MOLECULAR ECOLOGY

Major host transitions are modulated through transcriptome-wide reprograming events in *Schistocephalus solidus*, a threespine stickleback parasite

1   **Major host transitions are modulated through transcriptome-wide reprograming events in**

2   *Schistocephalus solidus***, a threespine stickleback parasite**

3

4   François Olivier HÉBERT[1], Stephan GRAMBAUER[2], Iain BARBER[2], Christian R LANDRY[1],

5   Nadia AUBIN-HORTH[1]*

6

7   1. Institut de Biologie Intégrative et des Systèmes (IBIS), Département de Biologie, Université

8   Laval, 1030 avenue de la Médecine, G1V 0A6, Québec (QC), Canada.

9   Phone: 1-418-656-3316

10   Fax: 1-418-656-7176

11

12

13   2. Department of Neuroscience, Psychology and Behaviour, Adrian Building, University of

14   Leicester, University Road, Leicester, LE1 7RH, UK.

15

16   * Corresponding author:

17   Nadia Aubin-Horth

18   Nadia.Aubin-Horth@bio.ulaval.ca

19   Phone: 1-418-656-3316

20   Fax: 1-418-656-7176

21

22

23   **Running title**

24   Parasite transcriptome reprogramming

25

26   **Keywords**

27   Parasite, cestode, *Schistocephalus solidus*, threespine stickleback, bird, Transcriptomics

28

29

30   **ABSTRACT**

31

32   Parasites with complex life cycles have developed numerous phenotypic strategies, closely

33   associated with developmental events, to enable the exploitation of different ecological niches

34   and facilitate transmission between hosts. How these environmental shifts are regulated from a

35   metabolic and physiological standpoint, however, still remain to be fully elucidated. We

36   examined the transcriptomic response of *Schistocephalus solidus*, a trophically-transmitted

37   parasite with a complex life cycle, over the course of its development in an intermediate host, the

38   threespine stickleback, and the final avian host. Results from our differential gene expression

39   analysis show major reprogramming events among developmental stages. The final host stage is

40   characterized by a strong activation of reproductive pathways and redox homeostasis. The

41   attainment of infectivity in the fish intermediate host – which precedes sexual maturation in the

42   final host and is associated with host behaviour changes – is marked by transcription of genes

43   involved in neural pathways and sensory perception. Our results suggest that un-annotated and *S.*

44   *solidus*-specific genes could play a determinant role in host-parasite molecular interactions

45   required to complete the parasite's life cycle. Our results permit future comparative analyses to

46   help disentangle species-specific patterns of infection from conserved mechanisms, ultimately

47   leading to a better understanding of the molecular control and evolution of complex life cycles.

48

INTRODUCTION

Parasites with multiple hosts commonly undergo dramatic phenotypic transformations and endure major environmental shifts over the course of their life cycle (Wilbur 1980; Poulin 2011), yet very little is known about how these are orchestrated at the molecular and physiological levels, or how conserved they are across species (Auld & Tinsley 2014). Among the key insights yet to be gained is a detailed understanding of the metabolic and developmental regulation of parasites associated with infection, survival and development in each host. Characterising patterns of gene expression can inform the study of how physiological functions are modulated, a task otherwise difficult to achieve for organisms such as parasites that need to be cultured and studied inside other animals. Gathering this information for multiple host-parasite systems will allow general comparisons to be drawn between species. These comparisons will ultimately help disentangle species-specific patterns from common mechanisms that have promoted the evolution of complex life cycles.

We dissected the genome-wide transcriptional activity of the cestode *Schistocephalus solidus*, a model parasite with a complex life cycle (Barber 2013), to uncover how biological functions are regulated in different developmental stages, and how they relate to the completion of the parasite's life cycle (Figure 1a). *S. solidus* successively parasitizes a cyclopoid copepod, a fish – the threespine stickleback, *Gasterosteus aculeatus* – and a piscivorous endotherm, typically a bird (Clarke 1954). We aimed to determine the functional changes happening when infecting the final host – where reproduction occurs – and identify differences in gene expression between pre-infective and infective forms of the plerocercoid stage within the second intermediate host. The first developmental stages – free swimming coracidia – occur in freshwater and hatch from eggs released with the faeces of the avian host. Each coracidium ingested by cyclopoid copepods then develops further into the procercoid stage. When a threespine stickleback feeds on infected copepods, the parasite is released from the copepod and penetrates the intestinal mucosal wall of the fish after 14-24 hours, before developing into the plerocercoid stage (Hammerschmidt & Kurtz 2007). However, the newly developed plerocercoid is not initially infective to the final bird host. The status of infectivity is defined as the development stage at which the parasite can successfully mature and reproduce (Tierney & Crompton 1992). During the early plerocercoid phase, the host immune system is not activated by the presence of the worm inside its body cavity

80   (Scharsack *et al.* 2007). *S. solidus* spends the next 50-60 days in an exponential growth phase,
81   gaining up to 20 times its initial mass (Barber *et al.* 2004). When the plerocercoid eventually
82   reaches infectivity, a phase that could be determined by sufficient glycogen reserves (Hopkins
83   1950), drastic phenotypic changes occur in the fish host (Barber & Scharsack 2010). These
84   changes include an activation of the immune system (Scharsack *et al.* 2007) and a loss of anti-
85   predator response (Barber *et al.* 2004). Following the ingestion of infected fish by an avian
86   predator, the parasite experiences a temperature of 40°C in the bird's digestive tract – compared
87   to a maximum of 15-18°C in the ectothermic intermediate hosts – as well as chemical attack by
88   digestive enzymes. These conditions trigger the parasite's development to the sexually mature
89   adult in ca. 36 hours (Smyth 1950). The adult parasite reproduces during the next 3-4 days with
90   eggs being released into the water with the avian host's faeces (Hopkins & Smyth 1951).
91   Adjusting to these host switches and life history transitions requires many physiological changes
92   that are expected to recruit the activity of different genes at each phase.

93

94   One of the major transitions expected to affect worm physiology is the transition from somatic
95   growth to reproduction. Histological and physiological studies suggest that gametogenesis only
96   occurs when the parasite reaches the final (bird) host (Hopkins 1950; Schjørring 2003). Despite
97   the advanced ('progenetic'; Smyth & MacManus 2007) development of reproductive organs in
98   infective plerocercoids, only an elevated temperature of 40°C in semi-anaerobic conditions can
99   trigger meiosis and reproductive behaviours (Smyth 1952; Schjørring 2003). Previous work on
100 the anaerobic activity of key enzymes involved in the catabolism of carbohydrates in *S. solidus*
101 also suggests that while carbohydrate breakdown is very slow in pre-infective and infective
102 plerocercoids, this rate increases several-fold upon maturation (Körting & Barrett 1977; Beis &
103 Barrett 1979). Energetic resources used during the adult stage mainly come from glycogen
104 reserves accumulated during growth of the pre-infective plerocercoid (Hopkins 1952; Körting &
105 Barrett 1977). One thus expects that plerocercoids cultured at 40°C will show an up-regulation of
106 glycogen-related pathways.

107

108 The adult stage interacts with its environment to time these developmental steps. The anatomical
109 structure more likely to achieve this task is the tegument, a very active and complex tissue that
110 behaves like a true epidermis (Lee 1967). The adult stage of *S. solidus* exhibits numerous

111   vacuolate vesicles packed with electron-dense or electron-lucent content. These small structures

112   are evenly distributed in *S. solidus* syncytial tegument (Charles & Orr 1968). Their role could be

113   related to both nutrition and defence, as they allow rapid internalization of environmental

114   nutrients and antigen-antibody complexes (Hopkins *et al.* 1978; Threadgold & Hopkins 1981).

