CHARLES UNIVERSITY

Faculty of Science

Department of Parasitology

Study programme: Biology

Field of study: Parasitology

Bc. Lukáš Konečný

Analysis of secretome from *Trichobilharzia regenti* cercariae and characterisation of selected peptidases

Analýza sekretomu cerkárie *Trichobilharzia regenti*
a charakterizace vybraných peptidáz

MASTER'S THESIS

Supervisor: RNDr. Martin Kašný, Ph.D.

Prague, 2019

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně, a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne 29.4.2019 …….

Acknowledgements

First of all, I would like to thank my supervisor Martin Kašný for his valuable advices, encouragement and trust during my whole master's studies and also for the opportunity to undergo an internship in Northern Ireland for which he had the greatest credit.

I am grateful to the members of our helminthological team that they were patient, understanding and have shown a great tolerance to have me around. My biggest thanks work-wise belong to Hanka Dvořáková, Lucka Jedličková and Libor Mikeš for introducing me to the enigmatic world of nucleic acids and proteins and to Roman Leontovyč for helping me understand files full of weird letters. Also, I want to express my gratitude to Petr Horák for his willingness and making this whole thing possible.

I owe a big thanks to our colleagues from Masaryk's University in Brno, which carried out the mass-spectrometry analysis and especially to Jirka Vorel for helping me with data evaluation.

My warmest thanks belong also to all members of John Dalton's lab at Queen's University Belfast, particularly to John P. Dalton for his kindness and genuine interest, Krystyna Cwiklinski for mentoring, Carolina De Marco Verissimo for a great help with enzymatic assays and Heather Jewhurst for saving me over 400 kilometres of walking.

This would not be complete without acknowledging my friend and colleague Týna Peterková, with whom we helped each other for the whole 3 years.

Finally, I would like to express the deepest gratitude to my family, Škopky and Zikmund for the endless support ♥.

This study was supported by the ERASMUS+ and Czech Science Foundation (Grant No. 18-11140S).

Abstract (English):

Trichobilharzia regenti is a neurotropic parasite of birds from the family Schistosomatidae. Cercariae, the invasive stages of these trematodes actively penetrate the host skin employing excretory-secretory products (ESPs), which contain proteolytic enzymes able to disrupt host tissues and thus reach the successful transmission. The most abundant secreted enzyme responsible for cercarial penetration of the human schistosome *S. mansoni* is a cercarial elastase. This serine peptidase is well known for the degradation of skin proteins such as elastin, keratin, collagen or laminin. However, the active expression of the orthologue of this enzyme has never been found in the genus *Trichobilharzia*. For this reason, it was firmly believed, that cercaria of *T. regenti* uses mainly cysteine peptidases for the invasion of the host, particularly cathepsins, which were repeatedly identified in this life stage. To strengthen this hypothesis, we incubated *T. regenti* cercariae in the apparatus with the excised duck skin stimulating the release of their glands´ content. The collected ESPs were further analysed by shotgun mass-spectrometry and for the first time, the protein form of cercarial elastase was identified. Unfortunately, we failed to produce its active recombinant protein in yeast and bacterial expression system. However, our secretomic study together with additional qPCR analysis suggests, that cercarial elastase might be an important player during the penetration.

Key words: schistosome, *Trichobilharzia*, peptidase, protease, excretory-secretory products

Abstrakt (česky):

Motolice *Trichobilharzia regenti* je neurotropní parazit ptáků z rodiny Schistosomatidae. Cerkárie, invazivní stadium těchto motolic, využívají k aktivní penetraci hostitele exkrečně-sekreční produkty, které obsahují proteolytické enzymy schopné narušit tkáně hostitele zaručit tak úspěšný přenos. Nejvíce zastoupeným enzymem zodpovědným za penetraci cerkárií lidské schistosomy *S. mansoni* je cerkáriová elastáza. Tato serinová peptidáza je dobře známá pro degradaci kožních proteinů jako je elastin, keratin, kolagen nebo laminin, avšak, aktivní exprese ortologu tohoto enzymu nebyla u rodu *Trichobilharzia* nikdy prokázána. Z tohoto důvodu panovalo pevné přesvědčení, že cerkárie *T. regenti* používá k invazi hostitele zejména cysteinové peptidázy, zvláště katepsiny, které byly opakovaně identifikovány v tomto životním stádiu. Abychom posílili tuto hypotézu. inkubovali jsme cerkárie *T. regenti* v sestavě využívající kachní kůži jakožto stimulant pro vyprázdnění penetračních žláz. Získané ESPs byly analyzovány hmotnostní spektrometrií a poprvé se nám tak podařilo identifikovat proteinovou formu cerkáriové elastázy *T. regenti*. Produkce aktivního rekombinantního proteinu se bohužel nezdařila v bakteriálním ani kvasinkovém expresním systému, nicméně, naše sekretomická studie společně s dodatečnou qPCR analýzou naznačuje, že cerkáriová elastáza může hrát důležitou roli v průběhu penetrace hostitele.

Klíčová slova: schistosoma, *Trichobilharzia*, peptidáza, proteáza, exkrečně-sekreční produkty

Table of contents

1. List of abbreviations

AA	amino acid
cDNA	complementary deoxyribonucleic acid
C_T	cycle threshold
DEPC	diethyl pyrocarbonate
DNA	deoxyribonucleic acid
DOC	deoxycholate
EDTA	ethylenediaminetetraacetic acid
ESPs	excretory-secretory products
EST	expressed sequence tag
FhCL3	*Fasciola hepatica* cathepsin L3
FhCys	*Fasciola hepatica* cystatin
FPLC	fast protein liquid chromatography
GSP	gene-specific primer
HSP	heat shock protein
ICR	Institute of Cancer Research
IPTG	isopropyl β-D-1-thiogalactopyranoside
LC-MS	liquid chromatography-mass spectrometry
LED	light-emitting diode
MALDI-TOF;	matrix-assisted laser desorption/ionization-time-of-flight
mRNA	messenger ribonucleic acid
NDUFV2	NADH dehydrogenase flavoprotein 2
OD	optical density
PCR	polymerase chain reaction
PSMD	26S proteasome non-ATPase regulatory subunit 4
qPCR	quantitative polymerase chain reaction
RACE	Rapid Amplification of cDNA Ends
RNA	ribonucleic acid
RT	room temperature
SdCE	*Schistosomatium douthitti* cercarial elastase
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
ShCE	*Schistosoma haematobium* cercarial elastase
SjCE	*Schistosoma japonicum* cercarial elastase
SmCE	*Schistosoma mansoni* cercarial elastase
TPC2L	Trafficking protein particle complex subunit 2-like protein
TrCB2	*Trichobilharzia regenti* cathepsin B2
TrCE	*Trichobilharzia regenti* cercarial elastase

2. Introduction, aims and objectives

The neuropathogenic schistosome *T. regenti* is a peculiar organism within the whole family of Schistosomatidae and therefore, its unique biology characteristic with neurotropic behaviour is intensively studied for more than 20 years (Horák et al., 1998).

Besides the mechanisms by which the fluke migrates through the neural tissues of both specific and random hosts and associated pathogenicity, the great focus has always been placed on the mechanisms connected with the invasion of the definitive host. Following the pattern of the research on human schistosomes, the biggest emphasis has been placed on cercarial penetration glands' contents, especially peptidases – enzymes responsible for the degradation of skin proteins.

Series of molecular, biochemical and immunological experiments repeatedly suggested, that avian schistosomes from genus *Trichobilharzia* use mainly cysteine peptidases, namely cathepsins, to invade the host (Mikeš et al., 2005; Dolečková et al., 2007, 2009). On the other hand, cercarial elastase, penetration enzyme of anthropophilic schistosomes' cercariae was despite the great effort (including Laboratory of Helminthology, Prague) never identified in *T. regenti*. Therefore, it was concluded, that the orthologue of this enzyme is not present or actively transcribed. This hypothesis was even strengthened by the fact, that no transcripts coding cercarial elastase were found in recently published *T. regenti* cercarial and schistosomulum's transcriptome (Leontovyč et al., 2016).

With the rapid development and availability of complex mass-spectrometry analyses and release of *T. regenti* genome, interest in identifying and characterizing major players in cercarial penetration process has increased again. Clarifying the mechanisms responsible for the definitive host invasion is essential for a complete understanding of the unique biology of this neuropathogenic schistosome.

Last but not least, published literature to date is inconsistent about gene expression patterns within the sporocyst and cercariae of schistosomes. While some authors suggest, that the levels of gene expression in those stages are comparable (Parker-Manuel et al., 2011), others reject active gene expression in cercaria stage (Roquis et al., 2015).

Goals of this work are:

- To collect excretory-secretory products of *T. regenti* cercaria using different methods, compare these methods and identify most abundant peptidases

- To further characterize peptidases possibly responsible for parasite's invasion

- To compare the gene expression levels of the main peptidases coding genes in sporocyst and cercariae to determine expression patterns for these stages and to support the quantitative results of the mass-spectrometry analysis

3. Literary review

Platyhelminthes is a phylum of bilaterian soft-bodied invertebrates containing very important parasites of animals and people. This phylum is divided into Turbellaria, which are mostly non-parasitic animals, and three obligatory parasitic groups: Cestoda, Trematoda and Monogenea. Flatworms, which annually cause the majority of human deaths are trematodes from family Schistosomatidae which are unusual trematodes with their gonochorism, adult life in pairs and surviving in the vascular system of its host (WHO, 2019).

3.1. *Trichobilharzia regenti*

The model organism *Trichobilharzia regenti* is a fluke with dixenous life cycle belonging to class Digenea and family Schistosomatidae. Genus *Trichobilharzia* is the biggest within the family with more than 40 parasites of birds. This genus can be divided into two groups based on final localization of the adults within the host. While visceral species end their life cycle in internal organs of their hosts, nasal flukes from this genus finish their development in the nasal cavity. *T. regenti* belongs to the nasal species, but it is unique with its neuropathogenic migration inside the host's body, in comparison with vascular migration of other species from the same genus (Horák et al., 1999). This parasite is transmitted through water, where freshwater snails of the genus *Radix* (mainly *R. labiata* and *R. lagotis*) act as intermediate hosts and birds from family Anatidae as definitive hosts.

T. regenti was discovered and described by a team of prof. Petr Horák in 1998 in South Bohemia and after that, the distribution of this parasite was reported in Germany, Russia, Poland, France, Denmark, Iran and Island (Rudolfová et al, 2007; Jouet et al., 2008; Aldhoun et al, 2009; Prüter et al., 2010; Korsunenko et al., 2010; Christiansen et al., 2016, 2016; Fakhar et al., 2016) (**Figure 1.**).

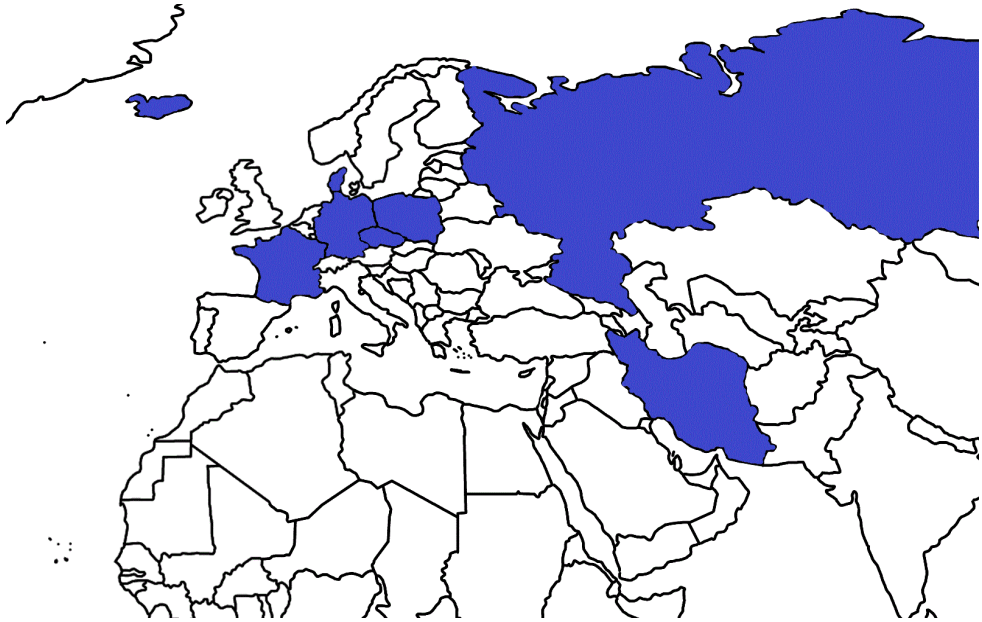

Figure 1.Distribution of *Trichobilharzia regenti*

3.2. Life cycle

After the copulation of adults, released fertilized eggs of *T. regenti* stay in the nasal cavity of the bird and next life stage, miracidium, hatches after the contact with fresh water within the host nasal tissues. This first larval stage is then released to the environment and actively seeks for its intermediate host, *Radix* snail. Miracidium then penetrates the snail, usually, the leg, using terebratorium for attachment and penetration glands. After successful penetration, miracidium transforms into the second larval stage, maternal sporocyst, which asexually produces daughter sporocysts. These stages then give rise to the fork-tailed cercariae which are released to the water periodically during the day based on natural light level correlating with the activity of the host (Haas, 2003). If the cercaria succeeds in finding a definitive host, it actively penetrates the skin, most often on the legs, which are in case of birds the most exposed to the parasite's larvae. After penetration, cercaria loses its tail and transforms into the schistosomulum – immature stage that migrates to the final localization within the host. Schistosomulum seeks for the peripheral nerves of the central nervous system, which is a unique type of migration typical for this species. Subsequently, these juvenile larvae approach the spine of the bird, where it can be found even the next day after infection. From the spine, the parasite continues to the brain, where it can be found 12 days after infection (Hrádková and Horák, 2002). Next day, fluke moves to the nasal area, where it matures, couples with a sex partner and first production of eggs occurs around 14. day after infection (Horák et al., 1999). The lifespan of *T. regenti* individuals is very short, it ranges between 23 and 25 days (Chanová and Horák, 2007) (**Figure 2.**).

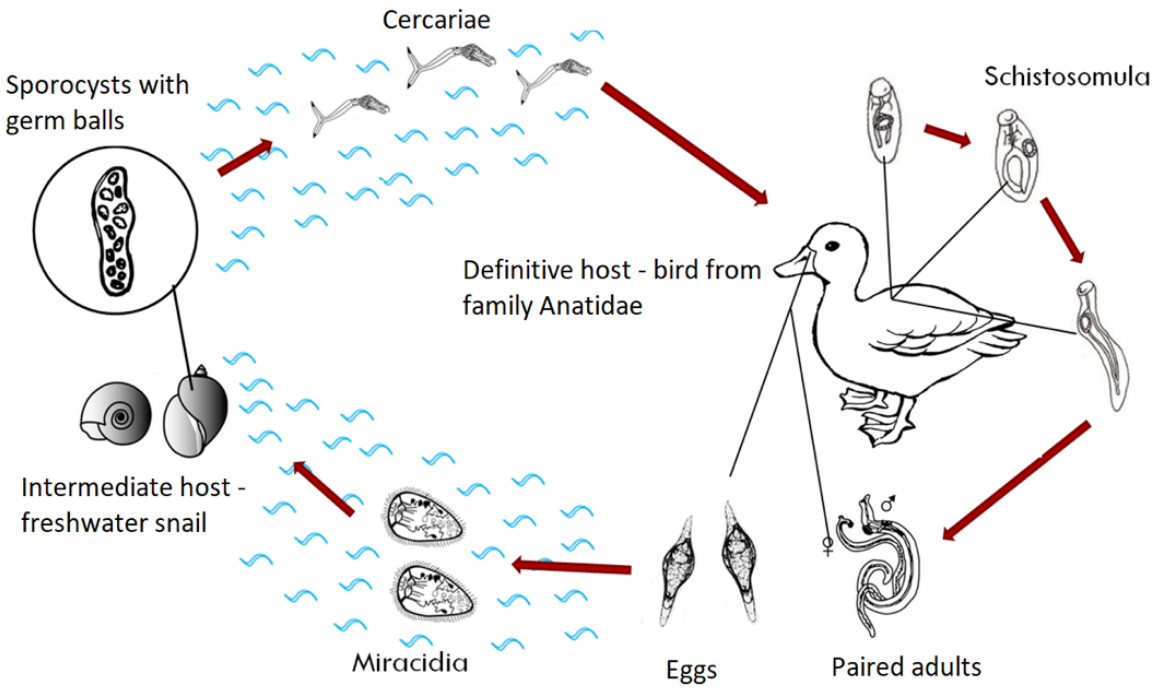

Figure 2. Life cycle of *T. regenti* (Horák et al. 1998; Walker 2011; www.cdc.gov; adjusted)

3.2.1. Characterisation of selected life stages

This chapter includes a basic characterization of selected life stages of *T. regenti* which are relevant for this thesis. All life stages are described in detail in Horák et al., 1998.

Similarly, as in human schistosomes, **sporocysts** of genus *Trichobilharzia* have two generations. The maternal sporocyst develops after the miracidium penetrates an intermediate snail host and undergoes dramatic physiological and morphological changes. This first generation sporocyst then produces large amounts of daughter sporocysts that often migrate to the nutritionally rich body sites such as gonads or hepatopancreas where they acquire nutrients by absorption through tegument. The schistosome cercariae then develop from germ balls within the daughter sporocyst. It has been also repeatedly suggested in human schistosomes, that proteins used by the cercaria for host entry are mostly transcribed and translated during the development in the snail. However, only very little is known about processes associated with the development of the germ balls that mature into cercariae (Parker-Manuel et al. 2011; Ingram et al. 2012).

Cercariae of *T. regen*ti is called furcocercariae. This type of cercariae has a distinctive bifurcated tail with a motion function in a water environment. The tail of cercariae is ca. 2/3 of the total length of 0,75 mm (Horák et al., 1998). Other significant structures are two black eye spots in the anterior part of the invasive larvae body. Although cercariae of this species do not take up any food and live from glycogen reserves, it still has the noticeable basis of short, bifurcated intestine. In certain ways, the most important visible structures of cercariae are penetration glands. These glands are represented by five pairs of glandular cells – two pairs of preacetabular (circumacetabular) glands surrounding ventral sucker and three pairs of postacetabular glands located posteriorly (see **Figure 3.**).

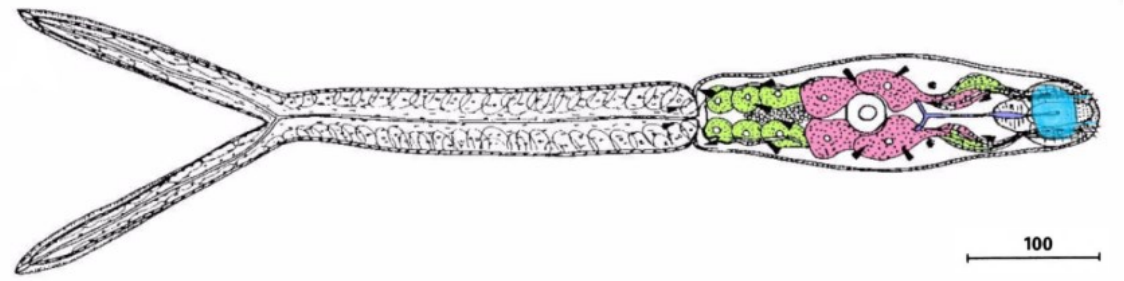

Figure 3. Cercaria of *T. regenti.* Colorized preacetabular (pink), postacetabular (green) and head (blue) glands (scale bar in μm) (Ligasová et al. 2011)

This setup corresponds with cercariae of other schistosome species (Ligasová et al. 2011). Gland ducts lead through head gland to the anterior part of the body, where they open. Excretory system of this larvae is represented by seven pairs of flame cells with collection canals leading to the end of the furca.

Cercariae of *T. regenti* show positive phototaxis, without any external stimulus they either float in the water column or rest on the bottom. They become active with a shadow or mechanical stimulus which indicates the presence of the host. The mechanical stimulus has a similar effect

on cercariae, which quickly after the stimulus swim to the water surface. After a successful finding of either specific or unspecific host, cercariae attach to skin surface and stimulation with fatty acids (linoleic acid, linolenic acid) from hosts skin initiates emptying of cercarial penetration glands. Proteolytic content of the glands then damages host tissues, larvae shed its tail and enters hosts body where it quickly undergoes major physiological changes resulting in the transformation to tailless migrating larva called schistosomulum (Horák et al., 1998).

3.2.2. Pathogenicity

Birds infected with *T. regenti* show balance and orientation disorders, in some cases epistaxis and massive petechiae in nasal mucosa were observed but these conditions are rather rare. Without apparent dependence on the infective dose, complete leg paralysis may also occur (Horák et al., 1999).

Since the main stimuli in host seeking process are fatty acids which are common for vertebrates' and thus not very specific, cercariae can invade even a non-specific host, which may be another vertebrate, including a human. Development and migration of immature stages of birds schistosomes in the incompatible host are similar as in compatible host, with the exception that parasites are not able to reproduce and die at different time points of this migration (Hrádková and Horák, 2002).

In experimentally infected laboratory mice, schistosomula of *T. regenti* can migrate to the spinal cord and the brain, where it damages the nervous tissue, and this infection leads to an inflammatory reaction. These infections, especially in immunosuppressed mice, are manifested by legs paralysis and can lead to death of the host. Natural infection of an immunosuppressed person has not yet been observed, however, this neurotrophic behaviour of *T. regenti* may present a potential hazard for these individuals. In the case of healthy people, repeated infections with *T. regenti* cercariae lead to neglected disease, cercarial dermatitis, which is currently on rise (Lichtenbergová and Horák, 2012). This disease is manifested by a severe maculopapular skin rash that is associated with intense itching. The itch can be negligible, but also very strong, almost unbearable. Red stains appear on skin within 20 minutes after infection and are replaced by pimples that occur between 10 and 15 hours after infection. Massive infections can also lead to fever, swollen limbs, nausea and diarrhoea. Skin lesions only occur on parts of the body that have been in contact with cercariae. Symptoms of infection usually disappear between 4 and 10 days after the infection (Kolářová et al., 2013).

In a mild climate, this disease is linked to the summer period, when the majority of cercariae are released from the intermediate hosts the most, while the number of people bathing in natural water reservoirs and streams is rising strongly (Kolářová et al., 2013).

3.3. Schistosomes and "-omics"

"-omics" is a suffix labelling a variety of biological approaches such as genomics, proteomics, transcriptomics or secretomics that produce a large informative datasets about biomolecules that translate into the structure, function, and dynamics of an organism. Besides the huge step forward in the understanding of schistosome biology and pathogenesis, these novel approaches contribute significantly to the exploitation and innovation of new drugs, vaccines as well as diagnostic tools for schistosomiasis (S. Wang and Hu, 2014).

3.3.1. Genomics

Genomics aims at the collective characterization of genes present in an organism as well as at the analysis of whole genomes through uses of advanced DNA sequencing and bioinformatics, which allow to assemble and analyse their functions and structure.

By 2012, the genomes of the three most important schistosomes infecting humans were published, however, this type of data for avian schistosomes was not available then (Sajid, et al., 2009; Zhou et al. 2009). With the introduction of 50 Helminth Genomes Project, an early draft of *T. regenti* genome was also published (Coghlan et al., 2019). Even though the data are not yet complete, they offer insights into early events in the evolution of the animals, including the development of a body pattern as well as they provide us very valuable tool for investigating the functions of the genes responsible for the essential attributes of the parasite translatable into practice (Howe et al., 2017).

For an overview of current progress in genomics of schistosomes, see **Table 1.**

Table 1. Overview of genomics in schistosomes.

Species	Key publications
Schistosoma bovis	Oey et al., 2019
Schistosoma curassoni	Coghlan et al., 2019
Schistosoma haematobium	Young et al., 2012
Schistosoma japonicum	Zhou et al., 2009
Schistosoma mansoni	Sajid et al., 2009; Protasio et al., 2012
Schistosoma margrebowiei	Coghlan et al., 2019
Schistosoma mattheei	Coghlan et al., 2019
Schistosoma rodhaini	Coghlan et al., 2019
Trichobilharzia regenti	Coghlan et al., 2019

3.3.2. Transcriptomics

Since not all genes that are present in the genome are also transcribed into mRNA at each stage of parasite or even at all, **transcriptomics** only deals with the transcribed genes. Transcriptome thus refers to all RNA molecules in one cell, tissue or whole organism - in this case, a parasitic life stage.

It follows from the definition above that this approach makes possible to monitor the gene expression levels at individual stages and in addition, under certain conditions such as particular time point at a specific stage of parasitic infection or for instance, treatment. Therefore, such data are an irreplaceable well of knowledge, enabling in combination with bioinformatics to sort out genes into individual groups based on *in silico* predicted functions or localizations and thus reveal a much about parasite's or its individual life stage's biology (W. Hu et al., 2004). In addition to the comparison of gene expressions in individual life stages, it is also possible to map expression in individual tissues using elegant methods combining microdissections and microarray techniques (Gobert et al., 2009).

To date, the largest transcriptomic study of *S. mansoni* compared the expression profiles of four key time points in the parasite's life cycle – cercariae, 3-hour schistosomula, 24-hour schistosomula and adult worm. The results nicely reflected major changes in biology of these stages from aerobic metabolism and glycogen processing of free-living cercariae, across major increase in expression of heat shock proteins connected with rapid temperature rise, to high expression of antioxidant enzymes in adults protecting them against potentially damaging reactive oxygen species (Protasio et al., 2012).

To this day, cDNA libraries with generated ESTs (expressed sequence tags) of almost all stages of human schistosomes species were already published (S. Wang and Hu, 2014). The increasing number of transcriptomic experiments resulted in production of a huge amount of data and their accessibility in public databases, but there are very few integrated projects and meta-analysis that would decipher this data and create a comprehensive view of their informational value. In case of *T. regenti*, the recent transcriptomic study of cercariae and schistosomulum outlined the key differences in gene expression between these stages. It also serves as a valuable source of information on putative protein equipment of these developmental stages which can be exploited by subsequent proteomic approaches (Leontovyč et al., 2016).

For an overview of current progress in genomics of schistosomes, see **Table 2.**

.

Table 2. Overview of transcriptomic studies in schistosomes. E- egg**, M** - miracidium**. S -** sporocyst**, C -** cercaria**, Sch -** schistosomulum**, A -** adult

Species	**Life stages**	**Publications**
S. haematobium	E, A	Young et al. 2012
S. japonicum	E, M, S, C, Sch, A,	Fan et al. 1998; Fung et al. 2002; S. Hu et al. 2009; Peng et al. 2003; Gobert et al. 2009
S.mansoni	E, M, S, C, Sch, A,	Verjovski-Almeida et al. 2003; Shabaan et al. 2003; Chai et al. 2006; Ojopi et al. 2007; Taft et al. 2009;
T. regenti	C, Sch	Leontovyč et al. 2016

3.3.3. Proteomics

Since there is no direct proportionality between the transcriptional and translational level of one expressed gene, **proteomic analyses** play a major role in characterization of functions, abundance, and possible localization of the active translated gene products. A huge increase in the number of proteomic studies was possible due to the recent great technological progress in the field of mass spectrometry and bioinformatics.

Typically, the proteomic analysis comprises of protein separation by 1D or 2D polyacrylamide gel electrophoresis, protein bands or spots of interest isolation, protein digestion and either Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) or liquid chromatography-tandem mass spectrometry (LC-MS/MS). Data obtained from mass-spectrometry are then compared to theoretical protein masses present in a chosen database and the protein identities are acquired. For protein identification, genomes, transcriptomes or proteomes of the examined or related species can be used. The type and quality of the chosen protein database are crucial for outcomes of the experiment and thus should be carefully considered (van Hellemond et al., 2007).

A new approachs called shotgun proteomics is now able to identify proteins in a complex mixture using bottom-up strategy. The process starts with the digestion of the protein sample, the resulting peptides are then separated by liquid chromatography and identified using tandem mass spectrometry using multiple steps of selection with subsequent fragmentation between the steps (Alves et al., 2007). Similarly as in transcriptomics, proteomic analyses allow to describe sample protein composition and to compare it between the samples which can be for example different tissues, life stages or even different organism species.

