; all other parameters were run by default. Finally, comparisons regarding sets of genes expressed between different stages of *F. hepatica* were made by means of Jaccard index calculation,

$$Jaccard\ index = \frac{A \cap B}{A \cup B}$$

where A and B were two sets to be compared [108], each one being the genes expressed at life cycle stages A and B, respectively.

Functional annotation and enrichment analysis

Functional annotation was performed using the e-mapper tool v2.1.5 [109] against the eggNOG v5.0 database [110], and transferring results for Gene Ontology [111], KEGG [112], PFAM v34.0 [113], COG (Bacteria), KOG (Eukarya), and arKOG (Archaea) [114]. To avoid transference of functional information from distant homologs, annotation transfer was restricted to one-to-one homolog relationships (--target_orthologs one2one mode). Domain annotation was performed with the InterProScan v5 tool [115]. The resulting complete functional annotation is provided in Suppl. Table 2. Functional enrichment of Gene Ontology (GO) terms was performed with the GOATOOLS Python library [116], tested using a Fisher's exact test and false positives controlled through the Benjamini-Hochberg FDR adjustment, reporting only cases with a significance level of FDR < 0.05.

Enriched terms were grouped using hierarchical clustering based on their pairwise semantic similarity, calculated as their Wang semantic similarity score [117]. Only terms of moderate specificity (ontology levels 4 to 9) were analyzed. For visualization, a heatmap was generated using the clustermap function of seaborn [118], plotting the calculated enrichment values (-log10(p-value)) across all *Fasciola hepatica* stages to compare the conservation of observed patterns. Functional classification of enriched terms was manually added to enhance result visualization.

Functional enrichment analyses for KEGG pathways and modules were conducted using the gprofiler2 v0.2.0 library [119] in R, with the gost() function and an FDR < 0.05 as the significance threshold.

In all cases, functional enrichment was conducted using the expressed gene set in the respective stage (i.e., genes with expression values > 1 CPM in all samples of the stage) as the reference, thus correcting for stage-specific expression biases.

Differential expression analysis

Differential expression analyses were conducted using the Degust server, which implements tools from the edgeR library [120]. A gene was defined as differentially expressed (DEG) between two stages if $|\log2FC| \geq 2$ and FDR < 0.01. In all cases, the analysis was performed on the gene set expressed in at least one of the conditions being compared (i.e., TPM > 1 in all samples of the stage).

Two differential expression analyses were conducted. First, pairwise analysis of DEGs amongst miracidia and intra-snail developmental stages (IS15 and IS30) was performed. The IS21 samples were excluded to these analysis due to their origin from a different sequencing batch. This prevents us from determining whether expression changes between these samples and their closest temporal counterparts (IS15 and IS30) have a biological basis or respond to changes in experimental conditions. Although batch correction methods do exist, we opted for a more conservative approach (excluding IS21), as we consider that the performed analyses capture major variability trends during intra-snail transitions without the need of adjusting counts (which can result in observing artefactual results). Volcano plots were generated to visualize results employing the EnhancedVolcano v1.4.0 library [121], highlighting genes relevant to *F. hepatica* developmental stages or other parasites. KEGG pathway enrichment analysis was also performed for the detected DEG sets.

Second, global differential expression analysis was performed to identify highly variable DEGs (HVGs), which present stage-specific expression profiles associated with potential functional roles in each developmental stage. Expression levels across all developmental stages were considered, but given the heterogeneity of the dataset, a restrictive criterion was applied to minimize false positives when defining HVGs: a DEG was defined as an HVG if its expression level was at least 16 times higher in one stage compared to the average of the others, as described for *F. gigantica* [17] and *Schistosoma japonicum* [24]. In this sense, we included the IS21 samples on the assumption that the radical expression threshold used in HVG definition reflects primarily to biological factors rather than to batch effects.

HVGs were grouped using a *k*-means algorithm with the hclust function in the stats v3.6.1 library in R. The number of clusters (k = 6) was chosen based on groupings observed in multidimensional scaling analysis. For comparison, the same procedure was applied to detect HVG groups in *F. gigantica* using data published by [17].

Comparative genomic analyses

Predicted proteomes of trematodes available in WormBase ParaSite v15 [95] and the GCA_008360955.1 assembly of *Fasciolopsis buski* [60] available from NCBI were downloaded. Homolog groups were inferred using OrthoFinder v2 [122] running under default parameters. Multiple sequence alignments were performed with MAFFT v7.429 [123] running under the L-INS-I mode for greater accuracy. Given the large number of groups analyzed, phylogenetic inference was performed with FastTree v2.1.10 [124] running under default parameters.

Based on the literature, in some selected gene families a more detailed analysis was performed including homologous sequences from cestodes and free-living flatworms. The corresponding proteomes were downloaded from WormBase Parasite database (*Echinococcus granulosus*, PRJEB121; *E. multilocularis*, PRJEB122; *Taenia solium*, PRJNA170813; *Mesocestoides corti*, PRJEB510; *Hymenolepis. microstoma*, PRJEB124, *Schmidtea. mediterranea*, PRJNA379262; *Macrostomum lignano*, PRJNA371498). Homology inference was conducted through pairwise alignments using BLASTP v2.11.0+, defined as bidirectional hits with an e-value ≤ 1e-30. To reduce false positives, only alignments with an amino acid identity ≥ 35% and sequence coverage ≥ 70% were considered [125]. Multiple alignments were performed with MAFFT, evaluated using T-COFFEE v12.00 [126], running under the -score option, and visually inspected. Sequences with a conservation score < 50 were removed, and alignments were recalculated. Phylogenetic tree inference was performed using the maximum likelihood method with IQ-Tree v2.1.2 [127]. Evolutionary model selection was performed using ModelFinder [128] with the -m TEST option, 1000 SH-aLRT branch length evaluations, and 1000 ultrafast bootstrap replicates [129].

