Snails as intermediate hosts for parasitic infections: host-parasite relationships and intervention strategies

MPhil THESIS

Mgr. Lukas Konecny

UCL
Department of Biosciences

London, 2022

Statement of originality

I, Lukas Konecny confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

A fundamental prerequisite in the fight against medically and veterinary important parasites transmitted by intermediate host snails is a good knowledge of their life cycles, host specificity and geographical distribution. With scientists around the world collecting material from the wild and generating vast amounts of sequencing data, there is a huge opportunity to expand our knowledge of host-parasite relationships from the comfort of an office chair. With these motivations in mind, a bioinformatics tool was developed that has proven to be time efficient and accurate for the rapid identification of hidden parasites in publicly available datasets. Several dozen hidden parasite infections were discovered from the 2150 gastropod datasets tested, and some of these relationships have not yet been described. With our better understanding and the rapid progress in development of molecular and genetic methods, new avenues are opening for the control and eradication of diseases caused by vector-borne parasites. To study crucial parasite-snail interactions and eventually try to interfere with the infection, it is desirable to edit the host genome. Thus, in the framework of this work, preliminary experiments for the development of the CRISPR/Cas9 protocol in *Biomphalaria glabrata*, the intermediate host of the dangerous blood fluke *Schistosoma mansoni*, were also performed. The most significant findings in this case are the proof-of-concept of cultivation of *B. glabrata* embryos in glass capillaries using natural egg fluid and the demonstration that dilution of this fluid or complete replacement by other culture media are not suitable for successful cultivation. I also show that the *Diaphanous* gene, which has been used in the past to optimize CRISPR/Cas9 in another snail model, is not suitable for our model. The ultimate goal of the development of this molecular-genetic toolbox is the eradication of schistosomiasis by replacing susceptible populations in nature with resistant populations using gene drive technology. Although disrupted by COVID-19 pandemic, this work's contribution to progress in the fight against helminthic parasitic infections is considerable.

COVID-19 disruption statement

This Mphil thesis is the outcome of a prematurely terminated PhD study under the London Interdisciplinary Doctoral Programme (LIDo) scheme. After one rotation in the labs of Prof Mark Marsh and Prof Sandip Patel, I started a second rotation in the lab of my current supervisor Prof Max Telford in March 2020. The PhD project I chose was called "Development of CRISPR/Cas9 in *Biomphalaria glabrata*, an intermediate host for *Schuistosoma mansoni*". This project aimed to develop of a protocol for culturing the embryos of this snail, the subsequent optimization of microinjections of these embryos with different gene constructs and finally the selection of suitable genes as targets for the development of CRISPR/Cas9. The ultimate goal of this project was to put this method into practice not only for the usage in developmental biology but also for future development of a gene-drive system potentially useful in the fight against schistosomiasis.

Exactly at the time of my joining the lab, the COVID-19 epidemic started to plague European countries and the first lockdowns occurred. Because of the uncertainty that was tormenting us all, it was recommended that foreigners studying at UCL should return to their home countries. So, at the end of March, I left the UK and continued my project remotely. As I did not have access to a lab, I started to learn bioinformatics on my own with the help of lab members to use this time effectively and so the idea for the first mini project and chapter - "Computational analysis of the *diaphanous* (*Dia*) gene in Gastropoda as a potential gene for targeting CRISPR/Cas9 in *B. glabrata*" was born. The assumption was that by the time I finished this miniproject I would be able to go back to the lab to continue my original PhD project. When it became clear that this was not going to happen, in July of 2020 I managed to get in touch with Laboratory of molecular helminthology at the University of Life Sciences in Czech Republic that routinely breeds *B. glabrata*. The head of this lab Dr Jan Dvorak allowed me to physically work there on experiments to culture embryos for subsequent microinjections. I tried to do this for the next few months, but this was very challenging without the necessary equipment that my lab at UCL has and without the know-how in the framework of this project. Despite the partial success described in the chapter "*In vitro* culturing of embryos of *Biomphalaria glabrata*", I decided to interrupt my studies in September 2020 until the situation regarding the pandemic calmed down, to make the most of the remaining time of my PhD studies. When it became obvious that the situation would not improve in the near future, I restarted my studies with a substitute computational project that Prof Telford offered me, namely "Probing RNAseq datasets for hidden and yet undiscovered parasite species", which is the last and main chapter of this MPhil thesis. I worked on this project remotely for the next few months. After its completion, however, my original PhD project was so disrupted, and my personal situation changed to such an extent during the pandemic that I decided to stay in my home country and unfortunately terminate my PhD studies.

This thesis is therefore a set of three sub-projects, which, although thematically close, are not directly related, making it difficult to combine them into one logical whole. However, given the results that these projects have produced, it is worth at least attempting to do so. In this work I have also attempted to compose a really detailed introduction to the subject, which may not seem to be always directly relevant to the practical part of this research but aims at least partly to compensate for the lack of practical experiments and results for obvious reasons.

Impact statement

This work has an impact from two different perspectives, which also corresponds to its structure. The first part of the work is devoted to the development of a protocol for genetic engineering of the snail *Biomphalaria glabrata*, which is an important intermediate host of the most medically important helminthic parasite, *Schistosoma mansoni*. Even though the invention of CRISPR/Cas9 and related gene drive technologies is in the public spotlight and ultimately the development of such technologies is the future for modern vector-borne parasitic disease control, it is important to realize that their implementation depends on the painstaking work of optimizing individual steps. These include, for example, construct delivery methods, validation of successful modification and even the care of the engineered organisms themselves. The knowledge about the cultivation of *B. glabrata* embryos as well as the exclusion of specific genes from the selection for optimization of this molecular tool presented in this thesis can guide future researchers who will further develop this technology in our model organism and significantly save their time. In the long term, the implementation of the developed protocol might lead to a significant reduction of the disease burden and an increase in the standard of living of populations in endemic areas. Apart from its practical impact in the fight against medically important parasitosis, the findings in the framework of this thesis can also accelerate research in other biological fields such as developmental biology.

A completely new method for identifying hidden parasite species in publicly available sequencing data is responsible for the second part of the impact of this work. Using this new methodology, it was possible to identify completely new relationships between gastropod intermediate hosts and medically and veterinary important as well as apparently unknown parasitic worm species. This knowledge may have consequences not only in the description of new parasitic species and their life cycles, but also in the monitoring of the spread of parasitic infections into areas where they have not been detected before. This can again lead to a better understanding and control of the transmission of parasitic infections and thus improve the standard of living of people in affected areas. Furthermore, due to its universality, this methodology is transferable also to other scientific disciplines where identifying hidden species in already existing RNAseq dataset is desirable. This opens the door to an almost inexhaustible number of ideas that can be easily realised using this tool. These can be, for example, the study of symbionts and commensals, investigation of food webs, but also currently popular studies of the microbiome. With the breadth of coverage that this research can encompass, the potential impact of this work in academia and beyond is difficult to estimate, but certainly undeniable. Such impact could be brought about mainly through scientific publications and publicly available code hosting platforms such as GitHub, where the sources for the presented tool is stored.

Table of Contents

List of tables

List of figures

Figure 4. Schematic illustration of diversity of genetic material that may occur in the sample for RNA or DNA sequencing. 1.a-e An organism whose genetic information is the target of sequencing (1a) may also have parasites (1b), commensals (1c), random contaminants (1d) and food (1e) genetic material within itself or immediate vicinity. 2. a-c When the sample is homogenized, the genetic material isolated (2a) and sequenced (2b), the sequences of all organisms in the sample are present in the dataset. Figure created with BioRender.com.

Figure 5. Life cycle of three most important schistosomes (CDC, 2019).

Figure 6. Life cycle of *Angiostrongylus cantonensis* (CDC, 2019).

Figure 7. Strategies and pathways for helminth vaccine development.

Figure 8. CRISPR/Cas9. CRISPR/Cas9 system consists of Cas9 protein which serves as molecular "scissors" and sgRNA which serves to "find" target sequence that is supposed to be changed. This change typically involves nonhomologous end joining (NHEJ) that leads to loss of function. When donor template is provided, the target sequence is inserted by homology-directed repair (HDR).

Figure 9. CRISPR/Cas9 protocol presented in Abe & Kuroda, 2019. a) egg masses are collected and sterilised **b)** the egg fluid is collected from egg capsules and transferred to the capillary tubes **c)** and **d)** extracted 1-cell stage embryoes are microinjected with CRISPR/Cas9 construct **e)** modified embryoes are cultivated in the capillary tubes filled with egg fluid to the juvenile snails **f)** in the case of self-fertilization all offsprings of modified snails are modified, in the case of copulation with WT snail, the rules of mendelian genetics apply. Figure created with BioRender.com. **Figure 13. Modified rooted phylogram of Gastropoda clade (Zapata et al., 2014) showing the number of Dia genes and shell orientation for each subgroup.**

Figure 10. Flowchart showing the analysis pipeline for searching and identifying COX1 sequences in SRA datasets. General description of the individual steps is on the left side and specific modifications and details used for the actual final analysis on the right.

Figure 11. Regression analysis of concentration of the egg fluid in the media and day of death of cultured individuals.

Figure 12. Deformed embryo of *B. glabrata* cultured in pure Bge medium.

Figure 13. *In vitro* culture of embryos of *B. glabrata* in glass capillary tubes filled with egg fluid in different developmental phases.

Figure 14. Modified rooted phylogram of Gastropoda clade (Zapata et al., 2014) showing the number of Dia genes and shell orientation for each subgroup.

Figure 15. The effect of relaxing the parameters on duration of the analysis. While a) gap_extend (ge), b) gap_open (go) and c) penalty (p) have almost no effect on time running of the analyses, d) decreasing the word size (ws) significantly increases the duration of the MAGIC-BLAST mapping and thus the whole analysis.

Figure 16. Number of the reads mapping to different phyla under various parameters. Changing value of word_size has no effect, while decreasing the penalty value leads to increasing number of reads mapping to Mollusca itself, but also different phyla.

Figure 17. Changing two parameters of MAGIC-BLAST at once in relation to ratio of "true reads" mapping to specific target testing organism and total mapping reads. The lowest values of word_size or gap_extend in combination with penalty leads to the most reads mapping to different species in the provided database.

Figure 18. Proportional representation of COX1 sequences of individual target groups after merging the BOLD and MIDORI 2 databases and removing duplications. The largest part of the database is represented by species belonging to the Mollusca group. Nematoda and Platyhelminthes represented less than a tenth of the sequences represented in the database.

Data accessibility

The source code for the tool, used databases and list of SRA datasets analysed in the framework of this thesis can be found on FigShare: https://figshare.com/projects/Probing_RNAseq_datasets_for_hidden_or_as_yet_undiscovered_parasite_species/127322

1. Introduction

1.1. Parasitism

The word parasite comes from the Latin form of the Greek word *παράσιτος (parasitos)*, meaning “one who eats at the table of another”. In terms of biology, parasitism is a type of symbiotic relationship, or long-term relationship between two species, where one member, the parasite, increases its fitness at the expense of the fitness of the host. An alternative definition famously proposed by entomologist O.E. Wilson "predators that eat prey in units of less than one" can also very well describe this unique trophic strategy (Wilson, 2014). If one adopts a broad definition of parasitism, such as obligate feeding on a living organism without death to the host, then almost 50% of known animal species can be classified as parasites. Stricter definitions, in which an organism must be completely dependent on the host for a significant portion of its life to be considered a parasite, would yield more modest estimates. In any case, if taken to account that every free-living metazoan species harbours at least one species of its specific obligatory parasite, the spectrum of parasite species diversity and relationships is overwhelmingly rich and yet not fully appreciated (reviewed in Poulin & Morand, 2000). The following chapters will attempt to summarise the most important aspects of parasitism in the context of this thesis.

1.1.1 The evolution of parasitism

Transitions from a free-living existence to one adopting parasitic features probably outnumber any other type of major evolutionary conversion in life on Earth history. Parasitism has independently evolved at least 223 times in just 15 animal phyla, with the majority of identified independent parasitic groups occurring in the Arthropoda, at or below the level of family (**Figure 1.**). In spite of the fact that the study of helminths, an artificial group comprising worm-like parasites from various not always closely related groups, significantly dominates over the study of arthropod parasites, only 20% of independently derived metazoan parasite taxa belong to those groups, with numerous transitions also seen in Mollusca, Rotifera, Annelida and Cnidaria. Interestingly, parasitism is almost entirely absent from deuterostomes where parasitic species are only represented by Petromyzonidae (lampreys), fishes of the Carapidae (e.g. Fierasfer) and Trichomycteridae (e.g. *Vandelia*) and *Desmodus rotundus* (the vampire bat) all belonging to Vertebrata. This obvious imbalance in the number of representatives of this lifestyle across the tree of life suggests, that although worm-like morphology and host associations are widespread across Animalia, the dual symbiotic and trophic interactions required for parasitism may constrain its evolution from ancient consumer strategies such as generalist herbivore, omnivore and carnivore predators and filter feeders (Weinstein & Kuris, 2016).

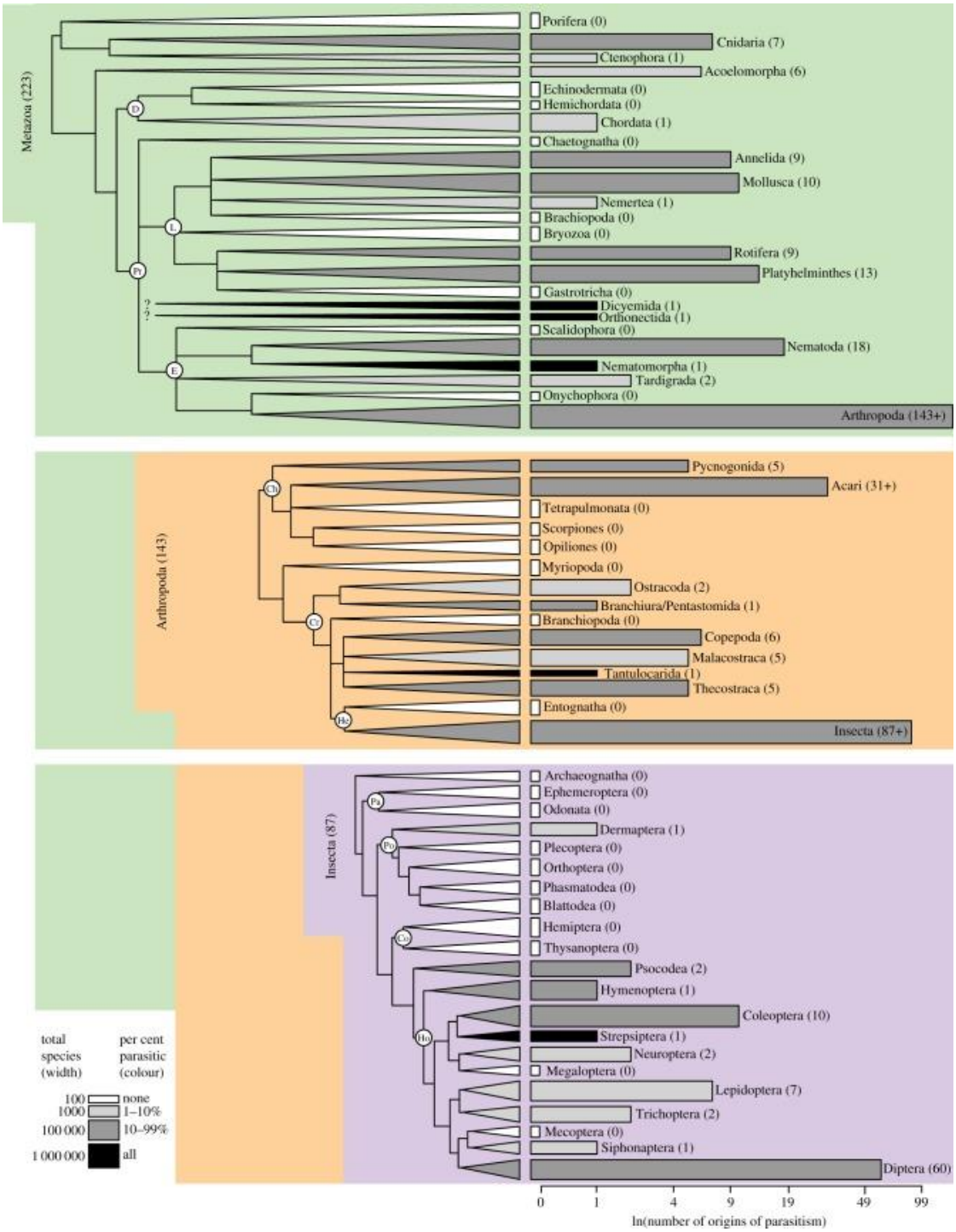

Figure 1. Origins of parasitism across Animalia. Bar plot of the natural log transformed number of independent transitions to parasitism within Metazoa, Arthropoda and Insecta, arranged according to taxonomic affiliation. The shading of the columns indicates the percentage of parasites in the group and the column width is proportional to the number of species in the group transformed to log 1000. D, Deuterostomia; Pr, Protostomia; L, Lophotrochozoa; E, Ecdysozoa; Ch, Chelicerata; Cr, Crustacea; He, Hexapoda; Pa, Palaeoptera; Po, Polyneoptera; Co, Condylognatha; Ho, Holometabola. Figure from Weinstein & Kuris, 2016.

