PLOS | NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Comparative Transcriptomic Exploration Reveals Unique Molecular Adaptations of Neuropathogenic *Trichobilharzia* to Invade and Parasitize Its Avian Definitive Host

Roman Leontovyč[1]*, Neil D. Young[2], Pasi K. Korhonen[2], Ross S. Hall[2], Patrick Tan[3,4], Libor Mikeš[1], Martin Kašný[1,5], Petr Horák[1], Robin B. Gasser[2]

1 Department of Parasitology, Faculty of Science, Charles University in Prague, Prague, Czech Republic, **2** Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Melbourne, Victoria, Australia, **3** Genome Institute of Singapore, Singapore, Republic of Singapore, **4** Cancer and Stem Cell Biology, Duke-NUS Graduate Medical School, Singapore, Republic of Singapore, **5** Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic

* leontovyc.roman@seznam.cz

Citation: Leontovyč R, Young ND, Korhonen PK, Hall RS, Tan P, Mikeš L, et al. (2016) Comparative Transcriptomic Exploration Reveals Unique Molecular Adaptations of Neuropathogenic *Trichobilharzia* to Invade and Parasitize Its Avian Definitive Host. PLoS Negl Trop Dis 10(2): e0004406. doi:10.1371/journal.pntd.0004406

Editor: Matty Knight, George Washington University School of Medicine and Health Sciences, UNITED STATES

Received: September 11, 2015

Accepted: January 4, 2016

Published: February 10, 2016

Data Availability Statement: This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GDKR00000000. The version described in this paper is the first version, GDKR01000000.

Funding: The study was financially supported by a grant from the Czech Science Foundation (http://gacr.cz/en/), (no. 13-29577S; P.H. et al.) and the Charles University in Prague (http://www.cuni.cz/UKEN-65.html), (PRVOUK P41, UNCE 204017 and SVV

Abstract

To date, most molecular investigations of schistosomatids have focused principally on blood flukes (schistosomes) of humans. Despite the clinical importance of cercarial dermatitis in humans caused by *Trichobilharzia regenti* and the serious neuropathologic disease that this parasite causes in its permissive avian hosts and accidental mammalian hosts, almost nothing is known about the molecular aspects of how this fluke invades its hosts, migrates in host tissues and how it interacts with its hosts' immune system. Here, we explored selected aspects using a transcriptomic-bioinformatic approach. To do this, we sequenced, assembled and annotated the transcriptome representing two consecutive life stages (cercariae and schistosomula) of *T. regenti* involved in the first phases of infection of the avian host. We identified key biological and metabolic pathways specific to each of these two developmental stages and also undertook comparative analyses using data available for taxonomically related blood flukes of the genus *Schistosoma*. Detailed comparative analyses revealed the unique involvement of carbohydrate metabolism, translation and amino acid metabolism, and calcium in *T. regenti* cercariae during their invasion and in growth and development, as well as the roles of cell adhesion molecules, microaerobic metabolism (citrate cycle and oxidative phosphorylation), peptidases (cathepsins) and other histolytic and lysozomal proteins in schistosomula during their particular migration in neural tissues of the avian host. In conclusion, the present transcriptomic exploration provides new and significant insights into the molecular biology of *T. regenti*, which should underpin future genomic and proteomic investigations of *T. regenti* and, importantly, provides a useful starting point for a range of comparative studies of schistosomatids and other trematodes.

260202/2015; P.H. et al.). Funding from the National Health and Medical Research Council (https://www.nhmrc.gov.au), (NHMRC) of Australia, the Australian Research Council and Melbourne Water Corporation is gratefully acknowledged (R.B.G. et al.). This project was also supported by a Victorian Life Sciences Computation Initiative (https://www.vlsci.org.au) (VLSCI; grant number VR0007) on its Peak Computing Facility at the University of Melbourne, an initiative of the Victorian Government (R.B.G.). N.D.Y. holds an NHMRC Early Career Research Fellowship. P.K.K. is the recipient of a scholarship (STRAPA) from The University of Melbourne. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Author Summary

Despite the clinical importance of *Trichobilharzia regenti* in bird hosts and as a cause of cercarial dermatitis in humans, almost nothing is known about the molecular aspects of this fluke and its interactions with its hosts. Here, we sequenced, assembled and annotated the transcriptome representing two life stages (cercariae and schistosomula) of *T. regenti* involved in the first phases of infection of the bird host. We identified key biological and metabolic pathways specific to each of these two developmental stages and also undertook comparative analyses using data available for related flukes. Detailed analyses showed the unique involvement of carbohydrate metabolism, translation and amino acid metabolism, and calcium in *T. regenti* cercariae during invasion and in growth and development, as well as cell adhesion molecules, microaerobic metabolism (citrate cycle and oxidative phosphorylation), peptidases (cathepsins) and other histolytic and lysozomal proteins in schistosomula during migration in neural tissues. These molecular insights into *T. regenti* biology should support future genomic and proteomic investigations of *T. regenti*, and comparative studies of flatworms.

Introduction

The bird fluke *Trichobilharzia regenti* is a member of the Schistosomatidae (= blood flukes; Class Trematoda), a family of parasitic flatworms of medical and veterinary importance [1,2]. *T. regenti* is widely distributed geographically and is highly prevalent, for instance, in parts of Europe (including Russia), New Zealand and Iran [3–6]. Like blood flukes of the genus *Schistosoma*, *T. regenti* is dioecious, has a two-host life cycle (including a lymnaeid snail of the genus *Radix*) and has an invasive furcocercarial stage that actively penetrates the skin of a definitive vertebrate host. Unlike members of the genus *Schistosoma*, *T. regenti* invades and migrates through skin and nerves to then establish within the nasal mucosa [7–9]. During its aquatic phase, *T. regenti* can accidently penetrate human skin and cause cercarial dermatitis. Cercarial dermatitis, caused by avian schistosomes, is regarded as an emerging disease [10–12] although global economic losses are not known, it is accepted that this condition can have a considerable impact on local, tourism-based economies, and may also represent a debilitating occupational disease of rice farmers (see Horák et al., 2015 for review [12]). As avian (including *T. regenti*) and human schistosomes can occur in the same water reservoirs, there are at least two issues of relevance in relation to the differential diagnosis of disease: (a) Based on clinical signs, cercarial dermatitis caused by avian schistosomes can be confused with that caused by human schistosomes [13]. (b) Prevalence surveys of hepatointestinal or urogenital schistosomiasis of humans might be influenced/affected by serological cross-reactivity resulting from exposure to cercariae of avian schistosomes [14].

To better understand *T. regenti* and the diseases that this parasite causes, considerable research has focused on exploring its life cycle. Once shed from the intermediate aquatic snail host, the cercariae survive only for a limited time in water (1 to 1.5 days in related avian schistosomes [15], consuming their glycogen reserves acquired from the intermediate host [16]. Upon contact with the skin of the definitive host, the cercariae release secretions containing proteolytic enzymes (peptidases) from their circumacetabular and postacetabular penetration glands [17], which enable tissue penetration [18–20]. During penetration, the cercariae transform to schistosomula within ~ 12 h [9,21]; for schistosomes, this process is accompanied by a loss of their tail, formation of a double (heptalaminar) membrane covering the tegument and a reduction of surface glycocalyx [21,22] as well as a switch from aerobic to anaerobic

metabolism, depending on the amount of accessible glucose [23,24] and the activation of metabolic processes in the parasite's gut [25]. In contrast to human schistosomes, *T. regenti* schistosomula do not migrate directly to blood vessels, but rather enter peripheral nerves, and migrate to the spinal cord and brain of the host, during which they feed on neural tissue [8,26]. Having reached the pre-adult stage in the meninges, the schistosomula start to feed on blood and then migrate into the nasal cavity, likely via an intravascular route [27]. The significant damage to nerve tissue caused by migrating schistosomula can lead to behavioural changes, disorientation, paralysis or even death in some hosts [7,28].

Despite the importance of cercarial dermatitis in humans caused by *T. regenti* and the unique neuropathogenic effects of this parasite on its permissive avian hosts as well as experimental rodent hosts, little is known about the molecular mechanisms underlying tissue penetration, transformation of cercariae to schistosomula, tissue invasion and parasite-host interactions. Here, we propose that exploring the developmental transcriptomes of cercaria and schistosomulum of *T. regenti* will provide vital insights into the fundamental molecular biology of this parasite, and identify essential pathways and protein classes linked to early tissue invasion. An analysis of the developmental transcriptome of *T. regenti* should also fill gaps in our knowledge of the parasite's biology, as, to date, major molecular investigations have focused mainly on human blood flukes. Therefore, in the present study, we (i) assembled and annotated the transcriptome of two consecutive life stages of *T. regenti* involved in the first phases of infection of the avian host, (ii) identified key biological and metabolic pathways specific to each of these two developmental stages, and (iii) undertook comparative analyses using data available for taxonomically related blood flukes of the genus *Schistosoma*.

Materials and Methods

Ethics statement

The maintenance and care of experimental animals was carried out in accordance with the European Directive 2010/63/EU and Czech law (246/1992 and 359/2012) for biomedical research involving animals. Experiments have been performed under legal consent of the Expert Commission of the Section of Biology, Faculty of Science, Charles University in Prague and the Ministry of Education, Youth and Sports of the Czech Republic (ref. no. MSMT-31114/2013-9).

Parasite materials

Trichobilharzia regenti was maintained in snail intermediate (*Radix lagotis*) and definitive (*Anas platyrhynchos* f. *domestica*; breed—Cherry Valley strain) hosts in the Laboratory of Helminthology, Faculty of Science, Charles University in Prague, using an established protocol [1]. Four separate groups of infected snails (n = 20 snails per group), each representing a distinct biological replicate, were established to obtain four independent biological replicates of pooled cercarial and schistosomula samples (Fig 1).

Schistosomulum replicates: Upon light stimulation for 2 h, cercariae (mixed-gender) shed from each snail group were collected and used to infect seven-day old ducks (n = 4 replicates; 2,500 cercariae per duck). Pooled schistosomula were collected from the spinal cord of each infected duckling seven days following inoculation using established methods [7,28]. Briefly, the spinal cord was carefully prised apart manually (using dissection needles) in phosphate-buffered saline (PBS) and exposed to bright light for 1 h. Schistosomula (n = 4 replicates; 150 to 400 individuals per duckling) were collected, washed extensively in PBS and stored in TRIzol reagent (Invitrogen) at -80°C until further processing. For each biological replicate, a sample of cercariae (shed from the same snail group used to infect the ducks) was also collected for RNA

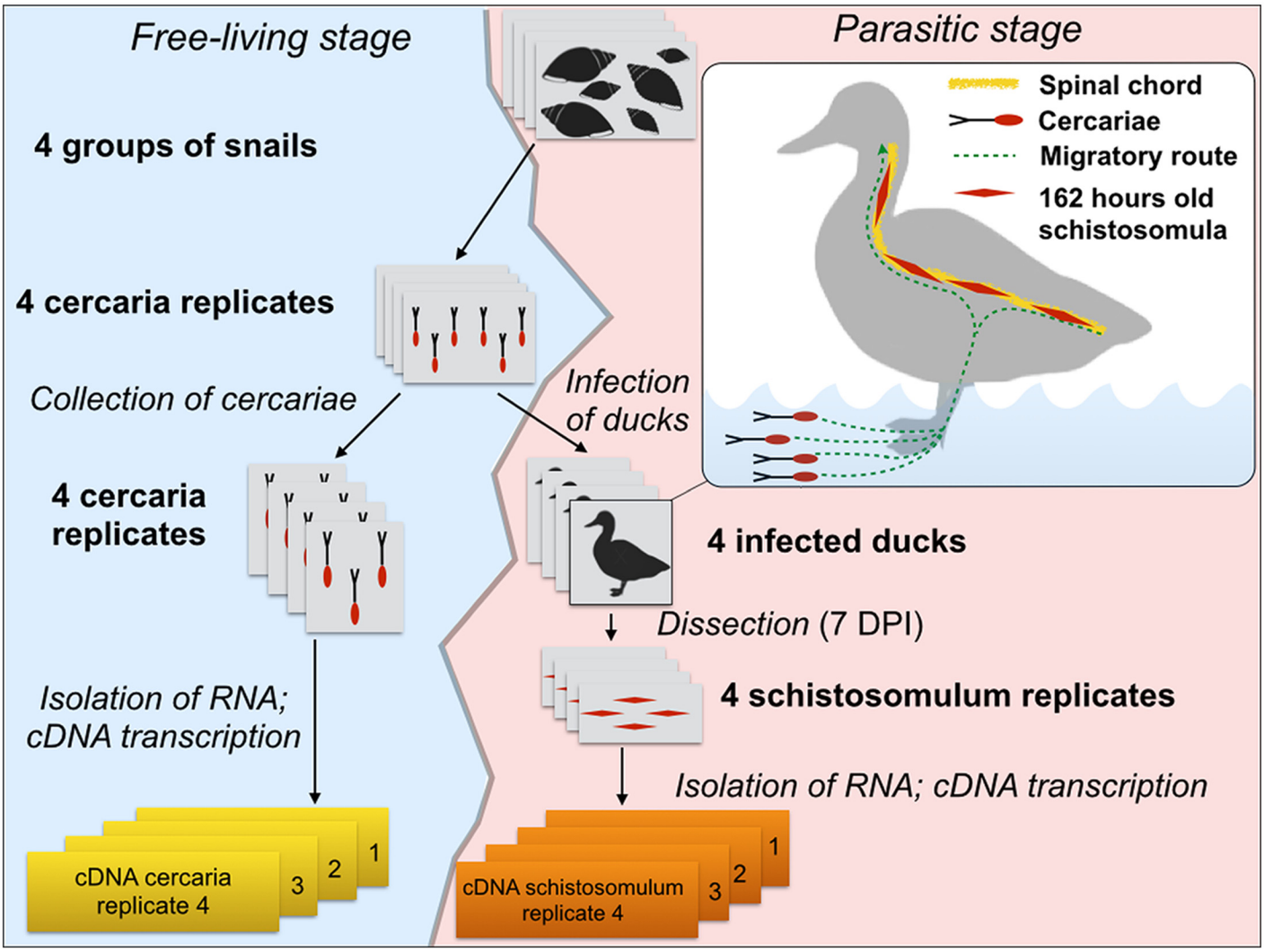

Fig 1. Experimental design for the production of *Trichobilharzia regenti* cercariae and schistosomula for the present study.

doi:10.1371/journal.pntd.0004406.g001

isolation. Cercaria replicates: Cercariae were collected every second day for one week, with each batch washed twice in tap water, centrifuged at 2,500 ×*g* at 4°C, and then stored in TRIzol reagent (Invitrogen) at -80°C.

RNA isolation, library preparation and sequencing

Total RNA was purified from individual biological replicates (four for both cercariae and schistosomula) using TRIzol, and residual genomic DNA removed (DNA-free kit, Invitrogen) following the manufacturer's instructions. The integrity and quality of total RNA were determined using a Bioanalyzer 2100 (Agilent) and Qubit RNA BR assay kit (Invitrogen). Messenger RNA (mRNA) was purified, and short-insert (330 bp) complementary DNA (cDNA) libraries constructed and barcoded according to the manufacturer's instructions (TruSeq RNA Sample Preparation v.2, Illumina). All cDNA library was paired-end sequenced (2 × 211 base reads) on a single line using the HiSeq 2500 platform (Illumina).

Assembly of the transcriptome

Sequencing adaptors and nucleotides with a Phred quality score of < 20 were removed using Trimmomatic v.0.3 [29]. The quality of filtered paired-end read data was manually assessed using FastQC [30] For each RNA-seq dataset, reads were corrected using Spades v.3.1.0 [31] and normalized digitally using the program khmer v.1.1 [32]. For each of the cercarial and schistosomula RNA-seq data sets, replicate datasets were pooled and used to assemble non-redundant transcriptomes using Oases v.0.2.8 [33], employing coverage cut-offs of 13 and 14, and *k*-mer values of 47 and 49, respectively. A final, merged transcriptome was generated by concatenating cercarial and schistosomula transcriptomes, and removing redundancy using CD-HIT-EST [34] using a nucleotide sequence identity threshold of 85%. Finally, coding domains were predicted using Transdecoder [35], and only transcripts encoding proteins of ≥ 30 amino acids were retained. The proportion of genome annotation represented by the non-redundant larval transcriptome was assessed using the program CEGMA [36].

Annotation of the transcriptome

Transcripts homologous (BLASTn; E-value cut-off: $< 10^{-5}$) to avian, bacterial or viral nucleotide sequences in the NCBI non-redundant nucleotide database [37] and translated proteins homologous (BLASTp; E-value cut-off: 10^{-5}) to transposable elements in the RepBase database [38] were quarantined as well as sequences shorter than 50 amino acids (aa) with no homology in public databases. The non-redundant, merged transcriptome was then annotated using an established pipeline [39]. Briefly, protein sequences inferred from transcripts were annotated using their closest homologues (BLASTp; E-value cut-off: $\leq 10^{-5}$) in the following databases: NCBI non-redundant protein [37]; SwissProt [40]; MEROPS peptidase and peptidase inhibitor [41]—with predicted peptidases of unknown catalytic type and inhibitor homologues of unassigned peptidases being excluded; Kyoto Encyclopedia of Genes and Genomes (KEGG) [42], excluding KEGG "Human Diseases" and "Organismal Systems" categories. Conserved domains and their associated gene ontology (GO) annotations were predicted using the program InterProScan [43]. The server REVIGO was used to summarize GO terms and define a representative subset of terms using a simple clustering algorithm that relies on semantic similarity measures [44]. Excretory/secretory (E/S) proteins were predicted based on the presence of a signal peptide and the lack of a transmembrane domain; in addition, proteins were subjected to analysis using the program MultiLoc2 [45] to predict their sub-cellular location. Only proteins predicted to be extracellular or lysosomal (score: > 0.5) were included from the final set of predicted E/S proteins.

Analysis of differential transcription

For each biological replicate, paired, trimmed and corrected reads were mapped to the final transcriptome using RSEM [46]. The expected counts predicted were rounded to the highest whole number, and used as counts per transcript for differential gene transcription analysis using edgeR v.3.6.7 [47] and R v.3.10 [48] software packages, with read counts being normalized to account for any GC bias [49] and using the trimmed mean of M-values (TMM) [50]. Transcripts with more than a $\log_2$ fold-change in transcription between the two developmental stages (i.e. cercaria and schistosomulum) and with a false discovery rate (FDR) of ≤ 0.01 were recorded as differentially transcribed. Enriched GO terms for transcripts recorded to be differentially transcribed in either developmental stage were tested using topGO [51] and the Fisher's exact test ($p \leq 0.05$); representative enriched GO terms were inferred using the program REVIGO [44]. In addition, transcripts specific to either the cercaria or the schistosomulum

were those with RSEM expected counts of > 2 in at least one of four replicates representing one but not the other developmental stage.

Results

Characterisation and annotation of the non-redundant transcriptome of *T. regenti*

The sequencing of eight *T. regenti* cDNA libraries (four representing each the cercaria and schistosomulum biological replicates) produced 146,921,480 high quality reads (75,733,168 for cercariae; 71,188,312 for schistosomula), with an average read length of 132 ± 30 bp (mean ± standard deviation; Tables 1 and S1). Pooled RNA-seq data were used to assemble a non-redundant transcriptome, which included 12,705 assembled transcripts (average nucleotide length: 2,697.1 ± 2,166.4 bp; 115 to 41,111 bp; N50 = 3,333), each encoding a predicted protein (average length: 514.5 ± 537.7 residues; range: 30 to 8,133 residues), excluding transcripts that did not code for a protein (n = 6,899). Of the selected coding regions, 9,514

Table 1. Characteristics of the transcriptomic and predicted proteomic datasets for the cercaria and schistosomulum stages of *Trichobilharzia regenti*.

Sequencing statistics	
Total raw reads sequenced (cercariae; schistosomula)	226,839,830 (102,569,960; 124,269,870)
Total reads trimmed (cercariae; schistosomula)	146,921,480 (75,733,168; 71,188,312)
Total reads trimmed, normalized (cercariae; schistosomula)	28,859,275 (10,953,728; 17,905,547)
Average read length of trimmed reads (mean ± standard deviation)	132±30 bases
Transcriptome assemblies	
Total number of reads for assembly	28,859,275
Contigs of ≥105 nucleotides (mean ± S.D.; range)	20,958 (2007.2±2063.13; 105–41,111)
Number of reads that mapped to contigs (%)	16,337,592 (67,9)
Number of non-redundant proteins of ≥30 amino acids predicted (mean ±S.D.; range)	12,705 (514.5±537.1; 30–8,133)
Complete/partial matches to 248 CEGMA-encoded proteins (%)	87.9/89.5
Numbers of proteins homologous (BLASTp; E-value cut-off off ≤ $10^{-0.5}$) to entries in various databases (01 July 2014) (% of predicted proteome)	
NCBI	10,900 (85.7)
SwissProt	8,347 (65.6)
MEROPS peptidase	318 (2.5)
MEROPS peptidase inhibitor	260 (2.0)
Numbers of proteins homologous (BLASTp; E-value cut-off of ≤ $10^{-0.5}$) to entries in KEGG databases (% of predicted proteome; number of conserved KEGG protein classes or pathways)	
KEGG BRITE	5,935 (46.7; 3,275)
KEGG PATHWAY	3,611 (28.4; 1,934)
Numbers of predicted proteins with conserved domains or GO annotations (% of predicted proteome; number of unique InterProScan domains or GO terms)	
InterProScan conserved domains	10,585 (83.3; 5,115)
GO terms (number of transcripts)	6,961 (54.8; 1,527)
Biological process	3,946 (31.0; 576)
Numbers of proteins predicted to be excreted/secreted (% of predicted proteome)	
Predicted E/S proteins	135 (1.1)

doi:10.1371/journal.pntd.0004406.t001

commenced with a start codon, and 11,434 terminated with a stop codon. Despite sequencing only two developmental stages, the *T. regenti* transcriptome includes 89.5% of the 358 conserved eukaryotic genes [36] (Table 1), and thus represents a substantial proportion of the gene set. On average, 61.5% and 74.4% of the sequence reads representing the cercaria and schistosomulum, respectively, mapped to the non-redundant *T. regenti* transcriptome (S1 Table). The final transcriptome and RNA-seq read data are available for download via the NCBI transcript reads archive and sequence read archive (SRA), respectively (BioProject ID: PRJNA292737).

Totals of 88 and 243 potential contaminants (avian, viral and bacterial) were removed from the cercaria and schistosomulum transcriptomes, respectively. Following the removal of these sequences, the larval transcriptome was shown to encode 10,900 (77.5%) and 8,347 (57.8%) predicted proteins that were homologous (BLASTp; E-value $\leq 1e^{-05}$) to those in the NCBI (non-redundant proteins) and SwissProt databases, respectively (Table 1). In addition, 5,935 predicted proteins were assigned 41 unique KEGG BRITE protein families (Fig 2A; Tables 1 and S2), and 3,611 predicted proteins were assigned 172 biological KEGG pathways (Tables 1 and S2). Using the MEROPS peptidase and inhibitors database, 318 peptidases, including key molecules recognized to be involved in cercarial penetration and schistosomulum migration, nutritional uptake and/or immune evasion [19] were identified (S3 Table). Transcripts encoding metallopeptidases ($n = 129$) were abundant, and included ubiquinol-cytochrome *c* reductase proteins ($n = 23$), kell blood-group proteins ($n = 8$) and leucine aminopeptidases ($n = 4$). Transcripts encoding cysteine peptidases ($n = 106$) including cathepsin B ($n = 11$), cathepsin L ($n = 6$) cathepsin C (= dipeptidyl-peptidase; $n = 3$), ubiquitin-specific peptidases ($n = 10$) and legumains/asparaginyl endopeptidases ($n = 2$) and a ubiquitin-specific peptidase, were also well represented. Transcripts encoding serine peptidases ($n = 62$) represented indeterminate peptidases ($n = 43$), cathepsin A (= carboxypeptidase A; $n = 3$) and mitochondrial inner membrane peptidase 2 ($n = 2$). Transcripts encoding threonine peptidases ($n = 15$) were represented by proteasome subunits ($n = 4$). Also identified was a small number of transcripts encoding aspartic proteases ($n = 6$), including one cathepsin D. In total, transcripts representing 260 inhibitors of peptidases were identified in the (non-redundant) transcriptome of *T. regenti*, most of which represented inhibitors of metallo- ($n = 110$), serine ($n = 77$) and cysteine peptidases ($n = 41$) (S3 Table).

A relatively large number of proteins predicted were homologous to those encoded in the genomes of other trematodes, including schistosomes (10,909 proteins; 85.6% being similar to *Schistosoma mansoni*, *S. japonicum* and/or *S. haematobium*) (Fig 2B) and Asian liver flukes (10,080 proteins; 79.3% being similar to *Clonorchis sinensis* and/or *Opisthorchis viverrini*) (Fig 2C). Of note were 1,722 transcripts that had no homology to other trematode species; only 12 of them (including collagen alpha-3(VI) chain-like, 40S ribosomal protein S7-like, ReO_6, membrane magnesium transporter 1-B-like, VWFA and cache domain containing protein 1 and UPF0729 protein C18orf32 homolog) shared homology to proteins in the NCBI database. In addition, collagen, type VI, alpha and small subunit ribosomal protein S7e were predicted to be involved in five biological pathways (S4 Table). In addition, the larval transcriptome encoded 135 ES proteins, including various peptidases and their inhibitors, phosphatases, kinases, transferases and ribonucleases (S5 Table).

Differentially transcribed genes, and biological pathways enriched in either cercariae or schistosomula

A total of 11,058 transcripts were shared by cercariae and schistosomula. Mapping results revealed 270 and 951 transcripts to be specific to the cercaria and schistosomulum stages, respectively (Fig 3A; S6 and S7 Tables). In total, 1,301 transcripts were up-regulated in

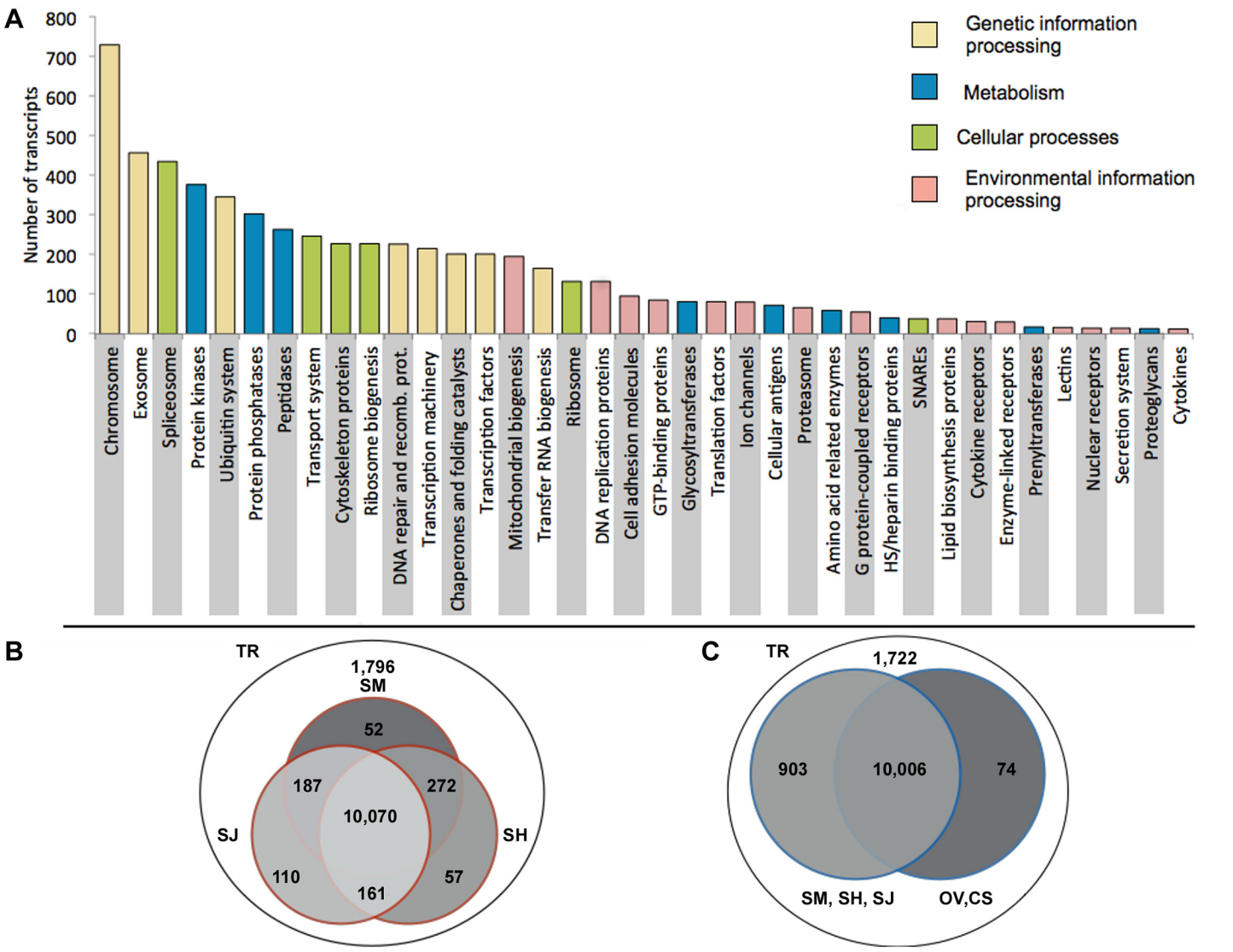

Fig 2. (A) Kyoto Encyclopedia of Genes and Genomes (KEGG) BRITE annotation of transcripts from cercaria and schistosomulum transcriptomes of *Trichobilharzia regenti*. (B, C) Venn diagrams displaying the results from homology-based comparisons of transcripts in the transcriptome of *T. regenti* cercariae or schistosomula with the genomes of *Schistosoma mansoni* (SM), *S. haematobium* (SH), *S. japonicum* (SJ), *Clonorchis sinensis* (CS) and *Opisthorchis viverrini* (OV).

doi:10.1371/journal.pntd.0004406.g002

cercariae and 1,876 in schistosomula (S8 and S9 Tables). The top twenty most differentially transcribed genes in cercariae encoded venom allergen-like protein 8, tegumental protein, calcium-binding proteins, glutamine synthetase and some uncharacterized proteins (with no homology to any sequences in public databases). The top twenty most differentially transcribed genes in schistosomula encoded peptidases (e.g. peptidase M26, cathepsin B1) and beta galactosidase, some uncharacterized proteins with homology to those of *S. mansoni*, two of which (Treg_015087, Treg_015334) were homologous to a saposin-like protein and beta hexosaminidase B based on InterPro classification respectively (Table 2). A number of up-regulated transcripts encoded proteins with conserved functional domains (1,178 and 1,699 for cercariae and schistosomula, respectively) and/or similarity to proteins in the KEGG database [605 and 786 (BRITE) (Fig 3B), and 389 and 487 (pathway) for cercariae and schistosomula, respectively].

