Charles University

Faculty of Science

Study programme: Biology

Field of study: Parasitology

Bc. Kristýna Peterková

Localization and quantification of mRNA coding digestive peptidases of *Fascioloides magna*

Lokalizace a kvantifikace mRNA kódující trávicí peptidázy *Fascioloides magna*

Master's thesis

Supervisor: RNDr. Martin Kašný, Ph.D.

Prague, 2019

Prohlášení: Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla 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

Kristýna Peterková

Acknowledgements

First of all, I wish to acknowledge the great help, enthusiastic guidance and patient useful consultations provided by my supervisor Dr Martin Kašný. Also, I would like to thank Hanka Dvořáková and Lucka Jedličková for their helpful assistance and advice during RNA *in situ* experiments. Furthermore, I wish to acknowledge Vláďa Šteiger for introducing me to quantitative PCR and Roman Leontovyč for providing me with various transcriptomic data and consultations about them. My grateful thanks are also extended to all members of the Laboratory of Helminthology along with its leader Prof. Petr Horák, who created friendly, helping and fun working environment, and of course to our technician Verča Harsová for her laboratory support. In addition, I wish to express my special thanks to Prof. John P. Dalton and to members of his laboratory in Belfast, for an unexceptionally friendly welcome and valuable opportunity to gain research experience in the frame of an international internship, which has greatly helped to enrich this thesis. This acknowledgement would not be complete, without expressing my thanks to a friend and lab colleague Lukáš Konečný, who has been my supportive fellow during the whole master study. Finally, I wish to thank my parents for their endless support and encouragement throughout my studies.

Abstract

Trematode peptidases are important molecules responsible for biocatalysis in many basal biological processes and are crucial in host-parasite interactions. Therefore, these enzymes are intensively studied in order to characterize their biological functions and to use them as potential diagnostic or therapeutic targets. Lately, investigation of transcriptome and secretome revealed, that adult *Fascioloides magna* (giant liver fluke) expresses and secretes a variety of peptidases. Thus, this thesis focuses on three newly identified enzymes: cathepsin L (FmCL), cathepsin B (FmCB) and cathepsin D (FmCD). In other trematode species, these cathepsins are being linked mainly with the digestion of host blood. We applied quantitative PCR (qPCR) to investigate relative expression levels of the three enzymes among three developmental stages – egg, miracidium and adult. It was revealed that all cathepsins have the highest expression in adult flukes in comparison to eggs and miracidia. We also localized the place of transcription of FmCL, FmCB and FmCD in adult fluke using RNA *in situ* hybridization. All of the peptidases were detected in gastrodermis, and in addition, they were localized in the reproductive system. The latter surprising finding is suggesting that these enzymes might have multiple functions in adult *F. magna*, not only in digestion but also in the development of cells within the reproductive system, such as vitelline cells, spermatozoa and oocytes.

Keywords: Fasciolidae, *Fascioloides magna*, trematode, peptidase, cathepsin, digestion, qPCR, RNA *in situ* hybridization, reproduction

Abstrakt

Peptidázy motolic jsou důležité molekuly zodpovědné za biokatalýzu v řadě bazálních biologických procesů a jsou klíčové pro interakci mezi hostitelem a parazitem. Tyto enzymy jsou proto intenzivně studovány za účelem charakterizace jejich biologických funkcí a jejich využití jako potenciálních diagnostických nebo terapeutických cílů. Nedávné výsledky transkriptomické a sekretomické studie ukázaly, že dospělci *Fascioloides magna* (motolice obrovské) exprimují a sekretují řadu peptidáz. Tato práce se proto zaměřuje na tři nově identifikované enzymy: katepsin L (FmCL), katepsin B (FmCB) a katepsin D (FmCD). U jiných druhů motolic jsou tyto katepsiny spojovány především s trávením hostitelské krve dospělci. Za použití kvantitativní PCR (qPCR) jsme stanovili hladiny relativní exprese mRNA těchto peptidáz mezi třemi vývojovými stádii – vajíčkem, miracidiem a dospělcem. Bylo zjištěno, že všechny katepsiny mají nejvyšší expresi u dospělých motolic ve srovnání s vajíčky a miracídii. Také jsme lokalizovali místo transkripce FmCL, FmCB a FmCD u dospělých motolic pomocí RNA *in situ* hybridizace. Všechny peptidázy byly detekovány v gastrodermis a navíc byly lokalizovány ve tkáních reprodukčního systému. Tento překvapivý nález naznačuje, že tyto enzymy mohou mít v biologii dospělé motolice více funkcí, a to nejen při trávení, ale také při vývoji buněk reprodukčního systému, jako jsou například vitelinní buňky, spermie a oocyty.

Klíčová slova: Fasciolidae, *Fascioloides magna*, motolice, peptidáza, katepsin, trávení, qPCR, RNA *in situ* hybridizace, reprodukce

Table of contents

Abbreviations:

AP	alkaline phosphatase
ATP5B	ATP synthase subunit beta
BACT	beta-actin
BMGY	buffered glycerol-complex medium
BMMY	buffered methanol-complex medium
CB	cathepsin B
CD	cathepsin D
cDNA	complementary DNA
CL	cathepsin L
CP	cysteine peptidase
DEPC	diethyl pyrocarbonate
DIG	digoxigenin
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ESPs	excretory/secretory products
FmCB	*Fascioloides magna* cathepsin B
FmCD	*Fascioloides magna* cathepsin D
FmCL	*Fascioloides magna* cathepsin L
GAPDH	G lyceraldehyde 3-phosphate dehydrogenase
GOI	gene of interest
Hb	haemoglobin
ISH	RNA *in situ* hybridization
mRNA	messenger RNA
NEJ	newly excysted juvenile
NF	normalization factor
NTC	no template control
PCR	polymerase chain reaction
SDS	sodium dodecyl sulfate
qPCR	quantitative polymerase chain reaction
rpm	revolutions per minute
RT	room temperature
TBP	TATA-binding protein
TBS	tris-buffered saline
TBST	mixture of tris-buffered saline with Tween 20
TUBBI	beta-tubulin
UBC	ubiquitin C
YPD	yeast extract peptone dextrose
YWHAZ	14-3-3 protein zeta/delta

1 Introduction and aims

Helminth peptidases are key proteins which are responsible for many biological processes within various developmental stages of parasites. They are also crucial in the frame of host-parasite interactions. To date, quite a number of peptidases have been identified and many of them have been intensively studied to reveal their functions. The research of helmith peptidases aims not only to characterize and understand their biological roles but also to use this knowledge for the development of new drugs and vaccines which would interfere with the functioning of peptidases (Kašný *et al*., 2009).

Fascioloides magna (giant liver fluke) belongs to the family Fasciolidae and it is a digenean blood-feeding trematode which infects ruminants in Europe and North America (Erhardová-Kotrlá, 1971). Although peptidases of closely related species such as *Fasciola hepatica* and *Fasciola gigantica* are relatively well described, there is a lack of information about peptidases of *F. magna*. The cornerstone and the main source of information about the molecular biology of *F. magna* are represented by a single transcriptomic and proteomic study of an adult fluke. It has been revealed that adult worm expresses a great number of peptidases and that many of them share significant similarity with peptidases that have been identified in related flukes (Cantacessi *et al*., 2012). We have therefore decided to shed light on three selected peptidases detected in *F. magna*. Two of them represent first and second most abundant peptidases found in the transcriptomic and secretomic data – cathepsin L and B, respectively. The third selected peptidase is less abundant cathepsin D, well known digestive peptidase in blood flukes (schistosomes), but almost not studied in fasciolid trematodes so far (Caffrey *et al*., 2004). According to the literature about these cathepsins, their main function in adult fluke is to degrade host blood proteins. In addition, cathepsins have diverged into many variants and it has been described, that particular cathepsins are developmentally regulated, for example, that infective larval stages express different cathepsins than adults, and that they use them for different purposes, such as excystment, penetration and migration through the host body (Cwiklinski *et al*., 2019).

Aims of this master's thesis are:

- to summarize current knowledge about peptidases of *F. magna* and related species with the focus on two cysteine peptidases – cathepsin L and B, and one aspartic peptidase – cathepsin D, and to report about available methods for studying gene expression in trematodes
- to measure expression levels of the three selected peptidases and compare them among various developmental stages of *F. magna*
- to localize place of transcription of the three cathepsins in the tissue of adult worm using RNA *in situ* hybridization method
- to express recombinant cathepsin B in the yeast expression system and subsequently characterize its biochemical properties

2 Literature review

The giant liver fluke, *Fascioloides magna* (Bassi, 1875), is a digenean trematode that belongs to the family Fasciolidae (Railliet, 1895) (Table 1). It infects a wide spectrum of wild and domestic ruminants. Definitive hosts are represented mostly by red deer (*Cervus elaphus*) in Europe and white-tailed deer (*Odocoileus virginianus*) in North America (Pybus, 1990). The obligatory intermediate host is an aquatic snail; in Europe, it is primarily *Galba truncatula* or less often *Radix* sp. (Erhardová-Kotrlá, 1971; Faltýnková *et al.*, 2006; Kašný *et al.*, 2012).

Originally, this parasite occurred in North America, but in 1865 it was introduced along with wapiti from Wyoming to Turin in Italy (Swales, 1935; Apostolo, 1996). In addition to this area, in 1910 *F. magna* was also reported in south-western Bohemia (Ullrich, 1930). Phylogenetic and statistical methods show that *F. magna* was introduced to Europe at least twice, first to Italy and then to the Czech Republic, from where it further spreads to Slovakia, Austria, Germany, Hungary, Poland and Croatia (Kašný *et al*., 2012).

Table 1**: Taxonomic classification of *F. magna.***

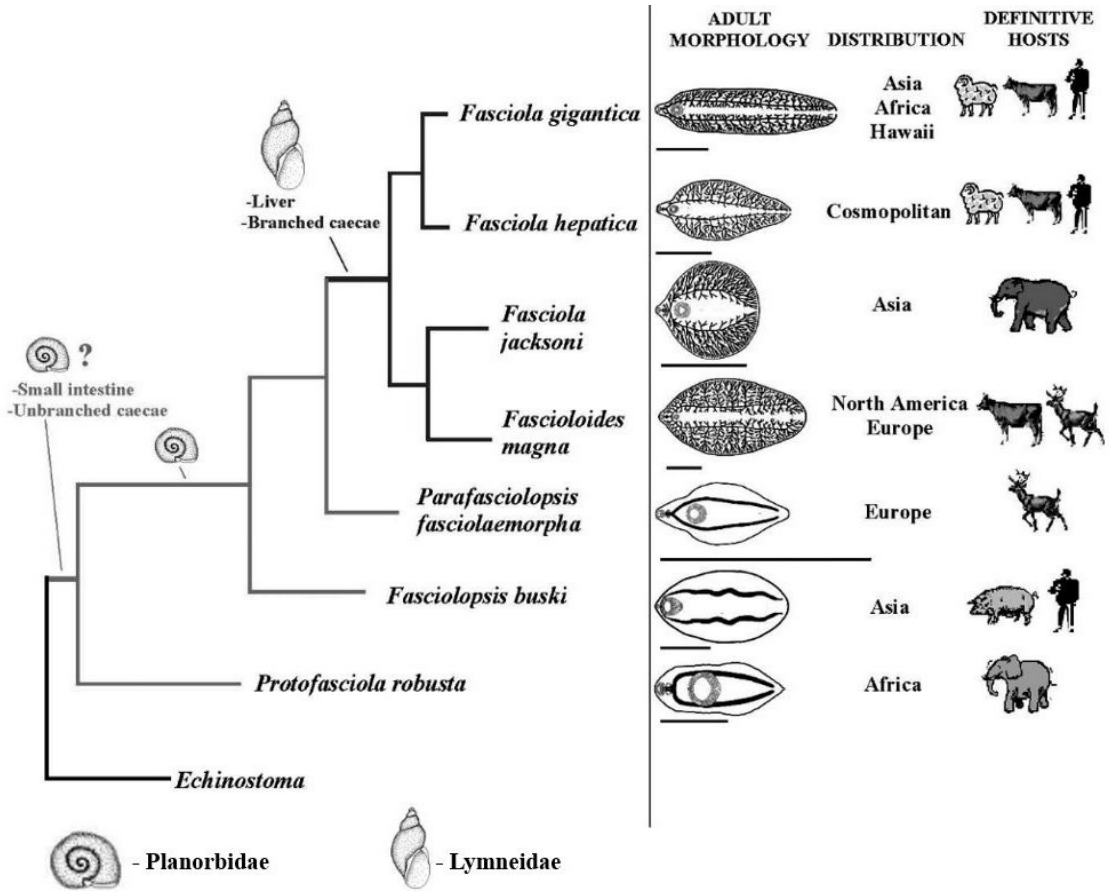
Kingdom	Animalia
Phylum	Platyhelminthes
Class	Trematoda
Order	Plagiorchiida
Family	Fasciolidae
Subfamily	Fasciolinae
Genus	*Fascioloides*
Species	***Fascioloides magna***

2.1 Systematics of the family Fasciolidae

The family Fasciolidae is a relatively small monophyletic group of trematodes that includes only 9 species (Figure 1). They usually infect large herbivores, however, *Fasciola hepatica* (Linnaeus, 1758), *Fasciola gigantica* (Cobbold, 1856) and *Fasciolopsis buski* (Lankester, 1857) frequently infect people and cause serious diseases (Lotfy *et al*., 2008).

F. magna is grouped with *F. hepatica*, *F. gigantica* and *Fasciola jacksoni* (Cobbold, 1869) into subfamily Fasciolinae (Railliet, 1895). Those species share many morphological features, such as branched intestinal caeca, dendritic testes, dendritic ovaries and branched vitellaria (Erhardová-Kotrlá, 1971). Lately, a study by Tkach *et al*. (2016) suggested that *F. magna* should be reclassified and placed into genus *Fasciola*, according to molecular phylogeny and because of rather feeble biological differences between *Fascioloides* and *Fasciola* genera.

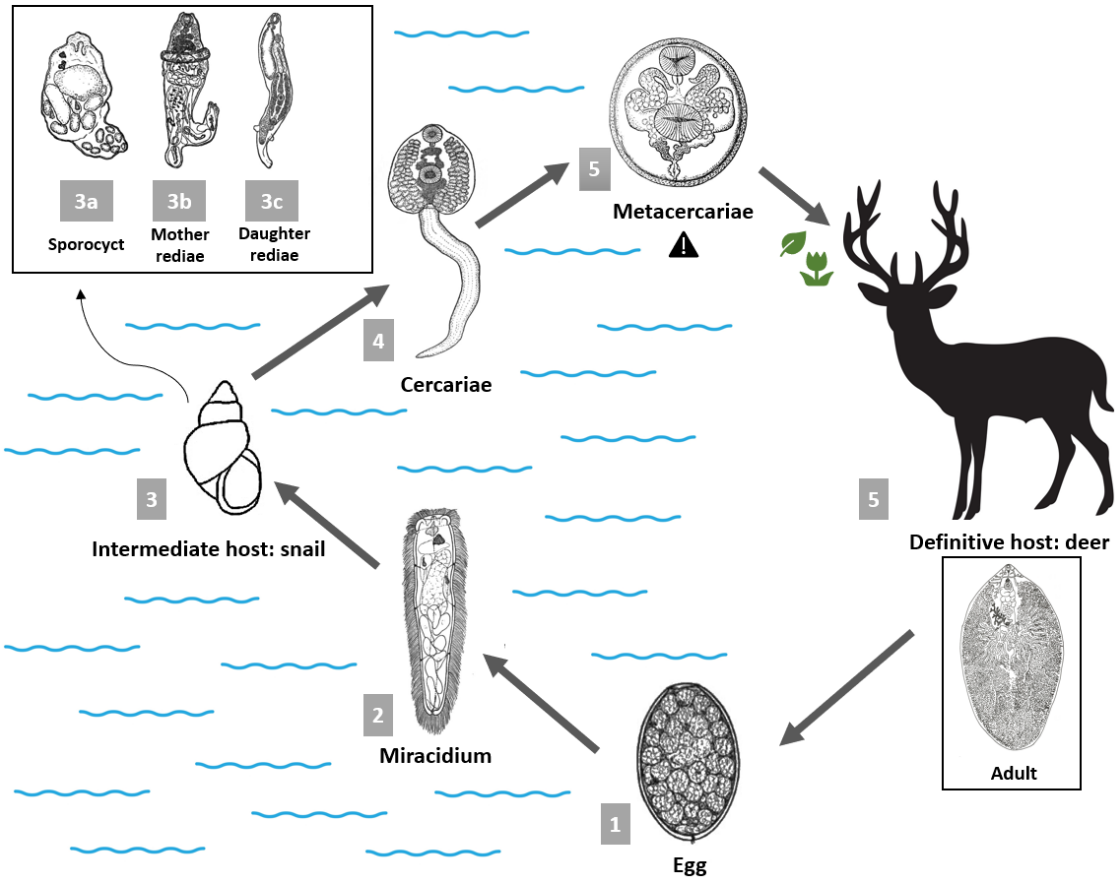
Figure 1: **Evolutionary tree of the family Fasciolidae**. Edited, original: Lotfy *et al*., 2008.



2.2 Life cycle and definitive hosts

The life cycle of *F. magna* is complex and consists of many developmental stages which have to survive in an outer environment and within two different hosts (Figure 2). Adult flukes are typically localized within the liver tissue of an infected animal. They can be usually found in groups of two or more individuals inside a fibrous pseudocyst which is formed as a defence reaction by the host. There, flukes produce eggs which are excreted through bile ducts and intestine to the outside environment. Within a month, larva called miracidium develops inside the egg and hatches. Free-swimming miracidium actively searches in water for the freshwater snail, then it penetrates into snail body and transforms into a larva called the sporocyst. This larva produces more larvae termed as mother rediae, which subsequently give rise to many daughter rediae. Finally, each daughter redia produces several cercariae which then emerge from the snail back to the aquatic environment. Altogether, the multiplication of a single miracidium may result in more than 1000 cercariae (Pybus, 1990). Afterwards, cercariae encyst on rocks or water vegetation as metacercariae. The definitive hosts (e.g. deer) become infected by drinking water or eating plants infested by metacercariae. After that, metacercariae are activated within the host gut and excysts. The newly excysted juveniles (NEJs) penetrate the intestinal wall, migrate to the liver, grows and fully develop into young flukes which search for one another to become encapsulated together in the pseudocyst (Erhardová-Kotrlá, 1971; Foreyt and Todd, 1976; Foreyt *et al*., 1977).

Figure 2: **The life cycle of *F. magna*.** (Source of schematic drawings 1-5: Swales, 1935; source of adult worm drawing: Jones *et al*., 2005; source of the snail and deer picture: www.vectorstock.com)



The progress and consequences of infection depend on the type of the definitive host. Historically, three groups of hosts are described for *F. magna*: specific definitive hosts, aberrant hosts and dead-end hosts. The *specific* definitive hosts are those in which *F. magna* matures within thin-walled pseudocysts in the liver and produces eggs that enter the outside environment through the bile ducts and intestine. Most of such infections are subclinical and chronic (Pybus, 1990). However, if the animal is infected with a greater number of flukes, extensive destruction of liver tissue, peritonitis, bleeding and even death may occur. This group includes, for example, red deer, fallow deer, white-tailed deer, wapiti, caribou, black-tailed deer and mule deer. In terms of *aberrant* host infections, the course and ultimate consequences are much more complicated. This concerns mainly sheep, goats and roe deer. In these cases, juvenile worms will never complete migration and do not induce the formation of pseudocysts. This results in extensive damage to the liver (up to hundreds of worms in pseudocysts in a single liver) and sometimes to other organs in the body (e.g. lungs). If untreated, such infections end always with death, which usually occurs 4-6 months after ingesting metacercariae (Pybus, 1990). In *dead-end* hosts, juvenile worms migrate to the liver, but then, thick-walled pseudocysts are formed around them. That usually prevents the release of eggs into the environment. Dead-end hosts are for instance cattle, moose, horses and wild boars (Pybus, 1990; McClanahan *et al*., 2005).

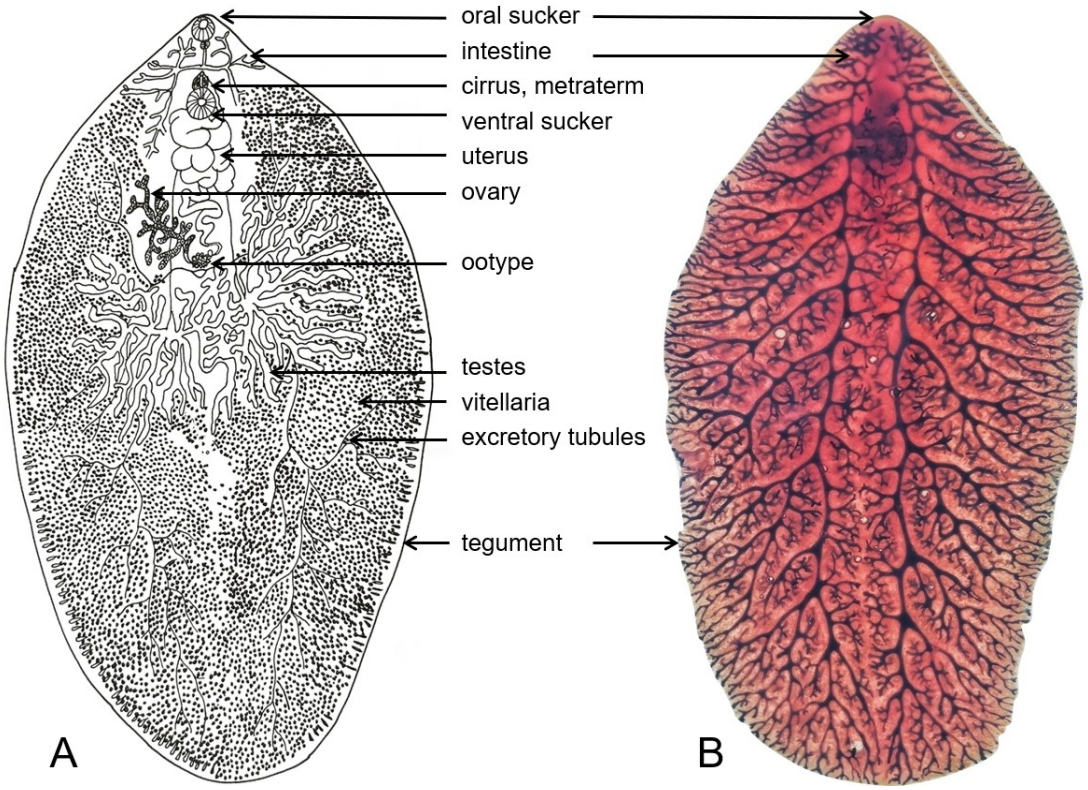
2.3 Characterisation of selected life stages

This chapter includes a basic characterization of selected life forms of *F. magna* – adult, egg, miracidium, which are relevant to the research part of the thesis. For the description of other developmental stages see Swales (1935) and Erhardová-Kotrlá (1971).

2.3.1 Adult fluke

Adult *F. magna* is dorsoventrally flattened and leaf-shaped. Its body is up to 8 cm long and about 3 cm wide (Figure 3). It has two cup-shaped muscular suckers. The oral sucker is located at the very front of the body and forms the beginning of the digestive system, the second one, ventral sucker (also called acetabulum), is larger and is situated on the ventral side in the first third of the body (Erhardová-Kotrlá, 1971). The external covering of *F. magna* forms an interface with the host environment and it is a site where biochemical, physiological and immunological interplay takes place (Smyth and Halton, 1983). The body surface is formed by syncytial epithelium called tegument which is equipped with papillae and spikes of various shapes (Naem *et al*., 2012). Under the basal lamina, there is a circular musculature, under it, there are longitudinal muscles and the third layer is formed by diagonal muscles (Trailović *et al*., 2015). The basic skeleton of the nervous system consists of paired cerebral ganglia, longitudinal cords and commissures. The excretory system is of a protonephridial type based on the flame cells which are distributed between parenchymal cells, connected with branched tubules which converge into two main excretory canals. Those canals end up in a single excretory pore which is located on the surface of the posterior and dorsal side of the body (Erhardová-Kotrlá, 1971; Naem *et al*., 2012). The digestive system starts with a mouth surrounded by the muscular oral sucker, continues with a short pharynx and then with an oesophagus. The gastrointestinal tract is then divided into the massively branched intestine that virtually goes through the entire body (Erhardová-Kotrlá, 1971). The gut epithelium (also called a gastrodermis) is formed by a layer of individual enterocytes that possess long microvilli on their apical side (Košťáková, 2013). Mature *F. magna* is a hermaphrodite; therefore, it has both female and male reproductive organs. The male system is composed of a pair of branched testes, vas deferens, seminal vesicle, prostatic glands, bursa cirri and cirrus. The female system consists of a single branched ovary (also termed as germarium) which is located on the left side (from the dorsal view) and in the first third of the body. An oviduct leads oocytes to an ootype which is connected to Mehlis' glands. Richly branched vitelline glands (vitellaria) produce vitelline cells which provide eggshell and "yolk" material for eggs. In the uterus, eggs are formed and leave the body through a metraterm (ending of the uterus) (Erhardová-Kotrlá, 1971).

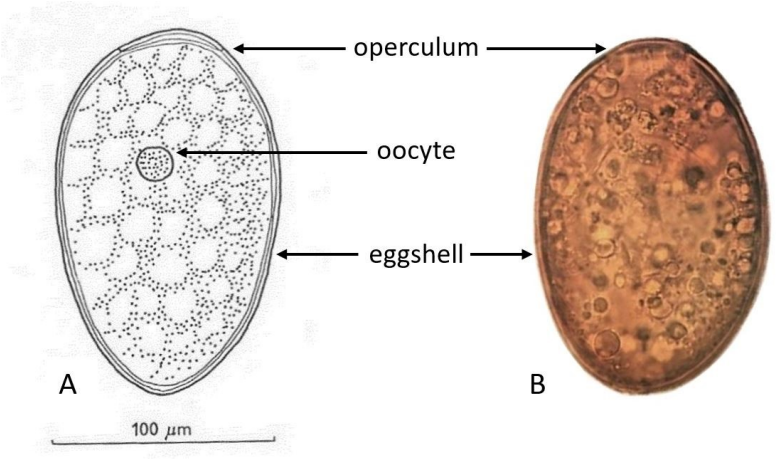
Figure 3: **A: Scheme of adult *F. magna*** (modified according to Jones *et. al*., 2005); **B: Photo of the adult worm from dorsal view.**



2.3.2 Egg

The egg of *F. magna* measures 124 – 175 µm in length and 81 – 117 µm in width, it is oval, smooth on the surface, yellowish brown in colour and closed with a lid called an operculum (Figure 4) (Erhardová-Kotrlá, 1971). Freshly laid egg contains one undivided germ cell (oocyte) and several vitelline cells (Swales, 1935). The latter provides shell precursors for the process of egg formation within ootype and also yolk as a source of nutrients for subsequent development (Smyth and Halton, 1983). The egg embryogenesis is temperature and humidity dependent. If the egg is constantly at temperature around 21 °C the morula stage originates the 8^{th} – 10^{th} day of incubation, and miracidium terminates the development on the 17^{th} – 21^{st} day. Hatching is stimulated by natural or artificial light and occurs 28^{th} – 30^{th} day (Erhardová-Kotrlá, 1971). Nevertheless, Swales (1935) observed that the optimal incubation temperature for the fastest maturation (14 – 21 days) of miracidia in laboratory conditions was between 24 and 28 °C. The temperature at 2 – 5 °C stops the development of the egg, which allows their storage for the purpose of laboratory experiments, and they can be deposited in this way at least nine months. (Swales, 1935).

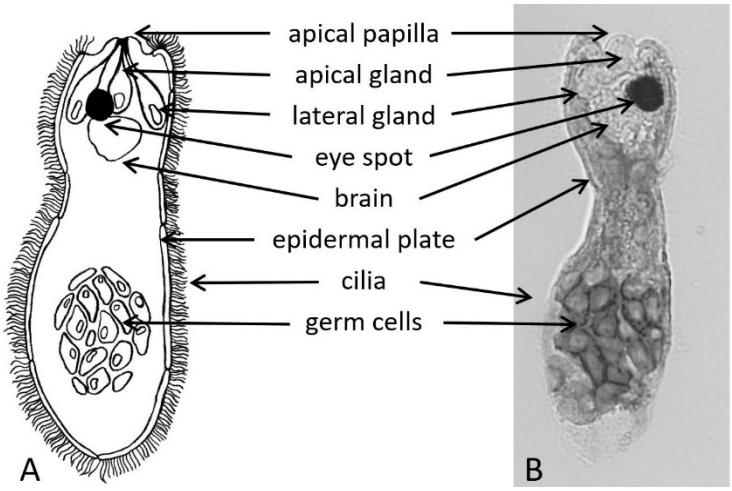
Figure 4: ***F. magna* egg. A:** scheme of an unembryonated egg (edited, source: Erhardová-Kotrlá, 1971); **B:** Photo of *F. magna* egg retrieved from the infected liver.



2.3.3 Miracidium

After hatching, miracidium is actively swimming in the aquatic environment and searching for a suitable freshwater snail. It is pyriform shaped larva and its surface consists of 21 flattened ciliated epithelial cells (Figure 5). In the area between those cells, in intercellular ridges, there are ciliated nerve endings and openings of the excretory system. In the first half of the body, there is an apparent black eye-spot serving as a photoreceptor, and central ganglion connected to muscles and sense endings via nerve fibres. At the anterior part of the body, there is also a retractable structure called the terebratorium (or the apical papilla) which bears openings of one apical and two lateral glands. Those glands contain material which probably helps with the attachment and penetration into the intermediate host (Coil, 1977, 1981). It is also worth mentioning that miracidium has no digestive system and is, therefore, completely dependent on the glycogen energetic supply. This has been experimentally demonstrated on miracidia of closely related species *F. hepatica* (Boyunaga *et al*., 2001). Finally, the posterior half of the body is filled up with several germ cells (Erhardová-Kotrlá, 1971).

Figure 5: **A: Scheme of *F. magna* miracidium; B: Section of *F. magna* miracidium.**



2.4 Peptidases of trematodes

2.4.1 Definition of peptidase

A peptidase, also termed as a proteolytic enzyme or protease, is a bioactive macromolecule which mediates cleavage of peptide bonds within a protein chain. Most proteolytic enzymes are either exopeptidases cleaving one or more amino acids from the N- (aminopeptidases) or C-terminus (carboxypeptidases) of the polypeptide, or endopeptidases that act inside the polypeptide chain. Endopeptidases are then primarily sorted to serine endopeptidases, cysteine endopeptidases, threonine endopeptidases, aspartic endopeptidases, metalloendopeptidases and glutamic endopeptidases based on their catalytic mechanism. The system of the peptidase classification by Rawlings and Barrett (1993) is further dividing them into families and clans according to the significant similarity in the primary structure (amino acid sequence) and the tertiary structure (three-dimensional). The definition of a unique peptidase is not trivial, but in general, a single peptidase is a set of proteins all of which display a particular kind of peptidase activity, and are closely related in sequence (Rawlings and Salvesen, 2012). The most up-to-date information on individual peptidases (including trematode peptidases), families and clans can be found in online MEROPS database (www.ebi.ac.uk/merops).

Table 2: **Example of peptidase classification for cathepsin L1.** (www.ebi.ac.uk/merops)

Clan	CA
Family	C1
Subfamily	A
Accession number	C01.033
MEROPS name	cathepsin L1 (*Fasciola* sp.)
Catalytic type	cysteine

2.4.2 Trematode peptidases

Trematode peptidases belong to key molecules responsible for biocatalysis in many basal biological processes and are crucial in host-parasite interactions. Trematodes use peptidases, for instance, during host penetration, body migration, nutrition acquisition and host immune modulation (Kašný *et al*., 2009). Therefore, these enzymes are intensively studied to characterize their functions and to use them as potential diagnostic or therapeutic targets. The main focus has been placed particularly on secreted peptidases found in excretory/secretory products (ESPs) as they play a pivotal role in immediate host-parasite interplay.