1.1.2 Adaptations to parasitic lifestyle

Many morphological and physiological adaptations are also associated with the parasitic way of life. From the aforementioned hundreds of independent transitions from a free-living existence to a parasitic mode of life, separate parasite lineages have converged over evolutionary time to share traits and exploit their hosts in similar ways. Thus far, genome-mining has yet to uncover a single universal gene, or subset of genes, associated with parasitism (Rödelsperger et al. 2013). With advances in sequencing technologies and comparative genomics, it is becoming more evident that the development of parasitism is not associated with a specific gene or a set of specific genes. (Rödelsperger et al. 2013). While parasite genomes do show similarities such as a reduction in genome size, multiplication of specific genes, or even upregulation of specific genes at the transcriptional level compared to wild species, these similarities are more likely to reflect selective pressures associated with parasitism than common genetic background. These pressures include metabolic processes, spatial constraints linked to the lifestyle as same as strategic molecules employed in invasion and other host-parasite interactions (Cavalier-Smith, 2005)

Although not universal (Kikuchi et al., 2011; Raffaele & Kamoun, 2012) nuclear genome reduction in parasites compared with their free-living relatives has been observed throughout the whole tree of life. Such a phenomenon concerns, for example in amoebozoans (Glöckner & Noegel, 2013), mites (Mounsey et al., 2012), nematodes (Rödelsperger et al. 2013) or parasitoid hymenopterans (Ardila-Garcia et al., 2010). As another striking example may serve the parasite *Intoshia variabili* from the Orthonectida group, recently identified as highly reduced Annelida worms, has the smallest known genome among the Metazoa with only 15.3 Mbp (Schiffer et al., 2018; Slyusarev et al., 2020). Regarding parasitic flatworms, the comparison is quite difficult due to the lack of genomic data of free-living flatworms. However, based on the Animal Genome Size Database, the amounts of DNA in trematode and cestode cells are generally smaller than those reported for free-living planarian flatworms (Gregory, 2013) thus suggesting consistency of parasitic genome reduction with most of other taxa.

An interesting correspondence with the reduced genome hypothesis connected to functional loss in parasitic organisms is a secondary reduction of mitochondria which is converted in many parasitic protists to significantly simpler mitochondria-derived organelles such as mitosomes and hydrogenosomes which do not contain any DNA. (Tovar et al., 1999; Benchimol, 2000; S. Liu et al., 2016; Leger et al., 2017). Mitochondrion is also reduced in size in metazoan parasites such as trematodes and cestodes (Thanh H Le *et al.*, 2002) and nematodes (Hu et al., 2003; G.-H. Liu et al., 2013). However, despite the reduction in size, **nematodes, trematodes and cestodes have conserved 36 of 37 genes typically found in animal mitochondria including cytochrome oxidase I (COX1)** which is important for DNA barcoding in frame of this thesis (e.g. Le et al., 2001; Park et al., 2007; G.-H. Liu et al., 2013). Other major losses that can be observed in lot of parasites such as general simplification of the body plan in cestodes correlated with the loss of several homeobox gene families in comparison with their closest free-living relatives appear to be rather indirect consequence of other evolutionary changes associated with parasitism (Tsai et al., 2013). **These simplifications of**

a body plan also significantly complicate the detection of many parasites, as well as their species identification using standard morphological keys.

1.1.3 Classification of parasites

As indicated in the previous chapter, parasites are not a taxonomic group *sensu stricto*, but a collective term for organisms with the common features in their lifestyle. Within the scientific discipline of classical parasitology, however, they are divided into many overlapping groups, often of an artefactual nature. These groups are formed, for example, on the basis of the size of the parasite (microparasites/macroparasites), the shape of the organism (helminths, amoebae, etc.), the definitive localization of the parasite (endo/ectoparasites), the mode of parasitism, etc. Nevertheless, from the point of view of classical parasitology, there are three main groups into which the vast majority of *sensu stricto* parasites belong: protists, helminths and arthropods, of which this thesis focuses only on helminths.

Helminths is an artificially created group of invertebrate parasites characterized by elongated, flat or round bodies (worm-like shape) but are not necessarily related evolutionarily. There is no real consensus on the taxonomy (or groupings) of the helminths, particularly within the nematodes. However, for practical considerations the term helminth is currently used to describe members of four phyla with superficial similarities: Platyhelminthes, Annelida, and Syndermata from the superpylum Lophotrochozoa and Nematoda belonging to Ecdysosoa (**Figure 2.**). In addition to the traditional classification into groups according to the true phylogeny, helminths are often subdivided for convenience according to the host organ in which they reside, e.g., blood flukes, intestinal nematodes, extraintestinal tapeworms etc. (Castro, 1996).

The phylum Platyhelminthes includes three classes with various veterinary and medical significance: cestodes (tapeworms), trematodes (flukes and blood flukes) and monogeneans.

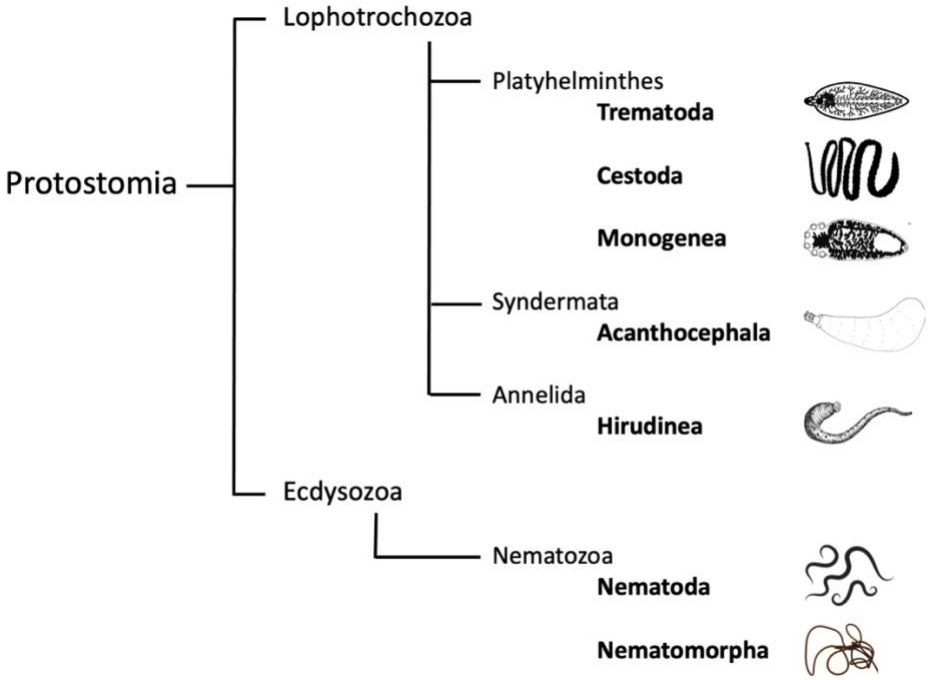

Figure 2. Classification of parasites belonging to the artificial group "helminths".

1.2 Snails as intermediate hosts for helminth parasites

Gastropoda, as the largest group within the phylum Mollusca, contains more than 65,000 species and inhabits virtually every place on planet earth. Terrestrial representatives of this group can be found in gardens, woodlands, deserts, or even on mountains; while aquatic ones can inhabit every niche from small ditches to rivers, lakes, estuaries, mudflats, the rocky intertidal, the sandy subtidal, to the depths of the oceans. Because of their extreme adaptability to diverse ecosystems, gastropods have long been known to serve as ideal hosts for a variety of parasites, often of medical or veterinary significance. Apart from serving as the exclusive intermediate hosts for all known species of the group Trematoda, which contains over 18,000 vertebrate parasites, individuals from this group also serve as hosts for dozens of species of metastrongylid nematodes (Littlewood & Bray, 2000).

1.2.1 Roles of snails in life cycles of helminths

Based on the roles of the snails and the developmental stages of the parasites they host, there are essentially six types of parasite developmental cycles where the snail plays an intermediate host role.

The first type is typical for parasitic nematodes using a snail as an intermediate host; a representative pathogen for which is *Angiostrongylus cantonensis.* The first-stage larvae (L1) of *A. cantonensis* travels through the digestive tract of the rat and are passed with the faeces. A terrestrial or freshwater snails become infected by ingesting the contaminated faeces or when these larvae penetrate their body wall or respiratory pores. L1 moult twice into a third-stage larvae (L3) that is being ingested with the infected snail by either broad range of paratenic hosts or the definitive rat host (Cowie, 2013; Thiengo et al., 2013).

The other five types are related to trematodes. The second, most notorious type of a parasite cycle involving snails can be represented by schistosomes, where the snail is infected by penetrating miracidia freshly hatched from egg release to freshwater with stool of the definitive hosts. In gastropods, the parasites replicate asexually through two generations of sporocysts and finally, thousands of cercariae emerge into the water column, and actively penetrate mammals including humans, who come into contact with the contaminated water (Grimes et al., 2015).

The third type includes two intermediate hosts, where snails become the first intermediate hosts by ingesting parasite eggs. In this case, cercariae released from the snail penetrate the second intermediate hosts, freshwater fish, where they reside as a dormant stage called metacercariae until they are consumed by the definitive hosts, fish eating mammals. *Clonorchis sinensis* is a typical example for this type of lifecycle (Zheng et al., 2017).

The fourth type includes crustacean as a second intermediate host. For example, after the penetration of the intermediate snail host, the miracidia of *Paragonimus westermani* develop into sporocysts, rediae and cercariae successively, then invade a second intermediate host, a crustacean where they also form metacercariae. When ingested by mammal, parasites mature and mate in lungs, and the eggs are then released with sputum or faeces when ingested (Q. Liu et al., 2008).

The fifth type of a cycle is exemplary for family Fasciolidae, where, after leaving the snail host, the cercariae encyst in the aquatic environment, typically on aquatic vegetation where they reside before ingested by the definitive, commonly herbivorous definitive host (Moazeni & Ahmadi, 2016).

The last, most unusual type of cycle, where molluscs play a major role can be found in Aspidogastrea. This small sister group of the Digenea (where all the aforementioned trematodes belong), numbers only about 80 species, which are capable of a direct life cycle in many bivalves and gastropods. In this case, the parasite develops from an ingested egg to adulthood in a single individual (Rohde, 1994).

1.2.2 Host specificity

A wealth of data are currently available on the relationships between key trematode parasites and their snail hosts. Conversely, the fundamental biological relationships between gastropods and nematode parasites are being elucidated.

Trematodes are highly specialized parasites and the relationships with their snail hosts are very specific, even though the assumption that one trematode species can only have one species of its specific intermediate host (Donald et al., 2004) is no longer valid. This fact can be demonstrated for example by *Schistosoma mansoni*, where, in addition to the originally recognized *Biomphalaria glabrata*, up to 22 species out of 34 belonging to this genus are susceptible to infection (reviewed in Lu et al., 2018). That the wide range of hosts for one species of flukes does not have to remain only with the genus of the intermediate host gastropod can be shown for example on the liver fluke *Fasciola hepatica*. Many different genera of gastropods of the family Lymnaeidae can serve as intermediate hosts for this fluke as has been found on the basis of natural but also experimental laboratory infections (Correa et al., 2010). On the other hand, host specificity may be undermined by incompatibility of individual strains of this parasite with other strains of the host susceptible species (Richards et al., 1992). A basic overview of the most important species of trematodes and their known intermediate hosts can be found in **Table 1**.

The situation is very different for gastropod-borne nematodes. There have been more than 160 species of snails and slugs recognized as first intermediate hosts for the nematodes from genus *Angiostrongylus,* and with the progressive expansion of these parasites across Southeast Asia, Australia, Pacific Islands, and the New World, the number of discovered susceptible hosts is likely to increase greatly (Kim et al., 2014; Giannelli et al., 2016). In addition to this huge number of intermediate hosts, these parasites also use a variety of paratenic hosts (e.g., shrimps, prawns, crabs, toads, planarians) which possess a danger for infection of humans (CDC, 2019).

This disparity and the many unknowns in the host specificity of helminth parasites complicate the effective fight against the diseases these organisms cause. The acquisition of information on host specificity for individual parasites and intermediate hosts is in this respect time and labour intensive and often depends on chance. The possibility of using freely available data obtained using omics technologies in elucidating these relationships as shown in this thesis could help massively in future control of gastropod-borne diseases

Table 1. The most important helminth species and their susceptible snail hosts identified so far. *Ac = Angiostrongylus cantonensis*; *Cs = Clonorchis sinensis*; *Fb = Fasciolopsis buski*; *Fh = Fasciola hepatica*; *Of = Opisthorchis felineus*; *Ov = O. viverrini*; *Pw = Paragonimus westermani*; *Sh = Schistosoma haematobium; Si = S. intercalatum; Sj = S. japonicum*; *Smal = Schistosoma malayensis*; *Sman = S. mansoni*; *Smek = S. mekongi*, Table reworked from and references in Lu et al., 2018.

Parasites	*Ac*	*Cs*	*Fb*	*Fh*	*Of*	*Ov*	*Pw*	*Sh*	*Si*	*Sj*	*Smal*	*Sman*	*Smek*
Achatinidae													
Achatina fulica	+	–	–	–	–	–	–	–	–	–	–	–	–
Ampullariidae													
Pila ampullacea	+	–	–	–	–	–	–	–	–	–	–	–	–
Pi. angelica	+	–	–	–	–	–	–	–	–	–	–	–	–
Pi. gracilis	+	–	–	–	–	–	–	–	–	–	–	–	–
Pi. pesmei	+	–	–	–	–	–	–	–	–	–	–	–	–
Pi. polita	+	–	–	–	–	–	–	–	–	–	–	–	–
Pi. scutata	+	–	–	–	–	–	–	–	–	–	–	–	–
Pi. turbinis	+	–	–	–	–	–	–	–	–	–	–	–	–
Pomacea canaliculata	+	–	–	–	–	–	–	–	–	–	–	–	–
Po. lineata	+	–	–	–	–	–	–	–	–	–	–	–	–
Ancylidae													
Ferrissia tenuis	+	–	–	–	–	–	–	–	–	–	–	–	–
Ariophantidae													
Girasia peguensis	+	–	–	–	–	–	–	–	–	–	–	–	–
Hemiplecta distincta	+	–	–	–	–	–	–	–	–	–	–	–	–
Microparmarion malayanus	+	–	–	–	–	–	–	–	–	–	–	–	–
Assimineidae													
Assiminea latericea	+	–	–	–	–	–	–	–	–	–	–	–	–
Bithyniidae													
Alocinma longicornis	–	+	–	–	–	–	–	–	–	–	–	–	–
Bithynia fuchsiania	–	+	–	–	–	–	–	–	–	–	–	–	–
Bit. funiculate	–	–	–	–	–	+	–	–	–	–	–	–	–
Bit. goniompharus	–	–	–	–	–	+	–	–	–	–	–	–	–
Bit. inflate	–	–	–	–	+	–	–	–	–	–	–	–	–
Bit. leachi	+	–	–	–	+	–	–	–	–	–	–	–	–
Bit. misella	–	+	–	–	–	–	–	–	–	–	–	–	–
Bit. siamensis	–	–	–	–	–	+	–	–	–	–	–	–	–
Bit. troscheli	–	–	–	–	+	–	–	–	–	–	–	–	–
Parafossarulus eximius	+	–	–	–	–	–	–	–	–	–	–	–	–
Pa. striatulus	–	+	–	–	–	–	–	–	–	–	–	–	–
Pa. sinensis	–	+	–	–	–	–	–	–	–	–	–	–	–
Pa. anomalospiralis	–	+	–	–	–	–	–	–	–	–	–	–	–
Pa. manchouricus	–	+	–	–	–	–	–	–	–	–	–	–	–
Bradybaenida													
Bradybaena despecta	+	–	–	–	–	–	–	–	–	–	–	–	–
Br. ravida	+	–	–	–	–	–	–	–	–	–	–	–	–
Br. circulus	+	–	–	–	–	–	–	–	–	–	–	–	–
Br. similaris	+	–	–	–	–	–	–	–	–	–	–	–	–
Euhadra quaesita	+	–	–	–	–	–	–	–	–	–	–	–	–
Plectotropis applanata	+	–	–	–	–	–	–	–	–	–	–	–	–
Buccinidae													
Clea helena	+	–	–	–	–	–	–	–	–	–	–	–	–
Camaenidae													
Satsuma mercatoria	+	–	–	–	–	–	–	–	–	–	–	–	–
Camaena cicatricosa	+	–	–	–	–	–	–	–	–	–	–	–	–
Cyclophoridae													
Pupina complanata	+	–	–	–	–	–	–	–	–	–	–	–	–
Helicarionidae													