3.3.4. Secretomics

An important part of proteome consists of secreted proteins. These proteins play an important role not only in parasite‘s body but also as they are secreted in the host environment, they actively interact with it and thus may be directly connected with immunomodulation and pathogenicity of the parasite (Cass et al., 2007; Hewitson et al., 2009). The term **secretome** was originally defined as all the factors secreted by a cell along with the secretory pathway constituents Tjalsma et al., 2000), however the definition was modified, and this term is currently used as a generic designation for all secreted proteins of one biological species (Meinken et al., 2015). Nevertheless, also this definition needs to be broadened - the individual life stages of schistosomes differ greatly in their biology that even the composition of secreted molecules changes significantly throughout the life cycle. Therefore, for the purpose of this work, the term secretome is understood to be a sum of all molecules secreted by an individual parasitic life stage.

The prediction of secretome depends on the type of secretory pathway. Classical secretory pathway proteins contain so-called hydrophobic signal peptide at the N-terminus. This peptide can be predicted using computational software (e.g., SignalP). Prediction of secreted proteins that do not contain a signal peptide and are still secreted (non-classical secretory pathways) is also possible by computer software (such as SecretomeP) and is based on various biochemical protein properties further described in the literature (Bendtsen et al., 2004; Cuesta-Astroz et al., 2017; Almagro Armenteros et al., 2019)

Recently, an extensive secretomic study compared features of *in silico* secretomes of 44 helminth species including parasites of plants, animals and even free-living worms. The secretomes of schistosomes have been shown to be the smallest in terms of protein counts and the proportion of secreted proteins in the proteome, which varied around 4 %, compared with an average of7.6 %. The authors suggest that the small amount of secreted proteins by schistosomes as animal endoparasites are associated with the evasion from the immune system of the host versus the enormous amount of secreted proteins in plant parasites (Cuesta-Astroz et al., 2017). However, this does not seem very convincing, as other trematodes, such as *Fasciola hepatica*, have a significantly higher number of secreted proteins (992 – 12,59 % of total valid sequences), than schistosomes as well.

For an overview of current -omics progress in schistosomes, see **Table 3..** For proteomic and secretomic studies focused on schistosomes, see **Table 4.**

Table 3. Summary of current -omics knowledge of selected schistosomes from WormBase ParaSite (Howe et al., 2017). It is worth to mention that listed numbers do not correspond to the exact numbers of genes and protein *per se*, but only the stage of knowledge to date.

Species	Genome size	Coding genes	Gene transcripts	Predicted proteome	Predicted secretome
S. mansoni	410 Mb	10, 144	14,528	11 774	431
S. japonicum	403 Mb	12, 738	12, 738	12, 738	476
S. haematobium	376 Mb	11, 140	11, 140	11, 140	379
T. regenti	702 Mb	22, 185	12, 705	12,705	?

Table 4. Overview of proteomic/secretomic studies in schistosomes. E - egg, **M** - miracidium. **S** - sporocyst, **C**- cercaria, **Sch** - schistosomolum, **A**- adult

Species	**Life stages**	**Publications**
S. haematobium	**A**	Higón et al. 2011
S. japonicum	**E, C, Sch, A,**	F. Liu et al. 2009; Hong et al. 2013; M. Liu et al. 2015; Cao et al. 2016
S.mansoni	**E, M, C, Sch, A,**	Curwen et al. 2004; Knudsen et al. 2005; Curwen et al. 2006; Cass et al. 2007; Hansell et al. 2008; Wu et al. 2009; E. Ridi and Tallima 2009; Mathieson and Wilson 2010; deWalick et al. 2011 Marcilla et al. 2012; Dvořák et al. 2015; Figueiredo et al. 2015; T. Wang et al. 2016; Sotillo et al. 2015, 2016;
T. regenti		No available literature

3.3.5. Collection of cercarial excretory-secretory products for proteomic analysis

Since the practical part of this work is focused on the collection of excretory-secretory products (ESPs), the description of various techniques and their advantages and disadvantages follows bellow.

The collection of cercarial secreted molecules and the analysis of its secretome is relatively specific task because the emptying of the parasite's penetration glands occurs simultaneously with the transformation to the next larval stage – schistosomulum. There are many approaches to induce transformation of the parasite in laboratory conditions, and there are many opinions on which of these methods is the most valid and describes the infection in natural conditions.

Mechanical transformations are simple methods leading to the emptying of penetration glands. These approaches involve parasite's tail removal by passing cercariae through the gauge needle or by vortexing (Curwen et al., 2006). Mechanical transformation is used in cultivation experiment because of the high survival rate of the parasites but it is also the most used method in secretomic studies of schistosome cercariae (Knudsen et al., 2005; Dvořák et al., 2008). Authors suggest that molecules released during this process best describe the content of acetabular glands (Knudsen et al., 2005). Despite the considerable advantages of this method, it is necessary to considerate the fact that during the mechanical transformation, bodies of the parasite are often damaged and so the sample is significantly contaminated with somatic proteins and it is then difficult to identify proteins of exclusively glands origin (Knudsen et al., 2005; Curwen et al., 2006).

Fatty acids commonly present on the host skin such as linoleic or linolenic acid are known to stimulate cercariae to release excretory-secretory products and shed their tails (Shiff et al., 2006). By the simple addition of a sufficient amount of these acids to the water with the larvae, the transformation to schistosomulum is achieved. This method uses a natural stimulus, however, during this type of transformation the cercaria loses glycocalyx, becomes sensitive to the osmotic pressure of water in the environment and soon after this stimulation dies (Salafsky et al., 1988). The death of the parasite then leads to rapid degradation processes, and the sample can be contaminated with somatic proteins. Furthermore, previous observations suggest, that stimulation with linoleic acid leads only to emptying of post-acetabular glands which contain a thick mucous substance, which facilitates attachment of cercariae to the skin surface, while pre-acetabular glands containing proteolytic enzymes for penetration are emptied after cercariae reach the 'spinous' layer of the middle part of the epidermis (Salter et al., 2000; McKerrow and Salter, 2002).

In newer secretomic studies the **skin of the host** is used as a stimulus for transformation. The purpose of this transformation method is to imitate the penetration process as it happens in natural conditions. Sections of the host skin are exposed to larvae and the water is then analysed. In addition, it is a great advantage to take secretory products at certain time intervals and then compare the spectrum of proteins at different stages of penetration. During this method, there is no mortality reported (Hansell et al., 2008; M. Liu et al., 2015) (**Figure 4.**).

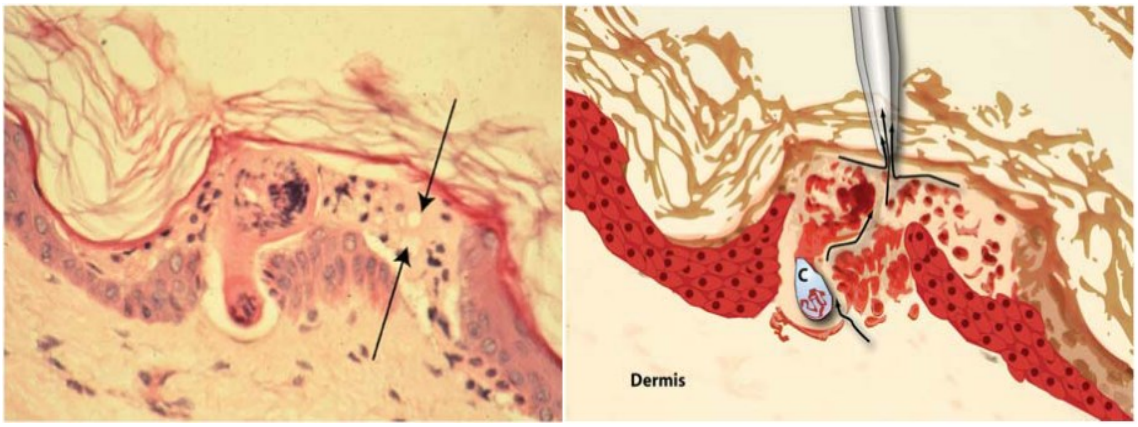

Figure 4. Cercaria and invasive tunnel in epidermis of human skin at 1/2-hour post invasion. A). The parasite larva is just entering the dermis. Note penetration tunnel (arrows) formed from destruction of epidermal cells by both acantholysis and apoptosis. B) Model of presumed acquisition of fluid from skin invaded by schistosome cercariae (Hansell et al., 2008).

Another method used for transformation of cercariae into the schistosomula is to incubate the parasites in a medium containing **the fresh or inactivated serum of the definitive host**. However, this method of transformation often leads to the death of larvae (W. Wang et al., 2006) and the incubation medium containing serum is not suitable for further processing by mass spectrometry.

Cercarial ESPs contain a complex mixture of proteins with various functions. As it results from the context, the major compounds of these products are peptidases which can make up to 50 % of cercarial products (Curwen et al., 2006). Their function is obvious – to disintegrate skin proteins. The presence of such a huge amount of peptidases is also associated with the presence of specific protease inhibitors, however, their function in the context of the penetration glands has not been satisfactorily explained. In all studies dedicated to cercarial secretions, calcium-binding proteins are also regularly found. These discoveries correspond to a high concentration of calcium in the acetabular glands of the parasite, but the purpose of this has not yet been elucidated (Dresden and Edlin, 1975). A significant part of the cercaria secretions consists of exclusively cytosolic proteins, such as glycolytic enzymes or heat shock proteins (HSPs). It was discussed that these proteins may get to the samples by contamination from damaged somatic cells, however, their presence in the cercaria secretions is a logical consequence of the holocrine mechanism of penetration glands secretion. The secretory glands are represented by individual cells and during the secretion, the complete content of these cells is extracted by muscle contraction (Curwen et al., 2006).

3.3.6. Peptidases and their main features

Peptidases (proteases/proteinases) are enzymes able to cleave proteins by hydrolysis of peptide bonds between individual amino acids in the peptide chain. It is a very diversified group of biomolecules, which play crucial roles in all known organism including viruses. These enzymes are involved in several physiological responses from simple digestion of proteins from ingested food to highly regulated biochemical cascades (for example blood clotting or activation of the prophenoloxidase cascade in the immune response of invertebrates). Peptidases can therefore act both destructively, as well as function activators or as part of signal pathways. Peptidases are encoded by approximately 2 % of the all known genes, and it is estimated that up to 14 % of all known human proteolytic enzymes are investigated as potential drug targets. Equally important is the regulation of the activity of these enzymes, which is most often mediated by proteins that specifically inhibit them. Peptidases and their inhibitors are therefore very important groups of molecules, and their knowledge and characterization are critical to understanding many biological processes (Neurath, 1986).

Proteolytic enzymes are also a highly studied group of molecules in helminthology. It has been repeatedly shown, that these enzymes play a crucial role in the life cycle of parasitic worms across all life stages. These include, for example, enzymes for host invasion (Salter et al., 2000), migration within the host (Law et al., 2003), but also digestive enzymes (Hall et al., 2011) or peptidases modulating the hosts immune responses (Jenkins et al., 2005). Equally important is their significance in the pathological manifestations of the diseases that these parasites cause. It is precisely peptidases that are often directly associated with pathogenic mechanisms. With the development of molecular, biochemical and immunological methods, the study of these biomolecules in parasitology becomes even more important. The pressure to discover the functions of individual peptidases and their role in parasite-host interactions has repeatedly led to the discovery of potential diagnostic markers or due to the antigenic properties of these molecules also to promising results in the field of vaccine development against severe helminthiases (Yan et al., 2005; J. Wang et al., 2013; Jiz et al., 2015).

There are three ways to divide proteolytic enzymes – based on "place of action" (**Table 5**.), type of catalysis or based on structure, homology and function. Based on the site in the chain where the enzyme acts, peptidases are divided into two broad groups – exopeptidases and endopeptidases. Exopeptidases cleave the individual amino acids from the ends of the peptides while the endopeptidases cleave the protein within the chain, significantly disrupting its structure, and are therefore much more important for the chemical processes in the organism (Barett et al., 2013). Further division is then based on the mechanism of peptidase catalysis. This is determined by the chemical properties of the amino acid side chains located at the so-called active site of the enzyme, which also determines peptidase's specificity.

Based on the primary structure (the order of AA in the chain), families are formed where each member shares a statistically significant similarity with at least one member of the family. Even though the similarity of the amino acid sequences of the individual peptidases is often not statistically significant, these families are divided into their superior clans based on their possible evolutionary relationship. This is based on the linear order of the residues of the active site in the

polypeptide chains and the tertiary structure of the protein. The clan belonging of each peptidase then expresses the uppercase letters defining the catalytic type (S, M, C, A, T, P, N, U, G) and the clan's lettering. Family is then expressed by the abbreviation of the catalytic type followed by the number (Kašný et al., 2009).

Table 5. Division of peptidases based on the character of the active site and its properties, according to the MEROPS 12.0 database (Rawlings et al., 2018).

Peptidase type	Letter	Mechanism of catalysis
Serine	**S**	Hydroxy group of serine in the active site as a nucleophile
Metallo-	**M**	Active water molecule activated with zinc or cobalt ions
Cysteine	**C**	The thiol group of cysteine in the active site as a nucleophile
Aspartate	**A**	Active water molecule of aspartic acid in active
Threonine	**T**	Hydroxy group of threonine in the active site as a nucleophile
Mixed peptidases	**P**	Hydroxy group of threonine or serine in the active site
Asparaginyl lyasis	**N**	Elimination reaction in presence of asparagine (Oda, 2012)
Unknown	**U**	Unknown
Glutamate	**G**	Active water molecule of glutamic acid in the active site

A very important feature of each peptidase is its substrate specificity. Peptidases specifically cleave protein substrates from the N– or C– terminus or somewhere in between of the substrate via binding of the active site to the amino acid side chains of the substrate flanking the cleavage site. Side chains of the active site in the peptidase are composed of adjacent pockets and each partial pocket of the active site binds to the corresponding chain in the substrate sequence. According to this definition, the amino acid chains in the substrate sequence are sequentially numbered outwardly from the cleavage sites such as P4–P3–P2–P1–P1'–P2'–P3'–P4 where the cleavage site is located between P1 and P1 ', while parts of the active site are appropriately labelled as S4–S3–S2–S1–S1'–S2'–S3'–S4' (**Figure 5.**). This numbering is important in describing the substrate specificity of each group of these biomolecules (Harper and Berger, 1972).

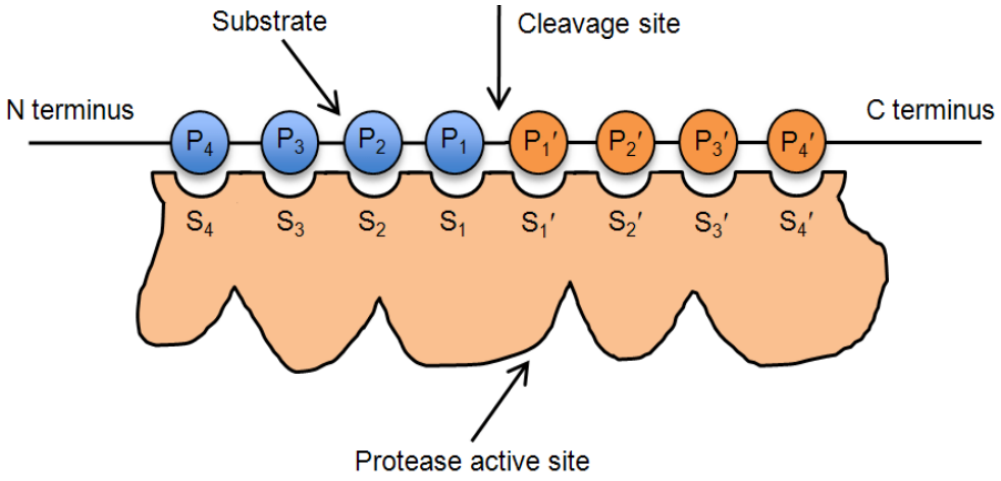

Figure 5. **The nomenclature of the substrate specificity of peptidase. (**https://prosper.erc.monash.edu.au)

3.4. Peptidases as penetration enzymes of schistosome cercariae

As mentioned in **Table 2**, there are 9 groups of peptidases. In this thesis, I will further describe only serine and cysteine peptidases since they are most explored within schistosome biology and have a proven role in cercarial host penetration.

3.4.1. Serine peptidases

As mentioned above, serine peptidases are the most numerous group of proteolytic enzymes. The largest and most significant family of these molecules is the S1 family belonging to the PA clan – P stands for clans containing families of more than one catalytic type (Dvořák and Horn, 2018). These enzymes are synthesized in the form zymogens – inactive proenzymes, which need to change the conformation by a specific fission process to their activation (Hedstrom, 2002). The active site of most of these peptidases consists of the side chains of histidine, aspartate, and serine, which are not linearly next to each other, but due to the spatial conformation of the protein, they are very close and form a catalytic triad. Three main types of peptidases from the S1 family have been described – trypsin-like, chymotrypsin-like and elastase-like, differing in the specificity of substrate cleaving. For trypsin-like serine peptidases, arginine or lysine is cleaved at position P1, for chymotrypsin-like peptidase one of the hydrophobic amino acids is at position P1, most commonly phenylalanine, and finally, elastase-like peptidase prefers small aliphatic amino acids, typically alanine at this position (Yang et al., 2015).

Cercarial elastases (clan PA, family S1)

The most studied group of serine peptidases of trematodes are cercarial elastases (CE). It is a group of 25–30 kDa peptidases with a typical chymotrypsin-like activity that is exclusive for schistosomes. These enzymes got their name when the pilot studies carried out on the excretory-secretory products of *S. mansoni* correctly determined the ability of these molecules to cleave elastin (Gazzinelli and Pellegrino, 1964) and other mammalian skin proteins (McKerrow et al., 1985). These properties, absolutely crucial for the successful penetration of cercaria and thus the transmission of parasites (Salter et al., 2000) have ensured these enzymes the attention of researchers to the present day. The long-presumed CE function in host penetration process has also been confirmed by studies dedicated to the expression and localization of this type of peptidase. The expression of the genes encoding these enzymes and the localization of these enzymes is limited exclusively to the stage of the cercaria and the germ balls – germ stage of the parasites where the cercariae develop within the sporocysts found most commonly in the hepatopancreatic gland of the snail (Fishelson et al., 1992; Ingram et al., 2012). The publication of the sequenced *S. mansoni* genome (Sajid et al., 2009) led to the discovery of 8 complete gene sequences corresponding to cercarial elastases. These isoforms were divided into two groups based on sequence similarity, SmCE1 and SmCE2, which contain the corresponding isoforms – SmCE1a.1, SmCE1a.2, SmCE1b, SmCE1c and SmCE2a.1, SmCE2a.2, SmCE2a.3. SmCE2b. Subsequent deeper investigations revealed that only 6 of them are translated into the protein form. SmCE1c is a pseudogene - DNA segment where transcription or translation is not present, and

the SmCE2b isoform was found only at the mRNA level; the protein form was not found by mass spectrometry analysis (Ingram et al., 2012). Besides the *S. mansoni* genome, other genomes of human schistosomes have also been published. In the case of *S. haematobium*, the genes for isoforms 1a, 1b, 2a and 2b have been discovered, of which two orthologs, ShCE2b and ShCE1a.1, contain a complete catalytic triad and also serve as primary penetration enzymes (Salter et al. 2002, Young et al. 2012). For the zoonotic *S. japonicum*, only one gene coding for this enzyme was found, the SjCE2b isoform (Zhou et al., 2009). Protein form encoded by this gene is found also in the ESPs of cercaria of this species, and very recent publication claims, that recombinant protein form displays similar enzymatic activity and substrate preference to *S. mansoni* cercarial secretions suggesting its possible role in penetration process (Zhang et al., 2018). This is in conflict with original general belief that the primary penetration enzymes of *S. japonicum* cercariae are cysteine peptidases, which are one the most dominant peptidase groups the ESPs (M. Liu et al., 2015). The last species, where the cercarial elastase isoform was discovered is *Schistosomatium douthitti*, schistosomal trematode of rodents, which causes cercarial dermatitis in humans (Malek, 1977). For phylogeny of cercarial elastases, see **Figure 6.**

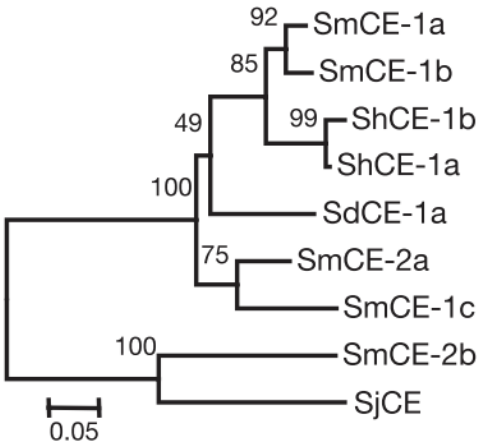

Figure 6. A phylogeny of the elastase family in schistosomes using the neighbour-joining method. Bootstrap values above branches. (Zhou et al., 2009).

These findings open the door to speculation that multiplicated cercarial elastase genes cohere to adaptation on the human definitive host (Dvořák and Horn, 2018). To this hypothesis also contributes to the fact that in species using cercarial elastases as dominant penetration enzymes, tissue migration is slower and acts less destructively compared to species considered using mainly cysteine peptidases for invasion (He et al. 2002; He et al. 2005). The low immunogenicity of cercarial elastases does not lead to strong skin proinflammatory reactions after penetration of cercaria as for zoophilic schistosomes possibly using cathepsins (Chlichlia et al., 2005; Dvořák et al. 2008; El-Faham et al. 2017) and it was experimentally proved, that cercarial elastases can degrade immunoglobulins IgE (Aslam et al., 2008).

Cercarial elastase of *T. regenti*

Biochemical, immunological, and immunohistochemical experiments did not confirm the presence of CE orthologs in *T. regenti* (Mikeš et al. 2005; Kašný et al. 2007; Dvořák et al. 2008). Neither PCR-based approaches with template cDNA obtained by reverse transcription of mRNA from cercarial germ balls did not succeed to find a gene corresponding to this type of enzyme (Dolečková et al., 2007). These findings have led to hypotheses that neither the zoophilic *S. japonicum*, nor ornithophilic *T. regenti* uses serine peptidases as a tool for penetration into a definitive host (Kašný et al., 2009). However, after the genome of *T. regenti* was released, a partial gene sequence sharing a 75 % identity with SjCE2b was found (Mikeš, 2016, personal communication).

3.4.2. Cysteine peptidases

Cysteine peptidases play a major role in life cycles either of zoophilic and anthropophilic schistosomes. They possibly participate in cercarial skin penetration (Dvořák et al., 2008), blood digestion (Caffrey et al., 2004) and even in immune evasion (Dalton and Brindley, 1997). Because of their immunogenicity and high levels of expression in the vertebrate stages, they have also become suitable serodiagnostic markers as well as suitable candidates for the development of vaccines against schistosomiasis and other trematodoses (McKerrow et al., 2006).

Papain-like peptidases (clan CA, family C1A)

Family C1A is the largest family of cysteine peptidases. Since the majority of trematodes are obligatory blood feeders, the major physiological role of papain-like peptidases of these parasites is related to the processing and digestion of blood protein components (Brindley et al., 1997; Caffrey et al., 2004), however some of these peptidases might play important roles in non-feeding larval stages such as penetration and early migration in the host tissues (Dvořák et al., 2008; El Ridi et al., 2014)

Cathepsin B

Today, over 20 genes coding cathepsin B1 and B2 sharing similar biochemical properties is known (Caffrey et al., 2002, 2004; Kašný et al., 2009). Since most flukes are obligatorily blood-feeding and the presence of these types of enzymes has been localized in the gut of schistosomules and adult worms, the main physiological role of cathepsins B of these parasites has been attributed to the processing and digestion of blood protein components (Brindley et al., 1997; Caffrey et al., 2004). Later, cathepsin B-like activity was measured in the ESPs of *T. regenti*, *S. japonicum* and *S. mansoni* cercariae indicating their possible role in the penetration process (Kašný et al., 2007; Dvořák et al., 2008). The following experiments showed that the recombinant forms of CB2 and the active fraction of cysteine peptidases of *T. regenti* successfully cleave elastin and are located in the preacetabular glands of the *T. regenti* (**Figure 7.**) and *S. japonicum* cercariae and thus their role in penetration process was accepted. Furthermore, regarding the seeming absence of the cercarial elastase orthologue in *T. regenti* and obvious dominance of cathepsin in *S. japonicum* cercariae it is strongly believed, that the cathepsin Bs substituted the functions of cercarial elastases in more zoonotic species (Dvořák et al. 2008; Dolečková et al., 2009; Liu et al., 2015).

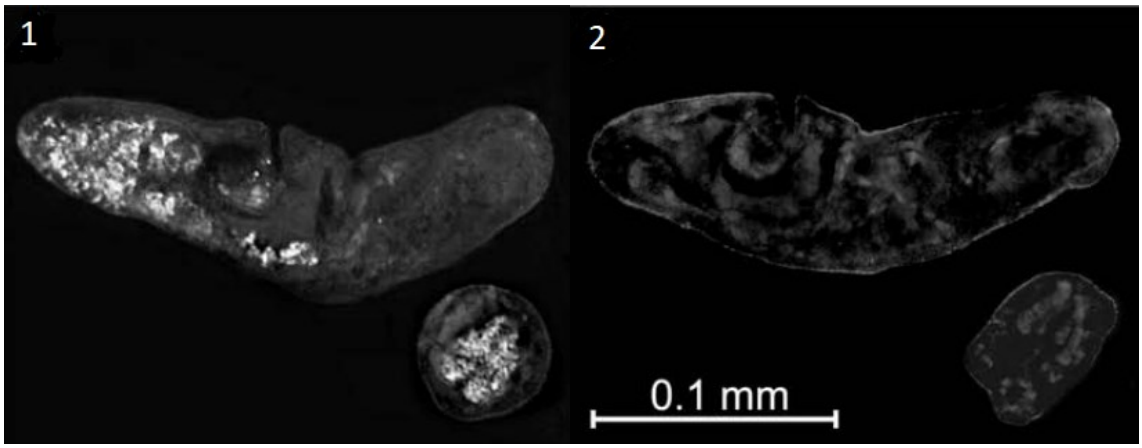

Figure 7. The immunolocalisation of cathepsin B2 in longitudinal and transverse sections of cercariae of *Trichobilharzia regenti* by indirect immuno-fluorescence microscopy. (1) Localization of cathepsin B2 in cercaria**. (2)** Control (Dolečková et al. 2009, adjusted)

4. Material and methods

4.1. Collection of the excretory-secretory products and their characterization

Collection of the samples was performed twice for both methods, analysed separately by mass spectrometry and results for each method were pooled.

4.1.1. Collection of *T. regenti* cercariae

Intermediate hosts, *Radix peregra* snails were transferred from breeding aquariums to a 200 ml glass filled with tap water and placed under a LED light source (10W). After 2 hours of light stimulation, cercariae were concentrated in concentration bottle using cercariae positive phototaxis and placed in 50 ml falcon tube. Larvae were then counted under a microscope Olympus CX21 (10x in 20 µl and extrapolated to the volume) and divided to new 50 ml falcon tubes with 15 ml of settled tap water by 5000 individuals.

4.1.2. Collection of ESPs after stimulation with linoleic acid

For a collection of ESPs with linoleic acid, a modified protocol described in Knudsen et al., 2005 was used. A liquid ≥99% linoleic acid (Sigma–Aldrich) was added to the falcon tube with cercariae to a final concentration of 10 µl/ml. The cercariae were incubated for 90 minutes in this solution. The water containing the ESPs, live cercariae and the cercarial debris was then incubated on ice and repeatedly centrifuged at 4 °C at a rate of 13.200 rpm to get rid of parasite bodies. After centrifugation, 10 ml of the supernatant was collected and filtered through Ultra-15 centrifugal filter unit (AMICON) of 10 kDa sensitivity by sequential transfer and centrifugation at 4 °C for 4x15 minutes at 6000 rpm. Concentrated samples were lyophilized (CT/DW 60E, Heto) overnight and sent to the Central Laboratory Proteomics in CEITEC research centre at Masaryk University in Brno for further processing on a mass spectrometer.