To compare groups with different expression trends, a set of genes with relatively stable expression throughout the life cycle was defined. A gene's variation was quantified using the coefficient of variation (CV = standard deviation/mean) of its expression values. A gene was defined as having stable expression ("SEG") if it presented a CV < 0.5.

Two types of comparative analyses were conducted between HVGs and SEGs. First, conservation (presence/absence) of each group of genes in both *F. hepatica* and *F. gigantica* was assessed. Homologous groups conserved in Metazoa (i.e., presence in ≥ 90% of clade species) were downloaded from the OrthoDB v10 database [130]. Homologous sequences were searched using the orthologr library in R [131], specifying the method of best reciprocal hit in DIAMOND alignments, with a minimum identity of 40% and sequence coverage ≥ 70% [132]. Second, sequence divergence patterns were analyzed for each category of genes by computing cophenetic distances between homologous genes, comparing trends observed in *F. hepatica* and *F. gigantica*. Distances obtained for different pairs of sequences were classified according to whether they involved HVGs or SEGs from

the two species. Cophenetic distance estimation was performed using the previously described gene trees and the get_distance method in the ete3 v3.1.2 library [133] in Python.

The concordance of homologous genes between the HVGs of *F. hepatica* and *F. gigantica* was analyzed using the Jaccard index. The Jaccard index measures the proportion of homolog groups with a detected member in both HVGs sets (numerator) relative to the total number of homolog groups detected in the pair (denominator). A similar approach was used to estimate the similarity between trematode stages based on orthologous gene expression [106].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12670-6.

Supplementary Material 1. Supplementary Table 1: Sample data, sequencing and mapping statistics. Samples from different life stages and/or time-points are indicates with the corresponding accession codes. Those corresponding to published data are referred in the last column by the DOI. Mapping statistics (total, trimmed and aligned reads by sample) are indicated.

Supplementary Material 2. Supplementary Table 2: *F. hepatica* gene annotation and differential expression dataAnnotation of all genes included in the analysis by different methods is indicated is indicated in several columns (green column titles), differential expression analysis results are resumed (yellow column titles) and mean cpm count by condition (red column titles) are indicated. A brief description of each column is indicated below the table. Total counts by column of associated items and blanks for each category are resumed at the top of the table (files 6 to 8).

Supplementary Material 3. Figure S1: Pipeline of the analysis performed. Sample preparation and sequencing is resumed in the first panel (blue), gene expression analysis is represented in the middle (red) panel and functional and phylogenetic analysis are resumed in the last (yellow) panel.

Supplementary Material 4. Figure S2: Transcript classification and expression across developmental stages. (A) Transcripts assembled with miracidium and intra-snail sample were classified based on the isoform classification system of gffcompare [99]. The proportions of each transcript category are shown in each case based on the transcripts with at least 10x coverage. (B) Gene expression of the newly generated data across all developmental stages. Genes detected in all stages analyzed (blue), expressed in at least 2 stages (green) restricted to a specific stage (red) are indicated.

Supplementary Material 5. Figure S3: Cophenetic distances for homologous gene pairs. Observed distributions for cophenetic distance (i.e., branch length) calculated for homologous gene pairs from *F. hepatica* and *F. gigantica*. Comparisons are classified according to whether they involve comparing HVG, SEG, or genes not classified in these categories.

Supplementary Material 6. Figure S4. Expression trends for developmental marker genes. Expression trends for marker genes associated with neoblast-like cell populations studied in larval stages of *S. mansoni* published by Wang et al. (2013) [37]. Homologous *F. hepatica* and *F. gigantica* genes (i.e., members of the same OrthoGroup) to those reported were selected. In each case, the normalized expression for each gene is shown, where the highest expression value is 1 and the lowest is 0. The expression data for *F. gigantica* correspond to those published by Zhang et al. (2019) [17].

Supplementary Material 7. Figure S5. Purine salvage pathways relevant to the development of trematodes in their parasitic stages. A. Expression trends of genes involved in purine salvage throughout the life cycle of *F. hepatica*. B. Proposed purine salvage pathway for Schistosoma mansoni. Taken from [134].

Supplementary Material 8. Figure S6. Fatty acid degradation pathway in liver fluke. Fatty acid degradation module from the KEGG database (Fatty acid degradation; M00087). Genes with expression detected in some larval stage associated with the snail host are highlighted in green (i.e., expression >= 1 CPM in all samples). The fatty acid beta-oxidation pathway is highlighted in pink. Schematic constructed using the Reconstruct option of the KEGG mapper tool.

Supplementary Material 9. Suppl. Data 1: Normalization criteria effects comparison.

Supplementary Material 10. Zenodo repository of supplementary files (https://doi.org/10.5281/zenodo.17204475). including: a. Sequence files. A Multifasta sequence file of the 11452 genes considered in this work are provided. MultiQC quality report of sequences analyzed. b. Supplementary Interactive Figures. Suppl. IP1. Interactive version of multidimensional analysis of samples as depicted in Fig. 2B Suppl. IP2. Interactive Volcano plot of miracidia and IS15 time-points DEGs as depicted in Fig, 6A. Suppl. IP3. Interactive Volcano plot of IS15 and IS30 time-points DEGs as depicted in Fig, 7A. c. Alignments and phylogenetic trees. Sequence alignments, phylogenetic trees, and expression data of sequences included in Figs. 8, 9, 10 and 11 are provided.

Supplementary Material 11. Github repository of scripts and notebooks. Scripts and Jupyter Notebooksfor all analyses performed in this study, as well as data employed and produced in it, are available atgithub (https://github.com/mauriciolangleib/fhepatica_snailseq).

Acknowledgements

We would like to thank Stephen Doyle for his generous exchange of ideas in the early stages of this work, particularly regarding the analysis of novel transcripts. We are also grateful to Obdulio for sustaining the work, and Jaime Roos, who provided support and inspiration to several members of the group during the development of this work.