Parasites	*Ac*	*Cs*	*Fb*	*Fh*	*Of*	*Ov*	*Pw*	*Sh*	*Si*	*Sj*	*Smal*	*Sman*	*Smek*
Lymnaeidae													
Fossaria cubensis	–	–	–	+	–	–	–	–	–	–	–	–	–
Galba cousin	–	–	–	+	–	–	–	–	–	–	–	–	–
G. glaticallsformis	–	–	–	+	–	–	–	–	–	–	–	–	–
G. pervia	+	–	–	+	–	–	–	–	–	–	–	–	–
G. truncatula	+	–	–	+	–	–	–	–	–	–	–	–	–
Lymnaea bulimoides	–	–	–	+	–	–	–	–	–	–	–	–	–
Ly. diaphana	–	–	–	+	–	–	–	–	–	–	–	–	–
Ly. fuscus	–	–	–	+	–	–	–	–	–	–	–	–	–
Ly. humilis	–	–	–	+	–	–	–	–	–	–	–	–	–
Ly. japonica	+	–	–	–	–	–	–	–	–	–	–	–	–
Ly. neotropica	–	–	–	+	–	–	–	–	–	–	–	–	–
Ly. obrussa	–	–	–	+	–	–	–	–	–	–	–	–	–
Ly. ollula	–	–	–	+	–	–	–	–	–	–	–	–	–
Ly. palustris	+	–	–	+	–	–	–	–	–	–	–	–	–
Ly. rupestris	–	–	–	+	–	–	–	–	–	–	–	–	–
Ly. tomentosa	–	–	–	+	–	–	–	–	–	–	–	–	–
Ly. viatrix	–	–	–	+	–	–	–	–	–	–	–	–	–
Ly. viridis	–	–	–	+	–	–	–	–	–	–	–	–	–
Omphiscola glabra	–	–	–	+	–	–	–	–	–	–	–	–	–
Pseudosuccinea columella	–	–	–	+	–	–	–	–	–	–	–	–	–
Radix auricularia	+	–	–	+	–	–	–	–	–	–	–	–	–
Ra. lagotis	+	–	–	+	–	–	–	–	–	–	–	–	–
Ra. natalensis	+	–	–	+	–	–	–	–	–	–	–	–	–
Ra. ovata (Ra. peregra)	+	–	–	+	–	–	–	–	–	–	–	–	–
Ra. plicatula	+	–	–	+	–	–	–	–	–	–	–	–	–
Ra. swinhoei	–	–	–	+	–	–	–	–	–	–	–	–	–
Stagnicola palustris	–	–	–	+	–	–	–	–	–	–	–	–	–
Physidae													
Physa acuta	+	–	–	–	–	–	–	–	–	–	–	–	–
Planorbidae													
Biomphalaria alexandrina	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. amazonica	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. andecola	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. arabica	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. camerunensis	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. choanomphala	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. glabrata	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. helophila	+	–	–	–	–	–	–	–	–	–	–	+	–
Bio. intermedia	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. kuhniana	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. obstructa	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. occidentalis	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. peregrine	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. pfeiffei	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. prona	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. schrommi	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. smithi	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. stanleyi	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. straminea	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. sudanica	–	–	–	–	–	–	–	–	–	–	–	+	–

Parasites	*Ac*	*Cs*	*Fb*	*Fh*	*Of*	*Ov*	*Pw*	*Sh*	*Si*	*Sj*	*Smal*	*Sman*	*Smek*
Planorbidae													
Bio. temascalensis	–	–	–	–	–	–	–	–	–	–	–	+	–
Bio. tenagophila	–	–	–	–	–	–	–	–	–	–	–	+	–
Bulinus africanus	–	–	–	–	–	–	–	+	–	–	–	–	–
Bu. bavayi	–	–	–	–	–	–	–	+	–	–	–	–	–
Bu. beccari	–	–	–	–	–	–	–	+	–	–	–	–	–
Bu. camerunensis	–	–	–	–	–	–	–	+	–	–	–	–	–
Bu. contortus	–	–	–	–	–	–	–	+	–	–	–	–	–
Bu. crystallinus	–	–	–	–	–	–	–	–	+	–	–	–	–
Bu. forakalii	–	–	–	–	–	–	–	+	+	–	–	–	–
Bu. globosus	–	–	–	–	–	–	–	+	+	–	–	–	–
Bu. liratus	–	–	–	–	–	–	–	+	–	–	–	–	–
Bu. nasutus	–	–	–	–	–	–	–	+	–	–	–	–	–
Bu. nyassanus	–	–	–	–	–	–	–	+	–	–	–	–	–
Bu. obtusispira	–	–	–	–	–	–	–	+	–	–	–	–	–
Bu. reticulatus	–	–	–	–	–	–	–	–	+	–	–	–	–
Bu. rohlfsi	–	–	–	–	–	–	–	+	–	–	–	–	–
Bu. senegalensis	–	–	–	–	–	–	–	+	–	–	–	–	–
Bu. tropicus	–	–	–	–	–	–	–	+	–	–	–	–	–
Bu. truncatus	–	–	–	–	–	–	–	+	+	–	–	–	–
Bu. ugandae	–	–	–	–	–	–	–	+	–	–	–	–	–
Bu. umbilicatus	–	–	–	–	–	–	–	+	–	–	–	–	–
Bu. wright	–	–	–	–	–	–	–	+	–	–	–	–	–
Gyraulus convexiusculus	+	–	+	–	–	–	–	–	–	–	–	–	–
Hippeutis cantori	+	–	+	–	–	–	–	–	–	–	–	–	–
H. umbilicalis	+	–	+	–	–	–	–	–	–	–	–	–	–
Indoplanorbis exustus	+	–	–	–	–	–	–	–	–	–	–	–	–
Lanistes carinatus	+	–	–	–	–	–	–	–	–	–	–	–	–
La. purpureus	–	–	–	–	–	–	–	+	–	–	–	–	–
Planorbarius metidjensis	–	–	–	–	–	–	–	+	–	–	–	–	–
Segmentina hemisphaerula	+	–	+	–	–	–	–	–	–	–	–	–	–
Seg. trochoideus	–	–	+	–	–	–	–	–	–	–	–	–	–
Pleuroseridae													
Semisulcospira amurensis	–	–	–	–	–	–	+	–	–	–	–	–	–
Sem. cancellata	+	+	–	–	–	–	+	–	–	–	–	–	–
Sem. kurodai	–	–	–	–	–	–	+	–	–	–	–	–	–
Sem. libertina	–	–	–	–	–	–	+	–	–	–	–	–	–
Sem. mandarina	–	–	–	–	–	–	+	–	–	–	–	–	–
Sem. peregrinorum	–	–	–	–	–	–	+	–	–	–	–	–	–
Sem. toucheana	–	–	–	–	–	–	+	–	–	–	–	–	–
Pomatiopsidae													
Neotricula aperta	–	–	–	–	–	–	–	–	–	–	–	–	+
Oncomelania hupensis	+	–	–	–	–	–	–	–	–	+	–	–	–
Robertsiella kaporensis	–	–	–	–	–	–	–	–	–	–	+	–	–
Subulinidae													
Allopeas kyotoensis	+	–	–	–	–	–	–	–	–	–	–	–	–
Opeas javanicum	+	–	–	–	–	–	–	–	–	–	–	–	–
Subulina octona	+	–	–	–	–	–	–	–	–	–	–	–	–
Thiaridae													
Melanoides tuberculata	–	+	–	–	–	–	+	+	–	–	–	–	–
Tarebia granifera	–	–	–	–	–	–	+	–	–	–	–	–	–

Parasites	*Ac*	*Cs*	*Fb*	*Fh*	*Of*	*Ov*	*Pw*	*Sh*	*Si*	*Sj*	*Smal*	*Sman*	*Smek*
Viviparidae													
Bellamya aeruginosa	+	–	–	–	–	–	–	–	–	–	–	–	–
Be. ingallsiana	+	–	–	–	–	–	–	–	–	–	–	–	–
Be. quadrata	+	–	–	–	–	–	–	–	–	–	–	–	–
Cipangopaludina chinensis	+	–	–	–	–	–	–	–	–	–	–	–	–
Filopaludina martensi martensi	+	–	–	–	–	–	–	–	–	–	–	–	–
F. sumatrensis polygramma	+	–	–	–	–	–	–	–	–	–	–	–	–
Sinotaia quadrata	+	–	–	–	–	–	–	–	–	–	–	–	–

1.2.3 Methods of identification in classical parasitology

Most of the studies focused on uncovering parasite specificity and diversity still rely on old fashioned time-consuming collection of the potential host *in situ*, physical examination and dissections combined with sequencing of the specimens found. Diagnosis of a parasitic infection traditionally involves methods such as optical microscopy, or molecular methods including polymerase chain reaction (PCR) using primers specific to the parasite (Tavares et al., 2011). However, these traditional methods can result in many issues. One such issue is the difficulty of detecting parasites and identifying their morphological features at a microscopic level, resulting in reduced sensitivity of identification (Tavares et al., 2011). Molecular methods of parasite identification have their own set of issues, as PCR tests are not always reliable and can produce false positive or false negative results depending on the presence of other parasitic contaminants (Pascar and Chandler, 2018). Additionally, due to the specificity of primers needed for the amplification, PCR can also overlook unexpected taxa that were not targeted (Wylezich et al., 2019). Given these limitations to traditional methods, a separate approach that is free of biases and able to accurately detect a wide variety of parasites may be useful in the diagnosis of infections. Since lot of helminth parasites cause of infectious diseases in animals and humans, the swift identification of parasites species in the intermediate hosts, definitive hosts or environment is of great interest (Wylezich et al., 2019).

1.2.4 Omics, databases, and DNA barcoding

The information provided by the genome and transcriptome analyses of an organism is currently an irreplaceable part of the research in almost any branch of life sciences. With the recent rapid and ongoing development of sequencing technologies, their price has also fallen, and the application of these methods is becoming increasingly accessible. Many research groups take the advantage of this opportunity to find answers to a variety of scientific questions thus producing a huge amount of data that are published and archived, making them publicly available (**Figure 3.**).

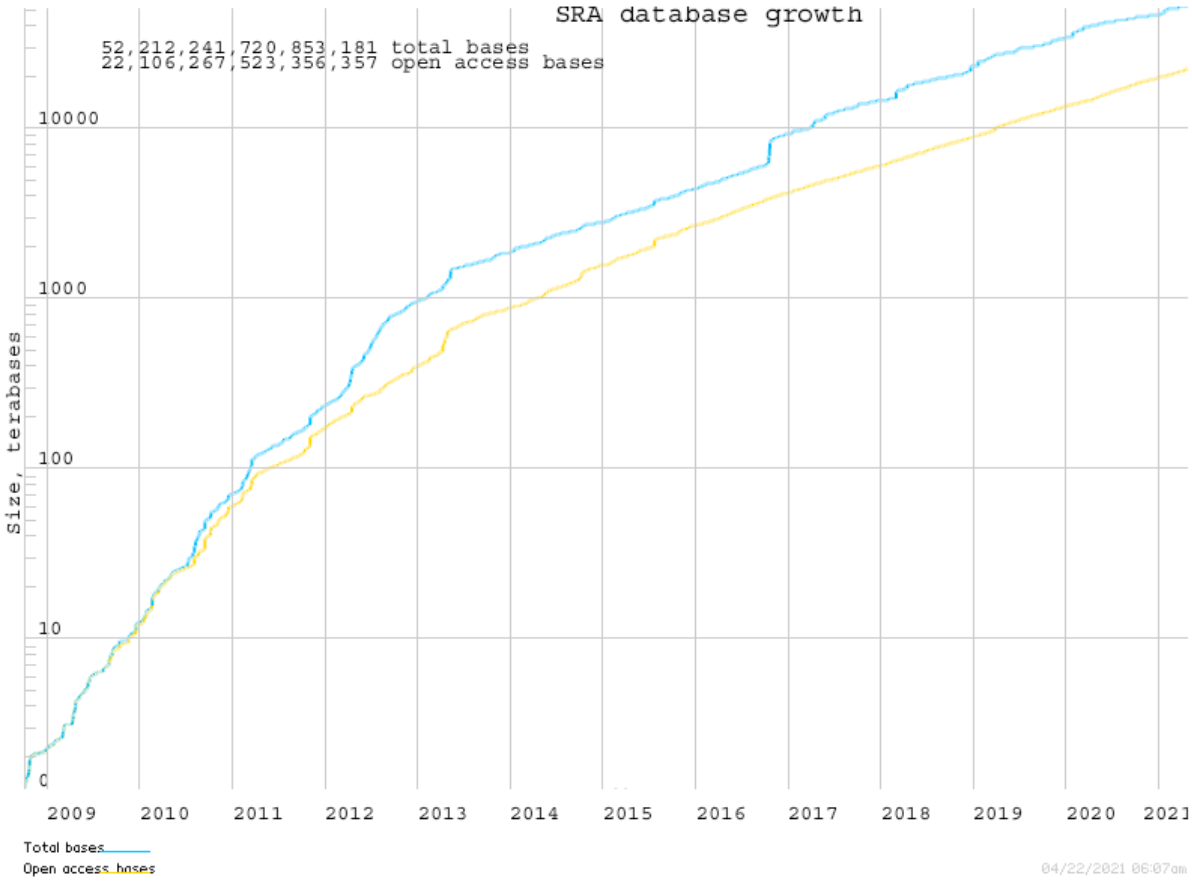

Figure 3. NCBI SRA database growth between 2009 and 2021.

Due to the nature of sequencing methods, it is inevitable that in addition to the organism that is the subject of the research, the sequences of its symbionts, commensals, food, but also ecto- and endoparasites also appear in the samples (**Figure 4.**). These "extra" sequences are usually filtered out during downstream analyses, although can be a source of a huge amount of valuable information about interspecies interactions, food preferences, host-parasite interactions but may also uncover hard to find or even previously undiscovered species. The method of species identification using a short section of DNA is called DNA barcoding. Different gene regions are used to identify the different organismal groups using barcoding.

The most commonly used barcode region for animals and some protists is a portion of the cytochrome *c* oxidase I (COX1) gene, found in mitochondrial DNA. The COX1 gene sequence is suitable for this role because its mutation rate is often fast enough to distinguish closely related species and also because its sequence is conserved among conspecifics. Contrary to the primary objection raised by sceptics that MT-COX1 sequence differences are too small to be detected between closely related species, more than 2% sequence divergence is typically detected between closely related animal species, suggesting that the barcode is effective for many animals. For instance, COX1 sequence divergence in congeneric species of insects range even from 6.6% to 11.5% (Hebert et al. 2003).

Other genes suitable for DNA barcoding are rRNA genes such as the internal transcribed spacer (ITS) rRNA often used for fungi, 16S rRNA gene widely used in identification of prokaryotes or 18S rRNA gene used for detecting microbial eukaryotes (Schoch et al., 2012; Lebonah et al., 2014; Guo et al., 2015). In the early days of DNA barcoding, scientists heavily relied on the public repositories of genetic data such as GenBank as their source for reference sequences of known species. These repositories contained many taxonomically erroneous sequences so the need of a community to create reliable curated databases of reference sequences with higher quality standards became more and more apparent. After the joint work of many scientists and institutions, two main databases of COX1 sequences are now available – Barcode Of Life Data System (**BOLD**) (Ratnasingham & Hebert, 2007) and **MIDORI** (Leray et al., 2018).

In 2004, The Consortium for the Barcode of Life (CBOL) was founded with the mission to build a barcode sequence library for all eukaryotic life. Three years later (BOLD) was launched. The system contains growing database of curated COX1 sequences with all records created using reliable PCR primers, have length ≥ 500 bp, and accepted by a COX1 Hidden Markov Model. The main advantage of BOLD is its specificity because of its strict curation protocol, but this also limits its growth. This cloud-based data storage and analysis platform currently contains barcode sequences for 326 680 formally described species covering animals, plants, fungi and protists with ~9.8 million specimens (in August 2021) (https://www.boldsystems.org).

The MIDORI database on the other hand, contains over 600,000 COX1 nucleotide records with species-level level taxonomic information selected from GenBank primarily on the basis of annotations. This allows more records to be deposited from GenBank, but also reduces the accuracy and credibility of records due to possible human error and propagation of errors by annotation software.

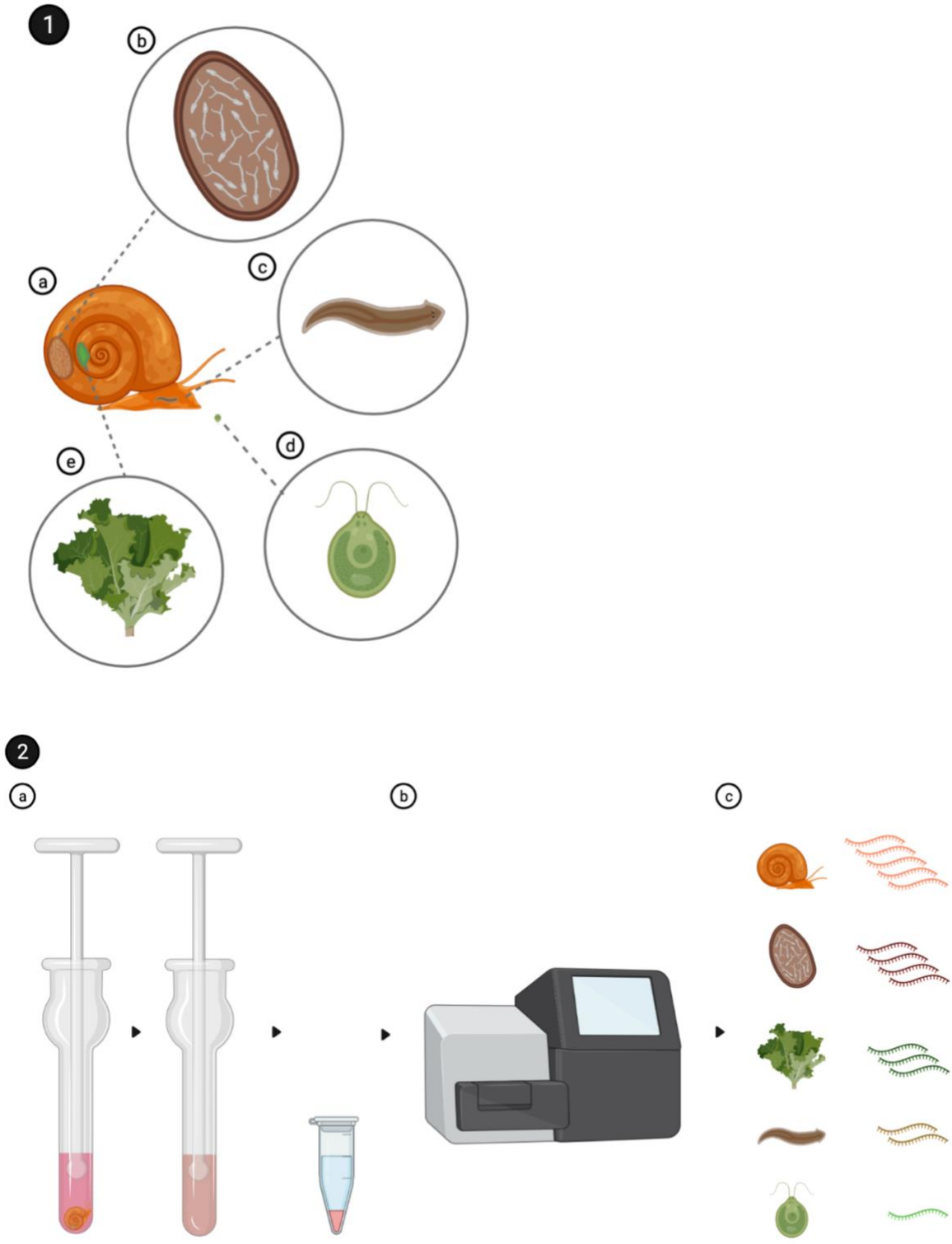

Figure 4. Schematic illustration of diversity of genetic material that may occur in the sample for RNA or DNA sequencing. 1.a-e An organism whose genetic information is the target of sequencing **(1a)** may also have parasites **(1b),** commensals **(1c),** random contaminants **(1d)** and food **(1e)** genetic material within itself or immediate vicinity. **2. a-c** When the sample is homogenized, the genetic material isolated **(2a)** and sequenced **(2b),** the sequences of all organisms in the sample are present in the dataset. Figure created with BioRender.com.

