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

François Olivier HÉBERT[1], Stephan GRAMBAUER[2], Iain BARBER[2], Christian R LANDRY[1], Nadia AUBIN-HORTH[1]*

1. Institut de Biologie Intégrative et des Systèmes (IBIS), Département de Biologie, Université Laval, 1030 avenue de la Médecine, G1V 0A6, Québec (QC), Canada.
Phone: 1-418-656-3316
Fax: 1-418-656-7176

2. Department of Neuroscience, Psychology and Behaviour, Adrian Building, University of Leicester, University Road, Leicester, LE1 7RH, UK.

* Corresponding author: Nadia.Aubin-Horth@bio.ulaval.ca

Running title

Parasite transcriptome reprogramming

Keywords

Parasite, cestode, *Schistocephalus solidus*, threespine stickleback, bird, complex life cycle, RNA-seq

ABSTRACT (205 words)

Parasites with complex life cycles have developed numerous phenotypic strategies, closely associated with developmental events, to enable the exploitation of different ecological niches and facilitate transmission between hosts. How these environmental shifts are regulated from a metabolic and physiological standpoint, however, still remain to be fully elucidated. We examined the transcriptomic response of *Schistocephalus solidus*, a trophically-transmitted parasite with a complex life cycle, over the course of its development in an intermediate host, the threespine stickleback, and the final avian host. Results from our differential gene expression analysis show major reprogramming events among developmental stages. The final host stage is characterized by a strong activation of reproductive pathways and redox homeostasis. The attainment of infectivity in the fish intermediate host – which precedes sexual maturation in the final host and is associated with host behaviour changes – is marked by transcription of genes involved in neural pathways and sensory perception. Our results suggest that un-annotated and *S. solidus*-specific genes could play a determinant role in host-parasite molecular interactions required to complete the parasite's life cycle. Our results permit future comparative analyses to help disentangle species-specific patterns of infection from conserved mechanisms, ultimately leading to a better understanding of the molecular control and evolution of complex life cycles.

INTRODUCTION

Parasites with multiple hosts commonly undergo dramatic phenotypic transformations and endure major environmental shifts over the course of their life cycle [1,2], yet very little is known about how these are orchestrated at the molecular and physiological levels, or how conserved they are across species [3]. 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 [4], 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 [5]. 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 [6]. 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 [7]. During the early plerocercoid phase, the host immune system is not activated by the presence of the worm inside its body cavity [8]. *S. solidus* spends the next 50-60 days in an exponential growth phase, gaining up to 20 times its initial mass [9]. When the plerocercoid eventually

reaches infectivity, a phase that could be determined by sufficient glycogen reserves [10], drastic phenotypic changes occur in the fish host [11]. These changes include an activation of the immune system [8] and a loss of anti-predator response [9]. Following the ingestion of infected fish by an avian predator, the parasite experiences a temperature of 40°C in the bird's digestive tract – compared to a maximum of 15-18°C in the ectothermic intermediate hosts – as well as chemical attack by digestive enzymes. These conditions trigger the parasite's development to the sexually mature adult in ca. 36 hours [12]. The adult parasite reproduces during the next 3-4 days with eggs being released into the water with the avian host's faeces [13]. Adjusting to these host switches and life history transitions requires many physiological changes that are expected to recruit the activity of different genes at each phase.

One of the major transitions expected to affect worm physiology is the transition from somatic growth to reproduction. Histological and physiological studies suggest that gametogenesis only occurs when the parasite reaches the final (bird) host [10,14]. Despite the advanced ('progenetic'; [45]) development of reproductive organs in infective plerocercoids, only an elevated temperature of 40°C in semi-anaerobic conditions can trigger meiosis and reproductive behaviours [14,15]. Previous work on the anaerobic activity of key enzymes involved in the catabolism of carbohydrates in *S. solidus* also suggests that while carbohydrate breakdown is very slow in pre-infective and infective plerocercoids, the rate of carbohydrate breakdown increases several-fold upon maturation [16,17]. Energetic resources used during the adult stage mainly come from glycogen reserves accumulated during growth of the pre-infective plerocercoid [16,18]. Based on this knowledge, one expects that plerocercoids cultured at 40°C will show an up-regulation of glycogen-related pathways.

The adult stage interacts with its environment to time these developmental steps. The anatomical structure more likely to achieve this task is the tegument, a very active and complex tissue that behaves like a true epidermis [19]. The adult stage of *S. solidus* exhibits numerous vacuolate vesicles packed with electron-dense or electron-lucent content. These small structures are evenly distributed in *S. solidus* syncytial tegument [20]. Their role could be related to both nutrition and defence, as they allow rapid internalization of environmental nutrients and antigen-antibody complexes [21,22]. However, the uptake of macromolecules by adult cestodes remains an open

question [23]. If, however, *S. solidus* performs pinocytosis or endocytosis at any developmental stage, specific transcripts involved in this biological activity are expected to be up-regulated at these stages. The existence of membrane-bound vesicles also suggests potential secretory/excretory functions that would allow the parasite to release various types of molecules in its host.

Pre-infective and infective plerocercoids are discrete developmental stages distinguished by their divergent growth and effects on the host immune system. Parasites grow rapidly in the first weeks of the stickleback host infection but growth rates tend to slow down as the parasite becomes infective to the final host [24]. Concurrently, empirical evidence shows that secretory/excretory products from pre-infective versus infective plerocercoids have different modulatory effects on the immune system of the fish host [25]. Small, pre-infective plerocercoids down-regulate the proliferation of host monocytes, but as soon as they attain infectivity they activate a strong respiratory burst activity [8]. From a transcriptional perspective, different functional programs between pre-infective and infective stages that reflect these divergent activities should be detectable. Distinct and specific gene expression profiles should characterize each developmental stage according to the biological activities that they need to perform to ultimately maximize the parasite's success in each host.

MATERIAL AND METHODS

Study design

Worm specimens used spanned three different development states and were extracted from the body cavities of laboratory-raised and experimentally infected threespine sticklebacks. We obtained parasite eggs through *in vitro* culture of mature plerocercoids [26] extracted from wild-caught threespine sticklebacks from Clatworthy Reservoir, England, UK. After a three week incubation period in tap water, eggs were hatched in response to exposure to daylight, and emergent coracidia used to infect copepods. Exposed copepods were screened after three to four weeks and those harboring infective procercoids were fed to healthy lab-bred threespine sticklebacks [27]. One hundred fish were exposed to infected copepods and maintained in laboratory aquaria for 16 weeks under controlled temperatures, where they were fed *ad libitum* and to excess with frozen chironomid larvae and subsequently euthanized by an overdose of 15

137 mM Benzocaine solution. Sequential extractions of the parasites from the fish body cavity were
138 scheduled to allow the sampling of plerocercoids that were pre-infective (i.e <50 mg) or infective
139 (i.e. >50mg). We obtained three adult specimens of *S. solidus* through *in vitro* culture of infective
140 plerocercoids extracted from wild-caught sticklebacks of the same population as the experimental
141 infections [28]. Briefly, infective plerocercoids were placed individually into a dialysis
142 membrane suspended in a medium composed of 50:50 RPMI:horse serum, at a temperature, pH
143 and oxygen tension to mimic the specific conditions experienced in the bird digestive track [12] –
144 for detailed protocol see [27]. Recovered adult worms had a body mass of 321-356 mg. Worms
145 were washed with ultra-pure RNase-free water, diced into small pieces of ~5 mm x 5 mm, placed
146 into RNALater (Ambion Inc., Austin, TX, USA) and kept at -80°C until later use.
147
148 *RNA extraction, library preparation and sequencing*
149 We used RNA samples from fourteen different worms to produce individual TruSeq Illumina
150 sequencing libraries (San Diego, CA, USA) according to the manufacturer's protocol. More
151 specifically, we produced libraries for seven pre-infective (<50mg) plerocercoids, four infective
152 plerocercoids (>50mg) and three adult (post-culture) worms [28]. cDNA libraries were then
153 sequenced on the Illumina HiSeq 2000 system at Centre de Recherche du CHU de Québec
154 (Québec, QC, Canada) with the paired-end technology (2X100 bp). In total, 75.8 Gb of raw data
155 was generated, which represents 375 million 2 x 100 bp paired-end sequences distributed across
156 the fourteen samples – deposited into the NCBI Sequence Read Archive (SRA) with accession
157 number SAMN04296611 associated with BioProject PRJNA304161 [29].
158
159 *Short-read alignment on reference transcriptome*
160 Raw sequencing reads were cleaned, trimmed and aligned on the reference transcriptome,
161 allowing the estimation of transcript-specific expression levels for each individual worm [28,30].
162 In summary, we aligned short reads from the 14 individual HiSeq libraries on the reference with
163 Bowtie 2 v.2.1.0 [31], allowing multi-mapping of each read. Transcripts showing similar
164 sequence, length and expression levels were then regrouped into clusters of unigenes by using
165 Corset v1.00 [32]. We obtained read counts for each unigene in each individual worm using the
166 mapping information contained in the SAM files [30]. We adjusted the algorithm parameters in
167 Corset so that isoforms, pseudogenes, alternative transcripts and paralogs would not be merged –