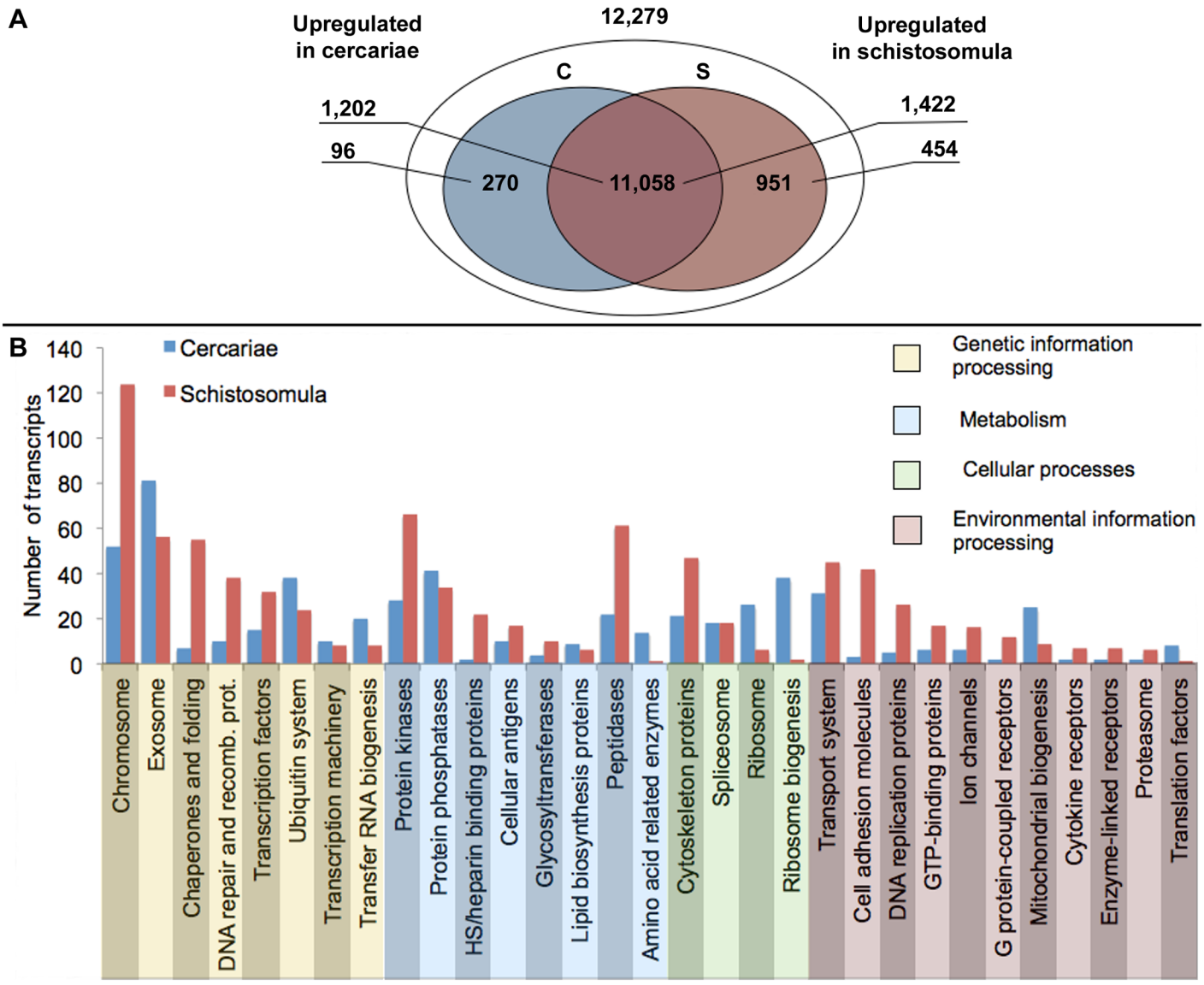

Fig 3. (A) Transcripts expressed exclusively in cercariae, schistosomula or in both developmental stages, based on expected counts derived from RNA-Seq by Expectation Maximization (RSEM) analysis. Cercariae (C) and schistosomula (S); heparan sulphate (HS). (B) Bar graph showing genes differentially transcribed between cercariae and schistosomula of *Trichobilharzia regenti* (Kyoto Encyclopedia of Genes and Genomes (KEGG) BRITE annotation).

doi:10.1371/journal.pntd.0004406.g003

Using these transcripts, we were able to identify enriched KEGG BRITE protein families, biological pathways and GO terms (biological process) in each of the two larval stages (Tables 3, 4, S10 and S11).

In cercariae, 162 up-regulated transcripts represented five enriched KEGG BRITE protein families (Tables 3 and S10), with most transcripts encoding exosome-related proteins, including calmodulin ($n = 10$), long-chain acyl-CoA synthetase ($n = 4$), creatine kinase ($n = 4$) and tropomyosin 1 ($n = 3$). Other protein families were associated with: (a) metabolic processes involving lipid biosynthesis proteins ($n = 9$) and amino acid-related enzymes ($n = 14$), and (b) cellular processes linked to protein translation, ribosomes ($n = 26$) and ribosome biogenesis ($n = 38$). In addition, 202 transcripts up-regulated in cercariae represented 21 enriched KEGG (biological) pathways (Tables 3 and S10), with almost one quarter of them ($n = 50$) linked to

Table 2. Twenty most differentially transcribed genes of *Trichobilharzia regenti* cercariae and schistosomula.

Transcript ID	Expected counts	Log2 fold change	NCBI annotation
Cercariae			
Treg_000002	33,582	18.4	NO HIT
Treg_000452	27,851	18.3	gi|159792922|gb|ABW98681.1| venom allergen-like protein 8 (*Schistosoma mansoni*)
Treg_000248	20,225	17	gi|256052720|ref|XP_002569901.1| tegumental protein (*S. mansoni*)
Treg_003587	19,928	17.1	NO HIT
Treg_003763	10,904	16.5	gi|256075345|ref|XP_002573980.1| Calcium-binding protein 2 (CaBP2) (*S. mansoni*)
Treg_000203	10,417	15.9	NO HIT
Treg_000153	10,336	15.4	gi|454247|emb|CAA82847.1| zinc finger protein (putative) (*S. mansoni*)
Treg_000941	7,094	15.3	NO HIT
Treg_000009	5,956	15.3	gi|256084831|ref|XP_002578629.1| hypothetical protein (*S. mansoni*)
Treg_000069	5,112	15.2	NO HIT
Treg_000112	3,731	15.1	gi|256074546|ref|XP_002573585.1| glutamine synthetase bacteria (*S. mansoni*)
Treg_003585	2,948	15.1	gi|256084831|ref|XP_002578629.1| hypothetical protein (*S. mansoni*)
Treg_000514	2,425	14.5	gi|256076875|ref|XP_002574734.1| Calmodulin (CaM) (*S. mansoni*)
Treg_000442	2,264	14.4	NO HIT
Treg_000878	1,116	14.2	NO HIT
Treg_003936	1,060	14.1	NO HIT
Treg_000031	868.185	14.0	gi|226485052|emb|CAX79803.1| Calcium-binding EF-hand,domain-containing protein (*Schistosoma japonicum*)
Treg_000619	731.25	13.9	NO HIT
Treg_000633	666.5	13.8	gi|256076881|ref|XP_002574737.1| calmodulin (*S. mansoni*)
Treg_000137	650	13.8	NO HIT
Schistosomula			
Treg_007215	7,386	16.5	NO HIT
Treg_015087	6,081	15.0	gi|353232479|emb|CCD79834.1| hypothetical protein Smp_194910 (*Schistosoma mansoni*); IPR011001|Saposin-like
Treg_006410	5,681	14.9	NO HIT
Treg_006851	3,719	14.3	gi|489236049|ref|WP_003144328.1| peptidase M26 (*Gemella haemolysans*)
Treg_006064	3,587	14.2	gi|350855260|emb|CAZ35935.2| MEG-4 (10.3) family (*S. mansoni*)
Treg_014548	2,670	14.2	NO HIT
Treg_015086	2,608	13.8	NO HIT
Treg_020247	1,984	13.5	gi|55793949|gb|AAV65885.1| cathepsin B1 isotype 5 precursor (*Trichobilharzia regenti*)
Treg_015412	1,771	13.3	gi|555955168|ref|XP_005889743.1| PREDICTED: beta-galactosidase-1-like protein isoform X2 (*Bos mutus*)
Treg_006816	1,537	13.3	NO HIT
Treg_015334	1,509	12.8	gi|56757485|gb|AAW26910.1| SJCHGC06873 protein *(Schistosoma japonicum*); IPR025705|Beta-hexosaminidase subunit alpha/beta
Treg_015853	1,500	12.8	gi|195729971|gb|ACG50796.1| cathepsin B1 (*Trichobilharzia szidati*)
Treg_008400	1,327	12.7	NO HIT
Treg_009334	1,323	12.7	gi|256078798|ref|XP_002575681.1| hypothetical protein (*Schistosoma mansoni*)
Treg_006699	1,082	12.6	gi|353231322|emb|CCD77740.1| MEG-8 family (*S. mansoni*)
Treg_006885	975	12.5	gi|226471160|emb|CAX70661.1| Saposin B domain-containing protein (*S. japonicum*)
Treg_008899	877	12.5	gi|4099279|gb|AAD00565.1| precursor anti-coagulant SAP-1 (*Schistosoma haematobium*)
Treg_007673	876	12.5	gi|226467490|emb|CAX69621.1| DM9 domain-containing protein (*S. japonicum*)
Treg_007426	747	12.4	NO HIT
Treg_006736	677	12.4	NO HIT

doi:10.1371/journal.pntd.0004406.t002

Table 3. Summary of enriched metabolic pathways and enzyme classes of cercaria stage (C) with comparison to schistosomulum (S) of *Trichobilharzia regenti* based on KEGG BRITE/pathway classification and Gene ontology (GO).

KEGG pathway/brite classification			
	No. of up-regulated transcripts (C/S)	No. of EC terms (C/S)	Total (C+S) No. of transcripts/EC terms
Metabolism			
KEGG pathways			
carbohydrate metabolism	50/16	35/12	180/87
amino acid metabolism	26/7	16/5	75/43
energy metabolism	25/3	23/3	92/67
metabolism of cofactors and vitamins	15/2	8/2	31/11
nucleotide metabolism;Purine metabolism	24/13	14/11	141/68
KEGG protein class			
lipid biosynthesis proteins	9/6	5/6	38/22
amino acid related enzymes	14/1	7/1	59/29
Genetic information			
KEGG pathways			
translation	42/3	38/3	183/128
KEGG protein class			
exosome	81/47	63/38	456/190
Enviromental information processing			
KEGG pathways			
signal transduction	21/7	11/6	97/40
Cellular processes			
KEGG pathways			
cell growth and death	23/12	11/9	97/38
transport and catabolism	10/2	6/2	26/11
KEGG protein class			
ribosome	26/1	24/1	132/100
ribosome biogenesis	38/2	33/2	227/156
Categories of biological processes of gene ontology (GO)			
		Number of transcripts	
Cellular respiration		115	
Carbohydrate metabolism		40	
Glucose metabolism		29	
Actomyosin structure organization		23	
Generation of precursor metabolites and energy		23	
Coenzyme metabolism		22	
Cofactor metabolism		22	
Pyridine-containing compound metabolism		9	

doi:10.1371/journal.pntd.0004406.t003

pathways associated with carbohydrate metabolism, including the tricarboxylic acid cycle ($n = 23$), glycolysis ($n = 21$) and pyruvate metabolism ($n = 13$). Other pathways enriched in the cercariae were linked to amino acid metabolism [alanine, aspartate and glutamate ($n = 11$); arginine and proline metabolism ($n = 11$); and glycine, serine and threonine metabolism ($n = 7$)]; energy metabolism [oxidative phosphorylation; $n = 21$; nitrogen metabolism ($n = 4$)]; the metabolism of cofactors and vitamins and nucleotide metabolism [nicotinate and nicotinamide metabolism ($n = 15$) and purine metabolism ($n = 24$)]; environmental information processing [calcium signaling pathway ($n = 23$) and two component regulatory system ($n = 4$)] (Tables 3 and S10). In addition to these enriched protein families and biological pathways, 161

Table 4. Summary of enriched metabolic pathways and enzyme classes of schistosomulum stage (S) with comparison to cercaria (C) of *Trichobilharzia regenti* based on KEGG BRITE/pathway classification and Gene ontology (GO).

KEGG pathway/brite classification			
	No. of up-regulated transcripts (S/C)	No. of EC terms (S/C)	Total (C+S) No. of transcripts/EC terms
Metabolism			
KEGG pathways			
glycan biosynthesis and metabolism	34/2	4/1	54/10
carbohydrate metabolism;Amino sugar and nucleotide sugar metabolism	32/7	3/4	73/23
KEGG protein class			
peptidases	61/22	31/14	263/135
heparan sulfate/heparin binding proteins	22/2	11/2	40/19
proteoglycans	6/0	6/0	13/9
Genetic information			
KEGG protein class			
chaperones and folding catalysts	55/7	12/7	20192
Enviromental information processes			
KEGG pathways			
signaling molecules and interaction	13/0	12/0	23/14
KEGG protein class			
cell adhesion molecules and their ligands	42/3	28/2	95/47
Cellular processes			
KEGG pathways			
lysosome	78/4	20/4	147/51
cell cycle	34/10	26/8	143/76
Categories of biological processes of gene ontology (GO)			
			Number of transcripts
Proteolysis			215
Cell adhesion			105
Biological adhesion			37
Developmental process			20
Multicellular organismal process			15

doi:10.1371/journal.pntd.0004406.t004

transcripts up-regulated in the cercariae were linked to 36 GO terms for 'biological process' that were enriched in the cercarial stage (Tables 3 and S10). These GO terms were represented by eight representative biological processes, including cellular respiration (115 transcripts linked to 17 GO terms, including ATP metabolic process, glycerol-3-phosphate metabolic process, nucleotide biosynthetic process, organophosphate biosynthetic process and oxidation-reduction process), glucose metabolism (29 transcripts representing five GO terms, including carbohydrate catabolic process, cellular carbohydrate metabolic process, glucose metabolic process and organic substance catabolic process), acto-myosin structure organization (23 transcripts representing five GO terms, including actomyosin structural organization, cell wall macromolecule metabolic process, cell wall organization or biogenesis, energy-coupled proton transport, down electrochemical gradient, and ribosome biogenesis) and coenzyme metabolism (22 transcripts representing four GO terms, including acetyl-CoA metabolic process, coenzyme catabolic process, coenzyme metabolic process and cofactor catabolic process). Other biological processes were represented by single GO terms and included carbohydrate metabolism (40

transcripts), cofactor metabolism (n = 22), generation of precursor metabolites and energy (n = 23) and pyridine-containing compound metabolism (n = 9).

In schistosomula, 145 up-regulated transcripts represented five enriched KEGG BRITE protein families (Tables 4 and S11), with most encoding peptidases (61 transcripts) including cathepsin B (n = 11), cathepsin D (n = 8), separase (n = 4), leucyl amino peptidase (n = 3) and legumain (n = 3). Other protein families enriched were associated with: (a) metabolic processes involving heparan sulfate/heparin binding proteins (n = 22) and proteoglycans (n = 6), (b) genetic information processing involving chaperones and folding catalysts (n = 54) and (c) environmental information processes represented by cell adhesion molecules and their ligands (n = 42). In addition, 128 up-regulated transcripts represented 10 enriched KEGG biological pathways (Tables 4 and S11). Metabolic pathways involved carbohydrate, amino sugar or nucleotide sugar metabolism (n = 32); glycan biosynthesis and metabolism, including glycosaminoglycan degradation (n = 32), and glycosphingolipid biosynthesis (n = 33). In each metabolic pathway, 30 different transcripts encoding hexosaminidase were inferred to be involved. Pathways associated with signalling were also significantly enriched in the schistosomula, and represented by cell adhesion molecules (n = 6) and extracellular matrix—receptor interaction molecules (n = 10). Other enriched pathways in the schistosomula were associated with cellular processes, such as cell cycle (n = 21) and processes linked to lysosome function (78 transcripts, including 30 encoding hexosaminidases). In addition to the enriched protein families and biological pathways, 332 transcripts were linked to 28 GO ('biological process') terms enriched in the schistosomulum stage (Tables 4 and S11). These GO terms represented five core processes: cell adhesion (n = 105 transcripts; 18 associated GO terms), proteolysis (n = 215; 7 GO terms), biological adhesion (n = 37; one GO term), multicellular organismal process (n = 15; one GO term) and developmental processes (n = 20; one GO term).

Discussion

To date, genomic and transcriptomic studies of schistosomatids have focused on *S. mansoni*, *S. japonicum* and *S. haematobium* [52–54]. In the present study, we characterized the first transcriptome of any avian fluke, *T. regenti*, undertook comparative studies with other schistosomatids and elucidated differences between two key developmental stages responsible for host invasion (cercaria) and migration through the neural tissues (schistosomulum) at the molecular level.

Comparison of *T. regenti* with other schistosomatids

Based on similarity searches, the proteins predicted from the *T. regenti* transcriptome have 85.9% sequence homology to those of human schistosomes (BLASTx, E-value 10^{-5}), although divergent molecules (14.2%) are proposed to relate to considerable differences in biology of bird and human schistosomes. *Trichobilharzia regenti* shares 110 transcripts exclusively with *S. japonicum*, with no homologues in either *S. mansoni* or *S. haematobium*. This finding is consistent with *S. japonicum* being the most similar of the three human schistosomes to *T. regenti* in terms of chemical tools serving for skin penetration. In particular, a shared feature of *T. regenti* and *S. japonicum* is the absence, at both the mRNA and protein levels, of cercarial elastase [18,55–57]. By contrast, *S. mansoni* and *S. haematobium* both use elastase for cercarial invasion of the definitive host [58,59]. While humans represent exclusive/dominant hosts of *S. haematobium* and *S. mansoni*, *S. japonicum* has an ability to infect a broader range of mammals (including pigs, water buffalo and water rats; [60]), but not birds as natural final hosts. In this case, therefore, the absence of elastase in both *S. japonicum* and *T. regenti* seems to be an interesting example of convergence. In other words, the data here indicate that *T. regenti* shares

more unique transcripts (n = 110) with *S. japonicum* than either *S. mansoni* (n = 52) or *S. haematobium* (n = 57) (Fig 2B); the apparent absence of cercarial elastase from the two former species might reflect their distinct host affiliations/broader spectra compared with the latter two. Interestingly, 1,722 (13.6%) of predicted proteins of *T. regenti* had no homology to those of any other trematode species, for which molecular data are available (none of which are parasites of birds) (Fig 2B). Of these 1,722 predicted proteins, only 151 had homologues in public databases. Remaining 1,571 transcripts represented orphans (genes with no homology to known domains or proteins). However, some of these orphans were highly transcribed in cercariae and/or schistosomula. Interestingly, 10 and 8 of 20 genes exhibiting the highest transcription in cercariae and schistosomula, respectively, encoded orphans. Such molecules might have unique particular biological functions or processes in *T. regenti* associated with the penetration of avian skin (cercaria), specific migration through neural tissue (schistosomulum) and an adaptation to the avian definitive host (both stages). This statement is supported by biological and clinical evidence [26], showing that *T. regenti* is not able to effectively establish infection in the mammalian hosts under natural conditions.

Molecular aspects unique to the cercarial stage of *T. regenti*

In digenetic trematodes, interestingly, the cercaria are considered to be less transcriptionally active than other developmental stages [61–63], although the high transcription of genes encoding proteins participating in particular metabolic pathways described here likely reflects the specific biological requirements of this larval stage.

Carbohydrate metabolism. After leaving their snail intermediate host, schistosome cercariae must find and penetrate the epidermis of their definitive host within hours [16], and thus rely exclusively on stored energy reserves (including glycogen) [16]. In *T. regenti* cercariae, the enriched carbohydrate metabolism, including glycolysis (linked to glucose-6-phosphatase, fructose-1,6-bisphosphatase, phosphoenolpyruvate carboxykinase and pyruvate carboxylase), pyruvate, citrate cycle and oxidative phosphorylation pathways indicate the importance of aerobic degradation of glucose from glycogen sources, consistent with observations in other free-living trematode stages [64–68]. However, interestingly, other pathways for the degradation of fructose, mannose, pentose phosphate, starch, sucrose, galactose, glyoxylate and dicarboxylate were also enriched in *T. regenti* cercariae and appear to be functional; the present results indicate that some enzymes likely participate in multiple metabolic pathways. For instance, citrate synthase is likely involved in glyoxylate and dicarboxylate metabolism as well as in the citrate cycle, the latter of which is regarded as a major metabolic pathway in *T. regenti* cercariae, corresponding to evidence for *S. mansoni* [69].

Translation and amino acid metabolism. The present findings show that *T. regenti* cercariae are less transcriptionally active than schistosomula, although the pathways linked to translational machinery, such as ribosome biogenesis, were significantly enriched, including 24 distinct large ribosomal subunit proteins. We speculate that cercariae synthesize ribosomes for immediate protein synthesis following their invasion of the definitive host, in which the energy and protein sources would appear to be unlimited. The limited pool of amino acids in schistosome cercariae might originate from the snail intermediate host and/or from biosynthesis [70]; in *T. regenti*, the transcripts encoding enzymes involved in the metabolism of amino acids, such as alanine (alanine transaminase), asparagine (aspartate-ammonia ligase), glutamine (glutamine synthetase) and cysteine (cystathionine gamma-lyase), were identified. All of these amino acids are derived from the citrate cycle or glycolysis, considered the main metabolic pathways in the cercarial stage. Although the synthesis of amino acids by parasitic stages of helminths is well recognized [71], this aspect has been seldom studied in free-living stages of schistosomes [62,72].

Key role for calcium during invasion. In *T. regenti* cercariae, calcium signalling may also be important for enriched cellular processes, as has been observed for *S. japonicum* [62]. For example, transcripts encoding calcium-binding protein (CaBP) represented the fifth most differentially transcribed gene (log-fold change: 16.5) between cercariae and schistosomula of *T. regenti* (cf. Table 2), which likely relates to CaBP being essential during the very active invasion process of larvae in the definitive host [73]. In other schistosomes, such as *S. mansoni*, there is a down regulation of CaBP in cercariae following epidermal penetration [73]. The preacetabular (circumacetabular) penetration glands of schistosome cercariae contain a high concentration of calcium [18,74,75]. This calcium has been suggested to regulate the activity of *S. mansoni* cercarial elastase [76,77] and of glycocalyx shedding by cross-linking endogenous postacetabular mucopolysaccharides [18,76]. Although there is no evidence here of elastase in the *T. regenti* cercaria, the highest transcription of any peptidase in this stage was exhibited by the calpain gene, which is strictly regulated by calcium ions. Calpain may regulate the surface membrane synthesis process, as it does in *S. mansoni* [78], but this proposals needs to be tested.

Molecular processes unique to the schistosomulum stage of *T. regenti*

During their invasion of the definitive host, the transformation of cercariae to schistosomula is associated with rapid and major morphological, biochemical and molecular changes. Unlike most schistosomes studied to date, the schistosomula of *T. regenti* do not enter the bloodstream, but rather seek out and migrate in nerves and consume neural tissue (as nutrition) during migration [26]. Compared with the circulatory system, the nervous system represents an immunologically and physiologically distinct environment (in terms of nutrients and immune responses). The present study investigated the schistosomula of *T. regenti* seven days after infection of the avian host and explored the molecular adaptations required for the parasite to establish and survive in this unique niche—the neural system.

Growth and development, and cell adhesion molecules (CAMs). In addition to playing critical roles in regulating normal cell integrity and cell-cell interactions, cell adhesion molecules (CAMs) can also regulate host-parasite interactions. For example, in protozoan parasites, CAMs have been reported to mediate the attachment of pathogens to host cells [79]. Whilst CAMs of *T. regenti* might participate in host-parasite interactions, it seems more likely that they regulate cell adhesion to maintain their multicellular structure [80]. Highly up-regulated transcription associated with CAMs in the schistosomula of *T. regenti* might indicate rapid growth and development of different organ structures within the definitive host. For instance, abundant transcription linked to neuroligin, netrin receptor and semaphorin might relate to the development of the nervous system in this developmental stage, based on evidence from other organisms [81,82].

Metabolism. Contrary to the situation in the cercariae of *T. regenti*, transcription of genes encoding enzymes involved in aerobic metabolic processes in the schistosomula is very low, and limited to the citrate cycle and oxidative phosphorylation. This finding suggests that the schistosomulum possesses a microaerobic metabolism seven days after infection of the definitive host, which is similar to observations made for human schistosomes [25]. However, in the schistosomulum stage of *S. mansoni*, for example, most energy is reported to originate via anaerobic glycolysis, with lactate as an end product [24,25,83], contrasting the situation in the *T. regenti* schistosomulum.

Proteolysis and histolysis during migration. The schistosomula of *T. regenti* need to migrate through the avian neural tissues to reach the nasal mucosa, where they mature to adults, mate and produce eggs [1]. The most abundant transcripts specifically in the schistosomulum stage encode proteolytic enzymes, including several cysteine peptidases, in particular,

transcripts encoding cathepsin B1.5 (Treg_006320) (RSEM-expected counts: 30,779), followed by cathepsin L (Treg_006337) (12,349), cathepsin B1.6 (Treg_004279) (8,216), cathepsin L-like peptidase (Treg_006792) (5,514), cathepsin C (Treg_014726) (5,145) and leucine amino peptidase (4,966) (Treg_015017) (S9 Table). While cercariae store and express proteolytic enzymes in penetration glands for invasion via the skin, schistosomula degrade various host tissues during migration [84], including neural tissue, specifically in the case of *T. regenti* [26], and evade or block host immune attack [75]. The present study indicates that *T. regenti* schistosomula employ a considerably broader arsenal of proteolytic enzymes than cercariae do, with more than twice as many peptidases (n = 31) predicted to be upregulated in the former than in the latter stages, respectively.

According to a previous study [85], the digestive peptidase cathepsin B1 of *T. regenti* schistosomula (TrCB1) represents at least 6 distinct isoforms (TrCB1.1 to TrCB1.6). The isoforms TrCB1.5 and TrCB1.6 have the catalytic cysteine substituted by glycin and are thus inactive, and were previously reported to be the least abundant members of all six known isoforms [85]. Surprisingly, the present transcriptomic data indicate that the two genes encoding TrCB1.5 and TrCB1.6 have the highest transcription of any cathepsin B detected here in *T. regenti*. In metazoan organisms, the existence of inactive enzyme forms (also of peptidases) is quite common, and the percentage of expression of inactive/active isoforms usually ranges from 10% to 15%; inactive paralogs of peptidases are speculated to have a regulatory function, achieved by a competition with active peptidase forms for their substrate(s) [86]. Also present among the highly transcribed peptidases was cathepsin B2 (TrCB2; RSEM-expected counts: 4,386). Cathepsins B2 as well as B1 were previously identified as dominant peptidases produced by *T. regenti* schistosomula, with a major capacity to degrade myelin basic protein, one of the principal components of spinal cord tissues, whereas haemoglobinolytic activity was negligible [20,85,87]. These findings may explain the ability of schistosomula to migrate through the white matter of the spinal cord, composed predominantly of neuronal axons, and cause axonal damage (cf. [26]).

Abundant transcription (RSEM-expected counts: 12,349) in *T. regenti* was also linked to cathepsin L. This enzyme has broad specificity in various parasitic flatworms, and can assume various functions. Importantly, it can cleave a range of substrates, including collagen, laminin, fibronectin, haemoglobin and immunoglobulins [88–90], and is likely intimately involved in the migration of *T. regenti* schistosomula and their ability to modulate or evade host immune attack. The immunomodulatory effect of proteolytic enzymes has been investigated in other schistosomes, and the aspartic peptidase, cathepsin D, is recognized as a key representative that cleaves host IgG from the tegumental surface of adult worms [91]. Besides this immunomodulatory role, cathepsin D is known to participate in the primary cleavage of haemoglobin [92]. In the present study, cathepsin D was represented by nine highly up-regulated transcripts in the schistosomula of *T. regenti*. As schistosomula of this species do not feed on blood, this enzyme in this developmental stage might assume an imunomodulatory role, but this hypothesis needs to be tested. In summary, the high level of transcription associated with cathepsins in the schistosomulum of *T. regenti* provides further support for the significance of this group of enzymes in parasitic flatworms. In addition to cathepsins, schistosomula also transcribe genes encoding a broad array of proteolytic enzymes that participate in numerous biological processes and likely have key roles during the parasite's invasion, survival and longevity in the definitive host.

Lysosomal proteins. In addition to the cysteine peptidases discussed, a variety of other proteins linked to lipid processing, including lysophospholipase III, Niemann-Pick C2 protein, palmitoyl-protein thioesterase and saposin-like proteins (SAPs), were encoded in the transcriptome representing the schistosomulum stage of *T. regenti*. Palmitoyl-protein thioesterase is also encoded in the genomes of human schistosomes and *C. sinensis* [52–54,93] and is likely involved

in the degradation of lipoproteins [94]. In relation to lipid trafficking, one transcript encoding a lipid-binding Niemann-Pick C2 protein, associated with cholesterol trafficking from the lysosome [95], was transcribed exclusively in *T. regenti* schistosomula. Genes encoding homologous proteins have been identified in the genomes of other trematodes, such as *S. mansoni*, *S. japonicum*, *S. haematobium*, *C. sinensis* and *O. viverrini* [52–54,93,96]. In addition, 10 different transcripts encoding SAPs were identified; SAPs can play diverse functional roles often through their interaction with lipids, lipid-degrading enzymes and lipid-presenting molecules [97], which may involve the activation of lipid-degrading enzymes, such as sphingolipid glycohydrolase [98]. Alternatively, SAPs might directly distrupt lipid membranes via the formation of pore-forming structures [99]. These proteins have been reported for trematodes including *S. mansoni* and *Fasciola* spp. and are presumed to facilitate the degradation of ingested host cells [100,101]. The high transcription of genes coding for proteins that interact with or bind to lipids might associate with the molecular machinery that *T. regenti* uses to degrade and then digest lipid-rich neural tissue, likely being critical for parasite migration and nutrition.

Interestingly, 38 different transcripts were predicted to encode hexosaminidase B, 30 were up-regulated in the schistosomula and 20 were uniquely transcribed in this stage. The analysis of enriched protein classes and pathways encoded by genes in schistosomula was biased by these different transcripts, which represent multiple KEGG classes, such that our interpretation is guarded. Although the function of hexosaminidase B in *T. regenti* is presently enigmatic, this enzyme is encoded in the genomes of human schistosomes [52–54] and *O. viverrini* [96]. We suspect that the high number of transcripts encoding this enzyme in *T. regenti* schistosomula might relate to multiple functions in this developmental stage. Based on independent evidence from humans, hexosaminidase can degrade sphingolipids (gangliosides) in neural tissue [102], and mutation of the hexosaminidase genes can lead to lethal neurodegenerative disorders (Tay-Sachs and Sandhoff or lysosomal storage disease) which are caused by an accumulation of gangliosides in the nervous system [103,104]. We suggest that an overexpression of hexosaminidase in *T. regenti* schistosomula leads to a degradation of neural tissue during their migration. In humans, only the alpha-subunit of hexosaminidase B is known to be involved in lipid degradation. As a corresponding alpha-subunit appears not to be encoded in *T. regenti*, the role of the hexosaminidase B should be explored. Although very little is known about hexosaminidases of helminths, homologs appear to have one or more roles in parasite-host interactions, for example, in the parasitic nematode *Trichinella spiralis* due to its specific sugar-binding property (lectin-like activity) or its specific glycohydrolase catalytic activity [105].

In conclusion, we describe here the first molecular exploration of neuropathogenic *Trichobilharzia* of birds and identify key biological pathways and proteins central to the invasion of the avian host and migration and development within neural tissues. Of particular significance is that we have been able to: (i) dissect the molecular differences between the cercarial stage (during cutaneous penetration) and the schistosomula stage (during neural migration and establishment within the avian host) and (ii) establish some biological distinctions between neuropathogenic *T. regenti* of birds and related blood flukes (schistosomes) of humans. The present annotated transcriptome for *T. regenti* provides a useful resource for comparative studies of schistosomatids and other trematodes, and should also underpin future genomic and proteomic investigations of *T. regenti*.