1.2.5 Gastropod-borne diseases (GBDs) – prevalence, prevention, and control strategies

Many parasites for which gastropods are intermediate hosts cause very serious diseases in humans and animals all over the world. Currently, diseases caused by gastropod-borne helminths (GBHs) are estimated to affect more than 300 million people worldwide and this number is likely to rise in the near future due to the spread of these diseases out of their endemic areas owing to global travel and climate change, and the continued neglect of these diseases as a global health problem (Giannelli et al., 2016). In addition to their devastating effects on the health of people living in these areas, these diseases also cause enormous economic damage primarily associated with the livestock industry.

The development of adequate control strategies against any disease heavily relies on a thorough understanding of the pathogen biology, ecology, and epidemiology. In the case of parasites with indirect life cycles, this includes a profound knowledge of the intermediate hosts. As examples I present here two GBDs, which differ fundamentally from each other in terms of causative agent, route of infection, but mainly in terms of relationships with intermediate host snails. In addition, the extent of our knowledge about each of them underlines the stark contrasts in their biology and potential information that we might get from wide scale analysis of their hosts. For a better understanding of the significance of this work and especially of the part dealing with the detection of these parasites using bioinformatic tools, I describe here two diametrically different parasites of importance in human medicine. Both use gastropods as intermediate hosts, but their host specificity, developmental cycles as well as their mode of detection in the intermediate host. are markedly different.

Schistosomiasis

Schistosomiasis is a neglected tropical disease caused by parasitic flatworms of the genus *Schistosoma.* This genus is characterized by gonochorism, distinctive sexual dimorphism, and localization in the vascular system of a wide range of vertebrates that serve as definitive hosts. As in other trematodes, the life cycle of schistosomes is complex (**Figure 5.**). Briefly, the snail intermediate host is actively penetrated by an invasive larva called miracidium, which hatches from the excreted egg immediately after contact with water. After penetration of the snail, the parasite quickly develops into the next stage call mother sporocyst which typically survives near the site of the invasion. Approximately two weeks after infection, the mother sporocyst begins to produce daughter sporocysts for up to 7 weeks which often migrate to more nutritionally rich areas of the host's body, which are typically the gonads or hepatopancreas (Okabe, 1964).). Each sporocyst produces dozens to hundreds of cercariae, an invasive stage infectious for the definitive host including humans, which are released from the body of the gastropod approximately 4 to 5 weeks after infection (Walker, 2011). Similar to miracidium hatching, a positive correlation between the time of cercariae release and the abundance of their hosts in the environment has been observed (Théron et al., 1977).

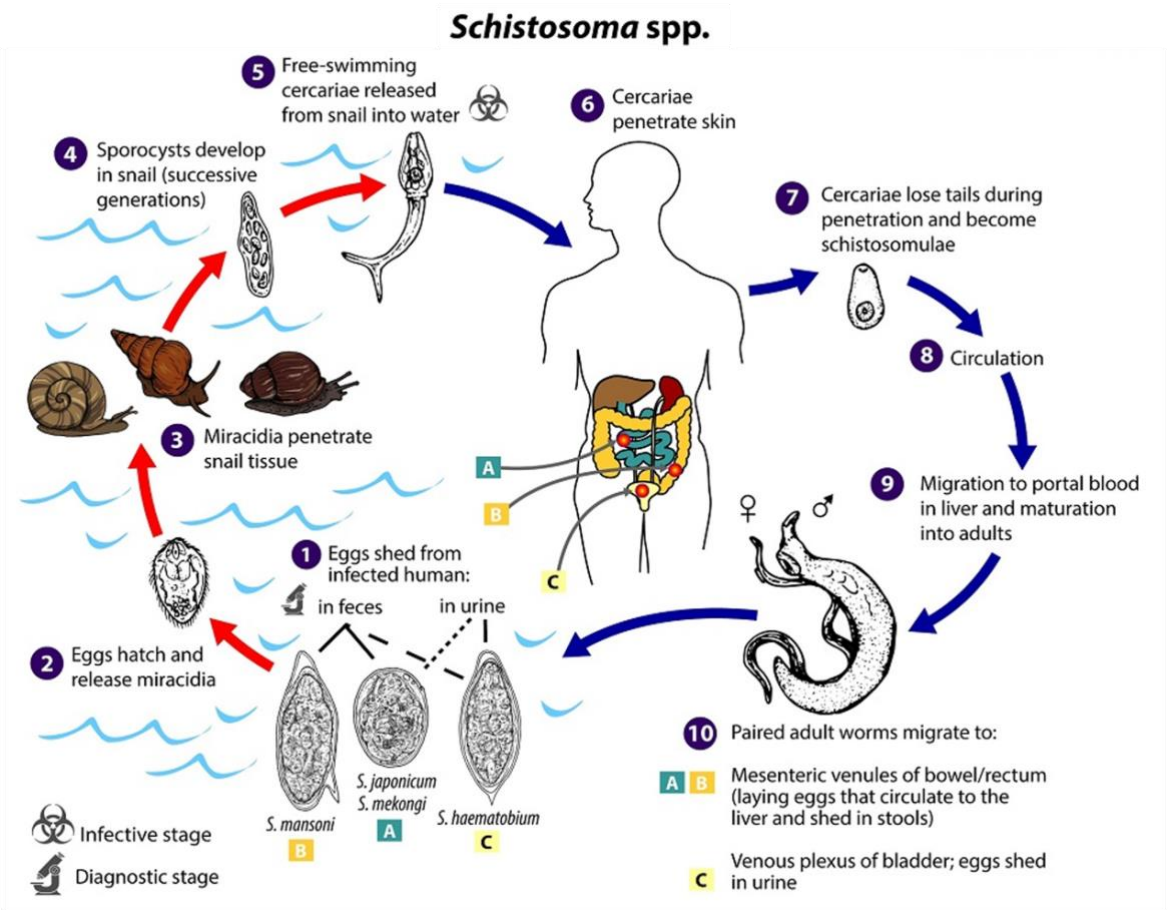

Figure 5. Life cycle of three most important schistosomes (CDC, 2019).

After a successful finding of either a specific or an unspecific host, cercariae attach to skin surface and stimulation with fatty acids (linoleic acid, linolenic acid) from the hosts skin initiates emptying of cercarial penetration glands. Proteolytic content of the glands then damages host tissues, the larva sheds its tail and enters hosts body where it quickly undergoes major physiological changes resulting in the transformation to tailless migrating larva called schistosomulum (Horák et al., 1998). Schistosomulum enters the circulation starting at the dermal lymphatics and venules and here, it feeds on blood, regurgitating the haem as hemozoin. The schistosomulum migrates to the lungs (5–7 days post-penetration) and then moves via circulation through the left side of the heart to the hepatoportal circulation (>15 days) where, if it meets a partner of the opposite sex, it develops into a sexually mature adult and the pair migrate together. At the site of schistosome final localization, which are typically small inferior mesenteric vessels for *S. mansoni* or venous plexuses around urinary bladder for *S. haematobium*, these parasites live in pairs and produce hundreds to thousands of eggs a day (Moore et al., 1956). The eggs endeavour to pass through the vessels and cross the intestinal or bladder wall into the organs lumen to be passed with the stool or urine to the aquatic environment where their snail intermediate hosts reside. Not all the eggs are successfully released from the host body, but about half of them are carried away from the mesenteric veins to the portal veins of the liver and disseminated to the surrounding tissues. Accumulation of these eggs in host tissues leads to serious pathologies such as spleno- and hepatomegaly, periportal fibrosis, portal hypertension, urinary obstruction, carcinomas, but also sterility, malnutrition or developmental retardation. Chronic infections can last for decades and may eventually lead to the death of the host (Elbaz & Esmat, 2013).

If there is a specific intermediate host on the site, the detection of the parasite in a location and intermediate host is straight-forward. The snails are collected, placed in separate containers and left in the direct light. The invasive stages, cercariae, react to this and leave the intermediate host. These are then visible to the naked eye. However, since these snails serve as intermediate hosts for more parasites, it is still necessary to determine the species of fluke based on the morphology of the cercariae Ouma et al., 1989). A very important limitation of this method of detection is the so-called **prepatent period**, which is the period between infection with the parasite and the detection of the parasite in the body, especially according to the recovery of the infectious form. In this case, this is the time between the penetration of the miracidium into the snail and the development of the first cercariae

Based on the recent report (WHO, 2020), schistosomiasis transmission has been reported from 78 countries where at least 240 million people required preventive treatment in 2018. The annual rate of deaths caused by schistosomiasis is estimated at 200 000 globally (Vos et al, 2016). Although occurrence of schistosomiasis has historically been limited to tropical and subtropical regions of Asia, Africa and South and Central America, first autochthonous human cases were identified in southern Europe, Corsica in 2013 and 2015 (Holtfreter et al. 2014, Berry et al. 2016). Interestingly, the causative agent of the infections was hybrid between mostly anthropophilic *S. haematobium* and strictly zoophilic *S. bovis*, which has been previously described infecting rodents in sub-Saharan Africa (Catalano et al., 2018, Kincaid-Smith et al., 2018) This phenomenon could thus indicate a gradual movement of schistosomes and their intermediate snail hosts to the northern regions of the globe and putting that even more people at risk.

Angiostrongyloisis

Angiostrongyliasis is an infection by a roundworm of the from genus *Angiostrongylus sp.* The life cycle of these parasites is well documented (**Figure 6.**). The first stage of larval worms is excreted in rat faeces (definitive host)..These infected faeces are ingested by snails or slugs that serve as intermediate hosts, but it is also possible for the larvae to enter the snail by penetrating the body wall or respiratory pore. Larvae develop from the first larval (L1) to the third larval stage (L3) in snails and remain there until the snail is eaten by a definitive or dead-end host or dies. Once the third instar larval stage snails are eaten by the host, they enter the small intestine where they penetrate its wall and enter the bloodstream. They then move passively through the bloodstream and some of them eventually reach the central nervous system and the brain. After entering the brain, the larvae develop into the subadult stage. After reaching the subadult stage, the worms leave the brain, move into the venous circulation and from there into the right ventricle of the heart and into the pulmonary arteries. Here the worms grow and mature, mate and the females lay eggs. The eggs travel through the bloodstream to the lung tissue where they hatch into L1 larvae. Depending on the degree of infection, the rat may suffer significant damage to the arteries caused by the abundance of adult worms and to the lungs caused by inflammatory reactions to the larvae. These first-stage larvae then break through the walls of the bronchi and lung chambers, move into the trachea in respiratory secretions and are swallowed to be released in the faeces. This cycle takes approximately 45 days. Ingestion of infected hosts is also a route how humans acquire infection, which represents a dead end in the life cycle of the parasite. This happens most often by intentional eating of raw or undercooked snails, or other infected paratenic hosts, or by eating raw fruits or vegetables contaminated by a small, infected gastropods (Cowie, 2013).

In an accidental human host, infective larvae migrate to the brain and develop into juvenile worms or even adults, but they quickly die inciting the eosinophilic meningitis manifesting with symptoms such as severe headache, neck stiffness, paresthesia, vomiting and nausea. Eventually, severe infection often leads to ascending weakness, quadriparesis, areflexia, respiratory failure, and muscle atrophy, and ultimately death if not treated (Wang et al., 2012). The problem why these infections so often occur is that the L1, L2 and L3 larvae, which are found in the anterior interhost, measure a maximum of 0.5 mm and are 0.03 mm wide, so it is basically impossible to see them with the naked eye, and for the evaluation if the gastropod is infected, dissection and microscopic examination is needed (Cowie, 2013).

While studies on *Angiostrongylus*-gastropod interactions are currently limited, to date there are 29 known species of gastropods in Americas alone that can serve as intermediate hosts (Valente et al., 2020). The spread of *A. cantonensis* via newly introduced gastropod species has been repeatedly reported via terrestrial snails as observed in the Hawaiian Islands, or Gulf Coast region of the USA (i.e., *Achatina fulica*, *Zachrysia provisoria*, *Bradybaena similaris,* and *Alcadia striata*) (Kim et al., 2014; Stockdale-Walden et al., 2015). However, the spread of rats may be more important in spreading *of Angiostrongylus* nematodes, especially to areas to where typical gastropod intermediate hosts are not present, such as Jamaica and the Canary Islands, and the diversity of possible intermediate snail and slug hosts facilitates this (reviewed in Cowie, 2013).

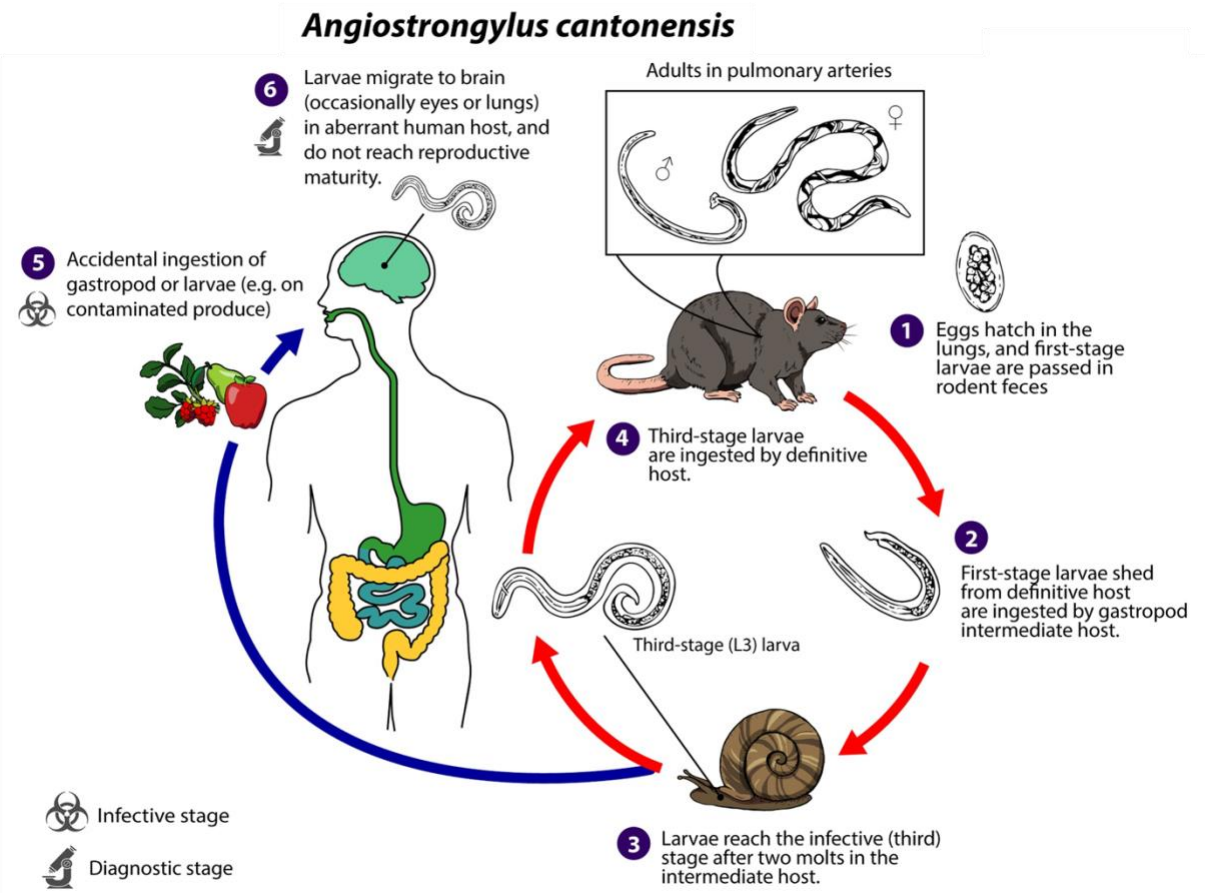

Figure 6. Life cycle of *Angiostrongylus cantonensis* (CDC, 2019).

Above all, global travel, climate change, and globalization act as major drivers for the emergence of *Angiostrongylus* infections worldwide. Furthermore, the recent identification of this nematode in Spain in native European gastropods together with the increase in suitable habitats for this pathogen due to climatic factors suggests that infection with *Angiostrongylus* may become a global health risk (Martin-Alonso et al., 2015). For this reason, extensive investigations of gastropod-angiostrongylid relationships need to be carried out to appropriately determine the risks and possibly implement preventive measures necessary to minimize the risk of infection.