115   However, the uptake of macromolecules by adult cestodes remains an open question (Conradt &

116   Peters 1989). If, however, *S. solidus* performs pinocytosis or endocytosis at any developmental

117   stage, specific transcripts involved in this biological activity are expected to be up-regulated at

118   these stages. The existence of membrane-bound vesicles also suggests potential

119   secretory/excretory functions that would allow the parasite to release various types of molecules

120   in its host.

121

122   Pre-infective and infective plerocercoids are discrete developmental stages distinguished by their

123   divergent growth and effects on the host immune system. Parasites grow rapidly in the first

124   weeks of the stickleback host infection but growth rates tend to slow down as the parasite

125   becomes infective to the final host (Barber & Svensson 2003). Concurrently, empirical evidence

126   shows that secretory/excretory products from pre-infective versus infective plerocercoids have

127   different modulatory effects on the immune system of the fish host (Scharsack *et al.* 2013).

128   Small, pre-infective plerocercoids down-regulate the proliferation of host monocytes, but as soon

129   as they attain infectivity they activate a strong respiratory burst activity (Scharsack *et al.* 2007).

130   From a transcriptional perspective, different functional programs between pre-infective and

131   infective stages that reflect these divergent activities should be detectable. Distinct and specific

132   gene expression profiles should characterize each developmental stage according to the biological

133   activities that they need to perform to ultimately maximize the parasite's success in each host.

134

135   **MATERIAL AND METHODS**

136   *Sampling*

137   Worm specimens spanning three development states were extracted from laboratory-raised and

138   experimentally infected threespine sticklebacks. To infect these fish, we obtained parasite eggs

139   through *in vitro* culture of mature plerocercoids extracted from wild-caught threespine

140   sticklebacks (Clatworthy Reservoir, England, UK). After a three-week incubation period in tap

141   water, eggs hatched in response to daylight exposure, and emergent coracidia were used to infect

142 copepods. Exposed copepods harboring infective procercoids after three to four weeks were fed

143 to healthy lab-bred threespine sticklebacks (Hébert *et al.* 2016a). One hundred fish were exposed

144 and maintained for 16 weeks (controlled temperatures, fed frozen chironomid larvae *ad libitum)*

145 and subsequently euthanized (overdose of 15 mM Benzocaine solution). Extractions of the

146 parasites from the fish were scheduled to obtain pre-infective (i.e <50 mg) and infective

147 (i.e. >50mg) plerocercoids. We obtained three adult specimens of *S. solidus* through *in vitro*

148 culture of infective plerocercoids extracted from wild-caught sticklebacks of the same population

149 as the experimental infections (Hébert *et al.* 2016b). Obtaining adult worms through *in vitro*

150 culture in a bird-gut model represents a standard method used for more than 50 years in

151 experimental parasitology applied to helminths. It offers quick and replicable sampling, as

152 compared to alternative methods such as *in vivo* infections of ducklings. Briefly, infective

153 plerocercoids were placed individually into a dialysis membrane suspended in a medium

154 composed of 50:50 RPMI:horse serum, at a temperature, pH and oxygen tension mimicking the

155 conditions experienced in the bird digestive track (Smyth 1950) – for detailed protocol see

156 (Hébert *et al.* 2016a). Adult worms had a body mass of 321-356 mg. Worms were washed with

157 ultra-pure RNase-free water, diced into small pieces of ~5 mm x 5 mm, placed into RNALater

158 (Ambion Inc., Austin, TX, USA) and kept at -80°C.

159

160 *RNA sequencing*

161 We used RNA samples from fourteen different worms to produce individual TruSeq Illumina

162 sequencing libraries (San Diego, CA, USA) according to the manufacturer's protocol. We

163 produced libraries for seven pre-infective (<50mg) plerocercoids, four infective plerocercoids

164 (>50mg) and three adult worms (Hébert *et al.* 2016). cDNA libraries were sequenced on a

165 Illumina HiSeq 2000 system (Centre de Recherche du CHU de Québec, Québec, QC, Canada)

166 with the paired-end technology (2X100 bp). In total, 75.8 Gb of raw data was generated, which

167 represents 375 million 2 x 100 bp paired-end sequences distributed across the fourteen samples –

168 deposited into the NCBI Sequence Read Archive (accession number SAMN04296611,

169 BioProject PRJNA304161, see Hébert *et al.* 2016c).

170

171 *Short-read alignment on reference transcriptome*

172   Raw sequencing reads were cleaned, trimmed and aligned on the reference transcriptome,

173   allowing the estimation of transcript-specific expression levels for each individual worm (Hébert

174   *et al.* 2016b; Hébert 2016b). In summary, we aligned short reads from the 14 individual HiSeq

175   libraries on the reference with Bowtie 2 v.2.1.0 (Langmead & Salzberg 2012), allowing multi-

176   mapping of each read. Transcripts showing similar sequence, length and expression levels were

177   then regrouped into clusters of unigenes by using Corset v1.00 (Davidson & Oshlack 2014). We

178   obtained read counts for each unigene in each individual worm using the mapping information

179   contained in the SAM files (Hébert 2016b). We adjusted the algorithm parameters so that read

180   counts with sequences of varying lengths (isoforms, pseudogenes, alternative transcripts,

181   paralogs) would not be merged (contig ratio test parameter switched on).

182

183   *Differential expression analysis*

184   We conducted downstream analyses using the R packages 'limma-voom' (Law *et al.* 2014) and

185   edgeR (Robinson *et al.* 2010). We imported the read count matrix into R v.3.3.2 (R Development

186   Core Team, 2008) for initial filtering of lowly expressed transcripts. We kept sequences with

187   more than 15 Counts Per Million (CPM) in at least three samples in any of the life stage. The

188   filtering threshold values were chosen based on a comparative analysis of multiple datasets

189   produced with different combinations of thresholds (Figure S1). This specific threshold allowed

190   filtering low-coverage transcripts and potentially several false positives, without losing too much

191   information on differentially expressed transcripts. Normalization of read counts was performed

192   using the method of Trimmed Mean of M-values (TMM), using edgeR default parameters,

193   followed by the voom transformation. Read counts were converted into CPM value and log2-

194   transformed. Next, each transcript was fitted to an independent linear model using the $\log_2(CPM)$

195   values as the response variable and the group – pre-infective plerocercoid, infective plerocercoid

196   and adult – as the explanatory variable. No intercept was used and all possible comparisons

197   between the three developmental stages were defined as our desired contrasts. Each linear model

198   was analyzed through limma's Bayes pipeline. This last step allowed the discovery of

199   differentially expressed transcripts based on a False Discovery Rate (FDR) < 0.001 (Hébert

200   2016a; Law *et al.* 2014).