Supporting Information

S1 Table. Summary of raw reads, trimmed reads, trimmed and normalized reads for the transcriptomic data of cercariae and schistosomula of *Trichobilharzia regenti*. Results of mapping of paired, trimmed, corrected, non-normalized reads from cercariae and

schistosomula to the final non-redundant transcriptome of *Trichobilharzia regent*i. P1—pair read 1; P2—pair read 2; STD—Standard deviation
(XLSX)

S2 Table. Summary of the classification of predicted proteins from transcriptome of cercariae and schistosomula of *Trichobilharzia regenti* based on homology (BLASTp; *E*-value $\leq 10^{-5}$) to annotated proteins in the Kyoto Encyclopedia of Genes and Genomes (KEGG) BRITE functional hierarchies database, and pathway maps for cellular and organismal functions.
(XLSX)

S3 Table. Summary of peptidases and their inhibitors encoded in the transcriptome representing *Trichobilharzia regenti*, classified into families and subfamilies, with type enzyme listed in the MEROPS peptidase database. UP—Unassigned peptidase
(XLSX)

S4 Table. Orphan transcripts of the combined cercaria/schistosomulum transcriptome of Trichobilharzia regenti with no homology to other trematodes and classified based on NCBI and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases.
(XLSX)

S5 Table. Transcripts of *Trichobilharzia regenti* predicted to encode excretory/secretory (ES) proteins, annotated using the National Center for Biotechnology Information (NCBI) database. Transcription level established based on RNA-Seq by Expectation Maximization (RSEM) analysis. Sorted in descending order based on "expected counts" for the schistosomula stage. Differential transcripts are marked with an asterisk.
(XLSX)

S6 Table. Summary of transcripts uniquely transcribed in each cercaria and schistosomulum of *Trichobilharzia regenti* based on read mapping to the (consensus) transcriptome. Annotation using the National Center for Biotechnology Information (NCBI) database, and level of transcription based on RNA-Seq by Expectation Maximization (RSEM) analysis.
(XLSX)

S7 Table. Transcripts uniquely transcribed in either the cercaria or schistosomulum of *Trichobilharzia regenti*, classified using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.
(XLSX)

S8 Table. Summary of transcripts upregulated in the cercaria stage of *Trichobilharzia regenti* based on read mapping to the (consensus) transcriptome with National Center for Biotechnology Information (NCBI) annotation. Sorted in descending based on log2-fold change of differential transcription.
(XLSX)

S9 Table. Summary of transcripts upregulated in the schistosomulum stage of *Trichobilharzia regenti* based on read mapping to (consensus) transcriptome, using National Center for Biotechnology Information (NCBI) annotation. Sorted in descending order based on log2-fold change of differential transcription.
(XLSX)

S10 Table. Summary of transcripts enriched in the cercaria stage of *Trichobilharzia regenti* based on Kyoto Encyclopedia of Genes and Genomes (KEGG) BRITE/pathway and gene

ontology (GO) classifications. Transcripts uniquely transcribed in the cercaria are marked with an asterisk.
(XLSX)

S11 Table. Summary of transcripts enriched in the schistosomulum stage of *Trichobilharzia regenti* based on Kyoto Encyclopedia of Genes and Genomes (KEGG) BRITE/pathway and gene ontology (GO) classifications. Transcripts uniquely transcribed in the schistosomulum are marked with an asterisk.
(XLSX)

Author Contributions

Conceived and designed the experiments: MK NDY RL. Performed the experiments: RL. Analyzed the data: NDY RL PKK RSH. Contributed reagents/materials/analysis tools: PH NDY LM PT RBG. Wrote the paper: RL NDY.

References

1. Horak P, Kolarova L, Dvorak J. Trichobilharzia regenti n. sp. (Schistosomatidae, Bilharziellinae), a new nasal schistosome from Europe. Parasite. 1998; 5: 349–357. PMID: 9879557
2. Horak P, Kolarova L, Adema CM. Biology of the schistosome genus *Trichobilharzia*. Adv Parasitol. 2002; 52: 155–233. PMID: 12521261
3. Jouet D, Skirnisson K, Kolarova L, Ferte H. Molecular diversity of Trichobilharzia franki in two intermediate hosts (Radix auricularia and Radix peregra): a complex of species. Infect Genet Evol. 2010; 10: 1218–1227. doi: 10.1016/j.meegid.2010.08.001 PMID: 20708105
4. Korsunenko A V, Chrisanfova GG, Ryskov AP, Movsessian SO, Vasilyev VA, Semyenova SK. Detection of European Trichobilharzia schistosomes (T. franki, T. szidati, and T. regenti) based on novel genome sequences. J Parasitol. 2010; 96: 802–806. doi: 10.1645/GE-2297.1 PMID: 20677938
5. Davis NE. Identification of an avian schistosome recovered from Aythya novaeseelandia and infectivity of its miracidia to Lymnaea tomentosa snails. J Helminthol. 2006; 80: 225–233. PMID: 16923264
6. Gohardehi S, Fakhar M, Madjidaei M. Avian schistosomes and human cercarial dermatitis in a wildlife refuge in Mazandaran Province, northern Iran. Zoonoses Public Health. 2013; 60: 442–447. doi: 10.1111/zph.12020 PMID: 23121919
7. Horak P, Dvorak J, Kolarova L, Trefil L. *Trichobilharzia regenti*, a pathogen of the avian and mammalian central nervous systems. Parasitology. 1999; 119 (Pt 6): 577–581. PMID: 10633919
8. Hradkova K, Horak P. Neurotropic behaviour of *Trichobilharzia regenti* in ducks and mice. J Helminthol. 2002; 76: 137–141. doi: 10.1079/JOH2002113 PMID: 12015826
9. Chanová M, Bulantová J, Máslo P, Horák P. In vitro cultivation of early schistosomula of nasal and visceral bird schistosomes (Trichobilharzia spp., Schistosomatidae). Parasitol Res. 2009; 104: 1445–52. doi: 10.1007/s00436-009-1343-y PMID: 19238442
10. Horák P, Kolářová L. Snails, waterfowl and cercarial dermatitis. Freshw Biol. Blackwell Publishing Ltd; 2011; 56: 779–790. doi: 10.1111/j.1365-2427.2010.02545.x
11. Soldanova M, Selbach C, Kalbe M, Kostadinova A, Sures B. Swimmer's itch: etiology, impact, and risk factors in Europe. Trends Parasitol. 2013; 29: 65–74. doi: 10.1016/j.pt.2012.12.002 PMID: 23305618
12. Horak P, Mikes L, Lichtenbergova L, Skala V, Soldanova M, Brant SV. Avian schistosomes and outbreaks of cercarial dermatitis. Clin Microbiol Rev. 2015; 28: 165–190. doi: 10.1128/CMR.00043-14 PMID: 25567226
13. Jelinek, Nothdurft, Loscher. Schistosomiasis in Travelers and Expatriates. J Travel Med. 1996; 3: 160–164.
14. Kourilová P, Kolárová L. Variations in immunofluorescent antibody response against *Trichobilharzia* and Schistosoma antigens in compatible and incompatible hosts. Parasitol Res. 2002; 88: 513–21. doi: 10.1007/s00436-002-0607-6 PMID: 12107473
15. Neuhaus W. [Biology and development of Trichobilharzia Szidati N. Sp. (Trematoda, Schistosmatidae), a parasite causing dermatitis in man]. Z Parasitenkd. Not Available; 1952; 15: 203–266. PMID: 14951656

16. Lawson JR, Wilson R a. The survival of the cercariae of Schistosoma mansoni in relation to water temperature and glycogen utilization. Parasitology. 1980. pp. 337–348. doi: 10.1017/S0031182000056079 PMID: 7443297

17. Ligasova A, Bulantova J, Sebesta O, Kasny M, Koberna K, Mikes L. Secretory glands in cercaria of the neuropathogenic schistosome Trichobilharzia regenti—ultrastructural characterization, 3-D modelling, volume and pH estimations. Parasit Vectors. 2011; 4: 162. doi: 10.1186/1756-3305-4-162 PMID: 21854564

18. Mikes L, Zìdková L, Kasný M, Dvorák J, Horák P. In vitro stimulation of penetration gland emptying by Trichobilharzia szidati and T. regenti (Schistosomatidae) cercariae. Quantitative collection and partial characterization of the products. Parasitol Res. 2005; 96: 230–41. doi: 10.1007/s00436-005-1347-1 PMID: 15868186

19. Kasny M, Mikes L, Hampl V, Dvorak J, Caffrey CR, Dalton JP, et al. Chapter 4. Peptidases of trematodes. Advances in parasitology. 2009. pp. 205–297. doi: 10.1016/S0065-308X(09)69004-7 PMID: 19622410

20. Doleckova K, Kasny M, Mikes L, Cartwright J, Jedelsky P, Schneider EL, et al. The functional expression and characterisation of a cysteine peptidase from the invasive stage of the neuropathogenic schistosome *Trichobilharzia regenti*. Int J Parasitol. 2009; 39: 201–211. doi: 10.1016/j.ijpara.2008.06.010 PMID: 18708063

21. Horak P, Kovar L, Kolarova L, Nebesarova J. Cercaria-schistosomulum surface transformation of *Trichobilharzia szidati* and its putative immunological impact. Parasitology. 1998; 116 (Pt 2): 139–147. PMID: 9509023

22. Mclaren DJ, Hockley DJ. Blood flukes have a double outer membrane. Nature.; 1977; 269: 147–149. PMID: 71658

23. Horemans AM, Tielens AG, van den Bergh SG. The reversible effect of glucose on the energy metabolism of *Schistosoma mansoni* cercariae and schistosomula. Mol Biochem Parasitol. 1992; 51: 73–79. PMID: 1565140

24. Skelly PJ, Stein LD, Shoemaker CB. Expression of *Schistosoma mansoni* genes involved in anaerobic and oxidative glucose metabolism during the cercaria to adult transformation. Mol Biochem Parasitol. 1993; 60: 93–104. doi: 10.1016/0166-6851(93)90032-S PMID: 8396206

25. Parker-Manuel SJ, Ivens AC, Dillon GP, Wilson RA. Gene expression patterns in larval Schistosoma mansoni associated with infection of the mammalian host. PLoS Negl Trop Dis. 2011; 5: e1274. doi: 10.1371/journal.pntd.0001274 PMID: 21912711

26. Lichtenbergova L, Lassmann H, Jones MK, Kolarova L, Horak P. *Trichobilharzia regenti*: host immune response in the pathogenesis of neuroinfection in mice. Exp Parasitol. 2011; 128: 328–335. doi: 10.1016/j.exppara.2011.04.006 PMID: 21554878

27. Chanova M, Horak P. Terminal phase of bird schistosomiasis caused by Trichobilharzia regenti (Schistosomatidae) in ducks (Anas platyrhynchos f. domestica). Folia Parasitol (Praha). 2007; 54: 105–107.

28. Kolarova L, Horak P, Cada F. Histopathology of CNS and nasal infections caused by *Trichobilharzia regenti* in vertebrates. Parasitol Res. 2001; 87: 644–650. PMID: 11511002

29. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30: 2114–2120. doi: 10.1093/bioinformatics/btu170 PMID: 24695404

30. FastQC. A quality control tool for high throughput sequence data. Babraham Bioinforma Web site http//www.bioinformatics.babraham.ac.uk/projects/fastqc/.

31. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol. 2013; 20: 714–737. doi: 10.1089/cmb.2013.0084 PMID: 24093227

32. Brown CT, Howe A, Zhang Q, Pyrkosz AB, Brom TH, Lansing E, et al. A Reference-Free Algorithm for Computational Normalization arXiv : 1203. 4802v2 [q-bio. GN] 21 May 2012.: 1–18.

33. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics. 2012; 28: 1086–1092. doi: 10.1093/bioinformatics/bts094 PMID: 22368243

34. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012. pp. 3150–3152. doi: 10.1093/bioinformatics/bts565 PMID: 23060610

35. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013; 8: 1494–512. doi: 10.1038/nprot.2013.084 PMID: 23845962

36. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007; 23: 1061–1067. doi: 10.1093/bioinformatics/btm071 PMID: 17332020

37. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2014; 42: D7–17. doi: 10.1093/nar/gkt1146 PMID: 24259429

38. Smit AFA, Hubley R GP. RepeatMasker Open-3.0. [http://www.repeatmasker.org] webcite.

39. Schwarz EM, Korhonen PK, Campbell BE, Young ND, Jex AR, Jabbar A, et al. The genome and developmental transcriptome of the strongylid nematode *Haemonchus contortus*. Genome Biol. 2013; 14: R89. doi: 10.1186/gb-2013-14-8-r89 PMID: 23985341

40. Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford). 2011; 2011: bar009. doi: 10.1093/database/bar009

41. Rawlings ND, Barrett AJ, Bateman A. MEROPS: the peptidase database. Nucleic Acids Res. 2010; 38: D227–33. doi: 10.1093/nar/gkp971 PMID: 19892822

42. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28: 27–30. PMID: 10592173

43. Zdobnov EM, Apweiler R. InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 2001; 17: 847–848. PMID: 11590104

44. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011; 6: e21800. doi: 10.1371/journal.pone.0021800 PMID: 21789182

45. Blum T, Briesemeister S, Kohlbacher O. MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. BMC Bioinformatics. 2009; 10: 274. doi: 10.1186/1471-2105-10-274 PMID: 19723330

46. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12: 323. doi: 10.1186/1471-2105-12-323 PMID: 21816040

47. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26: 139–140. doi: 10.1093/bioinformatics/btp616 PMID: 19910308

48. R Development Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria; 2008. Available: http://www.r-project.org

49. Risso D, Schwartz K, Sherlock G, Dudoit S. GC-content normalization for RNA-Seq data. BMC Bioinformatics. 2011; 12: 480. doi: 10.1186/1471-2105-12-480 PMID: 22177264

50. Dillies M-A, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief Bioinform. 2013; 14: 671–683. doi: 10.1093/bib/bbs046 PMID: 22988256

51. Alexa A and Rahnenfuhrer J. topGO: Enrichment analysis for Gene Ontology. R package version 2.20.0. [Internet]. 2010. Available: http://www.bioconductor.org/packages/release/bioc/html/topGO.html

52. Young ND, Jex AR, Li B, Liu S, Yang L, Xiong Z, et al. Whole-genome sequence of *Schistosoma haematobium*. Nat Genet. 2012; 44: 221–225. doi: 10.1038/ng.1065 PMID: 22246508

53. Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, et al. The genome of the blood fluke *Schistosoma mansoni*. Nature. 2009; 460: 352–358. doi: 10.1038/nature08160 PMID: 19606141

54. The Schistosoma japonicum genome reveals features of host-parasite interplay. Nature. Macmillan Publishers Limited. All rights reserved; 2009; 460: 345–351. Available: doi: 10.1038/nature08140

55. Kasny M, Mikes L, Dalton JP, Mountford AP, Horak P. Comparison of cysteine peptidase activities in *Trichobilharzia regenti* and *Schistosoma mansoni* cercariae. Parasitology. 2007; 134: 1599–1609. doi: 10.1017/S0031182007002910 PMID: 17517170

56. Doleckova K, Kasny M, Mikes L, Mutapi F, Stack C, Mountford AP, et al. Peptidases of Trichobilharzia regenti (Schistosomatidae) and its molluscan host Radix peregra S. Lat. (Lymnaeidae): construction and screening of cDNA library from intramolluscan stages of the parasite. Folia Parasitol (Praha). 2007; 54: 94–98.

57. Dvorák J, Mashiyama ST, Braschi S, Sajid M, Knudsen GM, Hansell E, et al. Differential use of protease families for invasion by schistosome cercariae. Biochimie. 2008; 90: 345–58. doi: 10.1016/j.biochi.2007.08.013 PMID: 17936488

58. Curwen RS, Wilson RA. Invasion of skin by schistosome cercariae: some neglected facts. Trends Parasitol. 2003; 19: 63–68. PMID: 12586470

59. Salter JP, Choe Y, Albrecht H, Franklin C, Lim K-C, Craik CS, et al. Cercarial elastase is encoded by a functionally conserved gene family across multiple species of schistosomes. J Biol Chem. 2002; 277: 24618–24624. doi: 10.1074/jbc.M202364200 PMID: 11986325

60. Gray DJ, Williams GM, Li Y, Chen H, Forsyth S, Li R, et al. The role of bovines in human *Schistosoma japonicum* infection in the Peoples' Republic of China. Am J Trop Med Hyg. 2009; 81: 301–301.

61. Jolly ER, Chin C-S, Miller S, Bahgat MM, Lim KC, DeRisi J, et al. Gene expression patterns during adaptation of a helminth parasite to different environmental niches. Genome Biol. 2007; 8: R65. doi: 10.1186/gb-2007-8-4-r65 PMID: 17456242

62. Gobert GN, Moertel L, Brindley PJ, McManus DP. Developmental gene expression profiles of the human pathogen *Schistosoma japonicum*. BMC Genomics. 2009; 10: 128. doi: 10.1186/1471-2164-10-128 PMID: 19320991

63. Verjovski-Almeida S, DeMarco R, Martins E a L, Guimarães PEM, Ojopi EPB, Paquola ACM, et al. Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. Nat Genet. 2003; 35: 148–157. doi: 10.1038/ng1237 PMID: 12973350

64. Tielens AG, van den Heuvel JM, van den Bergh SG. The energy metabolism of *Fasciola hepatica* during its development in the final host. Mol Biochem Parasitol. 1984; 13: 301–307. PMID: 6527693

65. Boyunaga H, Schmitz MG, Brouwers JF, Van Hellemond JJ, Tielens AG. *Fasciola hepatica* miracidia are dependent on respiration and endogenous glycogen degradation for their energy generation. Parasitology. 2001; 122: 169–173. PMID: 11272647

66. Prosdocimi F, Faria-Campos AC, Peixoto FC, Pena SDJSDJ, Ortega JMJM, Franco GRGR. Clustering of *Schistosoma mansoni* mRNA sequences and analysis of the most transcribed genes: Implications in metabolism and biology of different developmental stages. Mem Inst Oswaldo Cruz. 2002; 97: 61–69. doi: 10.1590/S0074-02762002000900014

67. Tielens AG, Van der Meer P, van den Heuvel JM, van den Bergh SG. The enigmatic presence of all gluconeogenic enzymes in *Schistosoma mansoni* adults. Parasitology. 1991; 102 Pt 2: 267–276. PMID: 1649428

68. Burenina EA. [Properties of gluconeogenesis enzymes from flatworms]. Zh Evol Biokhim Fiziol. 2001; 37: 85–91. PMID: 11452789

69. Van Oordt BE, Tielens AG, Van den Bergh SG. Aerobic to anaerobic transition in the carbohydrate metabolism of *Schistosoma mansoni* cercariae during transformation in vitro. Parasitology. 1989; 98 Pt 3: 409–415. PMID: 2771447

70. Tielens A. G. M., Hellemond van JJ. Unusual aspects of metabolism in flatworm parasites. Parasitic flatworms: molecular biology, biochemistry, immunology and physiology. CAB International; 2006. doi: 10.1079/9780851990279.0387

71. Barrett J. Amino acid metabolism in helminths. Adv Parasitol. ENGLAND; 1991; 30: 39–105.

72. Santos TM, Johnston D a, Azevedo V, Ridgers IL, Martinez MF, Marotta GB, et al. Analysis of the gene expression profile of *Schistosoma mansoni* cercariae using the expressed sequence tag approach. Mol Biochem Parasitol. 1999; 103: 79–97. PMID: 10514083

73. Ram D, Grossman Z, Markovics A, Avivi A, Ziv E, Lantner F, et al. Rapid changes in the expression of a gene encoding a calcium-binding protein in *Schistosoma mansoni*. Mol Biochem Parasitol. 1989; 34: 167–175. PMID: 2710168

74. Dorsey CH, Stirewalt MA. Schistosoma mansoni: localization of calcium-detecting reagents in electron-lucent areas of specific preacetabular gland granules. Z Parasitenkd. 1977; 54: 165–173. PMID: 605648

75. Modha J, Redman CA, Thornhill JA, Kusel JR. Schistosomes: unanswered questions on the basic biology of the host-parasite relationship. Parasitol Today. 1998; 14: 396–401. PMID: 17040829

76. Dresden MH, Edlin EM. *Schistosoma mansoni*: effect of some cations on the proteolytic enzymes of cercariae. Exp Parasitol. 1974; 35: 299–303. PMID: 4206936

77. McKerrow JH, Jones P, Sage H, Pino-Heiss S. Proteinases from invasive larvae of the trematode parasite *Schistosoma mansoni* degrade connective-tissue and basement-membrane macromolecules. Biochem J. 1985; 231: 47–51. PMID: 3904737

78. Siddiqui AA, Zhou Y, Podesta RB, Karcz SR, Tognon CE, Strejan GH, et al. Characterization of Ca(2+)-dependent neutral protease (calpain) from human blood flukes, Schistosoma mansoni. Biochim Biophys Acta. 1993; 1181: 37–44. PMID: 8457603

79. Rowe JA, Claessens A, Corrigan RA, Arman M. Adhesion of *Plasmodium falciparum*-infected erythrocytes to human cells: molecular mechanisms and therapeutic implications. Expert Rev Mol Med. 2009; 11: e16. doi: 10.1017/S1462399409001082 PMID: 19467172

80. Gumbiner BM. Cell adhesion: the molecular basis of tissue architecture and morphogenesis. Cell. 1996; 84: 345–357. PMID: 8608588

81. Larrivee B, Freitas C, Suchting S, Brunet I, Eichmann A. Guidance of vascular development: lessons from the nervous system. Circ Res. 2009; 104: 428–441. doi: 10.1161/CIRCRESAHA.108.188144 PMID: 19246687

82. Ou C-Y, Shen K. Setting up presynaptic structures at specific positions. Curr Opin Neurobiol. 2010; 20: 489–493. doi: 10.1016/j.conb.2010.04.011 PMID: 20471244

83. Skelly PJ, Shoemaker CB. Induction cues for tegument formation during the transformation of *Schistosoma mansoni* cercariae. Int J Parasitol. 2000; 30: 625–631. PMID: 10779576

84. Caffrey CR, McKerrow JH, Salter JP, Sajid M. Blood "n" guts: an update on schistosome digestive peptidases. Trends Parasitol. 2004; 20: 241–8. doi: 10.1016/j.pt.2004.03.004 PMID: 15105025

85. Dvorak J, Delcroix M, Rossi A, Vopalensky V, Pospisek M, Sedinova M, et al. Multiple cathepsin B isoforms in schistosomula of *Trichobilharzia regenti*: identification, characterisation and putative role in migration and nutrition. Int J Parasitol. 2005; 35: 895–910. doi: 10.1016/j.ijpara.2005.02.018 PMID: 15950230

86. Pils B, Schultz J. Inactive enzyme-homologues find new function in regulatory processes. J Mol Biol. 2004; 340: 399–404. doi: 10.1016/j.jmb.2004.04.063 PMID: 15210342

87. Dolecková K, Albrecht T, Mikes L, Horák P. Cathepsins B1 and B2 in the neuropathogenic schistosome *Trichobilharzia regenti*: distinct gene expression profiles and presumptive roles throughout the life cycle. Parasitol Res. 2010; 107: 751–5. doi: 10.1007/s00436-010-1943-6 PMID: 20556428

88. Berasaín P, Goñi F, McGonigle S, Dowd a, Dalton JP, Frangione B, et al. Proteinases secreted by *Fasciola hepatica* degrade extracellular matrix and basement membrane components. J Parasitol. 1997; 83: 1–5. Available: http://www.ncbi.nlm.nih.gov/pubmed/9057688 PMID: 9057688

89. Collins PR, Stack CM, O'Neill SM, Doyle S, Ryan T, Brennan GP, et al. Cathepsin L1, the major protease involved in liver fluke (Fasciola hepatica) virulence: propetide cleavage sites and autoactivation of the zymogen secreted from gastrodermal cells. J Biol Chem. 2004; 279: 17038–17046. doi: 10.1074/jbc.M308831200 PMID: 14754899

90. Yamakami K, Hamajima F, Akao S, Tadakuma T. Purification and characterization of acid cysteine protease from metacercariae of the mammalian trematode parasite *Paragonimus westermani*. Eur J Biochem. 1995; 233: 490–497. PMID: 7588793

91. Verity CK, Loukas A, McManus DP, Brindley PJ. *Schistosoma japonicum* cathepsin D aspartic protease cleaves human IgG and other serum components. Parasitology. 2001; 122: 415–421. PMID: 11315174

92. Delcroix M, Sajid M, Caffrey CR, Lim K-C, Dvorák J, Hsieh I, et al. A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. J Biol Chem. 2006; 281: 39316–29. doi: 10.1074/jbc.M607128200 PMID: 17028179

93. Wang X, Chen W, Huang Y, Sun J, Men J, Liu H, et al. The draft genome of the carcinogenic human liver fluke *Clonorchis sinensis*. Genome Biol. 2011; 12: R107. doi: 10.1186/gb-2011-12-10-r107 PMID: 22023798

94. Vesa J, Hellsten E, Makela TP, Jarvela I, Airaksinen T, Santavuori P, et al. A single PCR marker in strong allelic association with the infantile form of neuronal ceroid lipofuscinosis facilitates reliable prenatal diagnostics and disease carrier identification. Eur J Hum Genet. 1993; 1: 125–132. PMID: 7914464

95. Chikh K, Vey S, Simonot C, Vanier MT, Millat G. Niemann-Pick type C disease: importance of N-glycosylation sites for function and cellular location of the NPC2 protein. Mol Genet Metab. 2004; 83: 220–230. PMID: 15542393

96. Young ND, Nagarajan N, Lin SJ, Korhonen PK, Jex AR, Hall RS, et al. The *Opisthorchis viverrini* genome provides insights into life in the bile duct. Nat Commun. 2014; 5: 4378. doi: 10.1038/ncomms5378 PMID: 25007141

97. Bruhn H. A short guided tour through functional and structural features of saposin-like proteins. Biochem J. 2005; 389: 249–257. doi: 10.1042/BJ20050051 PMID: 15992358

98. Furst W, Sandhoff K. Activator proteins and topology of lysosomal sphingolipid catabolism. Biochim Biophys Acta. 1992; 1126: 1–16. PMID: 1606169

99. Leippe M, Tannich E, Nickel R, van der Goot G, Pattus F, Horstmann RD, et al. Primary and secondary structure of the pore-forming peptide of pathogenic *Entamoeba histolytica*. EMBO J. 1992; 11: 3501–3506. PMID: 1396552

100. Don TA, Bethony JM, Loukas A. Saposin-like proteins are expressed in the gastrodermis of *Schistosoma mansoni* and are immunogenic in natural infections. Int J Infect Dis. 2008; 12: e39–47. doi: 10.1016/j.ijid.2007.10.007 PMID: 18571965

101. Espino AM, Hillyer G V. Molecular cloning of a member of the *Fasciola hepatica* saposin-like protein family. J Parasitol. 2003; 89: 545–552. doi: 10.1645/GE-3113 PMID: 12880256

102. Mark BL, Mahuran DJ, Cherney MM, Zhao D, Knapp S, James MNG. Crystal structure of human beta-hexosaminidase B: understanding the molecular basis of Sandhoff and Tay-Sachs disease. J Mol Biol. 2003; 327: 1093–1109. PMID: 12662933

103. Sandhoff K, Harzer K. Gangliosides and gangliosidoses: principles of molecular and metabolic pathogenesis. J Neurosci. 2013; 33: 10195–10208. doi: 10.1523/JNEUROSCI.0822-13.2013 PMID: 23785136

104. Mahuran DJ. Biochemical consequences of mutations causing the GM2 gangliosidoses. Biochim Biophys Acta. 1999; 1455: 105–138. PMID: 10571007

105. Rhoads ML. Purification, characterization, and immunochemical studies of beta-N-acetyl-D-hexosaminidase from the parasitic nematode *Trichinella spiralis*. Mol Biochem Parasitol. 1988; 31: 57–69. PMID: 2972930

SCIENTIFIC REPORTS

OPEN

Molecular evidence for distinct modes of nutrient acquisition between visceral and neurotropic schistosomes of birds

Received: 14 August 2018
Accepted: 10 December 2018
Published online: 04 February 2019

Roman Leontovyč[1], Neil D. Young[2], Pasi K. Korhonen[2], Ross S. Hall[2], Jana Bulantová[1], Veronika Jeřábková[1], Martin Kašný[1], Robin B. Gasser[2] & Petr Horák[1]

Trichobilharzia **species are parasitic flatworms (called schistosomes or flukes) that cause important diseases in birds and humans, but very little is known about their molecular biology. Here, using a transcriptomics-bioinformatics-based approach, we explored molecular aspects pertaining to the nutritional requirements of *Trichobilharzia szidati* ('visceral fluke') and *T. regenti* ('neurotropic fluke') in their avian host. We studied the larvae of each species before they enter (cercariae) and as they migrate (schistosomules) through distinct tissues in their avian (duck) host. Cercariae of both species were enriched for pathways or molecules associated predominantly with carbohydrate metabolism, oxidative phosphorylation and translation of proteins linked to ribosome biogenesis, exosome production and/or lipid biogenesis. Schistosomules of both species were enriched for pathways or molecules associated with processes including signal transduction, cell turnover and motility, DNA replication and repair, molecular transport and/or catabolism. Comparative informatic analyses identified molecular repertoires (within, e.g., peptidases and secretory proteins) in schistosomules that can broadly degrade macromolecules in both *T. szidati and T. regenti*, and others that are tailored to each species to selectively acquire nutrients from particular tissues through which it migrates. Thus, this study provides molecular evidence for distinct modes of nutrient acquisition between the visceral and neurotropic flukes of birds.**

Parasitic flatworms of the family Schistosomatidae have complex biologies and life histories[1,2], and are usually significant pathogens[3]. Their life cycles involve aquatic snails as intermediate hosts, in which asexual reproduction takes place, and vertebrates such as mammals and birds as definitive hosts, in which sexual reproduction occurs[1,2].

On one hand, members of the genus *Schistosoma* infect humans and/or other mammals as definitive hosts. In humans, *Schistosoma mansoni*, *S. japonicum* and *S. haematobium* are the three predominant species and are significant pathogens in their own right. They affect ~230 million people worldwide and cause hepato-intestinal or urogenital illnesses, collectively referred to as schistosomiasis[4].

On the other hand, members of the genus *Trichobilharzia* infect birds as definitive hosts[2]. In water birds, *Trichobilharzia szidati* and *T. regenti* are key representatives that cause visceral and neurological forms of disease, respectively. Water birds release eggs containing larvae (miracidia) in the faeces (*T. szidati*) or miracidia from their nasal mucosa (*T. regenti*) into a freshwater environment. Here, motile miracidia seek out, find and infect a snail intermediate host, and undergo asexual reproduction, resulting in thousands of cercariae (invasive larvae) being shed into the water column. The cercariae find a definitive host (water bird) within 1 to 1.5 days[5]. Once they find this host, cercariae start to penetrate the skin using a cocktail of proteolytic enzymes (peptidases) and undergo morphological (metamorphosis) and molecular changes[6–8]. During penetration, cercariae lose their tail, shed their glycocalyx, form a tegumental double-membrane and switch from aerobic to facultative anaerobic metabolism[7,8]. Once inside the definitive host, transformed cercariae (schistosomula or schistosomules) of

[1]Department of Parasitology, Faculty of Science, Charles University, Prague, Czech Republic. [2]Department of Veterinary Biosciences, Melbourne Veterinary School, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria, 3010, Australia. Correspondence and requests for materials should be addressed to R.L. (email: leontovyc.roman@seznam.cz)

T. szidati and *T. regenti* use distinct migratory routes to establish infection in distinct tissues/organs and cause disease.

Trichobilharzia szidati is the species that causes visceral trichobilharziasis. The cercariae invade the avian host via skin, enter the circulatory system, and schistosomules reach the lungs within 2 days of infection and eventually develop to dioecious adults in the vessels of the intestinal wall, where they reproduce[9]. In the lungs, schistosomules can cause serious parasitic pneumonia, accompanied by lymphatic lesions and an influx of macrophages, heterophils and eosinophils, and/or death of the animal in severe cases[9]. Inside the host, *T. szidati* schistosomules evade the host's immune response and ingest blood to acquire nutrients[10], but the molecular basis for nutrient acquisition and their survival is unknown.

By contrast, *T. regenti* is the species that causes neural/nasal trichobilharziasis. Here, following cercarial penetration, schistosomules develop and migrate via the peripheral nerves, spinal cord and brain to the nasal mucosa, where the dioecious adults mate and reproduce. During their migration, schistosomules cause inflammation and neurodegenerative changes linked to motor neuronal malfunctions, and/or can lead to death of the animal, depending on the intensity of infection[11–13]. *Trichobilharzia regenti* schistosomules appear to feed on neural tissues rather than on blood[14].

Despite of the differences in biology between *T. szidati* and *T. regenti*, and in the diseases that they cause, almost nothing is known at the molecular level about the difference(s) in their host-parasite relationship and the way in which their schistosomules feed, acquire nutrients and survive during the migration phase in the avian host. To date, most molecular biological studies have focused on human schistosomes[15–17]. This contrasts markedly the situation for other schistosomatids, for which the potential to explore transcriptomes, proteomes and gene function is only now being realised[18]. No nuclear genomic sequences have yet been published for *Trichobilharzia* species (except of draft genome of *T. regenti* (parasite.wormbase.org)), and transcriptomic and proteomic data are scant. However, a recent transcriptomic investigation provided a first glimpse of the molecular biology of development, adaptation and host invasion of *T. regenti*[18]. In the present study, were used previous[18] and new data sets to compare the molecular basis of nutritional requirements of *T. regenti* and *T. szidati* schistosomules in their avian host.