An extensive source of peptidase related data has been recently provided by studies investigating parasite genomes, transcriptomes, proteomes and secretomes ("omics"). This approach has become routinely used as the sequencing technology is now relatively affordable. Lately, there has been an increasing number of bioinformatic studies focused on fasciolid trematodes (Table 3). Results of relevant bioinformatic data evaluation provide a useful source for further characterization of a gene (and protein) functions, understanding of

how helminth regulation of gene expression during their life cycle leading to discoveries and development of new drugs and vaccines against parasitic infections.

Table 3: **Available "omics" studies of the family Fasciolidae.**

Species	Genome	Transcriptome	Secretome/Proteome
F. hepatica	McNulty *et al*. (2017) Cwiklinski *et al*. (2015)	Cancela *et al*. (2010) Young *et al*. (2010) Cwiklinski *et al*. (2015) Cwiklinski *et al*. (2019)	Di Maggio *et al*. 2016) Cwiklinski *et al*. 2019) Wilson *et al*. (2011) Hernández-González *et al*. (2010) Robinson *et al*. (2009)
F. gigantica	Tripathi *et al*. (2018)	Young *et al*. (2011)	Huang *et al*. (2019)
F. magna	not available	Cantacessi *et al*., (2012)	Cantacessi *et al*., (2012)
F. buski	Biswal *et al*. (2018)	Biswal *et al*. (2018)	not available

In the case of *F. magna*, there is only one study so far, investigating transcriptome and secretome of an adult fluke by Cantacessi *et al*., (2012), and it is also the only invaluable source of information about peptidases of *F. magna*. The ESPs of *F. magna* adult are rich in peptidases, they make around 16% of the identified proteins (Table 4). There were even more peptidases identified in the transcriptome (e.g. cathepsin D, cathepsin L3, cathepsin B-like protease), but possibly due to low expression levels, they were not detected by the proteomic analysis. More than half of the identified mRNA molecules in the transcriptome shared significant similarity with those found in *F. hepatica* and *F. gigantica* transcriptomes (Cantacessi *et al*., 2012). In addition, peptidases detected in ESPs of *F. magna*, such as cathepsin L or cathepsin B, show significant homology to *F. hepatica* peptidases (Cantacessi *et al*., 2012).

Table 4: **Peptidases identified in ESPs of adult *F. magna*.** (Cantacessi *et al*., 2012)

Peptidase	Predicted signal peptide	Putative localization
cathepsin B2	yes	extracellular
cathepsin L1	yes	extracellular
cathepsin L-like proteinase	yes	extracellular/plasma membrane
cathepsin A	yes	extracellular/lysosomal
calpain	no	cytoplasm
legumain	yes	lysosomal/extracellular
leucine aminopeptidase	no	cytoplasm/mitochondria
lysosomal Pro-X carboxypeptidase	no	lysosome
lysosomal Pro-Xaa carboxypeptidase	yes	lysosome

In our laboratory, there is a work in progress on one of the major peptidases found both in transcriptome and secretome, cathepsin L, but there is no published work about *F. magna* peptidases so far. In the frame of this thesis, the focus is placed on three peptidases identified in transcriptome and/or in secretome: cathepsin L, B and D. Cathepsin L and B belong to the superfamily of cysteine peptidases, and cathepsin D belongs to the superfamily of aspartic peptidases.

2.4.3 Cysteine peptidases of trematodes

One of the 14 protease superfamilies is the group of cysteine proteases (CPs), which are characterized by a catalytic mechanism mediated by nucleophilic cysteine thiol in a catalytic triad or dyad (Barrett, 2013). To date, CPs are sorted into 11 clans and into more than a hundred families according to the MEROPS database. Cathepsin L and B belong to family C1 (papain-like) and subfamily C1A.

Trematode cysteine peptidases are part of housekeeping and basal cellular metabolism in cells and also represent prime enzymes responsible for digestion of host proteins. Their importance can be deducted by their high representation in trematode transcriptomes, for example, 15 % of transcripts of *F. hepatica* transcriptome, belong to cysteine peptidases (Dalton *et al*., 2006). Most trematode species that are of veterinary and medical significance feed on host blood. It has been experimentally proven that CPs can cleave haemoglobin (Hb) in many trematode genera, for example, *Schistosoma*, *Fasciola*, *Clonorchis*, *Opisthorchis* and *Paragonimus,* therefore, playing an important role in nutrient acquisition (Robinson *et al*., 2013; Caffrey *et al*., 2018). Trematodes also employ CPs to modulate host immune reaction in their favour. For instance, in the case of *F. hepatica*, it was observed that CPs cleave host immunoglobulins, thus cease eosinophil activation. Additionally, these enzymes can suppress Th1 proinflammatory response which advantages long-term survival of parasites in the host organs (Dalton *et al*., 2013). Furthermore, CPs are important for infective larval stages, such as cercaria, metacercaria, NEJs, because they facilitate excystment, penetration of the host body, and migration to the final localization (Chung *et al*., 1995; Kašný *et al*., 2007; Dvořák *et al*., 2008; McGonigle *et al*., 2008; Robinson *et al*., 2009). Finally, there are studies showing that CPs participate in trematode reproduction, specifically, that they play a function in egg formation and production (Meemon *et al*., 2004).

2.4.3.1 Cathepsin L

The dominant monophyletic group of cysteine endopeptidases expressed and secreted by fasciolid trematodes are cathepsins L peptidases (CLs). Bioinformatic data from adult *F. magna* show that CL (FmCL) is the most abundant protein encoded in the transcriptome and the second most abundant protein in ESPs, right after cysteine protease inhibitor cystatin (Cantacessi *et al*., 2012).

Historically, CLs belong to the most studied enzymes in fasciolid trematodes, especially in the case of *F. hepatica*. The predominance of CLs in the adult *F. hepatica* secretome, around 80% of the total secreted proteins, is pointing at their pivotal role in host-parasite interactions (Robinson *et al*., 2009). Latest genomic investigations revealed that there are 23 members of the CL gene family which are separated into five clades (FhCL1 – FhCL5) (Cwiklinski *et al*., 2019). Interestingly, each clade has its own particular functions and abilities throughout the life cycle, as the liver fluke migrates from intestine to liver to bile ducts where adult worms reside. Members of FhCL1, FhCL2 and FhCL5 are specific for adult flukes and members of FhCL3 and FhCL4 clade are exclusive for metacercariae and NEJs. (Dalton *et al.*, 2006; Robinson *et al.*, 2013; Cwiklinski *et al.*, 2015). The most important role of CLs for adult flukes is the cleavage of Hb and other blood proteins into small peptides in the parasite's gut lumen. It is believed, that those

small peptides are then absorbed by gut epithelial cells and further processed into free amino acids (Lowther *et al.*, 2009). The role of CLs in fasciolid infection is probably much wider and there are many other suggested functions for CLs, not only for adult flukes but also for other developmental stages (Figure 6).

2.4.3.2 Cathepsin B

Another important trematode cysteine peptidase group is formed by cathepsin B peptidases (CBs). Cathepsin B (FmCB) was also identified in the transcriptomic data from adult *F. magna* and it accounts for the second most expressed peptidase transcript, right after FmCL. In addition, FmCB was also abundant in its ESPs (Cantacessi *et al*., 2012).

The most up to date investigation of *F. hepatica* genome revealed that there are 11 members of CB gene family (Cwiklinski *et al*., 2019). In *F. gigantica* and *F. hepatica,* it has been observed that CBs are to some extent stage-specific (like it has been described for CLs) and predominantly expressed by juvenile parasites like metacercaria, NEJs and immature flukes, and less by adult flukes (Figure 7) (Meemon and Sobhon, 2015). Laboratory experiments revealed that CBs are crucial for successful infection within the definitive host because RNAi-mediated silencing of CB lowers the ability of NEJs to penetrate the intestinal wall (McGonigle *et al*., 2008). Nevertheless, some CBs were also localized within the adult's digestive system, which can possibly mean that they might have a function in the nutrition acquisition too (Meemon *et al*., 2004).

2.4.4 Aspartic endopeptidases of trematodes

Aspartic endopeptidases have a different catalytic mechanism from cysteine endopeptidases. They use an activated water molecule as a nucleophile that attacks the peptide chain. Usually, two catalytic residues consisting of aspartic acids act as ligands for the activated water molecule, but in some aspartic endopeptidases, one aspartic acid is substituted by another amino acid (Rawlings and Salvesen, 2012).

2.4.4.1 Cathepsin D

Cathepsin D (CD) belongs to AA clan and A1 family of aspartic endopeptidases. To date, no CD was studied in detail in the family Fasciolidae. The mRNA sequence of cathepsin D was found in the transcriptome of adult *F. magna*, but it was not identified in ESPs (Cantacessi *et al*., 2012).

The first described aspartic protease from any liver fluke was cathepsin D-like peptidase (*Ov*-APR-1) of *Opisthorchis viverrini.* The protein form of *Ov*-APR-1 was confirmed to be expressed in the adult worm, and it was detected in the ESPs, gut and reproductive system. The *Ov*-APR-1 mRNA was also identified in eggs and metacercariae. Therefore, it was suggested that *Ov*-APR-1 partake in the digestion of host proteins, protein turnover within reproductive organs and has a function in other developmental stages too. It is secreted extracellularly and it also functions intracellularly as a lysosomal peptidase (Suttiprapa *et al*., 2009; Young *et al*., 2014). Unlike in fasciolid trematodes, cathepsin D was relatively well studied in blood flukes. *Schistosoma japonicum* cathepsin D (SjCD) and *Schistosoma mansoni* cathepsin D (SmCD) are crucial enzymes in digestive cascade, possibly the initial proteases involved in the Hb cleavage. This claim is supported by the localization of CD in both species in gastrodermis. In addition, SjCD is able to degrade human IgG, serum albumin and complement C3 which might be useful for parasite's evasion from host immune response to infection. Interestingly, mRNA of SjCD was detected across all developmental stages, suggesting that it could have various functions throughout the life cycle apart from digestion (Caffrey *et al*., 2004; Morales *et al*., 2008).

Figure 6: **Available information related to CLs from the family Fasciolidae**. Localization was determined either by immunolocalization or/and RNA *in situ* hybridization. **Cu**: clade unspecific, **A**: adult, **NEJ**: newly excysted juvenile, **M**: metacercaria, **E**: egg.

Gene (clade)	Species	Life stage	Localization	Ability to cleave substrates	Proposed function	Reference
FhCL1	*F. hepatica*	A	gastrodermal epithelial cells	Hb, fibrinogen, plasmin, Toll-like receptor	liver migration, nutrition acquisition, inhibiting host cellular response	Collins *et al*. (2004); Robinson *et al*. (2008); Lowther *et al*. (2009); Robinson *et al*. (2009, 2013); Mebius *et al*. (2018)
FhCL2	*F. hepatica*	A	no data	host native collagen, fibrin, plasmin	liver migration, nutrition acquisition	Robinson *et al*. (2009, 2011); Mebius *et al*. (2018)
FhCL3	*F. hepatica*	M, NEJ	gut	host native collagen, fibrin	intestine and liver penetration	Robinson *et al*. (2009, 2011); Mebius *et al*. (2018); Cwiklinski *et al*. (2019)
FhCL4	*F. hepatica*	NEJ	no data	no data	no data	Robinson *et al*. (2008)
FhCL5	*F. hepatica*	A		Hb, immunoglobulin, laminin, fibronectin	liver migration, nutrition acquisition, immune evasion	Robinson *et al*. (2009); Norbury *et al*. (2011)
FhCL (Cu)	*F. hepatica*	A, M, E	no data	no data	no data	McNulty *et al*. (2017)
FgCatL1H	*F. gigantica*	NEJ, M	gut	collagen I, laminin, IgG	excystment, tissue penetration, migration, immune evasion	Sansri *et al*. (2013)
FgCatL1G	*F. gigantica*	NEJ	no data	immunoglobulin, laminin, fibronectin, collagen I	excystment, tissue migration, immune evasion	Norbury *et al*. (2011)
FgCatL-A	*F. gigantica*	A	gut	no data	nutrient digestion	Meemon *et al*. (2010)

Figure 7: **Available data related to CBs from the family Fasciolidae**. Localization was determined either by immunolocalization or/and RNA *in situ* hybridization. **Cu**: clade unspecific, **A**: adult, **NEJ**: newly excysted juvenile, **M**: metacercaria, **E**: egg.

Gene	Species	Life stage	Localization	Ability to cleave substrates	Proposed function	Reference
FhCB1	*F. hepatica*	A, NEJ, M	gut	no data	intestine penetration, migration	Beckham *et al*., (2009); Cwiklinski *et al*. (2015)
FhCB2	*F. hepatica*	A, NEJ, M	no data	no data	intestine penetration, migration	Cwiklinski *et al*. (2015)
FhCB3	*F. hepatica*	A, NEJ, M	no data	no data	intestine penetration, migration	Cwiklinski *et al*. (2015)
FhCB4–10	*F. hepatica*	A, NEJ, M	no data	no data	no data	Cwiklinski *et al*. (2015)
FhCB (Cu)	*F. hepatica*	A, M, E	no data	no data	no data	McNulty *et al*. (2017)
FhCB (Cu)	*F. hepatica*	A, NEJ, M	no data	no data	excystment, intestine penetration, internally within cells – protein turnover, catabolism	McGonigle *et al*. (2008); Robinson *et al*. (2009); Cwiklinski *et al*. (2019)
FgCatB1	*F. gigantica*	A, NEJ, M	gut epithelial cells, tegumental cells, prostate gland, Mehlis' gland, vitelline gland, testis, eggs	no data	general digestive function	Meemon *et al*. (2004)
FgCatB2	*F. gigantica*	NEJ, M	no data	collagen I, fibronectin, immunoglobulins	migration through the liver, immune evasion	Meemon *et al*. (2004)
FgCatB3	*F. gigantica*	NEJ, M	gut	collagen I, fibronectin	excystment, penetration of intestinal wall, migration through the liver	Meemon *et al*. (2004); Sethadavit *et al*. (2009); Anuracpreeda *et al*. (2011)
FgCB5	*F. gigantica*	A	gut	laminin, fibronectin, collagen, bovine serum albumin, Hb, IgG	active in the endolysosomal pathway, stimulating an immune response	Siricoon *et al*. (2015)

2.5 Methods of investigating gene expression in trematodes

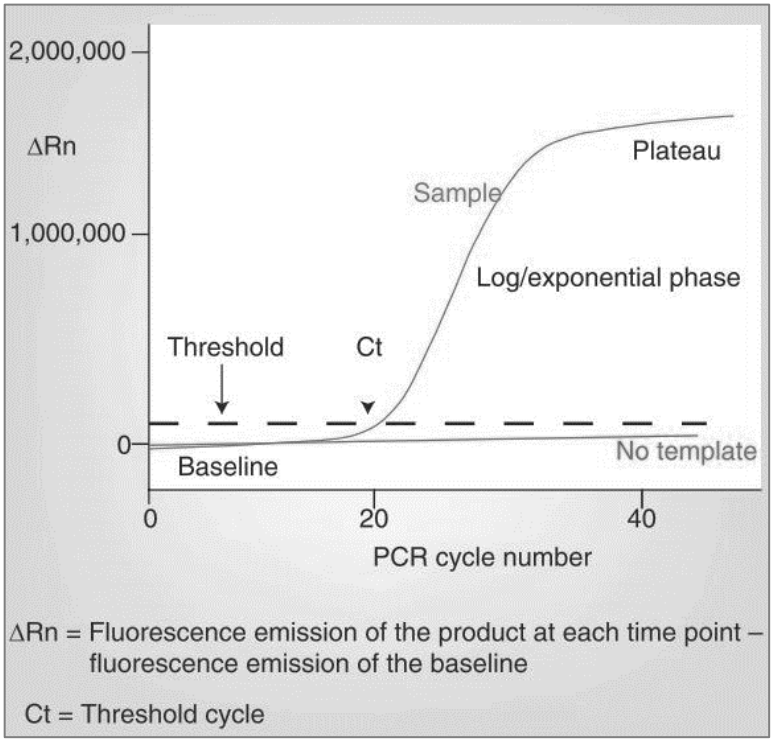
Research of gene expression is an important part of molecular biology mainly aiming to investigate messenger RNA (mRNA) as a mediator between DNA information and proteins (San Segundo-Val and Sanz-Lozano, 2016). In trematodes, the study of gene expression aims to clarify gene functions, to understand their biological roles, and to dissect which genes are crucial for successful infection of the host. Consequently, such understanding can help in the development of novel diagnostic tools, control approaches and vaccines (Cwiklinski and Dalton, 2018). Nowadays, most commonly used methods for detection and quantification of mRNA in trematodes are quantitative PCR (chapter 2.5.1), expression microarrays and RNA sequencing (chapter 2.5.3). To localize a certain place of mRNA transcription in an organism, RNA *in situ* hybridization is the method of choice (chapter 2.5.2). Recently, RNA interference (RNAi) became a very useful method which allows silencing of a target mRNA in living parasites, therefore, researchers are able to study direct effects on phenotype (chapter 2.5.4).

2.5.1 Quantitative PCR

Quantitative PCR (qPCR) is based on principles of conventional PCR with the difference that it can measure amplification of DNA molecules in real time, therefore allowing to estimate starting quantities of the DNA in a sample. The qPCR reaction contains extra component: a fluorophore (specific or unspecific) that emits a fluorescence every time it binds to a newly amplified double-stranded DNA. This signal is measured in every amplification cycle by a sensitive camera in a qPCR thermal cycler, which collects the data during the entire run. The basic component of qPCR results is a C_T value, which is a cycle number at which the fluorescence (amount of DNA) is greater than the minimal detection level. It means that, when comparing two samples with a different starting amount of DNA copies, C_T values will be lower for the sample with more starting DNA and vice versa (Figure 8). The starting amount of DNA can be calculated while the amplification is in the exponential phase. The qPCR method can be employed when measuring DNA (for example in diagnostics) or RNA, when reverse transcribed to complementary DNA (cDNA) (Arya *et al*., 2005; Segundo-Val and Sanz-Lozano, 2016).

To quantify the level of a gene of interest (GOI) expression, one has to choose the right strategy of proceedings suitable for a research purpose, organism and available material. The appropriate methodology should prevent sample-to-sample variations as much as possible. There are several considerations one should think through before even starting the experiment to ensure that final results would be strong, transparent and reproducible. Profusely cited methodology of MIQE (Minimum Information for Publication of Quantitative Real-Time PCR Experiments) guidelines summarize such considerations and crucial points worth being aware of while designing qPCR experiments (Bustin *et al*., 2009).

Figure 8: **Model of a single amplification curve illustrating the basic principles and terms for qPCR.** (Arya *et al*., 2005)



Quantitative PCR is commonly used in studies investigating gene expression of trematodes or their infected hosts. This technique is often employed for comparison of gene expression levels among various life stages (e.g. infective larvae/adult parasite) to reveal the importance and function of particular genes during infection process or ontogenetical development of the parasite. For example, *F. hepatica* mRNA cathepsin levels were observed during the development of NEJ (Cwiklinski *et al.*, 2019). FhCL3 and FhCB2 expression were upregulated 6 hours after NEJ excystment, but in contrast, FhCB1/3 expression was downregulated at the same time point. Based on such findings, we can elucidate, that FhCL3 and FhCB2 are peptidases important for an early migration of the NEJ (Cwiklinski *et al.*, 2019). Another study investigated expression levels of two similar cathepsin L3s of *F. hepatica* (FhCL3-1 and FhCL3-2) during development of metacercaria and NEJ, and revealed, that only FhCL3-2 was upregulated during the excystment of NEJ, and FhCL3-1 did not show any significant changes and had low expression levels during the given period (Zawistowska-Deniziak *et al*., 2013). In addition, this approach was employed to compare expression levels of fours cathepsins B of *Clonorchis sinensis* among egg, metacercaria, excysted juvenile and adult, and results revealed, that 3 of 4 CBs has the highest expression in adult flukes (Chen *et al*., 2014). Cathepsins are not the only genes studied by qPCR, for instance, differential expression levels have been investigated for ferritin or saposin-like proteins for eggs, miracidia, NEJs and adults of *F. hepatica* (Cabán-Hernández *et al*., 2012, 2013). The same approach to investigating expression profiles has been employed also for blood flukes (genus *Schistosoma*). For example, Cathepsin L3 of *S. mansoni* (SmCL3) is predominantly expressed in life stages infecting mammal hosts, such as schistosomula and adult flukes and it has low expression in eggs, sporocysts and cercariae (Dvořák *et al*., 2009). That suggests, that the

function of SmCL3 is mainly related to developmental stages in a definitive host. In the same study, relative expression levels of SmCL3 in adults were also compared to other digestive peptidases know in this species, and the transcription was 50 to 1000 fold lower for SmCL3, which could mean, that the contribution of SmCL3 to proteolysis in the digestive system is less important than the other peptidases (Dvořák *et al*., 2009).

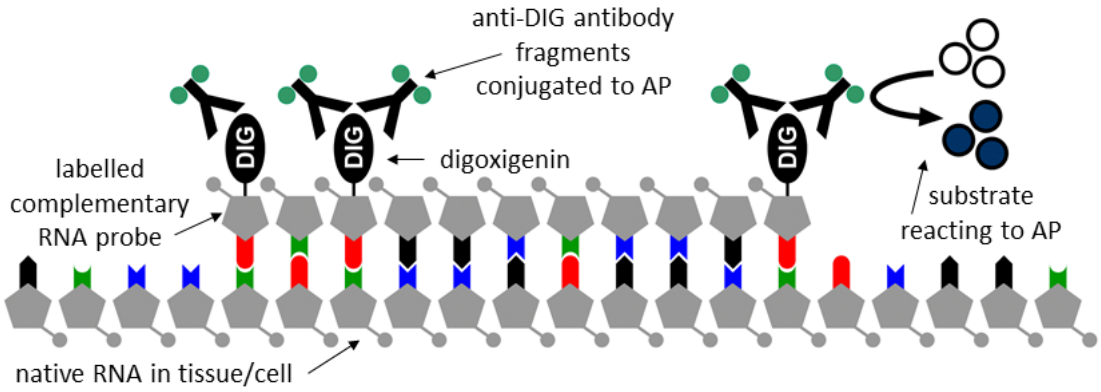
Nevertheless, using the qPCR method for such purposes demands adjustments to studied trematode species. Mainly, it is desirable to use relative quantification with a validated group (at least 3) of reference genes, which are stably expressed among studied samples (life stages, biological replicates), usually coding essential cellular housekeeping proteins (Yoo *et al*., 2009). There is a shortage of information in the literature about reliable reference genes for trematodes, and frequently, results of gene expression studies are normalized only according to one reference gene (often not even validated) which makes such results somewhat questionable. Good news is that lately, there have been published few articles focusing solely on analysis of reference genes for particular trematode species, for example, *Clonorchis sinensis* (Yoo *et al*., 2009) and *S. japonicum* (Liu *et al*., 2012), which established a cornerstone for future gene expression studies about these species.

2.5.2 RNA *in situ* hybridization

The localization of RNA in organisms and tissues can be studied using *in situ* hybridization (ISH). It can be performed on sections or in whole objects (whole-mount, WISH) by application of labelled RNA probes which are complementary (anti-sense) to RNA molecules within the studied organism. It is a powerful method which allows RNA detection even within a single cell. Traditionally, probes were labelled radioactively for example with ^{3}H, ^{35}S, ^{125}I, ^{32}P isotopes. However recent techniques rely mainly on nonradioactive, such as biotin, digoxigenin, fluorescein or alkaline phosphatase labelled probes (Lehmann and Tautz, 1994; Jin and Lloyd, 1997). ISH can be employed as an alternative of immunohistochemistry, which requires more time and resources, for example, production of recombinant protein, its purification and production of antibodies (Cogswell *et al.*, 2011).

RNA probes can be prepared by commercially available kits which allow easy *in vitro* transcription using template cDNA and reverse transcriptase. Using this procedure, “anti-sense” RNA probes (complementary to native RNA) and “sense” probes serving as a negative control (transcribed in the opposite direction) can be synthesized and labelled with digoxigenin (DIG). During the ISH experiment, DIG serves as a ligand for anti-DIG antibodies conjugated with alkaline phosphatase (AP). Afterwards, the reaction of AP and suitable substrate results in a coloured signal in the place of target gene expression (Figure 9).

Figure 9: **Illustration of a basic principle and components of ISH reaction.** (edited; source of the original scheme: www.laboratoryequipment.com)



The RNA *in situ* hybridization method is being employed also for the study of trematodes. For example, cathepsin B1 transcripts of *F. gigantica* (FgCat-B1) were localized in the caecal epithelial cells and cells underlining the proximal part of the digestive system (pharynx, oral sucker, ventral sucker) in juvenile and adult worm (Meemon *et al.*, 2004). Furthermore, mRNAs of the same cathepsin were detected in tissues of the reproductive system (vitelline glands, cells underlining uterus, eggs, Mehlis' gland, testes, prostate gland) in adult *F. gigantica*. Thus, suggested functions based on these findings are that FgCat-B1 can work as a digestive enzyme, as a secreted peptidase helping with evasion from host immune response, and it can possibly play a role in egg production (Meemon *et al.*, 2004). In the same species, RNA *in situ* hybridization was also applied to localize cathepsin L mRNA (FgCat-L) in adult fluke (Meemon *et al.*, 2010). In contrast to FgCat-B1, cathepsin L was detected only in the caecal epithelium, which can correspond to suggestions, that this peptidase is responsible for the digestion of the host blood (Meemon *et al.*, 2010). Closely related species *F. hepatica* was investigated for the place of cathepsin L (FhCL) transcription, and mRNA of this peptidase was detected solely in epithelial cells lining the parasite's gut. This observation further supports the assumption that FhCL is responsible for the degradation of ingested host tissue and blood (Dalton *et al*., 2003). In all of the previously mentioned studies, RNA *in situ* hybridization was performed on sections, but this methodology is quite labour intensive and requires the investigation of many sections to detect the localization of transcripts in all organs. Thus, a less laborious variation of ISH was developed for *S. mansoni* by using whole-mount ISH, which allows easier observation of mRNA localization in complex tissue architecture (Cogswell *et al*., 2011). The WISH technique has been employed to localize cathepsin L in *S. mansoni* (SmCL) adults and schistosomula (Dillon *et al*., 2007) and it was confirmed, that SmCL mRNA is located in the gut epithelium of adult worms and in the embryonic gut of schistosomula.

2.5.3 RNA sequencing

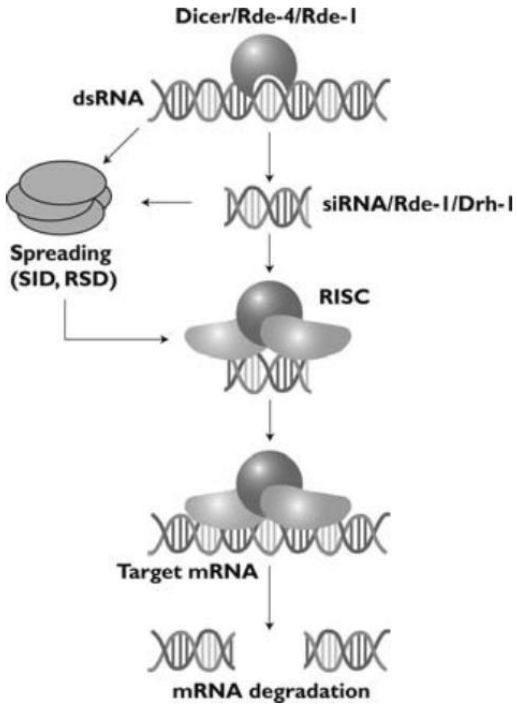
RNA sequencing (RNA-Seq) is a deep-sequencing method allowing transcriptomic profiling. It provides not only the reading of RNA sequences but also their precise quantification. Sequencing is performed via high-throughput sequencing technology and results in large datasets of typically 30-400 bp long reads. After sequencing, reads are aligned to reference genome/transcriptome or assembled *de novo*. The final dataset is called a transcriptome, which is (ideally) a set of all RNA molecules in studied cell or organism at a certain time and under certain conditions. A great advantage of this technique is the possibility of quantification of every molecule and comparing those expression levels to other samples/experiments (Wang *et al*., 2009). This feature is useful for the study of trematode developmental stages and for comparing expression levels of particular genes between life stages (Protasio *et al*., 2012; Cwiklinski *et al*., 2015). An interesting approach is to apply RNA-seq and subsequently transcriptomic data for characterization of the host response to helminth infection. By meta-analysis of multiple transcriptomic datasets, a common pattern in gene expression changes during infection in the host was discovered among various helminths, including trematode infections. Distinct changes were observed mainly in the expression of genes involved in immune regulation, Th2 immunity and inflammatory responses (Zhou *et al*., 2016). For effective exploitation of transcriptomic data of helminths, there has been established several databases which collect transcriptomes (Mangiola *et al*., 2013). Apart from general databases like GenBank®, there are also specialized ones for parasitic helminths which includes genomic, transcriptomic and proteomic data, for example, WowmBase (Yook *et al*., 2012), SchistoDB (Zerlotini *et al*., 2008), HelmCop (Abubucker *et al*., 2011), HelmDB (Mangiola *et al*., 2013) and Helminth.net (Martin *et al*., 2018). In addition, a recently introduced technique of laser microdissection microscopy (LMM) in conjunction with RNA-seq, offers a new possibility of identifying mRNAs in specific helminth organs and tissues (Jones *et al*., 2004).Therefore we can investigate transcriptomes of individual organs, such as gut, tegument or parts of the reproductive system. This approach was, for example, employed for investigation of intestinal transcriptomes of nematodes *Necator americanus* and *Ancylostoma caninum* (Ranjit *et al*., 2006).

2.5.4 RNA interference

Even though sequencing technologies provide large genome and transcriptome datasets, the function of the vast majority of identified genes remains unexplained. A relatively new approach of functional genomics called RNA interference (RNAi) or more precisely ds-induced RNA silencing became a beneficial method for unravelling the functions of genes, by switching them "off". RNAi mechanism was first discovered in plants (in 1990) as a defence system against viruses (Schepers, 2005). The main principle of this reverse genetics method is that the double-stranded RNA (dsRNA) is introduced into an organism, then it is processed into 21–23bp small inhibitory RNA molecules which facilitate the degradation of homologous mRNA molecules native to the organism (Figure 10) (Arenz and Schepers, 2003). Therefore, by performing RNAi, scientists can investigate the consequences of this intervention in living organisms and clarify gene functions while observing its changed phenotype. Such experiments can be carried out for a wide spectrum of eukaryotic organisms including parasitic helminths. First RNAi experiment in

trematodes was published in 2003 and it was performed for cercariae of *S. mansoni*, silencing its cathepsin B, and several studies followed (Skelly et al., 2003; Geldhof et al., 2007). This method was later also introduced for *F. hepatica* when McGonigle *et al.* (2008) successfully silenced two major secreted peptidases FhCL and FhCB.

Figure 10: **Model illustrating the RNAi mechanism of *Caenorhabditis elegans*.** Introduced dsRNA is cleaved into small inhibitory RNAs (siRNAs) by Dicer. RISC complex then binds to siRNAs which ultimately results in a degradation of homologous endogenous mRNA (Geldhof et al., 2007).



3 Materials and methods

3.1 Cathepsin mRNA expression profiles

For estimating and comparison of mRNA expression levels of FmCL, FmCB and FmCD among three developmental stages of *F. magna*, and for comparison of cathepsin expression among themselves in each life stage, quantitative PCR (qPCR) was performed. For this aim, total RNA was isolated from 3 biological replicates from each of *F. magna* eggs, miracidia, adults. Samples were treated with DNAse and reverse transcribed. Three sets of target gene (cathepsin) primers were designed and tested together with 17 sets of reference gene primers for specificity and amplification efficiency. Stability of reference genes was investigated using geNorm analysis and finally, expression levels of target genes from all biological replicates and life stages were measured and normalized.

3.1.1 Material acquisition

Adult worms and eggs were collected from the infected liver of wild deer. Livers were carefully sliced and examined for adults in pseudocysts. For the molecular methods involving RNA research, solely live worms were selected. Adults were washed in a physiological salt solution and immediately frozen with liquid nitrogen and stored in -80 °C. For a collection of eggs, livers were thoroughly washed and acquired suspension was cleaned over sieves several times and concentrated. Eggs were then stored in tap water at 4 °C.