4.1.3. Glands staining

After the collection of ESPs using linoleic as a stimulant, transformed parasites' glands were stained and examined to evaluate the success of the method. Circumacetabular glands and postacetabular glands were stained with alizarin and lithium carmine respectively as described in Holická, 2009:

2% alizarin solution was prepared by boiling alizarin saturated aqueous solution and diluted with water and 1% lithium carmine was prepared by adding 2.5 grams of carmine to 100 ml of boiling saturated Li_2CO_3 solution and diluted.

2% alizarin was added to the cercariae for 3 minutes, cercariae were then fixed with 70% ethanol, 1% lithium carmine was added to the cercariae and the stained larvae were observed under the light microscope (Olympus CX21).

4.1.4. Collection of ESPs released in the presence of host (duck) skin

For a collection of ESPs with duck skin, a modified protocol described in Hansell et al. 2008 was used. Legs of a freshly killed duck *Anas platyrhynchos f. domestica* were amputated and gently washed with clean tap water. The skin was excised and cut to approximately 25 cm^2 pieces which were then attached with a rubber band dermal side up to the neck of the falcon tube containing cercariae to completely cover the top. Falcon tubes were placed upside down in a Petri dish containing the Basal Medium Eagle (Gibco) to avoid degradation of the skin tissue during the experiment and incubated for 90 minutes in room temperature (RT) (**Figure 8**). After incubation, skins were removed from the necks and kept for histological examination (**4.1.5.**). The contents of the falcon tubes were then processed as in point **4.1.2**.

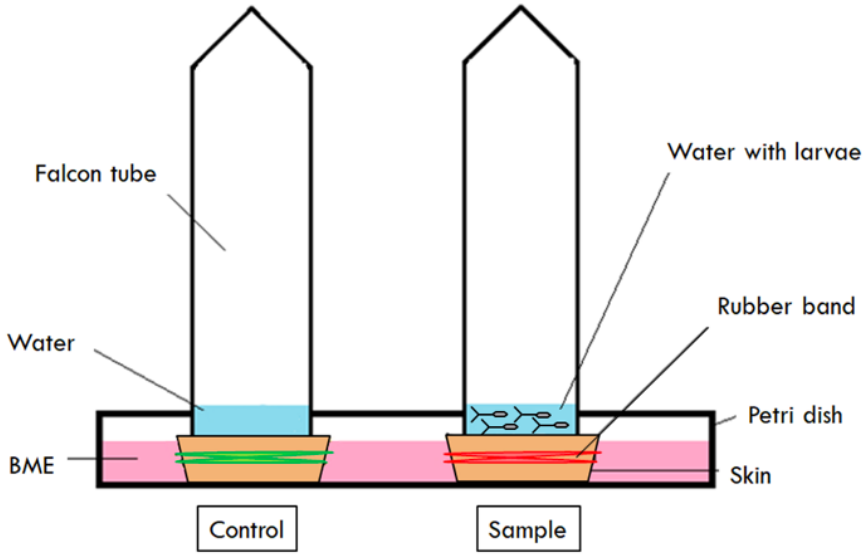

Figure 8. Collection of ESPs released in the presence of host (duck) skin

4.1.5. Histological examination

The skin was fixed in Bouin's fixative (Sigma) and embedded into a JB–4 (Polysciences, Inc.) hard resin according to the manufacturer's instructions. The solid resin block was then cut with a Thermo-Shandon Finesse ME microtome (Thermo Scientific) on 4 µm thick sections with a homemade glass knife. The sections were then placed on drops of distilled water prepared on the glass with an extra adhesive surface (X–tra thermal glass, Leica). After droplets dried up on the hot plate (VD 2n, P–LAB) at 40 °C, the sections were stained routinely with Hematoxylin – Eosin (Abcam).

Hematoxylin stain solution was pipetted on the slide to completely cover the sections and the sections were stained for 5 minutes. After staining, the slide was rinsed under running tap water until the extra stain was completely washed out. Eosin stain solution was pipetted on the slide to completely cover the sections and the sections were stained for 30 seconds. After the staining, the slide was rinsed under running tap water until the extra stain was completely washed out and observed with a light microscope (Olympus CX21)

4.1.6. Mass-spectrometry and ESPs protein identifications

All ESPs samples were analysed by Central Laboratory Proteomics in CEITEC research centre at Masaryk University in Brno. Briefly, lyophilized ESPs samples were thawed, solubilized, digested with trypsin at 37 °C and analysed with LC-MS/MS analyser Orbitrap Fusion Lumos (Thermo Scientific). LC-MS data were processed with MaxQuant software (version 1.6.0.13) using search engine Andromeda. Several protein databases were used for identification of proteins in *T. regenti* cercariae ESPs samples (**Table 6.).**

All sample contaminants (from duck skin and the environment) were filtered and the sequences sets of all identified *T. regenti* proteins for both linoleic acid and skin induced samples were separately searched with BLASTp with standard E-value < 10^{-5} (Altschul et al., 1990) using MEROPS 12.0 database (Rawlings et al., 2018). Resulting identifications for both sample groups were additionally searched with BLASTp with standard E-value < 10^{-5} using UniRef100 protein database (Bateman et al., 2017). Identified peptidases were then identified by sequence similarity and E-value.

Table 6. Protein databases used for identification of proteins in ESPs samples.

Type of database	Source	Link
Contaminant database	MaxQuant documentation	http://lotus1.gwdg.de/mpg/mmbc/maxquant_input.nsf/7994124a4298328fc125748d0048fee2/$FILE/contaminants.fasta
Anas platyrhynchos predicted proteome from genome	UniProt reference proteomes	ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/reference_proteomes/Eukaryota/UP000050795_157069.fasta.gz
T. regenti predicted proteome from genome	WormBase ParaSite (Howe et al., 2017)	ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS12/species/trichobilharzia_regenti/PRJEB4662/trichobilharzia_regenti.PRJEB4662.WBPS12.protein.fa.gz
T. regenti predicted proteome from transcriptome	(Leontovyč et al., 2016)	(Leontovyč et al., 2016)

4.2. Completion of *T. regenti* cercarial elastase whole gene sequence

A partial nucleotide sequence of *T. regenti* cercarial elastase was identified in the genome (see **Table 46.** in **Attachment**)

4.2.1. Primer design

The primers were designed based on a partial genomic sequence (233bp) of the TrCE gene with a sequence identity of 75 % with cercarial elastase SjCE2b of *Schistosoma japonicum*. The primers were then diluted to a concentration of 100μM according to the manufacturer's instructions and stored in a freezer at 20 °C.

4.2.2. mRNA samples preparation and isolation

As a source of mRNAs, sporocysts, cercariae and the healthy snail were used for further experiments.

Sporocyst material was collected as described by Dolečková et al. 2009 by excision of the entire digestive gland of the infected snail. Furthermore, the hepatopancreatic gland of the healthy snail was also excised as a control. About 50,000 cercariae were collected the same way as described in section **4.1.1.** The larvae were concentrated with Cell Strainer 40μl Nylon (Corning).

Infected organ, healthy organ and the concentrated larvae were separately placed to 1 ml of TRIzol reagent (Life technology) and all samples were homogenized with a sonicator (Vibracell 72405) 6x10s on ice. The rest of the procedure was performed according to the manufacturer's protocol, and total RNA was eluted to 50 μl of DEPC-treated water (Invitrogen). The concentration of the isolated RNA samples was then measured on a NanoDrop ND 1000 (Thermo Scientific) and all samples were stored at −80 °C to ensure the stability of nucleic acids.

4.2.3. Reverse transcription

cDNA synthesis was carried out with 1 μg of total RNA from each sample using oligo$(dT)_{18}$ primer and commercial Transcriptor First Strand cDNA Synthesis Kit (Roche) following the manufacturer's instructions. The concentration of the synthesized cDNA was measured on a Nanodrop ND 1000 (Thermo Scientific) and stored at –20 °C.

4.2.4. PCR to identify cercarial elastase transcript

To identify the cercarial elastase transcript, the PCR reaction was performed with all three synthesised cDNA samples as a template (**Table 7.**). cDNA samples were diluted to a concentration of 100 ng/μl. The PCR mixture was prepared under sterile conditions in a flow box cleaned with PCR clean (Minerva Biolabs). All reaction components were thawed, vortexed and centrifuged. After all, components were added to a sterile 0.2 ml PCR tubes (**Table 8.**), the samples were again vortexed, briefly spun, and inserted into the C 1000 Thermal Cycler (Bio-Rad) (**Table 9.)**

Table 7. Designed primers for a partial sequence of *T. regenti* cercarial elastase

Primer	Sequence
TrCE Forward	5'– ATG ATT CCT ACT TCG TCG CTG C – 3'
TrCE Reverse	5'– ACC CGG TTG ATT GAT TTC GGT – 3'

Table 8. PCR mixture composition

Component	25 µl reaction
EmeraldGT Mastermix (Takara)	12.5 µl
cDNA (100ng/µl)	1 µl
Forward primer	0.5 µl
Reverse primer	0.5 µl
ddH_2O	10.5 µl

Table 9. Thermal cycler programme used for PCR with EmeraldGT Mastermix

Step	Reaction	Temperature	Time
1	Denaturation 1	94 °C	300s
2	Denaturation 2	94 °C	30s
3	Primer annealing	54 °C	30s
4	Extension	72 °C	30s
Step 2–4 repeated 35x			
5	Final extension	72 °C	360s

4.2.5. DNA electrophoresis

2% agarose gel electrophoresis under a voltage of 100 V was performed. Agarose gel was prepared with 40 ml of TAE buffer (Thermo Scientific) and 0.8 g of powdered agarose (Sigma-Aldrich). 40 µl SYBR Safe DNA Gel Stain (Invitrogen) was also added to the gel. The marker GeneRuler DNA Ladder mix (Thermo Scientific) was used to identify the products size amplified in PCR and the products were visualized in the UV chamber.

4.2.6. DNA isolation

The visible amplicon was then excised from the gel with a sterilized scalpel and DNA was extracted using Zymoclean Gel DNA Recovery Kit (Zymo Research) according to the manufacturer's instruction and eluted with sterile water. The concentration of the sample was measured on NanoDrop 1000 (Thermo Scientific).

4.2.7. Cloning into pGEM®-T Easy Vector System

DNA was directly inserted into the pGEM®-T Easy Vector System (Promega) according to the protocol provided by the manufacturer (**Table 10.**) The ligation reaction was incubated at 4 °C overnight.

Table 10. Ligation reaction composition

Component	8 µl reaction
2x Rapid Ligation Buffer	4 µl
pGEM®-T Easy Vector	1 µl
PCR product	2 µl
T4 DNA Ligase	1 µ

4.2.8. Transformation of One Shot TOP10 *Escherichia coli* competent cells

2 µl of the ligation reaction was added to the 50 µl thawed of the *Escherichia coli* competent cells (One Shot TOP10; Thermofisher Scientific) and the procedure was carried out following manufacturer's protocol. Heat shock was performed by incubation of the cells for 42 seconds at 42 °C and cells were incubated in 250 µl of SOC media (Invitrogen) at 37 °C for 1.5 hours. 100 µl of transformed cells were spread with glass spreader on LB/agar plates treated with 12.5 µl of ampicillin (100mM).

4.2.9. PCR colony screening for the presence of the DNA insert

The success of ligation and transformation was confirmed by the disintegration of the bacterial cells, which served as a DNA template for following PCR reaction. Small portions of grown colonies were transferred from the culture LB/agar plate to the 0.2 PCR tubes containing 9 µl of dH_2O. The tubes were then placed to the thermal cycler to disintegrate bacterial cells (cell crack, see **Table 12.)**. After that, the rest of the reaction components was added, and PCR with M13 primers was performed (**Table 11.**, **Table 12.**, **Table 13.**).

Table 11. Designed primers for partial sequence of T. regenti cercarial elastase

Primer	Sequence
M13 Forward	5' – TGT AAA ACG ACG GCC AGT – 3'
M13 Reverse	5' – CAG GAA ACA GCT ATG ACC – 3'

Table 12. Thermal cycler programme to perform cell crack

Step	Reaction	Temperature	Time
1	Heating	96 °C	300s
2	Cooling	50 °C	110s
3	Heating	96 °C	110s
4	Cooling	45 °C	60s
5	Heating	96 °C	60s
6	Cooling	40 °C	60s
7	Chilling	4 °C	120s

Table 13. Composition of PCR mixture

Component	23 µl reaction
EmeraldGT Mastermix (Takara)	12 µl
Crack product	10 µl
Primer M13 forward	0.5 µl
Primer M13 reverse	0.5 µl

Table 14. Thermal cycler programme used for PCR with EmeraldGT Mastermix

Step	Reaction	Temperature	Time
1	Denaturation 1	94 °C	300s
2	Denaturation 2	94 °C	30s
3	Primer annealing	55 °C	30s
4	Extension	72 °C	30s
Step 2–4 repeated 35x			
5	Final extension	72 °C	360s

4.2.10. Plasmid DNA propagation and isolation

A portion of positive colony was picked with a sterile pipette tip and incubated in 15ml falcon tube with 5 ml of LB medium containing 5 µl of 100 mg/ml ampicillin. at 37 °C overnight.

Plasmid DNA isolation was performed using the Hybrid-Q plasmid Rapidprep kit (GeneAll®) following the manufacturer's instructions and eluted into 50 µl of sterile distilled water. Plasmid DNA was then sequenced with M13 primers (DNA Sequencing Laboratory, Faculty of Science, Charles University in Prague).

4.2.11. Rapid Amplification of cDNA Ends (RACE) of cercarial elastase

After verification of the partial gene sequence, Rapid Amplification of cDNA Ends (RACE) to obtain the whole gene sequence was performed.

Amplification of the 3' end of the gene was performed using the GeneRacer® Kit (Invitrogen). Gene-Specific Primer (GSP) was designed following the provided protocol (**Table 15.**). First-strand cDNA synthesis for RACE PCR was carried out with 1 µg of cercarial mRNA using oligo$(dT)_{20}$ primer and SuperScript® III reverse transcriptase (Invitrogen) following the manufacturer's instructions. Synthesised cDNA was diluted with sterile distilled water to a concentration of 100 ng/µl and directly used in gradient 3' RACE PCR (**Table 16.** and **Table 17.**).

The DNA from obtained amplicons was isolated as described above. Extracted DNA was directly cloned into pGEM®-T Easy Vector, TOP10 competent cells were transformed and plasmid DNA isolated was as described above. Plasmid DNA was then sequenced with M13 primers (DNA Sequencing Laboratory, Faculty of Science, Charles University in Prague).

Table 15. Designed Gene Specific Primer for 3'RACE PCR

Primer	Sequence
GSP	5'– GAC GCT GGT GTC CAC CCG AGC TGT GA – 3'

Table 16. Composition of 3' RACE PCR reaction

Component	25 µl reaction
Primestar MAX Matermix	12.5 µl
GSP (100mM)	0.5 µl
Generacer primer	1.5 µl
RACE cDNA (100 ng/µl)	1 µl
ddH_2O	9.5 µl

Table 17. Thermal cycler program used for 3'RACE PCR

Step	Reaction	Temperature	Time
1	Denaturation	98 °C	10s
2	Primer annealing	68, 70, 72 °C	30s
3	Extension	72 °C	60s
Step 1–3 repeated 35x			

4.2.12. TrCE gene sequence assembly

To assemble the whole gene sequence, Geneious R10.1 program was used. The sequence was then translated into the amino acid by the ExPASy Translate Tool (Artimo et al., 2012), the correct reading frame was determined. Sequence was also compared with the amino acid sequence of *S. japonicum* cercarial elastase 2b and the coding region of the *T. regenti* gene was determined.

4.3. *T. regenti* cercarial elastase recombinant protein expression in yeast

Since TrCE was identified in high abundance in ESPs of *T. regenti* and whole gene sequence was obtained, we decided to further characterize it by producing recombinant protein.

The PichiaPink Expression System (Thermofisher Scientific) was chosen for the protein expression in a yeast system. The whole procedure was performed according to the manufacturer's protocols, so below is just a brief overview of the main steps. All yeast expression experiments were carried out as a part of an internship at John Dalton's Laboratory at Queen's University Belfast under supervision of prof. John P. Dalton.

4.3.1. Expression gene design and propagation

The TrCE gene was analysed for the N-glycosylation sites with NetNGlyc 1.0, the signal peptide sequence was replaced with 5' *Mlyl* restriction site, His-tag was added starting with glycin-prolin for better accessibility, the stop codon was removed and placed before 3' *KpnI* restriction site. The gene was then synthesized and inserted into the pPINKα-HC vector by Eurofins Genomics company. Plasmid DNA was reproduced in *E. coli* TOP10 cells as described above and extracted using PureLink Quick Plasmid Miniprep Kit following the manufacturer's protocol.

4.3.2. Preparation of the transforming DNA

100 µg of isolated plasmid DNA was linearized with digestion following a modified protocol provided by the manufacturer of the used *SpeI* restriction enzyme (New England Biolabs) (**Table 18.**). The reaction was carried out at 43 °C for 90 minutes and terminated by heating up to 80 °C for 20 minutes. Complete linearization was confirmed on 1% agarose gel DNA electrophoresis.

Linearized DNA was collected by ethanol precipitation as described in (Zeugen and Hartley, 1985). Briefly, 1/10 of the reaction volume of 3M sodium acetate and 2.5 volumes of 100% ethanol was added to the sample and sample was placed to –20 °C for 30 minutes. DNA was pelleted by centrifugation, washed with 80% ethanol, air drought and finally resuspended in 10 µl of sterile distilled water.

Table 18. The reaction mixture for digestion with the restriction enzyme *SpeI*

Component	50 µl reaction
CutSmart buffer 10X	5 µl
TrCE pPINKα-HC (100µg)	33 µl
SpeI	2 µl
ddH_2O	9 µl

4.3.3. Preparation of the yeast cells for transformation

Small portions of PichiaPink strain #1 and #4 formed yeast cell colonies grown on YPD/agar plates were used for the initial culture. Cells were incubated in YPD media at 24 °C at 200 rpm until the OD_{600} of the culture reached between 1.2 – 1.5. The cultures were centrifuged at 1500g at 4 °C for 5 minutes, medium discarded and the pellets were sequentially resuspended with 250 ml of ice-cold, sterile distilled water; 50 ml of ice-cold, sterile distilled water and 10 ml of ice-cold sterile 1M sorbitol with centrifugation between every resuspension. After the last centrifugation, 300 µl of ice-cold sterile sorbitol was added making a final volume of approximately 600 µl after the final resuspension.

4.3.4. Transformation of the cells by electroporation

100 µl of yeast cells mixed with 100 µg of linearized plasmid DNA was electro-pulsed with MicroPulser Electroporator (Bio-Rad) according to the instrument manufacturer's instructions for yeast cells ("Fungi" program). Transformed cells were then immediately resuspended in 1ml of ice-cold YPDS media, incubated at 24 °C for 2 hours and 100–300 µl of transformed cells was spread on PAD selection plates. The plates were incubated at 24 °C for 3 days until pink and white colonies formed. White colonies were then transferred on the fresh PAD selection plate.

4.3.5. PCR colony screening using the microwave

White colonies were screened for the presence of the insert with the modified protocol of the PCR colony screening described in PichiaPink Expression System.

A portion of the colony was transferred to 1.5 ml eppendorf tube containing 75 µl of TE buffer (Invitrogen). The cells were repeatedly microwaved on the highest power (700W) for 3; 2; 1,5; 1 and 0,5 minutes and vortexed and briefly spun between every microwaving step. Microwaved cells were frozen at –80 °C for 10 minutes, thawed at 95 °C for 2 minutes in the heating block and centrifuged for 15 minutes at 2,500 rpm. The supernatant was used as a template in the screening PCR (**Table 19.** and **Table 20.**). Resulting DNA amplicons were isolated as described in **4.2.6** and sequenced with α-factor forward primer and CYC1 reverse primer by Eurofins Genomics company.

Table 19. The reaction mixture of the screening PCR

Component	**25 µl reaction**
DreamTaq Green PCR Master Mix	12,5 µl
α-factor forward primer	0,5 µl
CYC1 reverse primer	0,5 µl
supernatant	5 µl
ddH_2O	6,5 µl

Table 20. Programme of the thermal cycler

Step	Reaction	Temperature	Time
1	Denaturation 1	94 °C	300s
2	Denaturation 2	94 °C	30s
3	Primer annealing	54 °C	30s
4	Extension	72 °C	30s
Step 2–4 repeated 35x			
5	Final extension	72 °C	360s

4.3.6. Large expression screening - "Yeasternblot"

After confirmation, that the gene of interest was successfully inserted into the genomic DNA of yeast cells, slightly modified large expression colony screening of the "Yeastern blot" described previously by Cregg et al. 2009 was performed. Briefly, 231 white colonies together with positive (recombinant *Fasciola hepatica* cathepsin L3 provided by prof. John Pius Dalton) and negative controls (untransformed cells) were grown on YPD/agar plates, "printed" on nitrocellulose membrane and induced on the membrane for 36 hours on YNB/2% methanol agar plates. Cells were then lysed, washed out of the membrane and the membrane with bound proteins was then treated as in **4.3.10** (Western blot).

4.3.7. Methanol expression induction of yeast (*P. pastoris*)

Positive colonies were grown in 1 litre of the BMGY media under constant shaking 200 rpm at 28 °C until the culture reached OD_{600} 2–6. Cell pellets were collected with centrifugation at 5000g for 5 minutes at RT and resuspended 300 ml of BMMY media to start the induction. Cells were then induced at 16/28 °C at 200 rpm for 48 hours with 3 ml of 100% methanol feeding every 12/24 hours. After the induction, supernatants were carefully collected after centrifugation at 5000g for 10 minutes at RT, diluted with 900 ml of column buffer and filtered through filter paper with 0.45 µm pore diameter using a vacuum pump.

4.3.8. Affinity chromatography

The recombinant protein was purified with affinity chromatography (Ni-NTA agarose, Thermo Fisher Scientific). 500 µl bed volume of the beads was washed 10 times with 1 ml of the wash buffer in the 10ml column. After the wash steps, the filtered supernatant was passed through the column via gravity flow in the cold room at 4 °C overnight. For further analysis, 1 ml of run-through was collected. The column was washed 15 times with pH 8 and 1 ml of wash step was also collected for the analysis. Bound proteins were eluted with 5 ml of elution buffer pH 7 added to the column by a millilitre. The eluate was dialyzed against 1x PBS (Sigma-Aldrich) pH 7 in the cold room overnight. After purification of the protein, all the samples were analysed on SDS-PAGE, Western blot as described below (sections **4.3.9** and **4.3.10**). Also, enzymatic assays were carried out (**4.3.11**)

4.3.9. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis – SDS-PAGE

The reaction mixtures were prepared (**Table 21.),** then vortexed, briefly spun, heated to 95 °C for 10 minutes and loaded on the 4%–20%/12% precast (Mini Protean TGX, Bio-Rad) or hand cast 12% acrylamide gel together with protein marker (Precision Plus Protein All Blue Standards/Precious Plus Protein Dual Xtra Standards, Bio-Rad). The electrophoresis was performed on Mini-PROTEAN® Tetra Cell (Bio-Rad) in TGS buffer (Sigma-Aldrich) under the constant voltage of 150 V for 1 hour. Protein bands were then visualized by staining with Coomassie Brilliant Blue G-250 (Thermo Scientific) or Bio-Safe (Bio-Rad) Coomassie Stain. Alternatively, the gels were not stained but used for the western blotting.

Table 21. SDS-PAGE sample composition

Component	24 µl reaction
2x Laemmli buffer (Bio-Rad)	12 µl
500 mM DTT	2 µl
Protein sample	10 µl

4.3.10. Western blot

After SDS-PAGE, the gel was placed to blotting buffer for 5 minutes as well as blotting papers and nitrocellulose membrane (Nitrocellulose membrane 45 µm; Bio-Rad). "Blotting sandwich" was assembled, placed into Mini Trans-Blot® cell (Bio-Rad) filled with blotting buffer and proteins were transferred to the membrane under the constant voltage of 100 V for 1 hour.

Alternatively, methanol activated PVDF (ImmunBlot) was used and the blotting was performed in Trans-Blot® Turbo Transfer System (Bio-Rad) with "Mixed MW Turbo" programme (2.5A, 25V) for 15 minutes.

Membrane with bound proteins was then blocked with 5% non-fat dry milk (Bio-Rad) for 2 hours at RT (or at 4 °C overnight) and probed with monoclonal anti-polyHistidine antibody produced in mouse H1029 (Sigma-Aldrich) diluted according to the manufacturer's instructions for 1.5 hours. The membrane was then washed (5x 5 minutes) with TBS-T buffer and probed with goat anti-mouse IgG peroxidase conjugate antibody DC02L (Sigma-Aldrich) diluted according to the manufacturer's instructions for 1 hour. The membrane was then washed again, and reactions were detected with SIGMA*FAST* 3,3′ Diaminobenzidine tablets (Sigma-Aldrich) or Immun-Blot® Opti-4CN (Bio-Rad) according to the manufacturers' protocols. All reagents were dissolved in TBS-T buffer.

4.3.11. Enzymatic assays

To measure an activity and substrate specificity of the TrCE, multiple enzymatic assays were performed with significant contribution from Carolina De Marco Verissimo PhD (Queen's University Belfast, United Kingdom). Kinetic assays were performed on 96-well plate with 10, 20 or 50 µL of each protein/well at 37 °C for 1 hour. The activity was measured on POLARstar Omega (BMG LABTECH)

Buffers

• Buffer 1: 100mM Hepes, 500 mM NaCl, 0.01% Brij, pH 7.
• Buffer 2: 200mM Tris-HCl, 150mM NaCl, 0.1% BSA, pH 7.4

Substrates (at 20 µM)
• Ala-Ala-Pro-Phe
• Ala-Ala-Pro-Val
• Phe-Arg
• Gly-Pro-Arg
• Leu-Arg
• Val-Ile-Arg
• Ile-Glu-Gly-Arg

4.4. *T. regenti* cercarial elastase recombinant protein expression in bacteria (*E. coli*)

Protein expression was performed with vector pET-28a(+) (Novagen) and competent *Escherichia coli* BL21 cells (Novagen). Two different genes were designed for this purpose:

1. cercarial elastase DNA expression construct with His-tag on 5' end of the gene was designed and amplified by PCR, and then ligated into the expression vector to form proper gene prepared to be expressed in bacterial cells.

2. the expression vector with the TrCE expression construct with His-tag on 3'end of the gene was provided by the manufacturer (only transformation and expression steps were performed) during the internship at Queen's University Belfast.

4.4.1. Expression gene design and propagation

To include His-tag coding sequence at the N-terminus of the gene and the thrombin cleavage site, restriction endonucleases *XhoI* and *NheI* (both New England Biolabs) were chosen based on the pET-28a(+)vector map provided by a manufacturer (Novagen). NEBCutter and REBASE programs have been checked to ensure that the chosen enzymes do not cleave the inserted gene. The signal peptide was predicted using SignalP 4.1 (Nielsen, 2017) and removed.

Specific primers (**Table 22.**) were then designed to synthesise the gene construct and PCR was performed (**Table 23.** and **Table 24.**). Amplicons were isolated from the gel with Zymoclean Gel DNA Recovery Kit.

Since used PrimeSTAR MAX Mastermix (Takara) does not synthesize the A-ends of the PCR product, A-tailing was performed according to a pGEM®-T Easy Vector protocol (**Table 25.).** A-tailing reaction was carried out at 70 °C for 30 minutes and 2 µl of the reaction was then used to ligate the insert into pGEM®-T Easy Vector. TOP10 competent cells were then transformed, grown and propagated plasmid DNA, was isolated with HybridQ Plasmid MiniPrep kit (GeneAll) as described above.