Authors' contributions

ML: Conceptualization, data validation and analysis, scripting, visualization and figure generation, writing original draft. SF: Conceptualization, data validation, results discussion, manuscript drafting, revision. FD: Parasite and snail culture, snail dissection and RNA extraction and validation, results discussion, manuscript revision. GS: Library generation and sequencing. SW: Parasite and snail colony maintenance, snail dissection and sample collection. MB: Sequencing supervision, review and editing. GR: Conceptualization, sequencing organization, writing -reviewing & editing. AI: Conceptualization, methodology discussion, data analysis supervision, writing, reviewing and editing, funding acquisition, supervision. JFT: Conceptualization, methodology discussion, data analysis supervision, writing, reviewing and editing, funding acquisition, project supervision and administration.

Funding

This research was mainly funded by Comision Sectorial de Investigación Científica, Uruguay (CSIC-UdelaR) [Grant I + D-2016-516] awarded to JFT, and partially by Agencia Nacional de Investigación e Innovación, Uruguay (ANII) [Grant FCE_3_2016_1_125297] awarded to AI. The work was also supported in part by the Wellcome Trust [Grant numbers 206194 and 107475/Z/15/Z] and a UKRI Future Leaders Fellowship awarded to GR [Grant number MR/W013568/1].
M.L. was awarded a master's scholarship from the National Agency for Research and Innovation (code POS_NAC_2021_1_170401). M.L. is a student in the Programa de Desarrollo de las Ciencias Básicas (PEDECIBA), of which SF, JFT, and AI are members.
SF also was supported by ANII and PEDECIBA. SF, FD, GR, AI and JFT are members of the Sistema Nacional de Investigadores de Uruguay (SNI-ANII).

Data availability

The sequencing reads obtained for this work were deposited at SRA, accession PRJNA1039822.Sequence alignments, phylogenetic trees, RNA-seq mapping statistics, raw counts, functional annotation and interactive plots are available at Zenodo.org (DOI:10.5281/zenodo.17204476).Notebooks for all analyses performed in this study, as well as data employed and produced in it, are available at github (github.com/mauriciolangleib/fhepatica_snailseq).

Declarations

Competing interests
The authors declare no competing interests.

Author details
[1]Laboratorio de Biología Computacional, Departamento de Desarrollo Biotecnológico, Instituto de Higiene, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay
[2]Departamento de Genética, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay
[3]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK
[4]Museo Nacional de Historia Natural (MNHN), Ministerio de Educación y Cultura (MEC), Montevideo, Uruguay
[5]School of Infection & Immunity, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK
[6]Department of Biology, University of Oxford, Oxford, UK

Received: 17 December 2025 / Accepted: 16 February 2026
Published online: 09 March 2026