201

202   We performed hierarchical clustering among transcripts and samples using the limma-voom

203   transformed $\log_2$(CPM) values through the 'heatmap.2' function in the 'gplots' package v.2.17.0

204   (Warnes *et al.* 2016). For each life-stage, samples were clustered based on Euclidean distance

205   among transcript abundances and plotted on a heatmap by re-ordering the values by transcripts

206   (rows) and by samples (columns). We evaluated the robustness of each cluster of transcripts

207   identified through this method using the R package 'fpc' v.2.1.10 (Flexible Procedures for

208   Clustering, Hennig 2015), which implements a bootstrapping algorithm on values of the Jaccard

209   index to return a cluster stability index (Hébert 2016a). We only considered clusters with a

210   stability index greater than 0.50 with 1,000 bootstraps for downstream Gene Ontology (GO)

211   enrichment analyses. In total, 12 out of 16 clusters distributed across the two heatmaps were kept

212   for GO enrichment analysis. We considered each cluster satisfying the stability index threshold as

213   a module of co-expressed genes potentially bearing a broad functional status in accordance with

214   the results from the GO enrichment analysis.

215

216   We identified functional categories over-represented in each co-expression module to

217   characterize the biological functions associated with each life-stage. We used the Python package

218   'goatools' (Klopfenstein *et al.* 2015) to perform Fisher's exact tests on GO annotation terms

219   found in clusters of significantly differentially expressed gene. Annotation of GO terms for each

220   gene was based on the published transcriptome of *Schistocephalus solidus* (Hébert *et al.* 2016b).

221   GO terms over-represented in a given module, as compared to the reference transcriptome (FDR

222   ≤ 0.05), were labelled as putative 'transition-specific' biological functions.

223

224   *Ecological annotation*

225   We assigned an ecological annotation to transcripts exhibiting significant abundance changes

226   between life stages or showing stage-specific expression patterns (Pavey *et al.* 2012). Two

227   different types of ecological annotation were added to the dataset. First, we labelled un-annotated

228   transcripts according to their significant variation in abundance across life stages. Information on

229   GO terms over-represented in the cluster in which these transcripts could be found was also

230   added. Second, we labelled transcripts showing "on-off patterns of expression" among stages and

231   hosts as "stage-specific" or "host-specific". A pre-defined specificity threshold was chosen as the

232   $\log_2$(CPM) value representing the 5th percentile of the distribution of the $\log_2$(CPM) across all

233   transcripts. We identified stage-specific transcripts based on an average $\log_2$(CPM) above our

234   pre-defined specificity threshold ('ON') across at least two-thirds of the worms in only one of the

235   three life-stages. Similarly, we considered transcripts as host-specific if they showed an average

236      $\log_2$(CPM) above the specificity threshold across two-thirds of the worms in any of the two hosts.

237

238   **RESULTS AND DISCUSSION**

239   <u>Host transition as the main driver of genome reprogramming</u>

240   A total of 2894 genes (28% of transcriptome) are significantly differentially regulated (FDR <

241   0.001) over the course of the infection of the fish and bird hosts (Tables S1). A multidimensional

242   scaling analysis (MDS) performed on the top 1000 most differentiated genes in the dataset further

243   suggests that the main factor that drives the divergence among individual worms is host type (fish

244   vs. bird-gut model; Figure 1b). The first dimension of the MDS plot shows two distinct clusters:

245   one with adult worms and another regrouping pre-infective and infective worms. This analysis

246   also shows the grouping of pre-infective and infective worms into two different clusters on the

247   second dimension. The distance on the first dimension between host types is at least twice as

248   large as the distance separating pre-infective and infective worms, suggesting that host type is the

249   main driving factor. This may largely be explained by physiological acclimatisation of the

250   parasite to highly divergent thermal environments offered by the two hosts, or to other

251   differences such as oxygen tension, pH or osmotic pressure (Smyth 1950; Aly *et al.* 2009;

252   Oshima *et al.* 2011). The switch between these two hosts also correlates with rapid sexual

253   maturation, reproduction and changes in energy metabolism (Clarke 1954). Altogether, these

254   factors contribute to a major reprogramming of the worm transcription profile between hosts.

255

256   *Biological activities focused towards reproductive functions*

257   The development of the adult stage in the avian host requires the parasite to shift most of its

258   biological activities from growth and immune evasion (Hopkins & Smyth 1951; Hammerschmidt

259   & Kurtz 2005) to reproduction and possibly starvation (Hopkins 1950; Smyth 1954). In

260   accordance with these life-history changes, sexual maturation pathways and reproductive

261   behaviours were dominant functions in the transcriptional signature of the final host-switch, as

262   supported by GO terms significantly enriched in co-expression modules (Figure 2a, Table S1).

263   The largest co-expression module identified in the transition from infective plerocercoid to adult

264   contains a total of 769 genes significantly over-expressed in adult worms (Figure 2a, cluster 4).

265   This module is enriched (FDR < 0.05) in biological processes related to reproductive functions

266   such as spermatid nucleus differentiation (GO:0007289), sperm motility (GO:0030317),

267   luteinizing hormone secretion (GO:0032275) and positive regulation of testosterone secretion

268   (GO:2000845) (Table S1). Early studies on the life cycle of *S. solidus* suggested that once the

269   worm reaches the final bird host, its energy is canalized into maturation and reproduction,

270   including egg-laying (Hopkins & Smyth 1951; Clarke 1954). Adult worms sampled in this study

271   were collected after five days of *in vitro* culture in a bird-gut model at 40°C, 3-4 days after the

272   onset of gamete production (Smyth 1946; Smyth 1954). The transcriptional signature confirms

273   this at the molecular level, as we have detected the induction of many genes involved in sperm

274   motility and cilium movement (Table S1).

275

276   *Re-organisation of the energy budget*

277   The transition from infective plerocercoid to adult is characterized by a significant shift in energy

278   metabolism (Barrett 1977). Empirical data suggests that during the first hours of maturation and

279   reproduction, worms utilize glycogen reserves accumulated in the fish host (Hopkins 1950).

280   Adult worms cultured *in vitro* are also capable of absorbing glucose after more than 40 hours at

281   40°C (Hopkins 1952), suggesting they can stop using their glycogen reserves and instead use

282   host-derived nutrients. Our results suggest a complex and subtle pattern of regulation in terms of

283   carbohydrate metabolism. In total, eleven steps of the glycolysis pathway were differentially

284   regulated between the infective plerocercoid and adult stages (Figure 3). The first step in

285   glycogen breakdown consists in converting glycogen to glucose-1-phosphate, a reaction catalysed

286   by the enzyme glycogen phosphorylase (Smyth & McManus 2007). This enzyme is strongly up-

287   regulated in adult worms (logFC = 5.9, FDR < 0.0001), suggesting an active use of glycogen

288   reserves at this stage. The first three major biochemical transformations leading to glucose

289   breakdown into more simple sugars are strongly down-regulated (Figure 3). Intriguingly, genes

290   coding for the enzyme that produce glyceraldehyde-3-phosphate (GADP) are consistently up-

291   regulated in adult worms. All three homologous genes identified as fructose-bisphosphate

292   aldolase in our dataset, the enzyme responsible for the production of GADP, were labeled as

293   being switched 'ON' in adults (see Materials and Methods for details). Most of the downstream

294   genes leading to the production of pyruvate are down-regulated, with the exception of enolase,

295   the enzyme responsible for the penultimate step of glycolysis, i.e. the conversion of glycerate-2-
296   phosphate into phosphoenol-pyruvate (Figure 3, Table S1). Consistent with the semi-anaerobic
297   conditions experienced by adult worms, we found a significant up-regulation of the gene coding
298   for L-lactate dehydrogenase (logFC = 5.8, FDR < 0.001), the enzyme responsible for the
299   conversion of pyruvate to lactate when oxygen supplies are low (Smyth & McManus 2007).