3.1.2 RNA isolation

The procedure for obtaining total RNA was adapted to the individual ontogenetic stages of the parasite. The entire process was carried out in the RNAase-free environment and solutions were made of ultrapure diethyl pyrocarbonate (DEPC) treated water to avoid RNA degradation. RNA was isolated from 3 biological replicates and each of them had a different origin (infected animals were from different geographical locations).

RNA isolation from eggs. Eggs were incubated in Petri dishes at 26 °C in dark for 5 days. We chose this timespan because after such long incubation, the embryogenesis has already started and a sufficient degree of gene expression (transcription) was presumed to be detectable, but also it is secured that no miracidia would by already developed at that time. After incubation, eggs were transferred to eppendorfs and centrifuged (5 minutes/4000 x g/RT). The supernatant was discharged and replaced by TRIzol™ Reagent (Thermo Fisher Scientific), the volume of the reagent was set according to the manufacturer protocol: 1 ml of TRIzol™ Reagent per 50–100 mg of tissue (eggs). Afterwards, eggs were mechanically homogenized with a pestle, and the eggshell disruption was examined under a microscope. The rest of the procedure went according to the manufacturer's protocol, and in the last step, total RNA was eluted to 50 µl Milli-Q water.

RNA isolation from miracidia. Firstly, eggs were incubated in Petri dishes for 14 – 21 days in the dark at 26 °C. After incubation, the petri dish was placed under a light source for 30 minutes to induce hatching. Miracidia exhibit positive phototaxis, therefore they were attracted to the light

cone at the edge of the petri dish and collected into falcon tubes placed on ice. Low temperature decreased the motion of miracidia, thus they sank to the bottom of the tube. Special care was given to avoid the collection of unembryonated eggs, which could alter final results. Afterwards, miracidia were centrifuged (5 minutes/4000 x g/4 °C) and the supernatant was discharged. Such as with eggs, 1 ml of TRIzol™ Reagent per 50 – 100 mg of tissue (suspension with miracidia) was added and miracidia were mechanically homogenized with a pestle. The rest of the procedure went according to the manufacturer's protocol, and in the last step, total RNA was eluted to 50 µl Milli-Q water.

RNA isolation from adults. The whole adult worm was removed from -80 °C, cut to small pieces and divided into eppendorf tubes. 1 ml of TRIzol™ Reagent per 50 – 100 mg of tissue (adult worm) was added into each tube and tissue was mechanically homogenized with a pestle. All suspensions of the homogenized worm were merged into one tube and mixed, to assure, that the dispersion of RNA molecules from all organs would be as even as possible. Only a portion of this mixture was then used for the following procedure and remained sample was stored in -80 °C freezer. The rest of the procedure went according to the manufacturer's protocol, and in the last step, total RNA was eluted to 50 µl Milli-Q water.

The concentration and purity of all acquired RNA samples were assessed via NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific) and stored at -80 °C.

3.1.3 DNAse treatment and cDNA synthesis

All RNA samples were treated with TURBO™ DNase (Thermo Fisher) according to the manufacturer's protocol to clear samples from DNA contamination. Before proceeding to cDNA synthesis, it was crucial to adjust concentrations of all RNA samples to the same level (1000 ng/µl) as the aim of this thesis was to compare the expression of cathepsins between different life stages and biological replicates. High capacity cDNA reverse transcription kit (ThermoFisher Scientific) was used for cDNA synthesis and the procedure went according to the manufacturer's instructions and the starting amount of RNA was 1 µg per reaction. Newly synthesized cDNA was stored at -20 °C.

3.1.4 Primer design and PCR

Primers for qPCR amplification of FmCL, FmCB and FmCD were generated and selected using online free software Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/) (Koressaar and Remm, 2007; Untergasser *et al*., 2012). Characteristics of the primers were: same length of amplified product (149bp), similar primer melting temperatures (Tm) (53 – 56 °C) and very similar %GC content of the amplification product (FmCL: 45.6%, FmCB: 46.3%, FmCB: 48.3%). Primers for reference genes (9 reference genes, 17 sets of primers) were designed according to the master thesis of a former student of our laboratory Romana Šašková (Šašková, 2015). Primer sequences and parameters are shown in Table 5. Gene mRNA sequences of both cathepsins and reference genes were identified by Roman Leontovyč in *F. magna* transcriptome (Cantacessi *et al*., 2012). All original mRNA sequences are listed in the Supplementary data chapter 8.

PCR experiments were performed to confirm primer functioning and specificity. PCR reactions were prepared as shown in Table 6 and cDNA from chapter 3.1.3 was used. Amplification was performed in C1000 Thermal Cycler (Bio-Rad) and its cycling conditions are shown in Table 7. After amplification, the mixture was loaded on 2% agarose gel and samples ran 30 minutes/100V. Finally, the gel was viewed under ultraviolet light and DNA fragments were cut from the gel, cleaned (QIAquick Gel Extraction Kit – Qiagen) and sent to be sequenced in the Laboratory of DNA sequencing (Charles University, Faculty of Science, Biocev). Obtained DNA sequences were analysed using Geneious software. For the determination of the best annealing temperatures for all sets of primers, gradient PCR was performed.

Table 5: **Sequences and amplicon sizes of cathepsin and reference gene primers**. Abbreviations: **CL**: cathepsin L; **CB**: cathepsin B; **CD**: cathepsin D; **ATP5B**: ATP synthase subunit beta; **BACT**: beta-actin; **GAPDH**: Glyceraldehyde 3-phosphate dehydrogenase; **TBP**: TATA-binding protein; **TUBBI2**: beta-tubulin 2; **TUBBI3**: beta-tubulin 3; **TUBBI4**: Beta-tubulin 4; **UBC**: Ubiquitin C; **YWHAZ**: 14-3-3 protein zeta/delta.

gene	forward primer (5'3' direction)	reverse primer (5'3' direction)	amplicon
FmCL	AAGGATCAGGGACAATGTGG	TTACATCCGCTGTTGCCATA	149bp
FmCB	AGGATGGTGTGCCTTATTGG	TTCCGATGTTGCTCAGAGAC	149bp
FmCD	ATTGGTTCCTGCTGTGCTTT	GGATAGAGCCGTTGCTTCCT	149bp
FmATP5B	GGTGTGGGCAAAACAGTGCTT	AGATCGTTACCTTCTCGAGTA	112bp
FmBACT	TGCGACGTCGATATTCGAAAG	GCAGTAATTTCCTTCTGCATG	103bp
FmGAPDH	TCAGAAGGTCGTTGATGGTCC	TCGGGAATGACTTTGCCCACA	123bp
FmTBP1	CCAGAACATGGTAGGGTCGTG	CATGCGGTAAATCAAACCGGG	121bp
FmTBP2	CTTCCCCGGTTTGATTTACCGC	CCGGTCAGTACAATTTTCCCCG	81bp
FmTBP3	TGTGTACCGGAGCTAAGAGTG	ATCACACGACCCTACCATGTTC	131bp
FmTUBBI2.1	CAACTCTTCCGCCCCGATAACT	GCCTCTTTTCGTACCACGTCCA	124bp
FmTUBBI2.2	CGGATCGCATCATGAACACGTT	GACCAGCTGGTGAACAGAGAG	99bp
FmTUBBI3.1	ACTGGGACTTACCACGGAGAC	ATTGTGCCCGGCTCCAAATCAA	121bp
FmTUBBI3.2	GTTGATTGGAGCCGGGCACA	CAGCTCCCTCGGTGTAATGTCC	140bp
FmTUBBI4.1	TCATATCCGCGAGCATGGCA	CGATGAGAATGGAGGTGGAA	131bp
FmTUBBI4.2	CGTTCCACGTGCCATTCTCATC	CCAATTATTACCCGCGCCCGTA	126bp
FmTUBBI4.3	CATACGGGCGCGGGTAATAATT	AACACAGAGTCCACAAGTTGG	71bp
FmUBC3	TGCATCTTGTTTTGCGACTTCG	CTCCTTGTCCTGAATCTTCGC	131bp
FmUBC4	GCGACTTCGTGGTGGTATGCA	GCTTCCCTGCAAATATCAACCG	135bp
FmYWHAZ1	CCGACGAGAGCAGAGTGTTCT	GATGCAGACCACTTGCGATTTC	150bp
FmYWHAZ2	GAGGTGCAGCTGAAAGCCACT	CATGTAGAACACTCTGCTCTCG	141bp

Table 6: **Composition of PCR reaction.**

component	volume	final concentration
DreamTaq Green PCR Master Mix (2X) (ThermoFisher)	12.5 µl	
template cDNA	1 µl	
10µM forward primer	0.5 µl	0.2µM
10µM reverse primer	0.5 µl	0.2µM
nuclease-free water	10.5 µl	
total volume	25 µl	

Table 7: **PCR thermal cycling conditions.**

step	temperature °C	time	number of cycles
initial denaturation	95	2 min	1
denaturation	95	30 s	35
annealing	53	30 s	35
extension	72	1 min	35
final extension	72	10 min	1

3.1.5 qPCR experiments

Sensitive quantitative PCR (qPCR) method was applied for the purpose of measuring and comparing expression levels of selected cathepsins. Principles of this method and its applications are described in chapter 2.5.1. All qPCR experiments were performed in a Rotor-Gene thermocycler (Qiagen) and reactions were loaded and measured in 0.1ml Strip Tubes (Qiagen).

qPCR amplification efficiency. Firstly, the efficiency of PCR amplification had to be determined for each set of primers by generating standard curves from five 10-fold serial dilutions of cDNA. The first dilution was, therefore, 100ng of cDNA in reaction, the second was 10ng etc. For these experiments, original cDNA isolated from adult *F. magna* was used. The composition of the reaction is shown in Table 8 and the cycling conditions in Table 9. Each reaction was measured in 3 technical replicates and negative controls without template cDNA (NTC) were measured too. After cycling programme, dissociation heating process followed to capture melting curves which were checked for potential contamination or unspecific amplification. Standard curves were generated from obtained C_T values for all sets of primers in Microsoft Office Excel to determine amplification efficiency. Efficiency (E) was estimated by the formula: $E = (10^{-1/slope} - 1) \times 100$.

Table 8: **Composition of qPCR reaction.**

component	volume	final concentration
2X SYBR™ Green PCR Master Mix (ThermoFisher)	10 µl	
template cDNA	1 µl	
10µM forward primer	2 µl	0.8µM
10µM reverse primer	2 µl	0.8µM
nuclease-free water	5 µl	
total volume	20 µl	

Table 9: **qPCR thermal cycling conditions.**

step	temperature °C	time	number of cycles
initial denaturation	95	10 min	1
denaturation	95	10 s	40
annealing	53	15 s	40
extension	72	20 s	40
dissociation heating program for melting curves analysis			

Evaluation and selection of reference genes. Since there are no reliable reference genes determined for *F. magna* in literature so far (neither for any other species in Fasciolidae family), it was necessary to perform evaluating experiments and analysis to test reference genes suitable for subsequent data normalization. Five pairs of primers for reference genes were tested with cDNA from every developmental stage. Reactions were prepared according to Table 8 and all cDNA samples were diluted 2x from the original stock before adding them to the reaction. Reactions were measured in 3 technical replicates and negative controls without template cDNA (NTC) were also included. Cycling program is shown in Table 9. Afterwards, the threshold in Rotor-Gene Q software (QIAGEN) was set to 0.02, and this threshold was employed for all subsequent measurements too. C_T values were then imported to qBase+ software (Biogazelle) and evaluated in geNorm. Employed analysis settings are listed below:

- Limit for variation (v) in C_T values among technical replicates was set to $v < 0.5$
- Average C_T from 3 technical replicates was calculated from the arithmetic mean
- The range of valid C_T values was set from 5 to 35 cycles
- Limit for C_T values of negative controls: no C_T detected with threshold 0.02

Selection of the reference genes was carried out according to geNorm M value, which is "…the average M value of all remaining reference genes upon stepwise exclusion of the most unstable reference gene (highest M value)", as it is described in the qbase+ manual (version 2018-09-24). Originally, the calculation of M value for this analysis was described by Vandesompele *et al.* (2002).

Measuring cathepsin expression levels. Two types of gene expression experiments were performed. The first type aimed to compare how are FmCL, FmCB and FmCD expressed in relation to each other in a single developmental stage. Since three life stages were studied, three different comparisons were made for each life stage. For this experiment, it was important, that target gene-specific primers were designed to amplify product with the same length so that SYBR Green would bind equally to newly synthesized double-stranded cDNAs of all studied cathepsins. The second type of experiment aimed to compare how is each cathepsin gene expression regulated among 3 studied life stages (egg, miracidium, adult), therefore three (for 3 target genes) different comparisons were made. For all experiments, original cDNA stock of all samples (3 life stages x 3 biological replicates) was 2x diluted. qPCR reactions were prepared as described in Table 8 in 3 technical replicates and NTCs were measured too. Reactions with reference genes primers were measured in the same run as target genes reactions. Thermal cycling temperatures and timing were set as described in Table 9. After cycling, melting curves and NTCs were checked, and the threshold for C_T values was set to 0.02.

Data analysis. Limit for variation (v) of C_T values among technical replicates was $v < 0.5$. Average C_T from 3 technical replicates was calculated as the arithmetic mean. Normalization factor (NF) was calculated as the geometric mean from C_T values of 3 validated reference genes (Vandesompele *et al*., 2002). Normalized expression levels for GOIs were calculated using the $2^{-\Delta Ct}$ method (Livak and Schmittgen, 2001).

3.2 RNA *in situ* hybridization

The chromogenic RNA *in situ* hybridization was employed to localize FmCL, FmCB and FmCD mRNA transcripts in adult worm. Specific complementary anti-DIG RNA probes were synthesized and applied to paraffin sections to detect coloured signals under a light microscope. Principles and applications of this method are described in chapter 2.5.2.

3.2.1 Histological preparation of adult *F. magna* worms

Paraffin-embedded adult worms were acquired from Mgr. Romana Šašková and the process of worm acquisition, fixation and embedding are described in her master's thesis (Šašková, 2015). As it is written in Czech, a brief description of the process follows. Live worms were fixed in Bouin solution for 3 days, then, worms were washed from picric acid and went through baths of progressively more concentrated ethanol to remove the water from the tissue. Next step was clearing the tissue with xylene, and finally embedding it in paraffin (Šašková, 2015). Worms in paraffin were cut to 7-micrometre-thick sections using Leica Microtome. Then they were placed on extra adhesive microscope slides and stored at 4 °C.

3.2.2 Preparation of anti-DIG specific RNA probes

Primers for FmCL, FmCB and FmCD cDNA synthesis were designed the same way as for qPCR experiments (see chapter 3.1.4), except that amplicon sizes were ranging from 400 – 500bp (Table 10). Specificity and functioning of primers were tested with PCR (for reaction composition and cycling conditions see Table 6 and Table 7, respectively) and gel amplicons were visualised by UV light and checked by sequencing.

Table 10: **Primer sequences and amplicon sizes for RNA probe synthesis.**

gene	forward primer (5'3' direction)	reverse primer (5'3' direction)	amplicon
FmCL	CCTCATCGTCAGTGTGTTCG	ATCCGCTGTTGCCATAGTTC	494bp
FmCB1	TGCTCGTCAAGATGAACTGG	GTGGATAGCCACCTTTGCAT	495bp
FmCB2	CCAGGCAGCAAGATTAGACC	CCCGAAACCACAAATACCAC	409bp
FmCD	GTTCCTGCTGTGCTTTGTGA	CACCTCAACTCCACCAACCT	432bp

In order to have enough cDNA to synthesize RNA probes, it was necessary to clone cDNA into a plasmid vector and transform it into competent cells, grow them, isolate plasmids and digest them. Firstly, the clean PCR product was ligated to pGEM®-T Easy Vector System I (Promega) according to the manufacturer's protocol. The composition of the reaction is shown in Table 11 and ligation was carried out overnight at 4 °C.

Table 11: **Composition of the ligation reaction.**

component	volume
2X Rapid Ligation Buffer	5 µl
pGEM®-T Easy Vector (50ng)	1 µl
PCR product	3 µl
T4 DNA Ligase	1 µl
total volume	10 µl

Next step was a transformation of competent TOP10 cells with the plasmid. TOP10 cells were taken from -80 °C and placed on ice to slowly thaw. Afterwards, 2 µl of the ligation reaction was added to 50 µl of cells and this mixture was incubated 20 minutes on ice. Subsequently, the mixture was exposed to heat shock for 45 seconds at 42 °C and then immediately placed back on the ice. 250 µl of SOC media was added to cells and samples were incubated for 90 minutes, at 37 °C on a shaker set on 180 rpm. Meanwhile, LB plates were treated with 12,5 µl of ampicillin. After incubation, 100 µl of cell mixture was spread on plates which were then incubated for 12 hours at 37 °C. Afterwards, newly grown transformant colonies were transferred to new LB plates and incubated for 12 hours at 37 °C again. To confirm successful ligation and transformation, colony PCR screening (also called colony cracking) was performed. Using sterile tip, small pieces of several colonies were transferred to 0,2ml tubes containing 9 µl of nuclease-free water. Then they were placed into BioRad C1000 Thermal Cycler (Bio-Rad) and cycling programme described in Table 12 was performed. As a next step, the PCR reaction was prepared, and amplification was carried out according to Table 13 and Table 14, respectively. Samples were loaded to gel, cut out after electrophoresis, cleaned and sequenced same way like in chapter 3.1.4. Sequences had to be checked for the direction of the insert in the vector, and according to that, suitable polymerase (T7 or SP6) and restriction enzyme were selected for the synthesis of RNA probes.

Table 12: **Thermal conditions for „colony cracking “.**

step	temperature °C	time	number of cycles
1	96	5 min	1
2	50	2 min	1
3	96	2 min	1
4	45	1 min	1
5	96	1 min	1
6	40	1 min	1
7	4	2 min	1

Table 13: **Composition of PCR reaction for colony PCR screening.**

component	volume	final concentration
EmeraldAmp GT PCR Master Mix (Clontech)	12.5 µl	
mixture of competent cells	9 µl	
10µM M13 forward primer	0.5 µl	0.2µM
10µM M13 reverse primer	0.5 µl	0.2µM
nuclease-free water	2.5 µl	
total volume	25 µl	

Table 14: **Thermal cycling conditions for colony PCR screening.**

step	temperature °C	time	number of cycles
initial denaturation	95	2 min	1
denaturation	95	30 s	35
annealing	55	30 s	35
extension	72	1 min	35
final extension	72	10 min	1

Positive colonies were chosen to inoculate 30 ml of JB medium treated with 15 µl of ampicillin and this culture was incubated for 16 hours at 37 °C in a shaker set in 220rpm. Afterwards, plasmids were isolated from these cultures using Genopure Plasmid Midi Kit (Roche), the isolation was carried out according to manufacturer's protocol ("Procedure for High Copy Number Plasmids") and at the last step, plasmids were eluted to 50 µl of nuclease-free sterile water. The concentration and purity of plasmid samples were estimated with NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific). After that, plasmid had to be digested with suitable restriction enzyme in order to be in the correct direction for preparation of anti-sense RNA probes, which are complementary to native sense mRNAs localized in *F. magna* tissue. For that purpose, either NcoI or SacI (New England Biolabs) were employed for digestion reaction (see Table 15) along with appropriate buffers. Before that, sequences were tested for restriction sites via online software NEBcutter 2 to prevent unwanted digestion at additional sites. Digestion was carried out in C1000 Thermal Cycler (Bio-Rad) at 37 °C for 3 hours. Immediately after digestion, the reaction was stopped by increasing temperature to 80 °C for NcoI and to 65 °C for SacI for 20 minutes. As the last step, samples were cleaned and concentrated by ethanol precipitation, and their concentration was estimated by Nanodrop. Using DIG RNA Labelling Kit (SP6/T7) (Roche), anti-sense and control sense RNA probes were synthesized for FmCL, FmCB and FmCD according to manufacturer's manual, and were then stored at -20 °C.

Table 15: **Composition of the digestion reaction.**

component		volume
NcoI	SacI	1 µl
3.1 NEBuffer	1.1 NEBuffer	5 µl
plasmid DNA		10 µl
nuclease-free water		34 µl
total volume		50 µl

3.2.3 RNA *in situ* hybridization protocol

Preparation. After preparing histological sections and anti-DIG RNA probes, it was also necessary to make several solutions (for recipes see Table 18). Since this experiment involves work with RNA, all solutions were made of DEPC treated water (Diethyl pyrocarbonate; Sigma-Aldrich) and most of them were autoclaved to avoid RNA degradation by RNases. Laboratory tools, like tweezers, incubation boxes, cuvettes etc., were autoclaved too. Laboratory desk was treated with RNaseZap™ RNase Decontamination Solution (Invitrogen). The ISH procedure involves certain steps requiring overnight incubation, therefore, the entire experiment lasts 3 days.

First day: At each step, slides were completely immersed in the following solutions (if not mentioned differently, incubation was carried out at room temperature):

- Deparaffinization of slides: 2x15 min xylene; 2x5 min 100% EtOH; 5 min 96% EtOH; 5 min 70% EtOH; 5 min 50% EtOH; 5 min 30% EtOH; 5 min DEPC H_2O

- Proteinase K treatment: 20 min 0,2M HCl; 5 min proteinase K buffer at 37 °C, afterwards add 10,75 µl/100 ml of proteinase K (Roche) to proteinase K buffer with slides for 17 min at 37 °C; 15 sec 20% acetic acid at 4 °C (on ice); 5 min PBS
- Hybridization: 15 min 20% glycerol; short wash in 2x SSC; incubation of slides on the slide heater at 70 °C for 10 min; transfer of slides onto a frozen board, adding 2 µl of RNA probe to 80 µl hybridization mix 1 (HM1) (Table 16); incubation of the mixture at 70 °C for 5 minutes; adding 120 µl of hybridization mix 2 (HM2) to the tube with HM1 (ratio of HM1/HM2=1/1,5); adding 200 µl of the hybridization solution to each slide; covering slides with cover glasses; transfer of slides into wet boxes; incubation at 45 °C overnight (16 – 20 hours)

Table 16: **Composition of hybridization solution.**

hybridization mix 1 (HM1)	volume/weight
20xSSC	0.63 ml
RNA from torula yeast (Sigma)	2.5 g
0.1M PBS	0.37 ml
total:	1 ml

hybridization mix 2 (HM2)	
deionized formamide	1.25 ml
10% Tween 20	0.25 ml
dextran sulfate	0.25 g
total:	1.5 ml

Second day: At each step, slides were completely immersed in the following solutions (if not mentioned differently, incubation was carried out at room temperature):

- Application of anti-DIG alkaline phosphatase: 15 min in 2xSSC + 0.1% Tween 20; 15 min in 1xSSC; incubation 15 min in 0.5xSSC; incubation 15 min in 0.1xSSC; incubation 5 min in MAB, adding 200 µl of 4% blocking buffer (Table 17) to each slide, incubation for 30 min in wet boxes; adding Anti-Digoxigenin-AP, Fab fragments (Roche) in dilution 1:500 to 2% blocking buffer; applying 200 µl of 2% blocking buffer with alkaline phosphatase on each slide; covering slides with cover glasses; transfer of slides to wet boxes and incubation at 4 °C overnight (16 – 20 hours)

Table 17: **Composition of blocking buffer.**

	4% blocking buffer	2% blocking buffer
inactivated horse serum	0.08 ml	0.04 ml
MAB	1.92 ml	1.96 ml
Total:	2 ml	2 ml

Third day. At each step, slides were completely immersed in the following solutions (if not mentioned differently, incubation was carried out at room temperature):

- Application of substrate: 2x20 min in MAB; incubation for 5 min in substrate buffer; application of 200 µl of the substrate to each slide (SIGMAFAST™ Fast Red TR/Naphthol AS-MX tablets, Sigma-Aldrich); stopping the reaction in distilled water
- Mounting slides: application of VECTASHIELD® Antifade Mounting Medium (Vector Laboratories) on each slide; covering with cover glasses

Table 18: **Recipes for CISH solutions.**

solution	composition	volume	pH	autoclave
DEPC treated water	1 l ultrapure water 1 ml DEPC	1 l		yes
MAB (Maleic Acid Buffer)	0.1 M maleic acid 0.15 M NaCl DEPC H_2O to 1l	1 l	7.5	yes
proteinase K buffer	100 mM Trizma Base 50 mM EDTA DEPC H_2O to 0.5 l	500 ml	8	yes
substrate buffer	0.1 M Trizma Base 0.4 M NaCl DEPC H_2O to 0.25 l	250 l	8.2	yes
0.1 M PBS (Phosphate-buffered saline)	125 ml of: 0.2 M Na_2HPO_4 + 0.2 M NaH_2PO_4 125 ml DEPC H_2O 1.4 g NaCl	250 l	7.2	yes
0.2% glycine	1 g glycin DEPC H_2O to 0.5 l	500 ml		yes
20% glycerol	50 ml glycerol 200 ml DEPC H_2O	250 ml		yes
10% Tween 20 (Bio-Rad)	0.5 l Tween 20 4.5 ml DEPC H_2O	5 ml		
20x SSC	3 M NaCl 0.3 M trisodium citrate DEPC H_2O to 500ml	500 ml	7.2	yes
20% acetic acid	24 ml acetic acid 96 ml DEPC H_2O	120 ml		

3.3 Expression of recombinant cathepsin B in yeast

One of a widely used expression yeast systems employs *Pichia pastoris* as an organism for expression of a foreign recombinant protein. This system has several advantages, for example, simple molecular genetic manipulation (similar to *Saccharomyces cerevisiae*), the ability to produce recombinants at large quantities, the capability to express the protein either intracellularly or extracellularly and the ability to perform posttranslational modifications (Cereghino and Cregg, 2000). As the FmCB was identified in *F. magna* adult secretome (Cantacessi *et al*., 2012), it was convenient to use *Pichia pastoris* as it is a suitable system for expression of rightly folded secreted protein. The expression experiment was carried out using the PichiaPink™ Expression System (Invitrogen) according to the manufacturer's instructions with some modifications. A similar procedure for expression of CB was also carried out by Beckham *et al*. (2006) and Sethadavit *et al*. (2009).

3.3.1 Gene modification and synthesis

FmCB mRNA sequence was acquired with the help of Mgr. Roman Leontovyč from transcriptomic data (Cantacessi *et al*., 2012). The sequence was modified to be suitable for the PichiaPink™ Expression System (Invitrogen). Modifications included: removing start codon, removing signal peptide (not identified in FmCB using online software SignalP-4.0), replacement of glycosylation sites, adding His tag and stop codon at the end of the sequence. This modified sequence was then sent to a commercial company (Eurofins Genomics) to be synthesized and inserted into pPink-alpha HC vector. Gene was also codon optimized by the company to enhance expression by the yeast cells. pPink-alpha HC vector possesses restriction sites at the beginning (5' end) and at the end (3' end) of the inserted gene to allow further manipulation and cloning. It also contains its own secretion signal sequence (alpha mating factor secretion signal sequence) to facilitate recombinant protein extracellular expression. The original transcriptome sequence and the final modified sequence is shown in Table 19.

3.3.2 Transformation of TOP10 cells

Obtained lyophilized vector with FmCB insert was diluted to a final concentration of 200 µg/ml. TOP10 competent cells (*Escherichia coli*) were put on ice to slowly thaw. The volume of 0,5 µl of the stock vector was added to competent cells, mixed gently and incubated on ice for 15 min. Afterwards, cells were put to thermoblock set on 42 °C for 30 seconds and immediately after, transferred back onto the ice. 250 µl of SOC media was added to cells and the mixture was placed into an incubator at 37 °C, shaking horizontally, 200 rpm for 1 hour. Finally, cells were plated onto LB/Agar Ampicillin plates and incubated at 37 °C overnight. Next day, colonies were screened for the presence of FmCB using PCR. The composition of the reaction and cycling conditions are described in Table 20 and Table 21, respectively. Afterwards, samples were run on 2% agarose gel for 30 minutes/100V. Finally, the gel was examined under ultraviolet light.

Table 19: **mRNA and protein sequences of original (transcriptomic) and modified FmCB. In grey**: an open reading frame; **dash**: stop codon.

	5' - 3' nucleotide sequence/protein sequence
FmCB original nucleotide (1126bp)	GATGAACTGGCCAGTTGTGTTTGCCGTCATCGCCGTTGTCCAGGCGATTAAACCGAATTACAAATCGC AATTCGAGGCGTTCTCTGATGAACTGATTCGATATGTGAATGAAGAATCCGGTGCCACTTGGAAAGCA GCACGGTCCGCCCAATTCAACAGTATTGAACAATTCAAAGTGCATTTGGGAGCCCTAGCTGAAACACC GGAACAACGAATTTCCCGGCGACCAACCATAAAACATCACGTATCGGCAAATGATCTACCAGAGTCTT TTGATGCTCGAAATCAATGGCCACATTGTTCGTCGATTCCGGAGATTCGCGACCAGGCCAGTTGTGGTT CCTGTTGGGCTACGGCTGCAGTCAGCGCAATGTGTGACCGCGTGTGTATTCATTCTAACGGTAAAATT ACACCACGTTTGAGCACAATGGATCCGCTCACTTGTTGCACAAATTGTGGAAACGGATGCAAAGGTGG CTATCCACCGAAGGCTTGGGATTATTGGAAGAGTGAAGGCGTTGTGACCGGTGGTACCTGGGAAAATC GAACTGGTTGCCAGCCTTTCATATTTGCCAAGTGCGATCACGTTGGTAACTCTCGAAAATATCCTCCTT GTCCTCACTCCATCTATCCGACACCTCCTTGTGCCAAAACTTGCCAAACTGGATATAACAAAACGCTTG AGAAAGATAAGTTTTACGGAAATTCATCTTACAATGTGGATGCACATGAGTCTAATATCATGTACGAG ATCATGAAAAACGGTCCAGTCGAAGCGAGTTTCGTCGTGTTCACGGACTTCAGCGTGTACAGAAGTGG GATTTACCATCACGTCTACGGTGGTCCCGTAGGCAGACATGCAATTCGCGTTATCGGTTGGGGTGTTG AGGATGGTGTGCCTTATTGGTTGATGGCAAATTCGTGGAATGAGGAATGGGGCGAGAATGGATACTTC CGAATTAAAAGGGGAATGAACGAATGTGGCATCGAAACCGAAGTTGTCGCCGGAATGCCACGTCTCT GAGCAACATCGGAAAGACTGACAAACCTCTCTGTCTAAATAACTGTTTGGGATTTGCAATACATTTAT TGCTAACTTCAAAAAAAAAAAAAAAAAAAAAAAAA
FmCB original translated to protein	MNWPVVFAVIAVVQAIKPNYKSQFEAFSDELIRYVNEESGATWKAARSAQFNSIEQFKVHLGALAETPEQR ISRRPTIKHHVSANDLPESFDARNQWPHCSSIPEIRDQASCGSCWATAAVSAMCDRVCIHSNGKITPRLSTM DPLTCCTNCGNGCKGGYPPKAWDYWKSEGVVTGGTWENRTGCQPFIFAKCDHVGNSRKYPPCPHSIYPTP PCAKTCQTGYNKTLEKDKFYGNSSYNVDAHESNIMYEIMKNGPVEASFVVFTDFSVYRSGIYHHVYGGPV GRHAIRVIGWGVEDGVPYWLMANSWNEEWGENGYFRIKRGMNECGIETEVVAGMPRL-ATSERLTNLSV- ITVWDLQYIYC-LQKKKKKKKK
FmCB modified (1047bp)	ATGAACTGGCCCGTTGTTTTCGCTGTAATAGCTGTCGTCCAAGCTATCAAACCGAACTACAAATCGCA GTTTGAGGCATTTTCAGACGAGTTGATCAGATATGTCAACGAAGAATCAGGAGCTACGTGGAAGGCA GCTAGATCTGCCCAGTTTAACTCCATTGAACAATTCAAAGTCCATCTTGGAGCCTTAGCTGAAACTCCT GAACAACGAATCTCTCGTAGACCAACCATTAAGCACCATGTATCTGCCAATGACTTGCCAGAATCTTT TGATGCCCGTAATCAATGGCCTCACTGTTCATCTATTCCCGAGATTAGGGATCAAGCAAGTTGTGGTTC ATGTTGGGCTACTGCAGCTGTTAGTGCCATGTGCGATAGAGTATGCATACATTCCAATGGCAAGATTA CACCGAGACTGAGTACCATGGATCCATTGACCTGTTGCACAAATTGCGGGAATGGGTGTAAGGGTGGT TATCCACCAAAAGCTTGGGATTACTGGAAAAGCGAAGGAGTTGTGACTGGTGGTACTTGGGAACAGA GAACTGGATGTCAGCCATTCATTTTCGCTAAATGCGACCATGTGGGAAACTCCAGAAAGTATCCTCCC TGTCCACATTCCATTTATCCTACACCACCTTGTGCTAAGACTTGTCAAACTGGCTATCAGAAAACCCTA GAAAAGGACAAATTCTACGGTCAATCGAGCTACAACGTTGATGCTCACGAATCCAACATCATGTATGA GATCATGAAGAATGGTCCTGTTGAGGCTTCATTCGTTGTGTTTACGGACTTTTCTGTTTACAGATCTGG TATTTACCATCACGTCTATGGAGGTCCTGTTGGACGTCATGCAATCAGAGTAATTGGCTGGGGTGTGG AAGATGGAGTGCCATACTGGCTTATGGCCAACAGTTGGAATGAGGAGTGGGGTGAGAACGGCTACTT TAGGATAAAGAGAGGGATGAATGAGTGTGGTATTGAGACAGAAGTCGTTGCAGGAATGCCACGATTG GGTCCTCATCACCATCACCATCACTAA
FmCB modified translated to protein	MNWPVVFAVIAVVQAIKPNYKSQFEAFSDELIRYVNEESGATWKAARSAQFNSIEQFKVHLGALAETPEQR ISRRPTIKHHVSANDLPESFDARNQWPHCSSIPEIRDQASCGSCWATAAVSAMCDRVCIHSNGKITPRLSTM DPLTCCTNCGNGCKGGYPPKAWDYWKSEGVVTGGTWEQRTGCQPFIFAKCDHVGNSRKYPPCPHSIYPTP PCAKTCQTGYQKTLEKDKFYGQSSYNVDAHESNIMYEIMKNGPVEASFVVFTDFSVYRSGIYHHVYGGPV GRHAIRVIGWGVEDGVPYWLMANSWNEEWGENGYFRIKRGMNECGIETEVVAGMPRLGPHHHHHH--

Table 20: **Composition of the PCR reaction for TOP10 cells colony screening.**

component	volume	final concentration
DreamTaq Green PCR Master Mix (2X) (ThermoFisher)	10 µl	
forward primer: α-Mating Factor	0.1 µl	0.2µM
reverse primer: CYC1	0.1 µl	0.2µM
nuclease-free water	9,8 µl	
total volume	20 µl	

Table 21: **PCR thermal cycling conditions for TOP10 colony cells screening.**

step	temperature °C	time	number of cycles
initial denaturation	95	2 min	1
denaturation	95	30 s	35
annealing	55	30 s	35
extension	72	1 min	35
final extension	72	2 min	1

Two positive colonies were selected for further cultivation in 1 l autoclaved LB media at 37 °C overnight. Next day, plasmids were isolated from cells using PureLink™ HiPure Plasmid Maxiprep Kit (Invitrogen). The procedure was carried out according to the manufacturer's protocol and at the last step, DNA pellet was dissolved in 300 µl of TE Buffer. Plasmid concentration was measured in POLARstar Omega (BMG LABTECH) and DNA was sent to Source BioScience (UK, Nottingham) for sequencing to check if there was any mutation during multiplication in bacterial cells.