References
1. Mas-Coma S, Valero MA, Bargues MD. Human and Animal Fascioliasis: Origins and Worldwide Evolving Scenario. Clin Microbiol Rev. 2022;35. https://doi.org/10.1128/cmr.00088-19.
2. Zerna G, Spithill TW, Beddoe T. Current Status for Controlling the Overlooked Caprine Fasciolosis. Animals. Multidisciplinary Digit Publishing Inst. 2021;11:1819. https://doi.org/10.3390/ani11061819.
3. Rosas-Hostos Infantes LR, Paredes Yataco GA, Ortiz-Martínez Y, Mayer T, Terashima A, Franco-Paredes C, et al. The global prevalence of human fascioliasis: a systematic review and meta-analysis. Ther Adv Infect Dis. 2023;10:20499361231185413. https://doi.org/10.1177/20499361231185413.
4. Mehmood K, Zhang H, Sabir AJ, Abbas RZ, Ijaz M, Durrani AZ, et al. A review on epidemiology, global prevalence and economical losses of fasciolosis in ruminants. Microb Pathog. 2017;109:253–62. https://doi.org/10.1016/j.micpath.2017.06.006.
5. Harrington D, Lamberton PHL, McGregor A. Human liver flukes. Lancet Gastroenterol Hepatol. 2017;2:680–9. https://doi.org/10.1016/S2468-1253(17)30111-5.
6. Robinson MW, Hanna REB, Fairweather I. Development of Fasciola hepatica in the mammalian host. Fasciolosis. 2021;65–111. https://doi.org/10.1079/9781789246162.0003.
7. Dawes B. Penetration of Fasciola gigantica Cobbold, 1856 into Snail Hosts. Nat 1960 1854705. Nat Publishing Group. 1960;185:51–3. https://doi.org/10.1038/185051b0.
8. Køie M, Christensen NØ, Nansen P. Stereoscan studies of eggs, free-swimming and penetrating miracidia and early sporocysts of Fasciola hepatica. Z Parasitenkd Berl Ger Z Parasitenkd. 1976;51:79–90. https://doi.org/10.1007/BF00380530.
9. Whitfield PJ, Evans NA. Parthenogenesis and asexual multiplication among parasitic platyhelminths. Parasitol Parasitol. 1983;86(Pt 4):121–60. https://doi.org/10.1017/S0031182000050873.
10. Dreyfuss G, Sindou P, Hourdin P, Vignoles P, Rondelaud D. Fasciola hepatica larval development within the intermediate host. Fasciolosis. 2021;23–64. https://doi.org/10.1079/9781789246162.0002.
11. 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. https://doi.org/10.1186/s13071-018-2952-z.
12. Young ND, Hall RS, Jex AR, Cantacessi C, Gasser RB. Elucidating the transcriptome of Fasciola hepatica - a key to fundamental and biotechnological discoveries for a neglected parasite. Biotechnol Adv. 2010;28:222–31. https://doi.org/10.1016/j.biotechadv.2009.12.003.
13. Cancela M, Ruétalo N, Dell'Oca N, da Silva E, Smircich P, Rinaldi G, et al. Survey of transcripts expressed by the invasive juvenile stage of the liver fluke Fasciola hepatica. BMC Genomics. 2010;11:227. https://doi.org/10.1186/1471-2164-11-227.
14. Cwiklinski K, Jewhurst H, McVeigh P, Barbour T, Maule AG, Tort J, et al. Infection by the Helminth Parasite Fasciola hepatica Requires Rapid Regulation of Metabolic, Virulence, and Invasive Factors to Adjust to Its Mammalian Host. Mol Cell Proteom. 2018;17:792–809. https://doi.org/10.1074/MCP.RA117.000445.
15. Cwiklinski K, Robinson MW, Donnelly S, Dalton JP. Complementary transcriptomic and proteomic analyses reveal the cellular and molecular processes that drive growth and development of Fasciola hepatica in the host liver. BMC Genomics. 2021;22:1–16. https://doi.org/10.1186/S12864-020-07326-Y/FIGURES/6.
16. Gramberg S, Puckelwaldt O, Schmitt T, Lu Z, Haeberlein S. Spatial transcriptomics of a parasitic flatworm provides a molecular map of drug targets and drug resistance genes. Nat Commun. 2024;15:8918. https://doi.org/10.1038/s41467-024-53215-3.
17. Zhang XX, Cwiklinski K, Hu RS, Zheng WB, Sheng ZA, Zhang FK, et al. Complex and dynamic transcriptional changes allow the helminth Fasciola gigantica to adjust to its intermediate snail and definitive mammalian hosts. BMC Genomics. 2019. https://doi.org/10.1186/s12864-019-6103-5.
18. Augot D, Rondelaud D, Dreyfuss G, Cabaret J. Fasciola hepatica: in vitro production of daughter rediae and cercariae from first- and second-generation rediae. Parasitol Res. 1997;83:383–5. https://doi.org/10.1007/S004360050267.
19. Rinaldi G, Loukas A, Sotillo J. Trematode Genomics and Proteomics. Adv Exp Med Biol. 2024;1454:507–39. https://doi.org/10.1007/978-3-031-60121-7_13.
20. Abu-Jamous B, Kelly S. Clust: automatic extraction of optimal co-expressed gene clusters from gene expression data. Genome Biol. 2018;19. https://doi.org/10.1186/S13059-018-1536-8.
21. Takahashi H, Mitsui Y, Awazawa T, Fujimaki Y, Aoki Y. Control of ciliary activities of Schistosoma mansoni miracidia using triton-extracted parasites. J Parasitol. 1995;81:747–52. https://doi.org/10.2307/3283966.
22. Yang J, Liu X, Yue G, Adamian M, Bulgakov O, Li T. Rootletin, a novel coiled-coil protein, is a structural component of the ciliary rootlet. J Cell Biol. 2002;159:431–40. https://doi.org/10.1083/JCB.200207153.
23. Osinka A, Poprzeczko M, Zielinska MM, Fabczak H, Joachimiak E, Wloga D. Ciliary Proteins: Filling the Gaps. Recent Advances in Deciphering the Protein Composition of Motile Ciliary Complexes. Cells 2019 Vol 8 Page 730. Volume 8. Multidisciplinary Digital Publishing Institute; 2019. p. 730. https://doi.org/10.3390/CELLS8070730.
24. Cheng S, Zhu B, Luo F, Lin X, Sun C, You Y, et al. Comparative transcriptome profiles of Schistosoma japonicum larval stages: Implications for parasite biology and host invasion. PLoS Negl Trop Dis. 2022;16:e0009889. https://doi.org/10.1371/JOURNAL.PNTD.0009889.
25. Sponholtz GM, Short RB. Schistosoma mansoni miracidia: stimulation by calcium and magnesium. J Parasitol. 1976;62:155–7. https://doi.org/10.2307/3279082.
26. Matsuyama H, Takahashi H, Watanabe K, Fujimaki Y, Aoki Y. The involvement of cyclic adenosine monophosphate in the control of schistosome miracidium cilia. J Parasitol. 2004;90:8–14. https://doi.org/10.1645/GE-52R1.
27. Katsumata T, Kohno S, Yamaguchi K, Hara K, Aoki Y. Hatching of Schistosoma mansoni eggs is a Ca2+/calmodulin-dependent process. Parasitol Res. 1989;76:90–1. https://doi.org/10.1007/BF00931079.
28. Kawamoto F, Shozawa A, Kumada N, Kojima K. Possible roles of cAMP and Ca2 + in the regulation of miracidial transformation in Schistosoma mansoni. Parasitol Res. 1989;75:368–74. https://doi.org/10.1007/BF00931132.
29. Patocka N, Sharma N, Rashid M, Ribeiro P. Serotonin Signaling in Schistosoma mansoni: A Serotonin–Activated G Protein-Coupled Receptor Controls Parasite Movement. PLOS Pathog. 2014;10:e1003878. https://doi.org/10.1371/JOURNAL.PPAT.1003878.
30. Dalton JP, Caffrey CR, Sajid M, Stack C, Donnelly S, Loukas A, et al. Proteases in trematode biology. Parasit Flatworms Mol Biol Biochem Immunol Physiol CABI. 2006;348–68. https://doi.org/10.1079/9780851990279.0348.
31. Robinson MW, Dalton JP, Donnelly S. Helminth pathogen cathepsin proteases: it's a family affair. Trends Biochem Sci. 2008;33:601–8. https://doi.org/10.1016/j.tibs.2008.09.001.
32. Myers J, Ittiprasert W, Raghavan N, Miller A, Knight M, DIFFERENCES IN CYSTEINE PROTEASE, ACTIVITY IN SCHISTOSOMA MANSONI-RESISTANT AND -SUSCEPTIBLE BIOMPHALARIA GLABRATA AND CHARACTERIZATION OF THE HEPATOPANCREAS CATHEPSIN B FULL-LENGTH cDNA. J Parasitol. 2008;94:659. https://doi.org/10.1645/GE-1410R.1.
33. Holstein TW, Watanabe H, Ozbek S. Signaling pathways and axis formation in the lower metazoa. Curr Top Dev Biol. 2011;97:137–77. https://doi.org/10.1016/B978-0-12-385975-4.00012-7.
34. Holzem M, Boutros M, Holstein TW. The origin and evolution of Wnt signalling. Nat Rev Genet. 2024;25:500–12. https://doi.org/10.1038/s41576-024-00699-w.