300

301   Interestingly, we found a testis-specific gene, coding for glyceraldehyde-3-phosphate
302   dehydrogenase, among the few up-regulated genes of the glycolysis pathway. The gene was
303   significantly over-expressed in adult worms, with a fold-change of 97 (logFC = 6.6, FDR <
304   0.0001). A homologous gene – with the same annotation, but not testis-specific – is conversely
305   down-regulated in adults (logFC = -1.6, FDR = 0.0003). These results suggest that late stages of
306   adult *S. solidus* may still be very active in terms of sperm production, even after several days in
307   the avian host. At this late stage, the adult parasite may direct all of its energetic activities
308   towards sperm production, in order to maximize rates of egg fertilization.

309

310   *Potential role for endocytosis in balancing energetic reserves*
311   How glucose is produced or acquired by adult *S. solidus* is unclear, but this activity could be
312   performed by molecular mechanisms such as endocytosis or pinocytosis (Hopkins *et al.* 1978).
313   This hypothesis led to the prediction that expression of genes specific to this pathway should be
314   induced. Results from the GO enrichment analysis show a significant over-representation of
315   biological processes related to endocytosis in adult worms. In total, 21 genes annotated with
316   functional terms such as clathrin coat assembly (GO:0048268), clathrin-mediated endocytosis
317   (GO:0072583), and regulation of endocytosis (GO:0006898) are co-regulated within the same
318   cluster as reproduction-specific genes (Figure 2a, cluster 4). All 21 genes are significantly over-
319   expressed in adult worms, as compared to the previous infective stage (Table S1). Even though
320   we detect an over-expression of certain genes in adult worms that are involved in general
321   mechanisms of endocytosis, we cannot determine where exactly these genes are expressed in the
322   worm, since our experiment was performed on whole worms. They could be over-expressed in
323   cells from the integumentary system, but also in other organs that are not involved in interaction
324   with the external environment of the worm.

325

326   *Regulation of redox pathways through novel species-specific genes*

327   Adult stages of cestodes like *S. solidus* are exposed not only to the reactive oxygen species

328   (ROS) produced by their own metabolism, but also to the ones generated by their host (Williams

329   *et al.* 2013). Considering the extensive muscular activity required during reproductive behaviours

330   (Smyth 1952; Clarke 1954) and the potential internalisation of host molecules by adult worms –

331   which could include ROS produced by the host – maintenance of redox homeostasis should be a

332   central activity performed at this stage. This scenario is reflected in the smallest co-expression

333   module characterising the passage to the simulated avian host (Figure 2a, cluster 6), which

334   harbours genes predominantly up-regulated in adults. This module does not exhibit significant

335   enrichment for a particular biological activity, but it is nonetheless associated with oxidative

336   stress and antioxidant metabolism, such as glutathione metabolic process (GO:0006749),

337   glutathione biosynthetic process (GO:0006750) and glutathione dehydrogenase (ascorbate)

338   activity (GO:0045174). Interestingly, of the 242 genes contained in this module, 174 (72%) are

339   turned 'ON' in adults and 'OFF' in pre-infective and infective plerocercoids. All of the genes

340   turned 'ON' in adult worms are found exclusively in this cluster (Table S1). Furthermore, 108

341   (62%) of these 174 'ON' genes find no homology to any known sequence database, nucleotides

342   or amino acids, while they are among the top differentiated genes in the final developmental

343   transition (Figure 2b). This module is thus mainly composed of unknown genes that are co-

344   expressed with oxidative stress genes being specifically up-regulated at the adult stage.

345   Experimental evidence on the metabolism of adult worms shows a significant increase in lactate

346   concentration at this stage (Beis & Barrett 1979), which is confirmed in our data by the increased

347   expression of lactate dehydrogenase (Figure 3). Higher intracellular lactate content is considered

348   as evidence for a more oxidised cytoplasm in mature worms (Beis & Barrett 1979). The redox

349   module identified in our data supports this hypothesis of increased oxidative stress in adult

350   worms, suggesting the importance of preventing the damage caused by ROS in late stages of

351   infection.

352

353   Detecting distinct developmental stages within the same host

354   *Early plerocercoids associated with growth and regulatory programs*

355   Evidence from physiological and morphological studies suggests that growth and organ

356   development are the major biological programs that differentiate pre-infective from infective

357 plerocercoids (Clarke 1954). *In vitro* experiments showed that in the first 48 hours following

358 infection, the number of proglottids – i.e. body segments – is definitive. Unlike most of the

359 cyclophyllidean tapeworms, *S. solidus* plerocercoids increase their bulk several hundredfold by

360 adding layers of muscle tissue rather than adding proglottids (Hopkins & Smyth 1951; Clarke

361 1954). This suggests that organ development and tissue differentiation are switched off at this

362 point, while muscle synthesis and growth are switched on. We examined if this developmental

363 turning point is detectable in regulatory patterns of gene expression when comparing pre-

364 infective versus infective plerocercoids.

365

366 Overall, three out of the four co-expression modules up-regulated in pre-infective plerocercoids

367 were strongly associated with growth, cell division and regulatory functions. The first module

368 contained 478 genes predominantly up-regulated in pre-infective plerocercoids (Figure 4a, cluster

369 6) and significantly enriched in GO terms related to DNA/RNA metabolism – e.g. replication,

370 transcription and translation (Table S1). The second cluster contained 388 genes, also up-

371 regulated in pre-infective plerocercoids (Figure 4a, cluster 4), and significantly enriched in

372 biological activities involved in the regulation of cell cycle and cell division (Table S1). Among

373 genes annotated with these GO terms, those exhibiting the largest expression difference between

374 pre-infective and infective plerocercoids (i.e. logFC > 1.5, FDR < 0.001) code for mRNA

375 splicing factors, DNA polymerase and key proteins involved in the regulation of mitosis (Table

376 S1).

377

378 Developmental trajectories involving mitotic replications, cellular growth and tissue

379 differentiation are often associated with specific regulatory processes that coordinate the timing

380 of these events. Our results show that a third co-expression module significantly enriched in

381 regulatory activities may perform this task. The module of 240 genes significantly up-regulated

382 in pre-infective plerocercoids (Figure 4a, cluster 2) contains 144 (60%) genes annotated with GO

383 terms. Among these, 17 genes (12% of annotated genes) have enriched GO terms related to

384 regulatory processes involved in cellular functions such as apoptosis, mitosis, phosphorylation

385 and cell division (Table S1). The genes with the largest difference in expression level are the

386 transcription factors Sox-19b and GATA-3, with logFC values of 3.0 and 2.6 respectively (FDR

387 < 0.001). Other regulatory genes include WNT4 and WNTG, with logFC values of 1.7 and 1.6

388   respectively (FDR < 0.001). Interestingly, these 17 regulatory genes are co-regulated, within this

389   module, with other genes associated to cell cycle and DNA/RNA metabolism. In total, 41 genes

390   (28% of annotated genes in the cluster) have enriched GO terms related to biological activities

391   such as DNA replication, cell cycle, and DNA biosynthetic processes. These results suggest that

392   small pre-infective plerocercoids activate a series of regulatory pathways in their intermediate

393   fish host. We propose that these regulatory changes could result in the rapid increase in overall

394   body mass through tissue differentiation, muscle fibre synthesis, organ formation and increased

395   organ size (Benesh *et al.* 2013).