Table 22. Primers for TrCE expression DNA construct synthesis

Primer	Sequence
rTrCE Forward	5'– CATAGCTAGCACATGGTTAGTTCGTAGTGGA – 3'
rTrCE Reverse	5'– CTGGAATTCATTATGTCTAATATCTGACTCGAGTATC – 3'
Legend	**auxiliary bases** **restriction site** **gene specific sequence** **stop codon**

Table 23. PCR with expression primers reaction composition

Component	50 µl reaction
PrimeSTAR Max (Takara) DNA Polymerase	25 µl
Forward primer (100 nmol)	1 µl
Reverse primer (100 nmol)	1 µl
Sporocyst cDNA (100 ng/µl)	2 µl
ddH_2O	21 µl

Table 24. Thermal cycler programme for PCR with expression primers

Step	Reaction	Temperature	Time
1	Denaturation	98 °C	10s
2	Primer annealing	68, 70, 72 °C	30s
3	Extension	72 °C	60s
Step 1–3 repeated 35x			

Table 25. Composition of the A-tailing reaction

Component	10 µl reaction
PCR product	5 µl
Taq DNA polymerase 10X Reaction Buffer	1 µl
dATP	0.2 µl (to 0.2mM)
Ex Taq HS polymerase	1 µl (5 units)
dH_2O	2.8 µl

4.4.2. Double digestion of the plasmid with insert and expression vector pET28a(+)

Digestion for both plasmids was performed according to the enzymes' manufacturer's instructions (New England Biolabs) in a single reaction because both restriction enzymes share the same DNA cleaving temperature of 37 °C and 100% efficiency of digestion in provided buffer without any mutual inhibition. Plasmid DNA samples were digested for 15 minutes (**Table 26.**).

Table 26. Double digestion reaction composition

Component	50 µl reaction
Plasmid DNA	1 µg
CutSmart Buffer 10x	5 µl
Xho1	10 units (1 µl)
Nhe 1	10 units (1 µl)
dH_2O	to 50 µl

4.4.3. Vector and insert purification using the DNA agarose electrophoresis

Purification of both digested expression vector and insert DNA was performed by electrophoresis on a 1% agarose gel at 100 V for 40 minutes. Both purified DNAs were then extracted from the gel and isolated using a Zymoclean Gel DNA Recovery Kit and with sterile distilled water.

4.4.4. Ligation and transformation

Purified insert DNA was ligated to the pET28a(+) vector DNA according to the pET28a(+) vector manufacturer's instructions. The amount of vector and insert DNA needed for the ligation reaction was calculated according to the formula also provided by the manufacturer (**Figure 9.** and **Table 27.**)

After ligation, 2 µl from the reaction were added to 40 µl of thawed BL21 competent cells (Sigma-Aldrich) and heat-shock transformation was performed as described above. 100 µl of transformed cells were spread with glass spreader on LB/agar plates treated with 12.5 µl of kanamycin (50 mg/ml). A successful transformation was verified by cell crack PCR (see **4.2.9**), followed by DNA isolation and sequencing. Sequence accuracy was also re-verified by isolation of plasmid DNA using a HybridQ Plasmid MiniPrep (GeneAll) kit and re-sequencing.

$$\frac{ng\ of\ vector\ x\ insert\ size\ in\ kb}{vector\ size\ in\ kb} \times molar\ ratio\ insert{:}\ vector = ng\ of\ insert$$

$$\frac{55\ ng\ x\ 0{,}85\ kb}{5{,}5\ kb} \times \frac{6{,}5}{1} = 55{,}25\ ng\ insert$$

Figure 9. Formula for the ligation reaction of the insert DNA to vector pET28a(+)

Table 27. Ligation reaction composition

Component	20 µl reaction
10X T4 DNA Ligase buffer	2 µl
vector DNA	55.25 ng
insert DNA	55.25 ng
T4 DNA ligase	1 µl
dH_2O	to 20 µl

4.4.5. IPTG expression induction of bacteria (*E. coli*)

Positive colonies were grown in 500 ml of the LB media with 500 µl of 50mg/ml kanamycin under constant shaking 200 rpm overnight at 37 °C. When the OD_{600} reached between 0.4–0.6 then the culture was induced 500µl of 1M IPTG (Isopropyl β-D-1-thiogalactopyranoside). The induction of the cultures was performed in different conditions to optimize the expression (**Table 28.**). After the induction, the cells were centrifuged at 10000g for 10 minutes, the supernatant was discarded, and each pellet was resuspended in 5 ml of ST buffer placed to –80 °C for 20 minutes.

Table 28. Induction conditions for expression optimization

Culture	Time (hours)	Temperature	Shaking
A	4	30 °C	180 rpm
B	6	30 °C	180 rpm
C	15	16 °C	220 rpm
D	15	30 °C	220 rpm
E	15	37 °C	220 rpm

4.4.6. Protein isolation from bacterial periplasm

Bacterial pellets were thawed on ice and 5µl of 0.25 mg/ml lysozyme was added per 5mls of pellet solution and incubated on ice for 30mins.

After the disintegration of the bacterial cell walls by lysozyme, 250µl of 10% sarcosine (N-Lauroylsarcosine) in ST Buffer was added to 5 ml of the pellet. The cell suspension was sonicated (6x 10s) and spun at 15000g for 30 minutes. 5 ml protein sample (supernatant) was diluted 1:4 in lysis buffer to final 25 ml and affinity chromatography using Ni-NTA beads followed by dialysis as described in **4.3.8** was performed.

The concentration of the purified protein was then measured with Pierce Coomasie Bradford Protein Assay kit (Thermo Scientific) following provided protocol and protein was then loaded on SDS-PAGE (**4.3.9)** tested with anti-His-tag Western blot (**4.3.10**) and enzymatic assays for activity (**4.3.11**) and substrate specificity.

4.4.7. Protein isolation from inclusion bodies

Solubilization of bacterial inclusion bodies with guanidium hydrochloride

After the cell lysis as described in **4.4.6,** 5 ml Tris buffer pH 7.8 containing 6M guanidium hydrochloride was added. The suspension was sonicated (5x 30 seconds) on ice, incubated with constant rocking for 72 hours at RT, again sonicated as above and centrifuged at 10000g for 10 minutes. The supernatant was collected and diluted with lysis buffer to 25 ml, and the protein isolation using Ni-NTA agarose beads was performed as in 4.5.9. Flow-through, two wash steps and elution step were then loaded on SDS-PAGE (**4.3.9**) and Western blot with anti-His-tag monoclonal antibodies was performed (**4.3.10**).

Mass spectrometry

Two most distinctive bands from bacterial pellet protein samples formed after SDS-PAGE were precisely cut out of the acrylamide gel, placed to the de-stain solution and sent to mass-spectrometry analysis to confirm the presence of the recombinant protein (Institute of Organic Chemistry and Biochemistry of Czech Academy of Sciences).

Purification of inclusion bodies using mild detergents

When the presence of the recombinant protein in the bacterial pellet was confirmed, inclusion bodies were with the modified protocol described in Khan et al. 1998. Briefly, thawed bacterial pellet was sequentially resuspended in 50 mM Tris-HCl buffer pH 8.5 containing 5 mM EDTA; 5M EDTA and 1% DOC (deoxycholate) and 2% DOC with sonication on ice (8x for 30s) and supernatant removal (11000 rpm for 20 minutes at 4 °C) before every step. In Tris buffer with 2% DOC the pellet was incubated in 37 °C for 24 hours. The pellet was resuspended in 50 mM Tris-HCl buffer pH 8.5, incubated for 30 minutes at RT and spun as above. The supernatant was discarded and1µg of purified inclusion bodies was dissolved in sample buffer and analysed on SDS-PAGE (**4.3.9**).

4.4.8. Production of specific antibodies

Because of the failure to solubilize and purify recombinant protein, two female ICR mice were immunized with protein bands from acrylamide gel corresponding to the recombinant protein (rTrCE). 100 ug of the protein in the form of inclusion bodies was solubilized in sample buffer and SDS-PAGE was performed. Protein bands containing recombinant protein confirmed by mass-spectrometry were precisely excised, unstained and homogenized with a sterilized razor blade followed by injecting through the syringe with a 21G needle. The gel fragments were mixed with 0.5 ml of sterile 0.9% saline solution and injected intraperitoneally. The procedure was repeated 14 days later.

After four weeks, mice were killed, sera were collected according to Šteiger, 2018 and used as primary antibody in Western blot analysis for antibodies specificity in dilution 1:50 – 1:10 000. Western blots were performed with inclusion bodies proteins as an antigen (SDS-PAGE as in **4.3.9**), and with proteins from cercarial homogenate.

To prepare the cercarial homogenate sample, a collection of cercariae was performed as in section **4.1.1.** Concentrated parasites were then repeatedly frozen in –20 °C and thawed and peptidase inhibitor Complete Mini EDTA-free (Roche) solubilized in 1 ml of water was added. The sample was then sonicated on ice four times for 10 seconds. Homogenized cercariae were spun for 20 minutes at 4 °C at 13 000g, the supernatant was collected and ran on SDS-PAGE (**4.3.9**). Resulting acrylamide gel was then used for Western Blot (**4.3.10**).

4.5. qPCR of *T. regenti* cathepsin B2 and cercarial elastase in cercariae and sporocysts

All qPCR experiments were carried out as a part of an internship at John Dalton's Laboratory at Queen's University Belfast under supervision of Dr Krystyna Cwiklinski.

4.5.1. Genes of interest – cathepsin B2 and cercarial elastase

TrCB2 and TrCE were chosen for the quantitative PCR analysis. The sequence of TrCB2 was obtained from (Dolečková et al., 2009) and TrCE sequence was obtained as described above.

4.5.2. Reference genes – PSMD, TPC2L and NDUFV2

Suitable reference genes most stably expressed across the life cycle were chosen. The selection of the *T. regenti* orthologue genes PSMD (26S proteasome non-ATPase regulatory subunit 4), TPC2L (Trafficking protein particle complex subunit 2-like protein) and NDUFV2 (NADH dehydrogenase flavoprotein 2) was based on the genome-wide microarray expression analysis of *S. japonicum* (S. Liu et al., 2012) (see **Table 47.**, **Table 48.**, and **Table 49.** in **10.2.**).

4.5.3. qPCR primers design

Primers for all genes were designed to amplify 100–105 bases long amplicons to assure, that the same amount of SYBR® Green PCR Master Mix (Applied Biosystems) is bound to amplified DNA segment. Also, the primer melting temperatures were considered to assure, that all genes are amplified within the single run (**Table 29.**) (Bustin and Huggett, 2017; Debode et al., 2017).

4.5.4. cDNA for qPCR preparation

RNA of both stages was isolated by TRIzol reagent as described above. cDNA synthesis was carried out with 1 µg of mRNA using oligo(dT)$_{18}$ primer and High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) following the manufacturer's instructions.

4.5.5. qPCR

After confirmation that designed primers are efficient and specific with conventional PCR (PCR as in **4.2.4**) and do not form primer dimers, the qPCR reactions were set **(Table 30.)**. Reaction mixes were prepared in volume for 4 reactions (80 µl) per gene and 3 reactions per gene were transferred to sterile qPCR tubes in and placed to Rotor-Gene Q (QIAGEN) (**Table 31**).

4.5.6. Data evaluation

The threshold line which is the point at which a reaction reaches a fluorescent intensity above background levels and where the exponential phase of the reaction began was set. The resulting three C_T values of technical replicates (the PCR cycle numbers at which sample's reaction curve intersects the threshold line) for each gene were averaged and its variance calculated if it is under 0.5 (Hellemans et al., 2007).

4.5.7. Comparison of gene expression between TrCB2 and TrCE in both stages

To determine the relative differential expression pattern for 2 genes of interest **in a single stage**, the averaged C_T values of the genes were compared using the Livak method (Livak and Schmittgen, 2001):

4.5.8. Comparison of relative gene expression between sporocyst and cercariae

Comparison of expression levels of the genes of interest **between these two stages** was performed by an alternative method combining several methodical publications. Variances of the measured C_T values for reference genes between the stages were averaged (average variance must be under 0.5) and subtracted from the sporocysts C_T values to compensate the difference in input RNA into reverse transcription. Resulting C_T values of the genes of interest for both stages were normalized using geometric averaging of reference genes. Normalized expression values were calculated and compared (Vandesompele et al. 2002; Hellemans et al. 2007; S. Liu et al. 2012).

Table 29. Designed qPCR primers to compare the expression of reference genes and genes of interest

PSMD4			
	Sequence	**Length**	**(^{Tm} °C)**
Forward primer	GCACCATCGACAACAACAAC	20	58.25
Reverse primer	ACACCAGGGAGACTTTGTAGA	21	58.03
Product size	101 bp		
NDUFV2			
Direction	**Sequence**	**Length**	**(T_m °C)**
Forward primer	CACAGAGTGGACAAGGTGG	19	57.69
Reverse primer	TCCGAACGAACTTTGAATCCA	21	57.89
Product size	100 bp		
TPC2L			
Direction	**Sequence**	**Length**	**(T_m °C)**
Forward primer	CCATTCTATAAATCAGGCACACC	23	57.30
Reverse primer	ACCAGAACTAGCATGTGTAGC	21	57.75
Product size	105 bp		
TrCE			
Direction	**Sequence**	**Length**	**(T_m °C)**
Forward primer	CATTGTGTGTGTGGACCGAA	20	58.70
Reverse primer	CCACCTTGATTGATGTCGCT	20	58.26
Product size	103 bp		
TrCB2			
Direction	**Sequence**	**Length**	**(T_m °C)**
Forward primer	AGCCTTGATGAATTGCCGAA	20	58.16
Reverse primer	GCCCAACAAGAACCACATGA	20	58.67
Product size	104 bp		

Table 30. Composition of qPCR reactions

Component	20 µl reaction
SYBR® Green PCR Master Mix	10 µl
forward primer (10mM)	2 µl
reverse primer (10mM)	2 µl
cDNA (from 1 µg of total RNA)	1 µl
dH_2O	5 µl

Table 31. Programme of the qPCR thermal cycler

Step	Reaction	Temperature	Time
1	Initial denaturation	94 °C	120s
2	Denaturation	94 °C	15s
3	Primer annealing, extension, and fluorescence read	54 °C	60s
Step 2–3 repeated 40x			

5. Results

5.1. Collection of the cercarial ESPs and its characterization

Cercarial ESPs samples were collected with two different methods – linoleic acid stimulation and skin of the host stimulation.

5.1.1. Collection of ESPs with linoleic acid as the native stimulus

The cercarial ESPs was successfully collected using linoleic acid. The concentration of the linoleic acid that reliably induces a transformation of the larvae was determined experimentally to be 10 µl/ml. Subsequent glands staining and microscopical examination confirmed, that the larvae have undergone the transformation and thus released excretory-secretory products (**Figure 10.**).

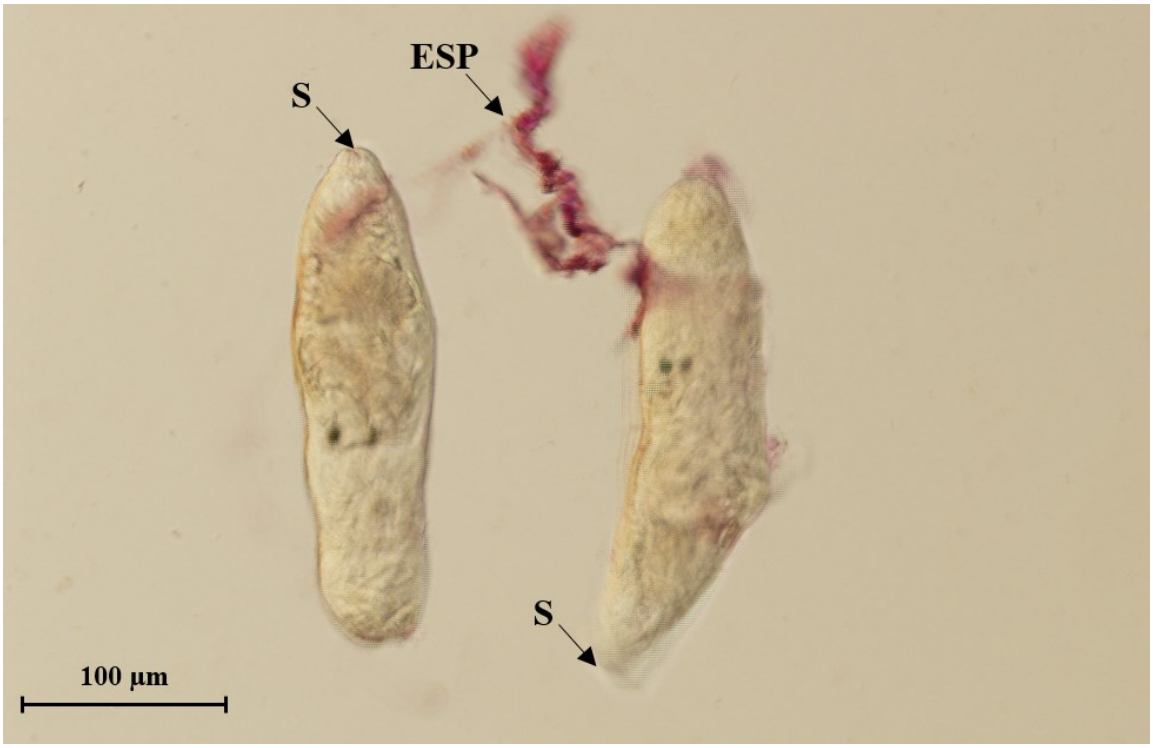

Figure 10. Transformed cercariae using linoleic acid. ESPs – excretory-secretory products, **S – "scar"** after tail abruption.

5.1.2. Collection of ESPs using the skin of the host

Excretory-secretory products were collected using the skin of the definitive host. Histological examination of duck skin after the collection of ESPs with subsequent hematoxylin-eosin staining confirmed the functionality of the collection method. Cercariae shed their tails, penetrated the duck skin and migrated through the skin tissue (see **Figure 11.**).

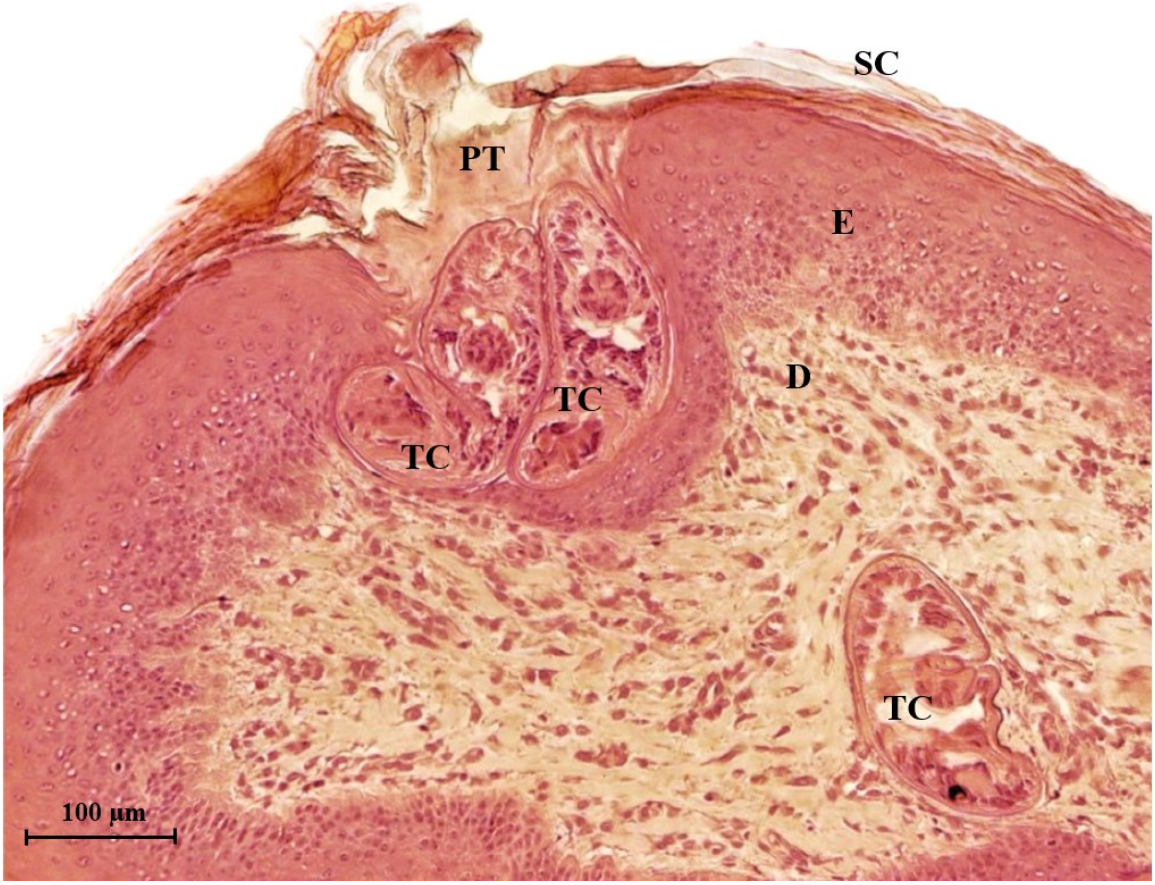

Figure 11. Transformed cercariae migrating in the host skin. PT - penetration tunnel, **SC** - *stratum corneum*, **E**- *epidermis*. **D** – *dermis*, **TC** – transformed cercaria

5.1.3. Mass-spectrometry and ESPs protein identifications

Mass-spectrometry and subsequent data analysis using genome and transcriptome of the parasite identified 163 proteins in ESPs collected using the skin of the host and 72 proteins in ESPs collected using linoleic acid. 118 of identified proteins were exclusive for "skin" samples, 27 for "linol" samples and 45 were present in both samples (**Figure 12.**).

BLASTp search (Altschul et al., 1990) using MEROPS 12.0 database (Rawlings et al., 2018) of the protein identifications found 62 and 29 annotated peptidases, unassigned peptidases or non-peptidase orthologues in "skin" and "linol" samples respectively. 48 were exclusive for skin samples, 14 for linol samples and 15 present in both groups (**Figure 13., Table 32., Table 33.** and **Table 34.**.).

Also, 5 most abundant peptidases, unassigned peptidases or non-peptidase orthologues from both samples were determined (**Table 35.** and **Table 36.**).

Two different cercarial elastases were found in skin induced samples, but no cathepsins were identified by mass spectrometry. The more abundant cercarial elastase (A0A183VJ85) was therefore chosen for further characterization.

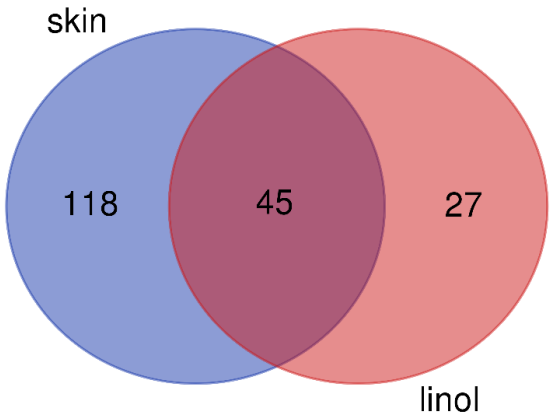

Figure 12. Numbers of identified proteins using two different methods of stimulation. skin (blue) – proteins identified in samples using skin of the host (blue) as a stimulant, **linol (red)**- proteins identified using linoleic acid as a stimulant.

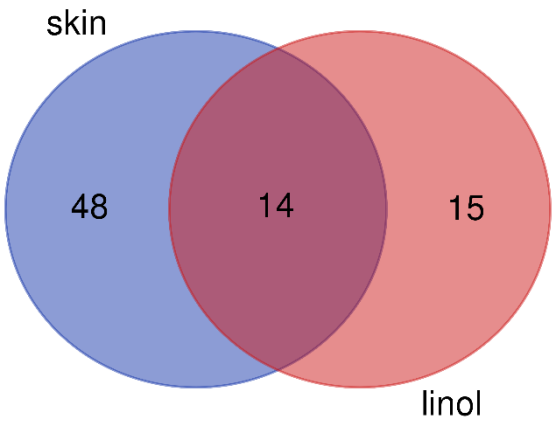

Figure 13. Numbers of identified peptidases, unassigned peptidases or non-peptidase orthologues using two different methods of stimulation. skin (blue)– MEROPS identifications in samples using skin of the host (blue) as a stimulant, **linol (red)**- MEROPS identifications using linoleic acid as a stimulant.

Table 32. Peptidases, unassigned peptidases or non-peptidase orthologues common for both ESPs samples. UniProt accession number is for the first BLASTp database hit. Protein identity is based on the most similar characterised/named protein based on E-value.

UniProt accession no.	Identified protein (organism)	Sequence similarity %	E-value
A0A183WRE6	Tubulin beta chain (*Trichobilharzia regenti*)	100.00	0,00E+00
G4VCN8	Dihydrolipoyl dehydrogenase (*Schistosoma mansoni*)	90.14	0,00E+00
A0A183VWJ5	Fructose-bisphosphate aldolase (Bilateria)	100.00	0,00E+00
A0A3Q0KC41	Heat shock protein HSP60, putative (*Schistosoma mansoni*)	93.00	0,00E+00
P53471	Actin (Platyhelminthes)	100.00	0,00E+00
P20287	Glyceraldehyde-3-phosphate dehydrogenase 3 (*Schistosoma mansoni*)	87.57	0,00E+00
A0A183W3Z6	Pyruvate kinase (*Trichobilharzia regenti*)	100.00	0,00E+00
A0A094ZPJ2	Citrate synthase (*Schistosoma haematobium*)	93.53	0,00E+00
A0A183WAU9	ATP synthase subunit beta (*Trichobilharzia regenti*)	100.00	0,00E+00
A0A183WKF3	Peptidyl-prolyl cis-trans isomerase (*Trichobilharzia regenti*)	100.00	1,00E−122
G4VD44	Tetraspanin (*Schistosoma mansoni*)	78.11	1,00E−88
A0A094ZX17	GLIPR1-like protein 1 (Fragment) (*Schistosoma haematobium*)	62.86	1,00E−79
A0A183WKA9	Thioredoxin (*Trichobilharzia regenti*)	100.00	4,00E−71
A0A183VTW0	Venom allergen-like (VAL) 10 protein (*Schistosoma mansoni*)	47.20	2,00E−32
A0A183WLL2	Glutathione S-transferase 26kDa (*Schistosoma mekongi*)	78.57	9,00E−30

Table 33. Peptidases, unassigned peptidases or non-peptidase orthologues identified exclusively in skin induced ESPs samples. UniProt accession number is for the first BLASTp database hit. Protein identity is based on the most similar characterised/named protein based on E-value.