35. Petersen CP, Reddien PW. Wnt Signaling and the Polarity of the Primary Body Axis. Cell. 2009;139:1056–68. https://doi.org/10.1016/J.CELL.2009.11.035.
36. Juliano CE, Swartz SZ, Wessel GM. A conserved germline multipotency program. Development. 2010;137:4113–26. https://doi.org/10.1242/DEV.047969.
37. Wang B, Collins JJ, Newmark PA. Functional genomic characterization of neoblast-like stem cells in larval Schistosoma mansoni. eLife. 2013;2013. https://doi.org/10.7554/ELIFE.00768.
38. Wurtzel O, Oderberg IM, Reddien PW. Planarian Epidermal Stem Cells Respond to Positional Cues to Promote Cell-Type Diversity. Dev Cell. 2017;40:491–e5045. https://doi.org/10.1016/J.DEVCEL.2017.02.008.
39. Armstrong R, Marks NJ, Geary TG, Harrington J, Selzer PM, Maule AG. Wnt/β-catenin signalling underpins juvenile Fasciola hepatica growth and development. PLOS Pathog. 2025;21:e1012562. https://doi.org/10.1371/journal.ppat.1012562.
40. Wang B, Lee J, Li P, Saberi A, Yang H, Liu C, et al. Stem cell heterogeneity drives the parasitic life cycle of Schistosoma Mansoni. eLife. 2018;7. https://doi.org/10.7554/ELIFE.35449.001.
41. Li P, Sarfati DN, Xue Y, Yu X, Tarashansky AJ, Quake SR, et al. Single-cell analysis of Schistosoma mansoni identifies a conserved genetic program controlling germline stem cell fate. Nat Commun 2021 121. 2021;12:1–12. https://doi.org/10.1038/s41467-020-20794-w.
42. Senft AW, Miech RP, Brown PR, Senft DG. Purine metabolism in Schistosoma mansoni. Int J Parasitol Int J Parasitol. 1972;2:249–60. https://doi.org/10.1016/0020-7519(72)90013-6.
43. McNulty SN, Tort JF, Rinaldi G, Fischer K, Rosa BA, Smircich P, et al. Genomes of Fasciola hepatica from the Americas Reveal Colonization with Neorickettsia Endobacteria Related to the Agents of Potomac Horse and Human Sennetsu Fevers. PLoS Genet. 2017. https://doi.org/10.1371/journal.pgen.1006537.
44. Tielens AGM, Heuvel JM, van den. Bergh SG van den. The energy metabolism of Fasciola hepatica during its development in the final host. Mol Biochem Parasitol. 1984;13:301–7. https://doi.org/10.1016/0166-6851(84)90121-X.
45. Tielens A, Metabolism. Dalton J, editors. Fasciolosis. CABI; 1999;277–305.
46. Khayath N, Mithieux G, Zitoun C, Coustau C, Vicogne J, Tielens AG, et al. Glyceroneogenesis: an unexpected metabolic pathway for glutamine in Schistosoma mansoni sporocysts. Mol Biochem Parasitol. 2006;147:145–53. https://doi.org/10.1016/J.MOLBIOPARA.2006.02.002.
47. Coustau C, Mitta G, Dissous C, Guillou F, Galinier R, Allienne JF, et al. Schistosoma mansoni and Echinostoma caproni excretory–secretory products differentially affect gene expression in Biomphalaria glabrata embryonic cells. Parasitology. 2003;127:533–42. https://doi.org/10.1017/S0031182003004049.
48. Sajid M, McKerrow JH, Hansell E, Mathieu MA, Lucas KD, Hsieh I, et al. Functional expression and characterization of Schistosoma mansoni cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. Mol Biochem Parasitol. 2003;131:65–75. https://doi.org/10.1016/s0166-6851(03)00194-4.
49. Galaktionov KV, Dobrovolskij AA. Organization of Parthenogenetic and Hermaphroditic Generations of Trematodes. In: Fried B, Graczyk TK, editors. The Biology and Evolution of Trematodes. Dordrecht: Springer; 2003. https://doi.org/10.1007/978-94-017-3247-5_1.
50. McVeigh P, Cwiklinski K, Garcia-Campos A, Mulcahy G, O'Neill SM, Maule AG, et al. In silico analyses of protein glycosylating genes in the helminth Fasciola hepatica (liver fluke) predict protein-linked glycan simplicity and reveal temporally-dynamic expression profiles. Sci Rep 2018 81. 2018;8:1–15. https://doi.org/10.1038/s41598-018-29673-3.
51. Cancela M, Santos GB, Carmona C, Ferreira HB, Tort JF, Zaha A. Fasciola hepatica mucin-encoding gene: expression, variability and its potential relevance in host-parasite relationship. Parasitology. 2015;142. https://doi.org/10.1017/S0031182015001134.
52. Ravida A, Cwiklinski K, Aldridge AM, Clarke P, Thompson R, Gerlach JQ, et al. Fasciola hepatica Surface Tegument: Glycoproteins at the Interface of Parasite and Host. Mol Cell Proteom. 2016;15:3139–53. https://doi.org/10.1074/MCP.M116.059774.
53. Caulfield JP, Yuan H, chang, Cianci CML, Hein A. Schistosoma mansoni: Development of the cercarial glycocalyx. Exp Parasitol. 1988;65:10–9. https://doi.org/10.1016/0014-4894(88)90102-6.
54. Berasaín P, Goñi F, McGonigle S, Dowd A, Dalton JP, Frangione B, et al. Proteinases secreted by Fasciola hepatica degrade extracellular matrix and basement membrane components. J Parasitol. 1997;83:1–5. https://doi.org/10.2307/3284308.
