



# **The transcriptomic analysis of *Planorbarius corneus* hemocytes (Gastropoda) naturally infected with *Bilharziella polonica* (Schistosomatidae)**

I A Orlov, G L Ataev, Benjamin Gourbal, A S Tokmakova, A V Bobrovskaya, E E Prokhorova

## **► To cite this version:**

I A Orlov, G L Ataev, Benjamin Gourbal, A S Tokmakova, A V Bobrovskaya, et al.. The transcriptomic analysis of *Planorbarius corneus* hemocytes (Gastropoda) naturally infected with *Bilharziella polonica* (Schistosomatidae). *Developmental and Comparative Immunology*, 2023, 140, pp.104607. <10.1016/j.dci.2022.104607>. <hal-04000378>

**HAL Id: hal-04000378**

**<https://hal.science/hal-04000378v1>**

Submitted on 22 Feb 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# The transcriptomic analysis of *Planorbarius corneus* hemocytes (Gastropoda) naturally infected with *Bilharziella polonica* (Schistosomatidae)

I.A. Orlov <sup>a</sup>, G.L. Ataev <sup>a</sup>, B. Gourbal <sup>b</sup>, A.S. Tokmakova <sup>a</sup>, A.V. Bobrovskaya <sup>a</sup>, E.E. Prokhorova <sup>a,\*</sup>

<sup>a</sup>Laboratory of Experimental Zoology, Department of Zoology and Genetics, Faculty of Biology, Herzen State Pedagogical University of Russia, Russia

<sup>b</sup>IHPE, Université de Montpellier, CNRS, IFREMER, Université de Perpignan Via Domitia, Perpignan, France

## ABSTRACT

The study of molluscan innate immunity is essential for understanding the evolution of the immune system. An advance in the knowledge of their immune system can be achieved by increasing the number of model species. Our study focuses on the immunity of *Planorbarius corneus*, a pulmonate snail widely distributed in Eurasia. These snails are intermediate hosts of many trematodes, including *Bilharziella polonica* (Schistosomatidae). In this paper we obtained and analyzed transcriptomes of hemocytes of uninfected snails *Planorbarius corneus* and snails naturally infected with *Bilharziella polonica*. The transcriptomes were found to contain transcripts encoding all major groups of immune factors previously described for other gastropods. Pathogen-recognition molecules were the most diverse group of immune factors. Comparison of the transcriptomes of the infected and the uninfected molluscs showed that the expression of some genes changed during infection. Our results extend the knowledge of immune responses of pulmonate snails to trematode invasion and promote *P. corneus* as a new model for the study of molluscan defence reactions

## KEYWORDS

*Planorbarius corneus* - *Bilharziella polonica* - Hemocytes - Transcriptome - Immune reactions - Natural invasion

## 1. Introduction

*Biomphalaria glabrata* is a popular object of immunological research and the best studied gastropod species in this respect. This is mainly due to the fact that these snails are the intermediate hosts of schistosomes, including *Schistosoma mansoni*, the agent of human schistosomiasis. Most of the information about gastropod immunity has been obtained in studies made on *Biomphalaria glabrata*. Therefore, it remains unknown how conservative and universal immunity-related mechanisms and factors reported in *B. glabrata* in different gastropod species. Moreover, the experiments are usually performed on laboratory molluscs, while disease-related events occur in natural settings, which can differ significantly from the laboratory conditions. This raises a question of whether we can confidently translate the laboratory findings to wild animals.

To answer this question, one should conduct studies on molluscs taken from the nature. Comparison of data obtained on molluscs from natural habitats and molluscs of laboratory strains can provide information about the impact of biological environmental constraints on their immune reactions. For a long time, the research community has lacked established models for studies of gastropod immunity, which hampered the progress in the field. A need for new model organisms was evident, and recently the situation has to change, albeit slowly.

Increasingly more studies have been focussed on the immunity in different gastropod species, such as a transcriptomic analysis of laboratory strains of *Lymnaea stagnalis* (Seppala et al., 2021) and molluscs from natural populations of *Littorina littorea* (Gorbushin and Borisova, 2015), *Biomphalaria pfeifferi* (Buddenborg et al., 2017), and *Physella acuta* (Schultz et al., 2020). In these studies, common groups of immune reaction factors are detected in all infected (more generally speaking, immunologically challenged) snails. However, there are also substantial differences between species, which may be due to physiological features, taxonomic position and the influence of the habitat, including the impact of pathogens that an organism encounters and coevolves with. It has become fairly obvious that in order to identify the general patterns of pulmonate immune reactions and to build a model of the gastropod immune system, it is necessary to expand the range of molluscan species under study and use molluscs from both laboratory lines and natural populations.

*Planorbarius corneus* is a promising model for the study of gastropod immunity. This species, common across Eurasia, is of particular interest because of its role as an intermediate host for a number of trematodes, including *Bilharziella polonica* and some other schistosomatids. *B. polonica* causes cercarial dermatitis, sometimes called the swimmer's itch. *Bilharziella polonica* – *Planorbarius corneus* host-parasite system has already been investigated in a number of studies (Zbikowska and Marszewska, 2018) and can be considered, with some reservations, as analogous to *Schistosoma mansoni* – *Biomphalaria glabrata* pair.

In our previous studies, we have obtained valuable information about *Planorbarius corneus* hemolymph, its cell composition, haematopoiesis and immune system reactions to exposure to various immunity-inducing agents (Ataev et al., 2016). Our data indicate that the immune mechanisms of *P. corneus*, especially the cell-based immunity, are in general very similar to those of *Biomphalaria glabrata*. These processes can now be productively explored in detail with the help of well-developed methods of next generation sequencing.

In this study, we performed a transcriptomic analysis of hemocytes, the main effectors of molluscan immune system, in *Planorbarius corneus* snails naturally infected with *Bilharziella polonica*. Our aims were to obtain the first ever transcriptomic data on *Planorbarius corneus* hemocytes, to provide an improved characterization of *P. corneus* as a model organism for gastropod immunity studies and to compare the processes occurring during natural infection with those recorded in laboratory snail strains. We described the main molecular factors of *P. corneus* immunity and recorded the response of these molluscs to *Bilharziella polonica* infection, thus significantly expanding our scarce knowledge about this non-model organism.

## 2. Materials and methods

### 2.1. Molluscs, hemolymph collection and RNA extraction

We collected snails *Planorbarius corneus* (n = 150) with the shell size ranging from 23 to 27 mm from Lake Finskoe in the outskirts of St. Petersburg (60.082, 30.285) in the end of July in 2018. To define the schistosome infection status, the snails were kept separately in glasses with artificial spring water for three days, and individuals emitting cercariae were considered as infected. Only snails shedding larvae characteristic of *Bilharziella polonica* were involved in the experimental group (first confirmation of infection). *B. polonica* infection was additionally confirmed in each case by morphological examination of sporocysts during dissection (second confirmation of infection) and by rDNA PCR. In addition, we checked the snails for infection with other parasites and assessed their general state. Only *B. polonica*-infected snails and those exhibiting typical *P. corneus* characteristics were incorporated into our experiment. Fragments of rDNA were amplified by polymerase chain reaction (for the description of the PCR parameters, see Prokhorova et al. (2015)), sequenced and annotated in GenBank (MK264353.1).

Before all manipulations, the snails were thoroughly washed with distilled water and then rubbed with 70% ethanol. Hemolymph was collected from a sinus in the head area using a sterile Pasteur pipette and immediately centrifuged at 1000 g for 10 min at 4 °C to sediment the cells. The hemolymph from each snail (approx. 500 µL) was collected and processed individually, separately from the other samples. For each snail the number of cells in the hemolymph was determined using the Goryaev chamber. The snails with a very low or a very high hemocyte count were not included in the experiment. Therefore, for both groups of snails (infected and uninfected) we selected individuals with an average concentration of hemocytes in the hemolymph (400–470 cells per 1 µL). After the second confirmation of infection with *B. polonica* and the hemocyte count, we formed two samples, each consisting of 10 pooled hemocyte specimens (with an average concentration of hemocytes) from the corresponding group: uninfected (control) and infected animals. RNA was extracted from pooled samples using the Trizol reagent (Invitrogen) and its quality was controlled by capillary electrophoresis on the Agilent Bioanalyzer.