1.2.6 Traditional strategies against GBDs

Basically, all GBDs are traditionally classified under the broader term neglected tropical diseases (NTDs), which are a group of parasitic, viral and bacterial infections that each year cause an estimated 534 000 deaths and a disease burden of 57 million disability-adjusted life–years (DALYs). The World Health Organization (WHO) advocates five strategies for preventing and controlling NTDs: preventive chemotherapy, intensified case management, control of disease vectors, provision of clean water and sanitation and veterinary public health measures (Hotez et al., 2006; WHO, 2010 in Kappagoda & Ioannidis, 2014)..

1.2.6.1 Drugs and vaccines against GBDs

Although effective and affordable therapeutics exist for many of these diseases, their availability to people living in highly endemic areas is often limited. Through humanitarian programmes, transmission of these diseases is often controlled through mass drug administration (MDA), defined as the provision of a therapeutic dose of an effective antipathogenic drug to the entire target population, regardless of infection status or disease symptoms. MDA is recommended as a potentially effective component of an integrated strategy to combat parasite reservoirs in many neglected tropical diseases, including malaria, onchocerciasis (river blindness), lymphatic filariasis or schistosomiasis (WHO, 2006, 2012, 2017, 2019), but its popularity is currently declining due to emerging pathogen resistance to these drugs and the toxicity of these therapeutics (Vale et al., 2017; Zuber & Takala-Harrison, 2018).

The lack of progress toward vaccination against helminth infections reflects scientific obstacles, such as the complexity and diversity of the helminth parasites. The complicated developmental cycles of helminths using snails as their intermediate hosts includes a various developmental life stages, with different gene expression profiles and thus plethora of targets for vaccination against these organisms. While this might be seen as an advantage, the opposite is true.

Even though identification of the protective antigens suitable for vaccine development got much easier with omics, the complexity of the parasite protein mixture to which the host body is exposed during invasion, migration, digestion and reproduction makes the choice of the right molecule a game of trial and error. But this is only one of the many obstacles that limit this research. Perhaps the most critical is identifying the protective mechanisms to monitor and optimize vaccine success. Helminths trigger a type 2 (Th2) profile immune response which has the potential to be effective against most helminth infections (Anthony et al., 2007). However, the ability of these parasites to promote host regulatory pathways to block this Th2 immunity poses a challenge of directing new vaccines to select protective and not counter-productive components of the immune system. One consequence of these immunomodulatory abilities that these helminth parasites possess, is that most individuals fail to develop protective immunity to infection. It is therefore essential to induce the correct Th2 immune response and, despite this being a major challenge, appropriate stimulation of the innate compartment is now considered crucial for the design of new vaccines (Pulendran & Ahmed, 2011). However, relatively little is yet understood about how type 2 immunity is selected, other than that it is promoted by certain adjuvants such as alum as was proven to be crucial for the successful vaccination against veterinary important nematode *H. contortus* (Piedrafita et al., 2013).

The last major obstacle is the theoretical possibility of building up resistance to the vaccine in case the person has already been infected with the parasite. This concern is mainly due to the fact that once the infection is initiated and the immune system is triggered, it is difficult to modulate the vaccination regimen or volume. In addition, previous helminth infection and its immunomodulatory nature may generate suppressive regulatory pathways that may block the proper immune response against the administered vaccine or cause severe

allergic reactions, as observed with in clinical testing a vaccine against hookworms (Diemert et al., 2012; Hewitson & Maizels, 2014)

All the above-mentioned problems with the development of vaccines against helminths are of course also valid in the context of GBDs such as schistosomiasis or fasciolosis. Even though vaccines based on recombinant antigens have led to a reduction of the worm burden by up to 70 percent in the case of schistosomiasis and 75 percent in the case of fasciolosis (Hota-Mitchell et al., 1997; Acosta et al., 2008), even here the development has met with only reasonable success but with the main difficulty of inducing complete immunity. The reason for this failure is likely to be the concomitant stimulation of mutually conflicting effector and regulatory pathways as has been shown in animal models (Rickard & Williams, 1982).

Even though the development of vaccines against GBDs is moving forward by leaps and bounds (

Figure 7.), major limitations such as targeted immunity modality, route of delivery or suitable adjuvants remain unresolved and thus do not currently seem like an intervention strategy that can be applied soon.

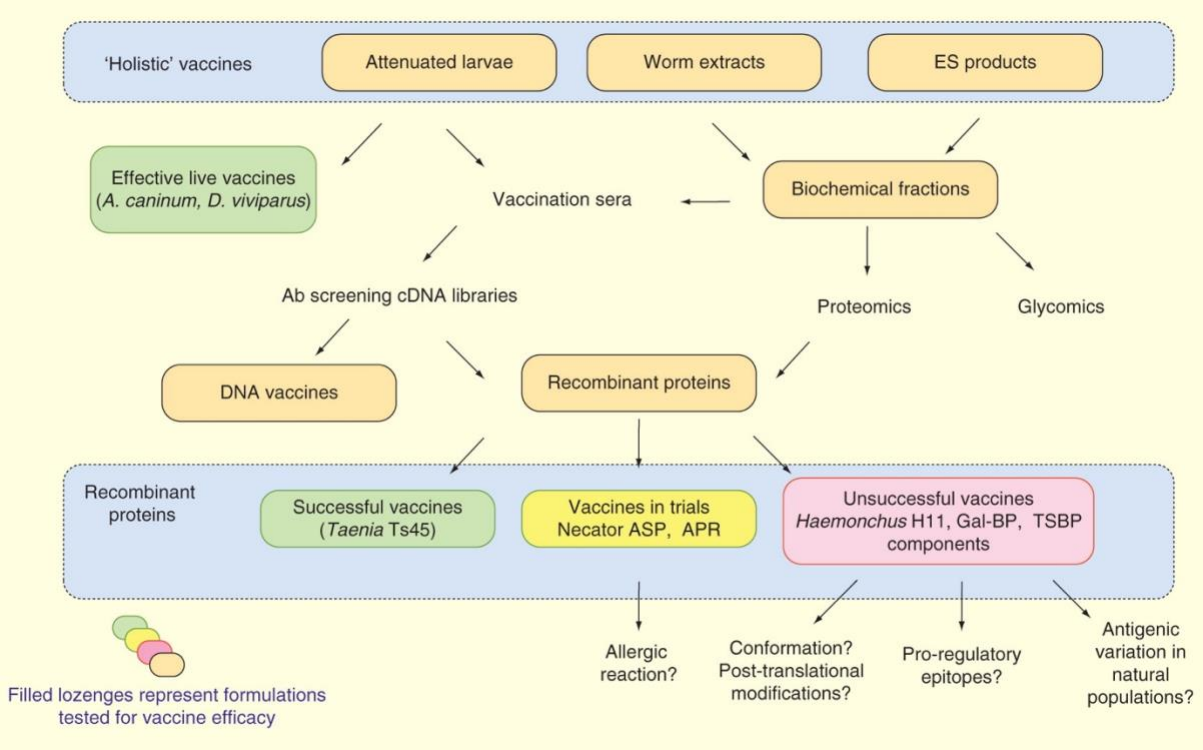

Figure 7. Strategies and pathways for helminth vaccine development.
APR: Aspartyl protease; ASP: Associated secreted protein; ES: Excretory–secretory; TSBP: Thiol Sepharose-Binding Protein. Figure from Hewitson & Maizels, 2014.

1.2.6.2 Water, sanitation, and hygiene (WASH)

Provision of safe water, adequate sanitation, and hygiene (WASH) is one of five key public health strategies to control, eliminate or eradicate neglected tropical diseases (NTDs). WASH plays a critical role, to varying degrees, in both prevention and management of all NTDs (2021). The importance of access to safe water, sanitation and hygiene in preventing transmission of soil transmitted helminth infections (STHs) or guinea worm is evident – the infection occurs when pathogenic agents are ingested, e.g. when drinking or consuming food prepared with contaminated water. The situation and importance of WASH is somewhat more complicated in GBDs.

GBDs are specific in that also contact, not just ingestion, with larvae is sufficient to acquire infection. Moreover, the danger in most water-borne diseases is posed by the faecal matter that contaminates the water surface, whereas in gastropod-borne helminths the infective stages for humans are not present in the faeces but are produced only in the presence of intermediate hosts. This leads to the fact that standard preventive measures such as hand washing after urination or defecation are irrelevant in the case of GBDs. Despite the fact that adequate sanitation such as defecating and urinating away from the water body will catch most helminth eggs and prevent miracidia from infecting intermediate host snails, sustained transmission requires only a few eggs to enter freshwater, and these do so without people directly defaecating or urinating into the water. Reductions in the input of eggs into freshwater thus may have no impact if this is not a limiting factor in overall transmission (Grimes et al., 2015).

The sum of these facts limits the practice of WASH to limiting contact with the water surface, which is often unrealistic given the socio-economic situation that often prevails in areas affected by schistosomiasis. Water is often a source of sustenance in the case of fishermen, hygiene in the context of washing clothes and personal hygiene, but it is also related to different cultural practices (Grimes et al., 2015). .

One of the promising practices under WASH is the use of soap or endod (a natural soap substitute). It was observed that these are toxic to cercariae, miracidia and specific freshwater snails, suggesting that their use during human water contact may protect or at least reduce the schistosome or other trematode infection by reducing infectivity of the invasive larvae and reduce intermediate snail numbers (Pacheco & Jansen, 1951 in Grimes et al., 2015; van Emden et al., 1974).

As is evident from this subchapter, WASH techniques are not entirely applicable to GBDs, but their adoption is certainly an important part of an otherwise very complex strategy for eradicating these diseases. Since the main focus of these strategies so far is on the use of preventive chemotherapy with first-line drugs (e.g. praziquantel for schistosomiasis), the role of WASH may be primarily very useful in tackling refractory transmission foci by preventing reinfection following this mass drug administration chemotherapy effort. Such WASH interventions must take into account many socio-economic *in situ* factors that will determine the appropriate measure and their potential effectiveness (Grimes et al., 2015).

1.2.6.3 The "old fashioned" snail control

As indicated above, control of parasitic infections is based on a combination of preventative measures which, in the case of gastropod-borne helminth diseases, rely on early diagnosis and treatment of infected people with effective drugs or potentially vaccines, improvement of life quality, and implementation of health education. Despite all efforts in these respects, the situation regarding the prevalence of these diseases in endemic areas is not significantly reduced, and it is becoming more and more clear that successful reduction of transmission or complete eradication of these diseases is not possible without effective snail control efforts, i.e. interruption of the parasite cycle by limiting its intermediate snail host (Lardans & Dissous, 1998).

In a recent meta-analysis of available, admittedly often observational data, snail control was found to be effective in schistosomiasis control, with measured reductions in both active cases (prevalence) and emergence of new cases (incidence) over time (Lo et al., 2018). This conclusion was then supported by more recent study that reported that a country-level control strategy with a focus on control of the intermediate snail hosts had larger reductions in infection prevalence than countries without it (Sokolow et al., 2016). Snail control has also notably achieved widespread success in GBDs control and elimination campaigns in many parts of Asia in the recent past (Rollinson et al., 2013). From this it is evident that snail control is an essential part of an integrated strategy against these parasitic diseases and should be intensively pursued. Snail control can be divided into four distinct strategies (McCullough et al., 1980; Lardans & Dissous, 1998; Sokolow et al., 2016; Maier et al., 2019).

1. **Chemical snail control**
2. **Environmental snail control**
3. **Biological snail control**
4. **Snail control through genetic engineering**

Chemical snail control

Chemical snail control is generally achieved using molluscicides which are chemicals or plant extracts specifically designed or simply used to kill snails.

Application of molluscicides have been the primary method used for controlling GBDs in past century with proving to be rapid and efficient mean of reducing or eliminating transmission (McCullough et al., 1980). These molecules can be divided into two classes: chemical and phytochemical compounds (Graf, 1989)..Although multiple synthetic drugs have been developed and used successfully in snail control, niclosamide is still considered to be the only safe chemical molluscicide recommended by the WHO to be used for snail control despite reported cases of resistance in molluscs after two decades of repeated use (Dai et al., 2015). Because of the difficulties in synthesizing molecules that would be effective against these intermediate hosts and at the same time environmentally friendly and, last but not least, safe for other aquatic organisms, the focus is more and more shifting to molluscicides of plant origin. Many plants were tested as a source of potential phytochemical molluscicides with mixed success (Schall et al., 2001; Lima et al., 2012; Augusto et al., 2017). *Euphorbia milii* var. *hislopii* was described as the most promising plant molluscicide, as its latex is able to kill snails at doses under 0.5 ppm in laboratory condition, it is biodegradable and it has been proved

to be less damaging to non-target organisms than niclosamide, meeting the requirements of WHO for use as a natural molluscicide (WHO, 2002; Lima et al., 2012)

Despite all the advantages, chemical-based methods for snail control are costly, extremely labour intensive, and most importantly, do not prevent repopulation of snails after treatment. Furthermore, chemical-based snail control can be toxic within the environment, which may lead to unintended ecological consequences (Rollinson et al., 2013; Lo et al., 2018)

Environmental snail control

Another method to reduce the population of intermediate hosts is to reshape the habitat of the ecosystem where the intermediate host occurs. However, as most parasitic worms are very host specific, almost every parasite, even from the same genus, needs a different intermediate host to complete its life cycle (e.g. genus *Biomphalaria*. for *S. mansoni*, genus *Bulinus* for *S. haematobium* and genus *Oncomelania* for *S. japonicum*). It is clear from this that there is no universal method of environmental snail control, as different gastropods need different ecological conditions to thrive. However, as they are almost exclusively aquatic snails, environmental control generally involves targeted drainage of aquatic ecosystems, water velocity control, channelization, seepage control, canal lining, canal relocation with deep burial of snails, vegetation removal, earth filling, mud removal from ponds and others, to reduce snail density and so on. However, as many metastudies show, environmental control is not as effective in reducing the prevalence and incidence of GBDs as, for example, administration of molluscicides or MDA to the population (Jobin, 1970; El-Emam & Madsen, 1982; Loreau & Baluku, 1991; Ofoezie & Asaolu, 1997; Khallaayoune et al., 1998).

Biological snail control

Biological snail control is understood as the reduction or elimination of an intermediate hosts with the help of its natural predators, competitors or even parasites and has long been discussed, with several reports documenting it as a possible way to reduce the burden of GBDs in endemic areas. This strategy is attractive primarily because of its potential safety and without the need to burden the ecosystem with large amounts of potentially harmful chemical agents. Although snails are on the diet of many species of organisms from a wide range of taxonomic units, there are relatively few outright snail predators that would target intermediate hosts to effectively control their populations.

Probably the most obvious group of these predators are fish, especially from the genera Cichlidae and Cyprinidae. Introduction of these fish led to a significant loss of intermediate hosts of schistosomes, but due to their low population density and broad-spectrum diet they were excluded as an effective biological control tool, even though the phenomenon of cercariophagy - targeted eating of invasive parasite larvae - was observed in these fish. In addition to fish, the use of crustaceans such as the prawn *Macrobrachium vollenhoveni* also seemed to be effective in reducing *Bulinus* snails, hosts for *S. haematobium*, but despite its partial success, there is a risk that the introduction of crustaceans into the ecosystem will also lead to the introduction of parasites that use these crustaceans as intermediate hosts (e.g. lung flukes *Paragonimus spp*.) (reviewed in Berg, 1973; Giannelli et al., 2016).

Apart from predators, an inexpensive, and promising, strategy has also been to use competing snails such as Ampullarids, but the benefits of disease transmission reduction are in

this case overshadowed by the environmental consequences of introducing these snails, which are often economically important agricultural pests. Intermediate snail hosts can also be effectively eliminated by introducing of pathogenic agents such as bacteria *Candidatus Paenibacillus glabratella* or nematodes of the genus *Daubaylia*, which cause high mortality of these snails, but their potential use in the fight against GBDs has not yet been practically confirmed (reviewed in Berg, 1973; Giannelli et al., 2016).

These key examples suggest that, today, the control of GBHs cannot rely solely on the use of biological agents, mainly because of the unfeasibility of such approaches on a global scale.

1.2.7 Control of GBDs using genetic engineering

As summarized in previous paragraphs, a variety of approaches for snail control has been more or less successfully used, including infrastructural, chemical, and biological interventions. However, in view of the shortcomings of snail control methods outlined such as molluscicide resistance, harmful effects of certain chemicals on the environment, logistical difficulties with environmental control and unfeasibility of biological control, new approaches to interrupting GBDs transmission at the intermediate host stage are needed. With recent technological advances in mollusc genetic engineering, the door is opening for the first time to control snail populations using gene drives, similar to what we see in the fight against malaria and its mosquito vectors (Kyrou et al., 2018).

For a long time, many researchers are pursuing a population-modification approach, which involves alteration, deletion, or insertion of genes that cause a parasite-resistant phenotype in an intermediate host or vector that would otherwise be fully capable of transmitting these pathogens (James et al., 1999; Isaacs et al., 2011; Gantz et al., 2015; Dong et al., 2018). The idea behind this approach is that, by introducing these modified individuals into the environment, this resistance gene would spread to the native vector population, reducing or eliminating disease transmission and leading to measurable impacts on morbidity and mortality in the affected area. Critical to this approach is the discovery of a gene that confers resistance to the transmission of the parasites, transgenesis tools for introducing the genes into the vector or intermediate hosts strains, and a mechanism to spread the genes at epidemiologically significant rates into the wild populations.

The following paragraphs will try to summarize the main gene editing technologies, their advantages, limitations, and applicability for the elimination of these diseases.