55. Corvo I, Cancela M, Cappetta M, Pi-Denis N, Tort JF, Roche L. The major cathepsin L secreted by the invasive juvenile Fasciola hepatica prefers proline in the S2 subsite and can cleave collagen. Mol Biochem Parasitol. 2009;167:41–7. https://doi.org/10.1016/J.MOLBIOPARA.2009.04.005.
56. Robinson MW, Corvo I, Jones PM, George AM, Padula MP, To J, et al. Collagenolytic Activities of the Major Secreted Cathepsin L Peptidases Involved in the Virulence of the Helminth Pathogen, Fasciola hepatica. PLoS Negl Trop Dis. 2011;5:e1012. https://doi.org/10.1371/JOURNAL.PNTD.0001012.
57. Tort J, Brindley PJ, Knox D, Wolfe KH, Dalton JP. Proteinases and associated genes of parasitic helminths. Adv Parasitol. 1999;43:161–266. https://doi.org/10.1016/S0065-308X(08)60243-2.
58. Cancela M, Acosta D, Rinaldi G, Silva E, Durán R, Roche L, et al. A distinctive repertoire of cathepsins is expressed by juvenile invasive Fasciola hepatica. Biochimie. 2008. https://doi.org/10.1016/j.biochi.2008.04.020.
59. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language, Bioinformatics. 2004;20(2):289–290. https://doi.org/10.1093/bioinformatics/btg412.
60. Choi YJ, Fontenla S, Fischer PU, Le TH, Costábile A, Blair D, et al. Adaptive Radiation of the Flukes of the Family Fasciolidae Inferred from Genome-Wide Comparisons of Key Species. Mol Biol Evol. 2020;37:84. https://doi.org/10.1093/MOLBEV/MSZ204.
61. Gevers D, Vandepoele K, Simillion C, Van de Peer Y. Gene duplication and biased functional retention of paralogs in bacterial genomes. Trends Microbiol. 2004;12:148–54. https://doi.org/10.1016/j.tim.2004.02.007.
62. Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. Selection in the evolution of gene duplications. Genome Biol. 2002. https://doi.org/10.1186/gb-2002-3-2-research0008.
63. Langleib M, Calvelo J, Costábile A, Castillo E, Tort JF, Hoffmann FG, et al. Evolutionary analysis of species-specific duplications in flatworm genomes. Mol Phylogenet Evol. 2024;199:108141. https://doi.org/10.1016/j.ympev.2024.108141.
64. Waterhouse RM, Zdobnov EM, Kriventseva EV. Correlating Traits of Gene Retention, Sequence Divergence, Duplicability and Essentiality in Vertebrates, Arthropods, and Fungi. Genome Biol Evol. 2011;3:75–86. https://doi.org/10.1093/GBE/EVQ083.
65. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 2010 112. 2010;11:97–108. https://doi.org/10.1038/nrg2689.
66. Nei M. Gene duplication and nucleotide substitution in evolution. Nature. 1969;221:40–2. https://doi.org/10.1038/221040A0.
67. Corvo I, O'Donoghue AJ, Pastro L, Pi-Denis N, Eroy-Reveles A, Roche L, et al. Dissecting the Active Site of the Collagenolytic Cathepsin L3 Protease of the Invasive Stage of Fasciola hepatica. PLoS Negl Trop Dis. 2013;7:e2269. https://doi.org/10.1371/journal.pntd.0002269.
68. Corvo I, Ferraro F, Merlino A, Zuberbühler K, O'Donoghue AJ, Pastro L et al. Substrate Specificity of Cysteine Proteases Beyond the S2 Pocket: Mutagenesis and Molecular Dynamics Investigation of Fasciola hepatica Cathepsins L. Front Mol Biosci. 2018;5. https://doi.org/10.3389/fmolb.2018.00040.
69. Wilson RA, Denison J. The parasitic castration and gigantism of Lymnaea truncatula infected with the larval stages of Fasciola hepatica. Z Für Parasitenkd Parasitol Res. 1980;61. https://doi.org/10.1007/BF00925458.
70. Lee FO, Cheng TC. Schistosoma mansoni: alterations in total protein and hemoglobin in the hemolymph of infected Biomphalaria glabrata. Exp Parasitol. 1972;31:203–16. https://doi.org/10.1016/0014-4894(72)90111-7.
71. Lee FO, Cheng TC. Incorporation of 59 Fe in the snail Biomphalaria glabrata parasitized by Schistosoma mansoni. J Parasitol. 1972;58:481–8. https://doi.org/10.2307/3278193.
72. Rinaldi G, Morales ME, Alrefaei YN, Cancela M, Castillo E, Dalton JP, et al. RNA interference targeting leucine aminopeptidase blocks hatching of Schistosoma mansoni eggs. Mol Biochem Parasitol. 2009;167:118–26. https://doi.org/10.1016/j.molbiopara.2009.05.002.
73. Chalmers IW, Hoffmann KF. Platyhelminth Venom Allergen-Like (VAL) proteins: revealing structural diversity, class-specific features and biological associations across the phylum. Parasitology. 2012;139:1231. https://doi.org/10.1017/S0031182012000704.
74. Chalmers IW, McArdle AJ, Coulson RMR, Wagner MA, Schmid R, Hirai H, et al. Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the Schistosoma mansoni venom allergen-like (SmVAL) gene family. BMC Genomics. 2008;9:89. https://doi.org/10.1186/1471-2164-9-89.
75. Hunt VL, Tsai IJ, Coghlan A, Reid AJ, Holroyd N, Foth BJ, et al. The genomic basis of parasitism in the Strongyloides clade of nematodes. Nat Genet. 2016;48:299–307. https://doi.org/10.1038/ng.3495.