168 i.e. contig ratio test parameter switched on – thus eliminating a common bias in RNA-seq
169 experiments in which read counts of sequences of varying lengths are merged [33].
170
171 *Differential expression analysis and clustering*
172 We conducted downstream analyses using the R packages 'limma-voom' [34] and edgeR [35].
173 We imported the read count matrix into R v.3.3.2 (R Development Core Team, 2008) for initial
174 filtering of lowly expressed transcripts. We kept sequences with more than 15 Counts Per Million
175 (CPM) in at least three samples in any of the life stage. The filtering threshold values were
176 chosen based on a comparative analysis of multiple datasets produced with different
177 combinations of thresholds (Figure S1). This specific threshold allowed to filter out many low-
178 coverage transcripts and potentially several false positives, without losing too much information
179 in terms of the number of differentially expressed transcripts. Normalization was performed
180 using the method of Trimmed Mean of M-values (TMM) as implemented using edgeR default
181 parameters. The voom transformation was performed on the resulting normalized read counts.
182 Raw read counts were converted into their respective CPM value and then log2-transformed.
183 Next, each transcript was fitted to an independent linear model using the $\log_2$(CPM) values as the
184 response variable and the group – pre-infective plerocercoid (<50mg), infective plerocercoid
185 (>50mg) and adult – as the explanatory variable. No intercept was used and all possible
186 comparisons between the three developmental stages were defined as our desired contrasts. Each
187 linear model was then analyzed through limma's Bayes pipeline. This last step allowed the
188 discovery of differentially expressed transcripts based on a False Discovery Rate (FDR) < 0.001
189 [34,36].
190
191 We performed hierarchical clustering among transcripts and samples using the limma-voom
192 transformed CPM values through the 'heatmap.2' function in the 'gplots' package v.2.17.0 [37]
193 in R v.3.2.2. For each life-stage transition, samples were first clustered based on Euclidean
194 distance among transcript abundance values and then plotted on a heatmap by re-ordering the
195 values by transcripts (rows) and by samples (columns). We evaluated the robustness of each
196 cluster of transcripts identified through this method using the R package 'fpc' v.2.1.10 (Flexible
197 Procedures for Clustering, Hennig 2015), which implements a bootstrapping algorithm on values
198 of the Jaccard index to return a cluster stability index [36]. We only considered clusters with a

199 stability index greater than 0.50 with 1,000 bootstraps for downstream Gene Ontology (GO)
200 enrichment analyses. In total, 12 out of 16 clusters distributed across the two heatmaps were kept
201 for GO enrichment analysis. We considered each cluster satisfying the stability index threshold as
202 a module of co-expressed genes potentially bearing a broad functional status in accordance with
203 the results from the GO enrichment analysis.
204
205 *GO enrichment*
206 We identified functional categories over-represented in each co-expression module in order to
207 characterize the biological functions that are associated with each life-stage transition. We used
208 the Python package 'goatools' [38] to perform Fisher's exact tests on GO annotation terms found
209 in clusters of significantly differentially expressed gene. Annotation of GO terms for each gene
210 was based on the published transcriptome of *Schistocephalus solidus* [28]. GO terms over-
211 represented in a given module, as compared to the reference transcriptome (FDR ≤ 0.05), were
212 labelled as putative 'transition-specific' biological functions.
213
214 *Ecological annotation*
215 We assigned an ecological annotation to transcripts exhibiting significant abundance changes
216 between life stages or showing stage-specific expression patterns [39]. Two different types of
217 ecological annotation were added to the dataset. First, we labelled un-annotated transcripts
218 according to their significant variation in abundance across life stages. Information on GO terms
219 over-represented in the cluster in which these transcripts could be found was also added. Second,
220 we labelled transcripts showing "on-off patterns of expression" among stages and hosts as "stage-
221 specific" or "host-specific". A pre-defined specificity threshold was chosen as the $\log_2$(CPM)
222 value representing the 5th percentile of the distribution of the $\log_2$(CPM) across all transcripts.
223 We identified stage-specific transcripts based on an average $\log_2$(CPM) above our pre-defined
224 specificity threshold ('ON') across at least two-thirds of the worms in only one of the three life-
225 stages – i.e. transcript is 'ON' in a given stage and 'OFF' in the two other stages. Similarly, we
226 considered transcripts as host-specific if they showed an average $\log_2$(CPM) above the same pre-
227 defined specificity threshold across two-thirds of the worms in any of the two hosts.
228
229 **RESULTS AND DISCUSSION**

230 Host transition as the main driver of genome reprogramming

231 A total of 2894 genes (28% of transcriptome) are significantly differentially regulated (FDR <
232 0.001) over the course of the infection of the fish and bird hosts (Tables S1 & S2). A
233 multidimensional scaling analysis (MDS) performed on the top 1000 most differentiated genes in
234 the dataset further suggests that the main factor that drives the divergence among individual
235 worms is host type (fish vs. bird; Figure 1b). The first dimension of the MDS plot shows two
236 distinct clusters: one with adult worms and another regrouping pre-infective and infective worms.
237 This analysis also shows the grouping of pre-infective and infective worms into two different
238 clusters on the second dimension. The distance on the first dimension between host types is at
239 least twice as large as the distance separating pre-infective and infective worms, suggesting that
240 host type is the main driving factor. This may largely be explained by physiological
241 acclimatisation of the parasite to highly divergent thermal environments offered by the two hosts,
242 or to other differences such as oxygen tension, pH or osmotic pressure [12,40,41]. The switch
243 between these two hosts also correlates with rapid sexual maturation, reproduction and changes
244 in energy metabolism [5]. Altogether, these factors contribute to a major reprogramming of the
245 worm transcription profile between hosts.

246

247 *Biological activities focused towards reproductive functions*

248 The development of the adult stage in the avian host required the parasite to shift most of its
249 biological activities from growth and immune evasion [13,42] to reproduction and possibly
250 starvation [10,43]. In accordance with these life-history changes, sexual maturation pathways and
251 reproductive behaviours were dominant functions in the transcriptional signature of the final
252 host-switch, as supported by GO terms significantly enriched in co-expression modules (Figure
253 2a, Table S1). The largest co-expression module identified in the transition from infective
254 plerocercoid to adult contains a total of 769 genes significantly over-expressed in adult worms
255 (Figure 2a, cluster 4). This module is enriched (FDR < 0.05) in biological processes related to
256 reproductive functions such as spermatid nucleus differentiation (GO:0007289), sperm motility
257 (GO:0030317), luteinizing hormone secretion (GO:0032275) and positive regulation of
258 testosterone secretion (GO:2000845) (Table S1). Early studies on the life cycle of *S. solidus*
259 suggested that once the worm reaches the final bird host, its energy is canalized into maturation
260 and reproduction, including egg-laying [5,13]. Adult worms sampled in this study were collected

261 after five days of *in vitro* culture at 40°C, 3-4 days after the onset of gamete production, as
262 observed in previous studies [26,43]. The transcriptional signature consequently confirms this at
263 the molecular level, as we have detected the induction of many genes involved in sperm motility
264 and cilium movement (Table S1).
265
266 *Re-organisation of the energy budget*
267 The transition from infective plerocercoid to adult is characterized by a significant shift in energy
268 metabolism [44]. Empirical data suggests that during the first hours of maturation and
269 reproduction, worms utilize glycogen reserves accumulated in the fish host [10]. Adult worms
270 cultured *in vitro* are also capable of absorbing glucose after more than 40 hours at 40°C [18],
271 suggesting they can stop using their glycogen reserves and instead use host-derived nutrients. Our
272 results suggest a complex and subtle pattern of regulation in terms of carbohydrate metabolism.
273 In total, eleven steps of the glycolysis pathway were differentially-regulated between the
274 infective plerocercoid and adult stages (Figure 3). The first step in glycogen breakdown consists
275 in converting glycogen to glucose-1-phosphate, a reaction catalysed by the enzyme glycogen
276 phosphorylase [45]. This enzyme is strongly up-regulated in adult worms (logFC = 5.9, FDR <
277 0.0001), suggesting an active use of glycogen reserves at this stage. The first three major
278 biochemical transformations leading to glucose breakdown into more simple sugars are strongly
279 down-regulated (Figure 3). Intriguingly, genes coding for the enzyme that produce
280 glyceraldehyde-3-phosphate (GADP) are consistently up-regulated in adult worms. All three
281 homologous genes identified as fructose-bisphosphate aldolase in our dataset, the enzyme
282 responsible for the production of GADP, were labeled as being switched 'ON' in adults (see
283 Materials and Methods for details). Most of the downstream genes leading to the production of
284 pyruvate are down-regulated, with the exception of enolase, the enzyme responsible for the
285 penultimate step of glycolysis, i.e. the conversion of glycerate-2-phosphate into phosphoenol-
286 pyruvate (Figure 3, Table S1). Consistent with the semi-anaerobic conditions experienced by
287 adult worms, we found a significant up-regulation of the gene coding for L-lactate dehydrogenase
288 (logFC = 5.8, FDR < 0.001), the enzyme responsible for the conversion of pyruvate to lactate
289 when oxygen supplies are low [45].
290