396

397   *Specific transcriptional signature of infectivity dominated by environment sensing and un-*

398   *annotated genes*

399   One of the proxies used to infer infectivity in *S. solidus* plerocercoids is its significant influence

400   on the immune system and behaviour of its fish host (reviewed in Hammerschmidt & Kurtz

401   2009), which implies communication between the two species (Adamo 2013). Consistent with

402   this, we find that environmental sensing is the dominant function represented in the

403   transcriptional signature of infectivity. The most compelling evidence comes from a large module

404   of co-expressed genes significantly up-regulated in infective plerocercoids (Figure 4a, cluster 1).

405   This module contains 407 genes significantly enriched in biological activities related to the

406   cellular response of the organism to various molecules from the external and internal

407   environment. More specifically, 84% of the 70 enriched GO terms in the module are involved in

408   cellular responses to drugs and neuromodulators, and secretion and transport of various

409   molecules through the cell membrane (Table S1). Of the 157 genes with a GO annotation in this

410   module, 28 (18%) have GO terms involved in environmental sensing and interactions. Among

411   these, those that exhibit the largest expression differences between pre-infective and infective

412   plerocercoids code for proteins including monocarboxylate transporter 7 (logFC = 4.9, FDR =

413   2.2e-06), solute carrier family 22 member 21 (logFC = 4.6, FDR = 1.2e-05), multidrug resistance

414   protein 1A (logFC = 4.4, FDR < 0.001), multidrug and toxin intrusion protein 1 (logFC = 4.4,

415   FDR = 0.0013) and neuropeptide FF receptor 2 (logFC = 2.4, FDR = 3.6e-05).

416

417   The cluster described above is particularly interesting because of the high proportion of genes

418   coding for unknown proteins differentially regulated between pre-infective and infective

14

419 plerocercoids. One of the key features of this module is that GO annotations could be assigned to
420 only 39% of the genes; hence it is the least annotated of all the modules characterizing the
421 transition of plerocercoids from pre-infective to infective. The top 15 most differentiated genes
422 between pre-infective and infective plerocercoids – with logFCs of 8-11, and FDRs < 0.00001 –
423 are all completely unknown (Figure 4b), and are all *S. solidus*-specific sequences, i.e. we find no
424 homology match to any known database except the *S. solidus* genome. These sequences also
425 exhibit valid open reading frames and are all highly expressed only in infective worms – i.e. they
426 are turned 'OFF' in pre-infective plerocercoids and adult worms. The only information that can
427 be used to assign a preliminary function to these genes is the ecological annotation that stems
428 from our transcriptomic analysis (see Materials and Methods for details). These sequences were
429 thus labelled as infective-specific and co-expressed with genes involved in environmental sensing
430 and interaction (table S1). They might hold important, yet hidden, functional aspects that would
431 allow a complete understanding of the interaction between infective plerocercoids and their fish
432 host (Koziol *et al.* 2016).

433

434 *Regulation of neural pathways could be essential for successful transmission*
435 Our findings regarding the strong signal detected for neural pathways, such as environmental
436 sensing, are further supported by another co-expression module. This module contains a total of
437 335 genes significantly up-regulated in infective worms (Figure 4a, cluster 5), among which 70%
438 have GO annotations. Our results indicate that 41 (17%) of annotated genes in the module have
439 GO annotations enriched in activities performed by the nervous system, while 124 (53%) of them
440 have GO annotations related to transmembrane structure and activity. Biological processes
441 associated with these genes include signal transduction, synaptic transmission, sensory receptor
442 activities and synaptic exocytosis (Table S1). An interesting candidate emerges as one of the top
443 differentiated genes in the module, with an expression fold change of 3.3 (FDR = 0.0003)
444 between pre-infective and infective plerocercoids. This candidate is 5-hydroxytryptamine A1-
445 alpha receptor, a serotonin receptor. Serotonin is an important regulator of carbohydrate
446 metabolism, host-parasite communication and rhythmical movements – in conjunction with other
447 related bioamines – in several cestodes (Marr & Muller 1995). Our results show a systematic up-
448 regulation of serotonin receptors, adenylate cyclase and sodium-dependent serotonin transporters
449 (SC6A4) specifically in infective worms (Figure 5). Functional studies showed that the signalling

450 cascade of serotonin stimulates muscle contraction and glycogen breakdown in *Fasciota hepatica*
451 and *Schistosoma mansoni* (Marr & Muller 1995). The ultimate downstream effect of serotonin
452 signalling would be a cellular response to catabolize glycogen, the main source of energy in
453 cestodes. In *S. mansoni*, it has been suggested that the main source of 5-HT is the host, even
454 though some of the enzymes involved in the process and recycling of 5-HT have been detected in
455 this species (Marr & Muller 1995). This is also the case with our dataset, in which we find at
456 least one enzyme that is capable of breaking down one of the metabolites required for serotonin
457 biosynthesis, i.e. tryptophan – indoleamine 2,3-dioxygenase 2, up regulated in infective worms
458 with logFC = -3.1 and FDR = 0.01. According to the current transcriptome annotation (Hébert *et*
459 *al.* 2016b), there is no sequence in the transcriptome of the pre-infective, infective and adult
460 stages that is annotated as part of the biosynthetic process of serotonin. If serotonin metabolism
461 plays such a central role in the success of *S. solidus* in its fish host without being synthesized by
462 the worm itself, we could consider the possibility that it progressively uses the host's supplies as
463 it grows.

464

465 Successful completion of a complex life cycle involves intricate interactions between the
466 parasite's developmental program and physiological parameters experienced in each host.
467 Investigating the transcriptomic signature of each developmental stage has led to the discovery of
468 multiple novel yet un-annotated transcripts. These transcripts hold significant co-regulatory
469 relationships with environmental interaction genes. Future functional characterization of these
470 parasite-specific sequences promise to reveal crucial insights on how developmental and
471 infection mechanisms evolved in different parasitic taxa.

472

473 **FIGURES**

474

475 **Figure 1. Developmental stages of *S. solidus* are characterised by different genome-wide**
476 **expression profiles.** A) Life cycle of *Schistocephalus solidus*. B) Multidimensional scaling
477 analysis (MDS) confirming the presence of three distinct phenotypes among samples (n=17). The
478 distance between two given points on the graph corresponds to the typical log2-fold-change
479 between the two samples for the top 1000 genes with the largest Euclidian distance.