Results and Discussion

Transcriptomes and homology-based comparisons. The transcriptomes of *T. szidati* ($n = 13{,}007$) and *T. regenti* ($n = 12{,}705$) contained similar numbers of transcripts and annotated transcripts (Table 1) and were 80.5% to 81.5% complete. In total, 728 of the BUSCOs in the transcriptomes of *T. szidati* ($n = 787$) and *T. regenti* ($n = 797$) were shared. Most (80%) transcripts in the transcriptomes of *T. szidati* ($n = 10{,}498$) and *T. regenti* ($n = 11{,}114$) shared amino acid sequence homology with proteins assigned one or more KEGG terms. Almost half of the transcripts ($n = 5{,}663$ and 5,935) associated with 3,065 and 3,275 KEGG terms with conserved protein family annotations (KEGG BRITE), respectively (Table 1), and 3,452 and 3,611 transcripts associated with 1,837 and 1,934 KEGG terms with KEGG biological pathway annotations, respectively (Fig. 1, panel a; Supplementary Table S1). In total, more than 460 excretory/secretory proteins were predicted from the transcriptomes of *T. szidati* ($n = 642$) and *T. regenti* ($n = 468$) (Supplementary Table S2).

An homology-based analysis showed that 5,886 orthologous protein groups were shared by *T. szidati*, *T. regenti* and *S. mansoni* (Fig. 1, panel b). Neurotropic *T. regenti* shared twice as many orthologous groups ($n = 757$) with *S. mansoni* than did visceral *T. szidati* ($n = 391$). Analyses of orthologues shared by *T. regenti* and *S. mansoni* revealed 145 transcripts assigned to "chromosome, DNA repair and recombination proteins, ubiquitin system, DNA replication proteins, protein kinases, chaperons and folding catalysts"; 112 transcripts of these orthologues were not detected in *T. szidati*, indicating differential transcription between species (upon pairwise comparison) rather than phylogenetic divergence. Transcripts that were unique to *T. szidati* ($n = 601$) and *T. regenti* ($n = 765$) linked to 20 and 18 unique protein families in the KEGG BRITE database, respectively, being assigned to one or more specific KEGG terms including spliceosome (7 for *T. szidati* and 11 for *T. regenti*), exosome (6 and 9), chromosome (5 and 9), peptidase (4 and 4) and kinase (2 and 6) (Supplementary Table S3).

Differential transcription. Differential transcription was investigated between the free-living, infective (cercaria) and the migratory, parasitic (schistosomule) stages of *T. szidati* and *T. regenti*. In *T. szidati*, 3,729 of all 13,007 known transcripts were differentially transcribed between these two larval stages (1,881 up-regulated in the cercaria, and 1,848 up-regulated in the schistosomule). In *T. regenti*, 3,396 of all 12,705 known transcripts were differentially transcribed between these two stages (1,301 up-regulated in cercaria, and 1,876 up-regulated in schistosomule). For each species, differential transcripts were then linked to enriched biological pathways and protein families using the KEGG database, and the numbers of species- and stage-specific enriched biological pathways and protein families established (Figs 2 and 3).

Pathways and protein families enriched in cercariae. A total of 283 KEGG pathway terms (99 for *T. szidati* and 82 for *T. regenti*; 102 shared by both) and 200 KEGG BRITE protein family terms (75 for *T. szidati* and 55 for *T. regenti*; 70 shared) were enriched (Figs 2 and 3) in the cercariae of both *Trichobilharzia* species. Compared with the schistosomules, the results showed that cercariae usually had greater enrichment for pathways or protein families associated with metabolism and translation. The former (i.e. metabolism) related predominantly to carbohydrate metabolism (linked to conserved pathways or processes, such as Krebs cycle, glycolysis, pentose phosphate pathway and pyruvate metabolism), and energy metabolism was reflected in an enrichment of proteins linked to oxidative phosphorylation. Enriched metabolic pathways linked specifically to Krebs cycle, glycolysis and oxidative phosphorylation reflect the aerobic metabolism of free-living cercariae, contrary to the microaerobic metabolism of schistosomules. The latter (i.e. translation) related predominantly

	T. szidati	*T. regenti*[18]
Transcriptome		
Total sequences	13,007	12,705
Minimum and maximum sequence lengths; N50 (bp)	160-39,264; 3,402	115-41,111; 3,682
Predicted proteome		
Total sequences	13,007	12,705
Minimum and maximum sequence lengths; N50 (bp)	30-11,108; 671	30-8,133; 805
Matches to conserved metazoan BUSCO genes		
Complete single copy orthologous groups[a]	603 (61.7)	659 (67.4)
Complete, duplicated orthologous groups[a]	97 (9.9)	81 (8.3)
Fragmented orthologous groups[a]	87 (8.9)	57 (5.8)
Total BUSCO orthologous groups[a]	978 (100)	978 (100)
Protein annotation		
NCBI nr database[b]	10,457 (80.4)	10,900 (85.8)
SwissProt[b]	7,633 (58.7)	8,120 (63.9)
MEROPS peptidases[b]	357 (2.7)	392 (3.1)
MEROPS inhibitors[b]	249 (1.9)	259 (2.0)
KEGG BRITE protein families[c]	5,663 (43.5; 3,065)	5,935 (46.7; 3,275)
KEGG pathway[c]	3,452 (26.5; 1,837)	3,611 (28.4; 1,934)
InterProScan[b]	8,171 (62.8)	9,642 (75.9)
Gene ontology annotation (InterProScan)[b]	6,730 (51.7)	6,961 (54.8)
Proteins predicted to be excreted/secreted	642 (4.9)	468 (3.7)

Table 1. Characteristics of the *Trichobilharzia szidati* and *T. regenti* transcriptomes and annotation of their predicted proteomes. [a]Number of proteins homologous to BUSCO orthologous groups (% of total groups). [b]Number of proteins homologous to entries in database (% of predicted proteome). [c]Number of proteins homologous to entries in the KEGG database (% of predicted proteome; number of conserved KEGG terms).

to ribosomal proteins and proteins involved in ribosome biogenesis, proteins associated with exosomes, protein phosphatases, amino acid related proteins and lipid biogenesis, with *T. szidati* having a greater enrichment for ribosomal proteins than *T. regenti* (Supplementary Table S4).

Pathways and protein families enriched in schistosomules. A total of 208 (54 for *T. szidati* and 83 for *T. regenti*; 71 shared by both) KEGG pathway terms and 387 (108 for *T. szidati* and 153 for *T. regenti*; 126 shared) KEGG BRITE protein family terms were enriched in the schistosomules of both species (Figs 2 and 3). The main biological pathways enriched in *T. regenti* and *T. szidati* schistosomules linked to signalling, cell growth and death, cell motility, transport and catabolism, membrane transport, and DNA replication and repair. A distinction between the two species was observed in an enrichment for signal transduction pathways (linked to Ras, MAPK, Rap1, Wnt and ErbB) that was exclusive to *T. regenti*. Enriched protein families were represented by chromosome-associated proteins, proteins linked with transcription, transport system, cell adhesion, cytoskeleton proteins, DNA replication, protein kinases and peptidases (Supplementary Table S4), and MEG-encoded secretory proteins and saposins were conspicuous.

Microexon gene (MEG)-encoded secretory proteins. As MEG-encoded secretory proteins are proposed to be involved in blood processing in schistosomes[19,20], we focused our attention on transcripts linked to this particular gene family in *T. szidati* and *T. regenti*. We detected 10 and 4 transcripts encoding MEGs in *T. szidati* and *T. regenti*, respectively. An analysis showed that predicted secretory proteins encoded by MEG orthologs of *T. szidati* and *T. regenti* clustered into four different groups (Fig. 4). *Trichobilharzia regenti* transcripts were represented in all four groups, suggesting that MEG-encoded secretory proteins are not strictly linked to visceral schistosomiasis[19]. While *S. mansoni* MEGs were assigned previously to three groups (MEG-3, MEG- 4 and MEG-8), some orthologs of *T. szidati* and *T. regenti* were not closely related to known MEGs of *S. mansoni* (Fig. 4), suggesting that they are specific to avian *Trichobilharzia* species. In *T. szidati*, MEG homologs were highly transcribed, with two of them (Tszi_001069 and Tszi_004986) being in the top-10 most upregulated transcripts in schistosomules. In *T. regenti*, MEG genes were transcribed at markedly lower levels than those in *T. szidati* (Fig. 4, panel b). Collectively, these results suggest that there is a differential involvement of MEG-encoded secretory proteins in feeding by the two *Trichobilharzia* species.

Saposins. These lysosomal proteins can play a key role in the degradation of some sphingolipids[21]. Two transcripts (Treg_014592, Treg_015261) encoding saposin B were represented in the ten most upregulated transcripts in *T. regenti* schistosomules. Saposin B degrades sulfatides, a major component of the myelin sheaths of nerves[22]. The specific detection of myelin in the guts of *T. regenti* schistosomules suggests that saposin B is central to lipid degradation and digestion during their migration through the spinal cord. Saposin-encoding transcripts were also identified in *T. szidati* schistosomules, but at very low levels. A role for saposins in blood digestion has been proposed for *S. mansoni* and for the liver flukes *Clonorchis sinensis* and *Fasciola hepatica*[23–25]. Taken together, this

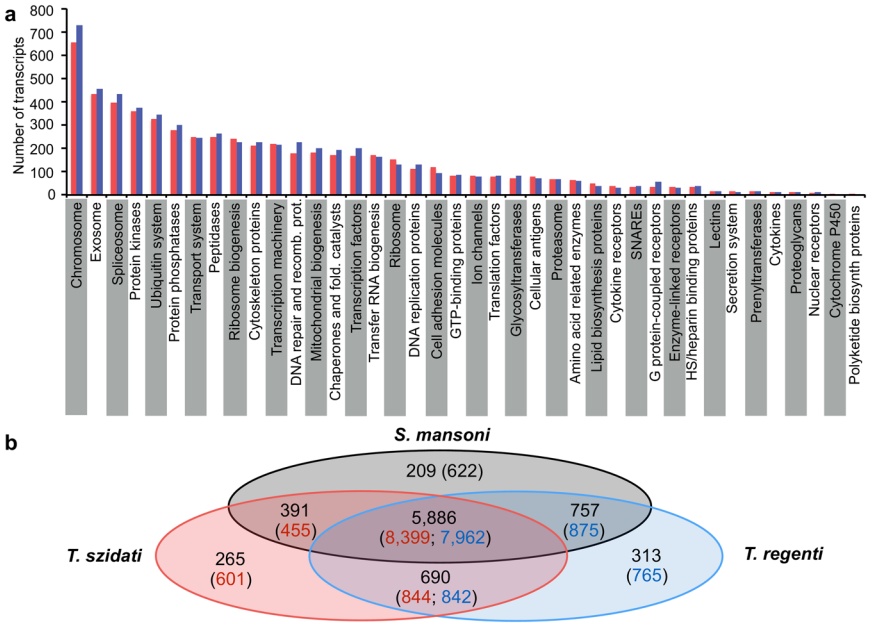

Figure 1. Annotation of protein-encoding genes transcribed in *Trichobilharzia szidati* and *T. regenti* larvae (cercaria combined with schistosomules) (**a**) Indicated are protein families represented in *T. szidati* (red) and *T. regenti* (blue) assigned *via* Kyoto Encyclopedia of Genes and Genomes (KEGG). Comparison of protein groups predicted from the larval transcriptomes among *T. szidati* and *T. regenti* and *Schistosoma mansoni* based on homology (OrthoMCL) (**b**) indicated are numbers of orthologous protein groups, followed by numbers of transcripts representing these groups (in brackets).

information suggests that saposins linked to the degradation of macromolecular substrates, depending on the nutrition requirements of a particular fluke.

Peptidases. Other evidence indicated a differing involvement of peptidases in the cercaria and schistosomule stages of *T. szidati* and *T. regenti*. Peptidases inferred to be enriched in the schistosomules of both species (Fig. 5; Supplementary Table S5) were principally the cathepsins (B, L, D, A, K and C), dipeptidyl-peptidase II, leucyl aminopeptidase, aspartyl peptidase, tissue plasminogen activator, legumain, calpain, tolkin, ADAMTS (a disintegrin and metalloproteinase with thrombospondin motifs) peptidases and carbamoyl-phosphate synthase. The transcription profiles representing these peptidases were similar between cercariae and schistosomules for both *Trichobilharzia* species (Fig. 5). Usually represented by relatively moderate to high numbers of transcripts ($n > 7$) were cathepsin B (11 for *T. szidati* and 12 for *T. regenti*), cathepsin D (11 and 8, respectively), cathepsin A (10 and 4, respectively) and calpain (9 and 7, respectively). Transcripts encoding calpain were abundant in both cercariae and schistosomules of both *T. szidati* and *T. regenti*, suggesting that this peptidase is multi-functional.

Following these analyses, we explored the abundance of transcripts encoding peptidases that were upregulated in the schistosomules of each species. In *T. szidati* schistosomules, transcripts encoding legumain were most abundant, followed equally by those encoding cathepsins L and B, followed by calpain, cathepsins D, K, C and A, leucyl aminopeptidase and dipeptidyl peptidase II (in order of decreasing abundance) (Fig. 5). In *T. regenti* schistosomules, transcripts representing cathepsin B were most abundant and twice as high as those representing cathepsin L, followed (in decreasing order) by transcripts representing cathepsins K and C, tissue plasminogen activator, leucyl aminopeptidase, dipeptidyl-peptidase II, cathepsin A, calpain and legumain. Transcripts representing selected peptidases were unique to *T. szidati* (i.e. disintegrin and metalloproteinase domain-containing protein 9) and to *T. regenti* (i.e. cathepsin S, caspase 7, disintegrin and metalloproteinase domain-containing protein 1 and anhydrolase domain-containing protein 5). These findings suggest that peptidases are the key enzymes involved the digestion and acquisition of nutrients from different macro-molecular substrates in schistosomules of *T. szidati* and *T. regenti*.

Although there were qualitative differences between the two species, the differential levels of transcription of genes encoding particular peptidases (taken as a snap-shot) indicate distinctive requirements of the schistosomules between the two species, to digest tissue types with differing macromolecular (protein) compositions (i.e. blood *versus* neural tissues). In particular, the high transcription of genes encoding cathepsins B and L and legumain (=haemoglobinase) are likely to be specifically linked to the degradation of blood[26,27], which accords with microscopic evidence showing the presence of blood metabolites specifically in the gut of schistosomules of *T. szidati*, but not *T. regenti* (Fig. 6). By contrast, genes encoding legumain (haemoglobinase) were transcribed at a substantially lower level in schistosomules of *T. regenti* compared with those of *T. szidati*. As peptidases are often multi-functional[28], it is possible that limiting amounts of legumain might activate (other) peptidases, such

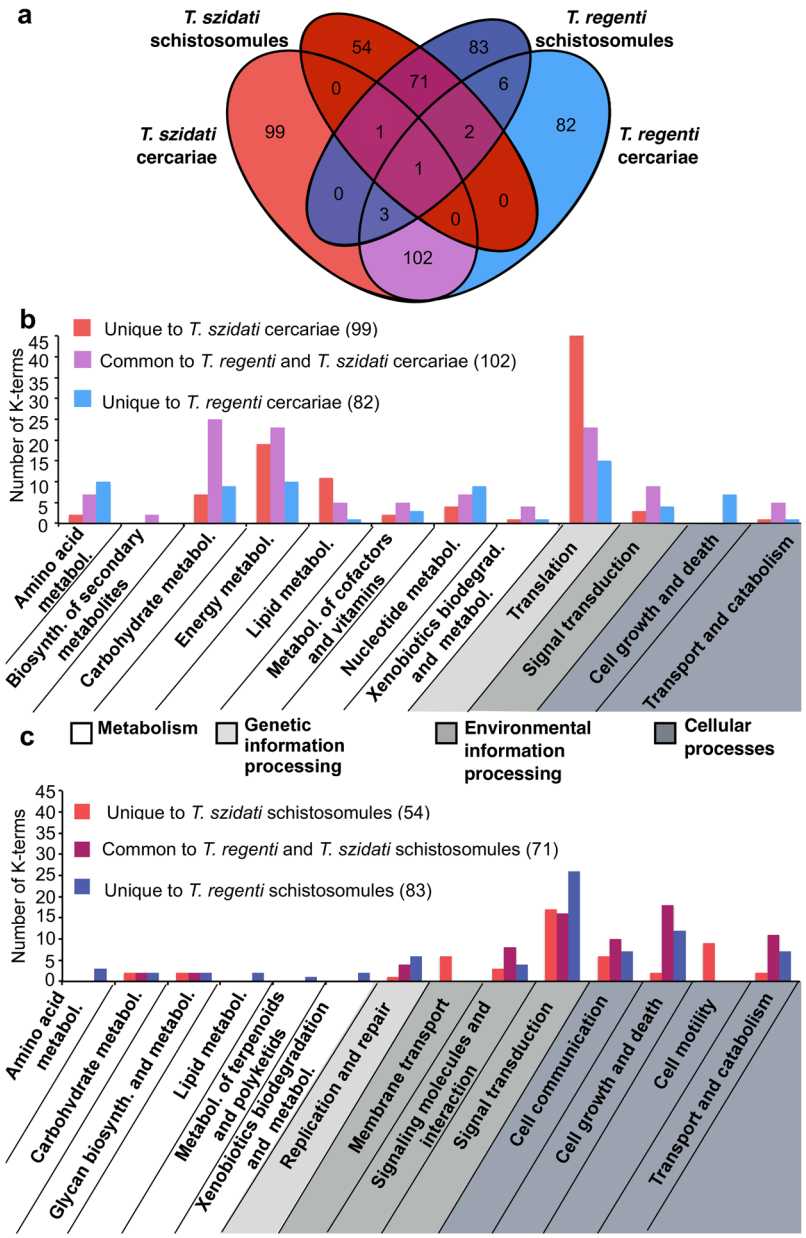

Figure 2. Comparison of enriched biological (KEGG) pathways common or specific to cercariae or schistosomules of *Trichobilharzia szidati* (red) and *T. regenti* (blue). Venn diagram-comparison of the numbers of K-terms that are common or specific to cercariae and/or schistosomules of *T. szidati* and/or *T. regenti* (**a**) Bar graph comparison of KEGG terms that are common or specific to either cercariae (**b**) or schistosomules (**c**).

as cathepsin B, which is known to degrade myelin[29] and is represented by the top transcribed peptidase gene (~3,300 CPM) in *T. regenti* (Fig. 5).

Conclusion

Using a bioinformatics workflow, we explored the transcriptomic landscape of the cercarial and schistosomule stages of *T. szidati* and *T. regenti*, which assume distinct migratory paths and predilection sites in the avian host. This study allowed us to explore the differences in biology between these flukes before they enter (cercariae) and as they migrate (schistosomules) through tissues in the host animal and aspects likely linked to fluke-host cross-talk. The exploration indicated that both *T. szidati* and *T. regenti* schistosomules have key molecular pathways and repertoires of enzymes critical for acquisition and degradation of macromolecules (including proteins, lipoproteins and lipids) from distinct tissue types during their distinct migratory paths. Such repertoires likely allow these migratory flukes to effectively use key components in blood (*T. szidati*), nerves (*T. regenti*) and/or other tissues (both species) to migrate, develop and live in quite different physiological conditions.

Following skin penetration of the avian host, cercariae lose their tails and metamorphose to schistosomules which then migrate through tissues in the definitive host. In the penetration phase, cercariae of both species are

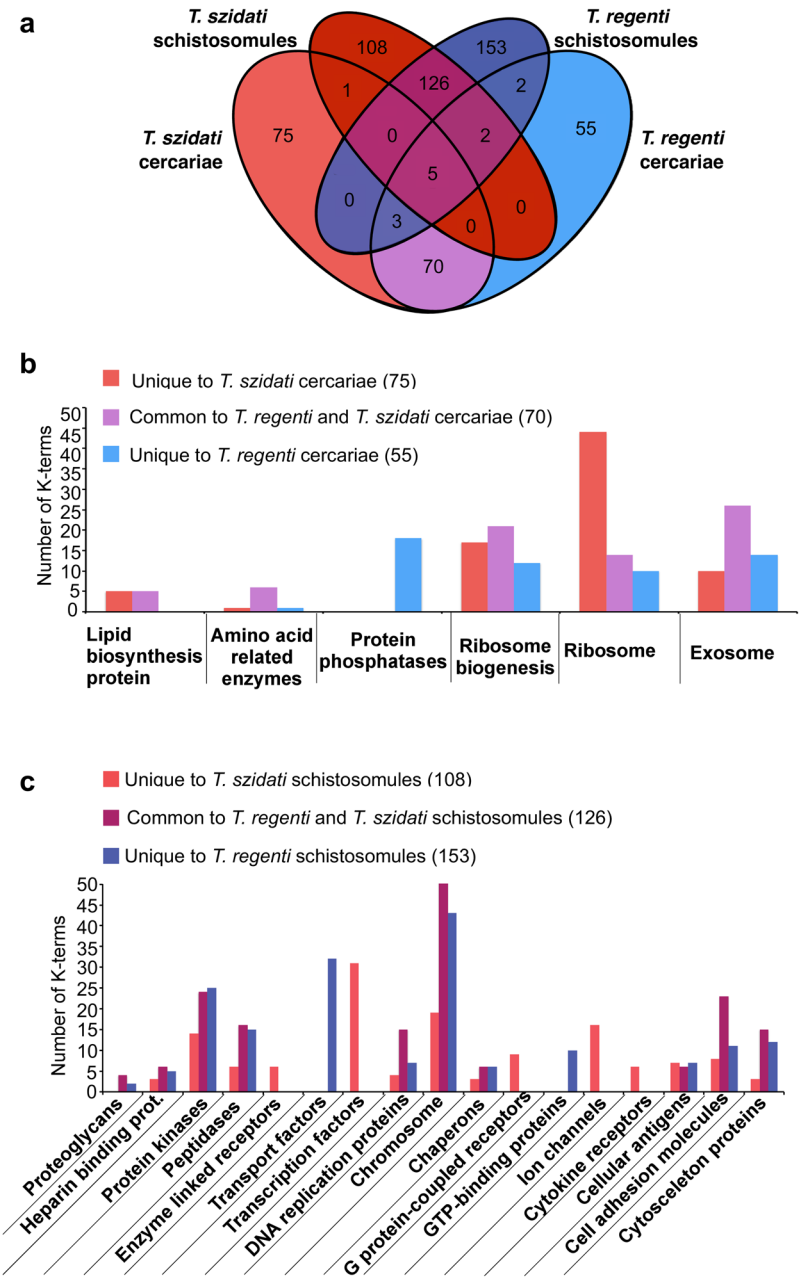

Figure 3. Comparison of enriched protein families (KEGG BRITE) common or specific to cercariae or schistsomules of *Trichobilharzia szidati* and *T. regenti*. Venn diagram-comparison of the numbers of KEGG terms that are common or specific to cercariae and/or schistosomules of *T. szidati* and/or *T. regenti* (**a**) Bar graph comparison of K-terms that are common or specific to either cercariae (**b**) or schistosomules (**c**).

enriched for pathways or molecules associated with carbohydrate metabolism (including glycolysis and Krebs cycle), oxidative phosphorylation and the translation of proteins including those required for ribosome biogenesis, exosome production and/or lipid biogenesis. In the migration phase, schistosomules of both species are enriched for pathways or molecules associated with processes including signal transduction, cell turnover and motility, DNA replication and repair, molecular transport and catabolism. The exclusive enrichment in signalling pathways linked to Ras, MAPK, Rap1, Wnt and ErbB in *T. regenti* appears to relate to processes including cell adhesion, proliferation, differentiation, migration, cell to cell communication and/or junction formation. These observations suggest the importance of cell adhesion molecules (CAMs) in this species[18]. CAMs are known to regulate host-parasite interactions, in addition to having crucial roles in the regulation of cellular integrity and interactions. The unique transcription profiles representing these molecules in schistosomules suggest an involvement in rapid growth and development of different organ structures in *T. regenti* within the avian host[30] and/or host-parasite interactions. The former aspect is supported by an abundant transcription linked to neuroligin, netrin receptor and semaphorin in *T. regenti*[18], which might relate to neural development in the schistosomule stage[31,32].

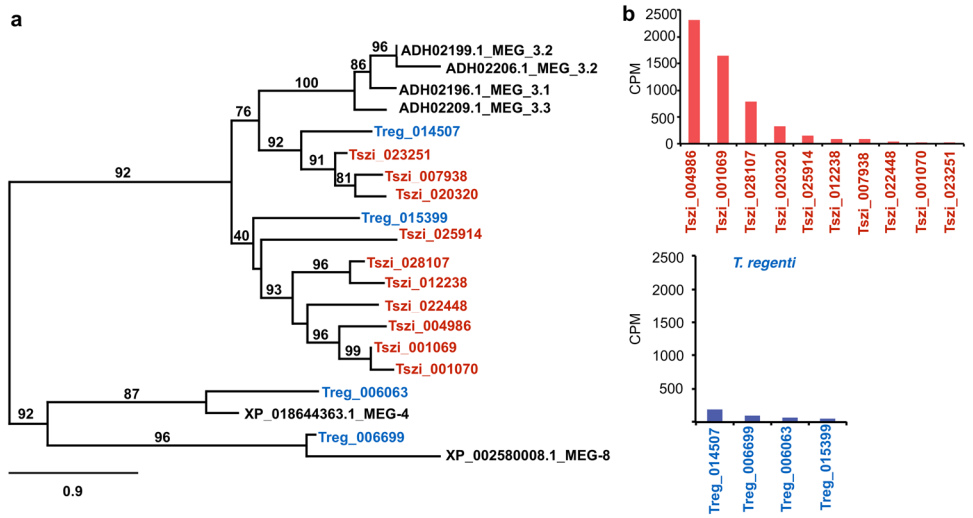

Figure 4. Phylogenetic relationships of protein sequences encoded by microexon genes (MEGs) among *Trichobilharzia szidati* (red), *T. regenti* (blue) and *Schistosoma mansoni* (reference taxon; black) constructed using a maximum likelihood (ML) tree building method, with nodal support values indicated (**a**), and a comparison of transcription levels (in counts per million reads, CPM) between *T. szidati* and *T. regenti* (**b**).

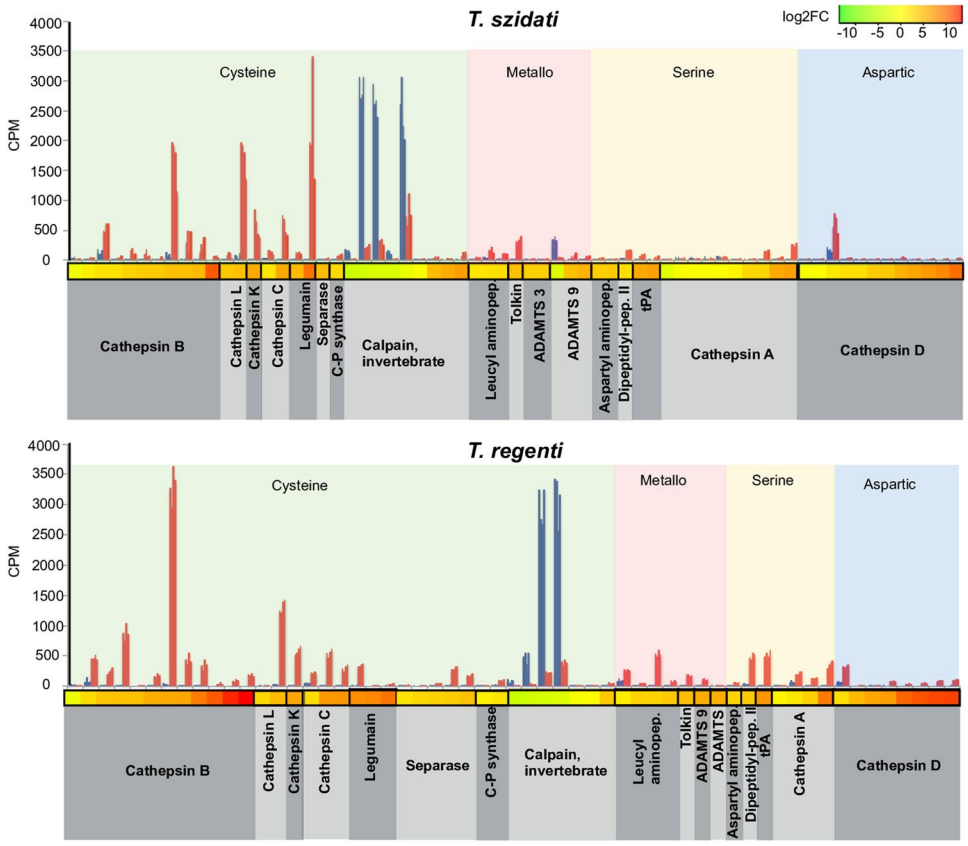

Figure 5. Transcription profiles of genes encoding peptidases (KEGG BRITE annotation) which were upregulated in cercariae (blue) or schistosomules (red) of *Trichobilharzia szidati* or *T. regenti*. Transcription levels, indicated by individual clusters of four biological replicates, are given in counts per million (CPM). Heat map colour indicates level of differential transcription between cercaria and schistosomule (log2-fold change).

In contrast to cercariae, which store and express proteolytic enzymes in their penetration glands for percutaneous invasion, schistosomules actively degrade various host tissue and cell types during migration[21] and inhibit or evade immune attack by the host[33]. The markedly higher levels of transcription of MEG genes in *T.*

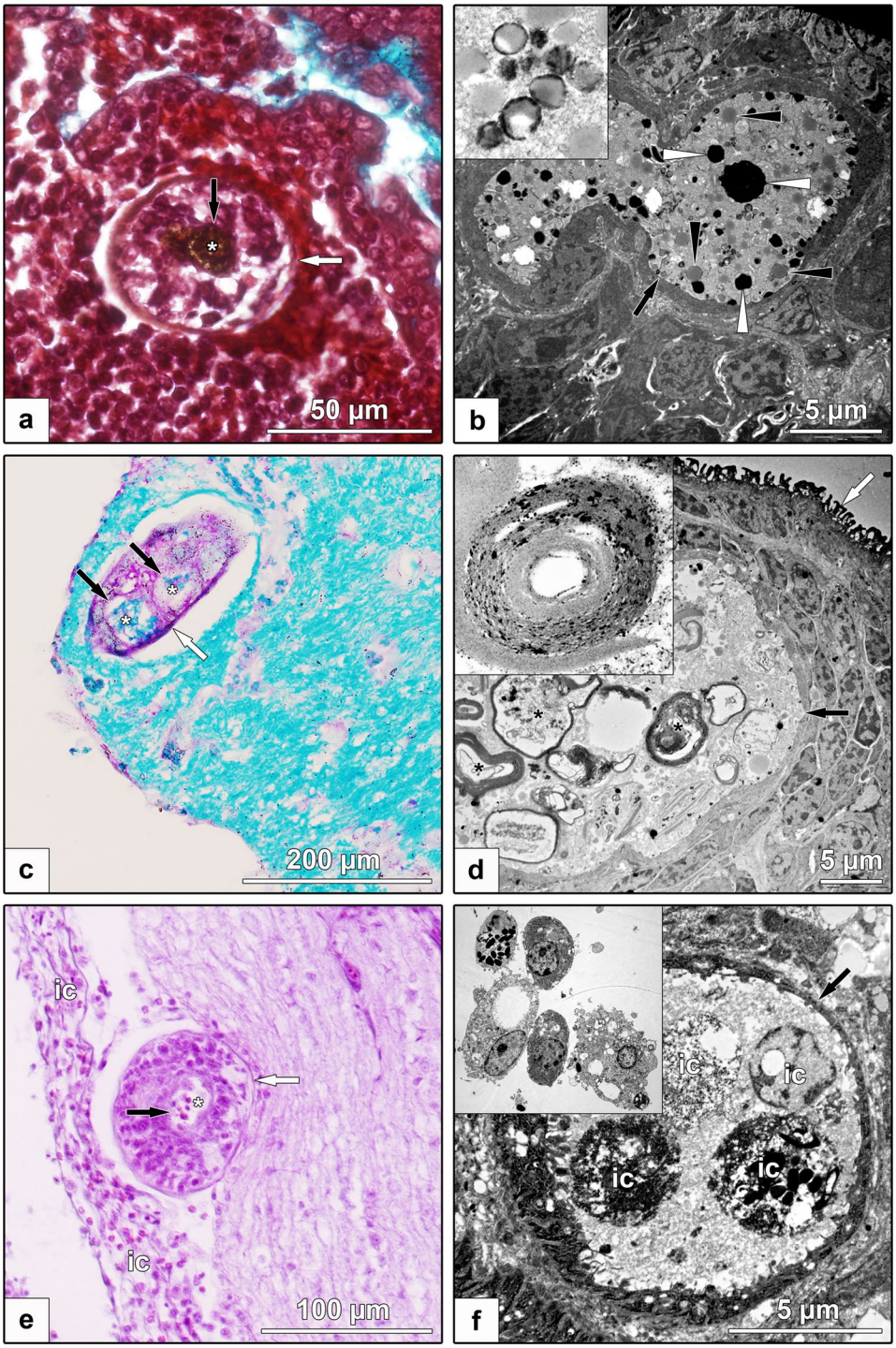

Figure 6. Light (left) and electron (right) microscopic examinations of schistosomules of *Trichobilharzia szidati* from the lungs and of *T. regenti* within the spinal cord from infected ducks. *T. szidati* schistosomules from/in lung (white arrow) with dark brown haem metabolites (white asterisk), crystals (white arrowheads) and/or lipid droplets (black arrowheads) in the gut lumen (black arrow). Detail of crystallization of hem metabolites is depicted in the box (**a**,**b**) *T. regenti* schistosomules (white arrow) migrating in the spinal cord with myelin (white asterisk) in the gut lumen (black arrows). Partially digested myelin sheath is depicted in detail in the box. (**c**,**d**) Immune cells surrounding (ic) or within the gut (black arrow, white asterisk) of *T. regenti* schistosomules (white arrow) in sub-meningeal space. Immune cells from sub-meningeal space are depicted in the box (**e** and **f**).

szidati compared with *T. regenti* suggests variation in nutrient acquisition or parasite-host cross-talk between the two species of fluke. Although many peptidases in both species would have broad substrate specificity to degrade many proteins and lipoprotein complexes, selected peptidases likely have relative specificity to degrade

blood (e.g., legumain) in *T. szidati* or myelin in nerves (e.g., cathepsin B) in *T. regenti*, reflecting an adaptation to distinct modes of nutrient acquisition between these visceral and neurotropic flukes of birds. The relatively high transcription of genes encoding a wide array of cathepsins and other proteolytic enzymes lends support for the involvement of this group of enzymes in numerous biological processes with key roles during the parasite's migration, feeding and survival in the definitive host.