291 Interestingly, we found a testis-specific gene, coding for glyceraldehyde-3-phosphate
292 dehydrogenase, among the few up-regulated genes of the glycolysis pathway. The gene was
293 significantly over-expressed in adult worms, with a fold-change of 97 (logFC = 6.6) and an FDR
294 < 0.0001. A homologous gene – with the same annotation, but this time not testis-specific – is
295 conversely down-regulated in adults (logFC = -1.6, FDR = 0.0003). These results suggest that
296 late stages of adult *S. solidus* may still be very active in terms of sperm production, even after
297 several days in the avian host. This opposite pattern is suggestive of a differential activity in
298 terms of energy metabolism between different tissues in adults. At this late stage, the adult
299 parasite may direct all of its energetic activities towards sperm production, in order to maximize
300 rates of egg fertilization.
301
302 *Potential role for endocytosis in balancing energetic reserves*
303 How glucose is produced or acquired by adult *S. solidus* is unclear, but empirical evidence
304 suggests that this activity could be performed by molecular mechanisms such as endocytosis or
305 pinocytosis [21]. This hypothesis led to the prediction that expression of genes specific to this
306 pathway should be induced. Results from the GO enrichment analysis show a significant over-
307 representation of biological processes related to endocytosis in adult worms. In total, 21 genes
308 annotated with functional terms such as clathrin coat assembly (GO:0048268), clathrin-mediated
309 endocytosis (GO:0072583), and regulation of endocytosis (GO:0006898) are co-regulated within
310 the same cluster as reproduction-specific genes (Figure 2a, cluster 4). All 21 genes are
311 significantly over-expressed in adult worms, as compared to the previous infective stage (Table
312 S1). Even though we detect an over-expression of certain genes in adult worms that are involved
313 in general mechanisms of endocytosis, we cannot determine where exactly these genes are
314 expressed in the worm, since our experiment was performed on whole worms. They could be
315 over-expressed in cells from the integumentary system, but also in other organs that are not
316 involved in interaction with the external environment of the worm.
317
318 *Regulation of redox pathways through novel species-specific genes*
319 Adult stages of cestodes like *S. solidus* are exposed not only to the reactive oxygen species
320 (ROS) produced by their own metabolism, but also to the ones generated by their host [46].
321 Considering the extensive muscular activity required during reproductive behaviours [5,15] and

322 the potential internalisation of host molecules by adult worms – which could include ROS
323 produced by the host – maintenance of redox homeostasis should be a central activity performed
324 at this stage. This scenario is reflected in the smallest co-expression module characterising the
325 passage to the final host (Figure 2a, cluster 6), which harbours genes predominantly up-regulated
326 in adults. This module does not exhibit significant enrichment for a particular biological activity,
327 but it is nonetheless associated with oxidative stress and antioxidant metabolism, such as
328 glutathione metabolic process (GO:0006749), glutathione biosynthetic process (GO:0006750)
329 and glutathione dehydrogenase (ascorbate) activity (GO:0045174). Interestingly, of the 242 genes
330 contained in this module, 174 (72%) are turned 'ON' in adults and 'OFF' in pre-infective and
331 infective plerocercoids. All of the genes turned 'ON' in adult worms are found exclusively in this
332 cluster (Table S1). Furthermore, 108 (62%) of these 174 'ON' genes find no homology to any
333 known sequence database, nucleotides or amino acids, while they are among the top
334 differentiated genes in the final developmental transition (Figure 2b). This module is thus mainly
335 composed of unknown genes that are co-expressed, with oxidative stress genes being specifically
336 up-regulated at the adult stage, while being turned off during the development phase in the fish.
337 Experimental evidence on the metabolism of adult worms shows a significant increase in lactate
338 concentration at this stage [17], which is confirmed in our data by the increased expression of
339 lactate dehydrogenase (Figure 3). Higher intracellular lactate content was then considered as
340 evidence for a more oxidised cytoplasm in mature worms [17]. The redox module identified in
341 our data supports this hypothesis of increased oxidative stress in adult worms, suggesting the
342 importance of preventing the damage caused by ROS in late stages of infection.
343
344 <u>Detecting distinct developmental stages within the same host</u>
345 *Early plerocercoids associated with growth and regulatory programs*
346 Evidence from physiological and morphological studies suggests that growth and organ
347 development are the major biological programs that differentiate pre-infective from infective
348 plerocercoids [5]. *In vitro* experiments showed that in the first 48 hours following infection, the
349 number of proglottids – i.e. body segments – is definitive. Unlike most of the cyclophyllidean
350 tapeworms, *S. solidus* plerocercoids increase their bulk several hundredfold by adding layers of
351 muscle tissue rather than adding proglottids [5,13]. This suggests that organ development and
352 tissue differentiation are switched off at this point, while muscle synthesis and growth are

353 switched on. We examined if this developmental turning point is detectable in regulatory patterns
354 of gene expression when comparing pre-infective versus infective plerocercoids.
355
356 Overall, three out of the four co-expression modules up-regulated in pre-infective plerocercoids
357 showed a transcriptional signature strongly associated with growth, cell division and regulatory
358 functions. The first module contained 478 genes predominantly up-regulated in pre-infective
359 plerocercoids (Figure 4a, cluster 6) and significantly enriched in GO terms related to DNA/RNA
360 metabolism – e.g. replication, transcription and translation. The second cluster contained 388
361 genes, also up-regulated in pre-infective plerocercoids (Figure 4a, cluster 4), and significantly
362 enriched in biological activities involved in the regulation of cell cycle and cell division. More
363 specifically, our results show that among the 416 genes (87%) annotated with GO terms in cluster
364 6, 203 genes (49% of annotated genes) have at least one enriched GO term related to DNA/RNA
365 metabolism (Table S1). The second co-expression module contains 322 annotated genes, of
366 which 118 (37%) exhibit significant enrichment for GO terms involved in cell cycle transitions
367 (cluster 4 in Figure 4a, Table S1). These genes harbour 59 (76%) of the 78 enriched GO terms in
368 the module. Among genes assigned these GO terms, those exhibiting the largest expression
369 difference between pre-infective and infective plerocercoids (i.e. logFC > 1.5, FDR < 0.001) code
370 for mRNA splicing factors, DNA polymerase and key proteins involved in the regulation of
371 mitosis (Table S1).
372
373 Developmental trajectories involving mitotic replications, cellular growth and tissue
374 differentiation are often associated with specific regulatory processes that coordinate the timing
375 of these events [47,48]. Our results show that a third co-expression module significantly enriched
376 in regulatory activities may perform this task. The module of 240 genes significantly up-
377 regulated in pre-infective plerocercoids (Figure 4a, cluster 2) contains 144 (60%) genes annotated
378 with GO terms. Among these, 17 genes (12% of annotated genes) have enriched GO terms
379 related to regulatory processes involved in cellular functions such as apoptosis, mitosis,
380 phosphorylation and cell division (Table S1). The genes with the largest difference in expression
381 level are the transcription factors Sox-19b and GATA-3, with logFC values of 3.0 and 2.6
382 respectively (FDR < 0.001). Other regulatory genes include WNT4 and WNTG, with logFC
383 values of 1.7 and 1.6 respectively (FDR < 0.001). Interestingly, these 17 regulatory genes are co-

384 regulated, within this module, with other genes associated to cell cycle and DNA/RNA
385 metabolism. In total, 41 genes (28% of annotated genes in the cluster) have enriched GO terms
386 related to biological activities such as DNA replication, cell cycle, and DNA biosynthetic
387 processes. These results suggest that small pre-infective plerocercoids activate a series of
388 regulatory pathways in their intermediate fish host. We propose that these regulatory changes
389 could result in the rapid increase in overall body mass through tissue differentiation, muscle fibre
390 synthesis, organ formation and increased organ size [49].
391
392 *Specific transcriptional signature of infectivity dominated by environment sensing and un-*
393 *annotated genes*
394 One of the proxies used to infer infectivity in *S. solidus* plerocercoids is its significant influence
395 on the immune system and behaviour of its fish host (reviewed in [50]), which implies
396 communication between the two species [51]. Consistent with this, we find that environmental
397 sensing is the dominant function represented in the transcriptional signature of infectivity. The
398 most compelling evidence comes from a large module of co-expressed genes significantly up-
399 regulated in infective plerocercoids (Figure 4a, cluster 1). This module contains 407 genes
400 significantly enriched in biological activities related to the cellular response of the organism to
401 various molecules from the external and internal environment. More specifically, 59 (84%) of the
402 70 enriched GO terms in the module are involved in cellular responses to drugs and
403 neuromodulators, and secretion and transport of various molecules through the cell membrane
404 (Table S1). Of the 157 genes with a GO annotation in this module, 28 (18%) have GO terms
405 involved in environmental sensing and interactions. Among these, those that exhibit the largest
406 expression differences between pre-infective and infective plerocercoids code for proteins
407 including monocarboxylate transporter 7 (logFC = 4.9, FDR = 2.2e-06), solute carrier family 22
408 member 21 (logFC = 4.6, FDR = 1.2e-05), multidrug resistance protein 1A (logFC = 4.4, FDR <
409 0.001), multidrug and toxin intrusion protein 1 (logFC = 4.4, FDR = 0.0013) and neuropeptide FF
410 receptor 2 (logFC = 2.4, FDR = 3.6e-05).
411
412 The cluster described above is particularly interesting because of the high proportion of genes
413 coding for unknown proteins differentially regulated between pre-infective and infective
414 plerocercoids. One of the key features of this module is that GO annotations could be assigned to