3.3.3 Transformation of PichiaPink yeast cells

Plasmid DNA was digested to be prepared for the transformation of yeast cells. The digestion was carried out in a total volume of 50 µl and consisted of 5 µl CutSmart® Buffer, 3 µl SpeI restriction enzyme (both New England BioLabs), 100 µg plasmid DNA and nuclease-free water to 50 µl. This mixture was incubated at 43 °C for 1,5 hours. A small aliquot of the digest was examined on 1,5% agarose gel electrophoresis for complete linearization. Afterwards, the digestion was heat deactivated at 80 °C for 20 minutes. Finally, the digest was purified by ethanol precipitation using 1/10 volume of 3 M sodium acetate and 2,5 volumes of 100% EtOH. Samples were put to -20 °C freezer for 30 minutes, then centrifuged for 5 minutes/12000 x g. DNA pellet was washed in 80% EtOH, air-dried, resuspended in 10 µl of sterile nuclease-free water and stored at -20 °C.

PichiaPink strain no. 1 cells from glycerol stock were streaked onto YPD/agar plates (for recipes see Table 22) and incubated for 3 – 5 days at 25 °C. Single colonies were picked from plates to inoculate 10 ml of YPD media in 125 ml sterile baffled flasks. Cultures were incubated shaking at 300 rpm at 28 °C for 1 – 2 days. This starter culture was used to inoculate 100 ml of YPD in a sterile 1-litre flask. Next incubation lasted 1 – 2 days until OD_{600} of 1.3 – 1.5 was reached. After that, cells were transferred into 0.5-l flasks and centrifuged at 1,500 x g at 4 °C for 5 minutes. The pellet was resuspended with 250 ml of ice-cold, sterile water. The cells were centrifuged again like in the previous step and the pellet was resuspended in 50 ml of ice-cold, sterile water. Centrifugation step was repeated again, and the pellet was resuspended in 10 ml of ice-cold 1 M sorbitol. The cells were centrifuged for the last time and the pellet was resuspended in 300 µl of ice-cold 1 M sorbitol. The cells were transferred into 2 ml tubes and kept on ice.

After preparation of the digested plasmid DNA and yeast cells, the next step was electroporation, as it is a recommended method of transformation for PichiaPink cells. Firstly, 100 µl of the electrocompetent PichiaPink cells were mixed with 100 µg of linearized plasmid DNA, transferred to an ice-cold 0.2 electroporation cuvette and incubated for 5 minutes on ice. Afterwards, cells we pulsed in a MicroPulser Electroporator (Bio-Rad) using the fungal mode. Immediately after pulsing, 1ml of ice-cold YPDS media was added to the cells and mixed by pipetting up and down. Then, the cuvette was incubated at 25 °C for two hours. After incubation, 100 – 300 µl of the cell mixture was spread onto PAD plates. Finally, plates were incubated for 3 – 10 days at 25 °C until white and pink single colonies were formed.

To check whether colonies contained inserted target gene, direct PCR screening of PichiaPink™ clones was performed for 24 single white colonies. Using a sterile pipette tip, a small amount of cells from each of 24 colonies was scraped from the original PAD plate and

transferred to 75 µl of TE buffer in a 1.5ml tube. The same colonies were re-plated to fresh PAD plates. Samples with the mixture of TE buffer and yeast cells were microwaved on high (800 V) in 4 rounds (for 3.5; 2; 1.5; 0.5 minutes), and after each round, tubes were briefly vortexed. After the last round of microwaving, samples were put in -80 °C freezer for 10 minutes. Then, the samples were incubated at 95 °C for 2 minutes and centrifuged at 2,500 rpm for 15 minutes. 5 µl of each sample supernatant was added to the PCR reaction described in Table 23 and thermal cycling conditions were the same as in Table 21. Samples were run on 2% agarose gel for 30 minutes at 100V. Finally, the gel was examined under ultraviolet light for right-size amplification products.

Table 22: **Media and solutions used for the transformation of yeast cells.**

solution/media	composition	volume	autoclave
YPD/agar plates	25 g YPD Broth 7.5 g Agar 500 ml dH_2O	0.5 l	yes
YPD media	50 g YPD Broth 1 l dH_2O	1l	yes
1 M sorbitol	91 g sorbitol 0.5 l dH_2O	0.5 l	yes
YPDS media	25 g YPD Broth 91 g sorbitol 0.5 l dH_2O	0.5 l	yes
PAD (Pichia Adenine Dropout) plates	20 g D-glucose (dextrose) 6.7 g YNB without amino acids 0.78 g SM2 – adenine 20 g agar 1 l dH_2O	1 l	yes

Table 23: **Composition of PCR reaction for yeast colony screening.**

component	volume	final concentration
DreamTaq Green PCR Master Mix (2X) (ThermoFisher)	10 µl	
forward primer: α-Mating Factor	0.1 µl	0.2µM
reverse primer: CYC1	0.1 µl	0.2µM
nuclease-free water	4.8 µl	
supernatant from yeast colony	5 µl	
total volume	20 µl	

3.3.4 Small-scale expression

In total, 24 single white colonies from the PAD plates were used to inoculate 5 ml of BMGY media in 24-well plates. Same colonies were also re-plated to fresh PAD plates to keep the colonies in case of a successful expression. Cells were incubated at 28 °C at 200 rpm overnight and next day, cultures were centrifugated at 1,500 x g for 5 minutes at room temperature. The supernatant was replaced by 2 ml of BMMY induction media and plates we incubated at 16 °C, 200 rpm overnight. Next two days each colony culture was fed by 100 µl of 40% methanol to induce expression, once every day. The third day, cells we centrifuged at 1,500 x g for 10 minutes and supernatants were examined for a presence of the protein using “pull-down” by

Ni-NTA Agarose affinity chromatography matrix (Quiagen), SDS-PAGE gel electrophoresis, western blotting and enzymatic assays.

Table 24: **Recipes for BMGY and BMMY media.**

solution/media	composition	volume	autoclave
YNB (13.4% Yeast Nitrogen Base with Ammonium Sulphate, without amino acids)	34 g YNB without ammonium sulphate and amino acids 100 g ammonium sulphate 1 l dH_2O	1 l	yes
0.02% biotin	20 mg biotin 100 ml dH_2O	100 ml	no – filter sterilize
5% methanol	5 ml 100% methanol 95 ml dH_2O	100 ml	no – filter sterilize
10% glycerol	100 ml 100% glycerol 900 ml dH_2O	1 l	yes
1 M potassium phosphate buffer, pH 6	132 ml 1 M K_2HPO_4 868 ml 1 M KH_2PO_4	1 l	yes
yeast extract & peptone	10 g yeast extract 20 g peptone 700 ml dH_2O	700 ml	yes
BMGY (Buffered glycerol-complex medium) / BMMY (Buffered methanol-complex medium)	700 ml yeast extract & peptone 100 ml YNB 2 ml 0.02% biotin 100 ml 1 M potassium phosphate buffer for BMGY: 100 ml 10% glycerol for BMMY: 100 ml 5% methanol	1 l	no

3.3.5 Recombinant protein detection

3.3.5.1 SDS-PAGE

Supernatants from induced colonies were processed using Ni-NTA Agarose affinity chromatography (Quiagen) to bind recombinant protein with His tag. 50 µl of Ni-NTA agarose beads per 1 ml of the sample was employed and this mixture was incubated for 1 hour while shaking horizontally. After incubation, samples were briefly centrifuged on benchtop centrifuge to pellet the beads and the supernatant was discarded. Beads were washed with 200 µl of wash buffer (50 mM sodium sulphate buffer with pH 8; 300 mM NaCl; 10 mM imidazole) and briefly centrifuged. The supernatant was replaced by 50 µl of loading mixture (1/9 volume of DTT: 8/9 volume of 2 x Laemmli Sample Buffer) and samples were incubated at 90 °C for 10 minutes. After this step, samples were centrifuged at 15,000 x g for 5 minutes. 20 µl of each sample was loaded to 4 – 20% Mini-PROTEAN® TGX™ Precast Protein Gel (Bio-Rad) and Precision Plus Protein™ Dual Color Standards (Bio-Rad) was used as a protein marker. Samples ran for 1 hour at 115 V in TGS (Tris-glycine-SDS) buffer using Mini-PROTEAN® Tetra Cell Systems and PowerPac™ HC High-Current Power Supply (both Bio-Rad). After electrophoresis, gels were stained using Bio-Safe™ Coomassie dye (Bio-Rad) for 50 minutes and then, washed by several baths in dH_2O.

3.3.5.2 Western blot

For the purpose of western blotting, samples in SDS-PAGE ran in duplicates and gels used for western blotting were not stained. For blotting, fibre pads and filter paper were soaked in transfer buffer (100 ml 10x Tris Glycine Buffer; 200 ml 100% methanol, 700 ml dH_2O). Gel/Nitrocellulose Membrane, 0.45 µm (Bio-Rad) sandwich was assembled and put into Mini-PROTEAN® Tetra Cell Systems (Bio-Rad) filled with transfer buffer to blotting mark. Blotting was set at 100 V for 1 hour. Afterwards, the membrane was incubated in several solutions for immuno-detection as described in Table 25 at room temperature. As a primary antibody was employed Mouse monoclonal 6X His tag® and Goat Anti-Mouse IgG H&L (HRP) was applied as a secondary antibody.

Table 25: **Procedure of immuno-detection for western blotting.**

Step	Composition of solution	duration
Blocking	5% blocking solution: 1 g of skimmed milk powder in 20 ml TBST (20 mM Tris BASE, 150 mM NaCl, 1% Tween filled to 1 l with dH_2O)	1 hour
Primary antibody	Primary antibody diluted to 1:3000 in 1% blocking solution	1 hour
Washing	TBST	4x5 min
Secondary antibody	Secondary antibody diluted to 1:5000 in TBST	1 hour
Washing	TBST	3x5 min
Visualisation	Solution of SIGMAFAST™ 3,3′-Diaminobenzidine tablets	variable
Stopping	dH_2O	1 min

3.3.5.3 Enzymatic assay

Samples from 24 white colonies were analysed for a specific activity. Supernatants were mixed with an activation buffer (see Table 26) to reach pH 4,5 and activated in 96-well plate for 30 minutes at 37 °C. Universal cysteine peptidase inhibitor E64 was added to control samples. After incubation, running buffer with benzoyl-Phe-Arg-7-amido-4-methyl coumarin substrate (Z-Phe-Arg-AMC) was added to samples. Running buffer was adjusted to a certain pH so that the final mixture with the sample resulted in pH 5.6. Immediately after adding the running buffer, an assay was run at 37 °C in POLARstar Omega (BMG LABTECH). All samples were measured in two technical replicates and blank sample (without substrate) was added too.

Table 26: **Composition of buffers for enzymatic assays.**

solution	composition of the solution	pH
Activation buffer	0.82 g sodium acetate (100 mM) 200 µl of 500 mM EDTA 4 ml of 5 M NaCl 2 ml of 500 mM DTT dH_2O to reach 100 ml legumain (for control samples: inhibitor E64)	Variable according to volume samples
Running buffer	0.82 g sodium acetate (100 mM) 1 ml 1% Brij® L23 200 µl of 500 mM EDTA 200 µl of 500 mM DTT dH_2O to reach 100 ml substrate Z-Phe-Arg-AMC: working concentration 20µM (add right before assay) For control samples add E64: working concertation:	Variable according to volume samples

3.3.6 Large-scale expression

To scale-up protein expression, yeast cells were grown and induced in larger volumes. Firstly, glycerol stocks were made of promising colonies. For that purpose, 200 ml of autoclaved YPD media were inoculated with a single colony from PAD plate using a sterile pipette tip. The culture was incubated at 28 °C shaking 200 rpm until OD_{600} of 3 – 6 was reached. After incubation, cells were harvested by centrifugation (1,500 x g; 16 °C; 5 min) and the supernatant was replaced by 10 ml of YPD/glycerol (autoclaved, ratio: 3:1). Finally, 1ml aliquots were pipetted to cryovials and stored at -80 °C.

For large-scale expression, 1 l of BMGY (Table 24) in the 5-litre sterile baffled flask was inoculated with a small piece of yeast cells from glycerol stock, and culture was incubated at 28 °C overnight shaking at 250 rpm. Next day, cells were transferred into sterile 0.5-l centrifuge flasks and were spun 1,500 x g at 16 °C for 5 min. The supernatant was discarded, and the pellet was dissolved in 300 ml of BMMY and transferred into 2-litre baffled flask. This induction culture was incubated at 16 °C shaking at 250 rpm. Each of the following two days, the culture was fed with 3 ml of 100% methanol. The third day, the purification procedure followed.

Cultures were centrifuged at 5,000 x g for 5 minutes, the supernatant was kept and transferred into a new flask. After each step (centrifugation, filtration, column purification, dialysis) 1 ml sample was collected for a final analysis on the gel. After centrifugation, the supernatant was filtered using vacuum filtration (0.45µm filters). 1 ml of Ni-NTA Agarose affinity chromatography beads (Quiagen) were transferred into 10ml filter column and beads were washed ten times with 1 ml of column buffer. After that, filtered supernatant was passed through the column via gravity flow in the 4 °C cold room. After all the supernatant passed through, the column was washed 15 x with 1 ml of wash buffer and then 5x with 1 ml of elution buffer. Eluted samples were transferred into a sealed dialysis membrane (3 kDa) and dialysed against PBS at pH 7 overnight. Finally, all samples were analysed as described in chapter 3.3.5.

Table 27: **Buffer recipes for large-scale protein purification.**

solution	composition of solution	pH
Column buffer	50 mM sodium phosphate buffer (pH 8) 300 mM NaCl 20 mM imidazole dH_2O to reach 1 l	8 (adjustment with HCl)
Wash buffer	50 mM sodium phosphate buffer (pH 8) 300 mM NaCl 10 mM imidazole dH_2O to reach 1 l	8 (adjustment with HCl)
Elution buffer	50 mM sodium phosphate buffer (pH 8) 300 mM NaCl 250 mM imidazole dH_2O to reach 1 l	7 (adjustment with HCl)

3.3.7 Wide colony screening ("Yeastern blot")

"Yeastern blot" is a technique allowing screening of many colonies for an extracellular protein expression on Petri dish plate using membrane and antibodies, similarly to western blot. The workflow followed principles described in Cregg *et al.* (2009) and was slightly adjusted to our laboratory. The procedure involves growing single white colonies, their induction, lysis and detection of the protein with anti-His antibodies. The PichiaPink strain no. 1 glycerol stock with confirmed ability to produce *F. hepatica* cathepsin L (FhCL) was employed as a positive control.

White individual colonies were transferred into YPD/agar plates (Table 22) in duplicate, along with positive expression control. Plates were inverted and incubated at 25 °C for 2 days. The nitrocellulose membrane was cut to size for a Petri dish and placed on top of the grown colonies onto YPD/agar plate. Then, the membrane was transferred onto a YNB/2% methanol/agar plate colonies side up, plates were inverted and incubated at 25 °C for 2 days. The nitrocellulose membrane was transferred colony-side up into Petri dishes containing filter paper soaked in lysis solutions as follows: 10 min in 10% SDS, 5 minutes in denaturing solution (0.5 M NaOH, 1.5 M NaCl), 5 min in neutralizing solution (1.5 M NaCl, 0.5 M Tris-HCl,pH 7.4), 5 min in a second fresh dish with a neutralizing solution, 15 min in 2X SSC. After that, the membrane was washed twice for 10 min in 15 ml of PBS/0.5% Tween-20, blocked in 15 ml of 5% milk in PBS/0.5% Tween-20 for 1 h, washed twice for 5 min in 15 ml of PBS/0.5% Tween-20, probed with 15 ml of a 1:1000 dilution of mouse monoclonal anti-His6 IgG in PBS/0.5% Tween-20 for 1 hour, washed twice for 5 min in 15 ml of PBS/0.5% Tween-20, probed with 15 ml of a 1: 1000 dilution of secondary anti-mouse antibody in PBS/0.5% Tween-20 for 1 hour and washed three times for 5 min in 15 ml of PBS/0.5% Tween-20. Finally, the membrane was developed by 10 ml solution of SIGMAFAST™ 3,3′-Diaminobenzidine tablets with dH_2O and the development was stopped in dH_2O.

Table 28: **Composition of solutions for "Yeastern blot".**

Solution/media	composition of solution	volume	autoclave
YNB/2%methanol/agar plates	7.5 g agar 250 dH_2O 50 ml YNB (Table 24) 200 ml 5% methanol	0.5 l	Yes, before adding YNB and methanol
10% SDS	10 g sodium dodecyl sulphate 100 ml sterile dH_2O	100 ml	no
denaturing solution	0.5 M NaOH 1.5 M NaCl	100 ml	no
neutralizing solution	1.5 M NaCl 0.5 M Tris-HCl	100 ml	no
20x SSC	3 M NaCl 0.3 M Trisodium citrate dihydrate	100 ml	no

4 Results

4.1 Cathepsin mRNA expression profiles

Total RNA from three biological replicates of *F. magna* adults, miracidia and eggs was isolated, and sample concentrations and purity were measured. 260/230 ratio was in the range of 1.8 – 2.2, and the 260/280 ratio was between 2.0 – 2.1, for all RNA samples, which indicates their solid purity. After the DNAse treatment and reverse transcription, cDNA samples (one from every ontogenetical stage) together with primers (3 target and 17 reference genes) were tested by PCR (Figure 11 and Figure 12) and sequencing. Primers for cathepsins were all confirmed to be specific and to amplify single product. Out of 17 primer pairs for reference genes, only 5 were confirmed to be specific and to amplify single product (FmGAPDH, FmBACT, FmYWHAZ2, FmTUBBI2, FmTUBBI3), other 12 sets either did not amplify any product or were not specific (data not shown).

Figure 11: **Confirmation of cathepsin's primer functionality.** Abbreviations: **L**: 100bp DNA Ladder (Promega); **A1**: adult 1 cDNA; **M1**: miracidia 1 cDNA; **E1**: eggs 1 cDNA.

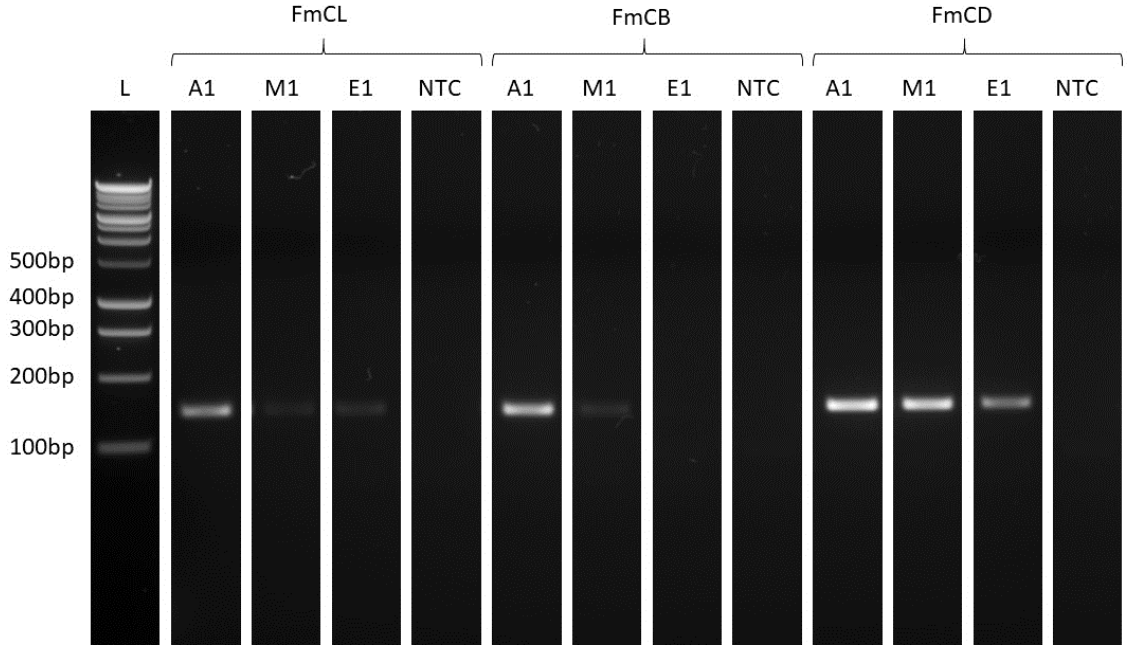
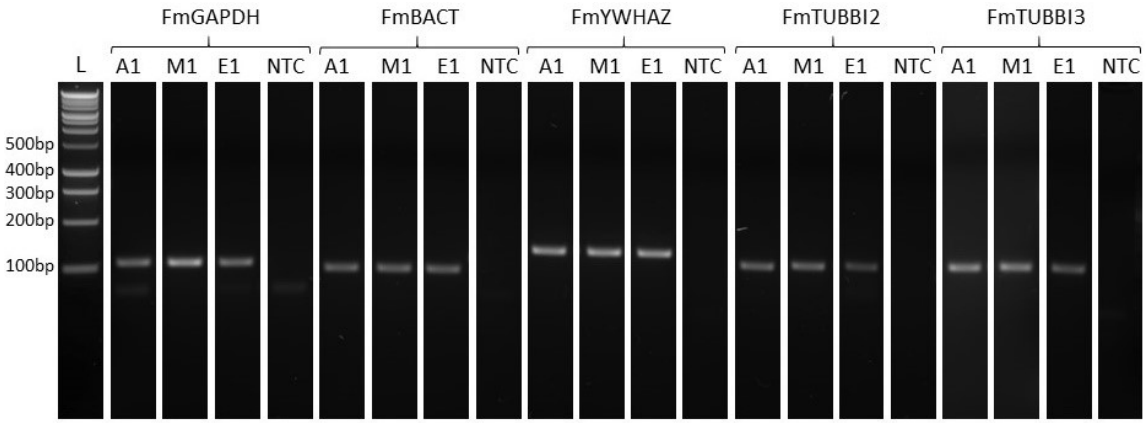


Figure 12: **Confirmation of primer functionality for reference genes.** Abbreviations: **L**: 100bp DNA Ladder (Promega); **A1**: adult 1 cDNA; **M1**: miracidia 1 cDNA; **E1**: eggs 1 cDNA.



A gradient PCR to determine the best annealing temperatures for primers was performed. As a result, a temperature of 53 °C was estimated as the best annealing temperature for all of the following qPCR experiments. Amplification efficiency was tested by serial dilutions for all pairs of primers in qPCR cycler. Resulting E (efficiency) values of all primers were ranging from 96% to 103%, therefore all of them were suitable for subsequent experiments.

To select suitable stable reference genes, the geNorm analysis was performed for 5 sets of primers. Stability is represented by M value for each reference gene, but it is only valid in comparison to all other tested reference genes and for cDNA from all three life stages. Final M values are shown in Figure 13 and Table 29. M values of stable reference genes should be < 1 if testing heterogenous tissues (which eggs, miracidia and adults definitely are). Even though all of the tested reference genes were by a small difference out of the ideal limit (see Table 29), it is recommended to use a combination of three reference genes with the lowest M value, which serves as a better normalization strategy than using single non-validated reference genes. Therefore, a combination of FmBACT, FmGAPDH and TUBBI3.1 was selected for subsequent measurements of cathepsin expression. It is important to point out, that these reference genes can be only employed for normalization in this combination of three because the stability is based upon the proportion in which their expression changes between different samples. So, the stability does not mean that the reference gene expression has the same levels in every ontogenetical stage, but the changes in their expression are stable in comparison to each other.

Figure 13: **geNorm analysis performed in qBase+ for 5 reference genes.**

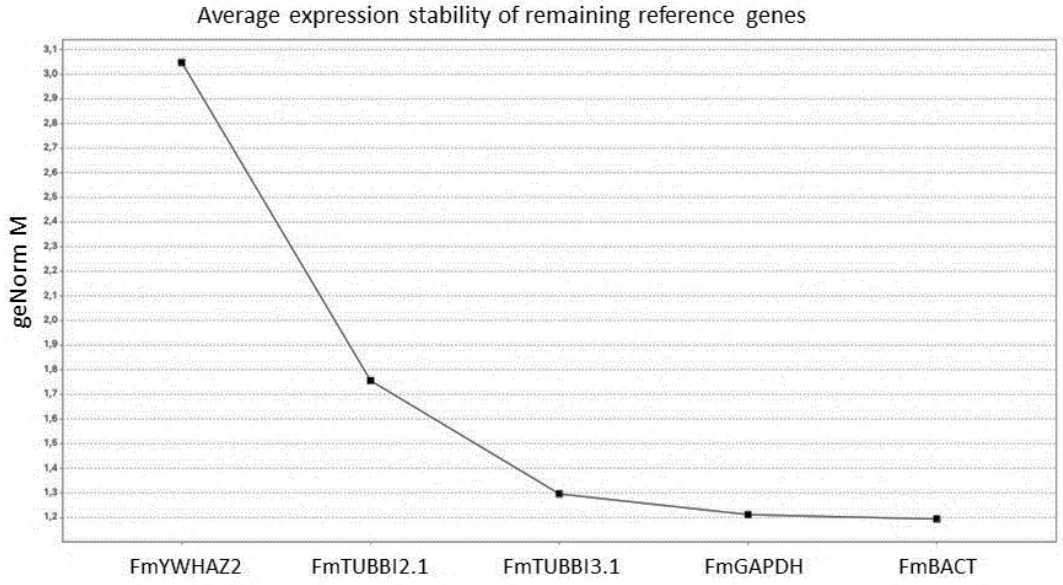


Table 29: **M values calculated in geNorm (qBase+, Biogazelle) for 5 reference genes.**

Reference gene	M value
FmBACT	1.19
FmGAPDH	1.21
FmTUBBI3.1	1.30
FmTUBBI2.1	1.76
FmYWHAZ2	3.05

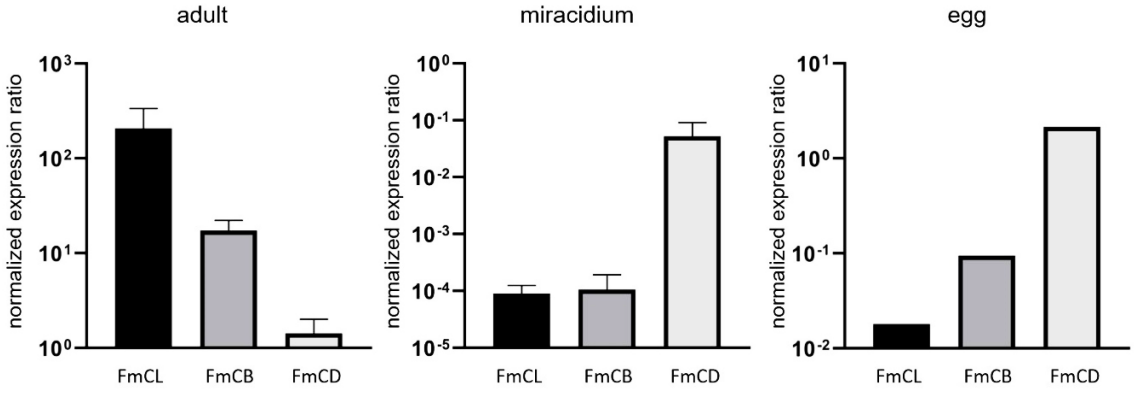
Gene expression levels of three peptidases (FmCL, FmCB, FmCD) were estimated for three life stages of *F. magna*. Relative expression levels were normalized against 3 reference genes (Table 30). For adults and miracidia three biological replicates were measured, and for eggs only one biological replicate because no amplification was detected for the other two egg cDNA samples.