480

481 **Figure 2. Differential patterns of gene expression reveal a strong stage-specific functional**
482 **signature dominated by reproduction-associated activities in adult worms.** A) Hierarchical
483 cluster analysis showing co-expression relationships between genes significantly differentially
484 expressed between infective plerocercoids and adult worms. Biological processes significantly

485   enriched in each module appear in white on the heatmap. B) Volcano plot showing genes
486   differentially expressed at three levels of FDR significance. Positive values of log2FoldChange
487   correspond to up regulated genes in adult worms. Data points circled on the graph represent 18 of
488   the top 30 most differentiated genes to which no annotation could be assigned. The un-annotated
489   genes are turned 'ON' in adult worms and are part of the redox homeostasis functional module
490   (heatmap cluster 6).
491
492   **Figure 3. Partial glycolysis KEGG pathway highlighting the biochemical steps for which**
493   **differential expression was detected between infective plerocercoids and adult worms.**
494   Boxes with a solid black line and white filling represent genes for which expression was detected
495   with no significant difference between developmental stages. Red boxes represent up regulated
496   genes in adult worms. Figure based on the complete KEGG pathway for
497   glycolysis/gluconeogenesis (http://kegg.jp).
498
499   **Figure 4. Differential patterns of gene expression suggest a significant role for**
500   **neuromodulatory pathways in the development of infectivity towards the final host.** A)
501   Hierarchical cluster analysis showing co-expression relationships between genes significantly
502   differentially expressed between pre-infective and infective plerocercoids. Biological processes
503   significantly enriched in each module appear in white on the heatmap. B) Volcano plot showing
504   genes differentially expressed at three levels of FDR significance. Positive values of
505   log2FoldChange correspond to up regulated genes in infective plerocercoids. Data points circled
506   on the graph represent the top 15 most differentiated genes, of which 100% are un-annotated.
507   These uncharacterised sequences are all species-specific and systematically turned 'ON' in
508   infective worms.
509
510   **Figure 5. Activation of serotonin-related genes in the transcriptional signature of**
511   **infectivity.** Each data point on the graph corresponds to the average gene expression level
512   (log2CPM) at a given developmental stage. Vertical bars represent the 95% confidence interval
513   of the geometric mean. Genes coding for serotonin receptors (5-HT1A), adenylate cyclase (AC)
514   and sodium-dependent serotonin transporters (SC6A4) are up-regulated specifically in infective
515   plerocercoids. Un-annotated genes co-expressed in the same modules as serotonin-related genes
516   and enriched in biological processes related to synaptic transmission and neural pathways show
517   very similar patterns of expression (open circles with dashed lines). These un-annotated genes
518   were labeled as 'secreted' based on the presence of a signal peptide in their sequence.
519

527
528

529 **REFERENCES**
530 Adamo SA (2013) Parasites: evolution's neurobiologists. *The Journal of experimental biology*,
531   **216**, 3–10.
532 Aly ASI, Vaughan AM, Kappe SHI (2009) Malaria parasite development in the mosquito and
533   infection of the mammalian host. *Annual Review of Microbiology*, **63**, 195–221.
534 Auld SK, Tinsley MC (2014) The evolutionary ecology of complex lifecycle parasites: linking
535   phenomena with mechanisms. *Heredity*, **114**, 125–132.
536 Barber I (2013) Sticklebacks as model hosts in ecological and evolutionary parasitology. *Trends*
537   *in Parasitology*, **29**, 556–566.
538 Barber I, Scharsack JP (2010) The three-spined stickleback-*Schistocephalus solidus* system: an
539   experimental model for investigating host-parasite interactions in fish. *Parasitology*, **137**,
540   411.
541 Barber I, Svensson PA (2003) Effects of experimental *Schistocephalus solidus* infections on
542   growth, morphology and sexual development of female three-spined sticklebacks,
543   *Gasterosteus aculeatus*. *Parasitology*, **126**, 359–367.
544 Barber I, Walker P, Svensson PA (2004) Behavioural Responses to Simulated Avian Predation in
545   Female Three Spined Sticklebacks: The Effect of Experimental *Schistocephalus solidus*
546   Infections. *Behaviour*, **141**, 1425–1440.
547 Barrett J (1977) Energy metabolism and infection in helminths. *Symposia of the British Society*
548   *for Parasitology*, **15**, 121-144.
549 Beis I, Barrett J (1979) The contents of adenine nucleotides and glycolytic and tricarboxylic acid
550 cycle intermediates in activated and non-activated plerocercoids of *Schistocephalus solidus*
551 (Cestoda: Pseudophyllidea) *International journal for parasitology*, **9**, 465–468.
552 Benesh DP, Chubb JC, Parker GA (2013) Complex life cycles: why refrain from growth before
553   reproduction in the adult niche? *The American Naturalist*, **181**, 39-51.
554 Charles GH, Orr TS (1968) Comparative fine structure of outer tegument of *Ligula intestinalis*
555   and *Schistocephalus solidus*. *Experimental Parasitology*, **22**, 137–149.
556 Clarke AS (1954) Studies on the life cycle of the pseudophyllidean cestode *Schistocephalus*
557   *solidus*. *Proceedings of the Zoological Society of London*, **124**, 257–302.
558 Conradt U, Peters W (1989) Investigations on the occurrence of pinocytosis in the tegument of
559   *Schistocephalus solidus*. *Parasitology Research*, **75**, 630–635.
560 Davidson NM, Oshlack A (2014) Corset: enabling differential gene expression analysis for *de*
561   *novo* assembled transcriptomes. *Genome biology*, **15**, 1–14.
562 Hammerschmidt K, Kurtz J (2005) Surface carbohydrate composition of a tapeworm in its
563   consecutive intermediate hosts: individual variation and fitness consequences. *International*
564   *journal for parasitology*, **35**, 1499–1507.
565 Hammerschmidt K, Kurtz J (2007) *Schistocephalus solidus*: Establishment of tapeworms in
566   sticklebacks – fast food or fast lane? *Experimental Parasitology*, **116**, 142–149.
567 Hammerschmidt K, Kurtz J (2009) Ecological immunology of a tapeworms' interaction with its
568   two consecutive hosts. In: *Advances in Parasitology* Advances in Parasitology. pp. 111–137.
569   Elsevier.
570 Hébert FO (2016a) Bulk codes for RNA-seq analysis. *Zenodo*.
571 Hébert FO (2016b) corset_pipeline: First complete release. *Github*.
572 Hébert FO, Grambauer S, Barber I, Landry CR (2016a) Protocols for "Transcriptome sequences
573   spanning key developmental states as a resource for the study of the cestode *Schistocephalus*
574   *solidus*, a threespine stickleback parasite." *Protocols.io*. (doi:10.17504/protocols.io.ew9bfh6)
575 Hébert FO, Grambauer S, Barber I, Landry CR, Aubin-Horth N (2016b) Transcriptome