415 only 39% of the genes; hence it is the least annotated of all the modules characterizing the
416 transition of plerocercoids from pre-infective to infective. Results from our transcriptomic
417 analysis nonetheless open the possibility to assign ecological annotations to these genes with
418 particular life-history transitions and co-expression information. The top 15 most differentiated
419 genes between pre-infective and infective plerocercoids – with logFCs of 8-11, and FDRs <
420 0.00001 – are all completely unknown (Figure 4b), and are all *S. solidus*-specific sequences, i.e.
421 we find no homology match to any known database except the *S. solidus* genome. These
422 sequence also exhibit valid open reading frames and are all highly expressed only in infective
423 worms – i.e. they are turned 'OFF' in pre-infective plerocercoids and adult worms. The only
424 information that can be used to try to assign a preliminary function to these genes is the
425 ecological annotation that stems from our transcriptomic analysis (see Materials and Methods for
426 details). These sequences were thus labelled as infective-specific and co-expressed with genes
427 involved in environmental sensing and interaction. They might in fact hold important, yet hidden,
428 functional aspects that would allow a complete understanding of the interaction between infective
429 plerocercoids and their fish host [52].
430
431 *Regulation of neural pathways could be essential for successful transmission*
432 Our findings regarding the strong signal detected for neural pathways, such as environmental
433 sensing, are further supported by another co-expression module. This strengthens the idea that
434 nervous system functions in infective plerocercoid carry an important role in the successful
435 completion of the life cycle. This module contains a total of 335 genes significantly up-regulated
436 in infective worms (Figure 4a, cluster 5), among which 235 (70%) have GO annotations. Our
437 results indicate that 41 (17%) of annotated genes in the module have GO annotations enriched in
438 activities performed by the nervous system, while 124 (53%) of them have GO annotations
439 related to transmembrane structure and activity. Biological processes associated with these genes
440 include signal transduction, synaptic transmission, sensory receptor activities and synaptic
441 exocytosis (Table S1). An interesting candidate emerges as one of the top differentiated genes in
442 the module, with an expression fold change of 3.3 (FDR = 0.0003) between pre-infective and
443 infective plerocercoids. This candidate is 5-hydroxytryptamine A1-alpha receptor, a serotonin
444 receptor. Serotonin is an important regulator of carbohydrate metabolism, host-parasite
445 communication and rhythmical movements – in conjunction with other related bioamines – in

446 several cestodes [53]. Our results show a systematic up-regulation of serotonin receptors,
447 adenylate cyclase and sodium-dependent serotonin transporters (SC6A4) specifically in infective
448 worms when compared to pre-infective and adult worms (Figure 5). Functional studies showed
449 that the signalling cascade of serotonin stimulates muscle contraction and glycogen breakdown in
450 *Fasciota hepatica* and *Schistosoma mansoni* [53]. The ultimate downstream effect of serotonin
451 signalling would be a cellular response to catabolize glycogen, the main source of energy in
452 cestodes, and more specifically in *S. solidus*. In *S. mansoni*, it has been suggested that the main
453 source of 5-HT is the host, even though some of the enzymes involved in the process and
454 recycling of 5-HT have been detected in this species [53]. This is also the case with our dataset,
455 in which we find at least one enzyme that is capable of breaking down one of the metabolites
456 required for serotonin biosynthesis, i.e. tryptophan – indoleamine 2,3-dioxygenase 2, up
457 regulated in infective worms with logFC = -3.1 and FDR = 0.01. According to the current
458 transcriptome annotation [28], there is no sequence in the transcriptome of the pre-infective,
459 infective and adult stages that is annotated as part of the biosynthetic process of serotonin. If
460 serotonin metabolism plays such a central role in the success of *S. solidus* in its fish host without
461 being synthesized by the worm itself, we could consider the possibility that it progressively uses
462 the host's supplies as it grows.
463
464 Successful completion of a complex life cycle involves intricate interactions between the
465 parasite's developmental program and physiological parameters experienced in each host.
466 Investigating the transcriptomic signature of each developmental stage has led to the discovery of
467 multiple novel yet un-annotated transcripts. These transcripts hold significant co-regulatory
468 relationships with environmental interaction genes. Future functional characterization of these
469 parasite-specific sequences promise to reveal crucial insights on how developmental and
470 infection mechanisms evolved in different parasitic taxa.
471
472 **ETHICS**
473 Fish were captured under UK Environment Agency permit and with the permission of the
474 landowner. All experiments were undertaken under a UK Home Office license (PPL80/2327)
475 held by IB, in accordance with local and national regulations and with ABS/ASAB guidelines for

476 the ethical treatment of animals in behavioral research (available online at
477 http://asab.nottingham.ac.uk/ethics/guidelines.php).
478

479 **DATA, CODE AND MATERIALS**

480 Raw datasets supporting the results of this article were deposited in a GigaDB repository made
481 publicly available [29]. All the sequencing data are available and associated with the NCBI
482 BioProject PRJNA304161. *In vitro*/*in vivo* culturing techniques and protocols are also available
483 via protocols.io [27]. Python and R codes used for data analysis are available through github
484 [30,36].
485

486 **COMPETING INTERESTS**

487 The authors declare that they have no competing interests.
488

489 **AUTHORS' CONTRIBUTIONS**

490 FOH, IB, CRL and NAH conceived the study. FOH and SG did the laboratory infections. FOH,
491 SG and IB undertook fish dissections and parasite culture. FOH extracted RNA, prepared the
492 sequencing libraries, performed bioinformatic analyses with supervision from NAH and CRL.
493 FOH, CRL and NAH drafted the manuscript with input from IB. All authors read and approved
494 the final manuscript.

495 **ACKNOWLEDGEMENTS**

496 The authors would like to thank BJ Sutherland for comments and discussion, and M. Caouette
497 and C. Tiley for precious technical work. This work was funded by a FRQ-NT grant to NAH and
498 CRL, a Natural Science and Engineering Research Council of Canada (NSERC) Discovery grant
499 to NAH, a NSERC Vanier Canada Graduate Scholarship and a Ressources Aquatiques Québec
500 (RAQ) International internship fellowship to FOH, and a UK BBSCR MITBP fellowship to SG.
501 CRL holds the Canada Research Chair in Evolutionary Cell and Systems Biology.
502

503 **REFERENCES**

504

505 1. Wilbur, H. M. 1980 Complex life cycles. *Annual review of Ecology and Systematics*
506 **11**,67-93 (doi:10.2307/2096903)

507 2. Poulin, R. 2011 *Evolutionary Ecology of Parasites* (Second edition). Princeton University
508 Press. Princeton, NJ (USA).

509 3. Auld, S. K. & Tinsley, M. C. 2014 The evolutionary ecology of complex lifecycle
510 parasites: linking phenomena with mechanisms. *Heredity* **114**, 125–132.
511 (doi:10.1038/hdy.2014.84)

512 4. Barber, I. 2013 Sticklebacks as model hosts in ecological and evolutionary parasitology.
513 *Trends in Parasitology* **29**, 556–566. (doi:10.1016/j.pt.2013.09.004)

514 5. Clarke, A. S. 1954 Studies on the life cycle of the pseudophyllidean cestode
515 *Schistocephalus solidus*. *Proceedings of the Zoological Society of London* **124**, 257–302.
516 (doi:10.1111/j.1469-7998.1954.tb07782.x)

517 6. Hammerschmidt, K. & Kurtz, J. 2007 *Schistocephalus solidus*: Establishment of
518 tapeworms in sticklebacks – fast food or fast lane? *Experimental Parasitology* **116**, 142–
519 149. (doi:10.1016/j.exppara.2006.12.013)

520 7. Tierney, J. F. & Crompton, D. W. 1992 Infectivity of plerocercoids of *Schistocephalus*
521 *solidus* (Cestoda: Ligulidae) and fecundity of the adults in an experimental definitive host,
522 Gallus gallus. *J. Parasitol.* **78**, 1049–1054.