### 2.2. Library preparation, sequencing and read preprocessing

mRNA was enriched and converted to cDNA by RT with polyT primers (Mint-2 kit, Evrogen). Sequencing library with mean fragment size of approximately 450 bp was prepared using QIAGEN QIAseq FX DNA Library Kit, 100bp sequencing in paired-end mode performed on Illumina HiSeq 2500 system. At first, we removed from the obtained reads library-specific adapters and polyA tails, trimmed N's and bases with phred score <5 from both ends, unconditionally clipped 8 bases from 5' and 5 bases from 3' as we noticed a bias in nucleotide distribution there and discarded all reads shorter than 20 bp using cutadapt (Marcel, 2011). We mapped reads resulting from the previous step to chimeric rRNA-coding cluster with hisat2 (Kim et al., 2015) using standard parameters and discarded all the aligned reads (the entire pair was excluded even if only one read aligned). For this purpose, we constructed the rRNA-coding cluster mentioned from the *Biomphalaria glabrata* sequence replacing ITS1, ITS2 with the *Planorbarius corneus* sequences previously obtained in our group (Prokhorova et al., 2015).

### 2.3. Contamination clearing

All the pre-processed reads were classified with kraken2 software (Wood et al., 2019) using NCBI NT database (<https://www.ncbi.nlm.nih.gov/nucleotide/>) with confidence parameter equal to 0.6. We used BBduk (<https://sourceforge.net/projects/bbmap/>) to filter out reads containing 27-mers from specially constructed database. This database consisted of sequences (genomic, mtDNA, rRNA) from the observed contaminating organisms or species closely related to them (supplementary file 1).

### 2.4. De-novo transcriptome assembly and annotation

Reads from both samples (control and infected) were merged and used for a *de novo* assembly of a reference transcriptome using Trinity (Grabherr et al., 2011). Assembly quality evaluation incorporated assessment of technical metrics such as ones obtained by TransRate (Smith-Unna et al., 2016) and biologically-relevant number of universal single-copy orthologs (USCOs) detected by BUSCO (Seppey et al., 2019; used odb10). Transcripts longer than 200 bp were clustered based on 95% sequence identity using CD-hitEST (Fu et al., 2012). ORFs were predicted and extracted with TransDecoder (<https://github.com/TransDecoder/TransDecoder/wiki>). Encoded protein sequences were then clustered based on 99% identity and subsequently annotated with BlastP, e-value threshold equal to  $1e-5$ , against NCBI NR database (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/>, accessed on August 02, 2021) and InterProScan, with default settings, utilized to predict functional domains. In this annotation we searched for proteins containing immune-related domains or having similarities to previously identified polypeptides with immune function with the Python script. To perform a pathway-based functional annotation we exploited the KEGG database with its related tool BlastKOALA on a genus level. Assigned K numbers were mapped to KEGG modules using KEGG Mapper – Reconstruct tool (<https://www.genome.jp/kegg/mapper/reconstruct.html>).

### 2.5. Differential expression analysis

To calculate read counts of each transcript we utilized Salmon (Patro et al., 2017). Obtained values alongside the offset matrix were imported into EdgeR (Robinson et al., 2010) package for R with the help of another package tximport (Soneson et al., 2015). To conduct differential gene expression analysis with likelihood ratio tests (glmLRT), a conservative value of biological coefficient of variation BCV = 0.4 was specified as suggested in the package manual (<https://bioconductor.org/packages/release/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf>). Transcripts satisfying the following conditions were considered differentially expressed (DE):  $|\log FC|$  (logarithm of fold-change)  $> 2$  and FDR (false discovery rate)  $< 0.05$ . Hemocytes from uninfected animals were considered as a reference for comparison, e.g., upregulated DEGs have a higher expression in the infected group compared to the uninfected molluscs. For DE transcripts we performed annotation with BlastX against NCBI NR database (e-value threshold equal to  $1e-5$ ).

### 2.6. Other methods

Alignment of identified FBG-containing proteins from *P. corneus* transcriptome to *B. glabrata* FRePs and calculation of corresponding percent identity matrices were performed with Clustal Omega (Sievers et al., 2011). Alignments were manually inspected using MEGA 7 software (Kumar et al., 2016).

## 3. Results and discussion

*Planorbarius corneus* is a promising alternative model for studying the immune response of pulmonates infected in nature or in the laboratory- controlled environment. As confirmed in numerous studies, the main mechanisms of the immune response in *P. corneus* do not differ from those in *Biomphalaria* spp. In particular, the main effector cells of the immune reactions in both these snails are circulating hemolymph cells, the hemocytes (Prokhorova et al., 2015, 2018). Trematode invasion is known to affect the ratio of hemocyte types (Ataev et al., 2016). Hemocytes are involved in cellular reactions such as phagocytosis of bacteria and encapsulation of metazoan parasites, especially sporocysts and metacercariae of trematodes. At the same time, they produce humoral immune factors which mark and opsonize any alien organism.

Gene expression analysis is a key tool for detailed functional activity studies of *Planorbarius corneus* hemolymph cells. We have previously analyzed the expression of several lectins in the hemocytes of infected and uninfected molluscs (Prokhorova et al., 2010) and observed the influence of infection on the expression of some genes in hemocytes and hepatopancreas cells. However, to obtain a more comprehensive picture of the role of hemocytes in the immune challenge and to compare our results with those obtained on other species, a transcriptomic analysis is indispensable. In this work, we sequenced and analyzed the transcriptome of *P. corneus* hemocytes for the first time.

To study the transcriptome of hemocytes of *P. corneus* from natural populations under conditions of the background level of immunity challenge and under conditions of trematode invasion, we collected molluscs and verified their infection status. Infection with *B. polonica* was ascertained by an experienced morphologist, who examined shed cercariae and intramolluscan sporocysts. These two stages of *Bilharziella polonica* are clearly distinguishable from those of other species on the basis of morphological characters (Combes et al., 1980). *B. polonica* usually starts to infect the snails actively in late May-early June, and we collected the snails in the end of July. Shedding cercariae was a necessary condition for the inclusion of the snail into the infected group. Therefore, we may be sure that all molluscs in the infected group had a well-established infection.

### 3.1. Pre-processing, decontamination and assembly

Raw and pre-processed library sizes are given in Table 1. All the steps of read pre-processing and transcriptome assembly were performed using merged reads, unless indicated otherwise. After the first technical step, we additionally filtered out probable rRNA contamination, as it can be present even in polyA-enriched RNA preparations. For this purpose, we constructed chimeric rRNA-coding cluster from rDNA cluster of *Biomphalaria glabrata* with ITS replaced by those of *Planorbarius corneus* (Prokhorova et al., 2015). We found that approximately 4% of pre-processed reads had rRNA origin, which is a modest value.

As already noted, before the sequencing experiment all the molluscs were dissected and scrupulously examined for the presence of visually detectable parasites and commensals. Nevertheless, since we worked with animals from a natural population, we additionally checked the experimental material for any possible contamination with the help of a bioinformatics method. For this, we separately classified pre-processed reads from both samples by kraken2 (Wood et al., 2019) using NCBI NT database. As expected, most of them were classified as having molluscan origin. Schistosomal reads (about 50% of all contaminating sequences in the sample from infected animals) could be detected only in the library from infected snails. Traces of bacterial and plant sequences were present

in both datasets. While the presence of parasite DNA in the infected sample could reflect the phagocytic activity of hemocytes, we also observed a significant and an almost equal portion of Annelida-originating sequences (approximately 20% of all contaminating sequences) in the infected and the control specimens. Annelid worms are well known to inhabit the mantle cavity of freshwater molluscs (Callisto et al., 2005). Therefore, we decided that the contamination occurred during hemolymph collection. Proportions of non-molluscan reads are shown in Fig. 1.