Methods

Ethics and study approval. Animal experimentation was approved (ref. no. MSMT-31114/2013-9 and MSMT-33740/2017-2) by Charles University and the Czech Ministry of Education, Youth and Sports, and was conducted in accordance with the European Directive 2010/63/EU and Czech Law for Biomedical Research (246/1992 and 359/2012).

Parasite procurement. *Trichobilharzia szidati* and *T. regenti* were maintained in separate laboratories in snail intermediate (*Lymnaea stagnalis* or *Radix lagotis*) and definitive (*Anas platyrhynchos* f. *domestica*; breed - Cherry Valley strain) hosts using well-established protocols[8,18,34]. Life cycle stages of *T. szidati* were produced here, and those of *T. regenti* had been raised previously[18]. For each *Trichobilharzia* species, four independent biological replicates of cercariae were collected from distinct groups of infected snails ($n = 20$). For each replicate, 10,000 cercariae were washed extensively in H_20, pelleted by centrifugation at 2,500 *g* (4 °C), suspended in TRIzol (Thermo Fisher Scientific) and then frozen at −80 °C. For each *Trichobilharzia* species, four biological replicates of schistosomules were produced in (7-day old, helminth-free) ducks; each of four ducks was infected percutaneously with 2,500 cercariae[8,18,29]. Schistosomules were collected by microdissection from the lungs (*T. szidati*) and spinal cord (*T. regenti*) at 90 h and 162 h following infection of ducks, respectively; these time points were selected to ensure that the development of the schistosomules of the two species (collected from the distinct tissues) was the same and thus comparable[2,35,36]. For each replicate, 300 schistosomules were washed extensively in phosphate-buffered saline (PBS, pH 7.0), pelleted, suspended in TRIzol and frozen in the same manner as for cercariae.

RNA-sequencing. RNAs of *T. szidati* were sequenced here, and those of *T. regenti* had been done previously[18]. For both *T. szidati* and *T. regenti*, total RNA was isolated from each of the four replicates of each cercariae and schistosomules and *DN*ase I-treated[18]. RNA quality was assessed using a Bioanalyser 2100 (Agilent), and the quantity was estimated using a Qubit 4 fluorometer (Invitrogen). From each replicate representing each species and developmental stage of *Trichobilharzia*, mRNAs were isolated to construct short-insert (330 bp) cDNA libraries, barcoded according to the manufacturer's instructions (TruSeq RNASamplePreparation v.2, Illumina). For each species, all eight cDNA libraries were each pair-end sequenced (2 × 211 bp reads) using the HiSeq. 2500 platform (Illumina)[18].

Transcriptome assembly and protein prediction. An established bioinformatic workflow system[18] was used for assembly. In brief, adaptors and nucleotides with Phred quality scores of ≤20 were removed from raw RNA-seq reads from each library using Trimmomatic[37] v.0.3. Reads were error-corrected using BayesHammer in the software package SPAdes[38] v.3.1.0 and normalised using khmer[39] v.1.1. Read datasets were mapped specifically to the duck genome (BioProject no. PRJNA46621) using TopHat[40] v.2.1.1 and host sequences removed.

For each *Trichobilharzia* species, RNA-seq datasets representing the cercariae and schistosomules were pooled and employed to assemble a non-redundant transcriptome using Oases[41] v.0.2.8. In order to achieve an optimum transcriptome for each species and each replicate, various *k*-mer (21 to 49) and coverage (5 to 21) threshold values were assessed during the assembly process. Based on number of contigs, length of contigs, and transcript redundancy, the best assemblies were achieved using *k*-mers of 47 to 49 and coverages of 9 to 14. For each species, the eight transcriptomes representing the individual replicates of each of the larval stages (cercaria and schistosomule) were pooled and any redundancy removed using CD-HIT-EST[42] employing a nucleotide identity threshold of 85%. Protein-coding regions were predicted from individual non-redundant transcripts using Transdecoder[43] v.3.0.0 employing a minimum length of 30 amino acids (aa).

Transcriptome annotation and curation. An established pipeline[44] was used for annotation. Predicted protein sequences were annotated using their closest homologues (BLASTp; E-value cut-off: $\leq 10^{-5}$) employing the NCBI non-redundant protein, Kyoto Encyclopedia of Genes and Genomes (KEGG), excluding the "organismal system" and "human disease" categories[45], and UniProt[46]. The proteomes predicted separately from the larval transcriptomes of *T. szidati* and *T. regenti* were compared with that of *S. mansoni*[47]. For each *Trichobilharzia* species, orthologous and unique protein groups were identified using OrthoMCL[48] (E-value cut-off: $\leq 10^{-5}$; similarity cut-off: 0.5). The completeness of each transcriptome was assessed by *b*enchmarking *u*niversal *s*ingle-*c*opy *o*rthologs (BUSCO)[49]. Excretory/secretory proteins were predicted using the prokaryotic and eukaryotic classical analysis of secretomes (PECAS) pipeline employing default settings[50].

Subsequently, each transcriptome was curated. Transcripts with high nucleotide identity (BLASTn; E-value: $\leq 10^{-5}$) to avian, bacterial or viral sequences present in the NCBI non-redundant nucleotide database (NCBI) were removed. Any conceptually translated protein with a high amino acid sequence homology (BLASTp; E-value cut-off: $\leq 10^{-5}$) to transposable elements in RepBase database[51] were also eliminated, as were translated sequences of ≤50 amino acids with no homology (at the nucleotide and protein levels) to sequences in current public databases. Additionally, transcripts with read counts of <10 (using RSEM)[52] in all replicates of individual developmental stages were removed from each assembled transcriptome.

Microexon gene (MEG) transcripts were annotated by comparison (BLASTp; E-value: $\leq 10^{-5}$) with reference sequences in the GenBank database (accession nos. ADH02196.1, ADH02199.1, ADH02206.1, ADH02209.1,

XP_018644363.1, XP_002580008.1). Amino acid sequences inferred from these transcripts were aligned using the program MUSCLE[53] v 3.8.31 and maximum likelihood (ML) phylogenetic tree built employing PhyML[54] v 3.1 using a WAG model of amino acids substitution and a shape parameter of gamma distribution of 3.758; the resultant tree was displayed using TreeDyn[55].

Differential transcription and enrichment analyses. An established method[18] was used for this analysis. For each *Trichobilharzia* species and each replicate, the trimmed, corrected, paired reads were mapped to the respective non-redundant, merged transcriptome using RSEM[52]. Differential gene transcription between cercariae and schistosomules was calculated using edgeR[56] v.3.6.7 and R[57] 3.2.3. Read counts (in counts per million, CPM) were normalised for G+C bias[58] and for the trimmed mean of M-values (TMM)[59]. The edgeR exactTest function was used to assess differential transcription. The multiplicity correction of *P*-values was then performed by applying the Benjamini Hochberg method to establish the false discovery rate. Transcripts were considered differentially transcribed if the latter rate was $\leq$0.01. For differentially transcribed molecules, enriched metabolic pathways and protein families were identified using the KEGG database[40]. Upregulated transcripts assigned to KEGG orthology (KO) terms were mapped to KEGG pathways and the KEGG BRITE database. A Fisher's exact test was used to calculate levels of statistical significance in differences between numbers of KO terms assigned to particular pathway/enzyme classes; a *P*-value of $\leq$0.05 was set as the threshold for significant enrichment in a particular developmental stage of each *Trichobilharzia* species studied.

Microscopic examinations. In the light microscopic examination, tissues containing schistosomules were collected from ducks infected for 4 days with *T. szidati* (lungs) or for 7 days with *T. regenti* (spinal cord) and immediately fixed in Bouin's solution (cat. no. HT10132, Sigma-Aldrich). These tissues were processed and stained with Masson's trichrome (lungs), haematoxylin-eosin or with Luxol fast blue (spinal cord) for histological examination using standard procedures. Sections were examined using a BX51 microscope (Olympus), and photographically documented using a DP72 camera (Olympus) and software Quick Photo Micro 3.1. (Promicra).

In the transmission electron microscopic examination, schistosomules of *T. szidati* (4 days) dissected from lung tissues and those of *T. regenti* (10 days) within spinal cord tissue blocks (4–6 mm^3) were immediately fixed in modified Karnovsky solution[60], fixed in OsO_4 and then embedded in Spurr resin (SPI-supplies) using established protocols[61]. Schistosomules and/or tissues were subjected to transmission electron microscopy (JEOL JEM-1011) and photographed using a CCD camera (Veleta) employing acquisition software (Olympus Soft Imaging Solution GmbH).

Data Availability

Short sequence reads and transcriptome analysed are available via Bioproject Submission SUB4374991.

References

1. Warren, K. S. Schistosomiasis: host-pathogen biology. *Rev. Infect. Dis.* **4**, 771–775 (1982).
2. Horák, P., Kolářová, L. & Adema, C. M. Biology of the schistosome genus *Trichobilharzia*. *Adv. Parasitol.* **52**, 155–233 (2002).
3. Colley, D. G., Bustinduy, A. L., Secor, W. E. & King, C. H. Human schistosomiasis. *Lancet* **383**, 2253–2264 (2014).
4. Rollinson, D. *et al*. Genetic diversity of schistosomes and snails: implications for control. *Parasitology* **136**, 1801–1811 (2009).
5. Neuhaus, W. Biology and development of *Trichobilharzia szidati* n. sp. (Trematoda, Schistosmatidae), a parasite causing dermatitis in man. *Z. Parasitenkd.* **15**, 203–266 (1952).
6. Horemans, A. M., Tielens, A. G. & van den Bergh, S. G. The reversible effect of glucose on the energy metabolism of S*chistosoma mansoni* cercariae and schistosomula. *Mol. Biochem. Parasitol.* **51**, 73–79 (1992).
7. Skelly, P. J., Stein, L. D. & Shoemaker, C. B. Expression of S*chistosoma mansoni* genes involved in anaerobic and oxidative glucose metabolism during the cercaria to adult transformation. *Mol. Biochem. Parasitol.* **60**, 93–104 (1993).
8. Horák, P., Kovář, L., Kolářová, L. & Nebesářová, J. Cercaria-schistosomulum surface transformation of *Trichobilharzia szidati* and its putative immunological impact. *Parasitology* **116**, 139–147 (1998).
9. Chanová, M., Vuong, S. & Horák, P. *Trichobilharzia szidati:* the lung phase of migration within avian and mammalian hosts. *Parasitol. Res.* **100**, 1243–1247 (2007).
10. Kašný, M. *et al*. Cathepsins B1 and B2 of *Trichobilharzia* spp., bird schistosomes causing cercarial dermatitis. *Adv. Exp. Med. Biol.* **712**, 136–154 (2011).
11. Dong, Y. & Benveniste, E. N. Immune function of astrocytes. *Glia* **36**, 180–190 (2001).
12. Kolářová, L., Horák, P. & Čada, F. Histopathology of CNS and nasal infections caused by *Trichobilharzia regenti* in vertebrates. *Parasitol. Res.* **87**, 644–650 (2001).
13. Rock, R. B. *et al*. Role of microglia in central nervous system infections. *Clin. Microbiol. Rev.* **17**, 942–964 (2004).
14. Lichtenbergová, L., Lassmann, H., Jones, M. K., Kolářová, L. & Horák, P. *Trichobilharzia regenti:* host immune response in the pathogenesis of neuroinfection in mice. *Exp. Parasitol.* **128**, 328–335 (2011).
15. Gobert, G. N., You, H. & McManus, D. P. Gaining biological perspectives from schistosome genomes. *Mol. Biochem. Parasitol.* **196**, 21–28 (2014).
16. Rinaldi, G. *et al*. New research tools for urogenital schistosomiasis. *J. Infect. Dis.* **211**, 861–869 (2015).
17. Hahnel, S. *et al*. Tissue-specific transcriptome analyses provide new insights into GPCR signalling in adult S*chistosoma mansoni*. *PLoS Pathog.* **14**, e1006718 (2018).
18. Leontovyč, R. *et al*. Comparative transcriptomic exploration reveals unique molecular adaptations of neuropathogenic *Trichobilharzia* to invade and parasitize its avian definitive host. *PLoS Negl. Trop. Dis.* **10**, e0004406 (2016).
19. Wilson, R. A. *et al*. The schistosome esophagus is a 'hotspot' for microexon and lysosomal hydrolase gene expression: implications for blood processing. *PLoS Negl. Trop. Dis.* **9**, e0004272 (2015).
20. Li, X.-H. *et al*. Microexon gene transcriptional profiles and evolution provide insights into blood processing by the S*chistosoma japonicum* esophagus. *PLoS Negl. Trop. Dis.* **12**, e0006235 (2018).
21. Vaccaro, A. M., Salvioli, R., Tatti, M. & Ciaffoni, F. Saposins and their interaction with lipids. *Neurochem. Res.* **24**, 307–314 (1999).
22. Jeon, S.-B., Yoon, H. J., Park, S.-H., Kim, I.-H. & Park, E. J. Sulfatide, a major lipid component of myelin sheath, activates inflammatory responses as an endogenous stimulator in brain-resident immune cells. *J. Immunol.* **181**, 8077–8087 (2008).
23. Lee, J.-Y. *et al*. Hemolytic activity and developmental expression of pore-forming peptide, clonorin. *Biochem. Biophys. Res. Commun.* **296**, 1238–1244 (2002).

24. Espino, A. M. & Hillyer, G. V. Molecular cloning of a member of the *Fasciola hepatica* saposin-like protein family. *J. Parasitol.* **89**, 545–552 (2003).
25. Don, T. A., Bethony, J. M. & Loukas, A. Saposin-like proteins are expressed in the gastrodermis of *Schistosoma mansoni* and are immunogenic in natural infections. *Int. J. Infect. Dis.* **12**, e39–47 (2008).
26. Caffrey, C. R., McKerrow, J. H., Salter, J. P. & Sajid, M. Blood 'n' guts: an update on schistosome digestive peptidases. *Trends Parasitol.* **20**, 241–248 (2004).
27. Delcroix, M. *et al*. A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. *J. Biol. Chem.* **281**, 39316–39329 (2006).
28. Kašný, M. *et al*. Chapter 4. Peptidases of Trematodes. *Adv. Parasitol.* **69**, 205–297 (2009).
29. Dolečková, K. *et al*. The functional expression and characterisation of a cysteine peptidase from the invasive stage of the neuropathogenic schistosome *Trichobilharzia regenti*. *Int. J. Parasitol.* **39**, 201–11 (2009).
30. Gumbiner, B. M. Cell adhesion: the molecular basis of tissue architecture and morphogenesis. *Cell* **84**, 345–357 (1996).
31. Larrivée, B., Freitas, C., Suchting, S., Brunet, I. & Eichmann, A. Guidance of vascular development: lessons from the nervous system. *Circ. Res.* **104**, 428–441 (2009).
32. Ou, G., Stuurman, N., D'Ambrosio, M. & Vale, R. D. Polarized myosin produces unequal-size daughters during asymmetric cell division. *Science* **330**, 677–680 (2010).
33. Modha, J., Redman, C. A., Thornhill, J. A. & Kusel, J. R. Schistosomes: unanswered questions on the basic biology of the host-parasite relationship. *Parasitol. Today* **14**, 396–401 (1998).
34. Horák, P., Dvořák, J., Kolářová, L. & Trefil, L. *Trichobilharzia regenti*, a pathogen of the avian and mammalian central nervous systems. *Parasitology* **119**, 577–581 (1999).
35. Horák, P. & Kolářová, L. Survival of bird schistosomes in mammalian lungs. *Int. J. Parasitol.* **30**, 65–68 (2000).
36. Blažová, K. & Horák, P. *Trichobilharzia regenti*: the developmental differences in natural and abnormal hosts. *Parasitol. Int.* **54**, 167–172 (2005).
37. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
38. Nurk, S. *et al*. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* **20**, 714–737 (2013).
39. Brown, C. T. *et al*. A Reference-Free Algorithm for Computational Normalization arXiv:1203.4802v2 [q-bio.GN], 1–18 (2012).
40. Trapnell, C. *et al*. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc.* **7**, 562–78 (2012).
41. Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086–1092 (2012).
42. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
43. Haas, B. J. *et al*. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–512 (2013).
44. Schwarz, E. M. *et al*. The genome and developmental transcriptome of the strongylid nematode *Haemonchus contortus*. *Genome Biol.* **14**, R89 (2013).
45. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–361 (2017).
46. Magrane, M. UniProt Consortium. UniProt knowledgebase: a hub of integrated protein data. *Database* **2011**, bar009 (2011).
47. Berriman, M. *et al*. The genome of the blood fluke *Schistosoma mansoni*. *Nature* **460**, 352–358 (2009).
48. Li, L., Stoeckert, C. J. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
49. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
50. Cortazar, A. R., Oguiza, J. A., Aransay, A. M. & Lavin, J. L. PECAS: prokaryotic and eukaryotic classical analysis of secretome. *Amino Acids* **47**, 2659–2663 (2015).
51. Bao, W., Kojima, K. K. & Kohany, O. RepbaseUpdate, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
52. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
53. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
54. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
55. Chevenet, F., Brun, C., Banuls, A.-L., Jacq, B. & Christen, R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7**, 439 (2006).
56. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
57. R Development Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria (2008).
58. Risso, D., Schwartz, K., Sherlock, G. & Dudoit, S. GC-content normalization for RNA-seq data. *BMC Bioinformatics* **12**, 480 (2011).
59. Dillies, M.-A. *et al*. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief. Bioinform.* **14**, 671–683 (2013).
60. Karnovsky, M. J. A formaldehyde-glutaraldehyde fixative of high osmolality for use in electron microscopy. *J. Cell Biol.* **27**, 137–8A (1965).
61. Ligasová, A. *et al*. Secretory glands in cercaria of the neuropathogenic schistosome *Trichobilharzia regenti* - ultrastructural characterization, 3-D modelling, volume and pH estimations. *Parasit. Vectors* **4**, 162 (2011).

Acknowledgements

In the Czech Republic, this study was supported through grants from the Czech Science Foundation (18–11140S), the European Regional Development Fund and Ministry of Education, Youth and Sports of the Czech Republic (CZ.02.1.01/0.0/0.0/16_019/0000759) and Charles University in Prague (PROGRES_Q43, UNCE/SCI/012-204072/2018 and SVV260432/2018) (P.H *et al*.), and computational resources were provided via CESNET LM2015042 and the CERIT Scientific Cloud LM2015085 (Large Research, Development and Innovation Infrastructure). In Australia, funding was from the Australian Research Council and the National Health and Medical Research Council (NHMRC) of Australia (R.B.G., N.D.Y. *et al*.), and computational resources were provided via the Melbourne Bioinformatics Platform. P.K.K. holds an Early Career Research Fellowship (ECRF) and N.D.Y. a Career Development Research Fellowship (CDF1) from NHMRC. The funders had no role in study design, data collection or analyses, decision to publish or preparation of the manuscript.

Author Contributions

Conceived, designed and/or supervised the study: P.H., R.B.G., N.D.Y. and M.K. Conducted the animal experiments and microscopy: R.L., J.B. and V.J. Carried out transcriptomic and bioinformatic analyses: N.D.Y., R.L., P.K.K., R.S.H. and R.B.G. Interpreted the findings: R.L., R.B.G., P.H. and N.D.Y. Planned and wrote the manuscript: R.L., R.B.G. and N.D.Y. All authors read and approved the final version of the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-37669-2.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 26 February 2020
doi: 10.3389/fcimb.2020.00066

Isoforms of Cathepsin B1 in Neurotropic Schistosomula of *Trichobilharzia regenti* Differ in Substrate Preferences and a Highly Expressed Catalytically Inactive Paralog Binds Cystatin

*Hana Dvořáková[1], Roman Leontovyč[1], Tomáš Macháček[1], Anthony J. O'Donoghue[2†], Ondřej Šedo[3], Zbyněk Zdráhal[3], Charles S. Craik[2], Conor R. Caffrey[4†], Petr Horák[1] and Libor Mikeš[1]**

[1] Department of Parasitology, Faculty of Science, Charles University, Prague, Czechia, [2] Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, San Francisco, CA, United States, [3] Central European Institute of Technology, Masaryk University, Brno, Czechia, [4] Center for Discovery and Innovation in Parasitic Diseases, Department of Pathology, University of California, San Francisco, San Francisco, CA, United States

Schistosomula (the post-infective stages) of the neurotropic schistosome *Trichobilharzia regenti* possess multiple isoforms of cathepsin B1 peptidase (TrCB1.1-TrCB1.6) with involvement in nutrient digestion. The comparison of substrate preferences of TrCB1.1 and TrCB1.4 showed that TrCB1.4 had a very narrow substrate specificity and after processing it was less effective toward protein substrates when compared to TrCB1.1. Self-processing of both isoforms could be facilitated by sulfated polysaccharides due to a specific binding motif in the pro-sequence. Trans-activation by heterologous enzymes was also successfully employed. Expression profiling revealed a high level of transcription of genes encoding the enzymatically inactive paralogs TrCB1.5 and TrCB1.6. The transcription level of TrCB1.6 was comparable with that of TrCB1.1 and TrCB1.2, the most abundant active isoforms. Recombinant TrCB1.6wt, a wild type paralog with a Cys^{29}-to-Gly substitution in the active site that renders the enzyme inactive, was processed by the active TrCB1 forms and by an asparaginyl endopeptidase. Although TrCB1.6wt lacked hydrolytic activity, endopeptidase, but not dipeptidase, activity could be restored by mutating Gly^{29} to Cys^{29}. The lack of exopeptidase activity may be due to other mutations, such as His^{110}-to-Asn in the occluding loop and Asp^{224}-to-Gly in the main body of the mature TrCB1.6, which do not occur in the active isoforms TrCB1.1 and TrCB1.4 with exopeptidase activity. The catalytically active enzymes and the inactive TrCB1.6 paralog formed complexes with chicken cystatin, thus supporting experimentally the hypothesis that inactive paralogs could potentially regulate the activity of the active forms or protect them from being inhibited by host inhibitors. The effect on cell viability and nitric oxide production by selected immune cells observed for TrCB1.1 was not confirmed for TrCB1.6. We show here that the active isoforms of TrCB1 have different affinities for peptide substrates thereby facilitating diversity in protein-derived nutrition for

OPEN ACCESS

Edited by:
Mario Alberto Rodriguez, National Polytechnic Institute, Mexico

Reviewed by:
César López-Camarillo, Universidad Autónoma de la Ciudad de México, Mexico
Anahid Jewett, University of California, Los Angeles, United States

***Correspondence:**
Libor Mikeš
mikes@natur.cuni.cz

†Present address:
Anthony J. O'Donoghue and Conor R. Caffrey, Center for Discovery and Innovation in Parasitic Diseases, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA, United States

Specialty section:
This article was submitted to Parasite and Host, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *30 September 2019*
Accepted: *07 February 2020*
Published: *26 February 2020*

Citation:
Dvořáková H, Leontovyč R, Macháček T, O'Donoghue AJ, Šedo O, Zdráhal Z, Craik CS, Caffrey CR, Horák P and Mikeš L (2020) Isoforms of Cathepsin B1 in Neurotropic Schistosomula of Trichobilharzia regenti Differ in Substrate Preferences and a Highly Expressed Catalytically Inactive Paralog Binds Cystatin. Front. Cell. Infect. Microbiol. 10:66. doi: 10.3389/fcimb.2020.00066

the parasite. The inactive paralogs are unexpectedly highly expressed and one of them retains the ability to bind cystatins, likely due to specific mutations in the occluding loop and the enzyme body. This suggests a role in sequestration of inhibitors and protection of active cysteine peptidases.

Keywords: peptidase, cathepsin B, processing, substrate specificity, occluding loop, cystatin, helminth, schistosome

INTRODUCTION

Flukes (Trematoda: Digenea) of the family Schistosomatidae include blood-dwelling flatworms that are pathogenic in birds and mammals, including man. Their infectious juveniles (cercariae) developing in intermediate snail hosts penetrate the skin of the definitive host in the water and enter the blood circulation while transforming to post-infective stages—the schistosomula (Kašný et al., 2016; Řimnáčová et al., 2017). In humans, the invasion of the skin can be accompanied by an allergic reaction called cercarial dermatitis (Horák et al., 2015; Macháček et al., 2018). However, schistosomula of the avian schistosome *Trichobilharzia regenti* use peripheral nerves and the central nervous system (CNS) as a migratory route to the final localization in the nasal mucosa of the duck host (Horák et al., 1998, 1999). The migration through the CNS can be manifested by neuromotor symptoms (leg paralysis and balance disorders) in birds and experimentally infected mice, occasionally leading to death of the host (Horák et al., 1999, 2008).

The migrating schistosomula actively feed on nervous tissue (Lichtenbergová et al., 2011; Leontovyč et al., 2019) and the adults in the nasal cavity ingest blood (Chanová and Horák, 2007). Indeed, both stages require gut-associated peptidases to digest host proteins. In *T. regenti*, isoforms of the cysteine peptidase cathepsin B1 (TrCB1) are expressed in the gut and represent a subset of a larger group of peptidases that are responsible for digestion (Dvořák et al., 2005; Dolečková et al., 2010; Leontovyč et al., 2016, 2019), much like in human *Schistosoma* species (Caffrey et al., 2004) and other helminths (Williamson et al., 2003, 2004; Delcroix et al., 2006; Caffrey et al., 2018). Cathepsin B (IUPAC: EC 3.4.22.1; MEROPS: Clan CA, Family C1) is unique among other papain-like cysteine peptidases due to its ability to act as both an endopeptidase and carboxy-exopeptidase (peptidyl dipeptidase) (Barrett and Kirschke, 1981). The latter activity is enabled due to the presence of an extra structural element termed "occluding loop," which occupies S′ subsites of the enzyme (Musil et al., 1991; Illy et al., 1997). Cathepsin B is synthesized as an inative precursor and a mature enzyme arises as a consequence of proteolytic removal of the N-terminal pro-peptide that sterically blocks the active site (Mort and Buttle, 1997). Six forms of TrCB1 (TrCB1.1 – TrCB1.6, GenBank: AY648119 – 24) were previously identified (Dvořák et al., 2005). Two of these, namely TrCB1.5 and TrCB1.6, are paralogs possessing a substitution of the catalytic cysteine Cys^{29} for a glycine Gly^{29} (mature TrCB1 numbering). Paralogs with similar single-amino acid mutations in the catalytic site are known for many parasite peptidases (Merckelbach et al., 1994; Caffrey et al., 2000; Holt et al., 2004; He et al., 2005; McCoubrie et al., 2007; Mendoza-Palomares et al., 2008; Pillay et al., 2010; Jedličková et al., 2018). Substitution of the active site cysteine need not always necessarily abolish the enzyme's activity, as shown in cases of Cys-to-Ser nucleophile substitutions (McCoubrie et al., 2007; Pillay et al., 2010), whereas other substitutions are expected to produce loss of peptidase activity, which is the case of TrCB1.5 and 1.6. It has been proposed that inactive paralogs of peptidases might regulate the activity of active forms by competing for substrates or inhibitors (Merckelbach et al., 1994; Holt et al., 2003, 2004; Dvořák et al., 2005). Moreover, they may alter host immune response through the inhibition of host complement pathways by binding the complement components C1q, mannose-binding lectin or properdin (Holt et al., 2004; Bergström et al., 2009; Reynolds et al., 2014). Active cathepsin B1 of *Schistosoma mansoni* (SmCB1) diminished the proinflamatory reaction of activated macrophages (Donnelly et al., 2009), provoked mixed Th1/Th2/Th17 immune response in mice and induced a transient T-helper 17 response in acute schistosome infection (Soloviova et al., 2019), suggesting it has a considerable immunogenic capacity. As for *T. regenti*, an inflammation-promoting activity was noticed in case of active TrCB1.1 which triggered the production of nitric oxide and proinflammatory cytokines interleukin-6 and tumor necrosis factor α by astrocytes and/or microglia (Macháček et al., 2016), i.e., the cells activated in mice infected experimentally by *T. regenti* (Lichtenbergová et al., 2011). However, no specific immunomodulatory mechanisms were tested so far with inactive peptidase paralogs of helminths.

Although the amino acid sequence identity among the six forms of TrCB1 is high (98 – 99% between the active forms TrCB1.1–TrCB1.4) (Dvořák et al., 2005), it seems that even a small change (a single amino acid substitution) in the sequences, especially in the S2 subsite, is likely to alter substrate specificity. Thus, the presence of multiple forms of a particular enzyme may indicate an adaptation of the parasite to physiological changes in the course of migration through the nervous tissue and final settlement in the nasal mucosa, enabling digestion of various host components.

In this study, we performed an in-depth comparison of substrate preferences of the two most divergent active homologs, TrCB1.1 and TrCB1.4, using positional scanning synthetic combinatorial peptide library and mass spectrometry. We also evaluated the ratios of relative expression levels of all six TrCB1 homologs using transcript-specific Illumina reads and stage-specific transcriptomes of *T. regenti* cercariae and schistosomula.We attempted to identify the possible biological roles of the cathepsin B paralogs employing heterologously

expressed wild type TrCB1.6wt and its site-directed mutant TrCB1.6G/C (retrograde Gly^{29}-to-Cys^{29} mutation). Various methods of enzymatic processing (pro-peptide sequence removal) were applied to all four zymogens included in this study. Finally, we assessed the effects of TrCB1.6wt paralog on cell viability and nitric oxide production using murine astrocytes, microglia and macrophages.