523 8. Scharsack, J. P., Koch, K. & Hammerschmidt, K. 2007 Who is in control of the
524 stickleback immune system: interactions between *Schistocephalus solidus* and its specific
525 vertebrate host. *P R Soc B* **274**, 3151–3158. (doi:10.1002/cyto.1171)

526 9. Barber, I., Walker, P. & Svensson, P. A. 2004 Behavioural responses to simulated avian
527 predation in female three spined sticklebacks: the effect of experimental *Schistocephalus*
528 *solidus* Infections. *Behaviour* **141**, 1425–1440.

529 10. Hopkins, C. A. 1950 Studies on cestode metabolism. I. Glycogen metabolism in
530 *Schistocephalus solidus* in vivo. *Journal of Parasitology* **36**, 384–390.

531 11. Barber, I. & Scharsack, J. P. 2010 The three-spined stickleback-*Schistocephalus solidus*
532 system: an experimental model for investigating host-parasite interactions in fish.
533 *Parasitology* **137**, 411. (doi:10.1017/S0031182009991466)

534 12. SMYTH, J. D. 1950 Studies on tapeworm physiology. V. Further observations on the
535 maturation of *Schistocephalus solidus* (Diphyllobothriidae) under sterile conditions in
536 vitro. *The Journal of parasitology* **36**, 371. (doi:10.2307/3273473)

537 13. Hopkins, C. A. & Smyth, J. D. 1951 Notes on the morphology and life history of
538 *Schistocephalus solidus* (Cestoda: Diphyllobothriidae). *Parasitology* **41**, 283–291.

539 14. Schjørring, S. 2003 *Schistocephalus solidus*: a molecular test of premature gamete
540 exchange for fertilization in the intermediate host *Gasterosteus aculeatus*. *Experimental*
541 *Parasitology* **103**, 174–176. (doi:10.1016/S0014-4894(03)00092-4)

542 15. Smyth, J. D. 1952 Studies on tapeworm physiology. VI. Effect of temperature on the
543 maturation *in vitro* of *Schistocephalus solidus*. *J. Exp. Biol.* **29**, 304–309.

544 16. Körting, W. & Barrett, J. 1977 Carbohydrate catabolism in the plerocercoids of
545 *Schistocephalus solidus* (Cestoda: Pseudophyllidea). *Int. J. Parasitol.* **7**, 411–417.
546 (doi:10.1016/0020-7519(77)90067-4)

547 17. Beis, I. & Barrett, J. 1979 The contents of adenine nucleotides and glycolytic and
548 tricarboxylic acid cycle intermediates in activated and non-activated plerocercoids of
549 *Schistocephalus solidus* (*Cestoda*: *Pseudophyllidea*). *Int. J. Parasitol.* **9**, 465–468.
550 (doi:10.1016/0020-7519(79)90050-X)

551 18. Hopkins, C. A. 1952 Studies on cestode metabolism. II. The utilization of glycogen by
552 *Schistocephalus solidus in vitro*. *Experimental Parasitology* **1**, 196–213.
553 (doi:10.1016/0014-4894(52)90011-8)

554 19. Lee, D. L. 1967 The Structure and composition of the helminth cuticle. In *Advances in*
555 *Parasitology Volume 4*, pp. 187–254. Elsevier. (doi:10.1016/S0065-308X(08)60450-9)

556 20. Charles, G. H. & Orr, T. S. 1968 Comparative fine structure of outer tegument of *Ligula*
557 *intestinalis* and *Schistocephalus solidus*. *Experimental Parasitology* **22**, 137–149.

558 21. Hopkins, C. A., Law, L. M. & Threadgold, L. T. 1978 *Schistocephalus solidus*:
559 pinocytosis by the plerocercoid tegument. *Experimental Parasitology* **44**, 161–172.

560 22. Threadgold, L. T. & Hopkins, C. A. 1981 *Schistocephalus solidus* and *Ligula intestinalis*:
561 pinocytosis by the tegument. *Experimental Parasitology* **51**, 444–456.

562 23. Conradt, U. & Peters, W. 1989 Investigations on the occurrence of pinocytosis in the
563 tegument of *Schistocephalus solidus*. *Parasitol Res* **75**, 630–635.
564 (doi:10.1007/BF00930961)

565 24. Barber, I. & Svensson, P. A. 2003 Effects of experimental *Schistocephalus solidus*
566 infections on growth, morphology and sexual development of female three-spined
567 sticklebacks, *Gasterosteus aculeatus*. *Parasitology* **126**, 359–367.
568 (doi:10.1017/S0031182002002925)

569 25. Scharsack, J. P., Gossens, A., Franke, F. & Kurtz, J. 2013 Excretory products of the
570 cestode, *Schistocephalus solidus*, modulate in vitro responses of leukocytes from its
571 specific host, the three-spined stickleback (Gasterosteus aculeatus). *Fish Shellfish Immun*
572 **35**, 1779–1787. (doi:10.1016/j.fsi.2013.08.029)

573 26. Smyth, D. J. 1946 Studies on tapeworm physiology, the cultivation of *Schistocephalus*
574 *solidus* in vitro. *J. Exp. Biol.* **23**, 47–70.

575 27. Hébert, F. O., Grambauer, S., Barber, I. & Landry, C. R. 2016. Protocols for
576 'Transcriptome sequences spanning key developmental states as a resource for the study of
577 the cestode *Schistocephalus solidus*, a threespine stickleback parasite'. *protocols.io*.

578 (doi:10.17504/protocols.io.ew9bfh6)

579 28. Hébert, F. O., Grambauer, S., Barber, I., Landry, C. R. & Aubin-Horth, N. 2016
580 Transcriptome sequences spanning key developmental states as a resource for the study of
581 the cestode *Schistocephalus solidus*, a threespine stickleback parasite. *Gigascience* **5**, 24.
582 (doi:10.1186/s13742-016-0128-3)

583 29. Hébert, F. O., Grambauer, S., Barber, I., Landry, C. R. & Aubin-Horth, N. 2016. Resource
584 for 'Transcriptome sequences spanning key developmental states as a resource for the
585 study of the cestode *Schistocephalus solidus*, a threespine stickleback parasite'.
586 *GigaScience Database* (doi:10.5524/100197)

587 30. Hébert, F. O. 2016 corset_pipeline: First complete release. *Github*.
588 (doi:10.5281/zenodo.50971)

589 31. Langmead, B. & Salzberg, S. L. 2012 Fast gapped-read alignment with Bowtie 2. *Nature*
590 *Methods* **9**, 357–359. (doi:10.1038/nmeth.1923)

591 32. Davidson, N. M. & Oshlack, A. 2014 Corset: enabling differential gene expression
592 analysis for de novo assembled transcriptomes. *Genome Biol* **15**, 1–14.
593 (doi:10.1186/s13059-014-0410-6)

594 33. Lin, Y., Golovnina, K., Chen, Z.-X., Lee, H. N., Negron, Y. L. S., Sultana, H., Oliver, B.
595 & Harbison, S. T. 2016 Comparison of normalization and differential expression analyses
596 using RNA-Seq data from 726 individual Drosophila melanogaster. *BMC Genomics* **17**, 1–
597 20. (doi:10.1186/s12864-015-2353-z)

598 34. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. 2014 voom: Precision weights unlock
599 linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**, R29.
600 (doi:10.1186/gb-2014-15-2-r29)

601 35. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. 2010 edgeR: a Bioconductor package
602 for differential expression analysis of digital gene expression data. *Bioinformatics* **26**,
603 139–140. (doi:10.1093/bioinformatics/btp616)

604 36. Hébert, F. O. 2016 Bulk codes for RNA-seq analysis. *Zenodo* (doi:10.5281/zenodo.60211)

605 37. Warnes, G. R. et al. In press. gplots: Various R programming tools for plotting data. *R*
606 *package version* ..

607 38. Klopfenstein, D., Pedersen, B., Flick, P., Sato, K., Ramirez, F., Yunes, J., Mungall, C. &
608 Tang, H. 2015 GOATOOLS: Tools for Gene Ontology. *Can. J. Fish. Aquat. Sci.*
609 (doi:10.5281/zenodo.31628)

610 39. Pavey, S. A., Bernatchez, L., Aubin-Horth, N. & Landry, C. R. 2012 What is needed for
611 next-generation ecological and evolutionary genomics? *Trends Ecol Evol* **27**, 673–678.
612 (doi:10.1016/j.tree.2012.07.014)

613 40. Aly, A. S. I., Vaughan, A. M. & Kappe, S. H. I. 2009 Malaria parasite development in the
614 mosquito and infection of the mammalian host. *Annu. Rev. Microbiol.* **63**, 195–221.
615 (doi:10.1146/annurev.micro.091208.073403)

616 41. Oshima, K. et al. 2011 Dramatic transcriptional changes in an intracellular parasite enable
617 host switching between plant and insect. *Plos One* **6**, e23242.
618 (doi:10.1371/journal.pone.0023242)

619 42. Hammerschmidt, K. & Kurtz, J. 2005 Surface carbohydrate composition of a tapeworm in
620 its consecutive intermediate hosts: individual variation and fitness consequences. *Int. J.*
621 *Parasitol.* **35**, 1499–1507. (doi:10.1016/j.ijpara.2005.08.011)

622 43. Smyth, J. D. 1954 Studies on tapeworm physiology. VII. Fertilization of *Schistocephalus*
623 *solidus in vitro*. *Experimental Parasitology* **3**, 64–67.