Uniprot accession number	Identified protein (organism)	Sequence similarity %	E-value
A0A183VM44	Myosin heavy chain (*Schistosoma mansoni*)	82.88	0,00E+00
A0A183VN33	Tubulin alpha chain (*Trichobilharzia regenti*)	100.00	0,00E+00
A0A183VQR7	Prolyl endopeptidase (*Schistosoma haematobium*)	71.67	0,00E+00
A0A183VUW0	Pyruvate dehydrogenase E1 component subunit alpha (*Trichobilharzia regenti*)	100.00	0,00E+00
A0A183WY32	Elongation factor 1-alpha (*Trichobilharzia regenti*)	100.00	0,00E+00
A0A3Q0KGR3	Aconitate hydratase, mitochondrial (S*chistosoma mansoni*)	88.66	0,00E+00
A0A3Q0KB95	Filamin (*Schistosoma mansoni*)	71.12	0,00E+00
C1LGM4	Asparaginyl-tRNA synthetase (*Schistosoma japonicum*)	75.82	0,00E+00
A0A094ZFT4	Glycogen debranching enzyme (Fragment) (*Schistosoma haematobium*)	80.63	0,00E+00
A0A183VSN1	L-lactate dehydrogenase (*Trichobilharzia regenti*)	100.00	0,00E+00
C1L502	Protein disulfide-isomerase (*Schistosoma japonicum*)	73.97	0,00E+00
A0A183WCE7	20 kDa calcium-binding protein (*Schistosoma haematobium*)	61.24	0,00E+00
G4V5G7	Glycogenin-related (*Schistosoma mansoni*)	81.25	0,00E+00
A0A183VVW7	Camp-dependent protein kinase type II-alpha regulatory subunit, (*Schistosoma mansoni*)	90.45	0,00E+00
A0A3Q0KL35	Calponin (*Schistosoma mansoni*)	95.01	0,00E+00
Q5DGY9	Heat shock 70kDa protein 9B (*Schistosoma japonicum*)	93.90	0,00E+00
A0A183VJ54	Phosphopyruvate hydratase (*Fasciola hepatica*)	76.79	0,00E+00
A0A183X423	Invadolysin (M08 family) (*Schistosoma mansoni*)	69.09	4,00E–168
G4V6B9	Triosephosphate isomerase (*Schistosoma mansoni*)	85.66	2,00E–159
C1LNQ1	Phosphoglycerate mutase (*Schistosoma japonicum*)	93.84	1,00E–146
A0A3P8LUE3	Malate dehydrogenase (*Trichobilharzia regenti*)	99.49	2,00E–139
A0A183WD86	Phosphoglycerate kinase (*Trichobilharzia regenti*)	100.00	9,00E–138

Q86F28	BAR,domain-containing protein (*Schistosoma japonicum*)	73.81	2,00E–136
A0A183WK30	Aspartate aminotransferase (*Trichobilharzia regenti*)	100.00	7,00E–134
Q6IV29	Transgelin (Platyhelminthes)	96.84	1,00E–132
A0A183WFN3	Uncharacterized protein (*Trichobilharzia regenti*)	100.00	2,00E–126
A0A183VJ85	Elastase 2b (*Schistosoma japonicum*)	65.02	9,00E–125
A0A2H1BUB8	ADP-ribosylation factor family protein (*Fasciola hepatica*)	87.29	1,00E–113
A0A094ZKP5	Adenylate kinase (*Schistosoma haematobium*)	80.20	1,00E–112
A0A183VU87	Rab-protein 10 (*Schistosoma japonicum*)	78.71	5,00E–109
V9V935	Scribble protein 3 (*Schistosoma japonicum*)	77.57	7,00E–109
A0A183WV85	Superoxide dismutase [Cu-Zn] (*Trichobilharzia regenti*)	100.00	4,00E–104
A0A183WVP1	Putative rap1 and (*Schistosoma mansoni*)	92.95	1,00E–101
A0A183VIP8	Uncharacterized protein (*Trichobilharzia regenti*)	97.79	5,00E–92
A0A183WZA9	GTP-binding nuclear protein (*Trichobilharzia regenti*)	100.00	1,00E–90
A0A183WBE5	Glutaminyl cyclase (M28 family) (*Schistosoma mansoni*)	61.26	9,00E–80
A0A183WIG8	Elastase 2b (*Schistosoma japonicum*) (not overlapping A0A183VJ85)	60.89	2,00E–72
A0A183VIV0	14-3-3 epsilon (*Schistosoma mansoni*)	84.55	4,00E–63
A0A183WP37	Aldo-keto reductase family 1, member B4 (*Schistosoma japonicum*)	75.19	2,00E–60
A0A183VVA1	Dipeptidyl peptidase 2 (*Schistosoma haematobium*)	42.71	9,00E–60
A0A183VZL0	Putative rho GTPase (*Schistosoma mansoni*)	95.56	4,00E–55
A0A074ZWL6	LIM domain protein (*Opisthorchis viverrini*)	89.89	2,00E–50
A0A183WKD6	Leucine aminopeptidase (*Schistosoma japonicum*)	79.55	5,00E–40
A0A183VPZ6	SJCHGC00739 protein (*Schistosoma japonicum*)	92.75	5,00E–39
A0A183WA20	Putative ras GTP exchange factor (*Schistosoma mansoni*)	75.00	1,00E–24
A0A183X423	Leishmanolysin-like peptidase (*Schistosoma haematobium*)	49.44	2,00E–20
A0A183WJL0	Acyl CoA binding protein (*Necator americanus*)	69.81	5,00E–20

Table 34. Peptidases, unassigned peptidases or non-peptidase orthologues identified exclusively in linoleic acid-induced ESPs samples. UniProt accession number is for the first BLASTp database hit. Protein identity is based on the most similar characterised/named protein based on E-value.

UniProt accession no.	Identified protein (organism)	Sequence similarity %	E-value
A0A183WCV9	Serine/threonine kinase (*Schistosoma mansoni*)	45.42	0,00E+00
A0A183X686	90 kDa heat shock protein (*Dugesia japonica*)	80.88	0,00E+00
Q5DDV5	Enolase (*Schistosoma japonicum*) (not overlapping A0A183VJ54)	89.63	0,00E+00
X2KKK0	Heat shock protein 70 (*Hyriopsis cumingii*)	89.33	0,00E+00
A0A3Q0KE68	Putative peptide n-glycanase (Pngase) (*Schistosoma mansoni*)	66.59	0,00E+00
A0A183W6L4	Triosephosphate isomerase (*Trichobilharzia regenti*)	100.00	3,00E–174
A0A183WYB0	Tubulin alpha chain (*Trichobilharzia regenti*)	100.00	6,00E–142
G4VAL2	Putative leucyl-tRNA synthetase (*Schistosoma mansoni*)	77.50	2,00E–115
A0A183VRC9	Calmodulin-alpha isoform X1 (*Penaeus vannamei*)	92.99	6,00E–96
A0A095B1V8	Protein farnesyltransferase/geranylgeranyltransferase type-1 subunit alpha (*Schistosoma haematobium*)	57.87	5,00E–93
A0A183VWA0	Phosphoglycerate kinase (*Trichobilharzia regenti*)	100.00	6,00E–93
A0A183VPN2	Histone H2A (*Trichobilharzia regenti*)	100.00	1,00E–80
A0A183W516	Histone H2B (*Trichobilharzia regenti*)	100.00	1,00E–79
A0A183VM96	Phosphoglycerate mutase (*Hymenolepis microstoma*)	53.04	6,00E–76

Table 35. 5 most abundant peptidases, unassigned peptidases or non-peptidase orthologues identified skin induced ESPs samples. UniProt accession number is for the first BLASTp database hit. Protein identity is based on the most similar characterised/named protein based on E-value. **RA** – relative abundance

RA	UniProt accession no.	Identified protein (organism)	Sequence similarity %	E-value
1.	A0A183VJ54	Phosphopyruvate hydratase (*Fasciola hepatica*)	76.79	0,00E+00
2.	A0A183VWJ5	Fructose-bisphosphate aldolase (Bilateria)	100.00	0,00E+00
3.	A0A183WKA9	Thioredoxin (*Trichobilharzia regenti*)	100.00	4,00E–71
4.	A0A183VJ85	Elastase 2b (*Schistosoma japonicum*)	65.02	9,00E–125
5.	A0A183VTW0	Venom allergen-like (VAL) 10 protein (*Schistosoma mansoni*)	47.20	2,00E–32

Table 36. 5 most abundant peptidases, unassigned peptidases or non-peptidase orthologues identified skin induced ESPs samples. UniProt accession number is for the first BLASTp database hit. Protein identity is based on the most similar characterised/named protein based on E-value. **RA –** relative abundance

RA	UniProt accession no.	Identified protein (organism)	Sequence similarity %	E-value
1.	A0A183WKA9	Thioredoxin (*Trichobilharzia regenti*)	100.00	4,00E–71
2.	A0A183WKF3	Peptidyl-prolyl cis-trans isomerase (*Trichobilharzia regenti*)	100.00	1,00E–122
3.	A0A183WRE6	Tubulin beta chain (*Trichobilharzia regenti*)	100.00	0,00E+00
4.	A0A183VWA0	Phosphoglycerate kinase (*Trichobilharzia regenti*)	100.00	6,00E–93
5.	A0A183WLL2	Glutathione S-transferase 26kDa (*Schistosoma mekongi*)	78.57	9,00E–30

5.2. Completion of *T. regenti* cercarial elastase whole gene sequence

5.2.1. RNA isolation and reverse transcription

´ The total RNA of the sporocysts, cercariae and hepatopancreas of the healthy snail was successfully isolated using TRIzol reagent and cDNA was generated by reverse transcription. The measured concentrations of the samples were sufficient, and 260/280 and 260/230 values were normal (**Table 37.** and **Table 38.**).

Table 37. Results of the RNA isolation

Sample	RNA concentration (µg/µl)	260/280 value	260/230 value
Infected snail 1	3,60	1,83	1,87
Healthy snail	3,87	1,87	2,01
Cercariae	1,20	2,04	1,85

Table 38. Results of the reverse transcription

Sample	cDNA concentration (µg/µl)	260/280 value	260/230 value
Infected snail	2,31	1,86	2,16
Healthy snail	2,01	1,83	2,07
Cercariae	2,15	1,99	1,92

5.2.2. PCR, colony PCR screening

Conventional PCR followed by agarose DNA electrophoresis has confirmed, that the gene coding cercarial elastase is transcribed either in sporocysts and cercariae (**Figure 14A.**). Hepatopancreas from healthy snail served as a control to confirm, that designed primers are specific only to *T. regenti* cercarial elastase gene.

PCR colony screening confirmed the successful transformation of the TOP10 cells with pGEM®-T Easy Vector containing a ligated insert. Sequencing of the isolated plasmid DNA elucidated a real partial sequence of the gene; thus, original genomic sequence was corrected accordingly (**Figure 14B**).

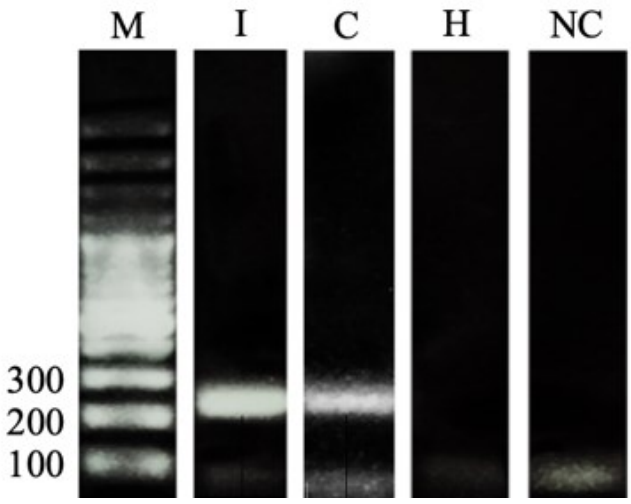

Figure 14A. Agarose gel after PCR for amplification of partial TrCE gene sequence. M - marker GeneRuler Ladder 100 bp, I – infected snail - sporocysts,

```
TrCE_PCR_produkt      ATGATTCCTACTTCGTCG--CTGCTCTTCTCATTGTTTCTGCTATGCAATCTAGTCTCTC  58
TrCE_partial_genomic  ATGATTCCTACTTCGTCGCTGCGCTCTTCTCATTGTTACTGCTATGCAATCTAGTCTCTC  60
                      ******************  *************** **********************

TrCE_PCR_produkt      TGAGTGTATCAACATGGTTAGTTCGTAGTGGAGAGACAGTGAGAGATAAGAATGAATTCC  118
TrCE_partial_genomic  TGAGTGTATCAACATGGTTAGTTCGTAGTGGAGAGACAGTGAGAGATAAGAATGAATTCC  120
                      ************************************************************

TrCE_PCR_produkt      CGTTTCTCGCTCTCATGGTGACAGACAGTTCGATGTGTACAGCGACGCTGGTGTCCACCC  178
TrCE_partial_genomic  CGTTTCTCGCTCTCATGGTGACAGACAGTTCGATGTGTACAGCGACGCTGGTGTCCACCC  180
                      ************************************************************

TrCE_PCR_produkt      GAGCTGTGATCACAGCTGGTCATTGTGTGTGTGGACCGAAATCAATCAACCGGGT      232
TrCE_partial_genomic  GAGCTGTGATCACAGCTGGTCATTGTGTGTGTGGACCGAAATCAATCAACCGGGT      234
                      ********************************************************
```

Figure 14B. Sequence alignment of original partial genomic sequence and sequence of product amplified by PCR. Red circle – two extra bases in genomic sequence

5.2.3. Rapid Amplification of cDNA Ends (RACE) of cercarial elastase

After a series of optimization steps, the amplification of the gene was finally successful using Superscript III reverse transcriptase in RACE cDNA preparation and Primestar MAX master mix. 3' ends were amplified with PCR at all three tested temperatures (**Figure 15.)**

Amplicons were successfully cloned into the pGEM Easy Vector and the plasmid DNAs were reproduced in TOP10 cells. Sequencing of the plasmid DNA confirmed, that the RACE PCR successfully amplified 3'end of the gene.

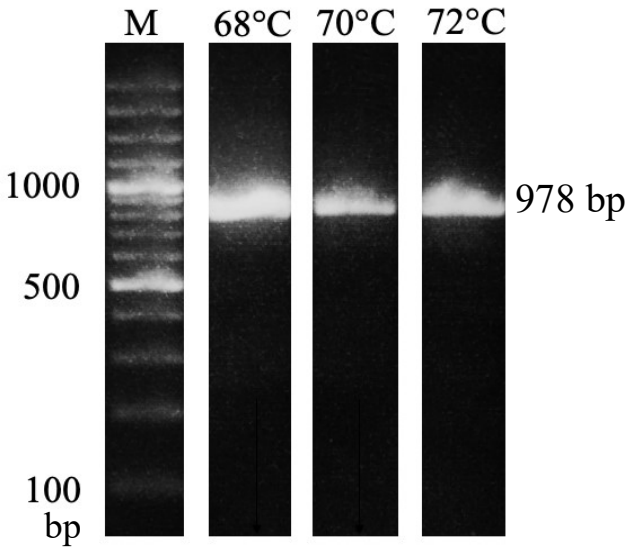

Figure 15. 3'RACE PCR results. M - marker GeneRuler Ladder 100 bp, **68°C, 70°C, 72°C** - annealing temperatures

5.2.4. TrCE gene sequence assembly

After the whole gene sequence was assembled, protein´s primary structure, size and the isoelectric point were predicted (**Table 39.**). Using MEROPS 12.0 database (Rawlings et al., 2018) it was also found out from the sequence, that the enzyme contains complete catalytic triad of histidine, aspartate, and serine typical for serine peptidases (Ingram et al., 2012).

Table 39. Complete nucleotide and amino acid sequence of *T. regenti* cercarial elastase. Catalytic triad (active site) highlighted green

***Trichobilharzia regenti* cercarial elastase complete nucleotide sequence – 792 bp**	
ATGATTCCTACTTCGTCGCTGCTCTTCTCATTGTTACTGCTATGCAATCTAGTCTCTCTGAG TGTATCAACATGGTTAGTTCGTAGTGGAGAGACAGTGAGAGATAAGAATGAATTCCCGTT TCTCGCTCTCATGGTGACAGACAGTTCGATGTGTACAGCGACGCTGGTGTCCACCCGAGC TGTGATCACAGCTGGTCATTGTGTGTGTGGACCGAAATCAATCAACCGGATCTCCTTCCAG ACACTGAGTGATTTCGACAAGCGGTCAGTGAACTATCTAGCGACATCAATCAAGGTGGCT CCAGAATATGATCCAGTGTGTCAATTGAAACGAGCCAGACGGCGTACGCGACAGTCATTC GGTGGTTATGATATGGCGATTGTGACATTGTCGGGGCTAGTGGATTTGTCGAGTGGAGTTC GAGTGATCAGTTTAGCGACGGCATCGGATATTCCTGTTCCGAATAGTATTGTTTTCATTGT TGGCTATGGGAAGGATTCGAAGGCGACGGATGATACGCGTAAATATGGTGGTGTATTGAA AAAGGGTCGTGCCACTGTAATGAAATGTATGCACAAAGTTGTTGGAAATCCAATCTGTAT CCGACCTGGTCCTGTTTCAGCACAAATTGTTGGTCCAGGTGATAGTGGTGGACCACTCCTG TTGACGCCTCAAGGACCGATTATTGGTGTCGCTTCAAATGGAGTGTTCCTTCCTGCTATTA ATGATTTGACTGTGGAATATGCAAGTGTTAGCAGGTCACTGGAATTCATTATGTCTAATAT CTGA	
Complete amino acid sequence 5'3' – 263 aa	**Theoretical pI/Mw: 9.33 / 28.09 kDa**
MIPTSSLLFSLLLLCNLVSLSVSTWLVRSGETVRDKNEFPFLALMVTDSSMCTATLVSTRAVIT AGHCVCGPKSINRISFQTLSDFDKRSVNYLATSIKVAPEYDPVCQLKRARRRTRQSFGGYDMA IVTLSGLVDLSSGVRVISLATASDIPVPNSIVFIVGYGKDSKATDDTRKYGGVLKKGRATVMK CMHKVVGNPICIRPGPVSAQIVGPGDSGGPLLLTPQGPIIGVASNGVFLPAINDLTVEYASVSRS LEFIMSNI-	

5.3. *T. regenti* cercarial elastase recombinant protein expression in yeast

The PichiaPink yeast expression system was used to produce rTrCE. According to the NetNGlyc 1.0 software output, no glycosylation sites were found in the amino acid sequence of TrCE. The cleavage site of the signal peptide was predicted with SignalP 4.0 between 28 and 29 amino acid and removed. Designed gene was then synthesised by Eurofins company (**Table 40.**) and inserted into the pPINKα-HC vector

Table 40. The final sequence used for expression in yeast

The final sequence of the gene for yeast expression – 759 bp
GAGTCTACATGGTTAGTTCGTAGTGGAGAGACAGTGAGAGATAAGAATGAATTCCCGTTT CTCGCTCTCATGGTGACAGACAGTTCGATGTGTACAGCGACGCTGGTGTCCACCCGAGCT GTGATCACAGCTGGTCATTGTGTGTGTGGACCGAAATCAATCAACCGGATCTCCTTCCAG ACACTGAGTGATTTCGACAAGCGGTCAGTGAACTATCTAGCGACATCAATCAAGGTGGCT CCAGAATATGATCCAGTGTGTCAATTGAAACGAGCCAGACGGCGTACGCGACAGTCATTC GGTGGTTATGATATGGCGATTGTGACATTGTCGGGGCTAGTGGATTTGTCGAGTGGAGTTC GAGTGATCAGTTTAGCGACGGCATCGGATATTCCTGTTCCGAATAGTATTGTTTTCATTGT TGGCTATGGGAAGGATTCGAAGGCGACGGATGATACGCGTAAATATGGTGGTGTATTGAA AAAGGGTCGTGCCACTGTAATGAAATGTATGCACAAAGTTGTTGGAAATCCAATCTGTAT CCGACCTGGTCCTGTTTCAGCACAAATTGTTGGTCCAGGTGATAGTGGTGGACCACTCCTG TTGACGCCTCAAGGACCGATTATTGGTGTCGCTTCAAATGGAGTGTTCCTTCCTGCTATTA ATGATTTGACTGTGGAATATGCAAGTGTTAGCAGGTCACTGGAATTCATTATGTCTAATAT C**GGACCA**CATCATCACCATCATCAC**TAG**GCCGGC
Restriction sites **GP coding site** **His-tag** **Stop codon**

5.3.1. Transformation of the yeast cells and PCR colony screening

After the successful transformation of the TOP10 cells, the concentrations of the extracted plasmid DNAs using PureLink Quick Plasmid Miniprep Kit were measured (**Table 41.**). The plasmid DNA was successfully digested with *SpeI* and yeast PichiaPink cells were transformed. The success of the transformation was confirmed not only by the appearance of hundreds of white colonies on the selection plates but also with PCR colony screening using the microwave (**Figure 16.**). Subsequent sequencing of the resulting amplicons also showed 100% match with originally designed gene sequence.

Table 41. Concentrations of the plasmid DNA after cloning

Sample	cDNA concentration (µg/µl)
TOP10 col. 5	2,82
TOP10 col. 6	3,03

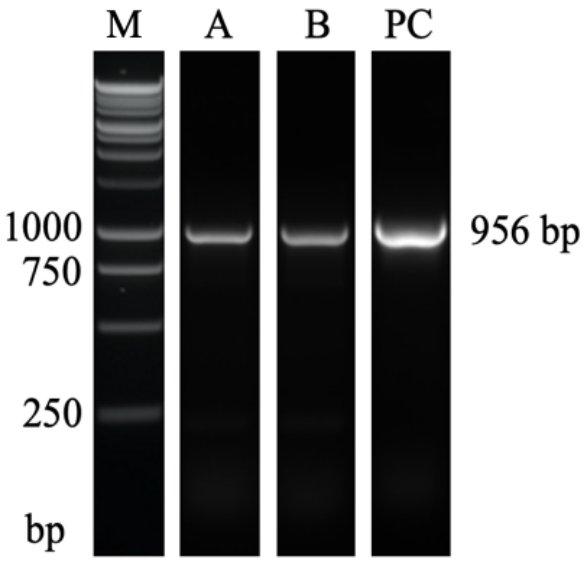

Figure 16. Yeast cells post-transformational PCR screening. M - marker Promega 1kb DNA Ladder, **A** – colony **5, B** - colony **6, PC** – positive control vector DNA

5.3.2. Large expression screening – "Yeastern-blot"

A total of 231 colonies (154 showed on the **Figure 17.**) were analysed for the presence of secreted recombinant protein. None of the tested colonies showed significant similarity in the strength of the signal on the membrane in comparison to the positive control FhCL3. Despite this apparent lack of expression in tested colonies, colonies with relatively strongest signal were selected for expression experiments.

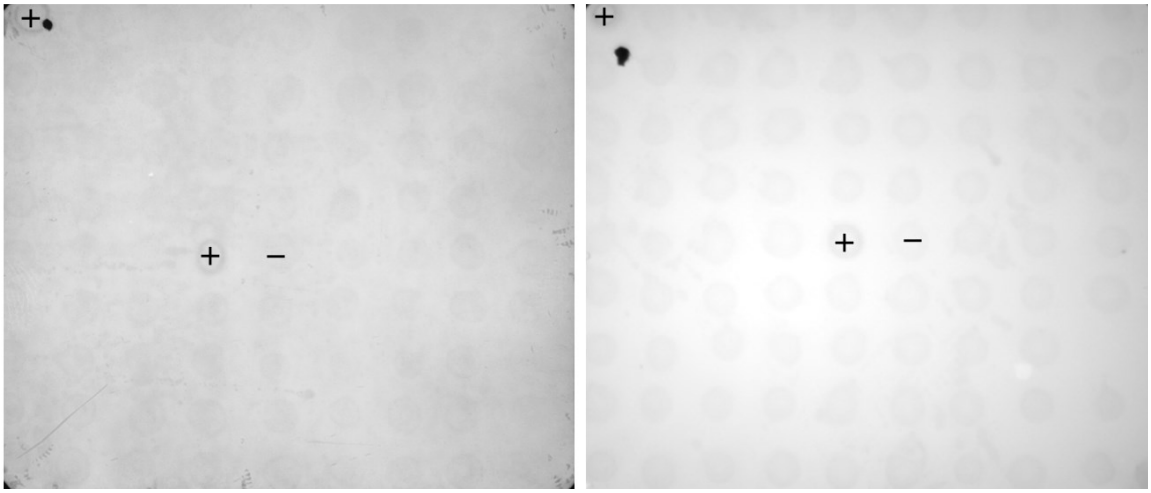

Figure 17. Results of "Yeastern-blots" – wide expression colony screening. + – positive control **(FhCL3)** – – negative control (uninduced colonies)

5.3.3. Expression in the yeast system

Due to failure to detect recombinant protein by small-scale expression experiments, large-scale expression experiments were performed to increase the final amount of produced protein and thus the chance to detect it and purify it.

Despite testing a large number of positive colonies of both PichiaPink yeast strains #1 and #4, different temperatures of induction and different dynamics of methanol feeding, SDS-PAGE analyses have not shown any protein bands corresponding to the size of recombinant cercarial elastase. More sensitive detection method, Western blot, using anti-his tag monoclonal antibodies with *Fasciola hepatica* cystatin (FhCys) as a positive control also have not confirmed a presence of desired recombinant protein.

Finally, the enzymatic assays were performed, but none on the test showed the enzymatic activity. From all the results above was concluded, that the protein expression in the yeast system was not successful despite numerous ways of troubleshooting.

5.4. *T. regenti* cercarial elastase recombinant protein expression in bacteria

Vector pET 28a(+) and *E. coli* BL21 competent cells were used for protein expression in a bacterial system. Firstly, the expression gene for insertion to vector was prepared.

5.4.1. Expression gene design

Analyses with REBsites and NEBCutter have confirmed, that none of the chosen endonucleases cleaves inside the sequence of the gene. PCR with primers designed for the amplification of TrCE gene was successfully performed. The right primary structure of obtained TrCE amplicon was confirmed by sequencing (**Figure 18.**) (gene sequence in **Table 42.**).

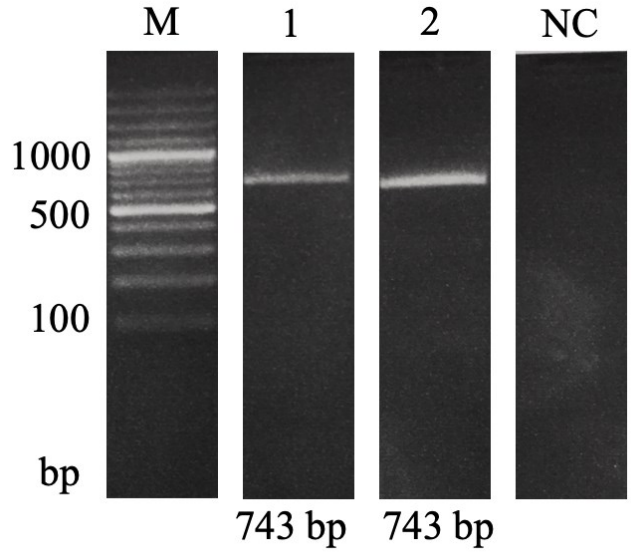

Figure 18. PCR to produce and amplify TrCE gene for expression. M – marker GeneRuler Ladder 100bp, **1,2** – amplified gene contruct. **NC** – negative control

Table 42. The final sequence used for expression in bacteria

The final sequence of the gene for bacterial expression – 792 bp	
ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCAT ATGGCTAGCACATGGTTAGTTCGTAGTGGAGAGACAGTGAGAGATAAGAATGAATTCCCG TTTCTCGCTCTCATGGTGACAGACAGTTCGATGTGTACAGCGACGCTGGTGTCCACCCGAG CTGTGATCACAGCTGGTCATTGTGTGTGTGGACCGAAATCAATCAACCGGATCTCCTTCCA GACACTGAGTGATTTCGACAAGCGGTCAGTGAACTATCTAGCGACATCAATCAAGGTGGC TCCAGAATATGATCCAGTGTGTCAATTGAAACGAGCCAGACGGCGTACGCGACAGTCATT CGGTGGTTATGATATGGCGATTGTGACATTGTCGGGGCTAGTGGATTTGTCGAGTGGAGTT CGAGTGATCAGTTTAGCGACGGCATCGGATATTCCTGTTCCGAATAGTATTGTTTTCATTG TTGGCTATGGGAAGGATTCGAAGGCGACGGATGATACGCGTAAATATGGTGGTGTATTGA AAAAGGGTCGTGCCACTGTAATGAAATGTATGCACAAAGTTGTTGGAAATCCAATCTGTA TCCGACCTGGTCCTGTTTCAGCACAAATTGTTGGTCCAGGTGATAGTGGTGGACCACTCCT GTTGACGCCTCAAGGACCGATTATTGGTGTCGCTTCAAATGGAGTGTTCCTTCCTGCTATT AATGATTTGACTGTGGAATATGCAAGTGTTAGCAGGTCACTGGAATTCATTATGTCTAATA TCTGA	
Amino acid sequence of desired protein – 264 aa	**Theoretical pI/Mw: 9.51/ 28,11kDa**
MGSSHHHHHHSSGLVPRGSHMASTWLVRSGETVRDKNEFPFLALMVTDSSMCTATLVSTRA VITAGHCVCGPKSINRISFQTLSDFDKRSVNYLATSIKVAPEYDPVCQLKRARRRTRQSFGGYD MAIVTLSGLVDLSSGVRVISLATASDIPVPNSIVFIVGYGKDSKATDDTRKYGGVLKKGRATV MKCMHKVVGNPICIRPGPVSAQIVGPGDSGGPLLLTPQGPIIGVASNGVFLPAINDLTVEYASV SRSLEFIMSNI–	
Start **His-tag** **Stop**	

5.4.2. Double digestion, ligation and transformation

After the A-tailing, ligation to the pGem Easy Vector and transformation of the TOP10 cells, the double digestion was performed. Insert DNA purified with 1% agarose gel was then successfully ligated to the expression vector pET 28a(+) and BL21 cells were transformed.

The success of the ligation and transformation of the BL21 cells was verified by CRACK, followed by DNA isolation from the gel and sequencing. Sequence accuracy was also re-verified by isolating plasmid DNA using a HybridQ MiniPrep kit and re-sequencing.