1.2.7.1 CRISPR/Cas9

Researchers have been trying to gain the ability to change DNA bases at precisely predetermined locations since the 1970s, when restriction enzymes were discovered (Danna & Nathans, 1971). To improve the robustness and specificity of these tools and their versatility for application across the tree of life, research has moved towards re-engineering naturally occurring endonucleases to alter their properties (Seligman et al., 2002). Even so, these methods were applicable to tiny fractions of known genomes. To this end, the discovery and use of eukaryotic zinc finger proteins started a new era in genome targeting and editing. Zinc fingers are small protein motifs that recognize 3-bp DNA sequences and can bind to it in a very specific manner (Klug & Rhodes, 1987). Unlike nucleases, zinc finger modules can be

combined to form a complex that could uniquely recognize up to 21 bp specific genomic sequence (Miller et al., 2007). When fused with specific endonuclease, this tool can introduce double-strand breaks (DSBs) in virtually any target site of a DNA molecule (Urnov et al., 2010). This technology was even improved with the discovery that transcription activator-like effector (TALE) proteins from *Xanthomonas* bacteria can specifically recognize one single base thus simplifying the design (Moscou & Bogdanove, 2009; Miller et al., 2011).

Despite the great enthusiasm about these revolutionary tools in the gene-engineering community, they were never really broadly adopted primarily because of the difficulty of cloning and engineering of these proteins, but also the need of redesigning and re-engineering the proteins for each individual targeting. All these obstacles and disadvantages were eliminated when CRISPR was discovered.

CRISPR stands for clustered regularly interspaced short palindromic repeat DNA sequences, which were first discovered in *Escherichia coli* (Ishino et al., 1987; Jansen et al., 2002). CRISPR sequences are associated with adjacent, highly conserved Cas genes and separated from each other by so-called spacers, which were later recognized to belong to viruses and other mobile genetic elements. After it was demonstrated that viral infection leads to the integration of new spacers derived from the phage genomic sequence into the genome of *Streptococcus thermophilus*, and that spacer sequences of CRISPR determine the targeting specificity of Cas enzymes, which provide defence against the phage, it became clear that this system could serve as an immune system to the bacteria (reviewed in Adli, 2018).

Within a year of this key discovery, it was shown that the activity of Cas enzymes is guided by short CRISPR RNAs (crRNA) transcribed from the spacer sequences and that it can block horizontal DNA transfer from bacterial plasmids. After it was shown, that Cas9 enzymes can be reprogrammed to target a desired DNA sequence in bacteria, and the guidance of the complex can be simplified by a single RNA molecule called single ?short? guide RNA (sgRNA) instead of two short RNAs as opposed to natural system (the mature crRNA and a trans-activating tracrRNA), the potential of CRISPR as a gene editing tool was recognized (Jinek et al., 2012) (**Figure 8.**). Further ground-breaking studies confirmed that this system can be also used to *in vivo* genome editing of eukaryotic cells as well which started a huge wave of experiments on all possible organisms was launched with often rather impressive results (reviewed in Adli, 2018). The implications of this discovery were immediately realised in the parasitological community as a hope in the fight against many tropical diseases such as malaria or leishmaniasis, and pilot experiments to change the genetic information of the vectors soon followed (Dong et al., 2018; Louradour et al., 2019). These experiments serve not only for the discovery of genes enabling the competence of the vector to become infected and transmit the parasite, but also for the development and potential introduction of resistant laboratory the vector into the wild with the aim of replacing susceptible populations and thus breaking the cycle of the parasite.

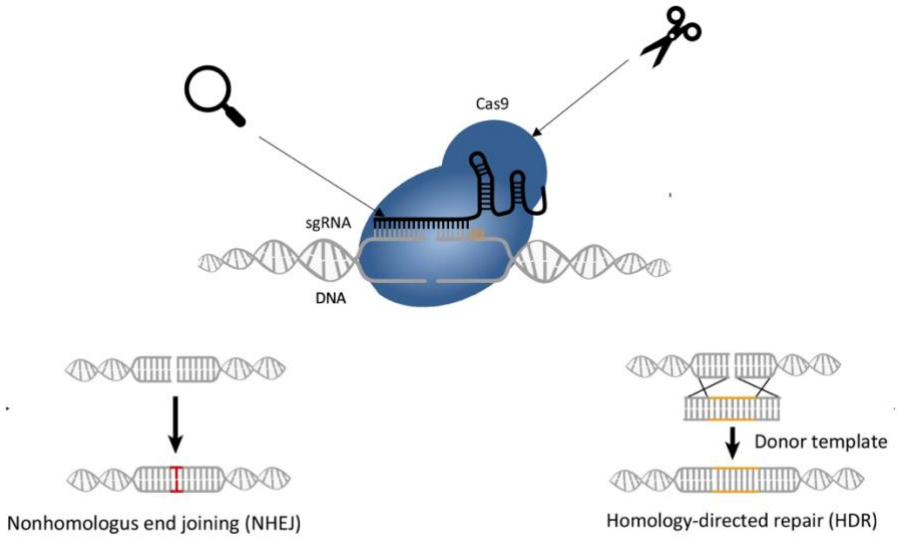

Figure 8. CRISPR/Cas9. CRISPR/Cas9 system consists of Cas9 protein which serves as molecular "scissors" and sgRNA which serves to "find" target sequence that is supposed to be changed. This change typically involves nonhomologous end joining (NHEJ) that leads to loss of function. When donor template is provided, the target sequence is inserted by homology-directed repair (HDR).

1.2.7.2 Delivery methods of CRISPR/Cas9

Since CRISPR permanently alters the genome, for the transmission of this information also to the descendants of the modified organism it is necessary that the future sex cells carry this modification. This is most easily achieved by delivering this system to a single cell developmental stage - the zygote, so that all further cells arising from it carry the same genomic information. To make this possible, there are three basic categories of delivery methods - physical, viral vectors and non-viral vectors. The most innocuous physical methods are microinjection and electroporation. Viral delivery vectors include specifically engineered adeno-associated viruses (AAV), full-sized adenoviruses or lentiviruses. These vectors are especially useful for *in vivo* work where they represent the most common CRISPR/Cas9 delivery methods. Non-viral vector systems are not currently widely used. They include systems such as lipid nanoparticles, cell-penetrating peptides (CPPs or gold nanoparticles (reviewed in Lino et al., 2018).

Microinjection is currently the gold standard for delivering CRISPR systems into cells with an efficiency close to one hundred percent (Horii et al., 2014). Its greatest advantage over viral and non-viral systems is the completely controlled quantities of the cargo, which is not limited by the molecular weight. In this method, using a microscope and a small diameter needle, a cell membrane is pierced and cargoes containing either plasmid DNA encoding both the Cas9 protein and the sgRNA, mRNA encoding Cas9 and sgRNA, or Cas9 protein with sgRNA are delivered directly to a target site within the cell. Ideally, the gene cargo can be injected directly to the nucleus and the Cas9 protein into the cytoplasm facilitating translation and movement of Cas9 to the nucleus. However, this requires two different microinjections, and due to the labour-intensive and difficult nature of this method, such an approach is impractical and rarely successful because cells often do not survive this process (H. Yang et al., 2013). Therefore, the strategy of microinjection of cargo into the cytoplasm, where it can be translated by the cell, is more often chosen. SgRNA in the cytoplasm is then bound by Cas9 and transported into the nucleus, allowing modification of the host DNA (reviewed in Lino et al., 2018).

1.2.7.3 CRISPR in Molluscs

Although molluscs are the second most abundant group of animals on the planet, CRISPR/Cas9 has only been adapted twice in this group and relatively recently. The first demonstration of CRISPR/ Cas9-mediated transgenesis in the Lophotrochozoa superphylum was achieved in the marine snail *Crepidula fornicata*. The aim of the study was to integrate the fluorescent gene mCherry into beta-catenin gene and to observe its expression during early embryonic development. Although transgenesis using microinjections was successful in principle, the success rate was only 11% and most embryos died at very early stages of development. Moreover, out of a total of 280 individuals that underwent microinjection of the construct, only two formed normal larvae, even though this snail species does not require any special culture conditions (Perry & Henry, 2015). A significantly more ambitious and more important breakthrough for parasitology was when Abe & Kuroda, (2019, 2020) "uncoiled" the snail *Lymnea stagnalis* and proved, that LsDia1 gene is responsible for axial symmetry by knocking it out of function using CRISPR/Cas9. In this study, the authors were able to raise the snail to adulthood after microinjection of embryos by culturing the snails in capillary tubes (

Figure 9.). This effectively leads to the possibility of raising an F1 generation of molluscs with permanently changed genomic information and virtually opens the door for applications in gene-engineering of gastropods serving as hosts for number of parasites of people and animals.

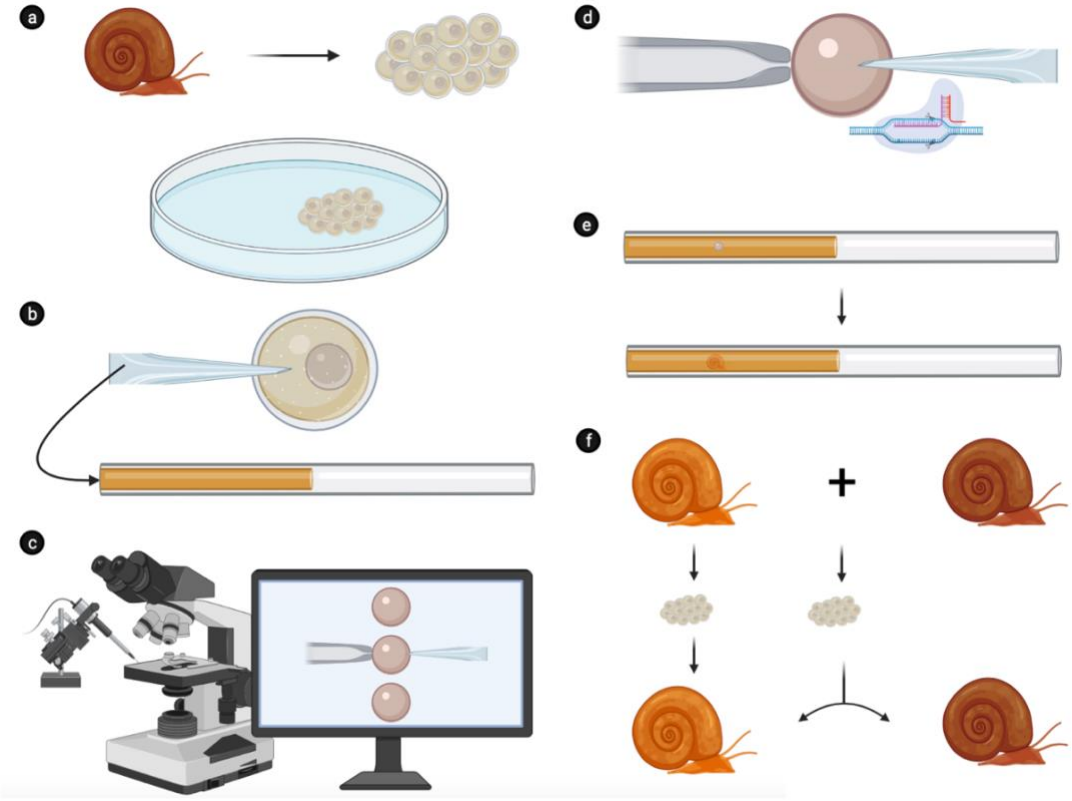

Figure 9. CRISPR/Cas9 protocol presented in Abe & Kuroda, 2019. a) egg masses are collected and sterilised **b)** the egg fluid is collected from egg capsules and transferred to the capillary tubes **c)** and **d)** extracted 1-cell stage embryoes are microinjected with CRISPR/Cas9 construct **e)** modified embryoes are cultivated in the capillary tubes filled with egg fluid to the juvenile snails **f)** in the case of self-fertilization all offsprings of modified snails are modified, in the case of copulation with WT snail, the rules of mendelian genetics apply. Figure created with BioRender.com.

1.2.7.4 Gene drives

Permanent population-modification through the introduction of genes that confer a schistosome-resistance phenotype to snail that otherwise would be fully capable of transmitting the parasite is an ultimate goal of such gene engineering efforts. The expectation is that the introduction of such an effector gene at a high enough frequency in a vector population would decrease or eliminate transmission and result in measurable impacts on overall prevalence and therefore morbidity and mortality. Critical to this approach is the discovery of a mechanism to spread the genes at epidemiologically significant rates into the target populations (Maier et al., 2019). *B. glabrata*, like most other organisms that reproduce sexually normally have a 50 percent chance of passing along a gene to an offspring. But this means that under normal circumstances, unless the genetic change that we induce significantly increases the fitness of the organism, this change does not spread further in the population until it can be completely lost in the population (Error! Reference source not found.**.**). Therefore, if we want to ensure that this change is always inherited to all offspring, we need to ensure that the second copy of the gene obtained from the WT parent is also changed as needed. The system that ensures this is called gene-drive. When introduced to the cell, Cas9 should cleave the genomic target at the site determined by the gRNA and then insert the Cas9/gRNA cassette into that locus via homology-directed repair (HDR). Cas9 and the gRNA produced from the insertion allele should then cleave the opposing allele followed by HDR-driven propagation of the Cas9/gRNA cassette to the companion chromosome (Gantz & Bier, 2015) (Error! Reference source not found.**.**). Even though the functionality of this principle has already been empirically verified in laboratory conditions in malaria-carrying mosquitoes, and future implementation is moving forward by leaps and bounds, the situation with snails is a bit more complicated. Even assuming that a protocol for genetic modification in schistosome-carrying snails can soon be established and suitable genes are found that prevent schistosome transmission, the population replacement strategy itself will not be straightforward. In hermaphroditic gastropods, such as *B. glabrata*, the ability to perform both cross-fertilisation and self-fertilisation must be considered. This means that WT individuals are not dependent on reproduction with modified populations and may never be eliminated from the population and thus not even eliminate the transmission of schistosomiasis. Within individual species and strains of *Biomphalaria*, preferences for one or the other type of reproduction have even been observed, further complicating epidemiological simulations.

2 Materials and methods

2.1 *In vitro* culturing of embryos of *Biomphalaria glabrata*

The first step in developing a protocol for the development of CRISPR/Cas9 in *Biomphalaria glabrata* is the ability to efficiently isolate the single-cell embryonic stages of the snail, which will be ultimately transfected/injected with the constructs and to culture them successfully to raise viable, fertile adults.

Adult hybrid snails *Biomphalaria glabrata/alexandrina* were used in this study. They have been maintained under a 16/8 h light/dark cycle at 28 °C and given frozen spinach leaves and artificial tropical fish food and aquarium shrimps. Pieces of polystyrene were placed on the surface of each tank and checked regularly for the presence of freshly laid eggs. Freshly laid eggs were removed from the polystyrene using a scalpel or razor blade and subjected to a series of sterilization steps by incubation in a 0.05% bleach solution, followed by gentamycin 50ng/ml with washes with sterile snail water in between. Sterilised eggs were then decapsulated and either freed from the egg fluid and transferred to a drop of complete *B. glabrata* embryonic cell culture media (Bge medium) on a Petri dish or kept in the egg fluid. The samples that were cultured in egg fluid were further diluted with different amounts of distilled water or Bge media (**Table 2.**). 5 embryos were used for each cultivation condition.

Eggs were then drawn into the end of a capillary tube which were left open to the air and maintained on a support over a layer of moistened paper towel in a closed Petri dish as published in Dickinson & Croll, (2001). These moist chambers were incubated at 28 °C, which is an optimal temperature for *B. glabrata* development as described in Joubert & Pretorius, 1985. In all the experiments, around 30% of the embryos died after 2 hours and were assumed to be injured during the extraction or transfer and were discarded. Moreover, despite all sterilizing steps and clean work, about 20% of all capillaries were obviously infected by bacterial, fungal, or protist colonies and were also excluded from the experiments and thus from the statistics. As a control for the development of individual cultured embryos and thus the timepoint when it is appropriate to stop cultivation and release potentially cultured juveniles into the water, a single egg mass was sterilized in the same way, but the embryos were not decapsulated. This egg mass was placed in a beaker with snail water and incubated at 28 °C simultaneously with the embryo in the capillaries. The viability of the embryos was assessed regularly under the microscope. Statistical relationship between concentration of egg fluid in the media and the survival of the embryos was assesed using regression statistical analysis with MS Excel Analysis ToolPak.

Table 2. Composition of culture media used for *in vitro* culturing of *B. glabrata* embryos.
The medium was always a mixture of egg liquid, Bge medium and distilled water in different ratios.

	Culture media composition		
No.	Egg fluid %	Bge media	Distilled water
1.	100	0	0
2.	75	25	0
3.	75	0	25
4.	50	50	0
5.	50	0	50
6.	25	75	0
7.	25	0	75
8.	0	100	0

2.2 Computational analysis of the *diaphanous (Dia)* gene in Gastropoda as a potential gene for targeting CRISPR/Cas9 in *B. glabrata*

As part of the development of the CRISPR/Cas9 protocol in *B. glabrata*, it is necessary to select a gene whose alteration or knock out will not be lethal to the organism but at the same time will lead to a clear phenotype to help evaluate the success of the experiment. Given the recent successful knockout of the shell-coiling gene in the great pond snail *Lymnaea stagnalis* (Abe & Kuroda, 2019), we opted to work on the same *Diaphanous* gene. *L. stagnalis* has two copies of this gene, LsDia1 and LsDia2, and knockout of only LsDia1 leads to a change in shell orientation from naturally dextral to sinistral. If *B. glabrata* also had both of these genes, which would be true orthologs to those of *L. stagnalis*, it could be assumed that they could play similar roles in our model organism. This situation was even more interesting, given that *B. glabrata* is naturally sinistral in a wild-type individual.