Table 30: **Normalized relative expression levels of FmCL, FmCB and FmCD of three life stages of *F. magna*.** n: number of biological replicates

	Adult (n=3)	Miracidium (n=3)	Egg (n=1)
FmCL	107,957277	0,000055	0,018053
	163,506862	0,000126	
	351,146137	0,000087	
FmCB	17,601787	0,000054	0,094629
	22,083149	0,000206	
	12,232051	0,000055	
FmCD	0,881266	0,025350	2,148636
	1,353355	0,097431	
	2,045700	0,034308	

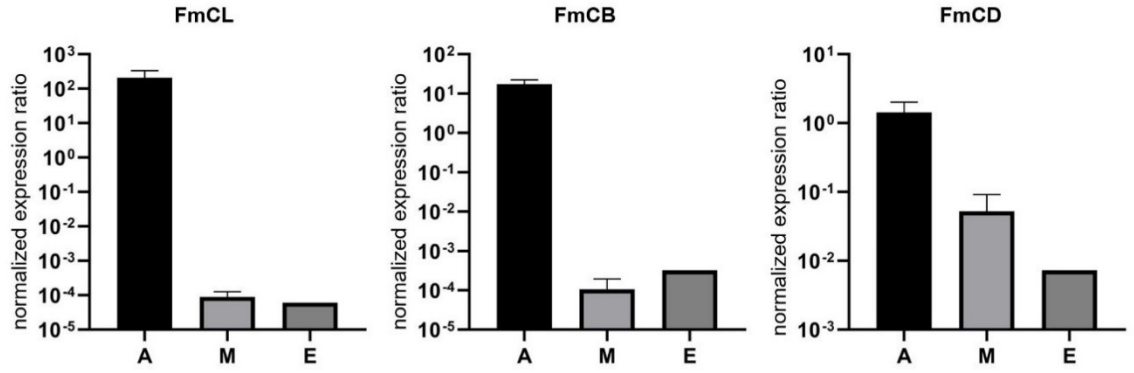
For the first type of the qPCR experiment, cathepsin expression levels were compared to each other in the frame of one developmental stage (Figure 14). FmCL is the most expressed peptidase in adults, 12 times more expressed than FmCB and almost 150 times more than FmCD. Cathepsin expression profiles of miracidia and eggs show an opposite trend. For both, the most expressed peptidase is FmCD. In miracidia, FmCB and FmCL are around 500 times less expressed than FmCD. Finally, FmCD in eggs has approximately 120x and 25x higher expression than FmCL and FmCB, respectively.

Figure 14: **Expression profiles of FmCL, FmCB and FmCD of three life stages of *F. magna*.** Error bars represent the standard deviation of biological replicates.



For the second type of qPCR experiment, relative gene expression levels of each cathepsin were compared among three life stages (Figure 15). FmCL was by far the most expressed in adults, with more than 10^6 times higher expression in adults when compared to miracidia and eggs. In the case of FmCB, expression levels are also highest in adults, and the expression of this peptidase is around 10^5 times lower in miracidia and eggs. Finally, FmCD reaches the highest expression levels in adults too, with almost 30 times and nearly 200 times higher expression in adults than in miracidia and eggs, respectively.

Figure 15: **Differences in expression levels of FmCL, FmCB and FmCD across three developmental stages**. A: adults, M: miracidia, E: eggs. Error bars represent the standard deviation of biological replicates.



4.2 RNA *in situ* hybridization

Three cathepsins (FmCL, FmCB and FmCD) were localized using anti-DIG RNA anti-sense probes, which were applied to paraffin sections of adult worm *F. magna*. Sense probes were employed as negative controls for each hybridization experiment and each cathepsin. A successful hybridization is represented by the permanent coloured (pink) signal, which can be detected at the cell level. This method has a qualitative, not quantitative character, therefore detected signals should be assessed rather like a positive/negative signal, and not by the means of the signal intensity. All pictures were taken under a light microscope and studied organs/tissues include gut, vitellaria, testes, ovary and uterus. The summary of detected localization all studied cathepsins is in Table 31 and detailed description of results follows in three chapters below.

Table 31: **Summary of ISH detected signals in various organs.**

	FmCL	FmCB	FmCD
gut	gastrodermis	gastrodermis, parenchymal cells underlining gut	gastrodermis, parenchymal cells underlining gut
vitellaria	nurse cells	nurse cells	nurse cells
testes	spermatogonia	spermatogonia	spermatogonia
ovary	oogonia/nurse cells	oogonia/nurse cells	oogonia/nurse cells
uterus	no signal	cells of newly forming eggs	uterus epithelium

4.2.1 Localization of cathepsin L

Cathepsin L mRNA was detected in epithelial cells of the gut. In positive sections (Figure 16), individual gastrodermal cells with a light nucleus can be recognized. The positive signal within the gastrodermis was observed in the proximal part, as well as in the distal part of the gut. FmCL anti-sense probe signal was also detected in vitellaria, specifically in cells (referred as “nurse cells”) located mainly in the peripheral zone of the vitelline gland (Figure 17). There was no signal detected within vitelline cells, which contain apparent yellowish shell globules. Vitelloducts which carry mature vitelline cells to ovovitelloduct were investigated as well, and no signal was detected there. Furthermore, a positive signal for FmCL was observed in testes (Figure 18). At the peripheral zone of the tubule, positively stained spermatocytes were distinguished. In addition, cells located in more of a central part of the tubules showed a weaker signal too. Spermatozoa could be observed as long and thin structures in bundles between spermatocytes in the central zone of testes, displaying no coloured signal. FmCL transcripts were also detected within the ovary in the peripheral zone of the ovarian tubule, where oogonia can be found (Figure 19). The central zone is packed with round oocytes which does not show any signal. Finally, the hybridization did not result in any positive signal in the uterus (Figure 20).

Figure 16: **FmCL in epithelial cells of the gut of adult *F. magna*.** Picture on the left: anti-sense FmCL probe showing a specific pink signal within gastrodermal cells. Picture on the right: sense FmCL probe showing no signal. **G**: gastrodermis, **L**: gut lumen. **P**: parenchyma.

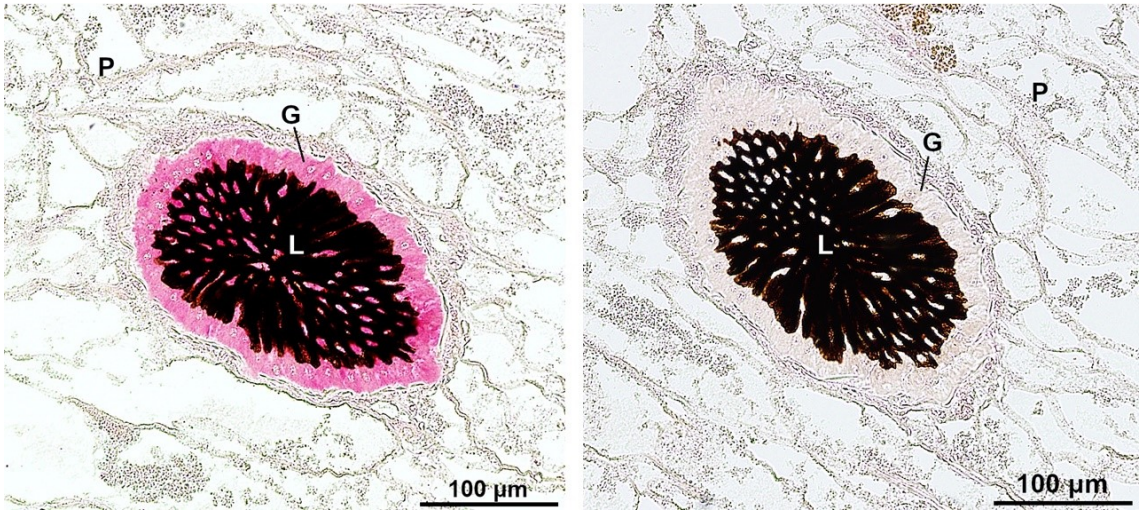


Figure 17: **FmCL in vitelline glands in adult *F. magna*.** Picture on the left: anti-sense FmCL probe showing pink signal in „nurse cells“ in vitelline glands. Picture on the right: control sense FmCL probe showing no specific signal. **NC**: „nurse cell“, **VC**: vitelline cell, **VG**: vitelline gland, **VD**: vitelloduct, **SG**: shell globules, **P**: parenchyma **A**: artefact.

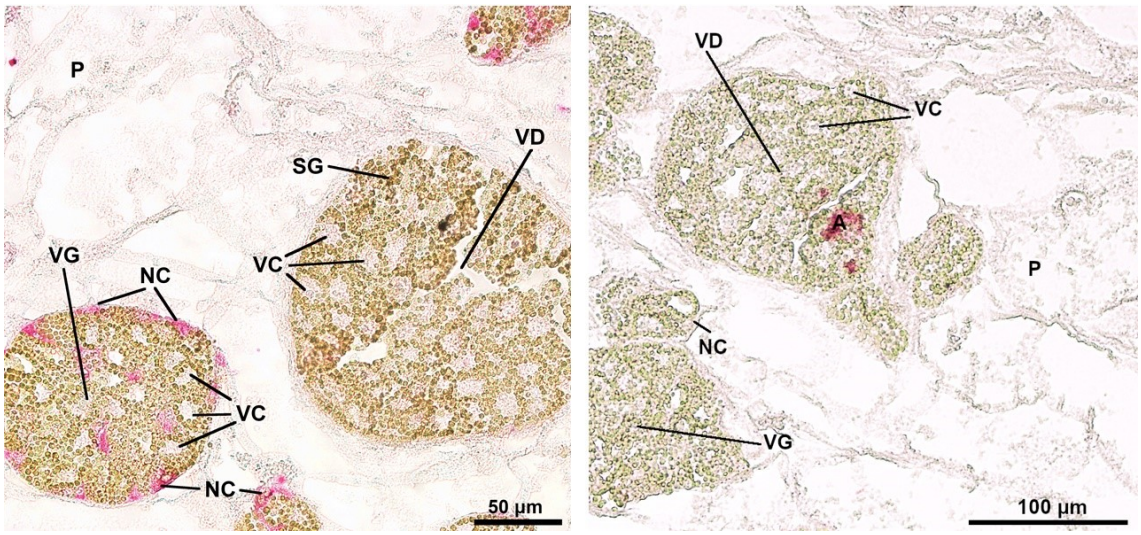


Figure 18: **FmCL in testes of adult *F. magna*.** Picture on the left: anti-sense FmCL probe showing pink signal in spermatogonia. Picture on the right: control sense FmCL probe showing no signal. **SG**: spermatogonia, **SZ**: spermatozoa, **ME**: meiotic division, P: parenchyma

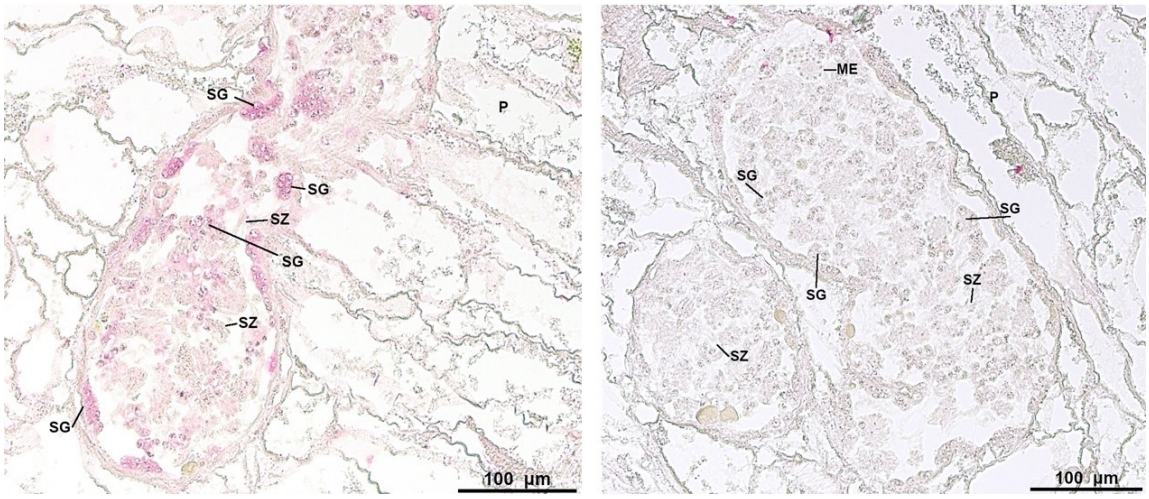


Figure 19: **FmCL in the ovary (germarium) of adult *F. magna*.** Picture on the left: anti-sense FmCL probe showing pink signal in the peripheral zone where oogonia can be found. Picture on the left: control sense FmCL probe showing no specific signal. **P**: parenchyma, **OC**: oocytes, **OG**: oogonia.

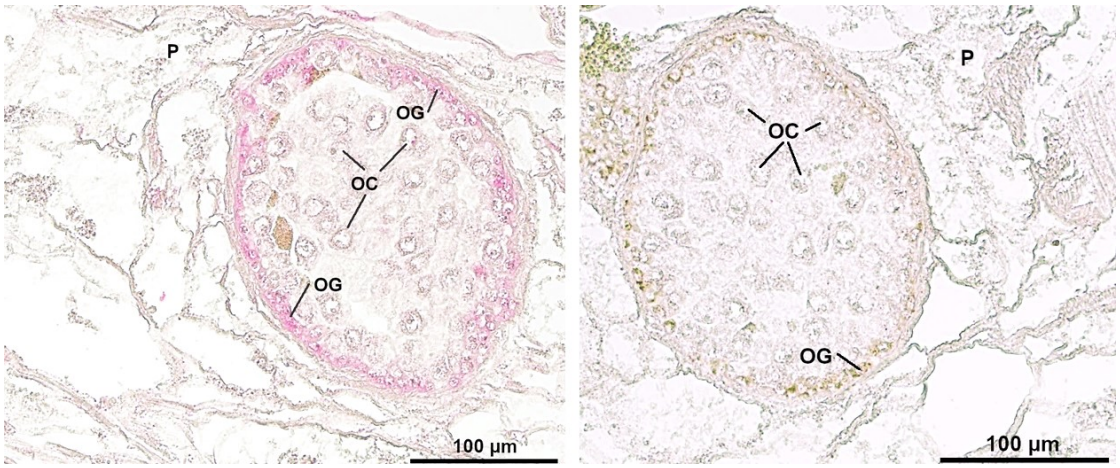
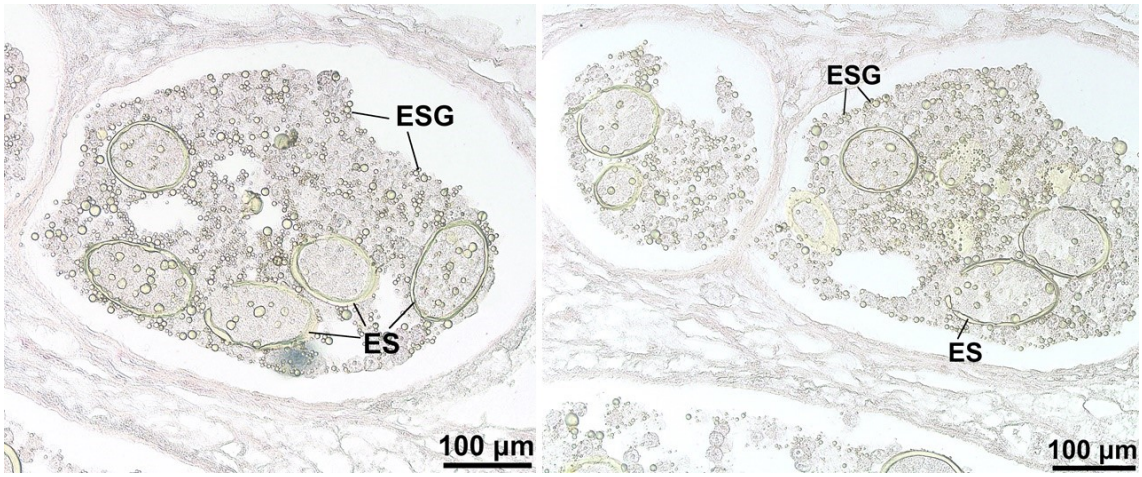


Figure 20: **FmCL in the uterus of adult *F. magna*.** The picture on the left: anti-sense FmCL probe showing no signal in cells of the uterus. Picture on the right: sense FmCL probe without signal. **ES**: eggshell, **ESG**: eggshell globules, **P**: parenchyma.



4.2.2 Localization of cathepsin B

Transcripts of FmCB were localized in the gastrodermis and a strong signal was also detected within parenchymal cells underlining the gut's musculature (Figure 21). In testes and vitellaria, the same signal as for FmCL transcripts was detected, which is in spermatocytes and "nurse cells", respectively (Figure 22), and similarly to FmCL, a positive signal was identified in oogonia within the ovary (Figure 23). Finally, FmCB mRNA was localized in the uterus, particularly within the cluster of vitelline cells and oocytes where the formation of new eggs takes place (Figure 24, Figure 25).

Figure 21: **FmCB in epithelial cells of the gut of adult *F. magna*.** Picture on the left: anti-sense FmCL probe showing a specific pink signal within gastrodermal cells and also in surrounding parenchymal cells. Picture on the right: sense FmCL probe showing no signal. **G**: gastrodermis, **L**: gut lumen, **PCG**: parenchymal cells underlining gut, **M**: musculature **P**: parenchyma.

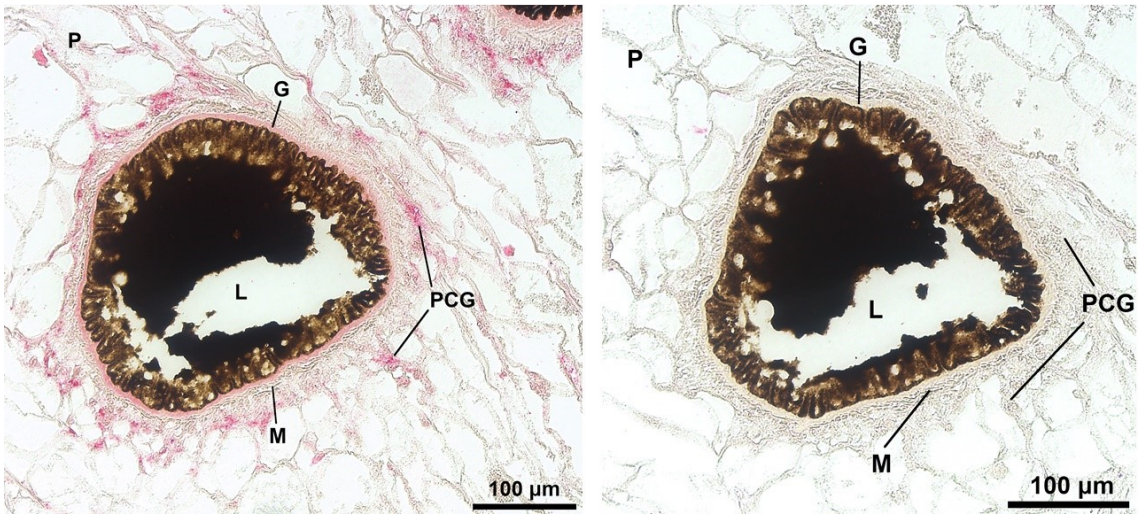


Figure 22: **FmCB in testes and vitellaria of adult *F. magna.*** Picture on the left: anti-sense FmCB probe showing pink signal in spermatocytes and in „nurse cells" within vitelline glands. Picture on the left: control sense FmCB showing no specific signal. **P**: parenchyma, **SG**: spermatogonia, **SZ**: spermatozoa, **VC**: vitelline cell, **NC**: „nurse cell", **ESG**: eggshell globules

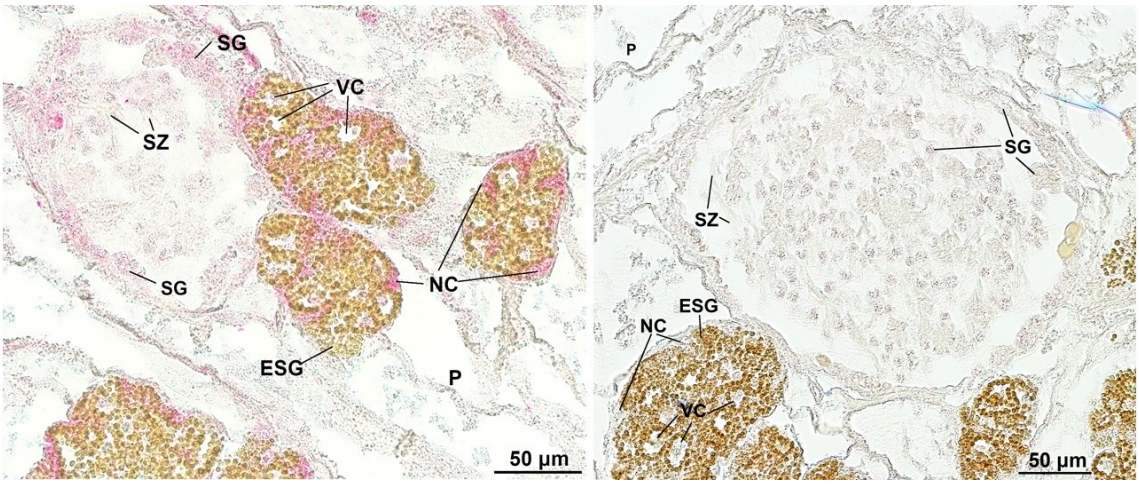


Figure 23: **FmCB in the ovary (germarium) of adult *F. magna*.** Picture on the left: anti-sense FmCB probe showing signal in the peripheral zone where oogonia are found. Picture on the left: control sense FmCB without signal. **OC**: oocytes, **OG**: oogonia, **P**: parenchyma.

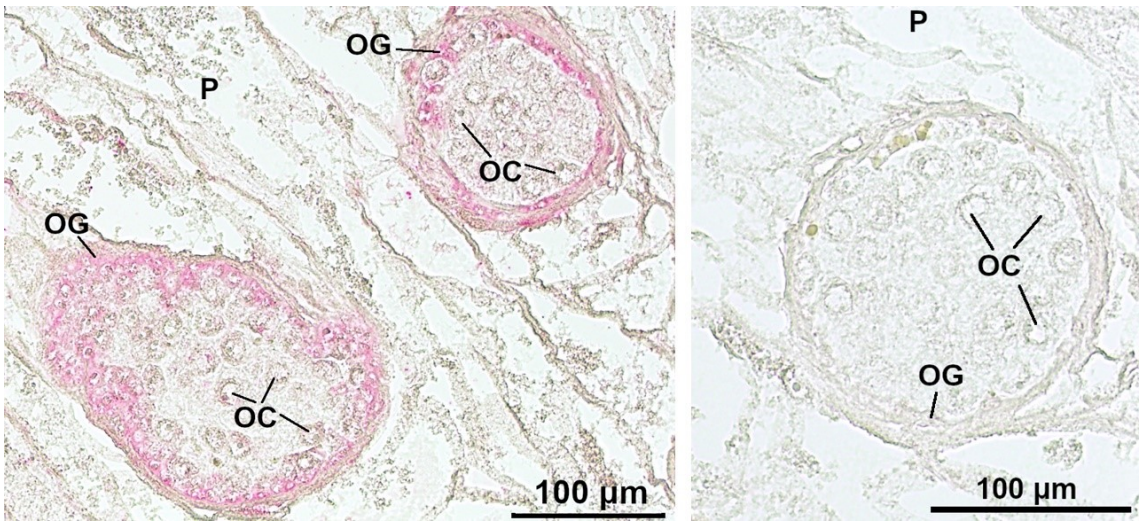


Figure 24: **FmCB in the proximal uterus of adult *F. magna*.** Picture on the left: anti-sense FmCB probe showing signal within cells of newly forming eggs. Picture on the right: control sense FmCB probe without signal. **ES**: eggshell, **ESG**: eggshell globules, **VC**: vitelline cells, **P**: parenchyma.

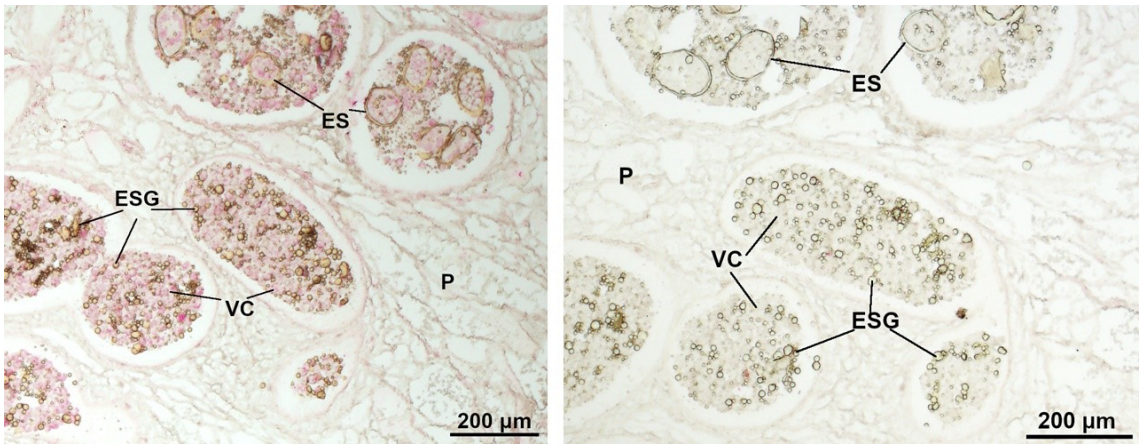
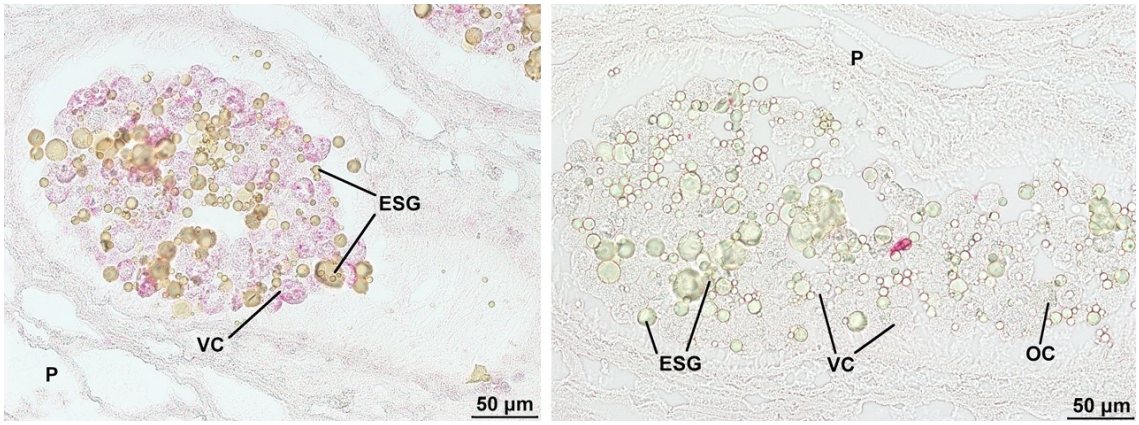


Figure 25: **FmCB in the proximal uterus of adult *F. magna* in detail.** Picture on the left: anti-sense FmCB probe showing pink signal within cells in the proximal uterus. Picture on the right: control sense FmCB probe showing no specific signal. **ES**: eggshell, **ESG**: shell globules, **VC**: vitelline cells, **OC**: oocyte. **P**: parenchyma.



4.2.3 Localization of cathepsin D

Transcripts of FmCD were localized in the gastrodermis (Figure 26). The parenchymal cells surrounding the gut showed coloured signal too, nevertheless not that strong as in the case of FmCB, where the signal was visibly more distinct in the parenchymal cells than in the gastrodermis. Positive signals were observed also in the "nurse cells" of vitellaria (Figure 27), in the spermatocytes within testes (Figure 27), and in oogonia found in the ovary (Figure 28), all just like for FmCB and FmCL signals. In the uterus, no specific signal was detected within the cluster of vitelline cells and oocytes, but there was a recognizable colouring in the epithelial cells (Figure 29, Figure 30).

Figure 26: **FmCD in epithelial cells of the gut of adult *F. magna*.** Picture on the left: anti-sense FmCD probe showing pink signal within gastrodermal cells. Picture on the right: sense FmCD probe showing no signal. **G**: gastrodermis, **PCG**: parenchymal cells underlining gut, **L**: gut lumen, **P**: parenchyma.

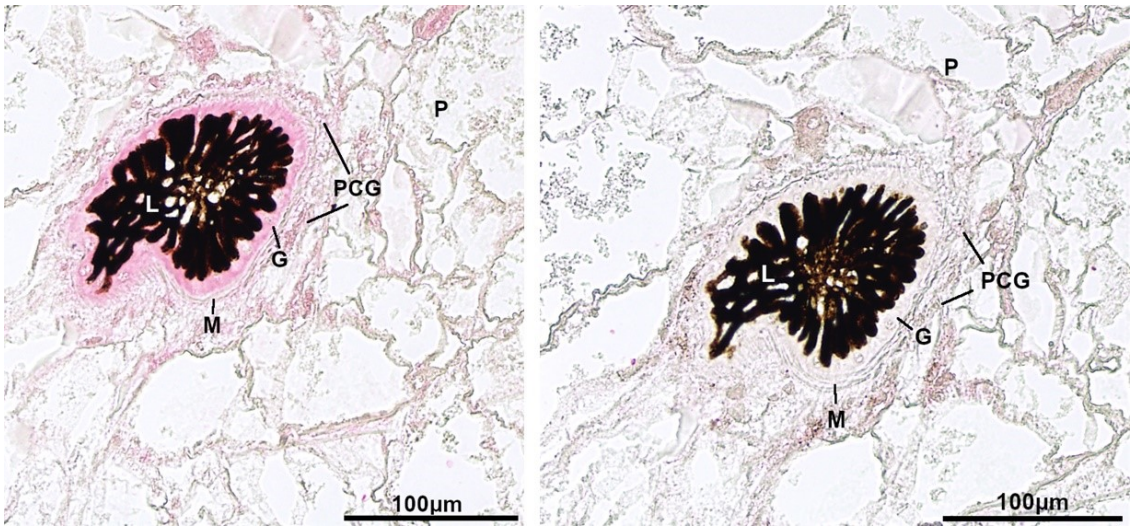


Figure 27. **FmCD in testes and vitellaria of adult *F. magna.*** Picture on the left: anti-sense FmCD probe showing pink signal in spermatocytes and in „nurse cells" within vitelline glands. Picture on the left: control sense FmCB showing no specific signal. **P**: parenchyma, **SG**: spermatocyte, **SZ**: spermatozoa, **VC**: vitelline cell, **NC**: „nurse cell", **ESG**: eggshell globules.

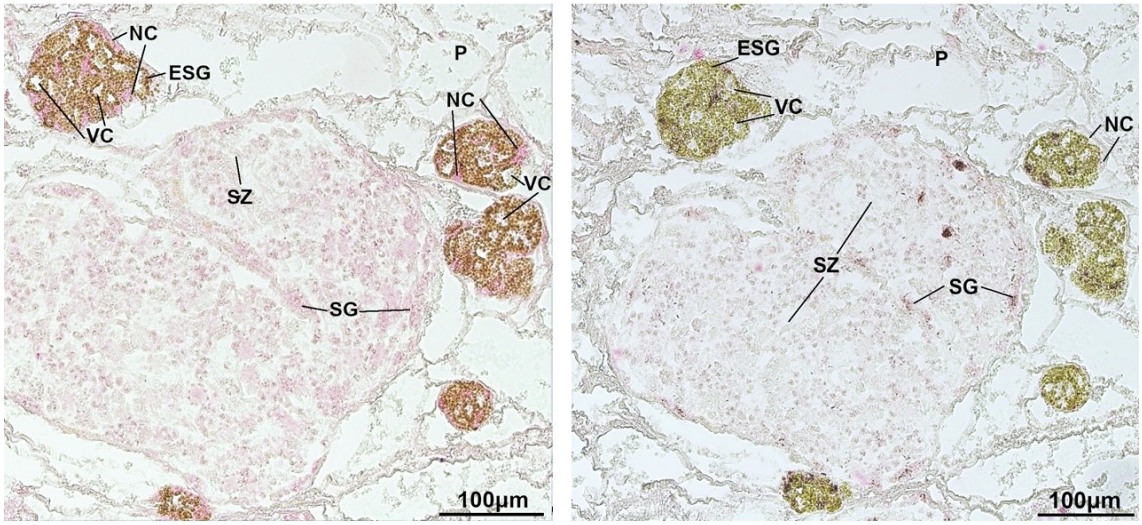


Figure 28: **FmCD in the ovary (germarium) of adult *F. magna*.** Picture on the left: anti-sense FmCD probe showing pink signal in the peripheral zone where oogonia are found. Picture on the left: control sense FmCB probe without signal., **OC**: oocytes, **OG**: oogonia, **P**: parenchyma.