5.4.3. Expression in a bacterial system

After testing the five different induction conditions listed above, colonies B and E (**Table 28.**) showed signs of recombinant protein production on SDS-PAGE. Colony B and colony E where the rTrCE seemed to be produced to bacterial periplasm and inclusion bodies respectively were identified and used for further processing.

5.4.4. Protein isolation from bacterial periplasm

After Ni-NTA beads affinity chromatography of the protein sample obtained by the disintegration of the bacterial cells, flow-through, wash steps and elution fractions were analysed using SDS-PAGE. The protein band corresponding to the rTrCE of size 28kDa was identified in elution fraction (**Figure 19.**).

By further adoption of Western blot, the presence of the reaction with anti-His-tag antibody was not observed (**Figure 19.**). Neither the enzymatic essays revealed TrCE activity and therefore it was concluded, that the protein expression was not successful. Mass-spectrometry analysis of produced protein was not applied.

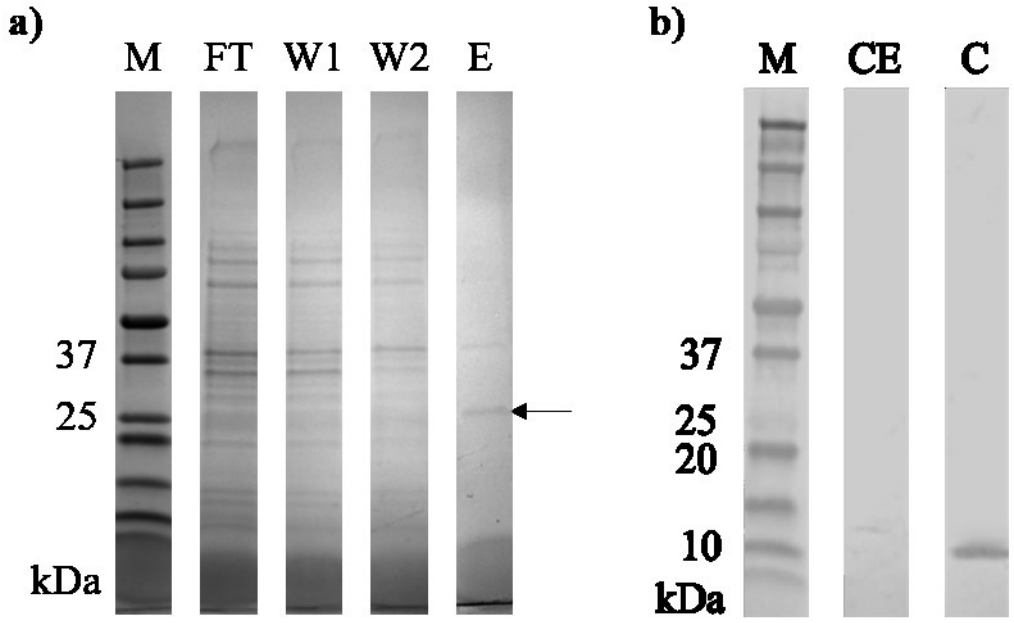

Figure 19. a) Ni-NTA affinity chromatography fractions separated on SDS-PAGE. M – Precious Plus Protein Dual Xtra Standards, **FT** – flow-through, **W1** – first wash step, **W2** – second wash step, **E** – elution**, arrow –** possible rTrCE. **b) Western blot analysis with anti-His-tag antibodies. M** – Precious Plus Protein Dual Xtra**. CE** – sample with possible cercarial elastase, **C –** positive control – His-tagged recombinant FhCys

5.4.5. Protein isolation from inclusion bodies

After Ni-NTA beads affinity chromatography of the protein sample obtained by solubilizing bacterial pellet with 6M guanidium hydrochloride, flow-through, wash steps and elution fractions were analysed using SDS−PAGE (**Figure 20.**). Electrophoresis indicated possible successful protein expression, because of the occurrence of two strong protein bands in flow-through not present in uninduced cells. Although the molecular weight was slightly higher than an expected molecular weight of rTrCE, further analyses were performed since SDS-PAGE determines approximate, not exact molecular weight. The second band with higher molecular weight might occur due to the insufficient reduction of the protein sample.

The protein bands were present in the flow-through. Western blot analysis also refuted the presence of functional His-tag labelling (**Figure 20.**). Due to unsuccessful attempts to solubilize the protein in conventional buffers, the protein could not be purified by Fast Protein Liquid Chromatography (FPLC). Protein bands were thus precisely cut out from the gel and sent to mass spectrometry analysis.

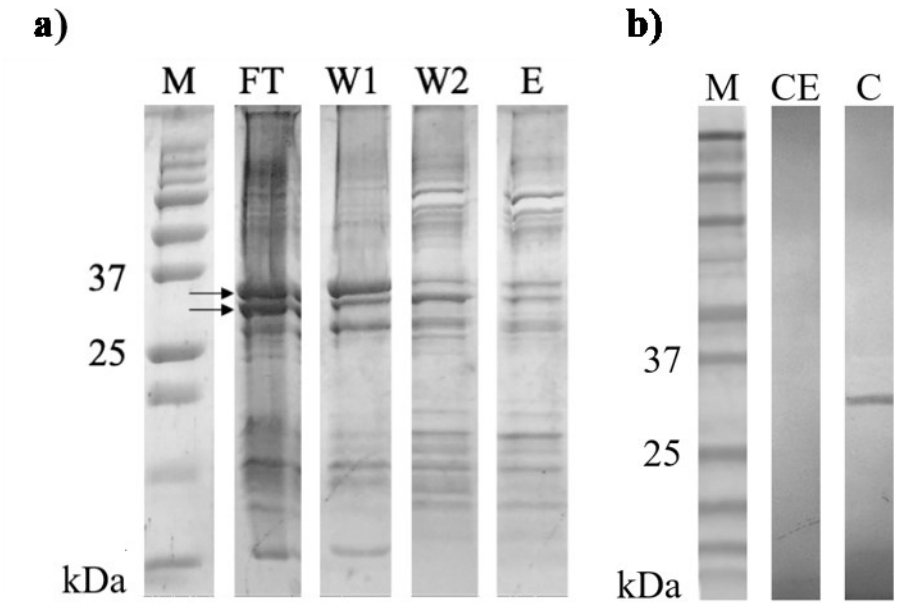

Figure 20. a) Ni-NTA affinity chromatography fractions separated on SDS-PAGE. M – Precious Plus Protein Dual Xtra Standards, **FT,** – flow-through **W1** – first wash step, **W2 –** second wash step, **E –** elution**, arrows** – possible rTrCE. **b) Western blot analysis with anti-His-tag antibodies. M** – Precious Plus Protein Dual Xtra**. CE** – sample with possible cercarial elastase, **C –** positive control – His-tagged recombinant *T. regenti* cathepsin B2

5.4.6. Mass spectrometry

Mass-spectrometry confirmed the presence of rTrCE in bacterial inclusion bodies in both analysed bands (**Table 43.**). Inclusion bodies were therefore purified and the acrylamide gel for immunization was prepared.

Table 43. The amino acid sequence of the rTrCE. Peptides identified by mass spectrometry in green

The amino acid sequence of the protein – 264 aa	**Theoretical pI/Mw: 9.51/ 28,11 kDa**
MGSSHHHHHHSSGLVPRGSHMASTWLVRSGETVRDKNEFPFLALMVTDSSMCTATLVSTRA VITAGHCVCGPKSINRISFQTLSDFDKRSVNYLATSIKVAPEYDPVCQLKRARRRTRQSFGGYD MAIVTLSGLVDLSSGVRVISLATASDIPVPNSIVFIVGYGKDSKATDDTRKYGGVLKKGRATV MKCMHKVVGNPICIRPGPVSAQIVGPGDSGGPLLLTPQGPIIGVASNGVFLPAINDLTVEYASV SRSLEFIMSNI–	

5.4.7. Production of specific antibodies

Protein profile of purified inclusion bodies showed an extra protein band corresponding to the size of rTrCE. This protein band was not previously analysed by mass-spectrometry and mice were therefore immunized with two protein bands confirmed to contain recombinant cercarial elastase (**Figure 21.**).

Both mice survived the immunization process. After collection of the blood sera, Western blot with inclusion bodies proteins as an antigen and with mice sera as a primary antibody was performed. After several experiments with different primary and secondary antibodies dilution ranging from 1:50 to 1:10000, an experiment in **(**Figure 22**.)** was considered as the most successful and reliable.

The Western blot analysis has shown that antibodies in mice sera have an affinity to many bacterial proteins but possibly also to proteins, that the mice were immunized with (**Figure 22.**).

To confirm that, Western blot analysis was also performed with proteins from cercarial homogenate. Not a single reaction occurred on the membrane with cercarial protein treated with immunized mice sera (**Figure 22.**). Consequently, it was concluded, that raising of specific antibodies against recombinant protein by immunization of mice with recombinant protein in acrylamide gel was not successful and therefore, immunolocalization of was not possible to be performed.

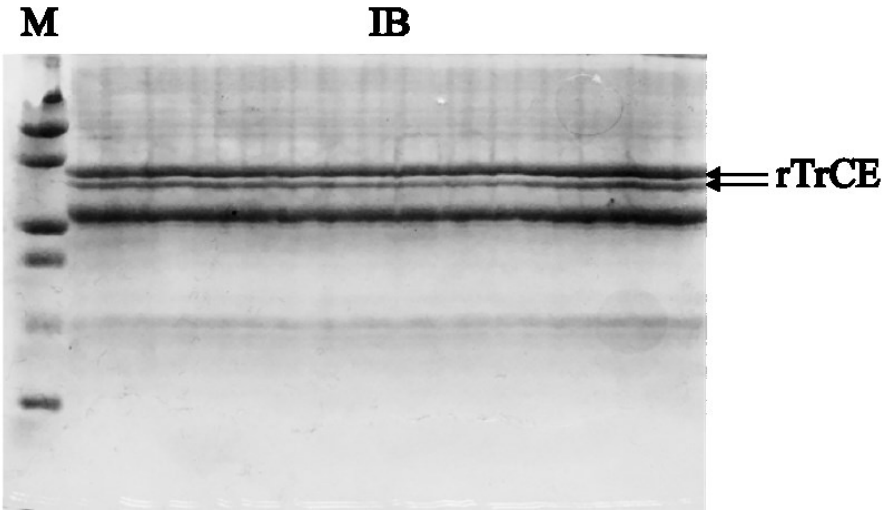

Figure 21. Acrylamide gel used for immunization of mice. M – Precious Plus Protein All Blue Standards, **IB –** inclusion bodies solubilized in sample buffer, **rTrCE** – recombinant cercarial elastase used for immunization

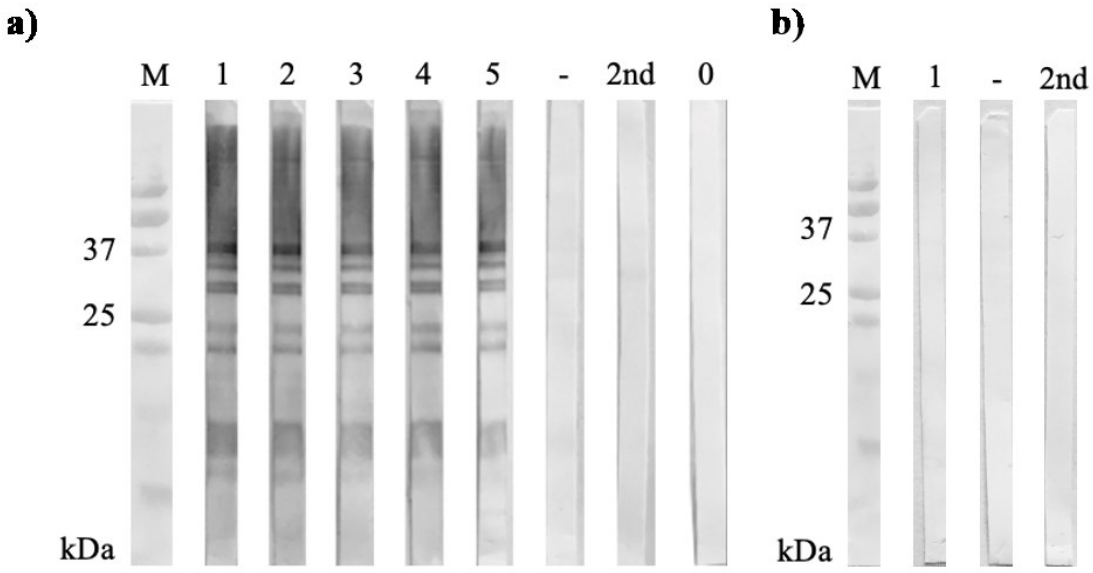

Figure 22. a) Western blot with inclusion bodies proteins as antigens and immunized mice sera as primary antibodies. M - Precious Plus Protein All Blue Standards, **1** – mice sera 1:500, **2** - mice **sera 1:1000,** 3 - mice sera 1:2000, **4** - mice sera 1:5000, **5** - mice sera 1:10000, **- -** pre-immunized mouse serum 1:500, **2nd –** no primary antibody, **0** – no serum, no secondary antibody**. b) Western blot with cercarial homogenate extract as antigens and immunized mice sera as primary antibodies. 1** – mice sera 1:500, **- -** pre-immunized mouse serum 1:500, **2nd** – no primary antibody

5.5. qPCR analysis of *T. regenti* cathepsin B2 and cercarial elastase

5.5.1. Comparison of relative gene expression between TrCB2 and TrCE in both stages

qPCR was performed to compare relative expression levels of cercarial elastase and cathepsin B2 in sporocyst and in cercariae. Relative expression fold change between the genes in a single stage was calculated from measured C_T values for both genes in both stages with Livak ($\Delta\Delta C_T$) method (Livak and Schmittgen, 2001). The variance within triplicates met the condition to be lower than 0.5 of a cycle. (**Table 44.**).

Calculated fold change between C_T values according to Livak method indicated, that the *T. regenti* cercarial elastase is more expressed then cathepsin B2 in both analysed stages – 8,11x in sporocyst and 77,17x in cercariae (**Figure 23.**).

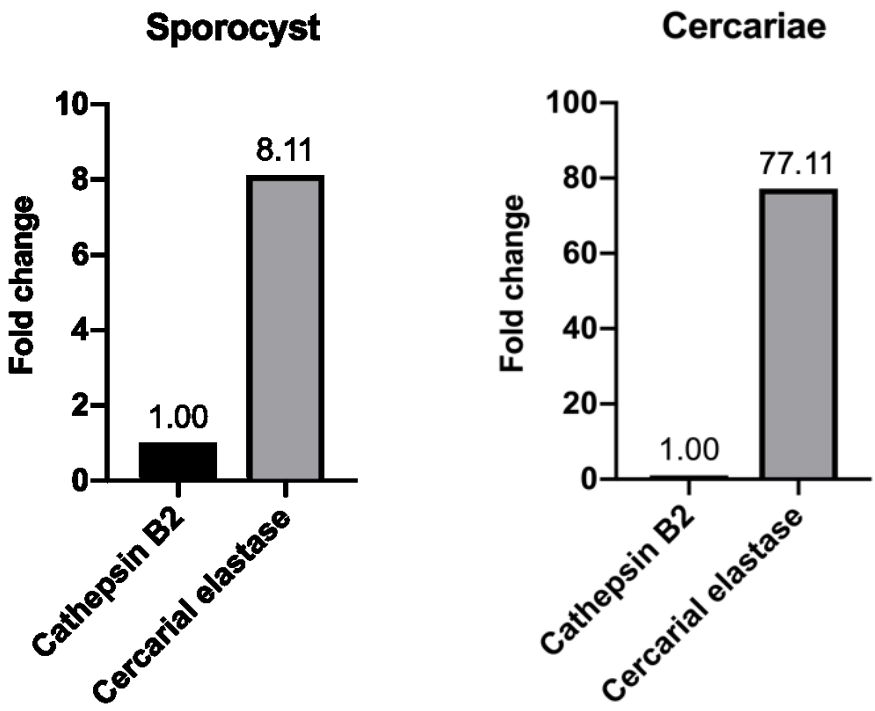

Figure 23. Relative gene expression comparison between *T. regenti* cathepsin B2 and cercarial elastase in sporocyst and in cercariae.

Table 44. Averaged measured C_T values for analysed genes in sporocyst and cercariae

Sample	TrCE C_T	TrCB C_T	PSMD4 C_T	NDUF2V C_T	TPC2L C_T
Sporocyst	18.11±0.04	21.13±0.03	21.27±0.03	22.97±0.09	21.51±0.21
Cercariae	13.23±0.02	19.50±0.23	17.11±0.08	18.40±0.12	17.38±0.23

5.5.2. Comparison of relative gene expression between sporocyst and cercariae

The average variance of reference genes (4,29) between the stages was subtracted from C_T values measured for sporocyst (**Table 45.**). Data for both stages were then normalized with the normalizing factor of 17,63 (geometric average of the reference genes C_T values). Relative expression values were then calculated. Calculated fold change between stages indicates, that the *T. regenti* cercarial elastase is **1.5x** more expressed in cercariae while cathepsin B2 is **6.29x** more expressed in sporocyst (**Figure 24.**).

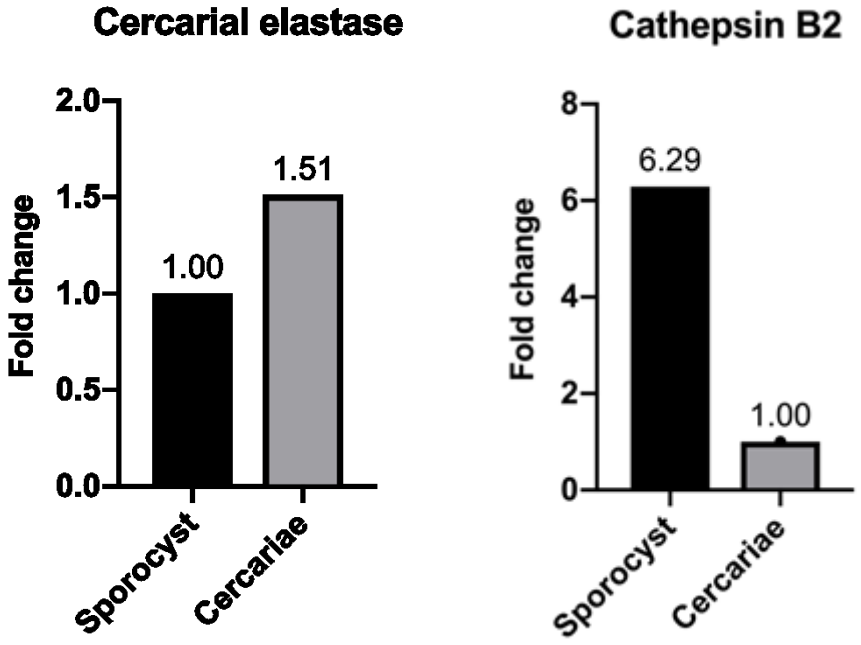

Figure 24. Relative gene expression comparison of *T. regenti* cathepsin B2 and cercarial elastase between sporocyst and cercariae.

Table 45. Averaged measured C_T values for analysed genes in sporocyst and cercariae after compensation of the input RNA for sporocyst

Sample	TrCE C_T	TrCB C_T	PSMD4 C_T	NDUF2V C_T	TPC2L C_T
Cercariae	13.23±0.02	19.50±0.23	17.11±0.08	18.40±0.12	17.38±0.23
Sporocyst – 4.29	13.82	16.84	16.98	18.68	17.22

6. Discussion

6.1. Collection of the excretory-secretory products and its characterization

The first goal of our work was to collect excretory-secretory products from *T. regenti* cercariae, compare the methods of ESPs collection and identify peptidases present in the samples. Resulting data are valuable for determination of the enzymes likely to be responsible for penetration of the definitive host.

We undertook first in-depth proteomic analysis of the peptidases released during the transformation of *T. regenti* cercariae into schistosomula. We confirmed microscopically and histologically, that both stimulants used for ESPs collection led to emptying of cercarial penetration glands as previously reported by other authors(Knudsen et al., 2005; Hansell et al., 2008). Using modified protocol from Hansell et al. 2008 using the skin as the stimulant for transformation, we identified 163 different proteins of which 62 were identified by the MEROPS database as peptidase-like proteins. Contrarily, authors of the original study using this method on *S. mansoni* cercariae identified only 23 proteins. A significant difference in a number of identified proteins could be explained by the database used for identifications – while we used the whole UniProt database UniRef100 in 2019, Hansell et al. used *S. japonicum* and *S. mansoni* entries present in the UniProt database in 2006. Using this method, we, for the first time, identified two protein orthologues of serine cercarial elastase (clan PA, family S1) in *T. regenti*. Furthermore, a more abundant orthologue of this peptidase was the fourth most abundant peptidase-like protein in the ESP. Since the first three most represented enzymes are ubiquitous throughout schistosome life-cycle, our results suggest that cercarial elastase is the most abundant peptidase potentially responsible for host penetration. This result is also in line with results of secretomic studies performed on the cercariae of *S. mansoni* (Knudsen 2005; Curwen 2006; Hansell et al. 2008). On the other hand, cercarial elastase is one of the least represented peptidases in cercarial ESPs of *S. japonicum*, where the most dominant peptidases are leishmanolysins and cathepsin B2.

The significantly lower number of identified proteins and absence of cercarial elastase in samples induced with linoleic acid suggests that this method of ESPs collection does not truly simulate the emptying of penetration glands as it happens in natural infections, as it was also previously suggested by Hansell et al., 2008. This is supported by the fact, that subsequent observation of cercariae after the ESPs collection using linoleic acid showed, that significant amount of the proteins was released after collection of the samples used for mass spectrometry processing analysis. This may be the result of prolonged exposure of larvae to this stimulant, which causes glycocalyx shedding and death of the parasites in the aqueous environment (Salafsky et al., 1988). This observation is also in line with the hypothesis, that content of post-acetabular glands is released after the linoleic acid stimulation, while pre-acetabular glands

containing penetration enzymes are completely emptied after the attachment of cercariae to the skin surface (Salter et al., 2000; McKerrow and Salter, 2002).

The high abundance of the *T. regenti* serine cercarial elastase in the skin induced samples also indicates, that this enzyme could be directly involved in the penetration process. This result is in stark contrast with results of Dvořák et al. 2008 who rejected the presence of serine peptidases in *T. regenti* secretions. It should be noted that the authors of the article used mechanical transformation to collect ESPs and determined the presence of peptidase types by their activity. Another surprising result with respect to mass spectrometric data analysis is the complete absence of cathepsins (CA, family C1A) in all analysed samples. The activity of cysteine peptidases in *T. regenti* cercarial secretions was repeatedly reported by members of our team (Mikeš et al., 2005; Kašný et al., 2007) and it was attributed to cathepsin CB2. Identification of this peptidase on mRNA level of *T. regenti* sporocyst, immunolocalization in cercarial penetration glands of cercaria (Dolečková et al., 2009) and seeming absence of cercarial elastase gene or protein resulted in hypothesis, that cathepsin B2 might be the major penetration enzyme of this life stage. On the other hand, cathepsins were never identified by mass-spectrometry in *T. regenti* cercariae. This opens up a space for speculation that cathepsin B2 is present in glands only in minor amounts but of great activity or alternatively, it is not present at all and reported cysteine peptidase activity corresponds to other cysteine peptidases present in the ESP. Our results together with similar research progress in *S. japonicum* where the situation used to be the similar (Dvořák et al., 2008) strengthen the position of cercarial elastase as a common dominant enzyme of schistosomes responsible for host penetration (Zhang et al., 2018).

It is also apparent from the results that the method used for the collection of ESPs has a direct influence on the number and spectrum of identified proteins in samples. We conclude that the induction of transformation using host skin represents more faithfully the gland contents rather than proteins released as a result of larvae reacting on a single stimulant. This also raises a question of what other factors play a role in complete glandular emptying – the answer to this question lies on the shoulders of future researchers.

6.2. *T. regenti* cercarial elastase gene

The second goal of our study was to further characterize peptidases possibly responsible for penetration of the definitive host.

For many years it has been assumed, that the active orthologue of cercarial elastase is not present in *T. regenti* (Dolečková et al. 2007; Dvořák et al. 2008; Dolečková et al. 2009). After the *T. regenti* genome was published as part of the 50 Helminth Genomes project in 2014, partial orthologue gene sequence of cercarial elastase was found. We, for a first time, confirmed active transcription of this gene by conventional PCR. Previous attempts of other researchers to amplify this gene with PCR were unsuccessful possibly due to the use of degenerate oligonucleotide

probes based on consensus regions of human schistosomes' cercarial elastase genes (Dolečková et al. 2007).

Surprisingly, we confirmed active transcription not only in sporocysts but also in cercariae which is in conflict with the *T. regenti* cercarial transcriptome (Leontovyč et al. 2016) where the transcripts of this gene are completely missing in all four biological replicates. The original idea to resolve this discrepancy was that our cercarial RNA sample from which the cDNA was subsequently synthesized was contaminated with genomic DNA. However, this hypothesis was refuted by our successful amplification of the whole gene sequence by the 3' RACE PCR. This method uses the natural polyA tail that exists at the 3' end of all eukaryotic mRNAs for priming during reverse transcription and thus gene has to be transcribed to be subsequently amplified by PCR (Yeku and Frohman, 2010).

After the personal communication with the authors of the transcriptome, the original unfiltered read data obtained from the sequencing (HiSeq 2500 platform, Illumina) were searched for the presence of cercarial elastase transcripts, but with no success. This led to the bold hypothesis, that the mRNA coding cercarial elastase of *T. regenti* is synthesised exclusively in germ balls, stored within the cercariae and translated to the protein form after the contact with the definitive host. This would be also in line with the recent publication, that claims that cercariae are transcriptionally silent (Roquis et al., 2015). The coordinated mRNA storage in ribonucleoprotein granules is conserved across all eukaryotes (Standart and Weil, 2018) and average protein synthesis time in eukaryotes is ca. 500 amino acids in 1 minute and 40 seconds which would be enough time to produce the protein after the finding the host (Fedorov and Baldwin, 2001). This hypothesis would also provide an alternative explanation of the absence of cercarial elastase in linoleic acid-induced ESPs samples. However, the recent mass-spectrometry analysis confirmed the presence of the protein form of cercarial elastase in uninduced *T. regenti* cercarial homogenate extract (Cwiklinski, 2019, unpublished results) and thus refuted this hypothesis. The absence of this transcript in the transcriptome remains to be elucidated.

6.3. *T. regenti* cercarial elastase features

According to the sequence obtained by 3´RACE PCR, the most similar protein to TrCE is SjCE-2b isoform with 65,78 % amino acid sequence similarity. In *S. japonicum* gene coding this isoform was discovered after the release of the parasite´s genome (Zhou et al., 2009) and it was speculated to be the most evolutionarily ancient among other cercarial elastases. This also lead to the assumption that this protein may have a different function in the skin invasion process and beyond the parasite life cycle (Ingram et al., 2012). Later, SjCE-2b was found in small abundance in ESPs of *S. japonicum* cercariae (M. Liu et al., 2015) but further investigation suggesting its features and the possible role was lacking. Finally, the most recent study showed, that

recombinant form of this peptidase displays similar enzymatic activity and substrate preference to *S. mansoni* cercarial secretions suggesting its same function (Zhang et al., 2018).

TrCE sequence similarity with other elastases able to cleave skin proteins, conserved full catalytic triad, abundant presence in ESPs and high expression levels again indicate its proteolytic role in the penetration process. To confirm or refute this hypothesis we decided to produce its recombinant form to determine its activity, substrate specificity and localization.

6.4. *T. regenti* cercarial elastase protein expression in yeast

To increase our chances of producing properly folded, a post-translationally processed and active recombinant protein secreted to the cultivation media, PichiaPink yeast expression system was selected. We experimentally confirmed that the gene construct was synthesised as we designed it and successfully integrated into the DNA of the yeast cells. Unfortunately, despite trying a large number of different expression conditions and approaches recommended by the manufacturer and experts in the field (prof. John Dalton's Laboratory members), we did not succeed to detect our protein by protein electrophoresis, Western blot, enzymatic assays or even by alternative methods such as "Yeastern blot". The failure of all these detection methods led us to the conclusion that protein expression has failed. There are several explanations of why this could have happened.