RESULTS

Expression of TrCB1 Forms

All recombinant pro-TrCB1 forms produced in *Pichia pastoris* X-33 were hyperglycosylated by the yeast. Each pro-enzyme migrated as a fuzzy band around 50 kDa on SDS-PAGE. After enzymatic deglycosylation and subsequent purification, they occurred as prominent ≈36 kDa bands (**Supplementary Figure 1**), which corresponds with their theoretical molecular weight (MW) (**Table 1**). N-terminal amino acid sequences NEMQF of pro-TrCB1.6wt and pro-TrCB1.6G/C, and ENEIQ of pro-TrCB1.1 and pro-TrCB1.4 corresponded to the beginning of the enzymes' predicted pro-sequences. In addition, a few minor bands of lower MW appeared in fractions of deglycosylated purified pro-TrCB1.6G/C, pro-TrCB1.1, and pro-TrCB1.4 (**Supplementary Figure 1**).

Processing of Recombinant *Pichia*-Derived Pro-enzymes

(A) Incubation of pro-TrCB1.1 and pro-TrCB1.4 at pH 4 or pH 5.5 for up to 16 h resulted in a decrease in abundance of the pro-protein but no detectable appearance of the lower MW mature enzymes (an example is presented in **Supplementary Figure 2**). However, addition of 20 μg/ml sulfated polysaccharides (SP)—dextran sulfate (DS) in the case of TrCB1.1 and TrCB1.4 or heparin sodium salt (HSS) with TrCB1.4—resulted in reliable production of ≈30 kDa processed enzymes at pH 4.5 (**Figures 1C,D**). N-terminal sequencing revealed a L^{88}EIPS sequence (pro-TrCB1 numbering) in TrCB1.4 that was processed in the presence of HSS (**Figure 2A**). The detected cleavage site was two amino acid residues upstream from the predicted beginning of the mature enzyme (Dvořák et al., 2005). Further efforts to determine N-terminal residues in ≈30 kDa SP-treated TrCB1.1 and DS-treated dsTrCB1.4 failed. Incubation of pro-enzymes with both SPs resulted in higher activity against the fluorogenic substrate Z-Phe-Arg-AMC, especially at pH 4.5 (**Figure 1A**). However, a rapid decline of the activity due to enzyme autodegradation could be observed between ∼1 and 4 h (**Figure 1B**), even when stored on ice.

(B) SDS-PAGE demonstrated that agarose-bound pepsin succesfully processed all pro-enzymes, since new single/double bands of ≈30–32 kDa appeared in the gel (**Figure 3A**). N-terminal sequencing of pepsin-treated pepTrCB1.1 and pepTrCB1.4 after 1-h incubation revealed that at least 3 products arose due to different cleavage sites within the pro-sequences (**Figure 2A**), but none of the products started with predicted N-terminal residues of mature enzymes. The pepTrCB1.1, pepTrCB1.4, and pepTrCB1.6G/C had increased activity toward Z-Phe-Arg-AMC compared to corresponding pro-enzymes. Maximum activities were observed within 30 min of incubation in samples of TrCB1.1 and 1.4 and within 5 min in the case of TrCB1.6G/G (**Figure 3B**). The pepTrCB1.1 and pepTrCB1.4 were stable even after longer (weeks) storage at −25°C. On the other hand, the pepTrCB1.6G/C had very limited stability. The fluorescent active-site probe DCG-04 bound to pepTrCB1.1 and pepTrCB1.4 (**Figure 4**). This reaction was blocked by the irreversible inhibitor of papain-like cysteine peptidases E-64. No binding of the probe to pepTrCB1.6wt and pepTrCB1.6G/C was observed (**Figure 4**). Unless otherwise noted, the pepTrCB1 forms (after 30 min incubation with pepsin) were used in subsequent experiments.

(C) Trans-processing of pro-TrCB.6wt by recombinant asparaginyl endopeptidase from the tick *Ixodes ricinus* (IrAE) led to a production of variously processed forms (**Figure 5A**). All tested pH values gave the same results (**Supplementary Figure 3**). Three cleavage sites identified by N-terminal sequencing were N^{23}/E^{24}, D^{59}/A^{60}, N^{85}/V^{86} (pro-TrCB1.6 numbering) (**Figure 2B**). Prolonged incubation at pH 5.5 (up to 36 h) resulted in the formation of a single ≈30 kDa band, but N-terminal sequencing failed in this case. The intensity of the upper band (TrCB1.6 pro-enzyme) present in the lane at 16 h decreased to undetectable levels compared to time 0 (**Figure 5A**).

(D) Overnight incubation of pro-TrCB1.6wt with pepTrCB1.1 or pepTrCB1.4 in an activation buffer of pH 4.5 supplemented with DS resulted in the conversion of the zymogens to a major band of ≈32 kDa, and a minor band of ca 30 kDa (**Figure 5B**). Processing was not observed when E-64 (10 μM) was added to the reactions (**Figure 5B**). The N-terminal sequence G^{65}VMRE (pro-TrCB1.6 numbering; **Figure 2B**) was detected in the more intense ≈32 kDa band after overnight incubation of pro-TrCB1.6wt with pepTrCB1.4, while N-terminal sequencing of the other products was unsuccessful.

TABLE 1 | Theoretical molecular weight and number of potential N-glycosylation sites of pro-enzymes/mature TrCB1 forms.

	Pro-enzyme (kDa)	Mature enzyme (kDa)	N-gly sites (pro-sequence/mature enzyme)
TrCB1.1	36.4	28.5	1/1
TrCB1.4	36.5	28.6	1/1
TrCB1.6wt	36.0	28.2	1/2
TrCB1.6G/C	36.1	28.3	1/2

Peptidase Activity and Inhibition Assays of TrCB1.6, TrCB1.1, and TrCB1.4

Assays were performed with two synthetic fluorogenic substrates: Z-Phe-Arg-AMC and Z-Arg-Arg-AMC. The pepTrCB1.6wt had no activity with both substrates (not shown). The activity optimum of pepTrCB1.6G/C with Z-Phe-Arg-AMC occurred at pH 6, i.e., it was shifted more to the neutral value than that of pepTrCB1.1 (pH 5.5) and pepTrCB1.4 (pH 5) (**Figure 6A**).

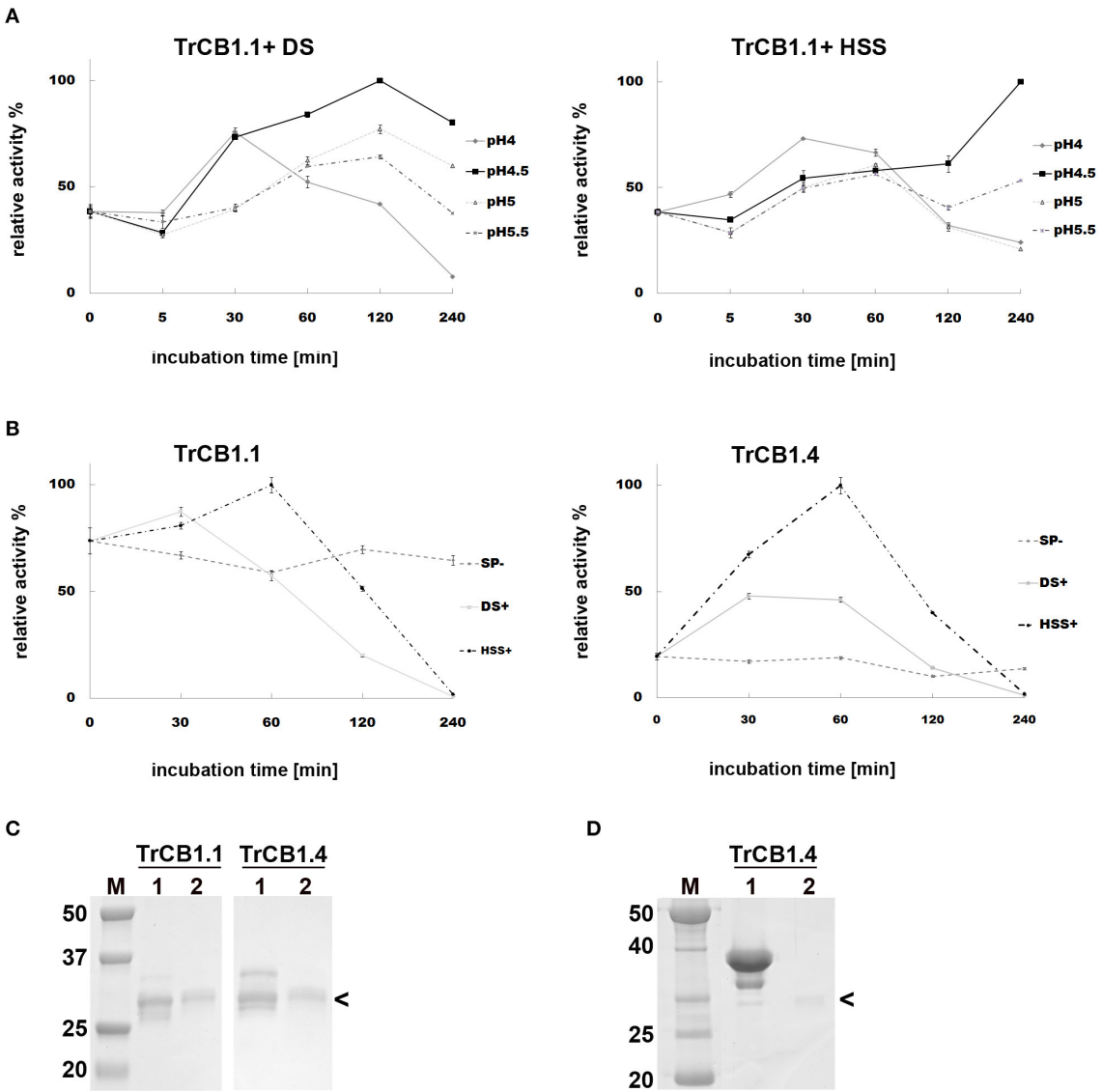

FIGURE 1 | Effect of sulfated polysaccharides on the autocatalytic activation of pro-TrCB1.1 and pro-TrCB1.4. Peptidolytic activities of TrCB1 forms generated after activation were determined in a kinetic assay with Z-Phe-Arg-AMC. The mean values ± standard deviation (SD) of two triplicate assays are expressed as percentage of relative activity in the sample at specified time points. **(A)** pro-TrCB1.1 was incubated in the presence of 10 μg/ml dextran sulfate (DS) or heparin sodium salt (HSS) at various pH values. **(B)** pro-TrCB1.1 and pro-TrCB1.4 were incubated at pH 4.5 in the presence or absence of 20 μg/ml sulfated polysaccharides (SP). **(C)** SDS-PAGE of (pro-)TrCB1.1 and (pro-)TrCB1.4 incubated in the presence of 10 μg/ml DS at pH 4.5 for 30 min (lanes 2). Lanes 1 contain the pro-enzyme without DS. **(D)** SDS-PAGE of (pro-)TrCB1.4 incubated in the presence of 10 μg/ml HSS at pH 4.5 for 30 min (lane 2). Lane 1 contains the pro-enzyme without HSS. M, markers of molecular size (kDa). Arrowheads point to processed enzymes, which were rapidly autodegraded under the experimental conditions. A gel for TrCB1.1 with HSS is not available.

The pepTrCB1.1 was also active toward Z-Arg-Arg-AMC with a similar optimum at pH 5.5 (not shown). No activity toward this substrate was detected for pepTrCB1.6G/C as well as for pepTrCB1.4.

The activities of pepTrCB1.1 and pepTrCB1.4 were inhibited by E-64 and the cathepsin B-specific irreversible inhibitor CA-074 (**Figure 6B**). In the case of pepTrCB1.1, <1% of the activity remained after inhibition by both inhibitors at 0.1 μM concentration. Approximately 1% activity of pepTrCB1.4 remained after treatment by 10 μM E-64, while 0.1 μM CA-074 completely suppressed the activity of TrCB1.4. On the other hand, E-64 and CA-074 had little and no effect, respectively, on the activity of pepTrCB1.6G/C (**Figure 6B**). Even a high concentration (100 μM) of E-64 reduced the activity of pepTrCB1.6G/C only to 34%.

The peptidyl dipeptidase activity of pepTrCB1 was investigated using Bz-Gly-His-Leu substrate (**Figure 7**). Both pepTrCB1.1 and pepTrCB1.4 exhibited carboxypeptidase activity

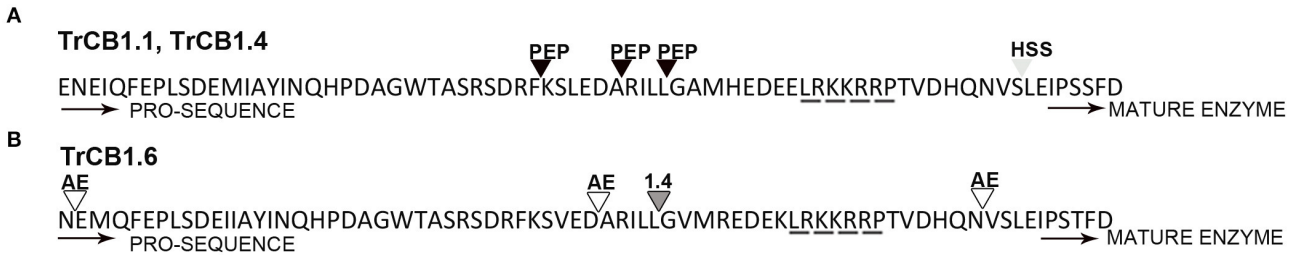

FIGURE 2 | Cleavage sites in the processed pro-sequences of TrCB1 forms. **(A)** cleavage sites determined by Edman degradation in pepsin-processed pepTrCB1.1 and pepTrCB1.4 (black triangles); cleavage site in TrCB1.4 auto-processing product resulted after 1-h incubation in the presence of heparin sodium salt (light gray triangle). **(B)** cleavage sites in rIrAE-activated TrCB1.6wt resulted after 1-h incubation at pH 5 (empty triangles); cleavage site in TrCB1.6wt after processing with pepTrCB1.4 (dark gray triangle). "Heparin-binding" (Horn et al., 2011) motif responsible for glycosaminoglycan binding to the pro-sequences is underlined by a dashed line.

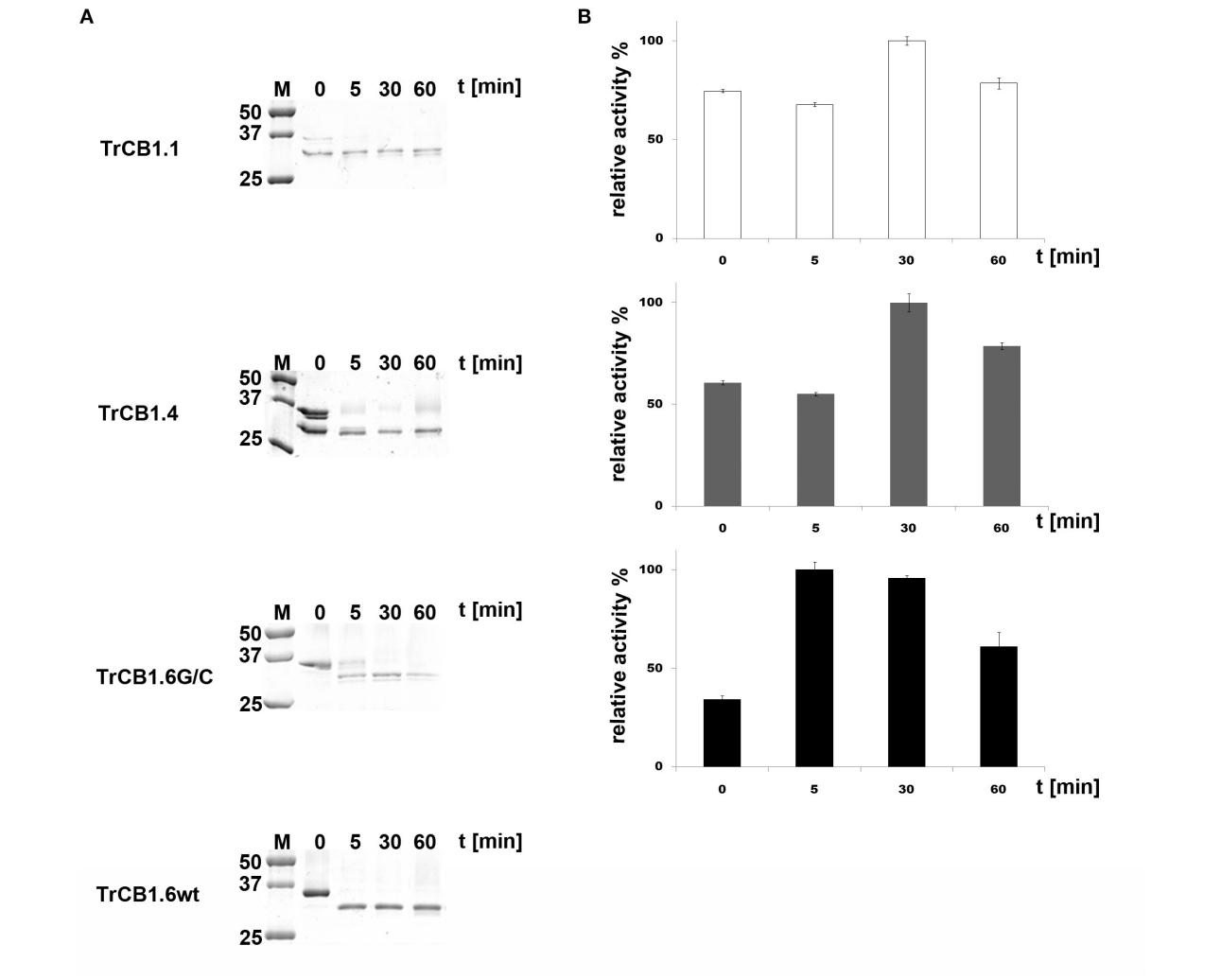

FIGURE 3 | Trans-processing of the TrCB1 zymogens by pepsin. TrCB1 forms were treated with pepsin immobilized on agarose beads at 37°C (TrCB1.6G/C at RT), and the reaction mixtures were analyzed at the indicated time points. **(A)** The processed forms were resolved by SDS-PAGE. M, markers of molecular size. **(B)** Peptidolytic activities of TrCB1.1, TrCB1.4, TrCB1.6G/C generated after processing were measured in a kinetic assay with the substrate Z-Phe-Arg-AMC (25 μM). The results are means ± SD of two triplicate assays and are expressed as percentage of relative activity in the sample.

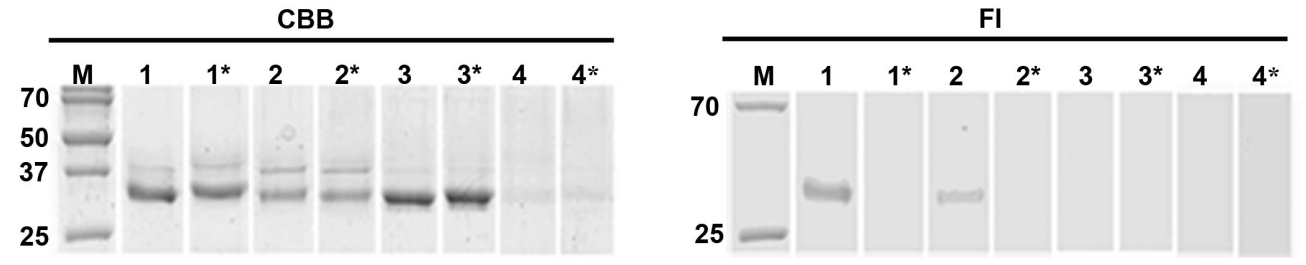

FIGURE 4 | Active site labeling of the pepTrCB1 forms by the cysteine peptidase-specific probe DCG-04. The pepTrCB1 forms were treated with the probe BODIPY green-DCG-04, resolved by SDS-PAGE, and stained with Coomassie Brilliant Blue (CBB) or visualized with a fluorescence imager (FI). (1) TrCB1.1, (2) TrCB1.4, (3) TrCB1.6wt, (4) TrCB1.6G/C. Samples marked by asterisk were pre-incubated with E-64 inhibitor prior to labeling. M, markers of molecular size (kDa). Note the autodegradation of TrCB1.6G/C (lanes 4 and 4*) during incubation with the probe/inhibitor, resulting in weak protein bands.

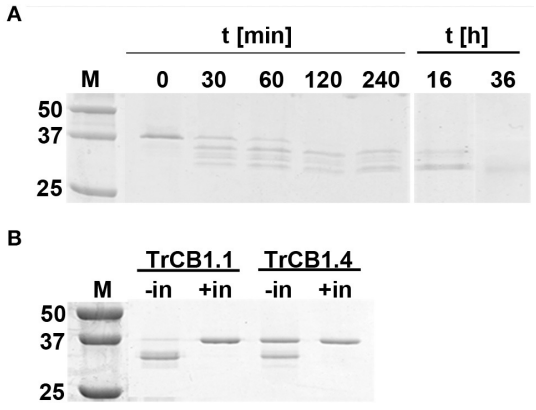

FIGURE 5 | Trans-processing of pro-TrCB1.6wt by IrAE, TrCB1.1, or TrCB1.4. **(A)** The pro-TrCB1.6wt was incubated at pH 5 with purified yeast medium containing activated IrAE for various time periods at 37°C. The TrCB1.6wt processing products were resolved by SDS-PAGE. **(B)** The TrCB1.6wt zymogen was incubated overnight at 37°C with pepTrCB1.1 or pepTrCB1.4 in the presence of dextran sulfate (10 μg/ml). Processing of pro-TrCB1.6wt was visualized by SDS-PAGE, gel stained with Coomassie Brilliant Blue. Processing was inhibited with 10 μM inhibitor of cysteine peptidases E-64. −/+ in denotes absence/presence of the inhibitor. M, markers of molecular size (kDa).

against this substrate with an optimum at pH 4.5 (**Figure 7A**). The activities were efficiently inhibited by 10 μM E-64 and CA-074 (**Figure 7B**). On the other hand, pepTrCB1.6G/C had very low activity against the exo-substrate at pH 5 and 5.5, and no activity occurred at pH ≤ 5. The pepTrCB1.6wt had no exopeptidase activity at any tested pH (not shown).

Repeated attempts to inhibit endopeptidase activity of activated TrCB1 forms (ds/pepTrCB1.1, ds/pepTrCB1.4, ds/pepTrCB1.6G/C) by chicken egg white cystatin and human cystatins B and C in the presence of Z-Phe-Arg-AMC substrate failed under the conditions used (not shown).

Formation of Peptidase Complex With Chicken Egg Cystatin

Overnight incubation at 4°C, but not a 15 min treatment at room temperature (RT) of auto-activated TrCB1 forms (dsTrCB1.1, dsTrCB1.4, dsTrCB1.6G/C) or pepTrCB1.6wt with chicken egg white cystatin resulted in the formation of new weak bands around ≈40 kDa that roughly matched the predicted size of the enzyme/cystatin complex (**Figure 8**). Only a small percentage of the enzymes bound to cystatin. On the other hand, chicken cystatin seemed to protect TrCB1.6G/C from autodegradation. Binding of human cystatins B and C to the pepsin-processed enzymes was not confirmed within 2 h of incubation at RT (not shown).

P1–P4 Specificity Profile of TrCB1.1 and TrCB1.4

The preference of TrCB1.1 and TrCB1.4 for amino acid residues at P1 – P4 positions of the substrates was determined using positional scanning synthetic combinatorial library (PS-SCL) (**Figure 9**). At P1 position, both enzymes preferred small basic amino acids Lys and Arg, but TrCB1.4 showed more distinct inclination to Arg. A stronger preference for Met, the non-natural norleucine (similar unbranched chain structure as Lys), and Lys at this position was also noted for TrCB1.4.

At P2 position, TrCB1.4 exhibited preference for non-polar aliphatic amino acids (Ala, Val, Ile), with overriding preference for Ala. In addition, TrCB1.4 was also able to accommodate polar residues (Ser, Thr and norLeu). TrCB1.1 showed greater promiscuity at P2, with a low preference for large non-polar residues, and could not accept Gly or Pro. However, it was able to hydrolyze substrates with Arg at P2.

Both enzymes displayed similar preferences for residues at P3. The highest preference was recorded for Leu, but the enzymes also well-accepted other amino acid residues (norLeu, Tyr, and Met) at this position. Screening at P4 position revealed a broader substrate specificity, with all amino acid residues accepted by both recombinant enzymes.

Hydrolysis of Protein Substrates by TrCB1 Forms

The pepTrCB1.1 and pepTrCB1.4 were able to degrade all given macromolecular substrates [albumin, IgG, fibrinogen, collagen type I and type IV, myosin, myelin basic protein (MBP) and hemoglobin (Hb)] at pH 5.5 (**Figure 10A**). While MBP was efficiently hydrolyzed between pH 4.5 and 6.5 with

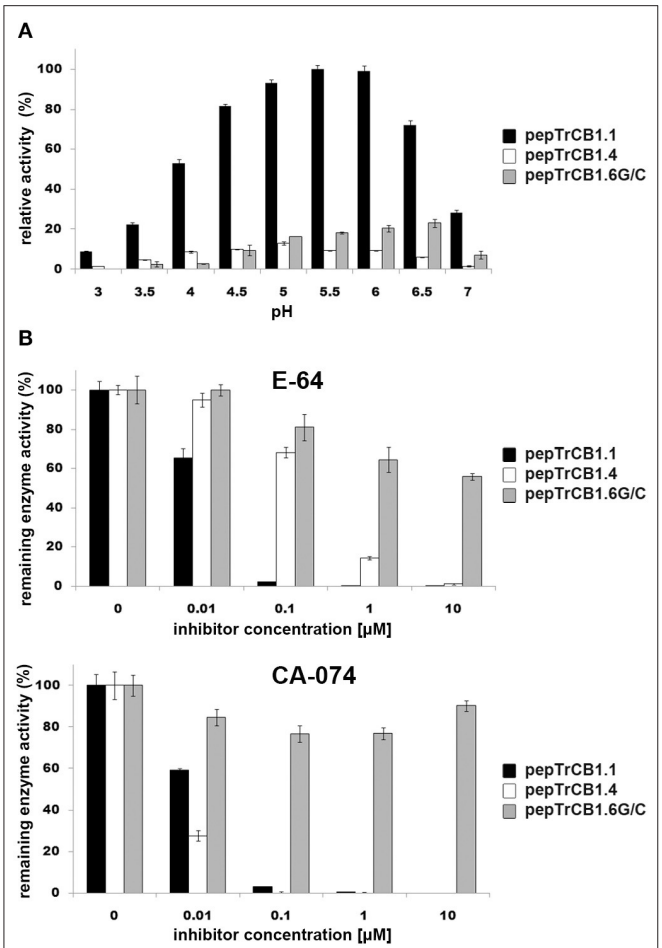

FIGURE 6 | pH optima of the pepTrCB1 forms and effect of inhibitors on enzyme endopeptidase activity. **(A)** The effect of pH on peptidolytic activity of pepTrCB1.1, pepTrCB1.4, pepTrCB1.6G/C toward the fluorogenic peptidyl substrate Z-Phe-Arg-AMC was measured in 50/100 mM CPB (pH range 3–7) with 2 mM DTT. Data are expressed as percentage of relative activity in the sample. **(B)** Effect of cysteine peptidase inhibitors (E-64, CA-074) on the activity of pepTrCB1.1, pepTrCB1.4, pepTrCB1.6G/C toward Z-Phe-Arg-AMC substrate. Inhibition assays were carried out at pH optimum of the particular enzymes. The values are mean ± SD of two triplicate assays.

the best result at pH up to 5.5, Hb was poorly degraded only at lower pH by both recombinant enzymes (**Figure 10B**). The pepTrCB1.6G/C could degrade most of the tested protein substrates (except for IgG) at pH 5.5 (**Figures 10A,B**), while pepTrCB1.6wt completely lacked such activity (not shown). Significant autodegradation of TrCB1.6G/C was noticed under the conditions of the experiment, which could to a certain degree distort the quantitative interpretation of the results.

Identification of Hemoglobin and MBP Fragments by Mass Spectrometry

The MS sequence analysis of peptides resulting from action of TrCB1 forms (pepTrCB1.1, pepTrCB1.4, and pepTrCB1.6G/C) on polypeptides of turkey Hb and human MBP was performed to determine enzymes' cleavage site specificities. Detected peptide fragments are listed in **Supplementary Table 1** (only the fragments which were not detected at time zero are included). The identified cleavage sites in the sequence of α- and β-subunit of Hb and in MBP are shown in **Figures 11A,B**, respectively. Several fragments were obtained with pepTrCB1.1, which did not occur in controls. After 5 min of reaction, eight cleavage sites (i.e., "initial cleavage sites") were identified. The cleavage map shows that TrCB1.1 obviously preferred Lys at P1 position in the two macromolecular substrates, followed by Arg, which corresponds with the results of PS-SCL (see above). Other amino acids could also be accommodated in this position (Leu>Thr/Glu>Phe/Ser>Gln/Ser/Gly/Met/Ala/Val/Phe). At the P2 position, the preference was not that distinctive: pepTrCB1.1 slightly preferred polar uncharged amino acids (Ser/Thr/Glu) and the positively charged Lys, closely followed by Phe/Ala/Leu/Pro/His/Asp, and less Arg/Thr/Ile/Glu. Although PS-SCL results have shown that Pro is unfavorable at P2, two of the fragments generated by pepTrCB1.1 contained Pro at this position. Only one fragment [amino acids 1 – 16 from Hbα (A) subunit] could be reliably detected after hydrolysis of Hb by pepTrCB1.4. While the same fragment generated by pepTrCB1.1 appeared after 5 min of incubation, the fragment produced by pepTrCB1.4 appeared as late as in 2 h. The preference for amino acids in both P1 and P2 positions of macromolecular substrates was not clearly defined and no obvious pattern was observed. Also, no cleavage sites distinct from those produced by TrCB1.1 were detected. Finally, no fragment was detected after incubation of Hb with TrCB1.6G/C and less fragments were also identified after cleavage of MBP comparing to the action of the other enzymes. Moreover, all the cleavage sites were identical with those produced by pepTrCB1.1 (**Figure 11**).

Differential Expression of TrCB1 Forms

Based on the differential transcriptional analysis, TrCB1.1, TrCB1.2, TrCB1.3, TrCB1.5, and TrCB1.6 were stated as being significantly upregulated in schistosomula, when compared to cercariae. TrCB1.4 was not detected in *T. regenti* juvenile transcriptomes at significant levels. Transcription profile of cercariae was limited only to forms TrCB1.1, TrCB1.2, and TrCB1.6 with low counts per million (CPM) rate. In schistosomulum stage, TrCB1.1, TrCB1.2, and TrCB1.6 genes were transcribed at a similar level, followed by TrCB1.3 and TrCB1.5 (**Figure 12**).

In vitro Exposure of Murine Astrocytes, Microglia and RAW 264.7 Macrophages to Pro-TrCB1.6wt

Production of NO by murine astrocytes, microglia and RAW 264.7 macrophages was examined after treatment with pro-TrCB1.6wt for 48 h. As revealed by Griess assay, pro-TrCB1.6wt neither triggered NO production in any of the cells nor reduced NO production in cells concurrently treated by LPS (**Supplementary Figure 4A**). Similarly, no effect on changes in cell viability was observed (**Supplementary Figure 4B**).

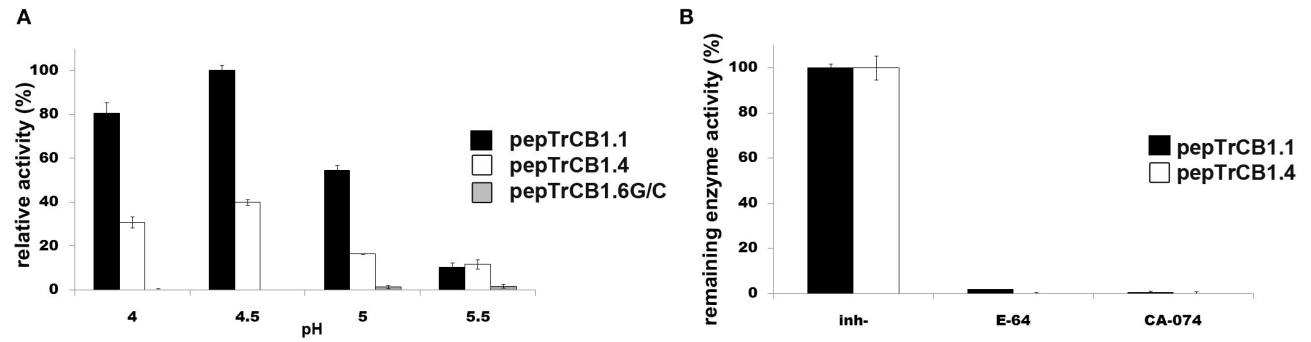

FIGURE 7 | Peptidyl-dipeptidase activity of the pepTrCB1 forms and the effect of inhibitors. **(A)** pepTrCB1 forms were incubated with Bz-Gly-His-Leu substrate in 50/100 mM CPB (pH 4–5.5) containing 2 mM DTT. The appearance of arising free amino groups was monitored by reaction with fluorescamine (0.05 mg/ml). **(B)** The effect of 10μM E-64 and CA-074 on carboxypeptidase activity of pepTrCB1.1 and pepTrCB1.4 assayed at pH optima of the particular enzymes. Values are means of two triplicate assays ± SD.