To eliminate contamination by schistosomes and annelids, we initially mapped the analyzed reads to 3 schistosome (*Schistosoma mansoni*, *S. japonicum*, *S. haematobium*) and 3 annelid (*Eisenia fetida*, *Helobdella robusta*, *Hirudo medicinalis*) genomes, discarding all the aligned sequences. Subsequent Kraken2 classification still showed a substantial level of contamination, even after modification of default parameters of splice-aware hisat2 (Kim et al., 2015). To maximally filter out reads from contaminating species we utilized kmer based approach applying BBDuk and custom sequence database. This database included genomic and/or most common sequences (mtDNA, rRNA) from contaminating organisms themselves (defined by Kraken2) or from species evolutionarily close to them in case the data for the original species were lacking. The resulting library consisted of 44744691 paired reads of 88bp. Due to a long evolutionary coexistence and a relatively small phylogenetic distance between gastropods and schistosomes and, especially, between gastropods and annelids, a substantial similarity of some of their genomic sequences may be expected. A double-edged filtration method relying on k-mer occurrence inevitably leads to the loss of some information due to the presence of evolutionary conserved sequences. Nevertheless, we chose this method, and minimized the number of constructed chimeric or non-molluscan transcripts by sacrificing a portion of “true” data.

From the plethora of available transcriptome assemblers, we tried Trinity (Grabherr et al., 2011), rnaSPAdes and Oases (Schulz et al., 2012). Unexpectedly, obsolete and non-supported Oases worked best for our data in terms of some technical metrics (average and median length) though not in terms of the fraction of mapped back reads (Table 2). Taking into account the modest amount of sequencing data, we finally chose Trinity-generated assembly as produced by the most conservative and reliable assembler, which is additionally supported by the highest number of mapped back reads.

Basic assembly parameters after 95% identity clusterization of transcripts can be found in Table 2. Our assembly included more than 90% of eukaryotic and metazoan (Table 2) universal single copy orthologs (USCOs) according to BUSCO (Seppey et al., 2019). The fact that there were only 55% molluscan-specific USCOs could be explained by the design of our experiment where we sequenced only one specialised tissue which is unlikely to express all the organism-level USCOs. However, the low proportion of molluscan-specific USCOs could also indicate that the assembly was incomplete. We detected a high duplication rate of USCOs despite 95% sequence identity clusterization, which is rather common for transcriptomic assemblies.

There were 60630 ORFs the assembled transcripts (28744 complete and the others partial or internal), which resulted in 53611 polypeptide products after 99% protein sequence identity clusterization.

### 3.2. Transcriptome annotation

Transcripts corresponding to all the main groups of immunity factors previously described in other gastropods were identified in hemocytes of *Planorbarius corneus* (Table 3). We detected transcripts encoding more than 160 different immune-relevant proteins from various functional classes, such as molecules involved in pathogen recognition, adhesion and intercellular interactions, immune effectors and immune regulators.

Pattern recognition receptors (PRRs) were one of the best represented groups in our data. The entire diversity of PRRs was identified. In particular, we found transcripts encoding ten members of Toll-like receptors (TLR) family and five members of scavenger receptors (SR) family. The diversity of TLRs defines their role as the main receptors recognizing pathogen-associated molecular patterns (PAMPs) in the theory of innate immunity. In addition, the presence of a certain set of TLR genes is considered as a possible mechanism of specific compatibility in the trematode-mollusc system (Pila et al., 2016).

In the widely studied snail species *Biomphalaria glabrata*, 56 genes encoding TLR of two different types (V-TLR and P-TLR) were identified (Adema et al., 2017). Analysis of hemocyte transcriptomes showed a higher level of constitutive TLR (BgTLR) gene activity in *B. glabrata* molluscs resistant to infection with *Schistosoma mansoni* (BS-90 line) compared with susceptible ones (M line). *S. mansoni* invasion leads to a 27-fold increase in the expression of this gene in the hemocytes of resistant molluscs, while the expression of this gene in the hemocytes of susceptible molluscs does not change (Pila et al., 2016). In our work, we did not detect any differences in TLR expression between infected and uninfected snails. It is important to note that the molluscs in our study were taken from nature, and so we do not know their genetic characteristics such as the degree of susceptibility to trematode invasion. However, we studied molluscs with mature infection, which indicates their susceptibility.

Transcripts corresponding to other PRR groups, as well as adhesion molecules, were also abundantly represented in the transcriptome of *Planorbarius corneus* by scavenger receptors, lectins of different groups, and adhesion molecules, including integrins, selectins, dermatopontins, etc. (Table 4). Besides, transcripts of all major groups of cytotoxicity factors of innate immunity cells were identified in hemocytes: antimicrobial proteins, homologues of complement system components, enzymes involved in the production of reactive oxygen and nitrogen species, enzymes of the phenoloxidase system, etc. These results confirm that hemocytes are universal cells of the innate immunity system of molluscs. Similarly to TLRs, there were no significant differences in expression between infected and uninfected molluscs for these transcripts. Although there is evidence that these factors are upregulated during experimental immunization of molluscs (Wang et al., 2018), the increase in the expression has a clear temporal dynamic (Chen et al., 2022). The main results regarding differential gene expression analysis are discussed in the corresponding section below.

The presence of PRRs and other molecules involved in innate immunity reactions is the basic characteristic of an immunocyte (an immunocompetent cell). The immune molecules mediating defense reactions have to be present in immunocytes in the amounts sufficient for an immediate execution of their function. An increase in the amount of immune factors in circulation is not necessarily due only to an increase in the number of transcripts in hemocytes. Their concentration in the hemolymph may be also elevated as a result of the increase in the number of hemocytes producing them during invasion.

We performed a large-scale functional annotation to reveal metabolic pathways whose elements are encoded by the assembled transcriptome with the help of KEGG tool BlastKOALA. We found 77 modules (metabolic pathways including specialised molecular complexes) that were complete and 33 modules that lacked one only block. Among the fully represented pathways there were some of the carbohydrate and energy metabolism including glycolysis, TCA cycle, pentose phosphate pathway and oxidative phosphorylation. Triacylglycerol, phosphatidylcholine, ceramide, sphingosine and fatty acid biosynthetic reactions enzymes as well as enzymes catalysing beta-oxidation of the latter were observed. There were also all the components of adenine, guanine and pyrimidine ribonucleotides synthesis pathways and complete sets of elements mediating a number of amino acid metabolism reactions including their synthesis, interconversion and degradation.

### 3.3. Proteins with immune-related domains

According to the InterProScan, we found 5385 immunity-related domains in the transcriptome-encoded proteome and 4405 unique proteins were predicted (Table 4). Slightly less than 30% of them were proteins with IgSF or different lectin-containing domains.

We did not detect any sequences bearing IgSF and fibrinogen domains simultaneously. This feature is a signature of fibrinogen-related proteins (FREPs), which are well-known and well-characterized effectors of immunity in pulmonate molluscs, especially *Biomphalaria glabrata* (Adema and Loker, 2015). Indeed, not all molluscan species rely on FREPs. Numerous other molluscs elaborated diversified lectins containing a single FBG domain named FREds (Adema and Loker, 2015). On the other hand, we have previously observed the expression of FREP-like proteins in this species under the burden of trematode invasion (Prokhorova et al., 2010). Taking into account both these facts, it is reasonable to suggest that while *Planorbarius corneus* does possess some classical FREPs in its arsenal, their amount might be lower than in case of the well-studied *Biomphalaria glabrata*. The main defence factors in *Planorbarius corneus* are probably FREds and other types of lectins, which are present in abundance. In particular, we found three polypeptides incorporating C-type lectin or fucoselectin domain alongside the motif with an immunoglobulin-like fold, which makes it possible to characterize them as CREPs (Gorbushin, 2019).

Taking into account the potential importance of FREPs for snail immunity, we took a closer look on 176 polypeptides with annotated FBG domain found in our transcriptomic data. Out of these proteins, 43 originated from complete, 37 from internal and 96 from partial predicted ORFs. Approximately in 15 percent of polypeptides with FBG domain it was adjoined towards the N-terminus with epidermal growth factor (EGF) or Ca-binding EGF-like domains. Another common neighbour was laminin domain. Some of these proteins included a motif indicating a possibility for aspartate or asparagine posttranslational modification by hydroxylation, which is often found in domains homologous to EGF. Other less frequently observed options were growth factor receptor domain, PAN/Apple domain, Kringle-like motif, and carbohydrate binding signatures such as galactose-binding and WSC domains. A couple of domains characteristic of complement system proteins, such as Sushi and Complement control module/SCR domain, were observed in conjunction with FBG domain as well.