The nucleotide sequences of *L. stagnalis* Dia genes 1 and 2 were obtained from GenBank (accession numbers KX387869, KX387871). These sequences were then used as a query for a BLAST search against *Biomphalaria*. The most recent assembly of the *B. glabrata* strain BB02 genome was obtained from VectorBase (Giraldo-Calderón et al., 2015), (GenBank accession GCA_000457365.1) and used as a database for BLAST searches. The resulting alignments were visualised in IGV (Robinson et al., 2011) and assessed. After discovering that naturally sinistral *B. glabrata* has only one copy of the *dia* gene, subsequent analysis was performed to determine whether duplication of this gene occurred in the lineage leading to *L. stagnalis* or whether the gene was secondarily lost in *B. glabrata* and this might have possibly led to change in the shell orientation. Transcriptomes of 40 gastropod species generated as a part of phylogenomic study analysing deep gastropod relationships (Zapata et al., 2014) (BioProject: PRJNA253054) were used for *Dia* gene duplication analysis (**Table 3.**).

First, individual RNAseq datasets were downloaded using SRA-tools (fastq-dump) (Leinonen et al., 2011) and their headers formatted according to library layout of the individual datasets (3 single end versus 37 paired end). The individual reads from each dataset were then trimmed with the Trimmomatic tool (Bolger et al., 2014), and the transcriptomes assembled using Trinity (Grabherr et al., 2011). The assembled Gastropoda transcriptomes were then used one by one as databases for a DIAMOND search. DIAMOND (Buchfink et al., 2015) allows blasting translated nucleotide sequence against a broader taxonomic group and millions of proteins and is thousands of times faster than BLASTP and BLASTX (Camacho et al., 2009) while maintaining a high degree of sensitivity. The identity threshold and e-value parameters of DIAMOND were tested and finally, identity of 70% and e-value of 1E-30 were chosen for the final search. *Dia* genes from *L. stagnalis* were then used as a query. Sequences of all hits output by DIAMOND were then retrieved from the database with seqtk software (https://github.com/lh3/seqtk). Discovered genes were counted and assigned to the appropriate species. The orientation of the shell of all gastropods that were part of this analysis was determined on the basis of the literary research. These data were then confronted with the existing phylogenetic tree of Gastropoda (Zapata et al., 2014.) and the conclusions were drawn.

Table 3. SRA datasets of 40 gastropod species used for the *Dia* genes analysis. For some species it was not possible to trace the part from which the RNA was obtained.

Organism	Run	Bases	Library layout	Tissue
Amphiplica gordensis	SRR1505101	13.78 G	PAIRED	three whole animals
Architectonica perspectiva	SRR1505102	3.49 G	PAIRED	foot
Astarte sulcata	SRR1505103	7.61 G	PAIRED	
Prodoris clavigera	SRR1505104	2.58 G	PAIRED	foot
Chaetoderma sp.	SRR1505105	1.62 G	PAIRED	whole animal
Clione limacina anatarctica	SRR1505107	4.06 G	PAIRED	two whole animals
Doris kerguelenensis	SRR1505108	2.66 G	PAIRED	foot
Fiona pinnata	SRR1505109	3.29 G	PAIRED	whole animal
Granata imbricata	SRR1505110	6.09 G	PAIRED	foot
Haminoea antillarum	SRR1505111	2.64 G	PAIRED	whole animal
Hinea brasiliana	SRR1505112	1.03 G	SINGLE	
Hydatina physis	SRR1505113	3.85 G	PAIRED	foot
Janthina janthina	SRR1505114	3.47 G	PAIRED	foot
Laevipilina hyalina	SRR1505115	5.40 G	PAIRED	three whole animals
Lepetodrilus fucensis	SRR1505116	3.63 G	PAIRED	gills
Mercenaria campechiensis	SRR1505117	3.44 G	PAIRED	
Microhedyle glandulifera	SRR1505118	1.36 G	PAIRED	180 whole animals
Monodonta labio	SRR1505119	3.43 G	PAIRED	most of one animal
Myochama anomioides	SRR1505120	7.78 G	PAIRED	
Neotrigonia margaritacea	SRR1505121	4.86 G	PAIRED	
Nerita peloronta	SRR1505122	3.66 G	PAIRED	foot
Onchidella floridana	SRR1505123	3.27 G	PAIRED	most of one animal
Ophicardelus sulcatus	SRR1505124	3.21 G	PAIRED	most of one animal
Oxynoe viridis	SRR1505125	4.00 G	PAIRED	foot
Paralepetopsis sp.	SRR1505126	379.25 M	PAIRED	
Phallomedusa solida	SRR1505127	5.14 G	PAIRED	most of one animal
Phasianella ventricosa	SRR1505128	5.39 G	PAIRED	most of one animal
Philine angasi	SRR1505129	2.01 G	SINGLE	foot
Pleurobranchaea californica	SRR1505130	4.87 G	PAIRED	
Euspira heros	SRR1505131	3.45 G	PAIRED	foot
Pomacea diffusa	SRR1505132	4.98 G	PAIRED	foot
Prothalotia lehmanni	SRR1505133	5.28 G	PAIRED	most of one animal
Pyropelta sp.	SRR1505134	3.45 G	PAIRED	
Rissoella caribaea	SRR1505135	4.96 G	PAIRED	whole animals
Rubyspira osteovora	SRR1505136	3.29 G	PAIRED	
Strubellia wawrai	SRR1505137	3.02 G	SINGLE	two whole animals
Titiscania limacina	SRR1505138	4.95 G	PAIRED	most of one animal
Turbonilla sp.	SRR1505139	5.86 G	PAIRED	most of one animal
Tylodina fungina	SRR1505140	3.53 G	PAIRED	foot
Urosalpinx cinerea	SRR1505141	3.72 G	PAIRED	

2.3 Probing RNAseq datasets for hidden and yet undiscovered parasite species

In order to discover new relationships between helminth parasites and their gastropod hosts, to discover hidden parasites and also to discover new parasitic species, the main goal of this sub-project was to develop a bioinformatics programme capable of identifying whether we could identify DNA from parasites (specifically their mitochondrial Cytochrome oxidase 1 gene) within the RNA extracted from a series of potential host species. We looked for parasite-derived COX1 sequences (from a large data base of metazoan COX1 sequences) in all species of gastropods present in publicly available RNA datasets. In short, we used MAGIC-BLAST (Boratyn et al., 2018) to search all snail RNAseq datasets on NCBI SRA Archive for the presence of all nematode and flatworm COX1 sequences present in the BOLD barcoding database.

This was to be achieved in several clear steps. First, it is necessary to build the largest possible database of publicly available sequences of selected barcoding genes, to format this database appropriately and to get rid of redundant records. Then, an extensive and complete list of SRA accession numbers of all gastropods present in the NCBI SRA archive needs to be obtained. After selecting the appropriate software, the next step is to use the prepared database to quickly map the relevant reads obtained from these individual SRA archives. The mapped reads must then be assembled into longest sequences possible, the correct reading frames selected, and these sequences translated. These assembled sequences can then be identified using publicly available tools and matched back to the datasets from which they originated using accession numbers. A great emphasis in the development of such a tool lies on time efficiency making the tool convenient and ideal sensitivity, which not only allows to identify already known species, but also to discover new species that are not present in the provided database. This computational project was undertaken when lab work was impossible due to Covid-19 restrictions.

2.3.1 Data accessibility

SRA identifiers of analysed RNAseq datasets were obtained from NCBI Sequence Read Archive (SRA). For initial analyses, the smallest molluscan dataset (ERR4193411 - RNAseq of oyster *Crassostrea gigas* gills) was selected to assure time effective testing of the tool. For assessing the effect of taxonomic resolution of the database analyses, the dataset with known parasitic infection was selected (planorbid snail *Biomphalaria glabrata* infected with trematode *Schistosoma mansoni* – BioProject: PRJNA383396, sample SRX2744891). The list of accession numbers of SRA datasets for the main screen of all gastropod RNAseq datasets was obtained with advanced filtering methods of NCBI SRA database using search parameters “Gastropoda”, “paired-end” and “RNAseq”. This procedure generated 2150 accession numbers of the gastropod transcriptomes.

2.3.2 Barcoding databases

Because Cytochrome oxidase I mitochondrial gene (COX1) is a conserved gene across all cestodes, trematodes and nematodes as described in chapter **1.1.2** and has been extensively sequenced for use in barcoding, it was selected as a gene for parasite species identification. For script optimisation experiments, MIDORI reference 2 database (Machida et al., 2017) was used and pruned to obtain only lophotrochozoan COX1 sequences. The header of each sequence in

FASTA format, originally containing all taxonomic units for particular species, was reformatted with a custom Python script so that the header contained only selected taxonomic unit of choice such as phylum, family, genus etc. For the rest of the experiments including assessing the effect of database taxonomic resolution and the final big analysis of Gastropoda SRAs, the lophotrochozoan MIDORI 2 database was filtered to contain Mollusca, Platyhelminthes and Nematoda sequences only. This reduced database was merged with the database of available COX1 sequences of the Mollusca, Platyhelminthes and Nematoda groups obtained from BOLD (Barcode Of Life Data), from which COX2 and rRNA genes were removed and formatted so that its headers corresponded to the format of the existing MIDORI 2 database and thus functioned as input database for the script. Finally, after merging the databases, this file was filtered using the CD-hit (Fu et al., 2012) programme with sequence identity threshold set to 1 removing duplicates from the database.

2.3.3 Computational analyses for identifying specific sequences in the sample

The process of identifying specific COX1 sequences in the SRA datasets is automated with the custom Python script using multiple publicly available command line tools.
Firstly, the script creates a BLAST nucleotide database from the provided FASTA COX1 database file and builds a dictionary of taxonomic groups present in the database to which individual mapping reads are later assigned.

MAGIC-BLAST then uses the publicly available SRA dataset from the provided list of SRA accession numbers and rapidly maps the reads to the BLAST reference database without the necessity of manual download of the individual SRA datasets. The sensitivity of this tool can be adjusted with various options having impact on the mapping specificity.

Mapped reads are then counted, assigned to the taxonomic groups stored in the dictionary, assembled by MEGAHIT (Li et al., 2015) to contigs based on overlapping regions and assigned to the taxonomic group that the reads mapped to.

Finally, Transdecoder (http://transdecoder.github.io) predicts the open reading frames (ORFs) and coding regions and translates the sequences.

This step was carried out for the purpose of potential downstream phylogenetic analysis of the discovered unknown species. When the run is finished, the tool automatically continues with analysis of the next SRA dataset in the provided list. At the beginning of each run, an appropriate folder is created in which all the resulting sequences and data from individual run are stored, including the SAM file with all mapping reads, potentially useful for advanced analysis. This folder also contains log files informing about the progress of the individual analysis steps useful for potential troubleshooting. If this extra data is not needed, it is possible to save disk space and "clean up" the folders using the optional clean argument. After completing the analysis of all SRA datasets from the list, COX1 sequences assembled in this way are then identified BOLD identification tool (https://github.com/linzhi2013/bold_identification). Schematic representation of the pipeline is shown in **Figure 10.**. The tool has four required arguments and six optional arguments (**Table 4.**).

Table 4. Table of required and optional input arguments for the described tool.

Input arguments for the tool	
Required arguments:	
-sra	The list of SRA accession numbers
-fi	input database of COI genes in FASTA format
-tpm	Transcripts per million threshold
-n	Number of CPU threads used for the analyses
Optional arguments:	
-O	Output folder name
-t	Timing of individual steps of the analysis
-e	"gap extend" parameter of MAGIC BLAST
-o	"gap open" parameter of MAGIC BLAST
-p	"penalty" parameter of MAGIC BLAST
-ws	"word size" parameter of MAGIC BLAST

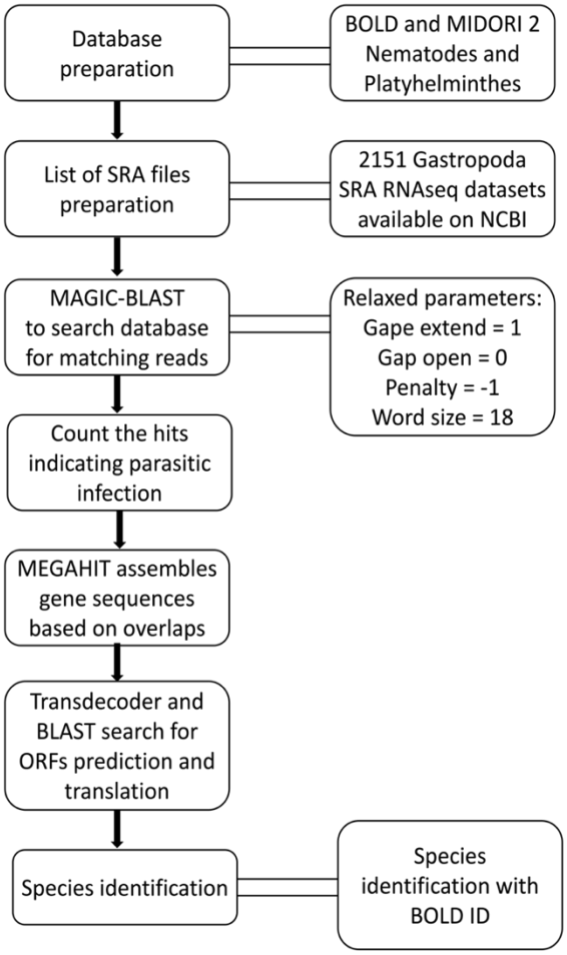

Figure 10. Flowchart showing the analysis pipeline for searching and identifying COX1 sequences in SRA datasets. General description of the individual steps is on the left side and specific modifications and details used for the actual final analysis on the right.

2.3.4 Altering MAGIC-BLAST parameters to evaluate sensitivity and time-effectivity of the tool

The sensitivity and the specificity of the analysis relies on the MAGIC-BLAST mapping tool. We wanted to establish the set of parameters for this step that produced the fastest and most accurate search. MAGIC-BLAST computes a local alignment by extending exact word matches (18-bases **word size** by default) between a read and a reference sequence. It computes an alignment score using the following system: 1 for each matching pair of bases, −4 for a base substitution (**penalty**), 0 for **gap opening** (either a read or reference sequence), and −4 for each base of **gap extension** (insertion or deletion). All these four parameters can be relaxed or tightened. To assess the effect of these parameters on the analysis, the iterations were performed over these values (default in **bold**):
word size {12, 13, 14, 15, 16, 17, **18**}
gap extension {**-4**, -3, -2, -1, 0}
gap open {**0**, 1, 2 , 3, 4}
penalty for mismatch {**-4**, -3, -2, -1}

Firstly, iterations were performed so that only one parameter was changed, while the rest of the parameters remained at the default values. Secondly, each combination of values of two parameters was subjected to analysis. The numbers of hits mapping to different taxonomic phyla belonging to Lophotrochozoa were counted and compared to the analysis with default parameters. In the case of analyses where two parameters were changed simultaneously, reads mapping to the taxonomic unit of the host organism were counted, then multiplied by 100 and divided by total number of mapping reads. This way the ratio of percentage of the reads mapping to the host taxonomic unit vs those mapping to other groups was calculated. In this way it was possible to project the results as a levelplot in R (version 3.6.2) and thus avoid impossibly large and complex heatmaps that were only useful in a single parameter analyses.

Analysis time was also measured for all individual simulations. Results of the simulations were plotted in R. This way we were able to evaluate which combination of parameters is relaxed enough to allow MAGIC-BLAST to map reads, which don't match with the COX1 sequences in the provided database perfectly, but at the same time don't slow down the analysis.

2.3.5 Effect of the absence of different taxonomic units in the database

We were interested in whether we would be able to detect a species that does not have a very close match in the database (e.g. a new species). To determine the effect of the absence of target COX1 sequences in the database on the resulting assembled COX1 sequences of the present organisms in the analysed samples, we assessed the ability of the tool to retrieve the COX1 sequences using incomplete database by removing different taxonomic units from the database when testing the sample. This analysis was performed on aforementioned SRA dataset (BioProject: PRJNA383396, sample SRX2744891) containing *B. pfeifferi* infected with *S. mansoni* to study the potential effect on both parasite and host retrieved sequences. The headers of the COX1 sequences in a database were formatted by custom script to contain all levels of taxonomic units they belong to and filtered with BBtools (https://sourceforge.net/projects/bbmap/) to exclude species (*B. pfeifferi*), genus (*Biomphalaria*) and finally even family (Planorbidae). The analyses were then performed with

relaxed MAGIC-BLAST parameters as described above and the resulting sequences were identified with the BOLD identification tool and verified with online BLAST tool against "standard databases" (Altschul et al., 1990).

2.3.6 Analysis of all accessible NCBI SRA Gastropoda datasets with relaxed parameters

After completing the simulations and selecting appropriate MAGIC-BLAST parameters for time-efficient but specific enough analysis and evaluating the appropriate resolution of the input database, all available gastropod transcriptomes on the SRA database were analysed for the presence of all Platyhelminthes and Nematoda COX1 sequences. The list of 2150 SRA accession numbers was divided into 22 lists of 100 and the analysis was independently run on each of these subsets separately to better distribute the performance of the computer, monitor the progress of the analysis, mitigate the effects of the possible crash of the analysis and then easier data sorting. At the end of the analysis, the potential parasite nucleotide sequences obtained from the individual snail SRA datasets were pooled into one FASTA file and identified using the BOLD identification tool. The resulting table was then processed in Microsoft Excel, filtered for non-matching sequences and manually evaluated.