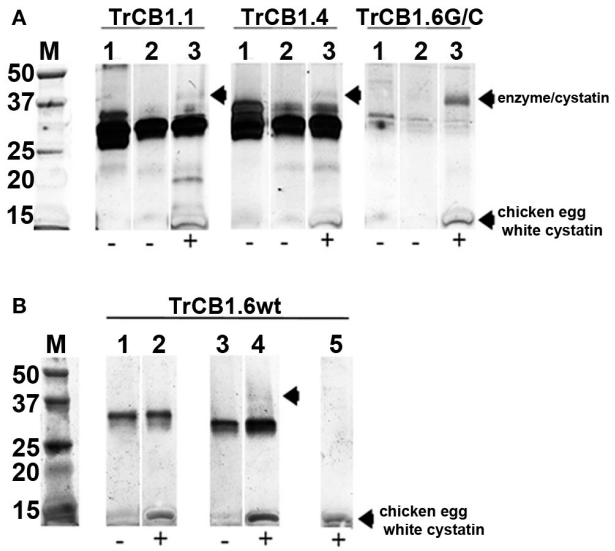

FIGURE 8 | Complex formation between TrCB1 forms and chicken egg white cystatin monitored by SDS-PAGE. **(A)** Auto-activated dsTrCB1.1, dsTrCB1.4, dsTrCB1.6G/C after overnight incubation at 4°C in the presence/absence of chicken egg white cystatin. After incubation, the samples were mixed with non-reducing sample buffer and resolved by SDS-PAGE. Proteins were visualized by silver staining. (1) pro-TrCB1, (2) dsTrCB1 without cystatin, (3) dsTrCB1 with cystatin. Note the autodegradation of TrCB1.6G/C (lanes 1 to 3) happening during incubation, resulting in weak protein bands.
(B) pepTrCB1.6wt after overnight incubation at 4°C in the presence/absence of cystatin. (1) pro-TrCB1.6wt, (2) pro-TrCB1.6wt with cystatin, (3) pepTrCB1.6wt, (4) pepTrCB1.6wt with cystatin, (5) control lane with cystatin only. Arrows point to peptidase/cystatin complex and/or to cystatin (in section B, lane 5). M, markers of molecular size (kDa).

DISCUSSION

Expression of TrCB1.1 and TrCB1.4 in *Pichia* led to production of zymogens which were partly auto-processed in the yeast media, producing a small proportion of partially activated intermediates. Further processing of the recombinant (pro-) enzymes was therefore necessary. Although Dvořák et al. (2005) obtained auto-processed TrCB1.4 with an eight-fold increase in activity after overnight incubation at acidic pH, we found an insufficient increase of activity in both TrCB1.4 and TrCB1.1 using this approach. Auto-activation of recombinant cathepsin B1 from *Schistosoma mansoni* (SmCB1) is forced by negatively charged polysaccharides, such as sulfated glycosaminoglycans. Their binding to the SmCB1 pro-sequence destabilizes the pro-sequence/enzyme core junction (Jílková et al., 2014) and is mediated by a specific α-helix that contains a "heparin-binding motif"-XBBXBBX- (B is a basic amino acid and X refers to a hydropathic residue) (Horn et al., 2011; Jílková et al., 2014). A sequence alignment revealed that this motif is specific for trematode cathepsins B involved in the digestion of host proteins (Horn et al., 2011). As this motif occurs in all TrCB1 pro-sequences (**Figure 2**), sulfated polysaccharides facilitated auto-activation of both TrCB1.1 and TrCB1.4 (**Figure 1**). However, unlike in SmCB1 (Jílková et al., 2014), the products were relatively unstable at all tested pH values. Therefore, we did not use them in most subsequent experiments. However, there seems to be similarity between SmCB1 and TrCB1 in terms of glycosaminoglycan-mediated acceleration of auto-activation *in vivo*.

In vitro trans-processing of pro-TrCB1 by pepsin was applied according to previously published experience (Lipps et al., 1996; Law et al., 2003). N-terminal sequencing showed that the pro-sequence of TrCB1 is cleaved at carboxyterminal side of Phe and Leu residues, which is typical for pepsin (Hamuro et al., 2008). Unfortunately, none of the cleavage sites resulted in fully mature enzymes. Even prolonged incubations or higher concentrations of pepsin did not give better results. On the other hand, partial removal of the TrCB1 pro-sequences was sufficient for hydrolytic activity toward macromolecular substrates. The advantage of this method was that agarose beads-immobilized pepsin could be easily removed by centrifugation. Therefore, pepsin-processed enzymes were used in most experiments. The asparaginyl endopeptidase (AE, legumain) was confirmed as an

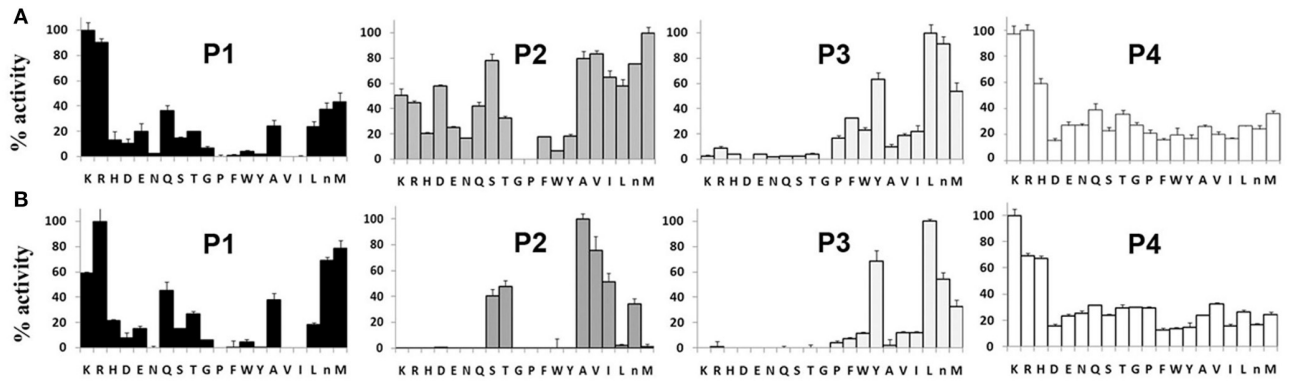

FIGURE 9 | P1–P4 amino acid preferences of pro-TrCB1.1 and pro-TrCB1.4 determined by positional scanning of a synthetic combinatorial library. **(A)** Substrate library of pro-TrCB1.1. **(B)** Substrate library of pro-TrCB1.4. P4-P1 represent the four amino acid side chains recognized by peptidase on the N-terminal side of the cleaved peptide bond. The X-axis indicates the 20 amino acids held constant at each position, designated by the single-letter code (n, norleucine). The Y-axis shows relative preference for the particular amino acid when 100% represents the most preferred amino acid residue.

endogenous *trans*-processing enzyme activating other digestive peptidases in human schistosomes (Sajid et al., 2003) and three transcripts coding this enzyme were also found in *T. regenti* transcriptome (Leontovyč et al., 2016, 2019). We were able to *trans*-process the pro-TrCB1 using tick IrAE, obtaining several intermediates during short-time incubations, but this method was not widely used in our experiments, since it provided low yields.

The substrate specificity of papain-like peptidases seems to be largely determined by the composition of the S2 pocket (Choe et al., 2006), in which the amino acid residues interact with an amino acid residue at P2 position of the peptide substrate (Schechter and Berger, 1967). From this point of view, it was interesting that amino acid preferences of TrCB1.1 and TrCB1.4 differed at the P2 position, while the preference for other positions was quite similar. Using PS-SCL, TrCB1.1 showed broader preferences at P2, having a relative preference for aliphatic amino acids (excepting Gly) over aromatic residues (Phe, Trp, and Tyr). Interestingly, TrCB1.1 had a P2 preference more similar to *S. mansoni* cathepsin B2 than to SmCB1 (Choe et al., 2006), although it has higher overall sequence identity and function similar with SmCB1. By contrast, TrCB1.4 displayed relatively narrow P2 specificity with an overriding preference for aliphatic Ala. In agreement with a previous study (Dvořák et al., 2005), TrCB1.4 did not show any noticeable activity toward substrates with Arg at P2. Surprisingly, the library assay also indicated that even Met is not a favored amino acid for TrCB1.4 at P2, which is not typical of cathepsins B (Choe et al., 2006). On the contrary, Met is the most favored amino acid accepted by TrCB1.1 at P2. Also, Z-Phe-Arg-AMC was a poor substrate for TrCB1.4, compared to TrCB1.1, wich corresponds with the results of PS-SCL.

Both TrCB1.1 and TrCB1.4 degraded a range of macromolecular substrates, such as albumin, IgG, fibrinogen, collagen type I, collagen type IV, myosin, myelin basic protein (MBP), and hemoglobin (Hb). Divergent specificities for the selected protein substrates were not observed between the two enzymes, except for MBP, which was obviously more efficiently hydrolysed by TrCB1.1. In line with previous work (Dvořák et al., 2005), both enzymes poorly degraded Hb. The authors assumed that this may indicate an adaptation to the unique migratory route of *T. regenti* through the CNS. On the other hand, Hb becomes an important source of nutrition for adult worms located in the nasal mucosa of ducks (Horák et al., 1999; Blažová and Horák, 2005; Chanová and Horák, 2007) and the level of TrCB1.1 gene transcription was comparable in *T. regenti* schistosomula and adults (Dolečková et al., 2010).

It is likely that active forms of TrCB1 can work in a cooperative network with other peptidases (e.g., cathepsins L and cathepsins D) in the parasite's gut, which could preferentially initiate hydrolysis of ingested host proteins (including Hb). Such a network was characterized in *S. mansoni*, where the cysteine peptidases (cathepsins L and B) initiate digestion of albumin, whereas cathepsin D is responsible for the primary cleavage of Hb, although cysteine peptidases may aid in this. Cathepsins B also act as potent (carboxy-) exopeptidases (Illy et al., 1997). Several transcripts of cathepsin L and D genes were found also in the transcriptome of *T. regenti* schistosomula (Leontovyč et al., 2016).

MS/MS analysis of protein substrate digests revealed that TrCB1.1 was more efficient in the cleavage of Hb than TrCB1.4 in terms of the number of produced cleavage sites. Nevertheless, intensive signals of undigested Hb were observed even after 2 h of incubation with the enzymes under given conditions (substrate:enzyme ratios). The cleavage sites produced by TrCB1.1 exhibited faintly defined specificity. Thus, it seems that TrCB1 enzymes may contribute to the degradation of Hb, likely after initial cleavage by other endopeptidase(s). Surprisingly, in spite of the differences in amino acid preferences at P2 of the substrates seen in PS-SCL, there was a 100% overlap in cleavage sites produced by TrCB1.4 in the molecule of MBP, with those made by TrCB1.1, which only produced a few more cuts. Unexpectedly, we were not able to detect more fragments of Hb

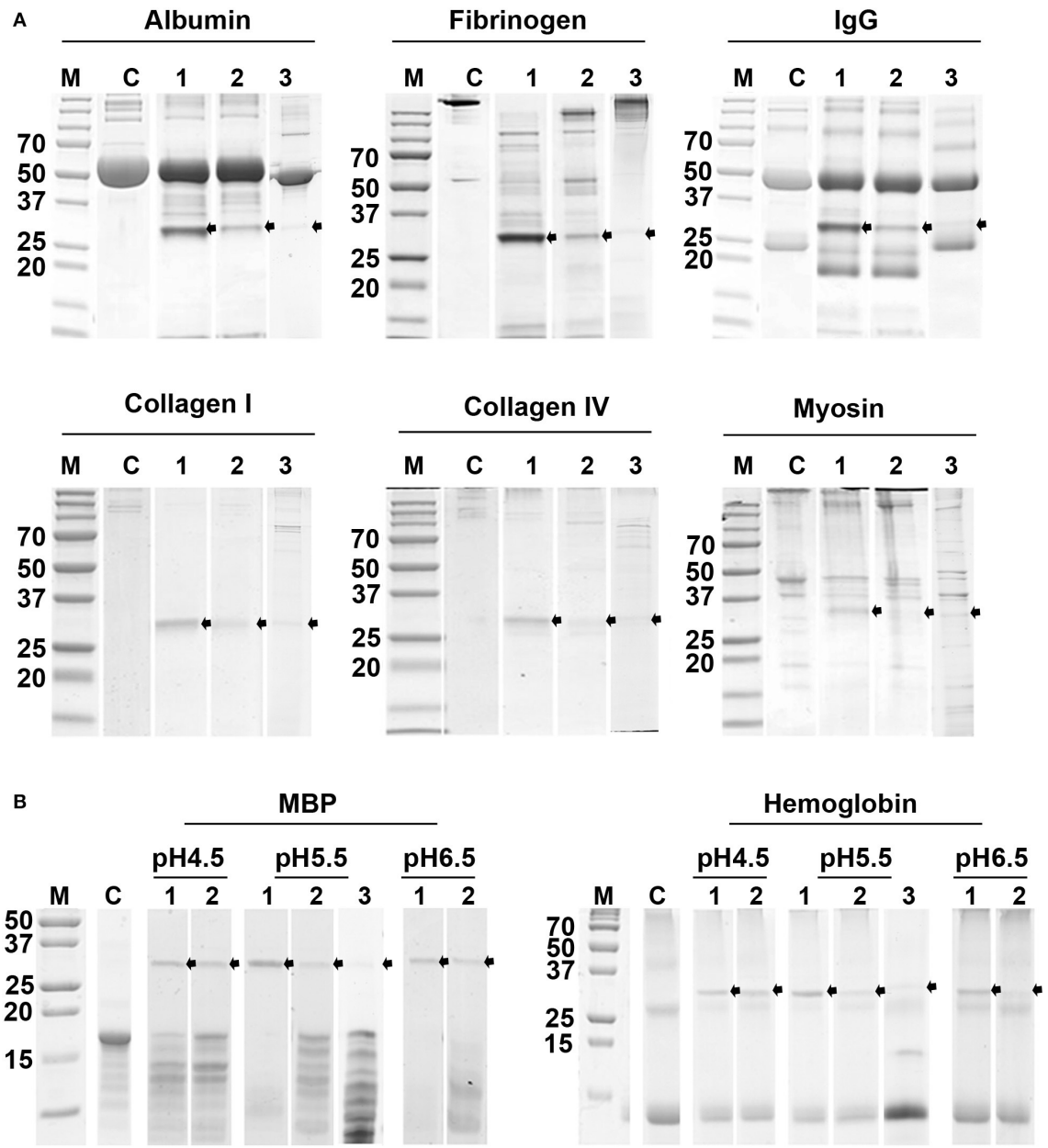

FIGURE 10 | Digestion of selected protein substrates by recombinant TrCB1 forms. The substrates (0.5 mg/ml) were incubated for 6 h with pepTrCB1 forms (1 μg) at 37°C. Aliquots (20 μl) of the resulting hydrolysates were separated by SDS-PAGE and stained with Coomassie Brilliant Blue. **(A)** Incubation of pepTrCB1.1, pepTrCB1.4 and pepTrCB1.6G/C (lanes 1–3, respectively) at pH 5.5 with albumin, fibrinogen, collagen type I, collagen type IV and myosin. **(B)** Incubation of pepTrCB1.1, pepTrCB1.4 and pepTrCB1.6G/C (lanes 1–3, respectively) at pH 4.5–6.5 with myelin basic protein (MBP) and hemoglobin. C, controls (substrates without enzyme). The starting amount of the peptidases was equal in all samples; note the autodegradation of the enzymes under given conditions, resulting in weaker bands especially in TrCB1.6G/C. Arrows indicate the visible peptidase bands in the gels. M, markers of molecular size (kDa).

produced by TrCB1.4, probably due to the low activity of the enzyme toward this substrate.

Although recombinant TrCB1.4 was previously obtained by employing mRNA from 5 to 6 days old schistosomula (Dvořák et al., 2005), our differential expression analysis revealed that transcripts of TrCB1.4 were not present at detectable levels in the transcriptome of cercariae or 7 days old schistosomula. This suggests that TrCB1.4 is not stably expressed in schistosomula and it is likely not essential for migrating worms in the definitive host. However, this, together with a distinct specificity toward amino acid residues of protein substrates, points to "a gene in waiting hypothesis." The redundancy of enzyme isoforms arisen by duplication and subsequent divergence of coding genes "provides a potential mechanism to alter the substrate

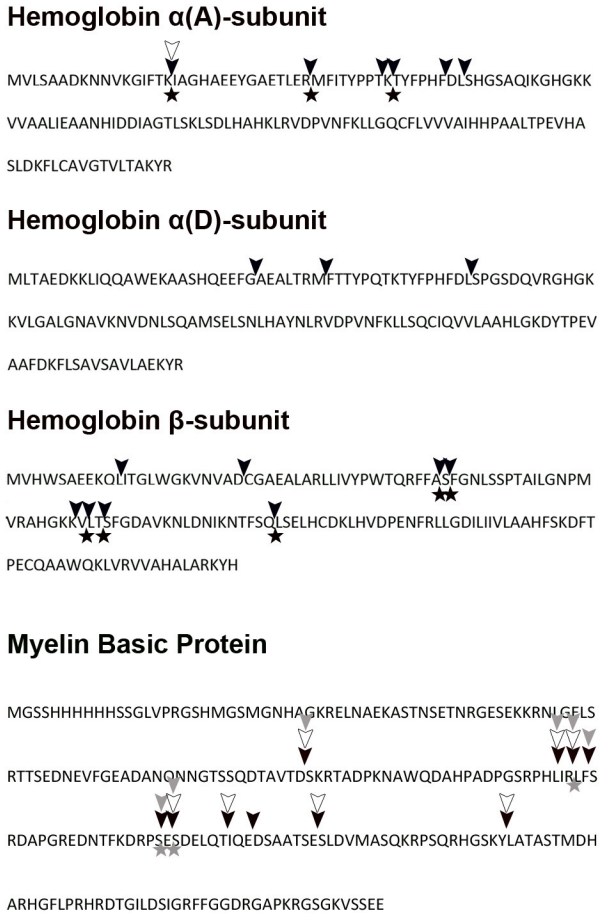

FIGURE 11 | Hemoglobin and myelin basic protein pepTrCB1 cleavage map. **(A)** Turkey hemoglobin and **(B)** recombinant human myelin basic protein were digested *in vitro* with pepTrCB forms at pH 5.5. The fragments were identified by mass spectrometry, and corresponding cleavage sites are indicated in hemoglobin and myelin basic protein sequences: cleavage by pepTrCB1.1 (black triangles), cleavage by pepTrCB1.4 (empty triangles), cleavage by pepTrCB1.6G/C (gray triangles). The cleavage map combines data from digests at three time points. The initial cleavage sites (after 5 min of reaction) are marked with asterisks. Only unique fragments which were not detected at time zero are included.

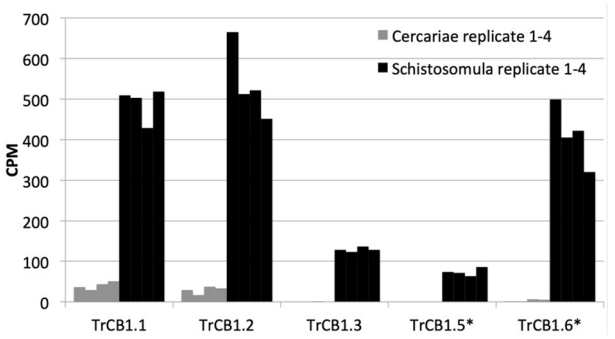

FIGURE 12 | Expression profiles of TrCB1 forms. Bar charts show the level of expression of different forms of cathepsin B1 in cercariae (gray) and schistosomula (black) of *T. regenti* based on mapping of RNA-seq data obtained from cercariae and 7-days old schistosomula. The level of expression is expressed in counts per million (CPM) of mapped reads. Each column is a cluster of four biological replicates. Inactive paralogs are indicated by asterisks.

specificity of the protease. Such a reservoir of genes would provide protection for the parasite against host adaptation or enhancement and expansion of host range" (Salter et al., 2002). Unfortunately, data on the level of TrCB1.4 expression in adult *T. regenti* is not available, thus currently it can be neither excluded nor confirmed if TrCB1.4 is employed by the mature worms in order to increase the total amount or broaden the specificity of peptidases produced in the gut.

Surprising and unexpected was the relatively high expression of TrCB1 inactive paralogs, especially TrCB1.6, which was comparable to that of the active enzymes such as TrCB1.1 in the stage of schistosomulum. This raises a serious question about the importance of the catalytically inactive forms for the biology of the parasite. As expected, recombinant pro-TrCB1.6wt was incapable of auto-processing. However, the presence of an Asp residue in the pro-sequence five amino acids upstream the beginning of the mature domain indicated a potential for processing by an asparaginyl endopeptidase (AE), similarly to SmCB1 (Sajid et al., 2003) and also TrCB1.1 (Dvořák et al., 2005). This was confirmed by incubation with recombinant IrAE (Sojka et al., 2007), which resulted in a single band of ≈30 kDa formed through several intermediates arisen by specific cleavage at sites with Asn and Asp at P1 position. This suggests that the transcripts coding for AE found in the transcriptome of *T. regenti* (Leontovyč et al., 2016, 2019) may be involved in the removal of the pro-sequence from pro-TrCB1.6 *in vivo*. Also, trans-processing of pro-TrCB1.6wt with pepsin-processed active forms pepTrCB1.1 and pepTrCB1.4 resulted in the formation of a major (≈32 kDa) and a minor (≈30 kDa) product in both cases. The 32-kDa product was generated by cleavage between Leu^{64} and Gly^{65}(pro-TrCB1.6 numbering). Unfortunately, the N-terminal sequence of the 30 kDa product could not be determined, but the size almost corresponded to that of the product obtained by IrAE processing. This data indicated that *in vivo* trans-processing of inactive forms of TrCB1 by the active ones is also possible.

For a more detailed view on the inactive TrCB1.6, a site-directed mutant TrCB1.6G/C (reverse $G^{29}C$ mutation) was expressed in *P. pastoris* system and partially characterized in comparison to the recombinant TrCB1.1 and TrCB1.4. The reverse mutation (Gly^{29} > Cys^{29}) of TrCB1.6wt resulted in the restoration of hydrolytic activity toward Z-Phe-Arg-AMC, but the mutant enzyme TrCB1.6G/C was not able to cleave cathepsin B-specific substrate Z-Arg-Arg-AMC, much like TrCB1.4. This feature is caused by uncharged Ala^{245} (TrCB1.4) or Leu^{245} (TrCB1.6) at the position, where negatively charged Glu^{245} typically occurs in the S2 pocket of cathepsins B (Dvořák et al., 2005).

Moreover, TrCB16G/C activity was not inhibited by CA-074 inhibitor (mostly specific for cathepsins B). Because the occluding loop of cathepsins B is important for the interaction with CA-074 (Yamamoto et al., 1997), in which the prolyl component of CA-074 forms hydrogen bonds with the conserved

His^{110} and His^{111} of the loop (bovine cathepsin numbering), the explanation may be that the latter histidine is substituted by an Asn residue in TrCB1.6, which causes the lack of susceptibility to CA-074, and also the virtual lack of exopeptidase activity. In human cathepsin B, two salt-bridge interactions between the occluding loop and the main body of the mature enzyme (Asp^{22}-His^{110} and Arg^{116}-Asp^{224}) anchor the loop down in the active site, and unpaired His^{111} forms a salt bridge with the C-terminus of an exo-substrate (Musil et al., 1991; Illy et al., 1997). According to another study (Krupa et al., 2002), His^{111} is not critical for exopeptidase activity, although the level of this activity might be somewhat impaired by the absence of this residue. This means that also other substitutions in the sequence of TrCB1.6 are probably responsible for the disruption of exopeptidase activity in the reverse $Gly^{29} > Cys^{29}$ mutant. The substitution $Asp^{224} > Gly^{224}$ in the sequence of TrCB1.6, when compared to the active TrCB1 forms or other cathepsins B, seems a likely candidate, since the removal of the salt bridges between the loop and the enzyme core effectively eliminates exopeptidase activity (Illy et al., 1997).

The presence of the occluding loop was proposed to be responsible for the generally lower affinities of cathepsins B to cystatins, compared to other endopeptidases of papain-like family with open active sites (Pavlova et al., 2000). Reduced electrostatic interaction between the loop and the body of the mature TrCB1.6 might cause greater affinity to cystatin. Based on this, we hypothesized that the inactive TrCB1.6 might sequester cystatins and thus protect active TrCB1 forms or other cysteine endopeptidases present in the vicinity of this paralog from inactivation by host cystatin(s). However, repeated attempts to inhibit the peptidolytic activity of TrCB1.1 and 1.4 toward Z-Phe-Arg-AMC with chicken and human cystatins failed. Nevertheless, SDS-stable enzyme/inhibitor complexes were formed after overnight incubation of all processed TrCB1 forms with chicken cystatin, including the catalytically inactive pepTrCB1.6wt. No complex was formed with the non-processed pro-TrCB1.6wt, which suggests that the pro-sequence blocked cystatin binding to the "active site" of the paralog. Binding of chicken cystatin to TrCB1.6G/C resulted even in larger proportion of the complex and prevented the enzyme from autodegradation. Surprisingly, human cystatins B and C did not form a complex with any of the TrCB1 forms after 2 h incubation.

The TrCB1.6G/C mutant degraded most of the protein substrates similarly with TrCB1.4. It seems that the binding pockets (except for the substitution in the active site) of TrCB1.6 are intact; thereby TrCB1.6 might still preserve substrate-binding properties. Thus, the possibility that the TrCB1.6 paralog competes with the active homologs for substrates and participates in this way in post-translational regulation of the activity of these enzymes cannot be omitted.

In a previous study, active TrCB1.1 and TrCB2 were shown to trigger nitric oxide production in murine astrocyte cultures, which could be harmful to both the parasite and the host (Macháček et al., 2016). Therefore, we wanted to test whether also the TrCB1.6 paralog has any effect on the immune response of the host. We exposed immune cells naturally occurring around *T. regenti* schistosomula migrating in murine spinal cord (astrocytes, microglia, and macrophages) (Lichtenbergová et al., 2011) to recombinant TrCB1.6wt *in vitro*. Surprisingly, it neither stimulated nor inhibited nitric oxide production in the naïve or LPS-treated cells, respectively. TrCB1.6wt also did not alter the viability of these cells. Perhaps, the hydrolytic activity of TrCB1.1 and TrCB2 was somehow responsible for the above stated effect.

In conclusion, the characterization of the two active forms TrCB1.1 and TrCB1.4 showed that the enzymes differ from each other in various properties. The levels of relative expression of TrCB1 paralogs in cercariae and schistosomula indicate that both the active forms and the inactive mutants are important for schistosomula migrating and developing in the definitive host, except for TrCB1.4. This enzyme form may be expressed even later in adult worms and this supposition needs to be further verified. The ability of the TrCB1.6 paralog with non-functional, Cys-to-Gly mutated active site to bind cystatin supports the hypothesis of inhibitor sequestration, although the efficacy appeared to be low in our experimental setup. Restoration of the active site of TrCB1.6 resulted in endopeptidase but not exopeptidase activity of the paralog. This elucidated important structural features in TrCB1.6 caused by other mutations in the occluding loop and the main body of the mature enzyme, which either regulate the exopeptidase activity of cathepsin B or may also adjust the affinity of cathepsin B to cystatins. Resolving of 3-D stucture of inactive TrCB1.6 + cystatin complex could help to elucidate the nature of the binding interactions, which is also the case of TrCB1.6G/C with restored active site. Finally, the inability of TrCB1.6 paralog to affect viability and production of nitric oxide by selected immune cells does not exclude other possible interactions with host physiology/immunity, which are a subject to further investigation in this unique and fascinating neurotropic schistosome.

MATERIALS AND METHODS

Parasite Material

Trichobilharzia regenti is routinely maintained in the Department of Parasitology, Charles University through domestic ducks (*Anas platyrhynchos* f. *dom.*) and lymnaeid snails (*Radix lagotis*) as definitive and intermediate hosts, respectively. Schistosomula were obtained from the spinal cord of infected ducks 9 days post-infection, washed in sterile phosphate-buffered saline (PBS) and used for RNA isolation and cDNA preparation as described previously (Dvořák et al., 2005).

Ethics statement: All procedures including animals were performed in concordance with the legislation of the Czech Republic (246/1992 and 359/2012) and the European Directive (2010/63/EU). All experiments were approved with the legal consent of the Professional Ethics Committee of the Faculty of Science, Charles University, i.e., the relevant institutional ethics committee for animal research, and of the Research and Development Section of the Ministry of Education, Youth, and Sports of the Czech Republic (approval no. MSMT-31114/2013-9). The animal facility, its equipment, animal welfare, and accompanying services, including the maintenance of experimental animals, have been approved by the Section of

Animal Commodities of the Ministry of Agriculture of the Czech Republic (approval no. 13060/2014-MZE-17214).

Expression and Purification of Recombinant TrCB1 Forms

Recombinant TrCB1.1 and TrCB1.4 were acquired from frozen glycerol stocks of *P. pastoris* transformants and expressed according to a previously published protocol (Dvořák et al., 2005). A column packed with Macro-Prep High S Support medium (Bio-Rad) was used for pre-purification from yeast media by cation exchange chromatography. TrCB1-containing fractions were treated with Endoglycosidase F1 (Calbiochem) following manufacturer's instructions and the pro-enzymes were further purified to homogeneity on a Mono S 5/50 GL column pre-equilibrated with 20 mM sodium acetate buffer pH 5.5 and employing 0–1 M linear NaCl gradient.

For the expression of recombinant wild-type TrCB1.6 (TrCB1.6wt), purified first-strand cDNA obtained from *T. regenti* schistosomula was used as a template for PCR with gene specific primers (**Supplementary Table 2**). PCR products were inserted into the expression vector pPICZα B (Thermo Fisher Scientific) and constructs were sequenced for verification (DNA Sequencing Laboratory, Faculty of Science, Charles University, Prague). To generate the Gly^{29}-to-Cys^{29} mutation (TrCB1.6G/C), the pPICZα B vector containing the sequence of TrCB1.6wt was mutagenized according to the instruction in QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene). The mutated vector was generated by PCR with oligonucleotide primers (each complementary to the opposite strand of the vector) containing the desired mutation (**Supplementary Table 3**). Following PCR, the product was treated with Dpn I endonuclease to digest methylated parental DNA template. Constructs were sequenced to verify the mutations.

The vectors containing sequences of TrCB1.6wt or TrCB1.6G/C were then electroporated into *P. pastoris* X-33 cells and recombinant proteins were produced as described above. Both TrCB1.6wt and TrCB1.6G/C were expressed with a polyhistidine affinity tag (6x His-tag) situated on the C-terminus. The proteins were purified by Ni^{2+} chelating chromatography (HisTrap^{TM}FF crude, GE Healthcare Life Science) equilibrated with 100 mM phosphate buffer pH 7 with 300 mM NaCl and 40 mM imidazole. A linear gradient of 100–500 mM imidazole in the same buffer was used for elution. The eluted proteins were deglycosylated (see above) and then purified to homogeneity by cation exchange FPLC (BioLogic, Bio-Rad) on a Mono S 5/50 GL column (GE Healthcare Life Science) equilibrated with 20 mM phosphate buffer pH 7.2 and using a linear gradient of 0–1 M NaCl.