76. Costábile A, Koziol U, Tort JF, Iriarte A, Castillo E. Expansion of cap superfamily proteins in the genome of Mesocestoides corti: An extreme case of a general bilaterian trend. Gene Rep. 2018;11:110–20. https://doi.org/10.1016/j.genrep.2018.03.010.
77. Cantacessi C, Hofmann A, Young ND, Broder U, Hall RS, Loukas A, et al. Insights into SCP/TAPS Proteins of Liver Flukes Based on Large-Scale Bioinformatic Analyses of Sequence Datasets. PLoS ONE. 2012;7:e31164. https://doi.org/10.1371/JOURNAL.PONE.0031164.
78. Farias LP, Chalmers IW, Perally S, Rofatto HK, Jackson CJ, Brown M et al. Schistosoma mansoni venom allergen-like proteins: phylogenetic relationships, stage-specific transcription and tissue localization as predictors of immunological cross-reactivity. Int J Parasitol. 2019;49:593–9. https://doi.org/10.1016/J.IJPARA.2019.03.003.
79. Fernandes RS, Barbosa TC, Barbosa MMF, Miyasato PA, Nakano E, Leite LCC, et al. Stage and tissue expression patterns of Schistosoma mansoni venom allergen-like proteins SmVAL 4, 13, 16 and 24. Parasit Vectors. 2017;10:1–13. https://doi.org/10.1186/S13071-017-2144-2/FIGURES/4.
80. Fernandes RS, Fernandes LGV, de Godoy AS, Miyasato PA, Nakano E, Farias LP, et al. Schistosoma mansoni venom allergen-like protein 18 (SmVAL18) is a plasminogen-binding protein secreted during the early stages of mammalian-host infection. Mol Biochem Parasitol. 2018;221:23–31. https://doi.org/10.1016/J.MOLBIOPARA.2018.02.003.
81. Perally S, Geyer KK, Farani PSG, Chalmers IW, Fernandez-Fuentes N, Maskell DR, et al. Schistosoma mansoni venom allergen-like protein 6 (SmVAL6) maintains tegumental barrier function. Int J Parasitol. 2021;51:251–61. https://doi.org/10.1016/J.IJPARA.2020.09.004.
82. Yoshino TP, Brown M, Wu XJ, Jackson CJ, Ocadiz-Ruiz R, Chalmers IW, et al. Excreted/secreted Schistosoma mansoni venom allergen-like 9 (SmVAL9) modulates host extracellular matrix remodelling gene expression. Int J Parasitol. 2014;44:551–63. https://doi.org/10.1016/J.IJPARA.2014.04.002.
83. Attenborough T, Rawlinson KA, Soria CLD, Ambridge K, Sankaranarayanan G, Graham J et al. A single-cell atlas of the miracidium larva of the human blood fluke Schistosoma mansoni: cell types, developmental pathways and tissue architecture. eLife. 2024;13. https://doi.org/10.7554/eLife.95628.1. [cited 2025 July 13].
84. Diaz Soria CL, Attenborough T, Lu Z, Fontenla S, Graham J, Hall C, et al. Single-cell transcriptomics of the human parasite Schistosoma mansoni first intra-molluscan stage reveals tentative tegumental and stem-cell regulators. Sci Rep. 2024;14:5974. https://doi.org/10.1038/s41598-024-55790-3.
85. Lu Z, Sankaranarayanan G, Rawlinson KA, Offord V, Brindley PJ, Berriman M, et al. The Transcriptome of Schistosoma mansoni Developing Eggs Reveals Key Mediators in Pathogenesis and Life Cycle Propagation. Front Trop Dis. 2021;2:713123. https://doi.org/10.3389/fitd.2021.713123.
86. Wendt GR, Collins JJ. Schistosomiasis as a disease of stem cells. Curr Opin Genet Dev. 2016;40:95. https://doi.org/10.1016/J.GDE.2016.06.010.
87. Wu XJ, Sabat G, Brown JF, Zhang M, Taft A, Peterson N, et al. Proteomic analysis of Schistosoma mansoni proteins released during in vitro miracidium-to-sporocyst transformation. Mol Biochem Parasitol. 2009;164:32. https://doi.org/10.1016/J.MOLBIOPARA.2008.11.005.
88. Wang B, Collins JJ III, Newmark PA. A <>Sánchez Alvarado editor 2013 Functional genomic characterization of neoblast-like stem cells in larval Schistosoma mansoni. eLife 2 e00768 https://doi.org/10.7554/eLife.00768.
89. Sarfati DN, Li P, Tarashansky AJ, Wang B. Single-cell deconstruction of stem-cell-driven schistosome development. Trends Parasitol. 2021;37:790–802. https://doi.org/10.1016/J.PT.2021.03.005.
90. Ittiprasert W, Brindley PJ. CRISPR-based functional genomics for schistosomes and related flatworms. Trends Parasitol. 2024;40:1016–28. https://doi.org/10.1016/j.pt.2024.09.010.
91. Gayo V, Cancela M, Acosta D. Maintenance of Life Cycle Stages of Fasciola hepatica in the Laboratory. Methods Mol Biol Clifton NJ. 2020;2137:1–14. https://doi.org/10.1007/978-1-0716-0475-5_1.
92. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20. https://doi.org/10.1093/bioinformatics/btu170.
93. Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Krueger C, Wingett S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
94. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–8. https://doi.org/10.1093/bioinformatics/btw354.
95. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite – a comprehensive resource for helminth genomics. Mol Biochem Parasitol. 2017;215:2–10. https://doi.org/10.1016/j.molbiopara.2016.11.005.
96. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15. https://doi.org/10.1038/s41587-019-0201-4.
97. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9. https://doi.org/10.1093/BIOINFORMATICS/BTP352.
98. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5. https://doi.org/10.1038/nbt.3122.
99. Pertea G, Pertea MGFF, Utilities. GffRead and GffCompare. F1000Research. 2020;9:304. https://doi.org/10.12688/f1000research.23297.1.
100. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34:525–7. https://doi.org/10.1038/nbt.3519.
101. Cwiklinski K, Dalton JP, Dufresne PJ, Course JL, Williams DJL, Hodgkinson J, et al. The Fasciola hepatica genome: gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution. Genome Biol. 2015;16. https://doi.org/10.1186/s13059-015-0632-2.
102. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research. 2015;4:1521. https://doi.org/10.12688/F1000RESEARCH.7563.1.
103. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 2010 113. BioMed Cent. 2010;11:1–9. https://doi.org/10.1186/GB-2010-11-3-R25.
104. ChenY, McCarthy D, Ritchie M, Robinson M, Smyth G, Hall E. edgeR: differential analysis of sequence read count data User's Guide. https://bioconductor.org/packages/devel/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf.
105. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
106. Nesterenko MA, Starunov VV, Shchenkov SV, Maslova AR, Denisova SA, Granovich AI, et al. Molecular signatures of the rediae, cercariae and adult stages in the complex life cycles of parasitic flatworms (Digenea: Psilostomatidae). Parasit Vectors. 2020;13:1–21. https://doi.org/10.1186/S13071-020-04424-4/FIGURES/6.
107. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12:2825–30.
108. Jaccard P, THE DISTRIBUTION OF, THE FLORA IN THE ALPINE ZONE.1. New Phytol. 1912;11:37–50. https://doi.org/10.1111/J.1469-8137.1912.TB05611.X.
109. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021;38:5825–9. https://doi.org/10.1093/molbev/msab293.
110. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47:D309–14. https://doi.org/10.1093/nar/gky1085.
111. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9. https://doi.org/10.1038/75556.
112. Kanehisa M. The KEGG databases at GenomeNet. Nucleic Acids Res. 2002. https://doi.org/10.1093/nar/30.1.42.
113. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49:D412–9. https://doi.org/10.1093/nar/gkaa913.
114. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res. 2021;49:D274–81. https://doi.org/10.1093/nar/gkaa1018.
115. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: Genome-scale protein function classification. Bioinformatics. 2014. https://doi.org/10.1093/bioinformatics/btu031.
116. Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Vesztrocy AW, Naldi A, et al. GOATOOLS: A Python library for Gene Ontology analyses. Sci Rep. 2018. https://doi.org/10.1038/s41598-018-28948-z.
117. Wang JZ, Du Z, Payattakool R, Yu PS, Chen CF. A new method to measure the semantic similarity of GO terms. Bioinf Oxf Acad. 2007;23:1274–81. https://doi.org/10.1093/BIOINFORMATICS/BTM087.