Yeast *Pichia pastoris* is known to secrete various peptidases during fermentation, which can result in degradation of the desired recombinant protein in the media. Although protease inhibitors could have been added to the medium during fermentation, it did not fit our goal to produce active serine peptidase. Some of the proteolytic activities of *Pichia pastoris* have been characterized and genes coding these peptidases (PEP4 and PRB1) has been knocked down in PichiaPink strain #4 (Gleeson et al., 1998) which we also tested for expression in our experiments. Sadly, with no success. However, proteolysis may occur even either during vesicular transport of recombinant protein by secretory pathway-resident peptidases or by cell wall-associated peptidases (Ahmad et al., 2014).

Another explanation can be in used signal sequence. Protein secretion requires the presence of a signal sequence on the expressed protein, coding signal peptide to target it to the secretory pathway. We have cloned and tried to express our gene construct in the frame of signal sequence "α-mating factor" that has been most successful previously (Dalton, 2018; personal communication) and is provided by the kit's manufacturer as the signal sequence of first-choice. We did not succeed; however, the expression system offers seven other signal sequence options that have not been tested. Moreover, due to the use of the α-mating factor signal sequence, we removed the cercarial elastase signal sequence predicted by SignalP 4.1 from the gene construct (Nielsen, 2017). Surprisingly, a new version of the programme SignalP 5.0 (Almagro Armenteros et al., 2019) that was released in 2019 suddenly refuses the presence of

a signal peptide in our gene. If the latter prediction is correct, the entire design of our gene construct is based on the wrong premise and therefore the final product could have been different. We have not found originally predicted signal peptide or at least its parts within the cercarial elastase peptides identified in ESPs samples, suggesting credibility of the original prediction, however, sequence coverage of cercarial elastase in mass spectrometry data was only 46.4 %.

Even though this enzyme has been extensively studied for over 40 years, there is no scientific publication to date reporting successful expression of this enzyme in a yeast expression system, thus creating a huge challenge for future scientists.

6.5. *T. regenti* cercarial elastase protein expression in bacteria

Recombinant protein expression experiments in *E. coli* BL21 competent cells were performed with two different gene constructs – with a His-tag sequence on 5' or 3' end of the gene. We confirmed the successful synthesis of both constructs and subsequent transformation to the bacterial cells by PCR colony screening. SDS-PAGE of all tested samples obtained with various induction conditions has shown promising results of expression to the periplasm and to inclusion bodies for 3'end His-tag and 5' end His-tag constructs respectively.

We failed to confirm periplasmic expression by Western blotting, therefore, enzymatic assays were performed to detect the protein by its activity. A recent publication about recombinant cercarial elastase of *S. japonicum* reports, that recombinant form of this enzyme is able to cleave Ala-Ala-Pro-Phe substrate in Tris-HCl buffer at 37 °C (Zhang et al 2018). Since cercarial elastase of *T. regenti* is closely related protein, we expected similar activity. Despite testing this and also different substrates and buffers, we have not achieved any positive results.

Finally, we confirmed recombinant protein expression in inclusion bodies. Firstly, SDS-PAGE revealed two distinct bands in samples of whole cell lysates which were later identified as a recombinant cercarial elastase. Although the His-tagged protein did not bind to the Ni-NTA beads and did not react on the Western blot, His-tag peptide was confirmed by mass spectrometry. Normally, we would assume that His-tag was incorporated into the tertiary structure of the protein and thus inaccessible to the Ni-NTA beads, but the 6M guanidine hydrochloride used to solubilize the inclusion bodies causes strong denaturation conditions (Hibbard and Tulinsky, 1978). Due to the failure of protein purification with affinity chromatography, we have planned to purify protein with FPLC, but again, we have failed to transfer soluble proteins from bacterial lysate to conventional buffers since the proteins precipitated during the buffer exchanges. In this case, immunization of mice with acrylamide gel is the particularly advantageous method when protein is insoluble without detergents. In addition, a further advantage of this method is an enhancement of the immune response, since polyacrylamide helps to retain the antigen in the animal and so acts as an adjuvant (Boulard and Lecroisey, 1982; Amero et al., 1988). Unfortunately, sera from

immunized mice reacted unspecifically on Western blot with recombinant protein and did not react with any protein from cercarial homogenate.

Failure to raise antibodies by immunization of mice can be explained by insufficient immunogenicity of cercarial elastases. Pilot study immunizing mice with native form of SmCE has shown only poor immunoreactive potential (Darani et al., 1997). Further investigations demonstrated that *S. mansoni*, *S. haematobium* and *S. japonicum* infection induces specific antibodies against cercarial secretions, but not against the cercarial elastases of these parasite species, implying that although CE is present in cercarial secretions, it is not apparently immunogenic during natural infection (Bahgat et al., 2001). An explanation of poor immunogenicity in natural infections could be the parasite's ability to cleave host IgG and IgE molecules to overcome the antibody response (Aslam et al., 2008)

However, only little is known of whether animals immunized with recombinant elastase exhibit immunity response. A more recent study showed that a recombinant SmCE fused to recombinant *S. japonicum* glutathione S-transferase 28 (SjGST28) resulted in smaller mean worm burden and lower tissue egg counts in immunized mice, but without statistical significance (El-Faham et al. 2017). Also, SjGST28 itself is long-term promising vaccine candidate (McManus and Loukas, 2008). The most recent study with recombinant refolded SjCE-2b has shown a statistically significant reduction in both adult worms and liver egg counts in *S. japonicum* infected immunized mice indicating this peptidase exhibited a certain level of immunoprotective effect upon cercaria invasion. Nevertheless, other factors could have a direct impact on the effectiveness of the mice immunization such as antigen amount given, adjuvants and the scheduling of dosages. From all this, we conclude that immunization has failed either due to the low immunogenicity of our protein, its denaturation or the immunization protocol used.

Recent successful *S. japonicum* cercarial elastase expression, refolding and mice immunization (Zhang et al., 2018) can serve as a guide to future researchers trying to characterize this protein.

6.6. qPCR analysis of cathepsin B2 and cercarial elastase

The third goal of our study was to compare relative expression levels of the main peptidases possibly involved in penetration of *T. regenti* cercariae and thus support of the mass-spectrometry analysis of cercarial ESPs.

To compare relative gene expression between both genes within sporocyst and cercariae, Livak's method was used. This method is mostly used to compare the relative gene expression of a gene with interest with the selected stably compressed reference gene, especially in "treated" and "untreated" sample studies (Livak and Schmittgen, 2001). We used this formula to express relative fold change between two genes when one gene serves as a "reference gene" to the other. Our data suggest, that cercarial elastase is more expressed than cathepsin B2 in both analysed life

stages with a significant increase after the cercariae leave snail intermediate host. This result is in line with our proteomic data, which show that cercarial elastase is more dominant peptidase in cercarial penetration glands therefore was identified by mass-spectrometry in high abundance while much less expressed cathepsin B2 remained unidentified.

Furthermore, our comparison of relative expression levels of the two genes between the developmental stages has shown, that expression of cercarial elastase is relatively stable during the differentiation of the cercariae into sporocyst and in cercariae themselves, which is different trend that has been observed in human schistosomes, where cercarial elastase is primarily expressed in sporocyst stage (Parker-Manuel et al., 2011). Cathepsin B2 has shown an opposite pattern in our study, matching the results previously delivered by members of our team suggesting that gene coding this peptidase is significantly more transcriptionally active in intra-molluscan stages than in cercariae (Dolečková et al., 2010).

To be able to compare relative expression levels of the genes of our interest between the developmental stages, the same amount of input RNA of both stages when preparing cDNA must be assured. Since sporocyst is surrounded by hepatopancreatic gland of its snail host, it is impossible to extract parasite material alone with simple dissection without the contamination with snail tissues and thus its RNA. The only reliable approach to extract pure parasite material from this stage is by using microdissection as previously described (Parker-Manuel et al., 2011) Due to the inability to repeat this procedure because of insufficient technical equipment, we have developed an alternative method to compensate for the difference in the input RNA and thus make the expression levels between the stages comparable. According to S. Liu et al. 2012, PSMD, NDUF and TPC2L of *S. japonicum* are the most stably expressed genes across the life cycle of the parasite. If this is true also for *T. regenti*, the variance between C_T values for these three genes between the mentioned two stages should be minimal. If so, the difference between the amount of parasitic RNA used for reverse transcription can be relatively calculated from the average variance of the housekeeping genes C_T values. Therefore, we subtracted this average variance from the C_T values for sporocysts genes of interest and the resulting C_T values should match the situation where the same amount of input RNA was used.

However, the results of qPCR and especially those from comparison of expression between the developmental stages need to be critically evaluated, given that each small difference in the amount of input RNA, reverse transcription efficiency, efficacy of PCR and even a slight pipetting error lead to a significant differences in final C_T values and associated fold changes (Bustin and Huggett, 2017; Debode et al., 2017). For this reason, we did not conduct the experiment to determine the exact and absolute differences between gene and stage expression levels, but mainly to clarify the main trends in different stages of the parasite, to support mass-spectrometric analysis results and to gain a broader perspective to evaluate our and others results. To make our qPCR analysis more trustworthy, it would be necessary to repeat the experiment with more biological and technical replicates, which, unfortunately, was not possible due to a lack of parasite material during the internship at Queen's University Belfast.

7. Conclusions

Cercaria is the first life stage of *Trichobilharzia regenti* that definitive host encounters. Employment of excretory-secretory products, especially peptidases released from penetration glands of this invasive larvae leads to degradation of skin proteins and thus successful transmission of the parasite. Identification of these peptidases and clarifying the mechanisms responsible for the penetration process is essential for a complete understanding of the unique biology of this neuropathogenic schistosome. The aim of this work was to perform a complex secretomic study using different methods of ESPs collection, to further characterize dominant peptidases possibly involved in the invasion of the host and compare the gene expression of these peptidases in and between different life stages.

In particular, these results were achieved in this work:

- Stimulation of emptying cercarial penetration glands using the host's skin as was considered to be a more appropriate method of ESPs collection than stimulation with linoleic acid due to the significantly higher number of identified proteins and peptidase-like proteins.

- Successful identification of peptidase-like proteins released during cercaria-schistosomulum transformation was achieved. One of the most abundant proteins in *T. regenti* cercarial ESPs was cercarial elastase – serine peptidase notoriously known to be responsible for penetration of *S. mansoni* cercariae, but not previously discovered in our species. Cathepsin B2, previously considered as the major penetration tool of *T. regenti* cercariae was not identified in any analysed sample in our study.

- Gene and protein sequence of *T. regenti* cercarial elastase is similar to cercarial elastases of human schistosomes and contains full catalytic triad which indicates its possible ability to cleave skin proteins.

- Cercarial elastase is more expressed in sporocyst and cercariae than cathepsin B2 which isindicating its dominant role in the penetration of the host.

In the frame of this master's research project, knowledge of the mechanisms responsible for successful parasite transmission to the definitive host has been deepened. The discovery of cercarial elastases opens the door for further research on these enzymes. Also, the role of cathepsin B2 must be reassessed and explored by further experimental work. In addition, the data obtained with our secretomic study can serve to determine and investigate other important proteins of schistosomes within the life cycle.

8. Recipes

Reagent	Recipe	Autoclave
Yeast expression		
BMGY	1% yeast extract, 2% peptone, 0,1M potassium phosphate buffer pH 6.0, 1,34% YNB, 0,00004% Biotin, 1% glycerol	yes
BMMY	1% yeast extract, 2% peptone, 0,1M potassium phosphate buffer pH 6.0, 1,34% YNB, 0,00004% Biotin, 0,5% methanol	yes
PAD plates	6,7g YNB w/o Amino Acids (Sigma-Aldrich), SMM-Adenine (Sigma-Aldrich), 20g agar (Millipore), 1 l of H_2O	yes
1M potassium phosphate buffer	132 ml of 1M K_2HPO_4 and 868 ml of 1M KH_2PO_4	yes
10x YNB	34g YNB w/o Amino Acids (Sigma-Aldrich), 100 g ammonium sulfate, 1 l of H_2O	yes
YNB plates	6,7g YNB w/o Amino Acids (Sigma-Aldrich), 20 g of powdered agar (Millipore), 1 l of H_2O	yes
YPD medium	50 g YPD broth (Millipore) 1l of H_2O	yes
YPD/agar plates	2% of powdered agar (Millipore), YPD broth 50 g, 1l of H_2O	yes
YPDS media	0,5M sorbitol, 50 g YPD broth,1l of H_2O	yes
Bacterial expression		
LB medium	25 g LB broth, (Sigma-Aldrich), 20 g of powdered agar (Millipore), 1 l H_2O	yes
LB plates	25 g LB broth, (Sigma-Aldrich), 1 l of H_2O	yes
ST buffer pH 8	10mM Tris, 150mM NaCl, 1 l of H_2O	no
Affinity chromatography		
Elution buffer	50mM NaH_2PO_4, 300mM NaCl, 250mM imidazole, 1 l of H_2O	no
Lysis buffer	50mM NaH_2PO_4, 300mM NaCl, 5mM imidazole, 1 l of H_2O	no
Wash buffer	50mM NaH_2PO_4, 300mM NaCl, 10mM imidazole, 1 l of H_2O	no
SDS−PAGE and Western blot		
De-stain solution	20% methanol, 10% glacial acetic acid, 70% of H_2O	no
TBS-T	20 mM Tris, 150 mM NaCl, 0.5% Tween 20, 1 l of H_2O	no
Transfer buffer	25 mM Tris, 192 mM glycine, 10% methanol, 1 l of H_2O	no

9. Bibliography

Ahmad M., Hirz M., Pichler H., Schwab H. (2014). Protein expression in *Pichia pastoris*: Recent achievements and perspectives for heterologous protein production. *Applied Microbiology and Biotechnology*, *98*, 5301–5317.

Aldhoun J. a, Kolářová L., Horák P., Skírnisson K. (2009). Bird schistosome diversity in Iceland: molecular evidence. *Journal of helminthology*, *83*, 173–180.

Almagro Armenteros J. J., Tsirigos K. D., Sønderby C. K., Petersen T. N., Winther O., Brunak S., von Heijne G., Nielsen H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, *37*, 420–423.

Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*, 403–410.

Alves P., Arnold R. J., Novotny M. V, Radivojac P., Reilly J. P., Tang H. (2007). Advancement in protein inference from shotgun proteomics using peptide detectability. *Pacific Symposium on Biocomputing*, 409–420.

Amero S. A., James T. C., Elgin S. C. R. (1988). Production of antibodies using proteins in gel bands In: Walker J.M. (eds) *New Protein Techniques. Methods in Molecular Biology, vol 3*. Humana Press, 355–362.

Artimo P., Jonnalagedda M., Arnold K., Baratin D., Csardi G., De Castro E., Duvaud S., Flegel V., Fortier A., Gasteiger E., Grosdidier A., Hernandez C., Ioannidis V., Kuznetsov D., Liechti R., Moretti S., Mostaguir K., Redaschi N., Rossier G., Xenarios I., Stockinger H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, *40*, 597–603.

Aslam A., Quinn P., McIntosh R. S., Shi J., Ghumra A., McKerrow J. H., Bunting K. A., Dunne D. W., Doenhoff M. J., Morrison S. L., Zhang K., Pleass R. J. (2008). Proteases from Schistosoma mansoni cercariae cleave IgE at solvent exposed interdomain regions. *Molecular Immunology*, *45*, 567–574.

Bahgat M., Francklow K., Doenhoff M. J., Li Y. L., Ramzy R. M. R., Kirsten C., Ruppel A. (2001). Infection induces antibodies against the cercarial secretions, but not against the cercarial elastases of *Schistosoma mansoni*, *Schistosoma haematobium*, *Schistosoma japonicum* and *Trichobilharzia ocellata*. *Parasite Immunology*, *23*, 557–565.

Bateman A., Martin M. J., O'Donovan C., Magrane M., Alpi E., Antunes R., Bely B., Bingley M., Bonilla C., Britto R., Bursteinas B., Bye-AJee H., Cowley A., Da Silva A., De Giorgi M., Dogan T., Fazzini F., Castro L. G., Figueira L., Garmiri P., Georghiou G., Gonzalez D., Hatton-Ellis E., Li W., Liu W., Lopez R., Luo J., Lussi Y., MacDougall A., Nightingale A., Palka B., Pichler K., Poggioli D., Pundir S., Pureza L., Qi G., Rosanoff S., Saidi R., Sawford T., Shypitsyna A., Speretta E., Turner E., Tyagi N., Volynkin V., Wardell T., Warner K., Watkins X., Zaru R., Zellner H., Xenarios I., Bougueleret L., Bridge A., Poux S., Redaschi N., Aimo L., ArgoudPuy G., Auchincloss A., Axelsen K., Bansal P., Baratin D., Blatter M. C., Boeckmann B., Bolleman J., Boutet E., Breuza L., Casal-Casas C., De Castro E., Coudert E., Cuche B., Doche M., Dornevil D., Duvaud S., Estreicher A., Famiglietti L., Feuermann M., Gasteiger E., Gehant S., Gerritsen V., Gos A., Gruaz-Gumowski N., Hinz U., Hulo C., Jungo F., Keller G., Lara V., Lemercier P., Lieberherr D., Lombardot T., Martin X., Masson P., Morgat A., Neto T., Nouspikel N., Paesano S., Pedruzzi I., Pilbout S., Pozzato M., Pruess M., Rivoire C., Roechert B., Schneider M., Sigrist C., Sonesson K., Staehli S., Stutz A., Sundaram S., Tognolli M., Verbregue L., Veuthey A. L., Wu C. H., Arighi C. N., Arminski L., Chen C., Chen Y., Garavelli J. S., Huang H., Laiho K., McGarvey P., Natale D. A., Ross K., Vinayaka C. R., Wang Q., Wang Y., Yeh L. S., Zhang J. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, *45*, 158–169.

Bendtsen J. D., Jensen L. J., Blom N., Von Heijne G., Brunak S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Engineering, Design and Selection*, *17*, 349–356.

Boulard C., Lecroisey A. (1982). Specific antisera produced by direct immunization with slices of polyacrylamide gel containing small amounts of protein. *Journal of Immunological Methods*, *50*, 221–226.

Brindley P. J., Kalinna B. H., Dalton J. P., Day S. R., Wong J. Y. M., Smythe M. L., McManus D. P. (1997). Proteolytic degradation of host hemoglobin by schistosomes. *Molecular and Biochemical Parasitology*, *89*, 1–9.

Bustin S., Huggett J. (2017). qPCR primer design revisited. *Biomolecular detection and quantification*, *14*, 19–28.

Caffrey C. R., McKerrow J. H., Salter J. P., Sajid M. (2004). Blood "n" guts: An update on schistosome digestive peptidases. *Trends in Parasitology*, *20*, 241–248.

Caffrey C. R., Salter J. P., Lucas K. D., Khiem D., Hsieh I., Lim K. C., Ruppel A., McKerrow J. H., Sajid M. (2002). SmCB2, a novel tegumental cathepsin B from adult *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, *121*, 49–61.

Cao X., Fu Z., Zhang M., Han Y., Han Q., Lu K., Li H., Zhu C., Hong Y., Lin J. (2016). Excretory/secretory proteome of 14-day schistosomula, *Schistosoma japonicum*. *Journal of Proteomics*, *130*, 221–230.

Cass C. C. L., Johnson J. R., Califf L. L., Xu T., Hernandez H. J., Stadecker M. J., Yates J. R., Williams D. L. (2007). Proteomic analysis of *Schistosoma mansoni* egg secretions. *Molecular and Biochemical Parasitology*, *155*, 84–93.

Chai M., McManus D. P., McInnes R., Moertel L., Tran M., Loukas A., Jonesa M. K., Gobert G. N. (2006). Transcriptome profiling of lung schistosomula,in vitro cultured schistosomula and adult *Schistosoma japonicum*. *Cellular and Molecular Life Sciences CMLS*, *63*, 919–929.

Chanová M., Horák P. (2007). Terminal phase of bird schistosomiasis caused by *Trichobilharzia regenti* (Schistosomatidae) in ducks (*Anas platyrhynchos f. domestica*). *Folia Parasitologica*, *54*, 105–107.

Chlichlia K., Schauwienold B., Kirsten C., Doenhoff M. J., Fishelson Z., Ruppel A. (2005). *Schistosoma japonicum* reveals distinct reactivity with antisera directed to proteases mediating host infection and invasion by cercariae of *S. mansoni* or *S. haematobium*. *Parasite Immunology*, *27*, 97–102.

Christiansen A., Olsen A., Buchmann K., Kania P. W., Nejsum P., Vennervald B. J. (2016). Molecular diversity of avian schistosomes in Danish freshwater snails. *Parasitology Research*, *115*, 1027–1037.

Coghlan A., Tyagi R., Cotton J. A., Holroyd N., Rosa B. A., Tsai I. J., Laetsch D. R., Beech R. N., Day T. A., Hallsworth-Pepin K., Ke H. M., Kuo T. H., Lee T. J., Martin J., Maizels R. M., Mutowo P., Ozersky P., Parkinson J., Reid A. J., Rawlings N. D., Ribeiro D. M., Swapna L. S., Stanley E., Taylor D. W., Wheeler N. J., Zamanian M., Zhang X., Allan F., Allen J. E., Asano K., Babayan S. A., Bah G., Beasley H., Bennett H. M., Bisset S. A., Castillo E., Cook J., Cooper P. J., Cruz-Bustos T., Cuéllar C., Devaney E., Doyle S. R., Eberhard M. L., Emery A., Eom K. S., Gilleard J. S., Gordon D., Harcus Y., Harsha B., Hawdon J. M., Hill D. E., Hodgkinson J., Horák P., Howe K. L., Huckvale T., Kalbe M., Kaur G., Kikuchi T., Koutsovoulos G., Kumar S., Leach A. R., Lomax J., Makepeace B., Matthews J. B., Muro A., O'Boyle N. M., Olson P. D., Osuna A., Partono F., Pfarr K., Rinaldi G., Foronda P., Rollinson D., Samblas M. G., Sato H., Schnyder M., Scholz T., Shafie M., Tanya V. N., Toledo R., Tracey A., Urban J. F., Wang L. C., Zarlenga D., Blaxter M. L., Mitreva M., Berriman M. (2019). Comparative genomics of the major parasitic worms. *Nature Genetics*, *51*, 163–174.

Cregg J. M., Tolstorukov I., Kusari A., Sunga J., Madden K., Chappell T. (2009). Chapter 13 Expression in the yeast *Pichia pastoris*. *Methods in Enzymology (1st ed., Vol. 463),* 169–189.

Cuesta-Astroz Y., Oliveira F. S. de, Nahum L. A., Oliveira G. (2017). Helminth secretomes reflect different lifestyles and parasitized hosts. *International Journal for Parasitology*, *47*, 529–544.

Curwen R. S., Ashton P. D., Johnston D. A., Wilson R. A. (2004). The *Schistosoma mansoni* soluble proteome : a comparison across four life-cycle stages. *Molecular and Biochemical Parasitology, 138*, 57–66.

Curwen R. S., Ashton P. D., Sundaralingam S., Wilson R. A. (2006). Identification of novel proteases and immunomodulators in the secretions of schistosome cercariae that facilitate host entry. *Molecular & Cellular Proteomics*, *5*, 835–844.

Dalton J.P., Brindley P. J. (1997). Proteases of trematodes. In *Advances in trematode biology*, 265–308.

Darani H. Y., Curtis R. H. C., McNeice C., Price H. P., Sayers J. R., Doenhoff M. J. (1997). *Schistosoma mansoni*: Anomalous immunogenic properties of a 27 kDa larval serine protease associated with protective immunity. *Parasitology*, *115*, 237–247.

Debode F., Marien A., Janssen É., Bragard C., Berben G. (2017). The influence of amplicon length on real-time PCR results. *Biotechnol. Agron. Soc. Environ*, *21*, 3–11.

Dolečková K., Kašný M., Mikeš L., Mutapi F., Stack C., Mountford A. P., Horák P. (2007). Peptidases of *Trichobilharzia regenti* (Schistosomatidae) and its mollusc an host *Radix peregra* s. lat. (Lymnaeidae): Construction and screening of cDNA library from intramolluscan stages of the parasite. *Folia Parasitologica*, *54*, 94–98.

Dolečková K., Kašný M., Mikeš L., Cartwright J., Jedelský P., Schneider E. L., Dvořák J., Mountford A. P., Craik C. S., Horák P. (2009). The functional expression and characterisation of a cysteine peptidase from the invasive stage of the neuropathogenic schistosome *Trichobilharzia regenti*. *International Journal for Parasitology*, *39*, 201–211.

Dolečková K., Albrecht T., Mikeš L., Horák P. (2010). Cathepsins B1 and B2 in the neuropathogenic schistosome *Trichobilharzia regenti*: Distinct gene expression profiles and presumptive roles throughout the life cycle. *Parasitology Research*, *107*, 751–755.

Dresden M., Edlin E. (1975). *Schistosoma mansoni*: Calcium content of cercariae and its effects on protease activity *in vitro*. *The Journal of Parasitology*, *61*, 398–402.

Dvořák J., Fajtová P., Ulrychová L., Leontovyč A., Rojo-Arreola L., Suzuki B. M., Horn M., Mareš M., Craik C. S., Caffrey C. R., O'Donoghue A. J. (2015). Excretion/secretion products from *Schistosoma mansoni* adults, eggs and schistosomula have unique peptidase specificity profiles. *Biochimie*, *122*, 1–11.

Dvořák J., Horn M. (2018). Serine proteases in schistosomes and other trematodes. *International Journal for Parasitology*, *44* , 333–344.

Dvořák J., Mashiyama S. T., Braschi S., Sajid M., Knudsen G. M., Hansell E., Lim K. C., Hsieh I., Bahgat M., Mackenzie B., Medzihradszky K. F., Babbitt P. C., Caffrey C. R., McKerrow J. H. (2008). Differential use of protease families for invasion by schistosome cercariae. *Biochimie*, *90*, 345–358.

El-Faham M. H., Wheatcroft-Francklow K. J., Price H. P., Sayers J. R., Doenhoff M. J. (2017). *Schistosoma mansoni* cercarial elastase (SmCE): differences in immunogenic properties of native and recombinant forms. *Parasitology*, *144*, 1356–1364.

El Ridi R., Tallima H., Dalton J. P., Donnelly S. (2014). Induction of protective immune responses against schistosomiasis using functionally active cysteine peptidases. *Frontiers in Genetics*, *5*, 1–7.

Fakhar M., Ghobaditara M., Brant S. V., Karamian M., Gohardehi S., Bastani R. (2016). Phylogenetic analysis of nasal avian schistosomes (*Trichobilharzia*) from aquatic birds in Mazandaran Province, northern Iran. *Parasitology International*, *65*, 151–158.

Fan J., Minchella D. J., Day S. R., McManus D. P., Tiu W. U., Brindley P. J. (1998). Generation, identification, and evaluation of expressed sequence tags from different developmental stages of the Asian blood fluke *Schistosoma japonicum*. *Biochemical and Biophysical Research Communications*, *252*, 348–356.

Fedorov A. N., Baldwin T. O. (2001). Cotranslational protein folding. *Molecular Biology*, *35*, 584–590.

Figueiredo B. C. P., Ricci N. D., de Assis N. R. G., de Morais S. B., Fonseca C. T., Oliveira S. C. (2015). Kicking in the guts: *Schistosoma mansoni* digestive tract proteins are potential candidates for vaccine development. *Frontiers in Immunology*, *6*, 1–7.

Fishelson Z., Amiri P., Friend D. S., Marikovsky M., Petitt M., Newport G., Mckerrow J. H. (1992). *Schistosoma mansoni*: Cell-specific expression and secretion of a serine protease during development of cercariae. *Experimental Parasitology*, *75*, 87–98.

Fung M. C., Lau M. T., Chen X. G. (2002). Expressed sequence tag (EST) analysis of a *Schistosoma japonicum* cercariae cDNA library. *Acta Tropica*, *82*, 215–224.

Gazzinelli G., Pellegrino J. (1964). Elastolytic activity of *Schistosoma mansoni* cercarial extract. *The Journal of parasitology*, *50*, 591–592.

Gleeson M. A. G., White C. E., Meininger D. P., Komives E. A. (1998). Generation of protease-deficient strains and their use in heterologous protein expression, In *Pichia protocols*., Humana Press, 81–94.