624 44. Barrett, J. 1977 Energy metabolism and infection in helminths. in: *Parasite invasion*.
625 *Symposia of the British Society for Parasitology*. Editors A.E.R. Taylor & R. Muller,
626 **15**:121-144.

627 45. Smyth, J. D. & McManus, D. P. 2007 *The Physiology and Biochemistry of Cestodes*.
628 Cambridge University Press.

629 46. Williams, D. L., Bonilla, M., Gladyshev, V. N. & Salinas, G. 2013 Thioredoxin
630 glutathione reductase-dependent redox networks in platyhelminth parasites. *Antioxid.*
631 *Redox Signal.* **19**, 735–745. (doi:10.1089/ars.2012.4670)

632 47. Brehm, K., Spiliotis, M., Zavala-Góngora, R., Konrad, C. & Frosch, M. 2006 The
633 molecular mechanisms of larval cestode development: first steps into an unknown world.
634 *Parasitol. Int.* **55 Suppl**, S15–21. (doi:10.1016/j.parint.2005.11.003)

635 48. Schaefke, B., Emerson, J. J., Wang, T.-Y., Lu, M.-Y. J., Hsieh, L.-C. & Li, W.-H. 2013
636 Inheritance of gene expression level and selective constraints on trans- and cis-regulatory
637 changes in yeast. *Molecular Biology and Evolution* **30**, 2121–2133.
638 (doi:10.1093/molbev/mst114)

639 49. Benesh, D. P., Chubb, J. C. & Parker, G. A. 2013 Complex life cycles: why refrain from
640 growth before reproduction in the adult niche? *The American Naturalist*
641 (doi:10.5061/dryad.60699)

642 50. Hammerschmidt, K. & Kurtz, J. 2009 Ecological Immunology of a Tapeworms'
643 Interaction with its Two Consecutive Hosts. In *Advances in Parasitology*, pp. 111–137.
644 Elsevier. (doi:10.1016/S0065-308X(08)00605-2)

645 51. Adamo, S. A. 2013 Parasites: evolution's neurobiologists. *J Exp Biol* **216**, 3–10.
646 (doi:10.1242/jeb.073601)

647 52. Koziol, U., Koziol, M., Preza, M., Costábile, A., Brehm, K. & Castillo, E. 2016 *De novo*
648 discovery of neuropeptides in the genomes of parasitic flatworms using a novel

649 comparative approach. *Int. J. Parasitol.*, 1–13. (doi:10.1016/j.ijpara.2016.05.007)

650 53. Marr, J. & Muller, M. 1995 *Biochemistry and Molecular Biology of Parasites*. Academic
651 Press Limited. London (England). 337pp.

652

653

654 **FIGURES**

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

660

661 **Figure 2. Differential patterns of gene expression reveal a strong stage-specific functional**
662 **signature dominated by reproduction-associated activities in adult worms.** A) Hierarchical
663 cluster analysis showing co-expression relationships between genes labeled as significantly
664 differentially expressed between infective plerocercoids and adult worms (FDR < 0.001).
665 Biological processes significantly enriched (FDR < 0.05) in each module appear in white on the
666 heatmap. B) Volcano plot showing genes differentially expressed at three levels of FDR
667 significance. Positive values of log2FoldChange correspond to up regulated genes in adult
668 worms. Data points circled on the graph represent 18 of the top 30 most differentiated genes
669 between infective plerocercoids and adult worms, to which no annotation could be assigned. The
670 un-annotated genes are turned 'ON' only in adult worms and they are part of the redox
671 homeostasis functional module (heatmap cluster 6).