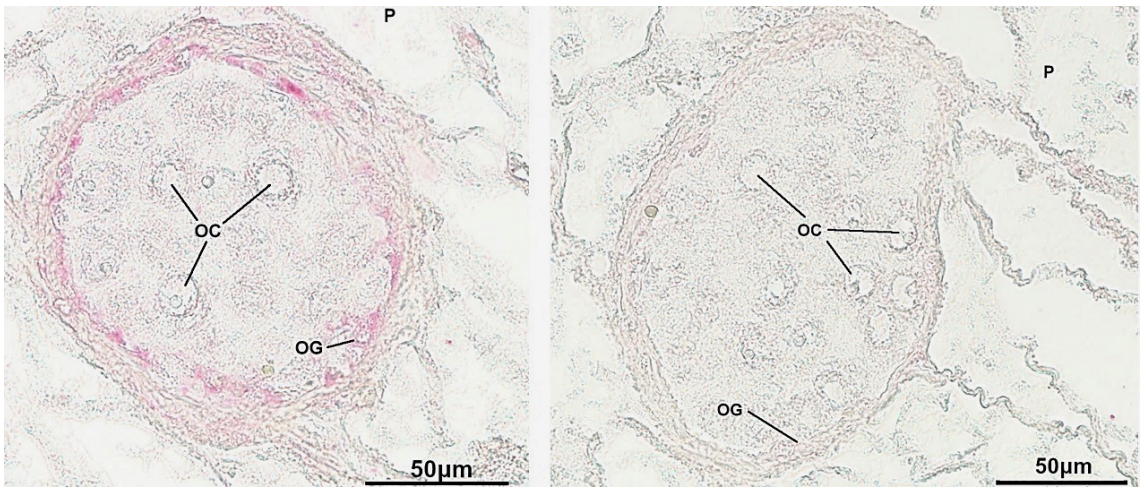


Figure 29: **FmCD in the proximal uterus of adult *F. magna*.** Picture on the left: anti-sense FmCD probe showing pink signal within the epithelium of the uterus. Picture on the right: control sense FmCD probe showing no specific signal. **ES**: eggshell, **SG**: shell globules, **UE**: uterus epithelium. **FE**: formation of the egg, **G**: gut, **P**: parenchyma.

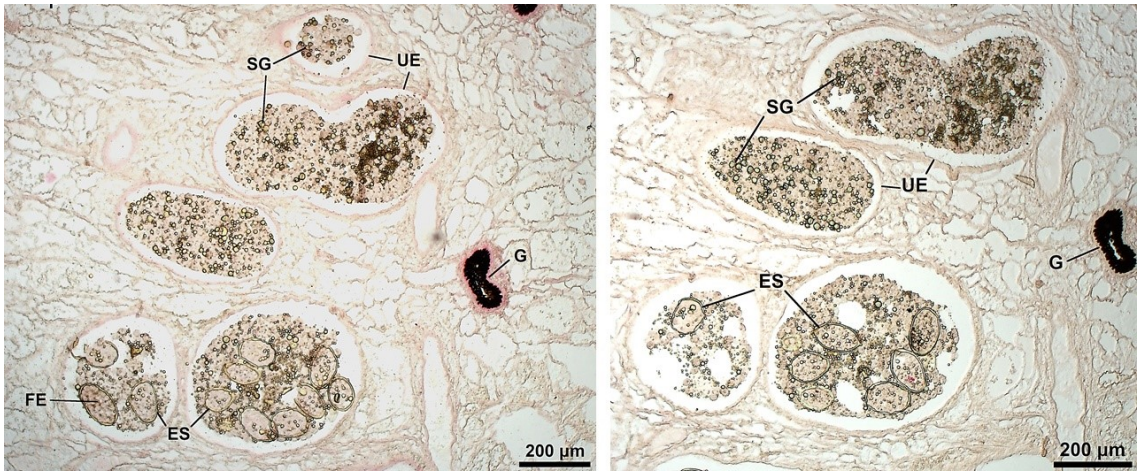
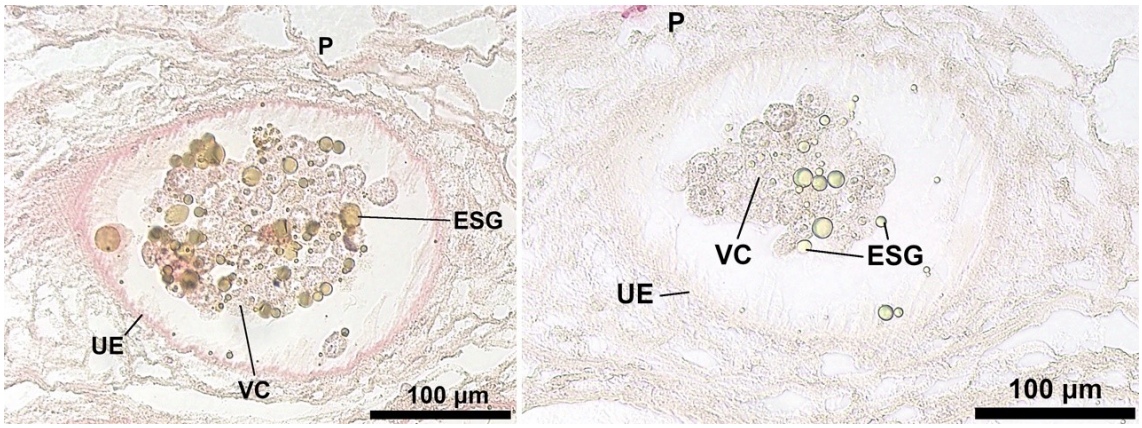


Figure 30 **FmCD in the proximal uterus of adult *F. magna* in detail.** The picture on the left: anti-sense FmCD probe showing pink signal within the epithelium of the uterus. Picture on the right: control sense FmCD probe showing no specific signal. **ESG**: shell globules, **UE**: uterus epithelium, **VC**: vitelline cell, **P**: parenchyma.



4.3 Expression of recombinant cathepsin B in yeast

For the expression of the recombinant FmCB, the PichiaPink™ Expression System (Invitrogen) was applied. Commercially synthesized FmCB sequence in the pPink-alpha HC vector was inserted into TOP10 competent cells (Figure 31). Plasmids were multiplied, isolated and successfully electroporated into PichiaPink™ cells strain no. 1 (Figure 32). Amplicons from TOP 10 cells and PichiaPink™ clones were sequenced and both sequences were verified to be correct, thus mutations during the multiplication process were ruled out.

Figure 31: **Colony PCR of TOP 10 competent cells colonies with FmCB insert.**

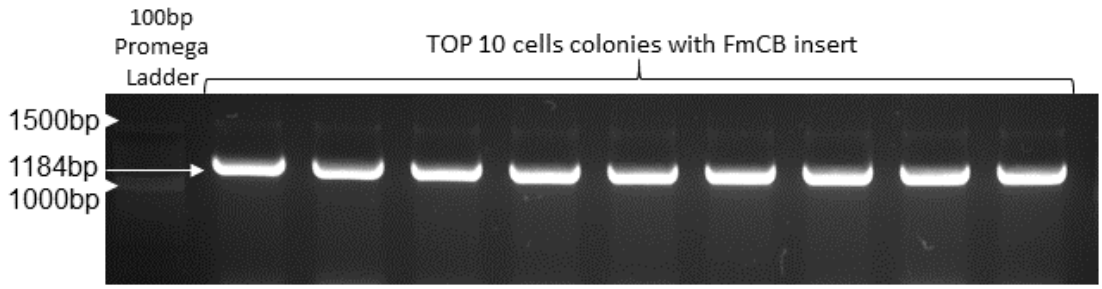
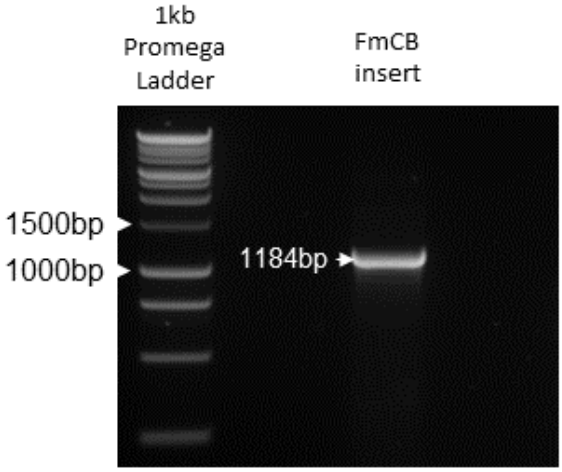
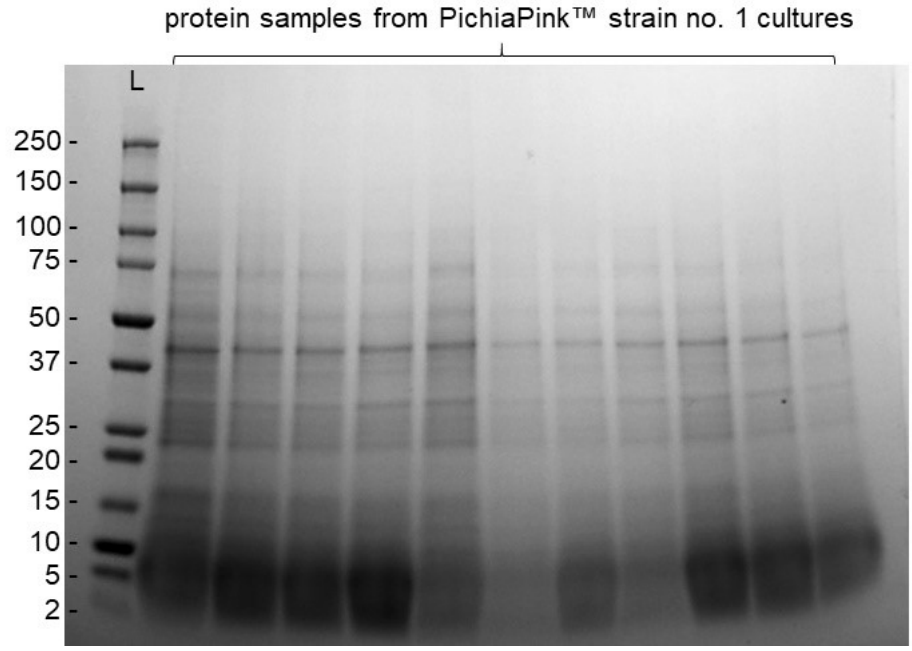


Figure 32: **Direct PCR screening of PichiaPink™ clones for FmCB insert.**



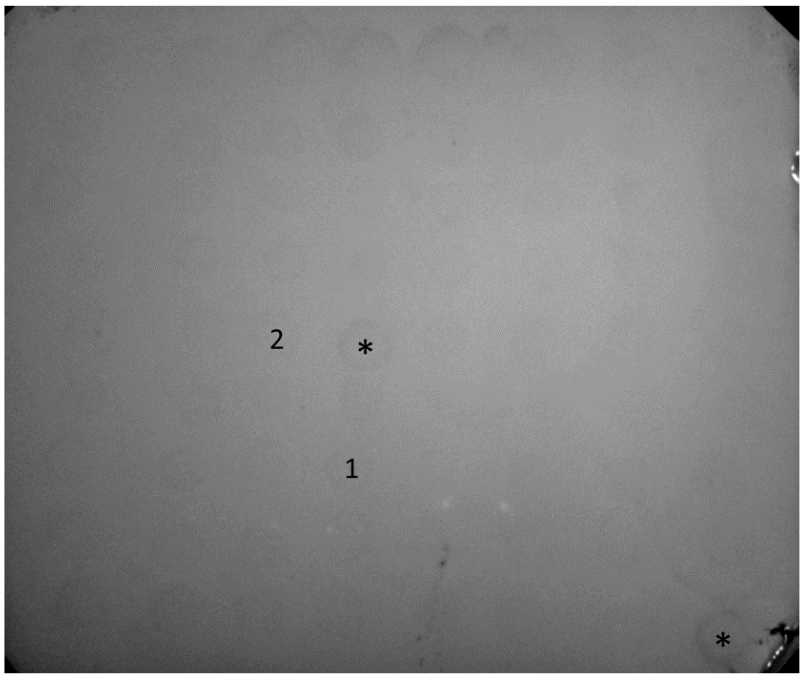
Supernatants from induced cultures were analysed for secreted FmCB using Ni-NTA Agarose affinity chromatography and SDS.PAGE. The molecular mass of the secreted FmCB was estimated using an online software tool (www.bioinformatics.org) and expected to be 39.1 kDa. However the protein gel (Figure 33) possessed a relatively distinct band at the molecular weight between 37 and 50 kDa, Western blot did not confirm any His tagged protein (data not shown). The enzymatic assays did not detect an activity either. The large scale expression experiment was employed anyway to raise the probability of protein detection in case of very low production rates. The supernatants were analysed, yet no protein was detected. Finally, PichiaPink™ strain no. 4, which is a strain derived from strain no. 1 with an extra feature of double knock-out of two yeast proteases, was employed. The analysis of the strain no. 4 protein production was the same as for the strain no. 1 and also the same in the means of negative results.

Figure 33: **Protein gel with supernatant samples from induced colonies of PichiaPink™ cells strain no. 1. L**: ladder



The wide colony screening method ("Yeastern blot") was performed for strain. No 1 and no. 4 to check a large number of colonies (precisely 140 colonies per each strain) for extracellular protein production (Figure 34). Even though the detected signals from tested colonies were not as strong as a positive control, two of them were selected for the large-scale expression, but no production was identified. Hence, the signal that had been observed on the membrane was probably nonspecific.

Figure 34: **„Yeastern blot" for 70 colonies of PichiaPink™ cells strain no. 1.** asterisk: positive control FhCL, **1, 2**: selected colonies for large scale production.



5 Discussion

5.1 Cathepsin mRNA expression profiles

The qPCR method was employed to validate reference genes for the subsequent normalization of FmCL, FmCB and FmCD expression levels measured in three developmental stages of *F. magna*. These levels were compared to each other using a combination of Livak method and normalization strategy using three most stable reference genes.

5.1.1 Analysis of reference genes

Although the selection of the potential reference genes was broad at the beginning (9 candidates, 17 sets of primers), the geNorm analysis could have been eventually done only for five reference genes due to dysfunctional primers. Despite that, three reference genes (FmBACT, FmGAPDH, FmTUBBI3.1) were evaluated as the best for the subsequent data normalization because their stability (M value) was nearly fitting the recommended limits for heterogeneous samples (Hellemans *et al*., 2007). Even though, stably expressed reference genes are crucial for comparisons of relative mRNA levels among parasitical developmental stages, there is a shortage of studies investigating the suitable ones. Often, researchers use only one non-validated reference gene, which can alter the results and even cause questionable conclusions (Vandesompele *et al*., 2002). To date, no evaluation of reference genes has been done for any of fasciolid trematodes. In contrary, Yoo *et al*. (2009) have analysed reference gene stability for *C. sinensis*, suitable for normalization of target genes between adults and metacercariae. Reference genes phosphoglycerate kinase, β-actin, and calcyphosine have shown the best stability. Nevertheless, these findings cannot be compared to ours for several reasons. Most importantly, it is a different species of a liver fluke and the analysis was carried out for different life stages. Our analysis is the first validation of reference genes for members of the family Fasciolidae. However, it can only be employed for a similar comparison among adult, miracidium and egg of *F. magna.* It is also worth mentioning, that the analysis has been conducted for 3 biological replicates of adults and miracidia, but only for one biological replicate from eggs. Because of that, in the future, it would be beneficial to complete the analysis with two more biological replicates for eggs and also to add more reference genes to geNorm analysis. The two failing egg biological replicates in our experiments might have been a consequence of a nonoptimal procedure of RNA isolation, where the crucial step is the disruption of egg-shell.

5.1.2 Gene expression profiles

The relative gene expression level of each cathepsin was compared among the three life stages. All studied cathepsins had the highest expression levels in adult flukes in comparison to eggs and miracidia. The difference between expression levels in adults and the other two life stages was enormous in the case of FmCL. This means that adult *F. magna* expresses high quantities of cathepsin L and that the importance of this peptidase for the adult worm is probably much greater than for eggs and miracidia. The same can be concluded for cathepsin B, where the difference between adult and miracidia with eggs was distinct too. The expression profile of cathepsin D shows a similar trend, but the difference between expression in adults and the other two

developmental stages is not that distinct. A similar qPCR study was performed for *S. mansoni*, Dvořák *et al*. (2009) found out that cathepsin L (SmCL3) is predominantly expressed in adult worms when comparing to eggs. Although this trend corresponds to ours, comparisons of gene expression in other trematode families should be considered critically. In the future, it would be interesting to measure the expression of cathepsins in other *F. magna* life stages, to see, if these peptidases would be upregulated for example in metacercaria or NEJ. Unfortunately, these life stages were not available in our laboratory due to the missing parasitical life cycle and specific intermediate hosts.

The second type of comparisons focused on a single life stage and differences in cathepsin expression levels within. Results suggest that FmCL is a dominant cathepsin for adult fluke, followed by cathepsin B in the second place. This corresponds to findings of Cantacessi *et al*. (2012) since according to the quantitative transcriptomic data, cathepsin L mRNA is the most expressed peptidase and cathepsin B is the second one. Our investigation also revealed, that the least expressed from studied peptidases in adult fluke was cathepsin D. Such results reflect the importance of these proteolytic enzymes in the biology of adult worm *F. magna*.

Interestingly, FmCD was distinctly the most expressed cathepsin in miracidia. Although it remains unclear, what is this peptidase responsible for, it is possible that it could function as a penetration enzyme or lysosomal enzyme playing role in intracellular protein catabolism, nevertheless, it is an open question and thus it would require further investigation. In *S. mansoni* miracidium proteome, cathepsin D was identified among secreted proteins and authors of the study hypothesize that this aspartic peptidase might play a role in the metabolism of miracidium, for example as a digestive enzyme prepared for a quick hydrolysis of snail's Hb, immediately after penetration into the host (Wang *et al*., 2016). Cathepsin D mRNA was also identified in *S. japonicum* miracidium, but no further investigation has been carried out on this (Verity *et al*., 1999).

FmCD was also the most expressed cathepsin in eggs, followed by FmCB and FmCL. In conjunction with results of ISH experiments, where FmCD was not detected in forming eggs within the uterus, it is likely, that the expression of cathepsin D is upregulated later, while the eggs are developing in outside environment (isolation of RNA from eggs was carried out after five days of incubation).

It is important to mention, that this experimental design is based on the presumption that SYBR Green, as it is an unspecific dye, binds equally to dsDNA amplicons with the same product length and similar AT% content. This expectation is based on several studies, proposing that SYBR Green has a preference for binding to AT-rich minor grooves on dsDNA and that longer amplification products yield in higher fluorescence signals (Zipper *et al*., 2004; Colborn *et al*., 2008). That is why we designed primers to amplify products with very similar GC% content (see chapter 3.1.4) and to be exactly the same in length (see chapter 3.1.4).

5.2 RNA *in situ* hybridization

The aim if this thesis was to identify a place of transcription of three peptidases (FmCL, FmCB and FmCD) within the body of adult *F. magna*. For that purpose, RNA *in situ* hybridization technique was employed. Localization of cathepsin mRNAs was successful and revealed that transcription of all studied cathepsins takes place in gastrodermal cells, as presumed according to the literature. Remarkably, cathepsin transcripts were also identified in the reproductive system, which means that the functions of these endopeptidases are not only in the digestion of host proteins in the gut but that they have a role in reproduction as well.

5.2.1 Localization of cathepsin L

Cathepsin L mRNA of *F. magna* was localized in **gastrodermal cells** lining the gut of the adult fluke. Unpublished data from our laboratory show that recombinant FmCL is able to cleave bovine Hb and it was also immunolocalized in adult's gut (Kašný *et al*., in preparation). The same localization in the adult worm was observed for *F. gigantica* cathepsin L transcripts by ISH and also by immunohistochemistry for its protein form (Meemon *et al*., 2010; Sansri *et al*., 2013). Identical place of transcription was observed for *F. hepatica* cathepsin L (FhCL) by Dalton *et al.* (2003). Detailed immunohistochemistry experiments by Collins *et al*. (2004) specified the localization of FhCL1 to be denser in the apical part of gastrodermal cells and the ultrastructure investigation revealed that the enzyme is stored in secretory vesicles within these cells. Given the accordance in the localization of cathepsin L among fasciolid adult trematodes, it seems CLs share similar functions in these flukes. The main function in the gut is indeed the digestion of blood and tissue proteins. The proposed digestion process in *F. hepatica* starts with CL proenzyme (inactive precursor) being stored in secretory vesicles within gastrodermal cells. After stimulation by ingested food, vesicles are released into the gut lumen, CLs become activated and can immediately function in protein catabolism (Collins *et al*., 2004)

Transcripts of FmCL were also localized in **vitellaria**. Moreover, it was revealed, that FmCL mRNA is present specifically in "nurse cells" found in the peripheral zone of the vitelline gland. Such localization has not been reported for other fasciolid trematodes so far. The term nurse cell was undertaken from the descriptions of *F. hepatica* reproductive system (Irwin and Threadgold, 1970; Hanna, 2015), on the basis of the similar morphological characteristics of these cells, and as there is no detailed study of the reproductive system of *F. magna* describing the morphology of vitellaria. Nurse cells have cytoplasmatic extensions which surround, supports and possibly nourishes the developing vitelline cells. It was also observed that nurse cells possess junctional complexes connecting them to parenchymal cells (Irwin and Threadgold, 1970). As a result of localizing FmCL in these cells, we can assume, that this cysteine peptidase can also play a role in the production of the vitelline cells. Nevertheless, till now, it is not clear what is the exact function of nurse cells, therefore the function of FmCL within these cells remains to be clarified as well.

In the **testes**, a distinct positive signal of FmCL mRNAs was detected in cells in the peripheral zone of the testicular tubule and less distinct signal in cells in the central part of the tubule. Same as for vitellaria, names of the cells were undertaken from studies about *F. hepatica* morphology (Hanna, 2015). Cells at the peripheral zone correspond to primary and secondary spermatogonia,

which are undifferentiated male germs cells. The cells more in the core of the tubule, with the less distinct positive signal, are similar to tertiary spermatogonia. After several mitotic and meiotic divisions, spermatogonia give rise to spermatocytes and ultimately to mature spermatozoa (Hanna, 2015). In addition, at the periphery of the testis tubules of *F. hepatica,* there has been described as a special layer called a sustentacular tissue (Hanna *et al*., 2012). This tissue envelopes and supports the neighbouring spermatogonia, it is syncytial and it is considered to be analogous to Sertoli cells in testes of vertebrates. The proposed function of the sustentacular tissue is that it is involved in support, nutrition and metabolic exchange for the process of spermatogenesis (Hanna *et al*., 2012). Therefore, if the same tissue would be identified also in *F. magna* and the localization of FmCL would correspond with this tissue, it is likely that cathepsin L could be involved in the heterophagy and digestion in sustentacular tissue. Indeed, for such hypothesis, it would be beneficial to perform detailed morphological studies of *F. magna* reproductive system. To date, cathepsin L has not been detected in testes in any other fasciolid species, but interestingly, CL is known to be secreted in testes by rodents and other vertebrates, and it is involved in formation and differentiation of spermatocytes (Griswold, 1995; Wright *et al.*, 2003).

Finally, FmCL transcripts were localized in the **ovary** within the peripheral cells lining the internal ovarian wall. Again, cells in the ovary were defined according to morphological studies in *F. hepatica* (Hanna, 2015). The positively stained cells were therefore identified as either oogonium, which is a germ cell giving rise to oocytes by mitotic division, or nurse cell. This is the first time when CL was identified in the ovary of fasciolid trematodes, thus there is no proposed function for this peptidase in ovary yet. Nevertheless, cathepsin L was localized and studied in the ovary of some vertebrates. For example, in a murine model, CL's hypothetical function is related to the development of oocytes (Robker *et al*., 2000; Oksjoki *et al*., 2001; Carnevali *et al*., 2006).

Even though CLs were not identified in female reproductive organs in other fasciolid flukes, a vaccine made of recombinant cathepsin L of *F. hepatica*, which was employed to induce protective immunity in cattle and sheep, had effects on egg production of adult flukes within the host. The reason for this might be due to interruption of the ability to feed, and thus to have enough nutrition to reproduce, but another explanation could also be that the antibodies against CL had a direct influence on the egg production which would confirm that CL play an important role in the reproductive system too (Dalton *et al*., 2003).

5.2.2 Localization of cathepsin B

Similarly like FmCL, cathepsin B transcripts were detected in **gastrodermal cells** lining the *F. magna* gut. This is in consensus with ISH localization of cathepsin B in *F. gigantica* adult's gastrodermis (Meemon *et al*., 2004) and also with immunolocalization of newly characterized cathepsin B5 of *F. gigantica* in the digestive tract. For the latter, it was detected that the recombinant protein cleaves Hb (Siricoon *et al*., 2015). Our hybridization also showed a distinct signal in cells under the gut, particularly in parenchymal cells under the gut's musculature. In contrast, Meemon *et al.* (2004) neither Siricoon *et al.* (2015) did not report similar localization. Therefore, the role of CB in the parenchymal cells close to the gut remains to be investigated. In

F. hepatica, CBs were not localized in adults so far, although they were identified in the transcriptome of adult fluke (Robinson *et al*., 2009). *S. mansoni* cathepsin B1 (SmCB1) was confirmed to be one of the main peptidase responsible for the degradation of Hb by adult flukes and was localized exclusively in the parasite's gut (Sajid *et al*., 2003; Caffrey *et al*., 2004). In addition, CB was also immunolocalized in *Clonorchis sinensis* within parasites gut (Chen *et al*., 2011). Given all the previous research of CBs in flukes, it is likely, that FmCB could also be involved in the blood protein digestion, although testing degradation abilities of host native proteins with recombinant FmCB would help to clarify this assumption.

In **vitellaria**, FmCB transcripts were detected in the peripheral nurse cells, identically to FmCL mRNAs. Such localization goes along with observations of Meemon *et al.* (2004) for cathepsin B in *F. gigantica*, although they only report cathepsin B transcripts from a vitelline gland as a whole, but did not specify a particular cellular location within. Authors suggest that cathepsin B present in the vitelline gland might be involved in the processing of the precursors for the eggshell proteins into a proper size before they could be employed for the formation of the mature eggshell proteins. Nevertheless, such a hypothesis needs to be investigated as well as the function of nurse cells themselves.

FmCB mRNAs were detected in the male reproductive organs, particularly in primary, secondary spermatogonia and possibly in sustentacular tissue within **testes**, equally as in the case of cathepsin L transcripts. In *F. gigantica*, cathepsin B transcripts were localized in various stages of spermatogenesis as well (Meemon *et al*., 2004). It is, therefore, possible, that CB along with CL is somehow involved in the spermatogenesis.

Cathepsin B transcripts were detected in oogonia/nurse cells within **ovary**, evenly as FmCL. This is a unique observation, not even Meemon *et al.* (2004) detected CB in *F. gigantica*, although most of our results are in agreement with their observations. Possibly, CB is also transcribed in *F. gigantica* ovary, but it has not been investigated or reported. According to the place of transcription, the function of FmCB in ovary might be related to early stages of oogenesis.

Interestingly, cathepsin B mRNAs were also detected in the **uterus**. Particularly, in the mass of vitelline cells and oocytes which were indistinguishable between each other. For better specification, it is worth mentioning, that the investigated part of the uterus was proximal (close to ootype) and the very proximal part of the uterus might be also termed as ootype, which is described as a part of the reproductive system where mature oocyte, vitelline cells and secretion products of Mehlis' gland meets (Erhardová-Kotrlá, 1971; Hanna, 2015). The same localization was also detected for cathepsin B in the uterus of *F. gigantica* (Meemon *et al*., 2004). As an explanation of such observation, authors suggest that CB might play a role in the cleavage of vitellogenin (precursor of yolk protein). This seems probable, as the mechanism of vitellogenin cleavage by CB was already described in other animals, for example in mosquitos (Cho *et al*., 1999), Drosophila (Medina *et al*., 1988) or even in teleost fish (Sullivan *et al*., 2003).

5.2.3 Localization of cathepsin D

Cathepsin D transcripts of *F. magna* were localized in **gastrodermal cells** (similarly to FmCL and FmCB) and also in parenchymal cells in the gut's proximity (similarly to FmCB). This is a completely new finding in the frame of the family Fasciolidae, in which cathepsin D, in general, has not been studied in detail so far. In contrary, cathepsin D expressed by *S. mansoni* and *S. japonicum* has been localized in the gut of adult flukes (Brindley *et al*., 2001). Cathepsin D in blood flukes is known to be involved in the digestive cascade. Moreover, it is considered to be the first peptidase initiating the Hb cleavage into peptides which are then degraded by cathepsin B and L (Delcroix *et al*., 2006). Cathepsin D-like aspartic protease was also immunolocalized in the gut of adult *O. viverrini* and recombinant enzyme had the ability to cleave Hb and albumin (Suttiprapa *et al*., 2009). Whether the cathepsin D in *F. magna* plays a similar role in digestion remains to be clarified. Although, when considering the results of measuring mRNA expression levels in adult worms (chapter 4.1), it is unlikely that FmCD would be more dominant digestive peptidase than FmCL and FmCB because the aspartic peptidase is the least transcribed one of them all.

In **vitellaria**, **testes** and **ovary**, the place of transcription of FmCD overlaps with the localization for FmCL and FmCB, it was found in nurse cells, spermatogonia/sustentacular tissue and oogonia/nurse cells, respectively. To our knowledge, such localization of CD was, from all trematodes, reported only for *O. viverrini* (Suttiprapa *et al*., 2009). Cathepsin D was immunohistochemically detected in vitellaria, in primary spermatogonia within testes and ovary of adult *O. viverrini*. It is, therefore, possible, that CD in these two flukes might be involved in the development and maturation of vitelline cells, spermatozoa and oocytes.

In addition, Suttiprapa *et al.* (2009) localized a protein form of CD in *O. viverrini* also in developing eggs within the **uterus**. On the contrary, developing eggs did not show any signal when detecting FmCD mRNA in *F. magna*, but interestingly it was detected in epithelial cells lining the uterus. This could mean that CD is involved in egg development, although possibly by different mechanisms in these two trematodes.

5.3 Expression of the recombinant FmCB in yeast

One of the aims of this thesis was to express a recombinant cathepsin B of *F. magna* and subsequently study its biochemical characteristics, such as catalytical activities, and to potentially produce antibodies and use them for immunolocalization. For this purpose, we employed the PichiaPink™ Expression System (Invitrogen) which is genetically modified yeast *Pichia pastoris*. Despite several attempts and various modifications of the expression procedure, we were not able to detect any FmCB expression.

One possible reason why the protein expression was not successful could lie at the beginning, while modifying the original transcriptomic FmCB sequence. The original sequence was analysed for a signal peptide using online free software SignalP-4.0 server, and no signal sequence was identified. After several months later, the new upgraded SignalP-5.0 server has been released and FmCB sequence was tested again. On the contrary, the upgraded version predicted a signal sequence within FmCB. Thus, the original FmCB signal peptide was not removed while modifying the sequence and it is possible, that the final FmCB sequence inserted into the vector was not rightly processed with the expression yeast cells because it would carry two signal peptides – yeast (α-mating factor) and *F. magna* signal peptides, and therefore the expression was not working.

The process of the expression of FmCB was modified several times, to try if any changes in the procedure could help. For example, two different strains of PichiaPink™ Expression System (Invitrogen) were tested for expression – "protease wild-type" PichiaPink™ Strain 1 and PichiaPink™ Strain 4 which is double knock-out for two yeast proteases prb1, pep4. The latter was employed in case of potential degradation of FmCB by proteases expressed by the "wild-type" strain 1. In addition, several temperatures were tested during the induction process by methanol. Nevertheless, no modifications had resulted in protein expression.

Although some modifications of the expression process have been made, there are several other options which could have solved the problem. For instance, trying different secretion signals (apart from α-mating factor), using unbuffered media during expression, switching from pPink-HC vector to pPink-LC (low-copy).