To check if any of FBG-containing proteins appear to be similar to already published FREPs we downloaded previously submitted protein sequences of *B. glabrata* FREPs from NCBI protein database (174 entries, supplementary file 2). In order to keep only molecules with both domains we annotated protein domains in the downloaded dataset with IPS *de novo*. After 99% identity clusterization we aligned each of our 176 predicted FBG-containing proteins to the chosen set of *B. glabrata* FREPs using Clustal omega, calculated percent identity matrices (Sievers et al., 2011). The same approach was utilized but using only IgSF domains from this set of FREPs.

After manual curation of calculated percentage identity matrices and the obtained multiple sequence alignments we inferred that the absolute majority of sequences shared a high similarity with *B. glabrata* FREPs only in FBG domain. We could find two notable sequences, which were isoforms of one Trinity gene (not to be confused with a gene in its general meaning), originating from predicted complete ORFs. Each of two contained a signal peptide at the N-terminus, and showed high similarity to IgSF domains of 13.1 and several K-type fibrinogen-related proteins of *B. glabrata* (alignments are in supplementary file 3 and 4).

Though FREPs are described in many molluscs, most of the characterized FREPs belong to one molluscan species, *Biomphalaria glabrata*. In a close species, *B. pfeifferi*, 249 sequences fitting the description of FREPs have been revealed. Twelve of them are full-sized homologues of *B. glabrata* FREPs (Buddenborg et al., 2017). Infection of *B. pfeifferi* with *Schistosoma mansoni* miracidia significantly changes the expression profile of FREP: expression of FREP2 increases gradually, that of FREP7 and 14 increases significantly on the 3rd day post infection. At the same time, the expression of FREP3, which strongly reacts to schistosomal infection in *Biomphalaria glabrata*, increases in *B. pfeifferi* slightly only in the beginning of the invasion. Conversely, FREP5 is expressed much more strongly in *B. pfeifferi* than in *B. glabrata* (Buddenborg et al., 2017). Thus, the change in FREP expression has a clear temporal dynamic.

The survey of FREPs in other gastropods has not shown such a broad variety of these proteins as in *Biomphalaria* spp. Only four FREPs genes have been identified in the genome of *Physella acuta* (Schultz et al., 2020). Infection with *Echinostoma paraense* trematodes and bacteria does not upregulate FREP expression but some other lectins respond positively to immunization. Similar results have been obtained for *Lymnaea stagnalis*, in which only two FREPs were detected (Seppala et al., 2021). One protein homologous to FREP3 was identified in *Helisoma trivolvis* (Zhang and Loker, 2004). Recently, genes encoding FREPs have been identified in deep-sea gastropods living in thermal springs: *Gigantopelta aegis* (18 genes) and *Chrysomallon squamiferum* (14 genes). The maximum expression of FREPs genes in these snails was shown not in hemocytes but in barrier structures: ctenidia and the mantle (Lan et al., 2021). Five genes encoding FREPs were identified in the genome of *Littorina littorea*, and two FREP genes were identified in the gastropod mollusc *Aplysia californica*. However, their structure differs significantly from the *Biomphalaria* FREPs (Gorbushin and Borisova, 2015).

Thus, though FREPs are common in molluscs, their diversity in *B. glabrata* exceeds that in any other species. It has been suggested that the diversity of FREPs in *Biomphalaria* could have been formed as a result of a long-term coevolution with *Schistosoma mansoni* (Bayne, 2009; Mitta et al., 2012, 2017; Adema and Loker, 2015). The parasite's attempts to "deceive" the immune system of the host with the help of an increased polymorphism and mimicry led to the development of mechanisms ensuring more effective recognition and destruction of the parasite by the immune system of the mollusc. Production of FREPs was one of these mechanisms. In other molluscan species interacting with other trematode species or other pathogens, the mechanisms of their recognition and elimination could be improved with the help of different PRRs and other components of immunity.

Thus, two explanations of the absence of FREPs in the assembled transcriptome of *Planorbarius corneus* can be offered. The first is that the mollusc can utilize other molecules for neutralization of the parasite. The second explanation is a constitutively low level of FREPs expression, which made them impossible to detect in our experiment. In order to ascertain whether FREPs are absent or present in *P. corneus*, it is necessary to analyze its genome and to study its transcriptome at different periods of infection.

### 3.4. Differentially expressed genes

With the help of EdgeR software (Robinson et al., 2010) we identified 189 differentially expressed genes (DEGs) under the condition of *B. polonica* invasion: 155 were upregulated and 34 were downregulated. After manual curation 60 and 11 DEGs were found to have sensible Blast annotation (not uncharacterized, putative unknown proteins etc.) for upregulated and downregulated groups, respectively (see supplementary file 5 for the full list of DEGs with sensible Blast annotation). It is important to note that we also included in the analysis sequences present only in one of the conditions as it seems to be highly probable that some of the genes are completely inactive in uninfected animals or, on the opposite, strongly inhibited by the parasite in the infected animals. Since there were no experimental replicates, we used the value of 0.4 for biological coefficient of variation parameter during differential expression analysis. This value was considered conservative enough to minimize expected false-positive results as advised by EdgeR manual. Keeping in mind all the details, we consider our analysis as qualitative but by no means quantitative.

Molecules of viral origin were detected in both down- and upregulated groups of transcripts (Table 5). Two downregulated transcripts were related to retrotransposon sequences coding endonuclease-reverse transcriptase and pol polyprotein. Two more transcripts, one with an increased presence and one with a decreased presence, resembled the replication-associated protein of CRESS ssDNA viruses infecting snails and insects. The level of one more transcript, encoding

snail nidovirus orf1ab polyprotein, showed an increase. It has been noted that retrotransposons are somehow involved into the relationship between trematodes and molluscs. It has even been suggested that they play a crucial role in the outcome of parasitic invasion or in the possible horizontal gene transfer between the host and the parasite (Smith et al., 2021). Our data obtained on *P. corneus* supports the idea that retrotransposons are involved in trematode-snail interaction, though they do not clarify the nature of this involvement.

Sixteen upregulated DEGs showed a high similarity to molluscan proteins or proteins from closely related species. There seem to be among them collagen alpha and cathepsin B-like polypeptide, which are known to be the effectors of anti-trematode snail immunity (Lockyer et al., 2008; Myers et al., 2008; Tetreau et al., 2017). The former is suggested to participate in hemolymph coagulation and sporocyst encapsulation while the latter, being an endosomal cysteine protease, recycles heterogeneous material acquired by the cell.

Three transcripts encoding myosin heavy chain-like proteins (II- and 6/7-like) had an increased expression in the infected animals. The times when myosin was considered a muscle-specific polypeptide are long gone and it is now generally accepted that members of the large myosin superfamily can be present in almost every compartment of various cells, mediating the processes of cell adhesion, migration, endocytosis, phagocytosis, chromatin remodelling and others (Vicente-Manzanares et al., 2009). Under conditions of endogenous parasite development, hemocytes perform encapsulation and phagocytosis, and undergo proliferation, accompanied with a change in gene expression patterns, probably exploiting myosin molecules at each of these steps. The role of myosin in immunity is still obscure, as is that of many other immune challenge-altered factors. There is evidence that myosin II heavy chain protein was less abundant in sporocyst-encapsulating hemocytes from the resistant *Biomphalaria glabrata* line compared to non-encapsulating cells (Dinguiard et al., 2018). It has also been reported that two myosin II heavy chain genes were expressed to a lesser degree in the hemocytes of susceptible molluscs exposed to miracidia compared to the hemocytes of the resistant strain (Lockyer et al., 2012). Upregulation of transcripts for translation elongation factor 1- $\alpha$  and polyubiquitin-B in the infected group can be considered as a sign of more active protein synthesis and degradation, which is to be expected in hemocytes under immunological challenge. An increased level of the same elongation factor-coding sequence has been described in hemocytes of trematode-resistant *B. glabrata* strain (Lockyer et al., 2008).