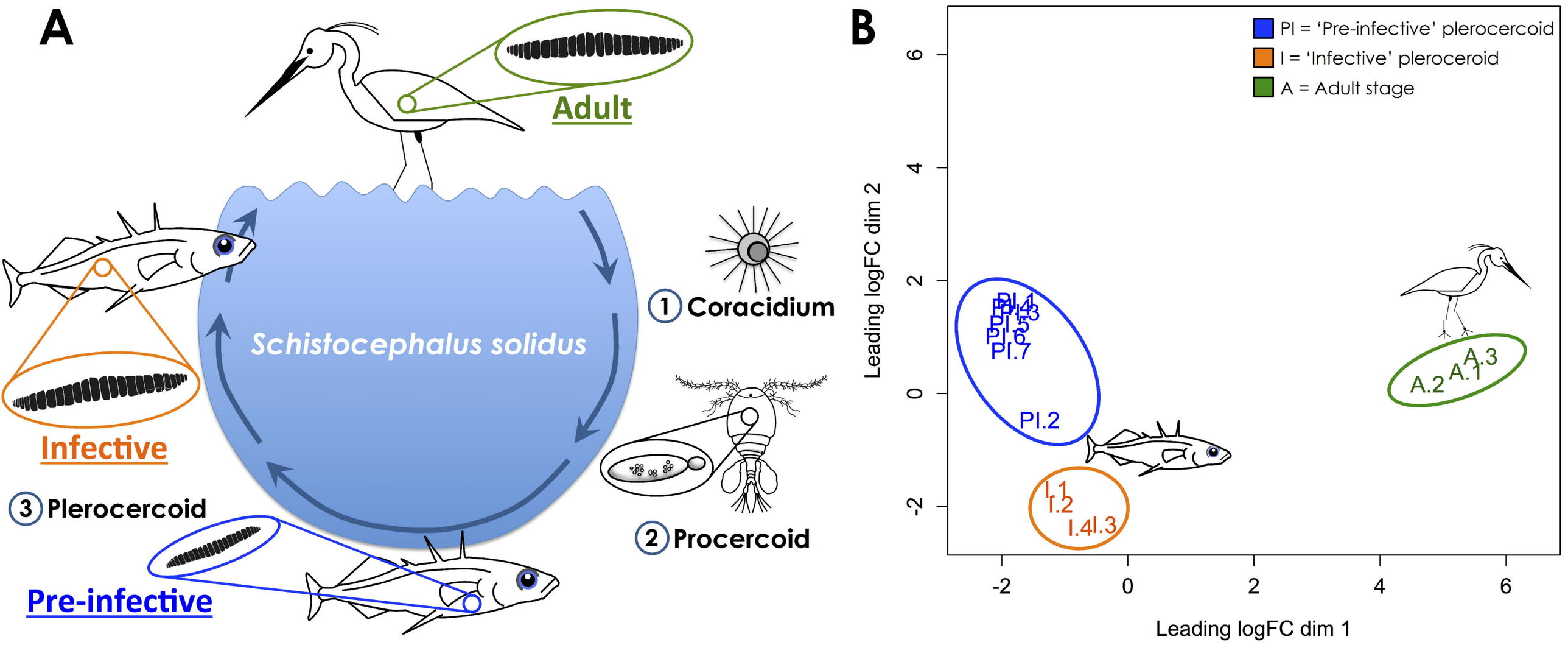
672

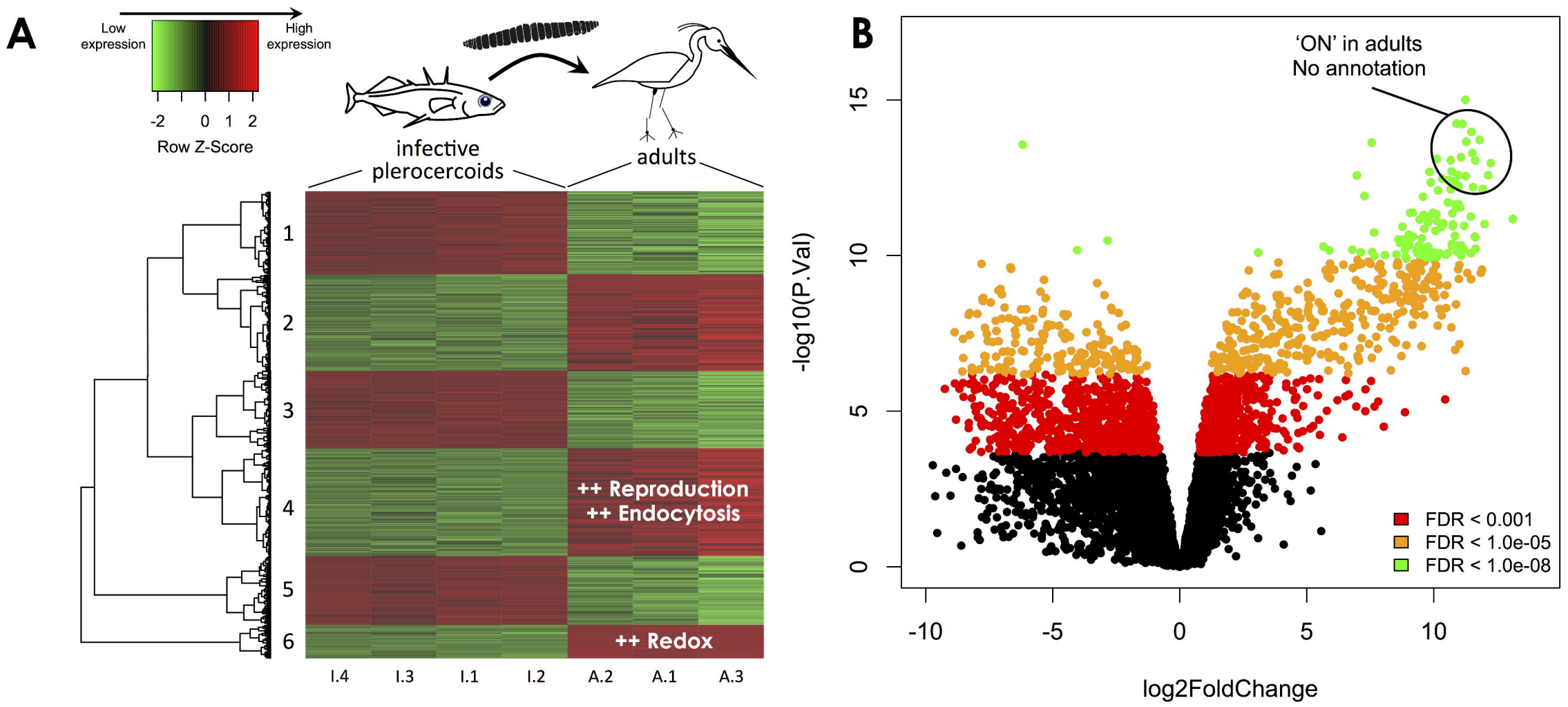
673 **Figure 3. Partial glycolysis KEGG pathway highlighting the biochemical steps for which**
674 **differential expression was detected between infective plerocercoids and adult worms.**
675 Boxes with a solid black line and white filling represent genes for which expression was detected
676 with no significant difference between developmental stages. Red boxes represent up regulated
677 genes in adult worms. Figure based on the complete KEGG pathway for
678 glycolysis/gluconeogenesis (http://kegg.jp).

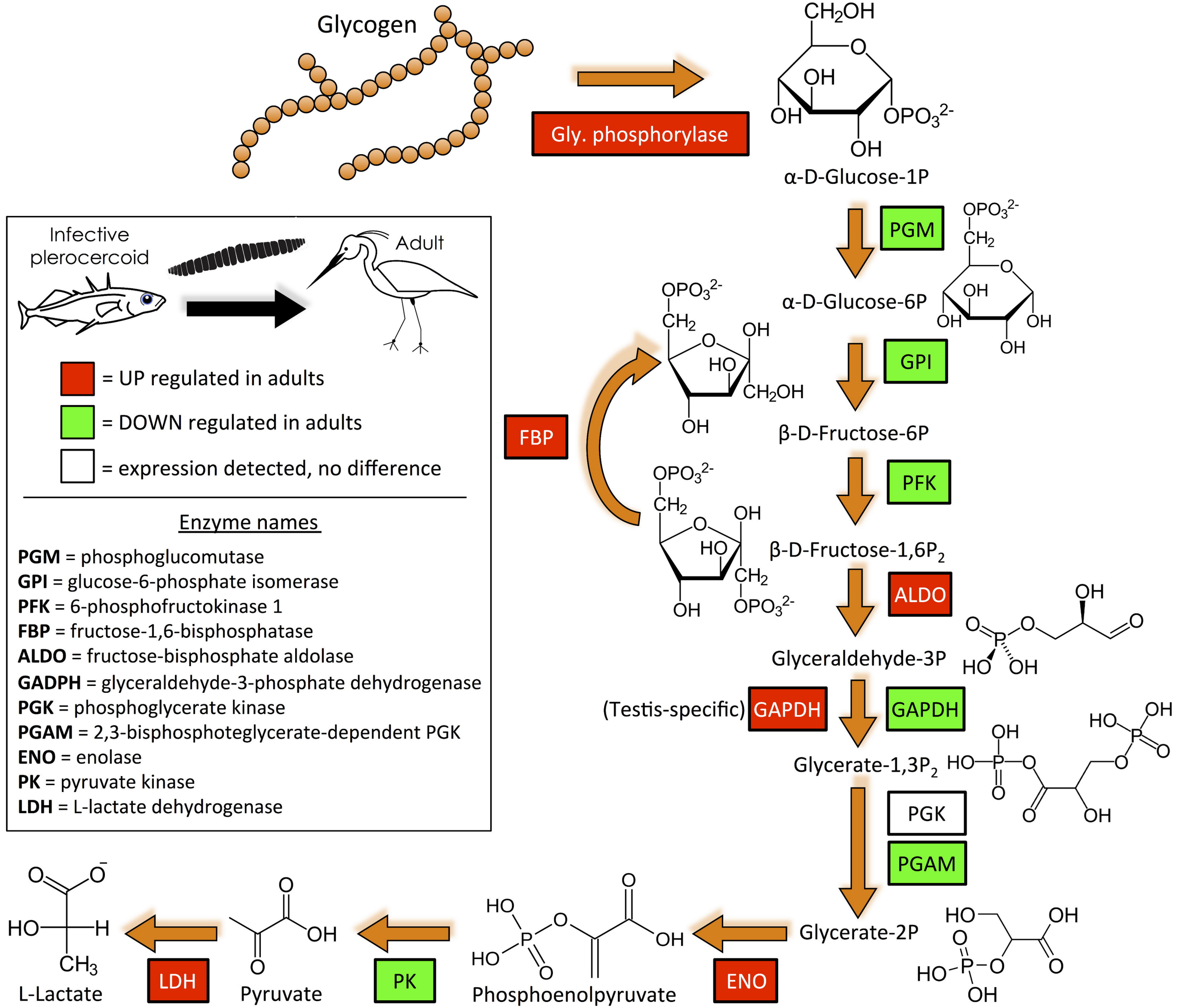
679

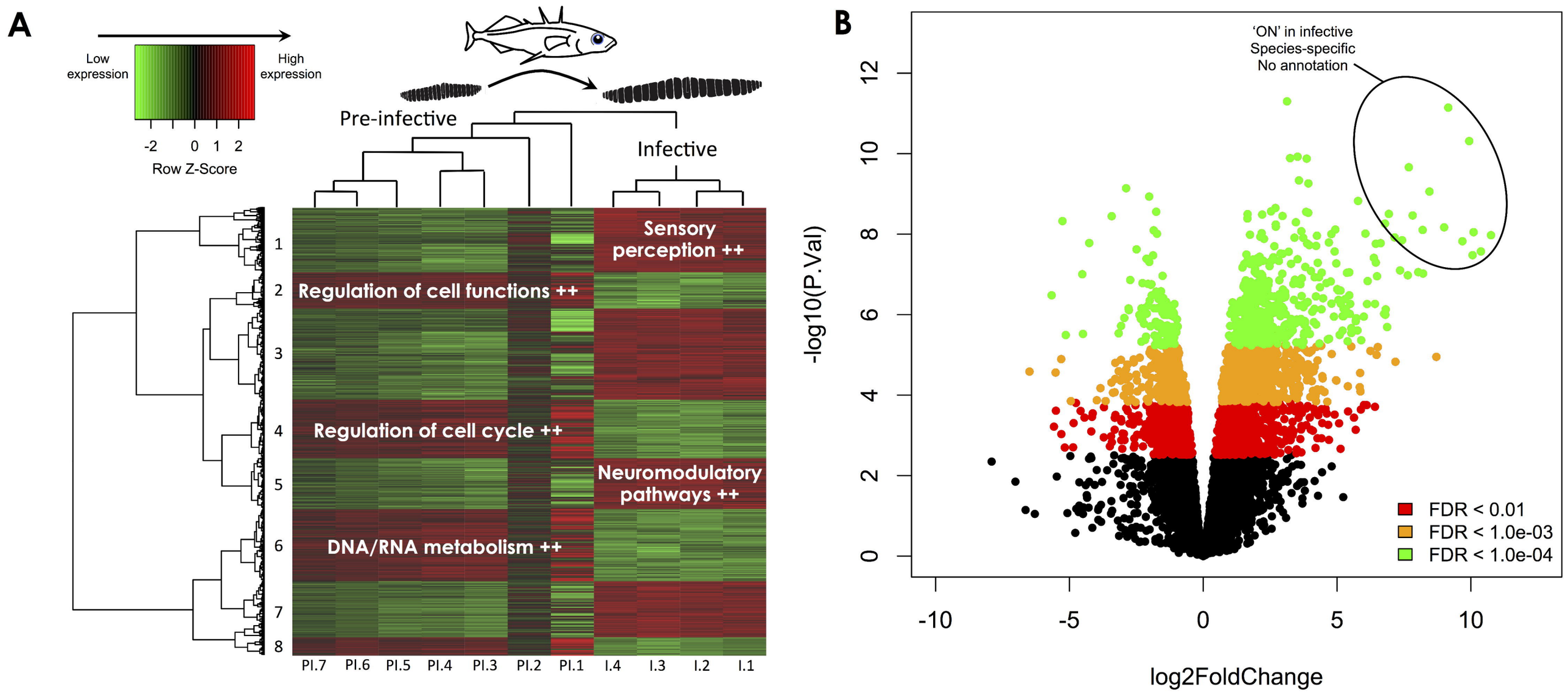
680 **Figure 4. Differential patterns of gene expression suggest a significant role for**
681 **neuromodulatory pathways in the development of infectivity towards the final host.** A)
682 Hierarchical cluster analysis showing co-expression relationships between genes labeled as
683 significantly differentially expressed between pre-infective and infective plerocercoids (FDR <
684 0.001). Biological processes significantly enriched (FDR < 0.05) in each module appear in white