Protein concentrations were measured by Quant-iT Protein Assay Kit (ThermoFisher Scientific). Molecular size of purified enzymes was estimated by SDS-PAGE and peptidolytic activity was measured with the peptidyl substrate Z-Phe-Arg-AMC (see section Peptidase Activity). Fractions after chromatography were stored at −80°C or lyophilized.

Theoretical MW/pI values and potential N-glycosylation sites were determined by the Compute Mw software and NetNGlyc 1.0 server available at the ExPASy web portal (https://www.expasy.org/tools) (Artimo et al., 2012).

Activation/Processing of Recombinant Enzymes

The possibilities of auto- and heterologous processing of pro-TrCB1 zymogens were investigated in order to produce mature (pro-sequence-free) fully active enzymes for subsequent analyses.

(A) The pro-TrCB1.1 and 1.4 zymogens (0.2 – 0.3 μg/μl) were incubated separately at 37°C for 5, 30, 60, 120, and 240 min in 50 mM sodium acetate buffer (pH 4, 4.5, 5, and 5.5) containing 2 mM DTT (Sigma Aldrich) either in the presence or absence of dextran sulfate (DS, MW 6.5–10 kDa; Sigma Aldrich) or heparin sodium salt (HSS, MW 6–30 kDa; Sigma Aldrich) at final concentrations of 10 or 20 μg/ml.

(B) The *trans*-processing of all recombinant pro-TrCB1 forms (0.2 – 0.3μg/μl) was catalyzed by pepsin immobilized on agarose beads (1U pepsin per 3.3 μg of recombinant pro-TrCB1; P0609 Sigma Aldrich) incubated at 37°C (pro-TrCB1.1, pro-TrCB1.4, and pro-TrCB1.6wt) or at RT (pro-TrCB1.6G/C) for 5, 30, 30, and 60 min in 50 mM sodium acetate buffer pH 4. The reaction was stopped by addition of 10 μM pepstatin A and the samples were centrifuged (1,000 × g, 5 min) to remove agarose-bound pepsin.

(C) The inactive pro-TrCB1.6wt was processed by recombinant asparaginyl endopeptidase of the hard tick *Ixodes ricinus*—IrAE (Sojka et al., 2007; kindly provided by Dr. Daniel Sojka, Institute of Parasitology, ASCR, Ceské Budějovice, Czech Rep.). One gram of the lyophilized yeast medium containing IrAE was dissolved in 7 ml of distilled water, concentrated and transferred to 500 μl of 50 mM sodium acetate buffer pH 4.5 containing 2 mM DTT using Amicon® Ultra 4 mL filters (MWCO 30 kDa; Millipore). IrAE was then incubated for 30 h at 37°C to allow self-processing, which was monitored fluorometrically with Z-Ala-Ala-Asn-AMC substrate (25 μM; Bachem). Subsequently, the pro-TrCB1.6wt (16.5 μg in 55 μl) was incubated with activated IrAE (5 μl, unknown concentration—see above) at 37°C in 50 mM sodium acetate buffer (pH4, 4.5, or 5) containing 2 mM DTT for 30,60, 120, and 240 min or 16 and 36 h.

(D) Alternatively, pro-TrCB1.6wt (0.2–0.3 μg/μl) was dissolved in 50 mM sodium acetate buffer pH 4.5 containing 2 mM DTT, 10 μM pepstatin A, and 10 μg/ml DS. Then the pepsin-activated pepTrCB1.1 or pepTrCB1.4 (2 μg/ml) was added overnight (up to 16 h) at 37°C. Controls contained 10 μM E-64.

Aliquots were taken from reaction mixtures at specified time intervals during incubations and catalytic activity of TrCB1 forms was measured with Z-Phe-Arg-AMC substrate (see section Peptidase Activity). Additionally, a shift in molecular weight was monitored by SDS-PAGE. N-terminal amino acid sequences of the proteins electroblotted onto PVDF membranes were identified by Edman degradation (Dr. Zdeněk Voburka, Laboratory of Medicinal Chemistry, IOCB, ASCR, Prague, Czech Rep.).

Active-site labeling of processed TrCB1 forms with a fluorescent probe BODIPY green-DCG-04 (Greenbaum et al., 2002; obtained from Dr. Doron C. Greenbaum, Tropical Disease Research Unit, Sandler Center for Basic Research in Parasitic Diseases, UCSF, USA—former affiliation) was carried out as described previously (Dvořák et al., 2005). The specificity of probe binding was verified by pre-incubation of controls with 10 μM irreversible cysteine peptidase inhibitor E-64 for 5 min.

Peptidase Activity

The peptidase activity of TrCB1 forms processed as described above (0.45 μg TrCB1 per well, i.e., 70 nM) was assayed with two synthetic fluorogenic substrates (25 μM, Bachem): Z-Phe-Arg-AMC, (cathepsin L/B substrate) and Z-Arg-Arg-AMC (cathepsin B-selective substrate) in 50/100 mM citrate/phosphate buffer (CPB), 2 mM DTT, pH 3–8 (final volume 200 μl). The reactions were performed in 96-well black flat bottom plates (Nunc) using Infinite M200 fluorometer (TECAN) with excitation and emission wavelengths set to 360 and 465 nm, respectively. The release of AMC was measured at RT in a 30 min kinetic cycle at 2 min intervals.

An exopeptidase (peptidyl dipeptidase) activity assay using Bz-Gly-His-Leu as a substrate was performed with pepTrCB1 forms employing a modified protocol (Sajid et al., 2003). The pepTrCB1 forms (1 μg, 160 nM) were incubated for 15 min with the substrate in 50/100 mM CPB pH 4–5.5 containing 2 mM DTT (final volume 100 μl). Spontaneous reaction of the emerging free amino groups of His-Leu with fluorescamine (0.05 mg/ml) was monitored in a fluorometer set to 390/475 nm excitation/emission wavelengths during a 30 min kinetic cycle at 2 min intervals.

Inhibition Assays

The effect of two inhibitors, E-64 and CA-074 (Sigma Aldrich) on endopeptidase activity of the pepTrCB1 forms (0.2 μg, 30 nM) was tested with Z-Phe-Arg-AMC. The concentrations of inhibitors ranged from 0.01 to 10 μM in the case of pepTrCB1.1 and pepTrCB1.4, and from 0.01 to 100 μM with pepTrCB1.6G/C. The samples were pre-incubated with inhibitors for 15 min prior to the measurement. Inhibition assays were carried out at the pH optima of the particular enzymes (pH 5.5 for TrCB1.1, pH 5 for TrCB1.4, and pH 6.5 for pepTrCB1.6G/C). The remaining relative enzyme activity was expressed in relation to control reactions performed in the absence of the inhibitors.

The inhibition assays of exopeptidase activity of the enzymes were performed with E-64 and CA-074 (10 μM) added prior to the substrate.

The effect of chicken egg white cystatin, human cystatin B, and human cystatin C (C8917, C4249, and H5041, Sigma Aldrich) on TrCB1 forms by was tested using dextran sulfate-activated and pepsin-activated enzymes ds/pepTrCB1.1, ds/pepTrCB1.4, ds/pepTrCB1.6G/C. Processed TrCB1 forms (0.2 μg, 30 nM) were pre-incubated with various concentrations of cystatins (0–1 μM) for 15 min prior to the measurement with 25 μM Z-Phe-Arg-AMC in the presence or absence of 2 mM DTT. All samples contained 0.01% non-ionic detergent SP BRIJ L23 (Croda).

Formation of Peptidase-Cystatin Complex

The dsTrCB1.1, dsTrCB1.4, dsTrCB1.6G/C, and pepTrCB1.6wt (3 μg, 3.75 μM) were incubated with chicken egg white cystatin (1 μg, 3 μM) in 50/100 mM CPB pH 5.5 containing 0.01 % BRIJ L23 for 15 min at RT or overnight at 4°C. The pepTrCB1.1, pepTrCB1.4, and pepTrCB1.6 were incubated with human cystatin B and C for 2 h at RT (concentrations as above). Controls did not contain cystatins. Then the samples were mixed with non-reducing sample buffer and allowed to stand for 15 min at RT. The samples were separated by SDS-PAGE under mild conditions (80 V).

Subsite Specificity Exploration of TrCB1.1 and TrCB1.4 by a Positional Scanning Synthetic Combinatorial Library (PS-SCL)

Purified non-processed pro-TrCB1.1 and pro-TrCB1.4 were used for the analysis of S1 – S4 subsite preferences toward P1–P4 residues of substrates. Since both pro-enzymes were able to cleave small peptide substrates, we had chosen this possibility to eliminate a potential risk of residual pepsin activity in pepsin-processed TrCB1 samples, which could bias the results. PS-SCL was applied as described previously (Choe et al., 2006). The peptidolytic activity of purified pro-enzymes (0.5 μg, 70 nM) was measured at RT in 50/100 mM CPB pH 5 containing 100 mM NaCl and 2 mM DTT.

Hydrolysis of Protein Substrates by TrCB1 Forms

The substrates (all Sigma-Aldrich) included turkey hemoglobin (Hb, H0142), bovine myelin basic protein (MBP, M1891), myosin (M7266), human albumin (A9511), collagen I (C7774), collagen IV (C8374), IgG (I4506), and fibrinogen (F3879) at final concentrations of 0.5 mg/ml, which were incubated for 6 h with pepTrCB1 forms (1 μg) in 50 μl of 50/100mM CPB pH 5.5 containing 2 mM DTT and 10 μM pepstatin A. Control reactions did not contain the enzymes. Aliquots of 20 μl of the reaction mixtures were separated by SDS-PAGE. Optimal pH for hydrolysis was ascertained using Hb and MBP as substrates in pH range 4.5–6.5.

Identification of Hb and MBP Fragments by Mass Spectrometry

For MALDI MS/MS analysis, turkey Hb (0.5 mg/ml) or recombinant human MBP (0.2 mg/ml, ab171675, Abcam) were incubated with 1 μg of pepTrCB1.1, pepTrCB1.4, and pepTrCB1.6G/C at 37°C in 25/50 mM CPB pH 5.5 containing 2 mM DTT and 10 μM pepstatin A, total reaction volume was 60 μl. Aliquots of 10 μl were taken after 5, 30, and 120 min and immediately treated with 0.5% trifluoroacetic acid. Control reactions did not contain the enzymes. Buffer employed for TrCB1 processing reactions that was incubated with pepsin-agarose beads was used as another control to assess possible residual activity of pepsin in the pepTrCB1-substrate samples.

MALDI-MS and MALDI-MS/MS were performed using Ultraflextreme instrument (Bruker Daltonics, Germany) operated in the linear or reflectron mode with detection of positive ions. Ferulic acid (12.5 mg/ml in water: acetonitrile:

formic acid, 50:33:17 v/v mixture) or alpha-cyano-4-hydroxycinnamic acid (saturated solution in water: acetonitrile: trifluoroacetic acid, 47.5:50.0:2.5 v/v mixture), respectively, were used as the MALDI matrices. Detected peptide ions were subjected to MS/MS analysis (in a LIFT arrangement) and the resulting fragments were then searched against the primary amino acid sequences of TrCB1 forms [GenBank: AY648119 – 24], porcine pepsinogen [AAA31095.1], turkey α-A-globin [AAB35442.1], α-D-globin [AAB35443.1], β-globin [P02113.1], and human MBP (sequence available on http://www.abcam.com/recombinant-human-myelin-basic-protein-ab171675.html) using MASCOT 2.4 search engine (MatrixScience). No enzyme specificity and oxidation (M) and deamidation (N, Q) as variable modifications were set for all searches.

Differential Expression of TrCB1 Forms

Full-length sequences of TrCB1 genes [GenBank: AY648119 – AY648124] (Dvořák et al., 2005) combined with *T. regenti* juvenile transcriptomes (cercariae and 7 days old schistosomula) (Leontovyč et al., 2016) have been used as the reference for the differential expression analysis (the data can be accessed via the NCBI website https://www.ncbi.nlm.nih.gov/ under BioProject database ID: PRJNA292737). All transcripts homologous to cathepsin B ($n = 14$) (BLASTp; E-value cut off: $< 10^{-5}$) have been removed from *T. regenti* transcriptome and replaced by all six TrCB1 sequences from NCBI database. Trimmed and corrected Illumina reads (HiSeq 2500 platform) from *T. regenti* cercariae and schistosomula were mapped to the reference using RSEM (Li and Dewey, 2011), with zero mismatch rate (i.e., only 100% identical reads were mapped to the reference). Predicted expected counts were rounded to the highest whole number and used for differential gene transcription analysis in edgeR v.3.6.7 (Robinson et al., 2010) and R v.3.2.3 (R Development Core Team, 2015) software packages. Cathepsin B sequences with more than $\log_2$ fold change and false discovery rate of $\leq$0.01 were stated as differentially expressed between cercariae and schistosomula.

In vitro Effect of Pro-TrCB1.6wt on Murine Astrocytes, Microglia and RAW 264.7 Macrophages

Primary murine astrocyte and microglia cultures were obtained from mixed glial cultures as described elsewhere (Macháček et al., 2016). These cells were used because they naturally occur and are activated in the vicinity of *T. regenti* schistosomula migrating in murine spinal cord (Lichtenbergová et al., 2011). Murine RAW 264.7 macrophages were received from Dr. Tereza Leštinová (Charles University, Prague). The cells were seeded in 96-well plates (Nunclon Delta Surface, Thermo Fisher Scientific) and exposed to 1 μg/ml of purified and deglycosylated pro-TrCB1.6wt with or without 0.5 μg/ml of lipopolysaccharide from *E. coli* 0127:B8 (LPS, Sigma Aldrich) for 48 h at 37°C. The medium itself or medium supplemented with 0.5 μg/ml of LPS was used as negative or positive control, respectively. Griess assay was used to analyze NO production by stimulated cells (Macháček et al., 2016). Cell viability after stimulation was assessed by fluorescein diacetate (Sigma Aldrich) according to manufacturer's instructions. The results were evaluated by ordinary one-way analysis of variance (ANOVA) followed by Tukey's multiple comparisons test, performed in GraphPad Prism, version 6.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI BioProject ID: PRJNA292737.

ETHICS STATEMENT

The animal study was reviewed and approved by Professional Ethics Committee of the Faculty of Science, Charles University.

AUTHOR CONTRIBUTIONS

HD performed expression of the enzymes, their processing, activity and inhibition assays, participated in protein purification, prepared samples for mass spectrometry, participated in interpretation of experimental data and preparation of figures, and wrote a larger part of the manuscript. RL obtained transcriptomic data, processed and interpreted the data on differential enzyme expression. TM performed immunological experiments and interpreted the data. AO'D and HD obtained and interpreted PS-SCL data. OŠ and ZZ obtained and interpreted mass spectrometry data. CSC provided know-how, facility and material for PS-SCL experiments, and performed critical reading of the whole manuscript. CRC and HD performed site-directed mutagenesis in TrCB1.6. AO'D and CRC provided critical comments on technical performance and interpretation of the results. PH revised the manuscript critically for important intellectual content. LM prepared conception, coordinated experiments, participated in protein purification, and cystatin-binding assays, contributed to interpretation of experimental data and finalized the manuscript. All authors contributed to writing or finalization of the manuscript.

FUNDING

The work on the project was funded by Czech Science Foundation (Grant Nos. 13-29577S and 18-11140S) and by the project Center for Research of Pathogenicity and Virulence of Parasites (No. CZ.02.1.01/0.0/0.0/16_019/0000759) funded by European Regional Development Fund and Ministry of Education, Youth and Sports of the Czech Republic. Charles University institutional support (PROGRES Q43, UNCE/SCI/012 - 204072/2018, and SVV 244-260432/2017) applied to RL, TM, PH, and LM. CIISB research infrastructure projects LM2015043 and LM2018127 funded by Ministry of Education, Youth and Sports of the Czech Republic are gratefully acknowledged for financially supporting the MALDI-MS measurements at the CEITEC Proteomics Core Facility. OS and ZZ thank for financial support of the project CEITEC 2020 (LQ1601) funded by Ministry of Education, Youth and Sports of the Czech Republic. Computational resources were

provided via the Melbourne Bioinformatics Platform and partly by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme Projects of Large Research, Development, and Innovations Infrastructures. CSC and AO'D were funded by P50 AI150476 and R01 GM104659 from the National Institute of Health.

ACKNOWLEDGMENTS

We thank Dr. Daniel Sojka (Institute of Parasitology, Biology Center of the Czech Academy of Sciences, Ceské Budějovice, Czech Republic) for providing recombinant tick legumain. We are also grateful to Dr. Tereza Leštinová (Department of Parasitology, Charles University, Prague, Czech Republic), who kindly provided murine RAW 264.7 macrophages, and Dr. Zdeněk Voburka (Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Prague, Czech Republic), who performed N-terminal protein sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2020.00066/full#supplementary-material

REFERENCES

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., et al. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 40, W597–W603. doi: 10.1093/nar/gks400

Barrett, A. J., and Kirschke, H. (1981). Cathepsin-B, cathepsin-H, and cathepsin-L. *Meth. Enzymol.* 80, 535–561. doi: 10.1016/S0076-6879(81)80043-2

Bergström, F. C., Reynolds, S., Johnstone, M., Pike, R. N., Buckle, A. M., Kemp, D. J., et al. (2009). Scabies mite inactivated serine protease paralogs inhibit the human complement system. *J. Immunol.* 182, 7809–7817. doi: 10.4049/jimmunol.0804205

Blažová, K., and Horák, P. (2005). *Trichobilharzia regenti*: the developmental differences in natural and abnormal hosts. *Parasitol. Int.* 54, 167–172. doi: 10.1016/j.parint.2005.03.003

Caffrey, C. R., Goupil, L., Rebello, K. M., Dalton, J. P., and Smith, D. (2018). Cysteine proteases as digestive enzymes in parasitic helminths. *PLoS Neglect. Trop. Dis.* 12:e0005840. doi: 10.1371/journal.pntd.0005840

Caffrey, C. R., Mathieu, M. A., Gaffney, A. M., Salter, J. P., Sajid, M., Lucas, K. D., et al. (2000). Identification of a cDNA encoding an active asparaginyl endopeptidase of *Schistosoma mansoni* and its expression in *Pichia pastoris*. *FEBS Lett.* 466, 244–248. doi: 10.1016/S0014-5793(99)01798-6

Caffrey, C. R., McKerrow, J. H., Salter, J. P., and Sajid, M. (2004). Blood 'n' guts: an update on schistosome digestive peptidases. *Trends Parasitol.* 20, 241–248. doi: 10.1016/j.pt.2004.03.004

Chanová, M., and Horák, P. (2007). Terminal phase of bird schistosomiasis caused by *Trichobilharzia regenti* (Schistosomatidae) in ducks (*Anas platyrhynchos* f. *domestica*). *Folia Parasit.* 54, 105–107. doi: 10.14411/fp.2007.014

Choe, Y., Leonetti, F., Greenbaum, D. C., Lecaille, F., Bogyo, M., Brömme, D., et al. (2006). Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. *J. Biol. Chem.* 281, 12824–12832. doi: 10.1074/jbc.M513331200

Delcroix, M., Sajid, M., Caffrey, C. R., Lim, K. C., Dvořák, J., Hsieh, I., et al. (2006). A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. *J. Biol. Chem.* 281, 39316–39329. doi: 10.1074/jbc.M607128200

Dolečková, K., Albrecht, T., Mikeš, L., and Horák, P. (2010). Cathepsins B1 and B2 in the neuropathogenic schistosome *Trichobilharzia regenti*: distinct gene expression profiles and presumptive roles throughout the life cycle. *Parasitol. Res.* 107, 751–755. doi: 10.1007/s00436-010-1943-6

Donnelly, S., O'Neil, S. M., Stack, C. M., Robinson, M. W., Turnbull, L., Whitchurch, C., et al. (2009). Helminth cysteine proteases inhibit TRIF-dependent activation of macrophages via degradation of TLR3. *J. Biol. Chem.* 285, 3383–3392. doi: 10.1074/jbc.M109.060368

Dvořák, J., Delcroix, M., Rossi, A., Vopálenský, V., Pospíšek, M., Šedinová, M., et al. (2005). Multiple cathepsin B isoforms in schistosomula of *Trichobilharzia regenti*: identification, characterisation and putative role in migration and nutrition. *Int. J. Parasitol.* 35, 895–910. doi: 10.1016/j.ijpara.2005.02.018

Greenbaum, D. C., Baruch, A., Hayrapetian, L., Darula, Z., Burlingame, A., Medzihradszky, K., et al. (2002). Chemical approaches for functionally probing the proteome. *Mol. Cell. Proteomics* 1, 60–68. doi: 10.1074/mcp.T100003-MCP200

Hamuro, Y., Coales, S. J., Molnar, K. S., Tuske, S. J., and Morrow, J. A. (2008). Specificity of immobilized porcine pepsin in H/D exchange compatible conditions. *Rapid Commun. Mass Sp.* 22, 1041–1046. doi: 10.1002/rcm.3467

He, W., Ohashi, K., Sugimoto, C., and Onuma, M. (2005). *Theileria orientalis*: Cloning a cDNA encoding a protein similar to thiol protease with haemoglobin-binding activity. *Exp. Parasitol.* 111, 143–153. doi: 10.1016/j.exppara.2005.06.003

Holt, D. C., Fischer, K., Allen, G. E., Wilson, D., Wilson, P., Slade, R., et al. (2003). Mechanisms for a novel immune evasion strategy in the scabies mite *Sarcoptes Scabiei*: a multigene family of inactivated serine proteases. *J. Invest. Dermatol.* 121, 1419–1424. doi: 10.1046/j.1523-1747.2003.12621.x

Holt, D. C., Fischer, K., Pizzutto, S. J., Currie, B. J., Walton, S. F., and Kemp, D. J. (2004). A multigene family of inactivated cysteine proteases in *Sarcoptes scabiei*. *J. Invest. Dermatol.* 123, 240–241. doi: 10.1111/j.0022-202X.2004.22716.x

Horák, P., Dvořák, J., Kolářová, L., and Trefil, L. (1999). *Trichobilharzia regenti*, a pathogen of the avian and mammalian central nervous systems. *Parasitology* 119, 577–581. doi: 10.1017/S0031182099005132

Horák, P., Kolářová, L., and Dvořák, J. (1998). *Trichobilharzia regenti* n. sp. (Schistosomatidae, Bilharziellinae), a new nasal schistosome from Europe. *Parasite* 5, 349–357. doi: 10.1051/parasite/1998054349

Horák, P., Mikeš, L., Lichtenbergová, L., Skála, V., Soldánová, M., and Brant, S. V. (2015): Avian schistosomes and outbreaks of cercarial dermatitis. *Clin. Microbiol. Rev.* 28, 165–190. doi: 10.1128/CMR.00043-14

Horák, P., Mikeš, L., Rudolfová, J., and Kolářová, L. (2008). Penetration of *Trichobilharzia* cercariae into mammals: dangerous or negligible event? *Parasite* 15, 299–303. doi: 10.1051/parasite/2008153299

Horn, M., Jílková, A.,Vondrášek, J., Marešová, L., Caffrey, C. R., and Mareš, M. (2011). Mapping the pro-peptide of the *Schistosoma mansoni* cathepsin B1 drug target: modulation of inhibition by heparin and design of mimetic inhibitors. *ACS Chem. Biol.* 6, 609–617. doi: 10.1021/cb100411v

Illy, C., Quraishi, O., Wang, J., Purisima, E., Vernet, T., and Mort, J. S. (1997). Role of the occluding loop in cathepsin B activity. *J. Biol. Chem.* 272, 1197–1202. doi: 10.1074/jbc.272.2.1197

Jedličková, L., Dvořáková, H., Dvořák, J., Kašný, M., Ulrychová, L., Vorel, J., et al. (2018). Cysteine peptidases of *Eudiplozoon nipponicum*: a broad repertoire of structurally assorted cathepsins L in contrast to the scarcity of cathepsins B in an invasive species of haematophagous monogenean of common carp. *Parasite. Vector.* 11, 1–17. doi: 10.1186/s13071-018-2666-2

Jílková, A., Horn, M., Řezáčová, P., Marešová, L., Fajtová, P., Brynda, J., et al. (2014). Activation route of the *Schistosoma mansoni* cathepsin B1 drug target: structural map with a glycosaminoglycan switch. *Structure* 22, 1786–1798. doi: 10.1016/j.str.2014.09.015

Kašný, M., Haas, W., Jamieson, B. G. M., and Horák, P. (2016). "Cercaria of *Schistosoma*," in *Schistosoma*: *Biology, pathology and control*, ed B. G. M. Jamieson (Boca Raton, FL: CRC Press; Taylor and Francis Group), 149–183. doi: 10.1201/9781315368900-9

Krupa, J. C., Hasnain, S., Nägler, D. K., Ménard, R., and Mort, J. S. (2002). S2' substrate specificity and the role of His110 and His111 in

the exopeptidase activity of human cathepsin B. *Biochem. J.* 361, 613–619. doi: 10.1042/bj3610613

Law, R. H. P., Smooker, P. M., Irving, J. A., Piedrafita, D., Ponting, R., Kennedy, N. J., et al. (2003). Cloning and expression of the major secreted cathepsin B-like protein from juvenile *Fasciola hepatica* and analysis of immunogenicity following liver fluke infection. *Infect. Immun.* 71, 6921–6932. doi: 10.1128/IAI.71.12.6921-6932.2003

Leontovyč, R., Young, N. D., Korhonen, P. K., Hall, R. S., Bulantová, J., Jeřábková, V., et al. (2019). Molecular evidence for distinct modes of nutrient acquisition between visceral and neurotropic schistosomes of birds. *Sci. Rep.* 9:1347. doi: 10.1038/s41598-018-37669-2

Leontovyč, R., Young, N. D., Korhonen, P. K., Hall, R. S., Tan, P., Mikeš, L., et al. (2016). Comparative transcriptomic exploration reveals unique molecular adaptations of neuropathogenic *Trichobilharzia* to invade and parasitize its avian definitive host. *PLoS Neglect. Trop. Dis.* 10, 1–24. doi: 10.1371/journal.pntd.0004406

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323

Lichtenbergová, L., Lassmann, H., Jones, M. K., Kolářová, L., and Horák, P. (2011). *Trichobilharzia regenti*: host immune response in the pathogenesis of neuroinfection in mice. *Exp. Parasitol.* 128, 328–335. doi: 10.1016/j.exppara.2011.04.006

Lipps, G., Fü, R., and Beck, E. (1996). Cathepsin B of *Schistosoma mansoni* - Purification and activation of the recombinant proenzyme secreted by *Saccharomyces cerevisiae*. *J. Biol. Chem.* 271, 1717–1725. doi: 10.1074/jbc.271.3.1717

Macháček, T., Panská, L., Dvořáková, H., and Horák, P. (2016). Nitric oxide and cytokine production by glial cells exposed *in vitro* to neuropathogenic schistosome *Trichobilharzia regenti*. *Parasite. Vector.* 9:579. doi: 10.1186/s13071-016-1869-7

Macháček, T., Turjanicová, L., Bulantová, J., Hrdý, J., Horák, P., and Mikeš, L. (2018). Cercarial dermatitis: a systematic follow-up study of human cases with implications for diagnostics. *Parasitol. Res.* 117, 3881–3895. doi: 10.1007/s00436-018-6095-0

McCoubrie, J. E., Miller, S. K., Sargeant, T., Good, R. T., Hodder, A. N., Speed, T. P., et al. (2007). Evidence for a common role for the serine-type *Plasmodium falciparum* serine repeat antigen proteases: implications for vaccine and drug design. *Infect. Immun.* 75, 5565–5574. doi: 10.1128/IAI.00405-07

Mendoza-Palomares, C., Biteau, N., Giroud, C., Coustou, V., Coetzer, T., Authié, E., et al. (2008). Molecular and biochemical characterization of a cathepsin B-like protease family unique to *Trypanosoma congolense*. *Eukaryot. Cell* 7, 684–697. doi: 10.1128/EC.00405-07

Merckelbach, A., Hasse, S., Dell, R., Eschlbeck, A., and Ruppel, A. (1994). cDNA sequences of *Schistosoma japonicum* coding for 2 cathepsin B-like proteins and Sj32. *Trop. Med. Parasitol.* 45, 193–198.

Mort, J. S., and Buttle, D. J. (1997). Cathepsin B. *Int. J. Biochem. Cell Biol.* 29, 715–720. doi: 10.1016/S1357-2725(96)00152-5

Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., et al. (1991). The refined 2.15 A X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO J.* 10, 2321–2330. doi: 10.1002/j.1460-2075.1991.tb07771.x

Pavlova, A., Krupa, J. C., Mort, J. S., Abrahamson, M., and Björk, I. (2000). Cystatin inhibition of cathepsin B requires dislocation of the proteinase occluding loop. Demonstration by release of loop anchoring through mutation of His110. *FEBS Lett.* 487, 156–160. doi: 10.1016/S0014-5793(00)02337-1

Pillay, D., Boulangé, A. F., and Coetzer, T. H. T. (2010). Expression, purification and characterisation of two variant cysteine peptidases from *Trypanosoma congolense* with active site substitutions. *Protein Expres. Purif.* 74, 264–271. doi: 10.1016/j.pep.2010.06.021

R Development Core Team (2015). *A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/

Reynolds, S. L., Pike, R. N., Mika, A., Blom, A. M., Hofmann, A., Wijeyewickrema, L. C., et al. (2014). Scabies mite inactive serine proteases are potent inhibitors of the human complement lectin pathway. *PLoS Neglect. Trop. Dis.* 8:e2872. doi: 10.1371/journal.pntd.0002872

Řimnáčová, J., Mikeš, L., Turjanicová, L., Bulantová, J, and Horák, P. (2017): Changes in surface glycosylation and glycocalyx shedding in *Trichobilharzia regenti* (Schistosomatidae) during the transformation of cercaria to schistosomulum. *PLoS ONE* 12:e0173217. doi: 10.1371/journal.pone.0173217

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

Sajid, M., McKerrow, J. H., Hansell, E., Mathieu, M. A., Lucas, K. D., Hsieh, I., et al. (2003). Functional expression and characterization of *Schistosoma mansoni* cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. *Mol. Biochem. Parasitol.* 131, 65–75. doi: 10.1016/S0166-6851(03)00194-4

Salter, J. P., Choe, Y., Albrecht, H., Franklin, C., Lim, K. C., Craik, C. S., et al. (2002). Cercarial elastase is encoded by a functionally conserved gene family across multiple species of schistosomes. *J. Biol. Chem.* 277, 24618–24624. doi: 10.1074/jbc.M202364200

Schechter, I., and Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27, 157–62. doi: 10.1016/S0006-291X(67)80055-X

Sojka, D., Hajdušek, O., Dvořák, J., Sajid, M., Franta, Z., Schneider, E. L., et al. (2007). IrAE - An asparaginyl endopeptidase (legumain) in the gut of the hard tick *Ixodes ricinus*. *Int. J. Parasitol.* 37, 713–724. doi: 10.1016/j.ijpara.2006.12.020

Soloviova, K., Fox, E. C., Dalton, J. P., Caffrey, C. R., and Davies, S. J. (2019). A secreted schistosome cathepsin B1 cysteine protease and acute schistosome infection induce a transient T helper 17 response. *PLoS Neglect. Trop. Dis.* 13:e0007070. doi: 10.1371/journal.pntd.0007070

Williamson, A. L., Brindley, P. J., Knox, D. P., Hotez, P. J., and Loukas, A. (2003). Digestive proteases of blood-feeding nematodes. *Trends Parasitol.* 19, 417–423. doi: 10.1016/S1471-4922(03)00189-2

Williamson, A. L., Lecchi, P., Turk, B. E., Choe, Y., Hotez, P. J., McKerrow, J. H., et al. (2004). A multi-enzyme cascade of hemoglobin proteolysis in the intestine of blood-feeding hookworms. *J. Biol. Chem.* 279, 35950–35957. doi: 10.1074/jbc.M405842200

Yamamoto, A., Hara, T., Tomoo, K., Ishida, T., Fujii, T., Hata, Y., et al. (1997). Binding mode of CA074, a specific irreversible inhibitor, to bovine cathepsin B as determined by X-Ray crystal analysis of the complex. *J. Biochem.* 121, 974–977. doi: 10.1093/oxfordjournals.jbchem.a021682

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.