Different toxic proteins like **biomphalysins** are exploited by the mollusc during infection to fight the parasite. Though we did not detect any increase in the expression of transcripts encoding biomphalysin, we found another potentially toxic protein resembling the so-called turriptide. Turriptide neurotoxins discovered in the marine snail from the superfamily Conoidea can inhibit acetylcholine receptors, harming the parasite in this way (Hernandez-Samano et al., 2020). Moreover, the turriptide-like protein found in our study contains kazal-type domain according to IPS, thus probably possessing serine protease inhibiting activity. Serine protease inhibitors are well established effectors of snail immunity helping to block the digestive enzymes of the invaders. Another member of the turriptide-like family has been reported to be overexpressed in non-resistant snails compared to the resistant line (Lu et al., 2022). The authors suggest that it is a susceptibility factor, which mediates tolerance to infection or perhaps contributes to animal survival. The same paper attributes a similar, tolerance-mediating role to temptin, a pheromone driving mating behaviour, which had an increased expression in hemocytes of infected molluscs in our research (Lu et al., 2022). At the same time, temptin expression was upregulated in *Mytilus galloprovincialis* hemocytes stimulated with lipopolysaccharides and in *Biomphalaria pfeifferi* 3 days after infection with *Schistosoma mansoni* (Buddenborg et al., 2017; Moreira et al., 2020).

**Apolipophorins** binding various lipids demonstrate immune-related properties in insects. Different members of this family are involved in pathogen pattern recognition, toxin detoxification and antimicrobial defence. The role of this family in molluscs is far from being understood but it has been shown that apolipophorin level is decreased in hemocytes of the *Biomphalaria glabrata* strain susceptible to trematode invasion (Dinguiard et al., 2018). In our study we reported an increase in the expression of apolipophorin-like protein-coding transcript, which is an evidence that it has some immunity-related functions in snails as well as in arthropods.

The secreted vertebrate protein **brorin** acts as the antagonist of BMP activity, thus participating in neurogenesis and neural tissue development. We observed an upregulation of brorin-like protein gene expression in hemocytes of infected *Planorbarius corneus*. A vWFC related protein (VREP) similar to brorin has been found to be abundant in coelomic fluid of starfish *Asterias rubens* several hours after an induced blood loss (Shabelnikov et al., 2019). The brorin-like protein detected in *Planorbarius corneus* might also mediate tissue repair processes aimed at minimizing the consequences of the ongoing infestation, similarly to what has been reported for the starfish.

Despite meticulous data pre-processing after the final assembly, we found in the hemocytes of the infected molluscs some transcripts encoding proteins with a very high similarity to schistosomal ones. We suggest that these sequences have a phagocytotic origin (see also above). It is likely that certain proteins of different trematodes share a high similarity, which is accompanied by a substantial nucleotide sequence divergence of the transcripts encoding them. This seems to be the reason behind such a high proportion of schistosome-related polypeptides in our assembled transcriptome even after a thorough read filtration. In principle, we could also apply a translational-based procedure during pre-processing, e.g. BlastX or DIAMOND. However, this approach would result in lots of false hits mainly due to conserved protein domains. For this reason, we eventually decided against using method and simply manually discarded schistosome proteins that could still be detected after the previous pre-processing steps.

Interestingly, we found a protein from Hsp70 family among other trematode housekeeping polypeptides such as major egg antigen, tubulin or COX III subunit. These proteins are known to localize in cercarial acetabular glands and play a role during the infection of the definitive host. Notably, Hsp70 with molluscan descent according to blast was found in the upregulated group as well. This protein in snail is related to the outcome of miracidium invasion and the course of the disease, but its role is yet unclear and probably depends on the stage of infection, the age of the mollusc, its genetic background and other factors (Ittiprasert and Knight, 2012). The data on Hsp70 expression in hemocytes are also controversial. While a decreased expression in both resistant and susceptible molluscs has been reported in some studies, with the amplitude being lower for resistant molluscs (Zahoor et al., 2010), a higher gene activity has been found in resistant snails in other studies, suggesting that Hsp70 exhibits protective properties (Lockyer et al., 2008). As indicated earlier, in our dataset Hsp70 demonstrated an elevated expression. Considering that we also observed some bacterial and protozoan transcripts in the hemocytes of the infected snails, which is an indirect evidence of compromised immunity, these finding suggest that an elevated snail Hsp70, despite its stress-alleviating properties, does not ultimately define the invasion failure.

The coding-transcript of **chitin deacetylase**, the protein of aminosugar metabolism, was downregulated. Hemolymph cells of some invertebrates contain chitin or molecules that bind to chitin-binding protein but their functionality is unclear (Heath-Heckman and McFall-Ngai, 2011). It is not the first time when proteins related to synthesis or catabolism of chitin are found to change their expression during trematode invasion. For instance, chitinase-like proteins have been found to exhibit a complex time-dependent behaviour during *Biomphalaria pfeifferi* infection with *Schistosoma mansoni* (Buddenborg et al., 2017). Another underrepresented transcript was that encoding a protein resembling ankyrin repeat domain-containing protein 17, which regulates the cell cycle and participates in antibacterial immune response in human cell lines (Menning and Kufer, 2013). Downregulation of its transcript in *Planorbarius corneus* may be indicative of *Bilharziella polonica* strategy to attenuate snail immunity. However, the role of both these transcripts remains unknown, and so does that of the downregulated transcripts that share some identity with craniofacial development protein 2, filaggrin 2, human pG1 protein.

To sum up, the set of differentially expressed genes reported in our study represents a somewhat mixed snail phenotype. While some upregulated immune-related genes seem to reflect ongoing attempts of the mollusc to suppress the parasite and its reproduction, even at the cercariae-producing stage, others seem to minimize the damage and promote survival of the mollusc with an established infection. The overall picture results from the superposition of the mollusc's efforts to heal its lesions and facilitate immune response and the parasite's efforts to inhibit and overcome them.

The heterogeneity of the population is crucial for its stability under conditions of the changing environment. Some evidence indicates an animal population maintained in the laboratory loses much of its diversity already after several reproduction cycles (unpublished data). Thus, some findings made on laboratory animals may be due to a genetic drift rather than some generalised processes occurring in a free-living population. This is a serious concern to keep in mind. Comparison of our results obtained on free-living animals with the results of experiments on laboratory animals are necessary and planned by our group. It would help to determine the factors defining the success or the failure of the invasion, choose the markers for infection monitoring and assess the extent to which laboratory lines differ from the snails in the nature.

In this study we analyzed and published the first transcriptome of hemocytes of a widespread freshwater pulmonate snail *Planorbarius corneus*, with an emphasis on their immunity function. We provided an improved characterization of a new organism for gastropod immunity studies, making a step towards going beyond the *Biomphalaria glabrata* model. Our findings, alongside with the deposited raw data, provide a basis for further research in the field of parasitology and comparative immunology. **Funding information**

The reported study was funded by RFBR according to the research project No. 20-54-15003; CNRS according to the research project No. 232465 and by the grant of the President of Russian Federation for state support of scholarly research by young scholars (project No. MK- 1015.2021.1.4).

#### **Availability of data and material**

Raw data generated in the current study is available for download from SRA: <https://www.ncbi.nlm.nih.gov/sra/PRJNA865135>.

#### **Ethics approval**

Not applicable.

#### **Declaration of competing interest**

The authors declare that they have no competing interests.

#### **Data availability**

Data will be made available on request.

#### **Acknowledgements**

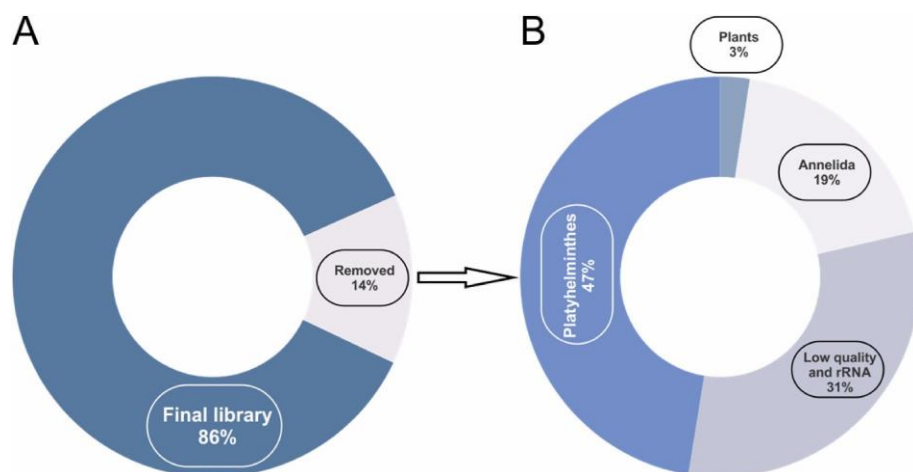
The authors are grateful to Natalia Lentsman for her help with the preparation of the English version of the manuscript.