118. Waskom ML. seaborn: statistical data visualization. J Open Source Softw Open J. 2021;6:3021. https://doi.org/10.21105/JOSS.03021.
119. Peterson H, Kolberg L, Raudvere U, Kuzmin I, Vilo J. gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000 Res. 2020;9:709. https://doi.org/10.12688/f1000research.24956.2.
120. Powell D. Degust: interactive RNA-seq analysis. Zenodo; 2019. https://doi.org/10.5281/zenodo.3501067.
121. Blighe K, Rana S, Lewis M, EnhancedVolcano. Publication-ready volcano plots with enhanced colouring and labeling DOI: 10.18129/B9.bioc.EnhancedVolcano; 2021. http://bioconductor.org/packages/EnhancedVolcano/. Accessed 22 June 2025.
122. Emms DM, Kelly S, OrthoFinder. Phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:1–14. https://doi.org/10.1186/S13059-019-1832-Y/FIGURES/5.
123. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–66. https://doi.org/10.1093/nar/gkf436.
124. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE. 2010;5. https://doi.org/10.1371/JOURNAL.PONE.0009490.
125. Rost B. Twilight zone of protein sequence alignments. Protein Eng Des Sel. 1999;12:85–94. https://doi.org/10.1093/PROTEIN/12.2.85.
126. Notredame C, Higgins DG, Heringa J. T-coffee: a novel method for fast and accurate multiple sequence alignment1. J Mol Biol. 2000;302:205–17. https://doi.org/10.1006/jmbi.2000.4042.
127. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37:1530–4. https://doi.org/10.1093/MOLBEV/MSAA015.
128. Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler AV, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods Nat Methods. 2017;14:587–9. https://doi.org/10.1038/NMETH.4285.
129. Hoang DT, Chernomor O, Haeseler AV, Minh BQ, Vinh LS. UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018. https://doi.org/10.1093/molbev/msx281.
130. Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, et al. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. 2019;47:D807–11. https://doi.org/10.1093/NAR/GKY1053.
131. Drost HG, Gabel A, Grosse I, Quint M. Evidence for Active Maintenance of Phylotranscriptomic Hourglass Patterns in Animal and Plant Embryogenesis. Mol Biol Evol. 2015;32:1221–31. https://doi.org/10.1093/MOLBEV/MSV012.
132. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods 2021 184. 2021;18:366–8. https://doi.org/10.1038/s41592-021-01101-x.
133. Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol Biol Evol. 2016. https://doi.org/10.1093/molbev/msw046.
134. Berens RL, Krug EC, Marr JJ. Purine and Pyrimidine Metabolism. Biochem Mol Biol Parasites. 1995;89–117. https://doi.org/10.1016/B978-012473345-9/50007-6.

Publisher's Note