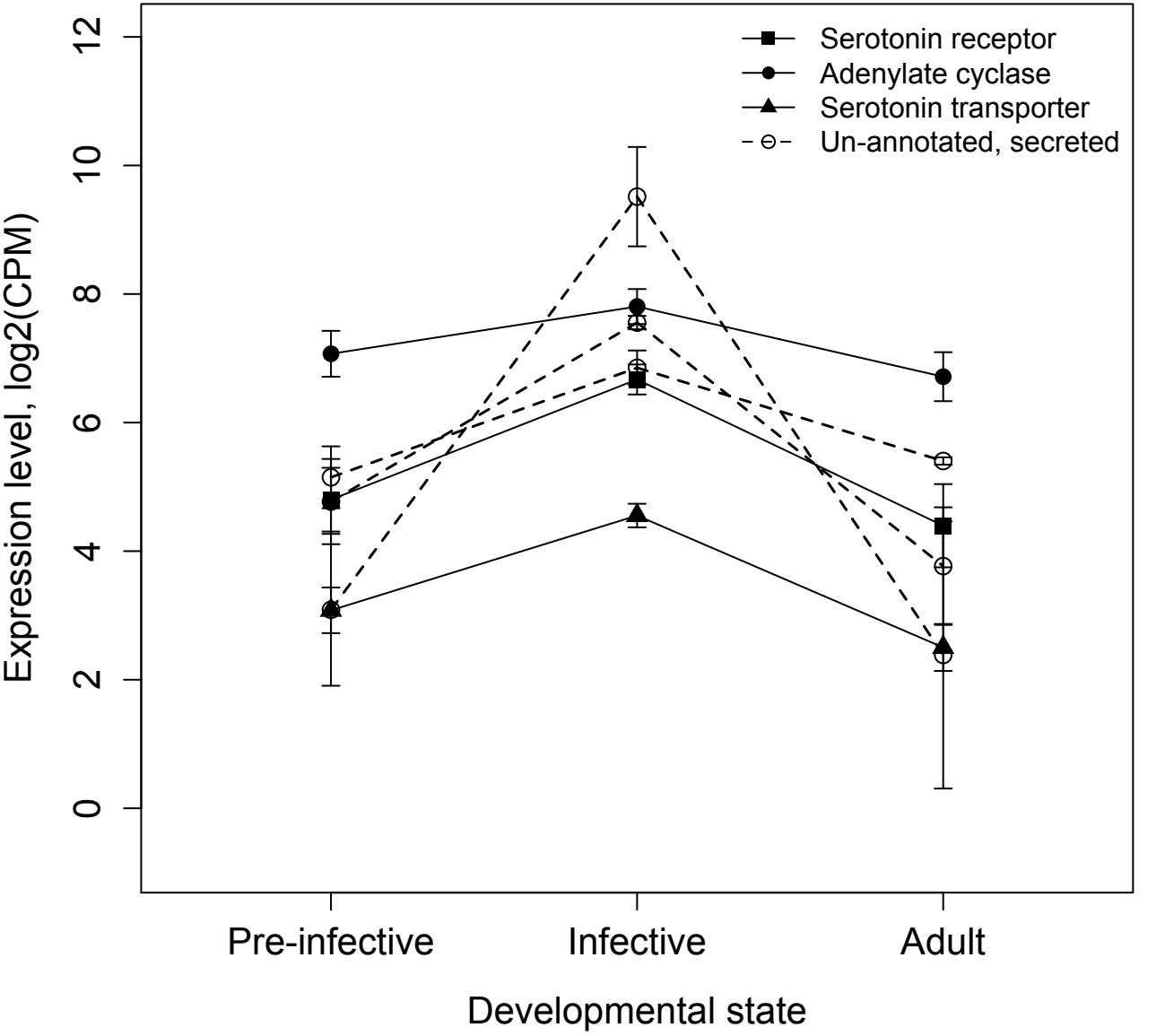
685 on the heatmap. B) Volcano plot showing genes differentially expressed at three levels of FDR
686 significance. Positive values of log2FoldChange correspond to up regulated genes in infective
687 plerocercoids. Data points circled on the graph represent the top 15 most differentiated genes
688 between pre-infective and infective plerocercoids, of which 100% are un-annotated. These
689 uncharacterised sequences are all species-specific and systematically turned 'ON' only in
690 infective worms.
691
692 **Figure 5. Gene expression data shows a systematic activation of serotonin-related genes in**
693 **the specific transcriptional signature of infectivity.** Each data point on the graph corresponds
694 to the average gene expression level (log2CPM) at a given developmental stage. Vertical bars
695 represent the 95% confidence interval of the geometric mean. Genes coding for serotonin
696 receptors (5-HT1A), adenylate cyclase (AC) and sodium-dependent serotonin transporters
697 (SC6A4) are up-regulated specifically in infective plerocercoids and down-regulated in pre-
698 infective plerocercoids and adult worms. Un-annotated genes co-expressed in the same modules
699 as serotonin-related genes and enriched in biological processes related to synaptic transmission
700 and neural pathways show very similar patterns of expression (open circles with dashed lines).
701 These un-annotated genes were labeled as 'secreted' based on the presence of a signal peptide in
702 their sequence.


A
Adult
Schistocephalus solidus
1 Coracidium
2 Procercoid
3 Plerocercoid
Infective
Pre-infective
B
PI = 'Pre-infective' plerocercoid
I = 'Infective' pleroceroid
A = Adult stage
Leading logFC dim 1
Leading logFC dim 2
-2
0
2
4
6
PI.1
PI.3
PI.5
PI.6
PI.7
PI.2
I.1
I.2
I.4
I.3
A.3
A.2
A.1


A
Low expression
High expression
-2
0
1
2
Row Z-Score
infective plerocercoids
adults
1
2
3
4
5
6
++ Reproduction
++ Endocytosis
++ Redox
I.4
I.3
I.1
I.2
A.2
A.1
A.3
B
'ON' in adults
No annotation
-log10(P.Val)
15
10
5
0
-10
-5
0
5
10
log2FoldChange
FDR < 0.001
FDR < 1.0e-05
FDR < 1.0e-08


Glycogen
Gly. phosphorylase
α-D-Glucose-1P
PGM
α-D-Glucose-6P
GPI
β-D-Fructose-6P
PFK
β-D-Fructose-1,6P2
FBP
ALDO
Glyceraldehyde-3P
(Testis-specific) GAPDH
GAPDH
Glycerate-1,3P2
PGK
PGAM
Glycerate-2P
ENO
Phosphoenolpyruvate
PK
Pyruvate
LDH
L-Lactate
Infective plerocercoid
Adult
= UP regulated in adults
= DOWN regulated in adults
= expression detected, no difference
Enzyme names
PGM = phosphoglucomutase
GPI = glucose-6-phosphate isomerase
PFK = 6-phosphofructokinase 1
FBP = fructose-1,6-bisphosphatase
ALDO = fructose-bisphosphate aldolase
GADPH = glyceraldehyde-3-phosphate dehydrogenase
PGK = phosphoglycerate kinase
PGAM = 2,3-bisphosphoteglycerate-dependent PGK
ENO = enolase
PK = pyruvate kinase
LDH = L-lactate dehydrogenase


A
Low expression
High expression
-2
0
1
2
Row Z-Score
Pre-infective
Infective
Sensory perception ++
Regulation of cell functions ++
Regulation of cell cycle ++
Neuromodulatory pathways ++
DNA/RNA metabolism ++
1
2
3
4
5
6
7
8
PI.7
PI.6
PI.5
PI.4
PI.3
PI.2
PI.1
I.4
I.3
I.2
I.1
B
'ON' in infective
Species-specific
No annotation
-log10(P.Val)
0
2
4
6
8
10
12
-10
-5
0
5
10
log2FoldChange
FDR < 0.01
FDR < 1.0e-03
FDR < 1.0e-04


Serotonin receptor
Adenylate cyclase
Serotonin transporter
Un-annotated, secreted
Expression level, log2(CPM)
0
2
4
6
8
10
12
Pre-infective
Infective
Adult
Developmental state