#### **Appendix A. Supplementary data**

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.dci.2022.104607>.



**Fig. 1.** Composition of final data used for *Planorbarius corneus* transcriptome assembly. **(A)** Proportion of removed sequences in initial data. **(B)** Reasons to discard sequences and the proportions of reads dropped for a particular reason from the total number of dropped reads. These reasons include resemblance to material from contaminating organisms inferred by Kraken2 (Platyhelminthes, Annelida, plants), rRNA origin or low quality of the data. Likeness of reads to the genetic information from contaminating species was defined as presence of 27-mers from the database made up of probable contaminating sequences (accession numbers can be found in supplementary file 1). To reveal reads of rRNA origin, the reads were mapped to chimeric rRNA-coding cluster (see Methods).



**Table 1**

Sizes of initial and final libraries. Final library was obtained after standard pre- processing steps (e.g., discarding reads with too many N's, trimming of low- quality bases), removal of reads mapped to chimeric rRNA-coding cluster (see Methods) and decontamination of the data. The latter was achieved by filtering out all the reads sharing 27-mers with sequences from contaminating organisms (accession numbers can be found in supplementary file 1) as detected by Kraken2 or species closely related to them.

	Control	Infected	Total
Raw library size, 101 read pairs	24225942	27633188	51859130
Final library size, 88bp read pairs	21317065	23427626	44744691

**Table 2**

Technical and biological metrics of assemblies obtained using Trinity, SPAdes or Oases. Fraction of mapped reads was defined by TransRate which uses snap aligner (Smith-Unna et al., 2016). BUSCO (Seppey et al., 2019) with odb10 was implemented to assess the presence of Universal Single-Copy Orthologs (USCOs) in the assembled data.

Parameter	Trinity	SPAdes	Oases
Number of transcripts	548525	489507	488655
Number of genes	478273	474345	–
GC %	35.21	35.05	33.24
Transcripts over 1 k	60550	73350	130406
Transcripts over 10 k	6	3	7
N50	665	831	1206
Average length, bp	538.07	609.91	825.77
Median length, bp	354	386	554
Fraction of mapped reads	0.63	0.61	0.35
Eukaryotic USCOs present	90.6%	92.2%	90.6%
Metazoan USCOs present	92.1%	92.7%	92.2%
Molluscan USCOs present	55.3%	56.0%	55.1%

**Table 3**

Functional groups of immune-related proteins encoded by identified in *Planorbarius corneus* hemocytes. Transcripts were checked for ORF presence, and one best ORF chosen for each transcript (as predicted by TransDecoder, <https://github.com/TransDecoder/TransDecoder/wiki>) was translated. The resulting proteome was annotated using BlastP against NR database (e-value<1e-5). In these results we searched for immunity-related proteins utilizing database comprising names of polypeptides with an immune function previously reported in the literature. We detected transcripts encoding more than 160 different immune-relevant proteins from various functional classes, such as molecules involved in pathogen recognition, adhesion and intercellular interactions, immune effectors and immune regulators.

Functional groups	Immune-relevant factors according to BLAST
Pathogen recognition, adhesion and intercellular interactions	
Toll-like receptors (TLR)	TLR 1, TLR 2, TLR 3, TLR 4, TLR 5, TLR 6, TLR 7, TLR 8, TLR 13, CD 180
Scavenger receptors (SR)	SR class A, SR class B, SR class F, SR cysteine-rich type 1 protein, CD36
Lectins	C-type lectin (C-lect) 1, C-lect 3, C-lect 4, C-lect 6, C-lect 10, C-lect 11, C-lect 18, C-lect-16a I-type lectin 1, I-type lectin 2, I-type lectin 3, I-type lectin 5 Collectin 10, collectin 11, collectin 12 Galectin 3, galectin 4, galectin 6, galectin 7, galectin 8 Fucoselectin-1, fucoselectin-4, fucoselectin-5, fucoselectin-7, F-lectin Tachylectin-4, techylectin-5a, techylectin-5b Pentraxin-1 Ficolin-1 Beta-1,3-glucan-binding protein Gram-negative binding protein Macrophage mannose receptor 1 Mannose-binding lectin Lipid-binding protein Ca <sup>2+</sup> -dependent lipid-bindin
FReDs (fibrinogen domain containing)	Fibrinogen-like protein A, fibrinogen C domain-containing protein 1, angiotensin-related protein 1, FREP 1, FREP 2, FREP 3, FREP 12, FREP B, FREP D, FREP J2
Selectins	E-selectin, polycystin-1, 2, bg selectin 1, e-selectin
Integrins	Integrin alpha-1, integrin alpha-2, integrin alpha-4, integrin alpha-8, integrin alpha-11, integrin beta-1, integrin beta-6
Dermatopontins	Dermatopontin 2, dermatopontin 3
<b>Immune effectors</b>	
Antimicrobial peptides	Achacin-like Aplysianin-a-like
Toxins	Biomphalysin 2, biomphalysin 3, biomphalysin 9, biomphalysin 10, biomphalysin 11, biomphalysin 13, biomphalysin 14, biomphalysin 15, biomphalysin 16, biomphalysin 17, Naegleriapore b pore-forming peptide precursor
Complement	C1q-like protein 4, c1q receptor-like, complement c3-like Thioester-containing protein 1, thioester-containing protein 2, thioester-containing protein 4
Proteases	Chymotrypsinogen, chymotrypsin B, chymotrypsin-2, chymotrypsin-like elastase Collagenase 3, elastase, pancreatic elastase Serine protease 42, serine protease 27, serine protease 30, serine protease alpha, serine protease alpha, transmembrane serine protease 3 Neurotrypsin, trypsin-1, trypsin-2, trypsin-3
Active oxygen and nitrogen production.	Catalase, glutathione peroxidase, glutathione reductase, thioredoxin reductase, thioredoxin, thioredoxin disulfide isomerase, nitric oxide synthase 2, nitric oxide synthase-interacting protein
Phenoloxydase system	Superoxide dismutase, Cu/Zn-superoxide dismutase, tyrosinase Laccase-4
Inhibitors	Alpha-2-macroglobulin, leukocyte elastase inhibitor, Kunitz-type serine protease inhibitor, Kazal-type serine protease inhibitor, subtilisin-chymotrypsin inhibitor

**Table 4**

List of immunity-related domains found in the proteome encoded by the assembled transcriptome of *Planorbarius corneus* hemocytes and number of hits for the indicated categories detected by BlastP. IPS with default setting was exploited to predict protein domains using the same proteome as for previously described BlastP annotation. With the help of database comprising names of immune-related proteins and folds we eventually found 5385 immunity-related domains predicted by IPS in 4405 unique polypeptides. Most of these domains were represented by immunoglobulin superfamily and lectin folds.