In other fasciolid trematodes CBs are considered to be mainly expressed by metacercaria and NEJs, thus playing an important role for the early migration in the definitive host body (Smooker *et al*., 2010). Although there has been identified CBs also in adult flukes, only a few have been expressed as a recombinant protein and studied in detail. An exception is represented by Siricoon *et al*. (2015). In this study, *F. gigantica* cathepsin B5 (FgCB5) was expressed in *Pichia pastoris* and its activity against native proteins was determined. FgCB5 displayed endopeptidase and also exopeptidase activity, and was able to cleave connective tissue proteins, bovine serum albumin, immunoglobulin G and Hb at the pH optimum 4.5 – 5.5. In the future, it would be beneficial, to compare FmCB with FgCB5 and investigate their similarities.

6 Conclusion

We investigated the gene expression of three cathepsins L, B and D of digenean fluke *F. magna*. In the beginning, we have analysed the stability of several reference genes using geNorm evaluation, and as a result, FmBACT, FmGAPDH, FmTUBBI3.1 were selected for the following normalization. The expression levels of the three peptidases were compared among three developmental stages – eggs, miracidia and adults. It was revealed, that all cathepsins were the most expressed in adult flukes, especially FmCL and FmCB, thus suggesting their great importance in adults compared to eggs and miracidia. These findings correspond to the localization of mRNA of all studied peptidases, identified by RNA *in situ* hybridization, in the gastrodermal cells of an adult. Such place of transcription supports the hypothesis, that cathepsins are participating in the hydrolysis of host blood proteins (e.g. Hb, albumin) in *F. magna*, similarly to observations made for related fasciolid trematodes. Taking into consideration that miracidia and eggs are life stages which do not feed at all, it is not surprising that studied cathepsins had low expression levels when compared to blood-feeding adults.

The quite surprising result was the detection of cathepsin transcripts in other tissues and organs apart from the gut. Particularly, all cathepsins were identified in nurse cells within vitellaria, which is a completely new observation among flukes and it suggests that these peptidases are involved in the development of vitelline cells. Another place of detected transcription common to all cathepsins was in spermatogonia within testes and also in oogonia within the ovary, which could mean that cathepsins function in the development of spermatocytes and oocytes, respectively. The diversity in localization of cathepsins was observed in the uterus. Cathepsin L was not detected in this organ at all, on the other hand, cathepsin B was localized within the mass of vitelline cells and oocytes, suggesting its role in the formation of eggs in the uterus. In contrast, cathepsin D was not detected in the same cells as FmCB, but it was localized within the epithelial cells lining the uterus. Such localization could mean that FmCD is also involved in the formation of eggs, but by a different mechanism than FmCB.

Despite the unsuccessful expression of recombinant FmCB in the yeast expression system, the results of this thesis significantly enriched the knowledge of *F. magna* peptidases. Mainly, we revealed that studied cathepsins are not only digestive peptidases but that they probably also play a role in *F. magna* reproduction. In the future, it would be interesting to investigate this feature further, and also to extend the research to more developmental stages, such as infective metacercaria and newly excysted juvenile.

7 Bibliography

Abubucker S., Martin J., Taylor C. M., Mitreva M. (2011). HelmCoP: an online resource for helminth functional genomics and drug and vaccine targets prioritization. *PLoS ONE*, *6*, e21832.

Anuracpreeda P., Songkoomkrong S., Sethadavit M., Chotwiwatthanakun C., Tinikul Y., Sobhon P. (2011). *Fasciola gigantica*: production and characterization of a monoclonal antibody against recombinant cathepsin B3. *Experimental Parasitology*, *127*, 340–345.

Apostolo C. (1996). The naturalistic aspects: flora, fauna and the environment. *La Mandria Storia e natura del Parco. Edizioni Eda, Torino*, 67–98.

Arenz C., Schepers U. (2003). RNA interference: from an ancient mechanism to a state of the art therapeutic application? *Naturwissenschaften*, *90*, 345–359.

Arya M., Shergill I. S., Williamson M., Gommersall L., Arya N., Patel H. R. (2005). Basic principles of real-time quantitative PCR. *Expert Review of Molecular Diagnostics*, *5*.

Barrett A. J. (2013). Introduction: the clans and families of cysteine peptidases. In *Handbook of Proteolytic Enzymes*. Academic Press, 1743–1773.

Beckham S. A., Law R. H.-P., Smooker P. M., Quinsey N. S., Caffrey C. R., McKerrow J. H., Pike R. N., Spithill T. W. (2006). Production and processing of a recombinant *Fasciola hepatica* cathepsin B-like enzyme (FhcatB1) reveals potential processing mechanisms in the parasite. *Biological Chemistry*, *387*, 1053–61.

Beckham S. A., Piedrafita D., Phillips C. I., Samarawickrema N., Law R. H. P. P., Smooker P. M., Quinsey N. S., Irving J. A., Greenwood D., Verhelst S. H. L. L., Bogyo M., Turk B., Coetzer T. H., Wijeyewickrema L. C., Spithill T. W., Pike R. N. (2009). A major cathepsin B protease from the liver fluke *Fasciola hepatica* has atypical active site features and a potential role in the digestive tract of newly excysted juvenile parasites. *The International Journal of Biochemistry & Cell Biology*, *41*, 1601–1612.

Biswal D. K., Roychowdhury T., Pandey P., Tandon V. (2018). De novo genome and transcriptome analyses provide insights into the biology of the trematode human parasite *Fasciolopsis buski*. *PLOS ONE*, *13*, e0205570.

Boyunaga H., Schmitz M. G. J., Brouwers J. F. H. M., Van Hellemond J. J., Tielens A. G. M. (2001). *Fasciola hepatica* miracidia are dependent on respiration and endogenous glycogen degradation for their energy generation. *Parasitology*, *122*, 169–173.

Brindley P. J., Kalinna B. H., Wong J. Y. M., Bogitsh B. J., King L. T., Smyth D. J., Verity C. K., Abbenante G., Brinkworth R. I., Fairlie D. P., Smythe M. L., Milburn P. J., Bielefeldt-Ohmann H., Zheng Y., McManus D. P. (2001). Proteolysis of human hemoglobin by schistosome cathepsin D. *Molecular and Biochemical Parasitology*, *112*, 103–112.

Bustin S. A., Benes V., Garson J. A., Hellemans J., Huggett J., Kubista M., Mueller R., Nolan T., Pfaffl M. W., Shipley G. L., Vandesompele J., Wittwer C. T. (2009). The MIQE Guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, *55*, 611–622.

Cabán-Hernández K., Gaudier J. F., Espino A. M. (2012). Characterization and differential expression of a ferritin protein from *Fasciola hepatica*. *Molecular and Biochemical Parasitology*, *182*, 54–61.

Cabán-Hernández K., Espino A. M. (2013). Differential expression and localization of saposin-like protein 2 of *Fasciola hepatica*. *Acta Tropica*, *128*, 591–597.

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., Goupil L., Rebello K. M., Dalton J. P., Smith D. (2018). Cysteine proteases as digestive enzymes in parasitic helminths. *PLoS neglected tropical diseases*, *12*, e0005840.

Cancela M., Ruétalo N., Dell'Oca N., da Silva E., Smircich P., Rinaldi G., Roche L., Carmona C., Alvarez-Valín F., Zaha A., Tort J. F. (2010). Survey of transcripts expressed by the invasive juvenile stage of the liver fluke *Fasciola hepatica*. *BMC Genomics*, *11*, 227.

Cantacessi C., Mulvenna J., Young N. D., Kašný M., Horák P., Aziz A., Hofmann A., Loukas A., Gasser R. B. (2012). A deep exploration of the transcriptome and "excretory/secretory" proteome of adult *Fascioloides magna*. *Molecular & Cellular Proteomics*, *11*, 1340–1353.

Carnevali O., Cionna C., Tosti L., Lubzens E., Maradonna F. (2006). Role of cathepsins in ovarian follicle growth and maturation. *General and Comparative Endocrinology*, *146*, 195–203.

Cereghino J. L., Cregg J. M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiology Reviews*, *24*, 45–66.

Chen W., Wang X., Li X., Lv X., Zhou C., Deng C., Lei H., Men J., Fan Y., Liang C., Yu X. (2011). Molecular characterization of cathepsin B from *Clonorchis sinensis* excretory/secretory products and assessment of its potential for serodiagnosis of clonorchiasis. *Parasites and Vectors*, *27*, 149.

Chen W., Wang X., Lv X., Tian Y., Xu Y., Mao Q., Shang M., Li X., Huang Y., Yu X. (2014). Characterization of the secreted cathepsin B cysteine proteases family of the carcinogenic liver fluke *Clonorchis sinensis*. *Parasitology Research*, *113*, 3409–3418.

Cho W.-L., Tsao S.-M., Hays A. R., Walter R., Chen J.-S., Snigirevskaya E. S., Raikhel A. S. (1999). Mosquito cathepsin B-like protease involved in embryonic degradation of vitellin is produced as a latent extraovarian precursor. *The Journal of Bological Chemistry*, *274*, 13311–13321.

Chung Y.-B., Kong Y., Joo I.-J., Cho S.-Y., Kang S.-Y. (1995). Excystment of *Paragonimus westermani* metacercariae by endogenous cysteine protease. *The Journal of Parasitology*, *81*, 137.

Cogswell A. A., Collins J. J., Newmark P. A., Williams D. L. (2011). Whole mount in situ hybridization methodology for *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, *178*, 46–50.

Coil W. H. (1977). The penetration of *Fascioloides magna* miracidia into the snail host *Fossaria bulimoides*. *Zeitschrift für Parasitenkunde*, *52*, 53–59.

Coil W. H. (1981). Miracidial penetration in *Fascioloides magna* (Trematoda). *Zeitschrift für Parasitenkunde*, *65*, 299–307.

Colborn J. M., Koita O. A., Krogstad D. J., Byrd B. D. (2008). Estimation of copy number using SYBR Green: confounding by AT-rich DNA and by variation in amplicon length. *The American Journal of Tropical Medicine and Hygiene*, *79*, 887–892.

Collins P. R., Stack C. N., O 'neill S. M., Doyle S., Ryan T., Brennan G. P., Mousley A., Stewart M., Maule A. G., Dalton J. P., Donnelly S. (2004). Cathepsin L1, the major protease involved in liver fluke (*Fasciola hepatica*) virulence. *Journal of Biological Chemistry*.

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

Cwiklinski K., Dalton J. P., Dufresne P. J., La Course J., Williams D. J. L., Hodgkinson J., Paterson S. (2015). The *Fasciola hepatica* genome: Gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution. *Genome Biology*, *16*, 1–13.

Cwiklinski K., Dalton J. P. (2018). Advances in *Fasciola hepatica* research using 'omics' technologies. *International Journal for Parasitology*, *48*, 321–331.

Cwiklinski K., Donnelly S., Drysdale O., Jewhurst H., Smith D., De Marco Verissimo C., Pritsch I. C., O'Neill S., Dalton J. P., Robinson M. W. (2019). The cathepsin-like cysteine peptidases of trematodes of the genus *Fasciola*. *Advances in Parasitology*, *104*, 113–164.

Cwiklinski K., Jewhurst H., Mcveigh P., Barbour T., Maule A. G., Tort J., O'neill S. M., Robinson M. W., Donnelly S., Dalton J. P. (2019). Infection by the helminth parasite *Fasciola hepatica* requires rapid regulation of metabolic, virulence, and invasive factors to adjust to its mammalian host. *Molecular & Cellular Proteomics*, *17*, 792–809.

Dalton J. P., McGonigle S., Rolph T. P., Andrews S. J. (1996). Induction of protective immunity in cattle against infection with *Fasciola hepatica* by vaccination with cathepsin L proteinases and with hemoglobin. *Infection and immunity*, *64*, 5066–5074.

Dalton J. P., Neill S. O., Stack C., Collins P., Walshe A., Sekiya M., Doyle S., Mulcahy G., Hoyle D., Khaznadji E., Moiré N., Brennan G., Mousley A., Kreshchenko N., Maule A. G., Donnelly S. M. (2003). *Fasciola hepatica* cathepsin L-like proteases: Biology, function, and potential in the development of first generation liver fluke vaccines. *International Journal for Parasitology*, *33*, 1173–1181.

Dalton J. P., Caffrey C. R., Sajid M., Stack C., Donnelly S., Loukas A., Don T., McKerrow J., Halton D. W., Brindley P. J. (2006). Proteases in trematode biology. In A. G. Maule & N. J. Marks (Eds.), *Parasitic Flatworms: molecular biology, biochemistry, immunology and physiology*. CAB International, 348–368.

Dalton J. P., Robinson M. W., Mulcahy G., O'Neill S. M., Donnelly S. (2013). Immunomodulatory molecules of *Fasciola hepatica*: Candidates for both vaccine and immunotherapeutic development. *Veterinary Parasitology*, *195*, 272–285.

Delcroix M., Sajid M., Caffrey C. R., Lim K.-C. C., Dvořák J., Hsieh I., Bahgat M., Dissous C., Mckerrow J. H., Dvorák J., Hsieh I., Bahgat M., Dissous C., Mckerrow J. H., Dvořák J., Hsieh I., Bahgat M., Dissous C., Mckerrow J. H. (2006). A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. *Journal of Biological Chemistry*, *281*, 39316–39329.

Di Maggio L. S., Tirloni L., Pinto A. F. M., Diedrich J. K., Yates III J. R., Benavides U., Carmona C., da Silva Vaz I., Berasain P. (2016). Across intra-mammalian stages of the liver fluke *Fasciola hepatica*: a proteomic study. *Scientific Reports*, *6*, 32796.

Dillon G. P., Illes J. C., Isaacs H. V., Wilson R. A. (2007). Patterns of gene expression in schistosomes: localization by whole mount in situ hybridization. *Parasitology*, *134*, 1589–97.

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.

Dvořák J., Mashiyama S. T., Sajid M., Braschi S., Delcroix M., Schneider E. L., McKerrow W. H., Bahgat M., Hansell E., Babbitt P. C., Craik C. S., McKerrow J. H., Caffrey C. R. (2009). SmCL3, a gastrodermal cysteine protease of the human blood fluke *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases*, *3*, e449.

Erhardová-Kotrlá B. (1971). *The occurrence of Fascioloides magna (Bassi, 1875) in Czechoslovakia*. Prague: Academia.

Faltýnková A., Horáčková E., Hirtová L., Novobilský A., Modrý D., Scholz T. (2006). Is *Radix peregra* a new intermediate host of *Fascioloides magna* (Trematoda) in Europe? Field and experimental evidence. *Acta Parasitologica*, *51*, 87–90.

Foreyt W. J., Todd A. C. (1976). Development of the large American liver fluke, *Fascioloides magna*, in white-tailed deer, cattle, and sheep. *The Journal of Parasitology*, *62*, 26–32.

Foreyt W. J., Samuel W. M., Todd A. C. (1977). *Fascioloides magna* in white-tailed deer (*Odocoileus virginianus*): Observations on the pairing tendency. *The Journal of Parasitology*, *63*, 1050–1052.

Geldhof P., Visser A., Clark D., Saunders G., Britton C., Gilleard J., Berriman M., Knox D. (2007). RNA interference in parasitic helminths: Current situation, potential pitfalls and future prospects HIV infection of *Schistosoma*. *Parasitology*, *134*, 609–619.

Griswold M. D. (1995). Interactions between germ cells and Sertoli cells in the testis. *Biology of Reproduction*, *52*, 211–216.

Hanna R., Moffett D., Brennan G. P. P., Fairweather I. (2012). *Fasciola hepatica*: A light and electron microscope study of sustentacular tissue and heterophagy in the testis. *Veterinary Parasitology*, *187*, 168–182.

Hanna R. (2015). *Fasciola hepatica*: Histology of the reproductive organs and differential effects of triclabendazole on drug-sensitive and drug-resistant fluke isolates and on flukes from selected field cases. *Pathogens*, *4*, 431–456.

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. *Genome Biology*, *8*, R19.

Hernández-González A., Valero M. L., Pino M. S. del, Oleaga A., Siles-Lucas M. (2010). Proteomic analysis of in vitro newly excysted juveniles from *Fasciola hepatica*. *Molecular and Biochemical Parasitology*, *172*, 121–128.

Huang S.-Y., Yue D.-M., Hou J.-L., Zhang X.-X., Zhang F., Wang C.-R., Zhu X.-Q. (2019). Proteomic analysis of *Fasciola gigantica* excretory and secretory products (FgESPs) interacting with buffalo serum of different infection periods by shotgun LC-MS/MS. *Parasitology Research*, *118*, 453–460.

Irwin S. W. B., Threadgold L. T. (1970). Electron-microscope studies on *Fasciola hepatica*: VIII. The development of the vitelline cells. *Experimental Parasitology*, *28*, 399–411.

Jin L., Lloyd R. V. (1997). In situ hybridization: methods and applications. *Journal of Clinical Laboratory Analysis*, *11*, 2–9.

Jones A., Bray R. A., Gibson D. I. (Eds.). (2005). *Keys to the Trematoda: Volume 2*. Wallingford: CABI Publishing and The Natural History Museum.

Jones M. K., Randall L. M., McManus D. P., Engwerda C. R. (2004). Laser microdissection microscopy in parasitology: microscopes meet thermocyclers. *Trends in Parasitology*, *20*, 502–506.

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). Peptidases of trematodes. In *Advances in Parasitology* (Vol. 69), 205–297.

Kašný M., Beran L., Siegelová V., Siegel T., Leontovyč R., Berankova K., Pankrac J., Košťáková M., Horák P. (2012). Geographical distribution of the giant liver fluke (*Fascioloides magna*) in the Czech Republic and potential risk of its further spread. *Veterinarni Medicina*, *57*, 101–109.

Koressaar T., Remm M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, *23*, 1289–1291.

Košťáková M. (2013). *Morfologie vývojových stádií motolice Fascioloides magna a histopatologické změny u vybraných druhů definitivních hostitelů*. Charles University.

Lehmann R., Tautz D. (1994). Chapter 30 In Situ Hybridization to RNA. *Methods in Cell Biology*, *44*, 575–598.

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.

Lotfy W. M., Brant S. V., DeJong R. J., Le T. H., Demiaszkiewicz A., Rajapakse R. P. V. J. V. J., Perera V. B. V. P. V. P., Laursen J. R., Loker E. S., Giay C., Hanoi V. ;, Stefan´ski W., Stefan´ski S., Thanh H. Le, Demiaszkiewicz A., Rajapakse R. P. V. J. V. J., Perera V. B. V. P. V. P., Laursen J. R., Loker E. S. (2008). Evolutionary origins, diversification, and biogeography of liver flukes (Digenea, Fasciolidae). *American Journal of Tropical Medicine and Hygiene*, *79*, 248–255.

Lowther J., Robinson M. W., Donnelly S. M., Xu W., Stack C. M., Matthews J. M., Dalton J. P. (2009). The importance of pH in regulating the function of the *Fasciola hepatica* cathepsin L1 cysteine protease. *PLoS Neglected Tropical Diseases*, *3*, e369.

Mangiola S., Young N. D., Korhonen P., Mondal A., Scheerlinck J.-P., Sternberg P. W., Cantacessi C., Hall R. S., Jex A. R., Gasser R. B. (2013). Getting the most out of parasitic helminth transcriptomes using HelmDB: Implications for biology and biotechnology. *Biotechnology Advances*, *31*, 1109–1119.

Martin J., Tyagi R., Rosa B. A., Mitreva M. (2018). A multi-omics database for parasitic nematodes and trematodes. *Methods in Molecular Biology*, *1757*, 371–397.

McClanahan S. L., Stromberg B. E., Hayden D. W., Averbeck G. A., Wilson J. H. (2005). Natural infection of a horse with *Fascioloides magna*. *Journal of Veterinary Diagnostic Investigation*, *17*, 382–385.

McGonigle L., Mousley A., Marks N. J., Brennan G. P., Dalton J. P., Spithill T. W., Day T. A., Maule A. G. (2008). The silencing of cysteine proteases in *Fasciola hepatica* newly excysted juveniles using RNA interference reduces gut penetration. *International Journal for Parasitology*, *38*, 149–155.

McNulty S. N., Tort J. F., Rinaldi G., Fischer K., Rosa B. A., Smircich P., Fontenla S., Choi Y.-J., Tyagi R., Hallsworth-Pepin K., Mann V. H., Kammili L., Latham P. S., Dell'Oca N., Dominguez F., Carmona C., Fischer P. U., Brindley P. J., Mitreva M. (2017). Genomes of *Fasciola hepatica* from the Americas reveal colonization with Neorickettsia endobacteria related to the agents of Potomac horse and human Sennetsu fevers. *PLOS Genetics*, *13*, e1006537.

Mebius M. M., Op Heij J. M. J., Tielens A. G. M., de Groot P. G., Urbanus R. T., van Hellemond J. J. (2018). Fibrinogen and fibrin are novel substrates for *Fasciola hepatica* cathepsin L peptidases. *Molecular and Biochemical Parasitology*, *221*, 10–13.

Medina M., León P., Vallejo C. G. (1988). Drosophila cathepsin B-like proteinase: a suggested role in yolk degradation. *Archives of biochemistry and biophysics*, *263*, 355–63.

Meemon K., Grams R., Vichasri-Grams S., Hofmann A., Korge G., Viyanant V., Upatham E. S., Habe S., Sobhon P. (2004). Molecular cloning and analysis of stage and tissue-specific expression of cathepsin B encoding genes from *Fasciola gigantica*. *Molecular and Biochemical Parasitology*, *136*, 1–10.

Meemon K., Khawsuk W., Sriburee S., Meepool A., Sethadavit M., Sansri V., Wanichanon C., Sobhon P. (2010). *Fasciola gigantica*: Histology of the digestive tract and the expression of cathepsin L. *Experimental Parasitology*, *125*, 371–379.

Meemon K., Sobhon P. (2015). Juvenile-specific cathepsin proteases in *Fasciola* spp.: their characteristics and vaccine efficacies. *Parasitology Research*, *114*, 2807–2813.

Morales M. E., Rinaldi G., Gobert G. N., Kines K. J., Tort J. F., Brindley P. J. (2008). RNA interference of *Schistosoma mansoni* cathepsin D, the apical enzyme of the hemoglobin proteolysis cascade. *Molecular and Biochemical Parasitology*, *157*, 160–168.

Naem S., Budke C. M., Craig T. M. (2012). Morphological characterization of adult *Fascioloides magna* (Trematoda: Fasciolidae): First SEM report. *Parasitology Research*, *110*, 971–978.

Norbury L. J., Beckham S., Pike R. N., Grams R., Spithill T. W., Fecondo J. V., Smooker P. M. (2011). Adult and juvenile *Fasciola* cathepsin L proteases: Different enzymes for different roles. *Biochimie*, *93*, 604–611.

Oksjoki S., Söderström M., Vuorio E., Anttila L. (2001). Differential expression patterns of cathepsins B, H, K, L and S in the mouse ovary. *Molecular Human Reproduction*, *7*, 27–34.

Protasio A. V, Tsai I. J., Babbage A., Nichol S., Hunt M. (2012). A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Negl Trop Dis*, *6*, 1455.

Pybus M. J. (1990). Liver Flukes. In *Parasitic diseases of wild mammals, WM Samuel, MJ Pybus, and AA Kocan (eds.). Iowa State Press, Ames, Iowa*, 121–149.

Ranjit N., Jones M. K., Stenzel D. J., Gasser R. B., Loukas A. (2006). A survey of the intestinal transcriptomes of the hookworms, *Necator americanus* and *Ancylostoma caninum*, using tissues isolated by laser microdissection microscopy. *International Journal for Parasitology*, *36*, 701–710.

Rawlings N. D., Barrett A. J. (1993). Evolutionary families of peptidases. *The Biochemical journal*, *290 (Pt 1)*, 205–18.

Rawlings N. D., Salvesen G. (2012). *Handbook of proteolytic enzymes*. Academic.

Robinson M. W., Brindley P. J., McKerrow J. H., Dalton J. P. (2013). Trematode cysteine endopeptidases. *Handbook of Proteolytic Enzymes*, 1941–1949.

Robinson M. W., Corvo I., Jones P. M., George A. M., Padula M. P., To J., Cancela M., Rinaldi G., Tort J. F., Roche L., Dalton J. P. (2011). Collagenolytic activities of the major secreted cathepsin L peptidases involved in the virulence of the helminth pathogen, *Fasciola hepatica*. *PLoS Neglected Tropical Diseases*, *5*.

Robinson M. W., Dalton J. P., Donnelly S. (2008). Helminth pathogen cathepsin proteases: it's a family affair. *Trends in Biochemical Sciences*, *33*, 601–608.

Robinson M. W., Menon R., Donnelly S. M., Dalton J. P., Ranganathan S. (2009). An integrated transcriptomics and proteomics analysis of the secretome of the helminth pathogen *Fasciola hepatica*: Proteins associated with invasion and infection of the mammalian host. *Molecular & Cellular Proteomics*, *8*, 1891–1907.

Robinson M. W., Tort J. F., Lowther J., Donnelly S. M., Wong E., Xu W., Stack C. M., Padula M., Herbert B., Dalton J. P. (2008). Proteomics and phylogenetic analysis of the cathepsin L protease family of the helminth pathogen *Fasciola hepatica*: expansion of a repertoire of virulence-associated factors. *Molecular & Cellular Proteomics*, *7*, 1111–1123.

Robker R. L., Russell D. L., Espey L. L., Lydon J. P., O'Malley B. W., Richards J. S. (2000). Progesterone-regulated genes in the ovulation process: ADAMTS-1 and cathepsin L proteases. *Proceedings of the National Academy of Sciences* (Vol. 97).

Sajid M., McKerrow J. H., Hansell E., Mathieu M. A., Lucas K. D., Hsieh I., Greenbaum D., Bogyo M., Salter J. P., Lim K. C., Franklin C., Kim J.-H., Caffrey C. R. (2003). Functional expression and characterization of *Schistosoma mansoni* cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. *Molecular and biochemical parasitology*, *131*, 65–75.

San Segundo-Val I., Sanz-Lozano C. S. (2016). Introduction to the gene expression analysis. In I. G. M. (Ed.), *Methods in Molecular Biology*. New York: Humana Press, 29–43.

Sansri V., Changklungmoa N., Chaichanasak P., Sobhon P., Meemon K. (2013). Molecular cloning, characterization and functional analysis of a novel juvenile-specific cathepsin L of *Fasciola gigantica*. *Acta Tropica*, *128*, 76–84.

Šašková R. (2015). *Expresní profil katepsinu L u jednotlivých vývojových stádií Fascioloides magna*. Charles University in Prague.

Schepers U. (2005). *RNA interference in practice:* Weinham: WILEY-VCH Verlag GmbH & Co. KGaA.

Sethadavit M., Meemon K., Jardim A., Spithill T. W., Sobhon P. (2009). Identification, expression and immunolocalization of cathepsin B3, a stage-specific antigen expressed by juvenile *Fasciola gigantica*. *Acta Tropica*, *112*, 164–173.

Siricoon S., Vichasri Grams S., Lertwongvisarn K., Abdullohfakeeyah M., Smooker P. M., Grams R. (2015). *Fasciola gigantica* cathepsin B5 is an acidic endo- and exopeptidase of the immature and mature parasite. *Biochimie*, *119*, 6–15.

Skelly P. J., Da'dara A., Harn D. A. (2003). Suppression of cathepsin B expression in *Schistosoma mansoni* by RNA interference. *International Journal for Parasitology*, *33*, 363–369.

Smooker P. M., Jayaraj R., Pike R. N., Spithill T. W. (2010). Cathepsin B proteases of flukes: The key to facilitating parasite control? *Trends in Parasitology*, *26*, 506–514.

Smyth J. D., Halton D. W. (1983). *The physiology of trematodes*. CUP Archive.

Sullivan C. V., Aquagyn C., Matsubara T., Nagae M., Ohkubo N., Andoh T., Sawaguchi S., Hiramatsu N., Sullivan C. V, Hara A. (2003). Multiple vitellogenins and their unique role in marine teleosts Multiple vitellogenins and their unique roles in marine teleosts. *Article in Fish Physiology and Biochemistry*, *28*, 295–299.

Suttiprapa S., Mulvenna J., Huong N. T., Pearson M. S., Brindley P. J., Laha T., Wongkham S., Kaewkes S., Sripa B., Loukas A. (2009). Ov-APR-1, an aspartic protease from the carcinogenic liver fluke, *Opisthorchis viverrini*: functional expression, immunolocalization and subsite specificity. *The international journal of biochemistry & cell biology*, *41*, 1148–56.

Swales W. E. (1935). The life cycle of *Fascioloides magna* (Bassi, 1875), the large liver fluke of ruminants in Canada with observations on the bionomics of the larval stages and the intermediate hosts, pathology of fascioloidiasis magna, and control measures. *Canadian Journal of Research*, *12*, 177–215.

Tkach V. V., Kudlai O., Kostadinova A. (2016). Molecular phylogeny and systematics of the Echinostomatoidea Looss, 1899 (Platyhelminthes: Digenea). *International Journal for Parasitology*, *46*, 171–185.

Trailović S. M., Marinković D., Trailović J. N., Milovanović M., Marjanović D. S., Aničić M. R. (2015). Pharmacological and morphological characteristics of the muscular system of the giant liver fluke (*Fascioloides magna* - Bassi 1875). *Experimental Parasitology*, *159*, 136–142.

Tripathi T., Ghosh A., Todur V. N., Kalita P., Rajendran V., Kalita J., Shukla R., Chetri P. B., Shukla H., Sonkar A., Lyngdoh D. L., Singh R., Chikara S. K., Tripathi T. (2018). Draft genome of the liver fluke *Fasciola gigantica*. *bioRxiv*, 451476.

Ullrich K. (1930). Über das Vorkommen von seltenen oder wenig bekannten Parasiten der Säugetiere und Vögel in Böhmen und Mähren. *Prager Archiv für Tiermedizin*, *10*, 19–43.

Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B. C., Remm M., Rozen S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Research*, *40*, e115–e115.

Vandesompele J., De Preter K., Pattyn I., Poppe B., Van Roy N., De Paepe A., Speleman R. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, *3*, 34.1-34.12.

Verity C. K., McManus D. P., Brindley P. J. (1999). Developmental expression of cathepsin D aspartic protease in *Schistosoma japonicum*. *International Journal for Parasitology*, *29*, 1819–1824.

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*, e0147247.

Wang Z., Gerstein M., Snyder M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*, 57–63.

Wilson R. A., Wright J. M., de Castro-Borges W., Parker-Manuel S. J., Dowle A. A., Ashton P. D., Young N. D., Gasser R. B., Spithill T. W. (2011). Exploring the *Fasciola hepatica* tegument proteome. *International Journal for Parasitology*, *41*, 1347–1359.

Wright W. W., Smith L., Kerr C., Charron M. (2003). Mice that express enzymatically inactive cathepsin L exhibit abnormal spermatogenesis. *Biology of Reproduction*, *68*, 680–687.

Yoo W. G., Kim T. I., Li S., Kwon O. S., Cho P. Y., Kim T.-S., Kim K., Hong S.-J. (2009). Reference genes for quantitative analysis on *Clonorchis sinensis* gene expression by real-time PCR. *Parasitology Research*, *104*, 321–328.

Yook K., Harris T. W., Bieri T., Cabunoc A., Chan J., Chen W. J., Davis P., De La Cruz N., Duong A., Fang R., Ganesan U., Grove C., Howe K., Kadam S., Kishore R., Lee R., Li Y., Muller H.-M. H.-M., Nakamura C., Nash B., Ozersky P., Paulini M., Raciti D., Rangarajan A., Schindelman G., Shi X., Schwarz E. M., Ann Tuli M., Van Auken K., Wang D., Wang X., Williams G., Hodgkin J., Berriman M., Durbin R., Kersey P., Spieth J., Stein L., Sternberg P. W., Tuli M. A., Van Auken K., Wang D., Wang X., Williams G., Hodgkin J., Berriman M., Durbin R., Kersey P., Spieth J., Stein L., Sternberg P. W. (2012). WormBase 2012: more genomes, more data, new website. *Nucleic Acids Research*, *40*, D735–D741.

Young N. D., Hall R. S., Jex A. R., Cantacessi C., Gasser R. B. (2010). Elucidating the transcriptome of *Fasciola hepatica* - A key to fundamental and biotechnological discoveries for a neglected parasite. *Biotechnology Advances*, *28*, 222–231.