576 sequences spanning key developmental states as a resource for the study of the cestode
577   *Schistocephalus solidus*, a threespine stickleback parasite. *GigaScience*, **5**, 24.
578 Hébert FO, Grambauer S, Barber I, Landry CR, Aubin-Horth N (2016c) Resource for
579   "Transcriptome sequences spanning key developmental states as a resource for the study of
580   the cestode *Schistocephalus solidus*, a threespine stickleback parasite." *GigaScience*
581   *Database*. (doi:10.5524/100197)
582 Hopkins CA (1950) Studies on cestode metabolism. I. Glycogen metabolism in *Schistocephalus*
583   *solidus* in vivo. *Journal of Parasitology*, **36**, 384–390.
584 Hopkins CA (1952) Studies on cestode metabolism. II. The utilization of glycogen by
585   *Schistocephalus solidus* in vitro. *Experimental Parasitology*, **1**, 196–213.
586 Hopkins CA, Smyth JD (1951) Notes on the morphology and life history of *Schistocephalus*
587   *solidus* (Cestoda: Diphyllobothriidae. *Parasitology*, **41**, 283–291.
588 Hopkins CA, Law LM, Threadgold LT (1978) *Schistocephalus solidus*: pinocytosis by the
589   plerocercoid tegument. *Experimental Parasitology*, **44**, 161–172.
590 Klopfenstein D, Pedersen B, Flick P *et al.* (2015) GOATOOLS: Tools for Gene Ontology.
591   *Zenodo*. https://zenodo.org/record/31628 (doi:10.5281/zenodo.31628)
592 Koziol U, Koziol M, Preza M *et al.* (2016) *De novo* discovery of neuropeptides in the genomes of
593   parasitic flatworms using a novel comparative approach. *International journal for*
594   *parasitology*, 1–13.
595 Körting W, Barrett J (1977) Carbohydrate catabolism in the plerocercoids of *Schistocephalus*
596   *solidus* (Cestoda: Pseudophyllidea). *International journal for parasitology*, **7**, 411–417.
597 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**,
598   357–359.
599 Law CW, Chen Y, Shi W, Smyth GK (2014) voom: Precision weights unlock linear model
600   analysis tools for RNA-seq read counts. *Genome biology*, **15**, R29.
601 Lee DL (1967) The Structure and Composition of the Helminth Cuticle. In: *Advances in*
602   *Parasitology Volume 4* Advances in Parasitology. pp. 187–254. Elsevier.
603 Marr J, Muller M (1995) *Biochemistry and Molecular Biology of Parasites*. Academic Press,
604   London.
605 Oshima K, Ishii Y, Kakizawa S *et al.* (2011) Dramatic transcriptional changes in an intracellular
606   parasite enable host switching between plant and insect. *Plos One*, **6**, e23242.
607 Pavey SA, Bernatchez L, Aubin-Horth N, Landry CR (2012) What is needed for next-generation
608   ecological and evolutionary genomics? *Trends in Ecology & Evolution*, **27**, 673–678.
609 Poulin R (2011) *Evolutionary Ecology of Parasites*. Princeton University Press, Princeton.
610 Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential
611   expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
612 Scharsack JP, Gossens A, Franke F, Kurtz J (2013) Excretory products of the cestode,
613   *Schistocephalus solidus*, modulate *in vitro* responses of leukocytes from its specific host, the
614   three-spined stickleback (*Gasterosteus aculeatus*). *Fish & Shellfish Immunology*, **35**, 1779–
615   1787.
616 Scharsack JP, Koch K, Hammerschmidt K (2007) Who is in control of the stickleback immune
617   system: interactions between *Schistocephalus solidus* and its specific vertebrate host.
618   *Proceedings Of The Royal Society B-Biological Sciences*, **274**, 3151–3158.
619 Schjørring S (2003) Schistocephalus solidus: a molecular test of premature gamete exchange for
620   fertilization in the intermediate host *Gasterosteus aculeatus*. *Experimental Parasitology*, **103**,
621   174–176.
622 Smyth DJ (1946) Studies on tapeworm physiology, the cultivation of *Schistocephalus solidus* in

623   vitro. *The Journal of experimental biology*, **23**, 47–70.

624   Smyth JD (1950) Studies on tapeworm physiology. V. Further observations on the maturation of
625   *Schistocephalus solidus* (Diphyllobothriidae) under sterile conditions in vitro. *The Journal of*
626   *parasitology*, **36**, 371.

627   Smyth JD (1952) Studies on tapeworm physiology. VI. Effect of temperature on the maturation
628   *in vitro* of *Schistocephalus solidus*. *The Journal of experimental biology*, **29**, 304–309.

629   Smyth JD (1954) Studies on tapeworm physiology. VII. Fertilization of *Schistocephalus solidus*
630   in vitro. *Experimental Parasitology*, **3**, 64–67.

631   Smyth JD, McManus DP (2007) *The Physiology and Biochemistry of Cestodes*. Cambridge
632   University Press, Cambridge.

633   Threadgold LT, Hopkins CA (1981) *Schistocephalus solidus* and *Ligula intestinalis*: pinocytosis
634   by the tegument. *Experimental Parasitology*, **51**, 444–456.

635   Tierney JF, Crompton DW (1992) Infectivity of plerocercoids of *Schistocephalus solidus*
636   (Cestoda: Ligulidae) and fecundity of the adults in an experimental definitive host, *Gallus*
637   *gallus*. *J Parasitol*, **78**, 1049–1054.

638   Warnes GR, Bolker B, Bonebakker L *et al.* (2016) gplots: Various R programming tools for
639   plotting data. *R package version 2.0.1*.

640   Wilbur HM (1980) Complex life cycles. *Annual review of Ecology and Systematics*, **11**, 67-93.

641   Williams DL, Bonilla M, Gladyshev VN, Salinas G (2013) Thioredoxin glutathione reductase-
642   dependent redox networks in platyhelminth parasites. *Antioxidants & redox signaling*, **19**,
643   735–745.

644

645   **ETHICS**
646   Fish were captured under UK Environment Agency permit and with the permission of the
647   landowner. All experiments were undertaken under a UK Home Office license (PPL80/2327)
648   held by IB, in accordance with local and national regulations and with ABS/ASAB guidelines for
649   the ethical treatment of animals in behavioral research (available online at
650   http://asab.nottingham.ac.uk/ethics/guidelines.php).

651

652   **DATA ACCESSIBILITY**
653   -   Raw sequencing data: Sequence Read Archive (SRA) accession number SAMN04296611
654   -   Final transcriptome sequences: NCBI BioProject PRJNA304161, uploaded with annotation.
655   -   All data associated with the transcriptome: http://gigadb.org/dataset/100197
656   -   Protocols: *In vitro*/*in vivo* culturing techniques and protocols available via protocols.io
657   (doi:10.17504/protocols.io.ew9bfh6).
658   -   Python and R codes used for data analysis are available through github (Hébert 2016a,b).

659

660   **AUTHORS' CONTRIBUTIONS**
661   FOH, IB, CRL and NAH conceived the study. FOH and SG did the laboratory infections. FOH,
662   SG and IB undertook fish dissections and parasite culture. FOH extracted RNA, prepared the
663   sequencing libraries, performed bioinformatic analyses with supervision from NAH and CRL.
664   FOH, CRL and NAH drafted the manuscript with input from IB. All authors read and approved
665   the final manuscript.

666

667   **SUPPORTING INFORMATION**
668   **Table S1** Integrated results from the differential gene expression analysis for the complete
669   transcriptome of *Schistocephalus solidus*. Summarises the information pertaining to the results

670   form all analyses for each transcript included in the transcriptome, including unigene ID, gene
671   product, GO annotation, association with co-expression modules, ON/OFF status, and differences
672   in expression levels for each life stage transition (logFC with corresponding FDR value).
673
674   **COMPETING INTERESTS**
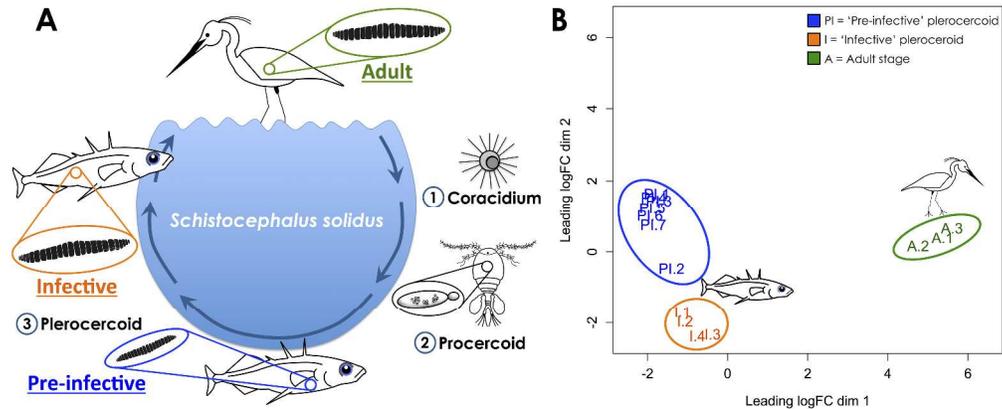675   The authors declare that they have no competing interests.
676