Gobert G. N., McManus D. P., Nawaratna S., Moertel L., Mulvenna J., Jones M. K. (2009). Tissue specific profiling of females of *Schistosoma japonicum* by integrated laser microdissection microscopy and microarray analysis. *PLoS Neglected Tropical Diseases*, *3*, e469.

Gobert G. N., Moertel L., Brindley P. J., McManus D. P. (2009). Developmental gene expression profiles of the human pathogen *Schistosoma japonicum*. *BMC genomics*, *10*, 128.

Haas W. (2003). Parasitic worms: strategies of host finding, recognition and invasion. *Zoology (Jena, Germany)*, *106*, 349–364.

Hall S. L., Braschi S., Truscott M., Mathieson W., Cesari I. M., Wilson R. A. (2011). Insights into blood feeding by schistosomes from a proteomic analysis of worm vomitus. *Molecular and Biochemical Parasitology*, *179*, 18–29.

Hansell E., Braschi S., Medzihradszky K. F., Sajid M., Debnath M., Ingram J., Lim K. C., McKerrow J. H. (2008). Proteomic analysis of skin invasion by blood fluke larvae. *PLoS Neglected Tropical Diseases*, *2,* e262

Harper E., Berger A. (1972). On the size of the active site in proteases: Pronase. *Biochemical and Biophysical Research Communications*, *46*, 1956–1960.

He Y. X., Chen L., Ramaswamy K. (2002). *Schistosoma mansoni*, *S. haematobium*, and *S. japonicum*: Early events associated with penetration and migration of schistosomula through human skin. *Experimental Parasitology*, *102*, 99–108.

He Y. X., Salafsky B., Ramaswamy K. (2005). Comparison of skin invasion among three major species of *Schistosoma*. *Trends in Parasitology*, *21*, 201–203.

Hedstrom L. (2002). Serine protease mechanism and specificity. *Chemical Reviews*, *102*, 4501–4524.

Hellemans J., Mortier G., De Paepe A., Speleman F., Vandesompele J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genomebiology*, *8*, R19.

Hewitson J. P., Grainger J. R., Maizels R. M. (2009). Helminth immunoregulation: The role of parasite secreted proteins in modulating host immunity. *Molecular and Biochemical Parasitology*, *167*, 1–11.

Hibbard L. S., Tulinsky A. (1978). Expression of functionaly of α-chymotrypsin. Effects of guanidine hydrochloride and urea in the onset of denaturation. *Biochemistry*, *17*, 5460–5468.

Higón M., Cowan G., Nausch N., Cavanagh D., Oleaga A., Toledo R., Stothard J. R., Antúnez O., Marcilla A., Burchmore R., Mutapi F. (2011). Screening trematodes for novel intervention targets: A proteomic and immunological comparison of *Schistosoma haematobium*, *Schistosoma bovis* and *Echinostoma caproni*. *Parasitology*, *138*, 1607–1619.

Holická M. (2009). Avian schistosomes of gastropods and birds in Europe. *Charles University*, *Master's thesis*

Hong Y., Sun A., Zhang M., Gao F., Han Y., Fu Z., Shi Y., Lin J. (2013). Proteomics analysis of differentially expressed proteins in schistosomula and adult worms of *Schistosoma japonicum*. *Acta Tropica*, *126*, 1–10.

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

Horák P., Dvořák J., Kolářová L., Trefil L. (1999). *Trichobilharzia regenti*, a pathogen of the avian and mammalian central nervous systems. *Parasitology*, *119 (Pt 6*, 577–581.

Howe K. L., Bolt B. J., Shafie M., Kersey P., Berriman M. (2017). WormBase ParaSite − a comprehensive resource for helminth genomics. *Molecular and Biochemical Parasitology*, *215*, 2–10.

Hrádková K., Horák P. (2002). Neurotropic behaviour of *Trichobilharzia regenti* in ducks and mice. *Journal of helminthology*, *76*, 137–141.

Hu S., Law P. ki, Fung M. C. (2009). Microarray analysis of genes highly expressed in cercarial stage of Schistosoma japonicum and the characterization of the antigen Sj20H8. *Acta Tropica*, *112*, 26–32.

Hu W., Brindley P. J., McManus D. P., Feng Z., Han Z. G. (2004). Schistosome transcriptomes: New insights into the parasite and schistosomiasis. *Trends in Molecular Medicine*, *10*, 217–225.

Ingram J. R., Rafi S. B., Eroy-Reveles A. A., Ray M., Lambeth L., Hsieh I., Ruelas D., Lim K. C., Sakanari J., Craik C. S., Jacobson M. P., McKerrow J. H. (2012). Investigation of the proteolytic functions of an expanded cercarial elastase gene family in *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases*, *6*.

Jenkins S. J., Hewitson J. P., Jenkins G. R., Mountford P. (2005). Modulation of the host ' s immune response by schistosome larvae. *Parasite Immunology*, *27*, 385–393.

Jiz M. A., Wu H., Olveda R., Jarilla B., Kurtis J. D. (2015). Development of Paramyosin as a Vaccine Candidate for Schistosomiasis. *Frontiers in immunology*, *6*, 347.

Jouet D., Ferté H., Depaquit J., Rudolfová J., Latour P., Zanella D., Kaltenbach M. L., Léger N. (2008). *Trichobilharzia* spp. in natural conditions in Annecy Lake, France. *Parasitology Research*, *103*, 51–58.

Kašný M., Mikeš L., Dalton J. P., Mountford A. P., Horák P. (2007). Comparison of cysteine peptidase activities in *Trichobilharzia regenti* and *Schistosoma mansoni cercariae. Parasitology*, *134*, 1599–1609.

Kašný M., Mikeš L., Hampl V., Dvořák J., Caffrey C. R., Dalton J. P., Horák P. (2009). Chapter 4 Peptidases of Trematodes. *Advances in Parasitology*.

Khan R. H., Appa Rao K. B. C., Eshwari A. N. S., Totey S. M., Panda A. K. (1998). Solubilization of recombinant ovine growth hormone with retention of native-like secondary structure and its refolding from the inclusion bodies of *Escherichia coli*. *Biotechnology Progress*, *14*, 722–728.

Knudsen G. M., Medzihradszky K. F., Lim K.C., Hansell E., McKerrow J. H. (2005). Proteomic analysis of *Schistosoma mansoni* cercarial secretions . *Molecular & Cellular Proteomics*, *4*, 1862–1875.

Kolářová L., Horák P., Skírnisson K., Marečková H., Doenhoff M. (2013). Cercarial dermatitis, a neglected allergic disease. *Clinical Reviews in Allergy and Immunology*, *45*, 63–74.

Korsunenko A. V., Chrisanfova G. G., Ryskov A. P., Movsessian S. O., Vasilyev V. A., Semyenova S. K. (2010). Detection of European *Trichobilharzia* schistosomes (*T. franki*, *T. szidati*, and *T. regenti*) based on novel genome sequences. *Journal of Parasitology*, *96*, 802–806.

Law R. H. P., Law R. H. P., Smooker P. M., Smooker P. M., Irving J. a, Irving J. a, Ponting R., Ponting R., Kennedy N. J., Kennedy N. J., Whisstock J. C., Whisstock J. C., Pike R. N., Pike R. N., Spithill T. W., Spithill T. W. (2003). Cloning and expression of the major secreted cathepsin B-like protein from juvenile. *Microbiology*, *71*, 6921–6932.

Leontovyč R., Young N. D., Korhonen P. K., Hall R. S., Tan P., Mikeš L., Kašný M., Horák P., Gasser R. B. (2016). Comparative transcriptomic exploration reveals unique molecular adaptations of neuropathogenic *Trichobilharzia* to invade and parasitize its avian definitive host. *PLOS Neglected Tropical Diseases*, *10*, e0004406.

Lichtenbergová L., Horák P. (2012). Pathogenicity of *Trichobilharzia* spp. for vertebrates. *Journal of parasitology research*, *2012*, 761968.

Ligasová A., Bulantová J., Sebesta O., Kašný M., Koberna K., Mikeš L. (2011). Secretory glands in cercaria of the neuropathogenic schistosome *Trichobilharzia regenti* – ultrastructural characterization, 3-D modelling, volume and pH estimations. *Parasites & vectors*, *4*, 162.

Liu M., Ju C., Du X. F., Shen H. M., Wang J. P., Li J., Zhang X. M., Feng Z., Hu W. (2015). Proteomic analysis on cercariae and schistosomula in reference to potential proteases involved in host invasion of *Schistosoma japonicum* larvae. *Journal of Proteome Research*, *14*, 4623–4634.

Liu S., Cai P., Hou N., Piao X., Wang H., Hung T., Chen Q. (2012). Genome-wide identification and characterization of a panel of house-keeping genes in *Schistosoma japonicum*. *Molecular and Biochemical Parasitology*, *182*, 75–82.

Livak K. J., Schmittgen T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods*, *25*, 402–408.

Malek A. (1977). Geographical distribution, hosts , and biology of *Schistosomatium douthitti*.

Marcilla A., Trelis M., Cortés A., Sotillo J., Cantalapiedra F., Minguez M. T., Valero M. L., Sánchez del Pino M. M., Muñoz-Antoli C., Toledo R., Bernal D. (2012). Extracellular vesicles from parasitic helminths contain specific excretory/secretory proteins and are internalized in intestinal host cells. *PLoS ONE*, *7(9)*: e45974

McKerrow J. H., Salter J. (2002). Invasion of skin by *Schistosoma* cercariae. *Trends in Parasitology*, *18*, 193–195.

McKerrow J. H., Jones P., Sage H., Pino-Heiss S. (1985). Proteinases from invasive larvae of the trematode parasite *Schistosoma mansoni* degrade connective-tissue and basement-membrane macromolecules. *The Biochemical journal*, *231*, 47–51.

McKerrow J. H., Caffrey C., Kelly B., Loke P., Sajid M. (2006). Proteases in parasitic diseases. *Annual Review of Pathology: Mechanisms of Disease*, *1*, 497–536.

McManus D. P., Loukas A. (2008). Current status of vaccines for schistosomiasis. *Clinical Microbiology Reviews*, *21*, 225–242.

Mikeš L., Zìdková L., Kašný M., Dvořák J., Horák P. (2005). In vitro stimulation of penetration gland emptying by *Trichobilharzia szidati* and *T. regenti* (Schistosomatidae) cercariae. Quantitative collection and partial characterization of the products. *Parasitology Research*, *96*, 230–241.

Neurath H. (1986). The versatility of proteolytic enzymes. *Journal of Cellular Biochemistry*, *32*, 35–49.

Nielsen H. (2017). Predicting secretory proteins with SignaIP. In D. Kihara (Ed.), *Methods in Molecular Biology* (Vol. 1611). Springer New York, 59–73.

Oda K. (2012). New families of carboxyl peptidases: Serine-carboxyl peptidases and glutamic peptidases. *Journal of Biochemistry*. *151* 13–25

Oey H., Zakrzewski M., Gravermann K., Young N. D., Korhonen P. K., Gobert G. N., Nawaratna S., Hasan S., Martínez D. M., You H., Lavin M., Jones M. K., Ragan M. A., Stoye J., Oleaga A., Emery A. M., Webster B., Rollinson D., Gasser R. B., McManus D. P., Krause L. (2019). Whole-genome sequence of the bovine blood fluke *Schistosoma bovis* supports interspecific hybridization with *S. haematobium*. *PLoS Pathogens*, *15*, e1007513

Ojopi E. P. B., Oliveira P. S. L., Nunes D. N., Paquola A., DeMarco R., Gregório S. P., Aires K. A., Menck C. F. M., Leite L. C. C., Verjovski-Almeida S., Dias-Neto E. (2007). A quantitative view of the

transcriptome of *Schistosoma mansoni* adult-worms using SAGE. *BMC Genomics*, *8*, 186.

Parker-Manuel S. J., Ivens A. C., Dillon G. P., Wilson R. A. (2011). Gene expression patterns in larval *Schistosoma mansoni* associated with infection of the mammalian host. *PLoS Neglected Tropical Diseases*, *5*.

Peng H. J., Chen X. G., Wang X. Z., Lun Z. R. (2003). Analysis of the gene expression profile of *Schistosoma japonicum* cercariae by a strategy based on expressed sequence tags. *Parasitology Research*, *90*, 287–293.

Protasio A. V., Tsai I. J., Babbage A., Nichol S., Hunt M., Aslett M. A., de Silva N., Velarde G. S., Anderson T. J. C., Clark R. C., Davidson C., Dillon G. P., Holroyd N. E., LoVerde P. T., Lloyd C., McQuillan J., Oliveira G., Otto T. D., Parker-Manuel S. J., Quail M. A., Wilson R. A., Zerlotini A., Dunne D. W., Berriman M. (2012). A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases*, *6*, e1455.

Prüter H., Sitko J., Krone O. (2017). Having bird schistosomes in mind—the first detection of *Bilharziella polonica* (Kowalewski 1895) in the bird neural system. *Parasitology Research*, *116*, 865–870.

Rawlings N. D., Barrett A. J., Thomas P. D., Huang X., Bateman A., Finn R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*, *46*, 624–632.

Ridi R. E., Tallima H. (2009). Schistosoma mansoni ex vivo lung-stage larvae excretory-secretory antigens as vaccine candidates against schistosomiasis. *Vaccine*, *27*, 666–673.

Roquis D., Lepesant J. M. J., Picard M. A. L., Freitag M., Parrinello H., Groth M., Emans R., Cosseau C., Grunau C. (2015). The epigenome of *Schistosoma mansoni* provides insight about how cercariae poise transcription until infection. *PLoS Neglected Tropical Diseases*, *9*, 1–22.

Rudolfová J., Littlewood D. T. J., Sitko J., Horák P. (2007). Bird schistosomes of wildfowl in the Czech Republic and Poland. *Folia Parasitologica*, *54*, 88–93.

Sajid M., Rogers J., Rajandream M. (2009). The genome of the blood fluke *Schistosoma mansoni*. *Nature*, *460*, 352–358.

Salafsky B., Fusco A. C., Whitley K., Nowicki D., Ellenberger B. (1988). *Schistosoma mansoni*: Analysis of cercarial transformation methods. *Experimental Parasitology*, *67*, 116–127.

Salter J. P., Lim K. C., Hansell E., Hsieh I., McKerrow J. H. (2000). Schistosome invasion of human skin and degradation of dermal elastin are mediated by a single serine protease. *Journal of Biological Chemistry*, *275*, 38667–38673.

Shabaan A. M., Mohamed M. M., Abdallah M. S., Ibrahim H. M., Karim A. M. (2003). Analysis of *Schistosoma mansoni* genes using the expressed sequence Tag approach. *Acta Biochimica Polonica*, *50*, 259–268.

Shiff C. J., Cmelik S. H. W., Ley H. E., Kriel R. L. (2006). The influence of human skin lipids on the cercarial penetration responses of *Schistosoma haematobium* and *Schistosoma mansoni*. *The Journal of Parasitology*, *58*, 476.

Sotillo J., Pearson M., Becker L., Mulvenna J., Loukas A. (2015). A quantitative proteomic analysis of the tegumental proteins from *Schistosoma mansoni* schistosomula reveals novel potential therapeutic targets. *International Journal for Parasitology*, *45*, 505–516.

Sotillo J., Pearson M., Potriquet J., Becker L., Pickering D., Mulvenna J., Loukas A. (2016). Extracellular vesicles secreted by *Schistosoma mansoni* contain protein vaccine candidates. *International Journal for Parasitology*, *46*, 1–5.

Standart N., Weil D. (2018). P-Bodies: Cytosolic droplets for coordinated mRNA storage. *Trends in Genetics*, *34*, 612–626.

Šteiger V. (2018). Molecular diagnostics of bird schistosomes during the infection of natural and accidental

hosts. *Charles University, Master's thesis*

Taft A. S., Vermeire J. J., Bernier J., Birkeland S. R., Cipriano M. J., Papa A. R., McArthur A. G., Yoshino T. P. (2009). Transcriptome analysis of *Schistosoma mansoni* larval development using serial analysis of gene expression (SAGE). *Parasitology*, *136*, 469–485.

Tjalsma H., Bolhuis A., Jongbloed J. D., Bron S., van Dijl J. M. (2000). Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiology and molecular biology reviews : MMBR*, *64*, 515–547.

van Hellemond J. J., van Balkom B. W. M., Tielens A. G. M. (2007). Schistosome biology and proteomics: progress and challenges. *Experimental Parasitology*, *117*, 267–274.

Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N., De Paepe A., Speleman F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology*, *3*, research0034.1–research0034.11

Verjovski-Almeida S., DeMarco R., Martins E. A. L., Guimarães P. E. M., Ojopi E. P. B., Paquola A. C. M., Piazza J. P., Nishiyama Jr. M. Y., Kitajima J. P., Adamson R. E., Ashton P. D., Bonaldo M. F., Coulson P. S., Dillon G. P., Farias L. P., Gregorio S. P., Ho P. L., Leite R. A., Malaquias L. C. C., Marques R. C. P., Miyasato P. A., Nascimento A. L. T. O., Ohlweiler F. P., Reis E. M., Ribeiro M. A., Sá R. G., Stukart G. C., Soares M. B., Gargioni C., Kawano T., Rodrigues V., Madeira A. M. B. N., Wilson R. A., Menck C. F. M., Setubal J. C., Leite L. C. C., Dias-Neto E. (2003). Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. *Nature Genetics*, *35*, 148.

Walker A. J. (2011). Insights into the functional biology of schistosomes. *Parasites & vectors*, *4*, 203.

Wang J., Zhao F., Yu C. X., Xiao D., Song L. J., Yin X. R., Shen S., Hua W. Q., Zhang J. Z., Zhang H. fang, He L. H., Qian C. Y., Zhang W., Xu Y. L., Yang J. (2013). Identification of proteins inducing short-lived antibody responses from excreted/secretory products of *Schistosoma japonicum* adult worms by immunoproteomic analysis. *Journal of Proteomics*, *87*, 53–67.

Wang S., Hu W. (2014). Development of "-omics" research in *Schistosoma* spp. and -omics-based new diagnostic tools for schistosomiasis. *Frontiers in Microbiology*, *5*, 1–11.

Wang T., Zhao M., Rotgans B. A., Strong A., Liang D., Ni G., Limpanont Y., Ramasoota P., McManus D. P., Cummins S. F. (2016). Proteomic analysis of the *Schistosoma mansoni* miracidium. *PLoS ONE*, *11*, 1–22.

Wang W., Kirschfink M., Ruppel A. (2006). *Schistosoma japonicum* and *S. mansoni* cercariae: Different effects of protein in medium, of mechanical stress, and of an intact complement system on in vitro transformation to schistosomula. *Parasitology Research*, *99*, 269–274.

World Heatlh Organization. (2019). Schistosomiasis. https://www.who.int/schistosomiasis/en/.

Yan Y., Liu S., Song G., Xu Y., Dissous C. (2005). Characterization of a novel vaccine candidate and serine proteinase inhibitor from *Schistosoma japonicum* (Sj serpin). *Veterinary Parasitology*, *131*, 53–60.

Yang Y., Wen Y. J., Cai Y. N., Vallée I., Boireau P., Liu M. Y., Cheng S. P. (2015). Serine proteases of parasitic helminths. *Korean J Parasitol*, *53531*, 1–11.

Yeku O., Frohman M. A. (2010). Rapid Amplification of cDNA Ends (RACE). In: Nielsen H. (eds) *RNA. Methods in Molecular Biology*, vol 703. Humana Press

Young N. D., Jex A. R., Li B., Liu S., Yang L., Xiong Z., Li Y., Cantacessi C., Hall R. S., Xu X., Chen F., Wu X., Zerlotini A., Oliveira G., Hofmann A., Zhang G., Fang X., Kang Y., Campbell B. E., Loukas A., Ranganathan S., Rollinson D., Rinaldi G., Brindley P. J., Yang H., Wang J., Wang J., Gasser R. B. (2012). Whole-genome sequence of *Schistosoma haematobium*. *Nature Genetics*, *44*, 221–225.

Zeugen J. A., Hartley J. L. (1985). Ethanol Precipitation of DNA. *Focus*, *7*, 1–2.

Zhang T., Mo X. J., Xu B., Yang Z., Gobert G. N., Yan S., Feng Z., Hu W. (2018). Enzyme activity of *Schistosoma japonicum* cercarial elastase SjCE–2b ascertained by in vitro refolded recombinant protein. *Acta Tropica*, *187*, 15–22.

Zhou Y., Zheng H., Chen Y., Zhang L., Wang K., Guo J., Huang Z., Zhang B., Huang W., Jin K., Dou T., Hasegawa M., Wang L., Zhang Y., Zhou J., Tao L., Cao Z., Li Y., Vinar T., Brejova B., Brown D., Li M., Miller D. J., Blair D., Zhong Y., Chen Z., Liu F., Hu W., Wang Z. Q., Zhang Q., Song H. D., Chen S. S., Xu X., Xu B., Ju C., Huang Y., Brindley P. J., McManus D. P., Feng Z., Han Z. G., Lu G., Ren S., Wang Y., Gu W., Kang H., Chen J., Chen X., Chen S. S., Wang L., Yan J., Wang B. B., Lv X., Jin L., Wang B. B., Pu S., Zhang X., Zhang W., Hu Q., Zhu G., Wang J. J., Yu J., Wang J. J., Yang H., Ning Z., Beriman M., Wei C. L., Ruan Y., Zhao G., Wang S., Liu F., Zhou Y., Wang Z. Q., Lu G., Zheng H., Brindley P. J., McManus D. P., Blair D., Zhang Q., Zhong Y., Wang S., Han Z. G., Chen Z., Wang S., Han Z. G., Chen Z., Consortium F. A. (2009). The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature*, *460*, 345–351.

10. Attachment

10.1. Completion of *T. regenti* cercarial elastase whole gene sequence

Table 46. Partial genomic sequence of *T. regenti* cercarial elastase

***T. regenti* cercarial elastase partial genomic sequence**
ATGCACACTGAAACAATGATTCCTACTTCGTCGCTGCTCTTCTCATTGTTACTGCTATGCAATCTA GTCTCTCTGAGTGTATCAACATGGTTAGTTCGTAGTGGAGAGACAGTGAGAGATAAGAATGAATT CCCGTTTCTCGCTCTCATGGTGACAGACAGTTCGATGTGTACAGCGACGCTGGTGTCCACCCGAG CTGTGATCACAGCTGGTCATTGTGTGTGTGGACCGAAATCAATCAACCGGGTGAGTGAGCATTGC TCTTCGGATGATGGTGATTGTGTTTTGAGATTTTTCGTGTTGTGCGTATATTTTTCTGACTTCCTTCC TATTTCTGTCTTGTTCAGATCTCCTTCCAGACACTGA

10.2. qPCR analysis of *T. regenti* cathepsin B2 and cercarial elastase

Table 47. *T. regenti* PSMD4 genomic sequence

***T. regenti* PSMD4**
ATGAGGAATGGGGATTTCTTCCCATCTCGTTTACAAGCACAGAATGATGCTGTAAGCTTGATTT GTCAAAGCAAACGCCAACGCAATCCCGAAAATACACTTGGTTTGTTGTCTCTTGCCAACACAG AGGTTTTGTGTACATTGACAAATGATGTGAGCAAAATATATAATCGTCTGCACCTTGTGGAAC CAAAAGGAAGTATTATTTTCTGCTCATCCATCAGGATTGCACACTTAGCCCTTCGTCATCGTCA ATTGAGGCATCAGAAAATGAGGATTGTGTGTTTTATCGGCAGCCCTATTATAGAAGACGAAAA AGAATTGATTAAACTCGCTAAACGACTCAAGAAAGAGAAAGTCAATGTGGACATTATTAATTT CGGTGAAAATGAAGTGAATGAGAAAAAATTGTCGGAGTTCATTGATACACTGAATGGGAAAG ATGGCACAGGTTCTCATCTTGTTTCTATTGCACCAGGGACAGTTTTGCATGATACTTTGGTGAC AAGTCCTATTGTTGCTGGCGAAGATGGTTCTGGACTTGCTGGTGCAGGTTTAGGACTAGAATTT GGATTAGATGCCGCAGAAGATCCAGATCTCTTATACGCACTCAGAGTATCTATGGAAGACCAG CGTATGAGACAAGAACATGAGGTTAATGCAGACGCTGGTAATGCGCCTACTGCAGCCACTTCT TTACCAGCCGGGTCAGGTACTTCTGAAGAGGTTATGCTTCAACAGGCTTTGGCTATGTCAATG CAAATGGATAATGCAGGATCGGCAGGATTACCTATGGATATTGATTTGGCAGCAATGTCTGAA GAAGATCAAATCGCCTATGCACTGCGTATGTCTCTTCAACAAATGGGAGAAGAAACTGCACA ACCAACTGCTACAGAGTCGAATAAGACTAATGTGGAAGCCTCGCCGATTACAACAGCTATGG CTATGGATGTAGATCAAACACCAACTAAAGCTGTAGAAAAGACAGAAGGATCTACAGTCTCT GGGGGATTGCCAACGGATACTGCTGCCGCCGTTACCACTACGAGCTCAGCAGCAGCACCATCG ACAACAACAACAACAGTTCATGCACCTGCTGATTTAGATGTTATCTATGATGTAGCATTTATTG AATCTGTTCTACAAAGTCTCCCTGGTGTAGATACACAAAATGAAGATGTACGCAAGGCGATAA GTGCACTGACTAAATCTCAATCTCAAGGGGGTGGTGCTGGTGGTGGCGGTAGTTCATCATCAA ATGATCATGATGAAGAGAATAAGAATAAGAAGATGAAGACGGCTGAGGAAGATGATAAACA GGATAAATGA

Table 48. *T. regenti* NDUFV2 genomic sequence

***T. regenti* NDUFV2**
ATGGTTAATTTTTATGTACTCGTAAGGCACTATTATTTAGGTTGGTTGCCAATATCTGCAA TGAACAAGGTTGCTGAAATTCTTAATGTCCCAGCAATGAGGGTATATGAAGTGGCTACCT TCTACACTATGTTTAACAGGGAACCCGTGGGAAAATATCATATTCAGATCTGTACTACTAC CCCGTGCATGTTGGGCGGTGCAGGCTCAGATGTTATTTTGGACGCCCTGAAGGAAAATCT CGGTATTGAACCAGGTCAAACAACAGAAGATAAAATGTTCACTCTGTCAGAGGTAGAATG TCTTGGAGCTTGTGTTAACGCACCAATGATGCAGATCAATGATGACTATTATGAGGACTTA ACAGTTGAAGATACAGTTCGTATTCTTGAAGAACTAAAAGCTGGAAAAAAGCCAAAACCT GGACCACAGAGTGGACAAGGTGGACGTTTTGCATCAGAACCAAAAGGTGGTTTAACCTCA CTGACAACAGATCCTAAAGGTCCTGGATTCAAAGTTCGTTCGGATTTATAA

Table 49. *T. regenti* TPC2L genomic sequence

***T. regenti* TPC2L**
ATGGCTGCCTGCGTCGTAGTTATTTCAGATACTAATCAGCTACTCTACCTTCGAACCGCTG AATGTCCTGATCCCCTGTTTTACCATTTCAAAGCACATTCAGCTTTAGATGTTATAGAGGA CAAGTTATCCAAGCGTACAACTAGTGGTAGTCCTGATCAACTGGAACAGTATCTTGGTCTT TTATACCCAATGGAAGATCATAGGATTTATGGTTATGTAACCAATACAAAAGTTAAGTTT ATTATGATCTTTGAGTCACAGAATCTGTCGCCCTCACAGTCAACTCAAACAAGCCTACCA AGTCATCAGCATAATCATCATATACGTGATGTAAATATTCGTGATATGTTTCAACGATTGC ATACAGCATATATCGATTTGGTGTGTTCACCATTCTATAAATCAGGCACACCGATTCAACC AGACCAGCTACCAGCTGCACGTCGATTCGATCAGCGTATCTCAACACTTTTAGCTACACAT GCTAGTTCTGGTTATTGTTCAACATTTACTAGCCTGATTGATGAAGCAAAAACACCGACTG GTGAAAGTTTAACAACAAGTATAATCTCCCCTAAAGCGTGA