	IPS	Blast
IgSF	643	71
Lectin	567	330
Serine proteases and inhibitors	374	112
LRR	314	316
Heat shock protein	300	166
Thioredoxin	257	67
C-type lectin	241	140
Metalloproteases_and_inhibitors	196	35
FRd	176	152
Caspase	171	52
Complement components	160	82
LBSG	158	20
TNF	155	88
LINE elements	143	209
Cyclin	140	141
Interleukin	137	84
TLR	76	139
GST	74	28
Integrin	74	77
Carbohydrate binding	72	2
CDK	65	102
programmed cell death	58	48
TNFR	52	35
MMIF	47	13
PAN-Apple	46	
Cytidine deaminase	44	5
MAPK	44	54
C1q	43	18
Biomphalysin	41	42
Bcl-2	39	29
Concavalin	35	12
F-type lectins	32	28
Galectin	32	27
Retinoic acid inducible	31	69
Scavenger receptors	31	29
Catalase	22	8
Ceruloplasmin	22	15
a2m	21	36
Peroxiredoxin	20	19
Dermatopontin	19	17
Serpin	19	22
LPS-binding	18	30
Cystatin	16	8
Antimicrobial peptide	16	
Superoxide dismutase	16	19
XIAP	15	77
Tyrosinase	13	18
Lysozyme	11	7
Peptidoglycan recognition	11	12
Glutathione peroxidase	9	10
Nitric oxide synthase	9	6
Selectin	9	11
CD-109	8	3
Apoptosis-inducing factor	6	8
Gram-negative binding	5	1
HTRA	5	5
Macrophage mannose receptor	5	32
Fas inhibitor	4	
Glutathione reductase	4	2
ITIM	4	3
Granulin	3	3
Laccase	3	6
Phenoloxidase	2	
Macin	1	
BGR2	1	1
Achacin		10
Aplysianin		12
Ficolin		146
Glucan binding		9
Junctional adhesion		1
MAPKBP		1
TEP proteins		24
Overall	5385	3404
Unique proteins	4405	2687

**Table 5**

BlastX (e-value < 1e-5) annotation of differentially expressed transcripts determined by EdgeR, transcripts with microorganismal, trematodal or plant origin (according to BlastX) are not shown, for the full list see supplementary file 5. Due to the absence of replicates the analysis is considered to be qualitative, and the data are presented as up- or downregulated without a concrete fold change for every transcript (though  $\text{Log}_2|\text{FC}| > 2$  was necessary to consider the transcript to be differentially expressed). EdgeR was provided with the conservative numerical value of 0.4 for biological coefficient of variation to conduct the analysis.

Upregulated transcripts	Downregulated transcripts
replication-associated protein orf1ab	endonuclease-reverse transcriptase
polyprotein	pol polyprotein/retrotransposon-derived presenilin
collagen alpha-1(I) chain-like	replication-associated protein
collagen alpha-1(I) chain-like	chitin deacetylase 8 ankyrin
cathepsin B-like myosin heavy chain 6/7 myosin heavy chain	repeat domain- containing
type II myosin heavy chain	protein 17
elongation factor 1-alpha 1-like	craniofacial development protein 2
polyubiquitin-B turripeptide	filaggrin-2-like isoform X33 G
OL11-like	patch domain-containing protein 1-like
temptin-like/tyramine beta-hydroxylase-like	KHDC3-like protein pG1 protein
apolipoproteins-like brorin-like isoform X3	small proline-rich protein 3-like
heat shock cognate 71 kDa protein-like HSP70	
family protein neurogenic locus notch homolog	
protein 1-like isoform X2 aldehyde dehydrogenase, mitochondrial	