Figure 1. Developmental stages of S. solidus are characterised by different genome-wide expression profiles. A) Life cycle of Schistocephalus solidus. B) Multidimensional scaling analysis (MDS) confirming the presence of three distinct phenotypes among samples (n=17). The distance between two given points on the graph corresponds to the typical log2-fold-change between the two samples for the top 1000 genes with the largest Euclidian distance.
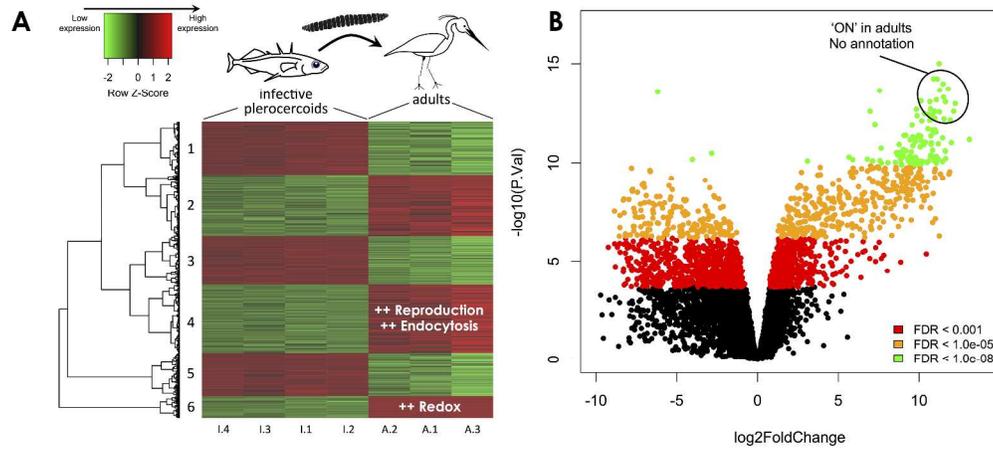
Figure 2. Differential patterns of gene expression reveal a strong stage-specific functional signature dominated by reproduction-associated activities in adult worms. A) Hierarchical cluster analysis showing co-expression relationships between genes significantly differentially expressed between infective plerocercoids and adult worms. Biological processes significantly enriched in each module appear in white on the heatmap. B) Volcano plot showing genes differentially expressed at three levels of FDR significance. Positive values of log2FoldChange correspond to up regulated genes in adult worms. Data points circled on the graph represent 18 of the top 30 most differentiated genes to which no annotation could be assigned. The un-annotated genes are turned 'ON' in adult worms and are part of the redox homeostasis functional module (heatmap cluster 6).
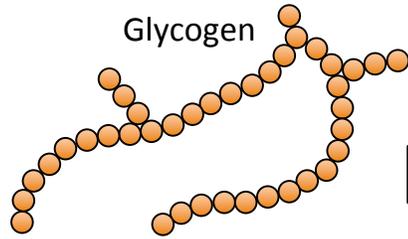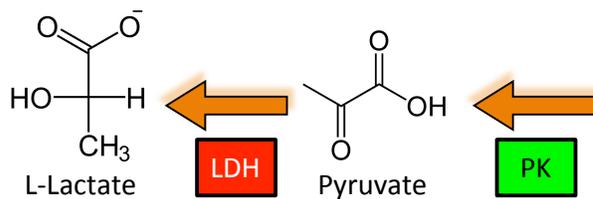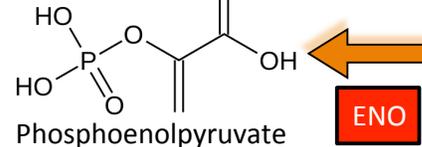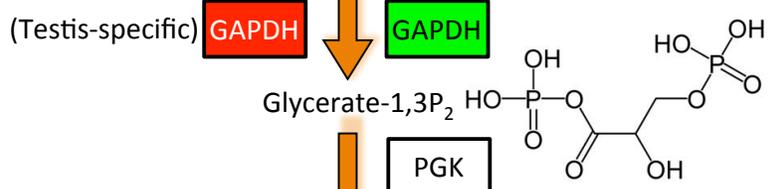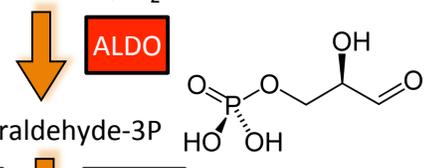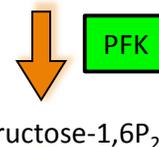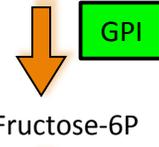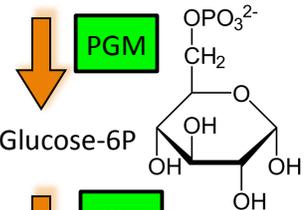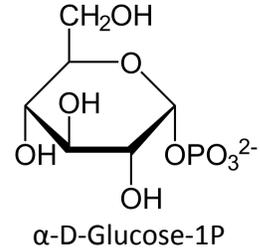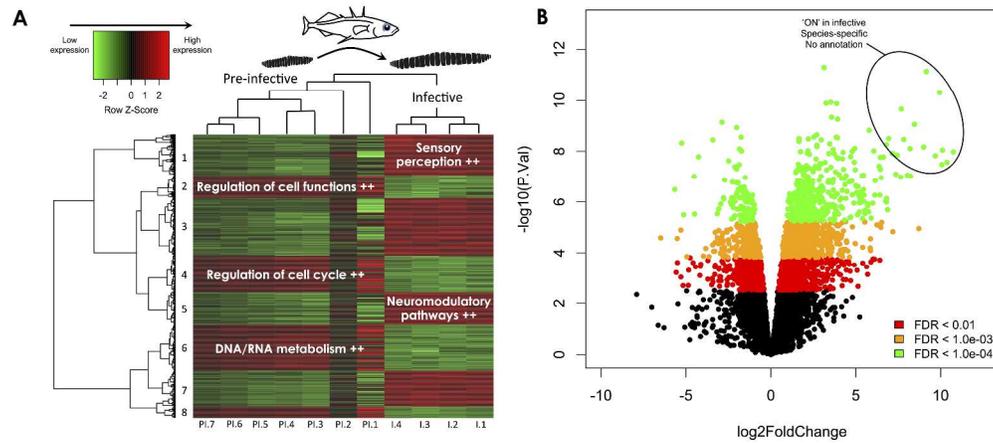
Figure 4. Differential patterns of gene expression suggest a significant role for neuromodulatory pathways in the development of infectivity towards the final host. A) Hierarchical cluster analysis showing co-expression relationships between genes significantly differentially expressed between pre-infective and infective plerocercoids. Biological processes significantly enriched in each module appear in white on the heatmap. B) Volcano plot showing genes differentially expressed at three levels of FDR significance. Positive values of log2FoldChange correspond to up regulated genes in infective plerocercoids. Data points circled on the graph represent the top 15 most differentiated genes, of which 100% are un-annotated. These uncharacterised sequences are all species-specific and systematically turned 'ON' in infective worms.