Young N. D., Jex A. R., Cantacessi C., Hall R. S., Campbell B. E., Spithill T. W., Tangkawattana S., Tangkawattana P., Laha T., Gasser R. B. (2011). A portrait of the transcriptome of the neglected trematode, *Fasciola gigantica*-biological and biotechnological implications. *PLoS Neglected Tropical Diseases*, *5*, e1004.

Young N. D., Nagarajan N., Lin S. J., Korhonen P. K., Jex A. R., Hall R. S., Safavi-Hemami H., Kaewkong W., Bertrand D., Gao S., Seet Q., Wongkham S., Teh B. T., Wongkham C., Intapan P. M., Maleewong W., Yang X., Hu M., Wang Z., Hofmann A., Sternberg P. W., Tan P., Wang J., Gasser R. B. (2014). The *Opisthorchis viverrini* genome provides insights into life in the bile duct. *Nature Communications*, *5*, 4378.

Zawistowska-Deniziak A., Wasyl K., Norbury L. J., Wesolowska A., Bień J., Grodzik M., Wiśniewski M., Baska P., Wedrychowicz H. (2013). Characterization and differential expression of cathepsin L3 allelesfrom *Fasciola hepatica*. *Molecular and Biochemical Parasitology*, *190*, 27–37.

Zerlotini A., Heiges M., Wang H., V Moraes R. L., Dominitini A. J., Ruiz J. C., Kissinger J. C., Oliveira G. (2008). SchistoDB: a *Schistosoma mansoni* genome resource. *Nucleic Acids Research*, *37*, 579–582.

Zhou G., Stevenson M. M., Geary T. G., Xia J., Yazdanbakhsh M. (2016). Comprehensive Transcriptome meta-analysis to characterize host immune responses in helminth infections. *PLoS neglected tropical diseases*, *10*, e0004624.

Zipper H., Brunner H., Bernhagen J. J., Vitzthum F. (2004). Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. *Nucleic Acids Research*, *32*, e103–e103.

8 Supplementary data

Cathepsin sequences

>EU877764.1_Fascioloides magna_cathepsinL1(CL1)mRNA_complete_cds

ATGCGATTGTTCATACTCGCTGTCCTCATCGTCAGTGTGTTCGCCTCGAATGATGAT
TTGTGGCATCAATGGAAGCGGAAGTACAATAAAGAATACAACGGAGCTGACGATG
AGCACAGACGAAATATTTGGGAACAGAATGTGAAACATATCCAAGAACACAATTT
ACGTCACGATCTCGGCCTAGTCACGTACACGTTGGGATTGAACCAATTCACTGATA
TGACATTCGAGGAATTCAAGGCCAAATATCTGTTCGAGATATCACCCAAATCTGAA
CTACTCTCACATTCTGGCATTTCGTACCAGGCAAAAGGGAATGACGTACCGGAGAG
TATAGACTGGCGTGATTATGGATATGTGACGGAGGTTAAGGATCAGGGACAATGTG
GTTCTTGTTGGGCTTTTTCATCCACCGGTGCTATGGAGGGACAATACATAAAAAAGT
TCCGAACCACGGTGTCATTTTCTGAGCAACAGTTAGTCGACTGTACCAGGAACTAT
GGCAACAGCGGATGTAATGGTGGATGGATGGAGCGAGCCTTTGAATATTTGAGACG
AAACGGATTGGAAACCGAATCCTCTTACCCGTACAGGGCTGTGGATGATCATTGCC
GATACGAGAGCCAGTTAGGAGTTGCCAAAGTGACTGGCTACTACACGGAGCATTCT
GGTAATGAGGTGTCGTTGATGAACATGGTCGGTGGAGAAGGACCGGTTGCAGTTGC
TGTGGATGTGCAAAGTGATTTCTCGATGTACAAGAGTGGTATTTATCAGAGCGAAA
CGTGTTCAACGTATTACGTGAATCATGCGGTGTTGGCTGTCGGTTATGGAACAGAA
TCTGGCACTGATTATTGGATTTTGAAAAACAGTTGGGGATCGTGGTGGGGTGATCA
AGGTTACATTCGGTTTGCTAGGAACCGAAATAATATGTGTGGAATTGCTTCGTATGC
CAGTGTCCCGATGGTCGAACGTTTTCCGTGAT

>FmCB_FM_transcriptom

ATGCACAGAGTTGGTGCTCGTCAAGATGAACTGGCCAGTTGTGTTTGCCGTCATCG
CCGTTGTCCAGGCGATTAAACCGAATTACAAATCGCAATTCGAGGCGTTCTCTGAT
GAACTGATTCGATATGTGAATGAAGAATCCGGTGCCACTTGGAAAGCAGCACGGTC
CGCCCAATTCAACAGTATTGAACAATTCAAAGTGCATTTGGGAGCCCTAGCTGAAA
CACCGGAACAACGAATTTCCCGGCGACCAACCATAAAACATCACGTATCGGCAAAT
GATCTACCAGAGTCTTTTGATGCTCGAAATCAATGGCCACATTGTTCGTCGATTCCG
GAGATTCGCGACCAGGCCAGTTGTGGTTCCTGTTGGGCTACGGCTGCAGTCAGCGC
AATGTGTGACCGCGTGTGTATTCATTCTAACGGTAAAATTACACCACGTTTGAGCAC
AATGGATCCGCTCACTTGTTGCACAAATTGTGGAAACGGATGCAAAGGTGGCTATC
CACCGAAGGCTTGGGATTATTGGAAGAGTGAAGGCGTTGTGACCGGTGGTACCTGG
GAAAATCGAACTGGTTGCCAGCCTTTCATATTTGCCAAGTGCGATCACGTTGGTAA
CTCTCGAAAATATCCTCCTTGTCCTCACTCCATCTATCCGACACCTCCTTGTGCCAA
AACTTGCCAAACTGGATATAACAAAACGCTTGAGAAAGATAAGTTTTACGGAAATT
CATCTTACAATGTGGATGCACATGAGTCTAATATCATGTACGAGATCATGAAAAAC
GGTCCAGTCGAAGCGAGTTTCGTCGTGTTCACGGACTTCAGCGTGTACAGAAGTGG
GATTTACCATCACGTCTACGGTGGTCCCGTAGGCAGACATGCAATTCGCGTTATCG
GTTGGGGTGTTGAGGATGGTGTGCCTTATTGGTTGATGGCAAATTCGTGGAATGAG
GAATGGGGCGAGAATGGATACTTCCGAATTAAAAGGGGAATGAACGAATGTGGCA
TCGAAACCGAAGTTGTCGCCGGAATGCCACGTCTCTGAGCAACATCGGAAAGACTG
ACAAACCTCTCTGTCTAAATAACTGTTTGGGATTTGCAATACATTTATTGCTAACTT
CAAAAAAAAAAAAAAAAAAAAAAAAA

>FmCD_FM_transcriptom

ATGAAGCAGCACATCTTGATTTGTTTATTGGTTCCTGCTGTGCTTTGTGATGTTTTAA
GGATTAAATTACGTCCATTCAAGACACTCCGTCAGGAGCTCAGCGAGTATGGCTCA
CTCGATTCGAATTCATCACCGCGGCTTTTTGAAAAATATGCAGGAAGCAACGGCTC
TATCCCTGAGCAGCTAAATAACTACTTAGACGCTCAATATTATGGTGAAATTGGTAT
TGGAACGCCACCACAAACTTTCAAAGTCATTTTCGACACTGGGTCGTCAAACTTGT
GGGTCCCGTCAAAACGCTGCAGCTATCTTAGCTGGGCTTGTTGGCTGCACAACAAA
TATAATTACGCCGCTTCCTCAACTTATCAACCTAATGGCACGGCTTTCAGTATCCGG
TACGGCAGTGGCAGTGTATCGGGTTTTATAAGCATCGATTCGTTTGAGGTTGGTGG
AGTTGAGGTGAAAAGCCAACCCTTTGGGGAAGCTATCAAGGAACCTGGTGTCGTTT
TTGTGTTTGCTAAATTCGACGGTATCCTTGGAATGGGATTTAGAAGCATATCTGTTG
ATGGTGTAATAACTGTTTTCGAGAACATGGTTGCCCAGGGCCTAGTACCTGAACCT
GTCTTCTCTTTTTACCTCAACAGAAATGCATCCGATCCTATGGGTGGCGAGCTCCTT
CTCGGAGGGGTCGATCCGGATTTCTATACCGGTGACATTACCTACGTGCCCGTCACT
CATGAAGCATATTGGCAATTCAAAGTCGACAAAATCGAGTTTCCCGGTGTTTCAAT
TTGTGCTGATGGTTGTCAAGCTATTGCTGATACAGGCACTTCACTCATTGCCGGCCC
AAAGAAGGAAGTTGATGCACTCAATGAACAAATTGGGGGTACTTGGATGCCTGGA
GGTATCTACGTCTTGAATTGTGACAAGGTTGATAGTCTTCCTGCTATTGCCTTTGTC
GTGGCCGGAAGAAAAATGGTGCTTGAGGCTAAAGACTATGTAATGAAGTTAAGTA
GCATGGGACGAACAGTGTGTATAACGAGCTTTATCGGTATGGACATTCCTGTTGGG
CCATTGTGGATCTTGGGTGATGTGTTCATCGGCAACTACTACAGTGTGTTTGATATG
GGACAAAAGCGGATCGGCTTTGCCACGGCGAAGCATCCCTCGGTTTCAAAACCACC
ACTTTCGATACCTATGATGCGACTTCAACCCGCCTTCCGTCGCCAAGAGGAACCGC
GCTCCTCTCCACCGAGAAACTTGCTCAACTTCTGGAATTTGTTGATTTGA

Reference genes sequences

>UBC_FM_transcriptom

ATATTTGCAGGGAAGCAGTTGGAAGATGGTCGTACGTTGTCAGATTACAATATCCA
GAAGGAGTCGACATTGCATCTTGTTTTGCGACTTCGTGGTGGTATGCAGATATTTGT
GAAGACGTTGACTGGTAAGACGATAACATTGGAAGTGGAGGCTAGTGATACGATT
GAGAATGTGAAGGCGAAGATTCAGGACAAGGAGGGCATTCCACCAGACCAGCAAC
GGTTGATATTTGCAGGGAAGCAGTTGGAAGATGGTCGTACGTTATCAGATTACAAT
ATCCAGAAGGAGTCAACATTGCATCTTGTACTTCGTCTTCGTGGTGGCAATTAATTA
TGAGGCATTCCTCAATGAAAATTTTTGTTTTGATTTATTTTCCAGAATAAAGGTAGT
TTAGTCTCTTGCGTTAAAAAAAAAAAAAAAAAAAAAAA

>TBC_FM_transcriptom

GAACTGATCTGGCGCCGAGTTACATTGGTAACGCTACGGACATATGAACCCCCTTT
CGCAGCAGCCCAGTGTGACAAGCATAGGCAGCCTTCTGGGTAACCTGCATGGACAG
AGCGAGCATCAACCACTTACCCAAGGTACACCTGCTCCCTTACCGTCTAATGCAAT
CAACCTCGGACTTACGGATCCTCACAACCCAGGAAGTGTACCCAGTCGGACCCCCT
ATACCCCAAATAGTTTGACCACGCCCCTTCATGATCGGGATGGAGCCATACCTAAA
CCGCAATTACAGAATATTGTTTGCACAGTGTACCTCGGATGTAAATTAGATCTTAAG
CAAATCGCCTTAAATGCGCGTAACGCCGAGTACAATCCGAAACGATTTGCCGCGGT
GATCATGCGAATTCGCGAACCGCGAACCACGGCTCTCATATTTTCCTCCGGTAAAA
TGGTGTGTACCGGAGCTAAGAGTGAGGAGCAAGCTCGTTTGGCCGCCCGGAAATAC
GCTCGTATCATTCAAAAACTGGGTTTTGAAGCTCATTTCAAAGAATTCAAAATCCA
GAACATGGTAGGGTCGTGTGATGTCCGGTTTCACATTCGTCTGGAAGGTCTTGTCTT
GTCTCAGAGTCATTTTGCCACCTACGAACCAGAACTCTTCCCCGGTTTGATTTACCG
CATGACAAAACCAAAGATCGTCCTGTTAATCTTTGTCTCGGGGAAAATTGTACTGA
CCGGTGCCAAGGTCCGCGAGGAGATTTATGAAGCCTTCCGAAACATTTACCCGATT
TTGAAGAGTTTCAAAAAACCTGAAAAAGATCTCCGAACCCTATCTAACAACTATCT
AGTTTCCTCTAGCTGAATGAAATTCCCTAATGTGCTACCTGCTCCCTCGTGATGTCA
ACTGGTTTTATGCCACCATGTTTCAGACACGGACTTGCTTGTTCCCATTTGTTACTTT
TCTCTATCCCACCACTAAGAAATTCTAATAAGGACTATTTACATCTTTCCAAATCTA
AATATCACTTGTAAATAATTTCTGACTCTTCCCAATACGCATCTGTCCATTACCGCT
CCCGGCTCAAAAGCACCTTGAAGAACCACATCAAATGGCAAATTCAACTCCATTGA
CGCATTAATGCACACCACGCACTACTGGTCTTTTGTATCCATCCATTCTGCAACGGA
TGATCACTTTCTTTTTTCTACCGTG

>YWHAZ_FM_transcriptom

TTAGCTTCTAAGGCGCGAATTAAGTGCTTAAATTGCCCTGCGACCGTCAGTTTCGCG
GCGTTTCACCATGTCTCCGTGCTGGCTAACGCAGTGTGACTTTAGCTCTGAGCAGCT
CGTCTTCTTGGCTAAGTGTGCCGAGAGGGCAGAGCGTTATGAGGATATGGCTGCAG
CTATGAAGATGCATACCGAGAAGATTGGGAAAAAGGAACTATCAAACGAGGAGCG
GAATCTCCTATCCGTCGCCTACAAGAATGTTGTGGGCGCCAGACGGTCCGCTTGGC
GTATCACTTGTGGTTCCATTAAGCCCGATTTGAAGGAGGTGCAGCTGAAAGCCACT
AAGGATTATATCAAGAGGATGGTTGTTGAACTGAACGATATTTGCAAGGAAGTCTT
GGACATTTTGCAAAATTATTTGATCCCGAATGCTGTTTCCGACGAGAGCAGAGTGTT
CTACATGAAAATGCAGGGTGATTATTATCGGTACTGCGCTGAAGTTGCTGACACCG
AGAATCGCCAGGCTATTGTCGACAGCTCCTTCGATGCTTACAAGAAGGCCATGGAA
ATCGCAAGTGGTCTGCATCCTACTCACCCTATCCGTTTGGGATTGGCCCTGAACTTT
TCGGTCTTTTACTATGAGATTAAAGACCAGCCCAAAGAGGCTTGCGATTTGGCTCA
GAAGTCCTTTGATGAAGCCCTAGCGGAACTCGATACTTTGGGGGAGGATTCATACA
AAGATAGCACCCTAATAATGCAGCTGCTCAGAGATAACCTAACGTTGTGGACAACT
GATCAACCAGAGAACGAGGCTGACGCTGGTGACAAGTAATATCAACTTATGTACTT
ACTTCGCTTTGTCGATATACTGTCGAACATCCAAAAAGGAGACCAAAAAGGAGAGC
AAAAAGGAGACC

>TUBBI2_FM_transcriptom

ATGTGGAAACCAAATTGGCGCGAAGGTAAGTTGGCCGTCGCCTGTCTGGTCGCTGA
ATTTTGGTTCTCCGTAATCCCAGTCAGATACCTATTGTAATTTTCCTTCAGTGCACCC
TTTATTTTGTGAACATCTGTGTATGATATAGCGTAACATTATCCGGACTTTTTATCCT
CGCAGTTCTGGGAGGTAATTTCGGATGAACATGGCATAGATCCAACTGGAACATAT
CACGGAGACTCCGATTTGCAGTTGGAACGCATCAATGTTTACTACAACGAAGCCAC
TGGTGGGAAGTACGTGCCCCGCGCCATCCTGGTCGACCTTGAACCCGGAACAATGG
ATAGCGTACGCGCCGGACCTTTCGGCCAACTCTTCCGCCCCGATAACTTTGTGTTCG
GCCAAAGTGGTGCTGGAAACAATTGGGCCAAGGGACATTATACTGAAGGGGCTGA
GCTCGTTGACTCCGTTCTGGACGTGGTACGAAAAGAGGCTGAGTCCTGTGACTGTC
TGCAAGGCTTCCAATTAACTCACTCACTTGGTGGTGGTACTGGTTCGGGTATGGGA
ACTCTACTTATCTCGAAAATTCGTGAGGAGTATCCGGATCGCATCATGAACACGTTC
TCAGTGGTTCCCTCACCAAAAGTATCCGATACTGTAGTGGAACCCTATAACGCAAC
ACTCTCTGTTCACCAGCTGGTCGAGAATACTGATGAGACGTACTGCATTGACAATG
AGGCGCTCTATGACATTTGCTTCCGTACTCTGAAGCTGACGAACCCAACCTACGGT
GATCTAAACCATCTGGTCAGTGCAACTATGTCTGGTGTGACCACCTGTCTGCGTTTC
CCTGGTCAGCTAAATGCAGATCTTCGGAAACTGGCGGTCAACATGGTACCGTTCCC
ACGGCTTCACTTCTTTATGCCGGGTTTCGCTCCGCTCACCAGCCGCGGCAGTCAACA
GTATCGCGCCCTGACCGTTCCTGAGTTGACTCAGCAGATGTTTGATGCCAAAAACA
TGATGGCTGCTTGTGACCCACGTCATGGACGCTACCTAACAGTCGCAGCTATGTTCC
GCGGTCGAATGTCGATGAAGGAAGTGGACGAGCAAATGTTGAATGTTCAAAATAA
AAATTCGAGCTATTTTGTTGAGTGGATCCCCAACAATGTGAAGACTGCCGTGTGTG
ACATTCCACCACGTGGCCTAAAGATGTCTGCCACCTTTGTTGGTAACAGTACTGCCA
TCCAGGAACTGTTCAAGCGTGTCTCCGAGCAGTTCACCGCTATGTTCCGTCGGAAG
GCCTTCTTGCATTGGTACACCGGTGAGGGTATGGACGAGATGGAGTTCACTGAAGC
CGAGTCTAATATGAATGATTTGGTCAGTGAGTACCAGCAGTATCAAGATGCAACGG
CCGAGGAGGAGGGTGAATTTGACGAAGAGGCTGAGGAAGCC

>TUBBI3_FM_transcriptom
ATGCGTGAGATTGTGCACATTCAGGCGGGTCAGTGTGGGAATCAGATTGGTGCTAA
GTTTTGGGAAGTTATTTCGGACGAACATGGAATTGATCCCACTGGGACTTACCACG
GAGACTCTGACTTGCAACTGGAGCGTATCAACGTCTATTACAATGAAGCAACTGGT
GGGAAATATGTCCCACGGGCTATCCTCGTTGATTTGGAGCCGGGCACAATGGACAG
TGTTCGCGCTGGCCCGTTCGGTCAACTGTTTCGTCCTGACAACTTTGTGTTCGGCCA
GAGCGGTGCCGGAAATAACTGGGCTAAAGGACATTACACCGAGGGAGCTGAATTG
GTTGACTCAGTTCTCGATGTTGTCCGTAAGGAAGCCGAGTCGTGCGACTGTCTGCA
AGGCTTCCAGCTAACCCATTCATTGGGTGGTGGTACCGGTTCTGGTATGGGTACTCT
GCTCATCTCTAAAATCCGGGAGGAGTATCCCGATCGGATCATGAACACTTTCTCCGT
GGTCCCCTCACCGAAGGTATCCGATACCGTTGTGGAACCGTATAATGCCACCCTAT
CTGTCCATCAGTTGGTTGAGAACACGGATGAGACGTATTGTATTGACAATGAGGCT
CTGTATGATATTTGCTTCCGCACTTTGAAGTTGACCACGCCAACCTACGGTGATTTG
AACCATTTGGTGAGTGCCACTATGTCCGGTGTGACTACTTGTCTGCGTTTCCCTGGC
CAACTGAATGCTGATCTCCGAAAGCTTGCCGTGAACATGGTTCCATTCCCCCGTTTG
CACTTCTTCATGCCTGGGTTCGCTCCACTGACTAGTCGAGGTAGTCAGCAATATCGT
GCCCTGACTGTACCCGAGTTGACTCAACAAATGTTTGATGCGAAGAATATGATGGC
CGCTTGTGACCCACGTCACGGGCGTTACTTGACAGTGGCTGCAATGTTCCGTGGTCG
CATGTCGATGAAGGAGGTGGACGAGCAGATGTTGAATGTGCAGAATAAGAATTCG
AGCTATTTCGTCGAATGGATCCCCAACAACGTGAAGACTGCTGTTTGTGACATTCCA
CCCCGTGGATTGAAGATGTCCGTCACATTCATCGGAAATAGTACTGCCATTCAGGA
ATTGTTCAAACGTGTCTCGGAACAGTTCACTGCCATGTTCCGTCGTAAAGCTTTCTT
GCATTGGTACACTGGTGAAGGAGTTTAT

>TUBBI4_FM_transcriptom

ATGCGGGAGATCGTGCATATGCAAGCCGGCCAATGTGGGAATCAGATCGGTTCGAA
GTTCTGGGAAGTCATATCCGACGAGCATGGCATTGATCCAAATGGAGTCTATCACG
GAGACTCGGACATGCAACTGGAACGAATCAACGTCTACTACAATGAAGCTATGGGC
GGTAAATACGTTCCACGTGCCATTCTCATCGATCTTGAACCGGGCACGATGGATAG
CGTGCGTGCTGGCCCGTTTGGTCAACTTTTTCGACCGGACAATTTCATTTTTGGTCA
TACGGGCGCGGGTAATAATTGGGCCAAGGGGCACTACACTGAAGGTGCCGAACTT
GTGGACTCTGTGTTGGACGTGGTTCGAAAAGAATCCGAATCGTGCGACTGTCTGCA
AGGGTTCCAACTGACCCACTCCCTTGGCGGAGGTACCGGATCTGGGATGGGTACAC
TGTTAATTTCGAAAATTCGTGAGGAATATCCTGACCGTATTATGACCACATTCTCTG
TCGTTCCATCACCAAAAGTCTCTGATACCGTTGTCGAACCGTACAATGCTACTCTTT
CGGTGCATCAGCTTGTGGAAAACACAGACGAAACATTCTGCATTGACAATGAAGCT
CTGTACGACATATGTTTCCGCACTCTGAAACTAACGAATCCGACTTACGGAGATCT
GAATCATCTGGTCAGTGCCACTATGTCCGGTGTGACAACTTGCTTGCGTTTCCCGGG
TCAGCTGAACGCCGATCTCCGGAAACTGGCAGTCAATATGGTCCCGTTTCCGCGTTT
GCACTTCTTTATGCCCGGTTTCGCGCCACTCACGAGCCGTGGAAGCCATCAGTACCG
CGCTGTGACGGTTCCCGAGTTGACTCAGCAAATGTTCGATGCGAAGAACATGATGG
CTGCATG

>ATP5B_FM_transcriptom

CTGTCATATAAGGCGCCGAGTTTCATTCCTGTCCGCCATCATGAGTCTGGCGCGGTC
GGTGTTTTCGGTGGCGAGGCTTCTGAAGCCGAGCGGTTCGCTAGCAGGATTCTGTG
CCAAGCCGCAGGTTGTTTCAACGCTTAATAGCAAACGCTGTTATGCTGCCGCGGCT
GCCGCAACTGCGCCTGCTGCAGCTGCTCCTGCTCGCAACTGCGCCTGCTGCAGCTGC
TCCTGCTGCAGCTGCGCCTGCTGGAGCACTCTCAACCGGTCGAGTGATTGCCGTCAT
TGGCGCTGTAGTTGACGTCCAGTTCGATGAAAACTTACCCCCTATTTTAAATGCACT
CGAAGTGAAAGGAAGGTCCCCCAGACTTATTCTTGAGATCGCCCAGCATCTTGGTG
AAAATACCGTTCGGACAATTGCCATGGATGGTACCGAGGGTTTGGTGCGTGGGCAA
CAATGTGTGGACACGGGCGGCCCAATATGCATTCCTGTTGGCCCCGAGACTCTCGG
AAGAATTATGAATGTTATCGGTGAACCTATCGATGAACGGGGTCCAATCTCAACCA
AATTGAAGGCAGGTATTCACCAAGATGCCCCTAGCTTTGTGGAAATGAGTACAGAT
CAAGAAATCCTGGAAACAGGAATCAAAGTTGTTGACCTGTTAGCTCCATATGTTAA
GGGTGGGAAAATTGGTCTCTTCGGGGGTGCTGGTGTGGGCAAAACAGTGCTTATCA
TGGAATTAATTAATAACATTGCCCGTGCTCACGGTGGCTACTCAGTGTTCGCCGGCG
TTGGCGAGCGTACTCGAGAAGGTAACGATCTCTACCATGAAATGATTACCACGGGT
GTAATCGATCTAAAGGGTGACAAGTCCAAGGTGTCACTCGTTTATGGTCAGATGAA
CGAACCCCCAGGGGCTCGTGCCCGTGTTGCTCTGTCTGCTTTGACTGTAGCAGAATA
CTTCCGTGATCAAGAGGGTCAAGATGTGCTCTTGTTCATTGATAACATTTTCCGATT
CACACAGGCTGGCTCAGAAGTTTCCGCTCTATTGGGTCGCATTCCGAGTGCTGTTGG
CTACCAACCAACGCTTGCAACGGATATGGGTACCATGCAAGAGAGGATTACGTCAA
CTAAAAAGGGGTCCATCACATCTGTACAGGCCATCTATGTGCCTGCCGATGATTTG
ACTGATCCCGCCCCAGCGACCACTTTCGCTCACTTAGATGCCACTACCGTGCTTAGC
CGAGCAATTGCTGAACTTGGGATCTATCCCGCTGTGGATCCTCTTGACTCTAACAGT
CGTATCCTGGACCCCAACATAGTTGGCGAAGAACACTACTCGGTTGCACGTGGTGT
CCAAAAGATCCTCCAAGACCATCGATCTCTTCAAGATATCATCGCTATTCTGGGTAT
GGATGAACTATCAGAAGAAGATCGGCTGACTGTCGCACGTGCTCGAAAAATTCAGC
GATTCCTGAGCCAACCTTTCCAAGTCGCCGAAGTGTTCATCGGTCGGGAGGGTAAA
CTAGTATCCTTGAAGGATTCCATTAAGGGTTTCAAGATGATCATCAATGGTGATTTG
GATCATATCCCAGAGGCTGCCTTTTTCATGGTTGGTGCCATCGAAGATGTGTTAGAG
AAGGCGGAACAATTGGCTAAAGAAAGTGCATGAGTCTTTCAACGTTACGTGGCCGG
CTGGATGCTAGTAATTTTAGCCATGTAAACACTGCTCGGTAGATCAAAAAATAAAA
CTAATAGGCTTTTCTACAGCTC

>BACT_FM_transcriptom

CGGCTCGTTAGCATTCTGCGGCGCTATGGGTGATGAAGACGTTGCTGCTCTAGTGAT
CGACAATGGTTCGGGTATGTGCAAGGCTGGATTTGCCGGCGACGATGCGCCTCGTG
CTGTGTTCCCGTCCATTGTTGGTCGACCAAGACACCAGGGTGTGATGGTTGGTATGG
GTCAGAAAGATAGCTACGTTGGAGATGAAGCTCAATCAAAACGCGGTATCCTGACA
CTGAAATATCCCATTGAGCATGGAATTGTGACGAACTGGGATGACATGGAGAAGAT
TTGGCATCACACTTTCTACAACGAATTGCGTGTTGCACCCGAGGAGCATCCGGTGCT
CTTGACCGAAGCCCCACTTAACCCTAAAGCCAACCGGGAAAAGATGACTCAGATCA
TGTTTGAGACCTTCAACGCACCGGCTATGTATGTGGCGATTCAAGCTGTTCTGTCTC
TTTACGCCTCTGGTCGCACCACTGGTATTGTGTTGGATTCTGGTGACGGTGTTACCC
ATACTGTGCCCATTTACGAGGGCTATGCTCTCCCTCACGCCATTCTGCGTCTGGACC
TCGCCGGTCGTGATCTGACTGATCATCTCATGAAGATCCTTACTGAGCGTGGCTACA
GCTTCACCACGACTGCCGAGCGTGAAATTGTGCGAGACATTAAGGAAAAGTTGTGC
TACGTTGCTCTTGATTTTGAGCAAGAAATGGGAACTGCGTCCGCCAGTTCTTCTCTG
GAGAAGAGCTATGAGCTTCCTGATGGACAAGTAATCACCATTGGCAACGAGCGGTT
CCGTTGTCCTGAGGCTTTGTTCCAACCTAGCTTCCTTGGTATGGAGTCAGCAGGTAT
CCATGAGACCACCTACAACTCCATCATGAAATGCGACGTCGATATTCGAAAGGATC
TGTACGCTAACACTGTCCTCTCCGGAGGCTCTACCATGTACCCAGGCATTGCTGACC
GCATGCAGAAGGAAATTACTGCACTGGCCCCGAGCACCATGAAGATCAAGATCGT
GGCGCCTCCAGAACGCAAATACTCTGTGTGGATTGGTGGTTCTATTCTGGCGTCGCT
TTCCACCTTCCAGCAAATGTGGATCTCGAAGCAGGAATACGATGAGTCTGGCCCTG
GCATCGTTCACCGCAAATGCTTCTAGATTGTTATTTCATTTTTTTTTTTTGGCAATCG
ACTGCAAGGAATGTATTAGATACAAAGAACAAAATGAAATAACAATCTAGAAGCA
TTTGCGGTGAACG

>GAPDH_FM_transcriptom

TACCACACGGCAGCTGTATCCAAATTCGTTATCGTACCTGAAACACGAGTGAGTGC
AACGAAGAAAACCAACTTACCACGAAATCAGCTTCACGAAGTTATCATTCAGACTA
ATTCCAGCTTGGGCATCGAAGACAGAGGAGCAGTGCGTACCCAGCATGTCCGAACT
AACCAGAAGTTCTTCGGTGTAATCCAGAATTCCCCTCATGGGTCCGTCAGCTGCGG
CCTTCACCGCGGCTTTGATTTCGTCGTACTTAGCAGCTTTGTTCAAGCGACAAGTGA
GATCCACAACGGAAACATCAGCAGTGGGAACACGGAAGGCCATGCCGGTGAGTTT
TCCATTCAGTTCGGGAATGACTTTGCCCACAGCCTTGGCTGCACCGGTGGAGGCAG
GTATAATGTTCTGACCTGCGCCCCTACCGTCTCTCCAAGCTTTGCTCGAGGGACCAT
CAACGACCTTCTGAGTAGCTGTGTACGAATGCACAGTTGTCATCAAGCCTTCAACA
ATGCCAAACTTGTCATTGATAACCTTTGCGAGAGGGGCAAGACAATTTGTTGTGCA
CGAAGCGTTGGACACCACAGACATATCCTTGCTGTATTTCTCGTGGTTCACTCCAAA
CACGAACATCGGCGCGTCGGCCGACGGGGCCGAAATAATAACCTTCTTGGCACCTC
CCTGAAGGTGAGCCTTGGCCTTATCGATAGTAGTAAATACGCCAGTGGATTCAACG
ACATATTCGGCTTGGCTGGCACCCCAAGGGATTTGAGTAGGATCACGCTCGTTGTG
AACAGCAATTGGGCTATTTCTCACAATCAGCTTGTTTTGATCAACGGTAATGTCGTA
CGGAAAACGACCGTGAGTGGAGTCATACTTGAGCATATATCGCATGTAATCAAGAT
CAATGAAGGGATCATTGACAGCTACGACGTCAACGGTGCCCTTCTCCACTGCTGCA
CGAAGAACGAGCCGGCCAATACGGCCAAATCCATTAATACCAACACGAGCCTTCCC
CATCGTTGCTAACACTAACCAAACGTGACGACGCTAACGAATATATGCTATGCGCT
TAGCCC