## References

- Adema, C.M., Loker, E.S., 2015. Digenean-gastropod host associations inform on aspects of specific immunity in snails. *Dev. Comp. Immunol.* 48, 275–283. <https://doi.org/10.1016/j.dci.2014.06.014>.
- Adema, C.M., Hillier, L.W., Jones, C.S., Loker, E.S., Knight, M., Minx, P., et al., 2017. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. *Nat. Commun.* 8, 1–11. <https://doi.org/10.1038/ncomms15451>.
- Ataev, G.L., Prokhorova, E.E., Kudryavtsev, I.V., Polevshchikov, A.V., 2016. The influence of trematode infection on the hemocyte composition in *Planorbium corneus* (Gastropoda, Pulmonata). *Invertebr. Surviv. J.* 13, 164–171. <https://doi.org/10.25431/1824-307X/isy.v13i1.164-171>.
- Bayne, C.J., 2009. Successful parasitism of vector snail *Biomphalaria glabrata* by the human blood fluke (trematode) *Schistosoma mansoni*: a 2009 assessment. *Mol. Biochem. Parasitol.* 165, 8–18. <https://doi.org/10.1016/j.molbiopara.2009.01.005>.
- Buddenborg, S.K., Bu, L., Zhang, S.M., Schilkey, F.D., Mkoji, G.M., Loker, E.S., 2017. Transcriptomic responses of *Biomphalaria pfeifferi* to *Schistosoma mansoni*: investigation of a neglected African snail that supports more *S. mansoni* transmission than any other snail species. *PLoS Neglected Trop. Dis.* 11, e0005984 <https://doi.org/10.1371/journal.pntd.0005984>.
- Callisto, M., Moreno, P., Gonçalves, J.F., Ferreira, W.R., Gomes, C.L.Z., 2005. Malacological assessment and natural infestation of *Biomphalaria straminea* (Dunker, 1848) by *Schistosoma mansoni* (Sambon, 1907) and *Chaetogaster limnaii* (K. Von Baer, 1827) in an urban eutrophic watershed. *Braz. J. Biol.* 65, 217–228. <https://doi.org/10.1590/s1519-69842005000200005>.
- Chen, H., Cai, X., Li, R., Wu, Y., Qiu, H., Zheng, J., et al., 2022. A novel toll-like receptor from *Crassostrea gigas* is involved in innate immune response to *Vibrio alginolyticus*. *Infect. Genet. Evol.* 97, 05159 <https://doi.org/10.1016/j.meegid.2021.105159>.
- Combes, C., Albaret, J.-L., Arvy, L., Bartoli, P., Bayssade-Dufour, C., Deblock, S., et al., 1980. M'emoires du Mus'eum National d'Histoire Naturelle. S'erie A, Zoologie. Atlas mondial des cercaires. edition du museum, Paris.
- Dinguirad, N., Cavalcanti, M.G.S., Wu, X.J., Bickham-Wright, U., Sabat, G., Yoshino, T. P., 2018. Proteomic analysis of *Biomphalaria glabrata* hemocytes during in vitro encapsulation of *Schistosoma mansoni* sporocysts. *Front. Immunol.* 9, 2773. <https://doi.org/10.3389/fimmu.2018.02773>.
- Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>.
- Gorbushin, A.M., 2019. Derivatives of the lectin complement pathway in Lophotrochozoa. *Dev. Comp. Immunol.* 94, 35–58. <https://doi.org/10.1016/j.dci.2019.01.010>.
- Gorbushin, A.M., Borisova, E.A., 2015. Lectin-like molecules in transcriptome of *Littorina littorea* hemocytes. *Dev. Comp. Immunol.* 48, 210–220. <https://doi.org/10.1016/j.dci.2014.10.007>.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., et al., 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. <https://doi.org/10.1038/nbt.1883>.
- Heath-Heckman, E.A.C., McFall-Ngai, M.J., 2011. The occurrence of chitin in the hemocytes of invertebrates. *Zoology (Jena)* 114, 191–198. <https://doi.org/10.1016/j.zool.2011.02.002>.
- Hernandez-Samano, A.C., Falcon, A., Zamudio, F., Ortiz-Arellano, M.A., Lopez-Vera, E., Aguilar, M.B., 2020. A turriptide from *Polystira nobilis* venom inhibits human  $\alpha\beta 2$  and  $\alpha 7$  nicotinic acetylcholine receptors. *Insect Biochem. Mol. Biol.* 124, 103416 <https://doi.org/10.1016/j.ibmb.2020.103416>.
- Ittiprasert, W., Knight, M., 2012. Reversing the resistance phenotype of the *Biomphalaria glabrata* snail host *Schistosoma mansoni* infection by temperature modulation. *PLoS Pathog.* 8, e1002677 <https://doi.org/10.1371/journal.ppat.1002677>.
- Kim, D., Langmead, B., Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. <https://doi.org/10.1038/nmeth.3317>.
- Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. <https://doi.org/10.1093/molbev/msw054>.
- Lan, Y., Sun, J., Chen, C., Sun, Y., Zhou, Y., Yang, Y., et al., 2021. Hologenome analysis reveals dual symbiosis in the deep-sea hydrothermal vent snail *Gigantopelta aegis*. *Nat. Commun.* 12, 1165. <https://doi.org/10.1038/s41467-021-21450-7>.
- Lockyer, A.E., Spinks, J., Kane, R.A., Hoffmann, K.F., Fitzpatrick, J.M., Rollinson, D., et al., 2008. *Biomphalaria glabrata* transcriptome: cDNA microarray profiling identifies resistant- and susceptible-specific gene expression in haemocytes from snail strains exposed to *Schistosoma mansoni*. *BMC Genom.* 9, 634. <https://doi.org/10.1186/1471-2164-9-634>.
- Lockyer, A.E., Emery, A.M., Kane, R.A., Walker, A.J., Mayer, C.D., Mitta, G., et al., 2012. Early differential gene expression in haemocytes from resistant and susceptible *Biomphalaria glabrata* strains in response to *Schistosoma mansoni*. *PLoS One* 7, e51102. <https://doi.org/10.1371/journal.pone.0051102>.
- Lu, L., Bu, L., Zhang, S.-M., Buddenborg, S.K., Loker, E.S., 2022. An overview of transcriptional responses of *Schistosoma*-susceptible (M line) or -resistant (BS-90) *Biomphalaria glabrata* exposed or not to *Schistosoma mansoni* infection. *Front. Immunol.* 12, 805882 <https://doi.org/10.3389/fimmu.2021.805882>.
- Marcel, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17, 10–12. <https://doi.org/10.14806/ej.17.1.200>.
- Menning, M., Kufer, T.A., 2013. A role for the Ankyrin repeat containing protein Ankrd17 in Nod1- and Nod2-mediated inflammatory responses. *FEBS Lett.* 587, 2137–2142. <https://doi.org/10.1016/j.febslet.2013.05.037>.
- Mitta, G., Adema, C.M., Gourbal, B., Loker, E.S., Th'eron, A., 2012. Compatibility polymorphism in snail/schistosome interactions: from field to theory to molecular mechanisms. *Dev. Comp. Immunol.* 37, 1–8. <https://doi.org/10.1016/j.dci.2011.09.002>.
- Mitta, G., Gourbal, B., Grunau, C., Knight, M., Bridger, M.J., Theron, A., 2017. The compatibility between *Biomphalaria glabrata* snails and *Schistosoma mansoni*: an increasingly complex puzzle. *Adv. Parasitol.* 97, 111–145. <https://doi.org/10.1016/bs.apar.2016.08.006>.
- Moreira, R., Romero, A., Rey-Campos, M., Pereiro, P., Rosani, U., Novoa, B., Figueras, A., 2020. Stimulation of *Mytilus galloprovincialis* hemocytes with different immune challenges induces differential transcriptomic, miRNomic, and functional responses. *Front. Immunol.* 11, 606102 <https://doi.org/10.3389/fimmu.2020.606102>.
- Myers, J., Ittiprasert, W., Raghavan, N., Miller, A., Knight, M., 2008. Differences in cysteine protease activity in *Schistosoma mansoni*-resistant and -susceptible *Biomphalaria glabrata* and characterization of the hepatopancreas cathepsin B Full-length cDNA. *J. Parasitol.* 94, 659–668. <https://doi.org/10.1645/GE-1410.1>.
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. <https://doi.org/10.1038/nmeth.4197>.
- Pila, E.A., Tarrabain, M., Kabore, A.L., Hanington, P.C., 2016. A novel Toll-like receptor (TLR) influences compatibility between the gastropod *Biomphalaria glabrata*, and the digenean trematode *Schistosoma mansoni*. *PLoS Pathog.* 12, 1005513 <https://doi.org/10.1371/journal.ppat.1005513>.
- Prokhorova, E.E., Tsybalko, N.V., Ataev, G.L., 2010. Expression of genes encoding defence factors in the snail *Planorbium corneus* (Gastropoda, Pulmonata) infested with trematodes. *Parazitologiya* 44, 310–325.
- Prokhorova, E.E., Zhemchuzhnikova, E.A., Ataev, G.L., 2015. Analysis of ITS1 and ITS2 of ribosomal DNA in populations of *Planorbium corneus* snails (Gastropoda) from the Leningrad and Kaliningrad oblasts of Russia. *Contemp. Probl. Ecol.* 8, 729–734. <https://doi.org/10.1134/S1995425515060141>.
- Prokhorova, E.E., Serebryakova, M.K., Tokmakova, A.S., Ataev, G.L., 2018. Hemocytes of mollusc *Biomphalaria glabrata* (Gastropoda, Pulmonata). *Invertebr. Surviv. J.* 15, 346–351. <https://doi.org/10.25431/1824-307X/isy.v15i1.346-351>.
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. <https://doi.org/10.1093/bioinformatics/btp616>.
- Schulz, M.H., Zerbino, D.R., Vingron, M., Birney, E., 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–1092. <https://doi.org/10.1093/bioinformatics/bts094>.
- Schultz, J.H., Bu, L., Kamel, B., Adema, C.M., 2020. RNA-seq: the early response of the snail *Physella acuta* to the digenetic trematode *Echinostoma paraensei*. *J. Parasitol.* 106, 490–505. <https://doi.org/10.1645/19-36>.
- Seppala, O., Walsler, J.-C., Cereghetti, T., Seppala, K., Salo, T., Adema, C.M., 2021. Transcriptome profiling of *Lymnaea stagnalis* (Gastropoda) for ecoimmunological research. *BMC Genom.* 22, 144. <https://doi.org/10.1186/s12864-021-07428-1>.
- Seppely, M., Manni, M., Zdobnov, E.M., 2019. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol. Biol.* 227–245. [https://doi.org/10.1007/978-1-4939-9173-0\\_14](https://doi.org/10.1007/978-1-4939-9173-0_14).
- Shabelnikov, S.V., Bobkov, D.E., Sharlamova, N.S., Petukhova, O.A., 2019. Injury affects coelomic fluid proteome of the common starfish, *Asterias rubens*. *J. Exp. Biol.* 222, jeb198556 <https://doi.org/10.1242/jeb.198556>.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. <https://doi.org/10.1038/msb.2011.75>.
- Smith, M., Yadav, S., Fagunloye, O.G., Pels, N.A., Horton, D.A., Alsultan, N., et al., 2021. PIWI silencing mechanism involving the retrotransposon nimbus orchestrates resistance to infection with *Schistosoma mansoni* in the snail vector, *Biomphalaria glabrata*. *PLoS Neglected Trop. Dis.* 15, e0009094 <https://doi.org/10.1371/journal.pntd.0009094>.
- Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J.M., Kelly, S., 2016. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* 26, 1134–1144. <https://doi.org/10.1101/gr.196469.115>.
- Soneson, C., Love, M.I., Robinson, M.D., 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 1; peer review: 2 approved]. *F1000Research* 4, 1521. <https://doi.org/10.12688/f1000research.7563.1>.
- Tetreau, G., Pinaud, S., Portet, A., Galinier, R., Gourbal, B., Duval, D., 2017. Specific pathogen recognition by multiple innate immune sensors in an invertebrate. *Front. Immunol.* 8, 1249. <https://doi.org/10.3389/fimmu.2017.01249>.
- Vicente-Manzanares, M., Ma, X., Adelstein, R.S., Horwitz, A.R., 2009. Non-muscle myosin II takes centre stage in cell adhesion and migration. *Nat. Rev. Mol. Cell Biol.* 10, 778–790. <https://doi.org/10.1038/nrm2786>.
- Wang, W., Song, X., Wang, L., Song, L., 2018. Pathogen-derived carbohydrate recognition in molluscs immune defense. *Int. J. Mol. Sci.* 19, 721. <https://doi.org/10.3390/ijms19030721>.

Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 257. <https://doi.org/10.1186/s13059-019-1891-0>.

Zahoor, Z., Davies, A.J., Kirk, R.S., Rollinson, D., Walker, A.J., 2010. Larval excretory- secretory products from the parasite *Schistosoma mansoni* modulate HSP70 protein expression in defence cells of its snail host, *Biomphalaria glabrata*. *Cell Stress Chaperones* 15, 639–650. <https://doi.org/10.1007/s12192-010-0176-z>.

Zbikowska, E., Marszewska, A., 2018. Thermal preferences of bird schistosome snail hosts increase the risk of swimmer's itch. *J. Therm. Biol.* 78, 22–26. <https://doi.org/10.1016/j.jtherbio.2018.08.023>.

Zhang, S.-M., Loker, E.S., 2004. Representation of an immune responsive gene family encoding fibrinogen-related proteins in the freshwater mollusk *Biomphalaria glabrata*, an intermediate host for *Schistosoma mansoni*. *Gene* 341, 255–266. <https://doi.org/10.1016/j.gene.2004.07.003>.