3 Results

3.1 *In vitro* culturing of embryos of *Biomphalaria glabrata*

Embryos were cultured for 13 days, which was the time period after which control embryos cultured in the original egg mass spontaneously hatched. The survival rate of the embryos was strongly dependent on the different concentration of egg fluid in the media. From the statistical analysis (**Table 6**. and **Figure 11.**) it is clear, that the higher the concentration of egg fluid in the culture medium, the more viable the cultured embryos were (p value < 0.002). On day 9 of culture, out of a total of 40 embryos, only three survived, two cultured in pure egg fluid and one in a combination of 75% egg fluid and 25% Bge medium. Only one embryo, cultured in 100% egg fluid was able to survive the entire culture period and develop into a juvenile gastropod (Error! Reference source not found.**.** and **Figure 13.**). Due to the hatching of control embryos on day 13 of culture and the stage of development of this cultured embryo in the capillary, it was released into the snail water and fed frozen spinach. this juvenile seemed viable after release, crawling on the bottom of the beaker and actively interested in the offered food. Unfortunately, this cultured juvenile died the day after release (**Figure 12.).**

Table 5. Days of death of *B. glabrata* decapsulated embryoes under various culturing conditions. The composition of individual media 1-8 can be found in a **Table 2.**, chapter **2.1**. The total of 5 embryos were cultured in each media for 13 days with viability assessment every second day of the culture. While in medium 1, which represented pure egg fluid, one embryo was able to survive until the end of culture, with decreasing concentration of this fluid the survival rate decreased already in the early stages of culture. In pure Bge medium, represented by number 8, one embryo was able to survive until day 9 of culture, but did not show signs of normal development during the culture

			Day of death						
Media	Egg fluid %	N	Day 1	Day 3	Day 5	Day 7	Day 9	Day 11	Day 13
1	100	5			X		XX	X	X
2	75	5		X	X	X	X		X
3	75	5		X		XX	X		X
4	50	5	XX	X			XX		
5	50	5	XX	X	X	X			
6	25	5	XXX		X		X		
7	25	5	XXX	X	X				
8	0	5	XX		XX			X	

Table 6. Regression analysis of concentration of the egg fluid in the media and day of death of cultured individuals. There is a moderate relationship between egg fluid and time to death. As egg fluid in the media increases, the lifespan increases (p-value <0.002).

Regression Statistics	
Multiple R	0,49009053
R Square	0,24018873
Adjusted R Square	0,21965329
Standard Error	3,39548715
Observations	39

ANOVA	df	SS	MS	F	Significance F
Regression	1	134,850576	134,850576	11,6963034	0,00154004
Residual	37	426,585321	11,529333		
Total	38	561,435897			

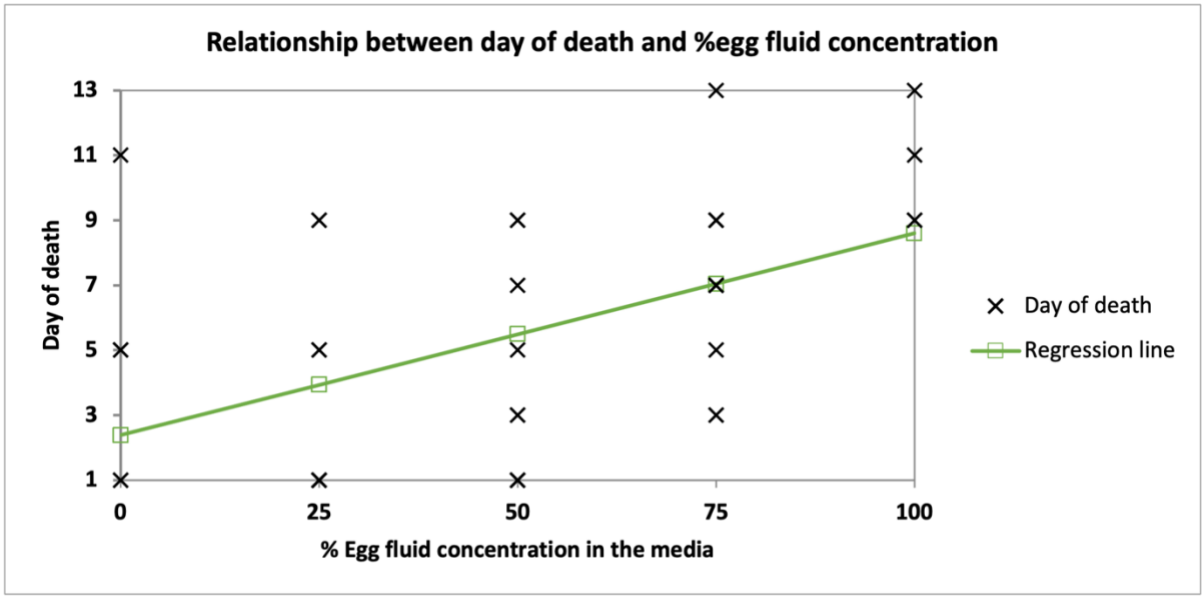

Figure 11. Regression analysis of concentration of the egg fluid in the media and day of death of cultured individuals. The regression line estimates, that the higher the concentration of the egg fluid in the media, the longer the embryos survive.

Although the relationship between concentration of the egg fluid in the culturing media and the survival of the embryos in our analysis is statistically significant, we are well aware that only a very small number of conditions have been tried and only a few dozen embryos have been tested. This is primarily due to the fact that obtaining egg fluid from the eggs of *B. glabrata* was extremely challenging due to their significantly smaller size compared to the snails that this method was successfully established on. Already after pilot experiments, where the culture medium did not contain egg fluid, it became evident that the snail embryos do not survive without it, which is consistent with the available literature, and I will deal with this problem in more detail in the discussion.

We noted one very interesting phenomenon that occurred in the cultivation of embryos in pure Bge medium. While 4 out of 5 embryos did not survive the fifth day of culture, one of the embryos survived until the ninth day of culture and went through a very bizarre development, which, due to unfortunate circumstances we unfortunately do not have better documented (**Figure 12.**). On the fifth day of cultivation the embryo had significantly increased in volume, but it had no obvious shape or recognizable signs of the normal development of this species of mollusc. Due to its size, it began to outgrow the diameter of the capillary and had to be moved to a 96-well plate, which did not pose a major problem due to the simplicity of the culture medium. The embryo was then monitored until day 9 of culture when it had unseen characteristics. This embryo had developed one eye spot and was covered with motile cilia all over its surface. There were various lobes protruding from its body with no apparent regularity. As some kind of "seeking" behaviour was observed, where this life form moved slowly around the area of the well, we decided to offer it food. After the addition of sterilized aquarium snail food, the embryo actively moved towards the food, which it visibly processed with increased movement of the cilia and overall increased activity. After this "feeding", the embryo was transferred to pure fresh Bge medium and further incubated. On day 11 of culture, unfortunately, the embryo did not show any signs of life, but due to the unusual nature of this event, it was left in the medium until day 13, when it became clear that it had died.

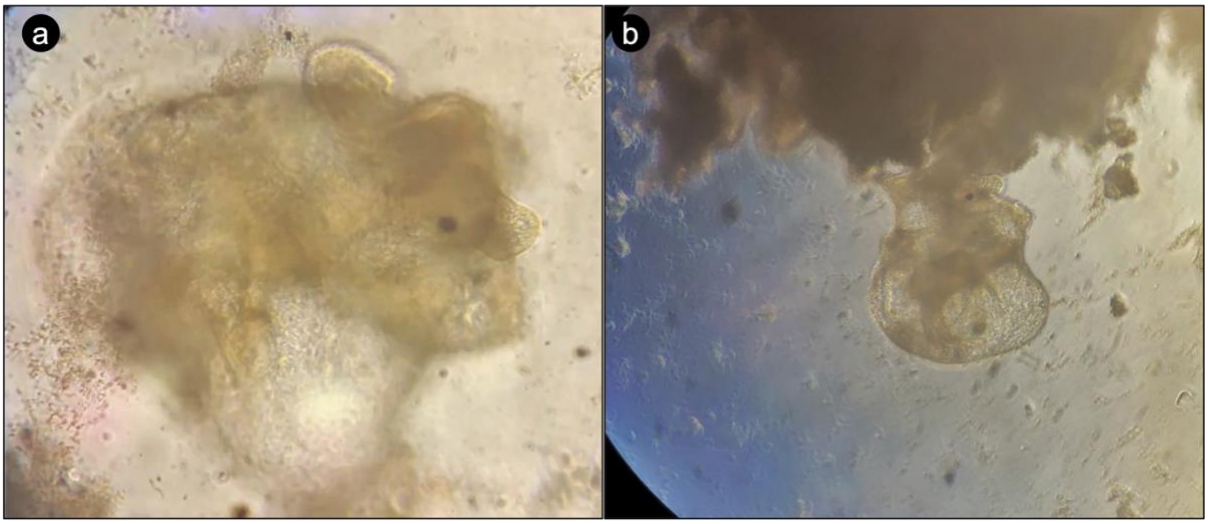

Figure 12. Deformed embryo of *B. glabrata* cultured in pure Bge medium. a) Deformed embryo on the day 9 of the culture. **b)** Deformed embryo "feeding" on a provided food on day 11.

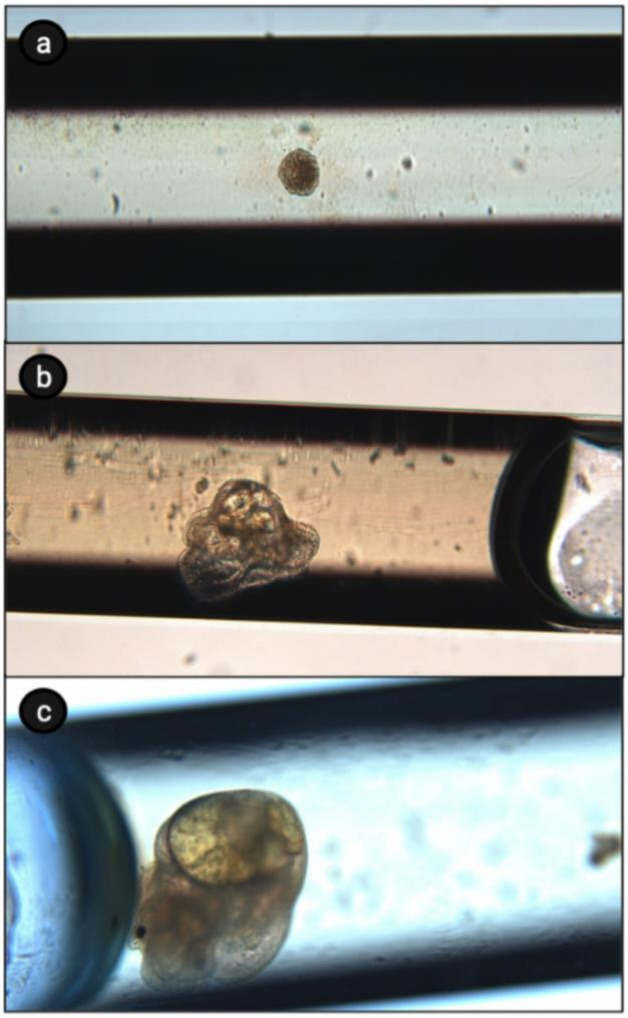

Figure 13. ***In vitro*** **culture of embryos of** ***B. glabrata*** **in glass capillary tubes filled with egg fluid in different developmental phases. a)** day 3 of cultivation **b)** day 5 of cultivation **c)** day 11 of cultivation

3.2 Computational analysis of the *Dia* gene in Gastropoda as a potential gene for targeting CRISPR/Cas9 in *B. glabrata*

Sequences for *Dia* genes were looked for in assembled RNAseq datasets as described in chapter **2.2** and **Table 3**. and were found in 24 species from the total of 40. Two *dia* genes seem to be expressed in only six species of all gastropods tested – *Ophicardelus sulcatus*, *Turbonila* sp., *Haminoea antilarum*, *Doris kerguelenensis*, *Lymnaea stagnalis*, and *Monodonta labio*. It is not clear from the phylogenetic tree with assigned numbers of *dia* genes discovered in each species and shell orientation information whether there is any relationship between these two factors. The presence of only one gene in i.e. dextral *Phallomedusa sollida*, closely related species to *B. glabrata*, might indicate that this relationship does not exist, since both have just one copy of the gene, but have different shell orientation. However, because *dia* genes were identified on the basis of transcriptomic data and not genomic data, it is difficult to draw any conclusions from these results, because many genes that may be present in these gastropods may have been unidentified in this analysis.

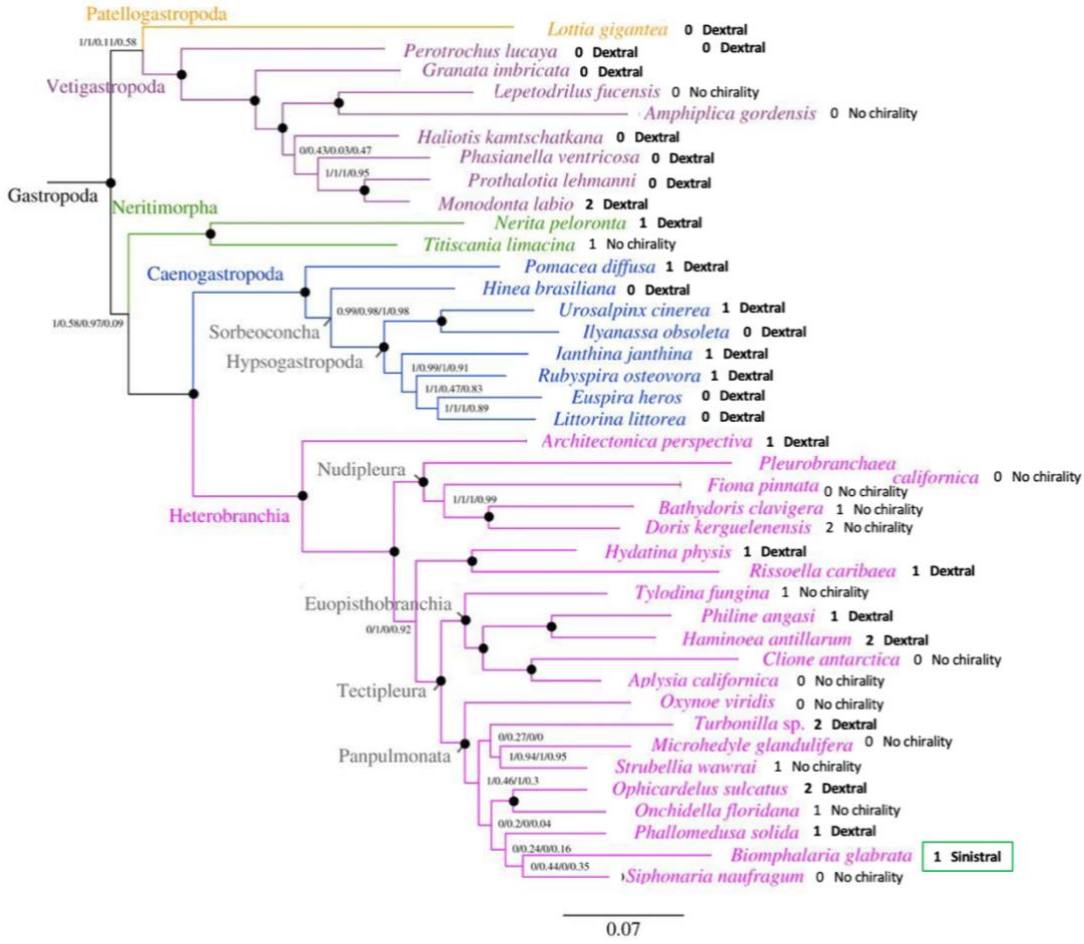

Figure 14. Modified rooted phylogram of Gastropoda clade (Zapata et al., 2014) showing the number of *Dia* genes and shell orientation for each individual species.

3.3 Probing RNAseq datasets for hidden or as yet undiscovered parasite species

With clearly defined steps that needed to be taken in order to identify selected COX1 sequences present in the provided SRA datasets and thus discover possibly new relationships between helminth parasites and their gastropod hosts, to discover hidden parasites but also to discover new parasitic species, we developed a relatively simple Python-based bioinformatic tool. We have carefully tested and optimized individual steps of this tool to create a programme both time efficient enough to analyse massive amounts of data in a reasonable time, and specific enough to reliably identify the species present in the provided RNAseq datasets. In addition to testing the general functionality of the tool, we primarily tested the effect of individual adjustable parameters on the overall time and sensitivity of the analysis and the effect of the completeness of the database on the results of this analysis. After this optimization, we used the program to identify all helminths from the Platyhelminthes and Nematoda phyla present in all available SRA datasets obtained from gastropods. We critically reviewed the results of this analysis manually and drew conclusions, which are then discussed in the Discussion of this thesis.

3.3.1 Altering MAGIC-BLAST parameters – time effectiveness

In order to optimise the pipeline, we wanted to know which parameters were most important to get an accurate result and which might be optimised to speed up the analysis. The stages of the pipeline we tested were MAGIC-BLAST running time, Transdecoder running time and then Total running time to see, if there are any other parts of the analysis, that may be affected by the changes in MAGIC-BLAST parameters. By iteration over all values of each parameter, which MAGIC-BLAST allows and measuring the effect of these changes on the analysis time showed that no change in the parameters affects the translation time. Even though it may seem so between gap_open 0 and gap_open 1, after a more detailed examination it was found that this effect is caused by the current workload of the computer and other tasks on which our analysis was running and thus there was a temporary drop in performance and thus a prolongation of the specific analysis. On the other hand, the simulations showed that if the value of the **word_size** parameter is decreased, the analysis time increases significantly up to a factor of five in case of the lowest value of word_size 12. The total analysis time and its possible increase then always corresponded to the extension of the analysis due to the longer MAGIC-BLAST step, which means that the other steps within the script were not affected by the any parameter change (**Figure 15.**).

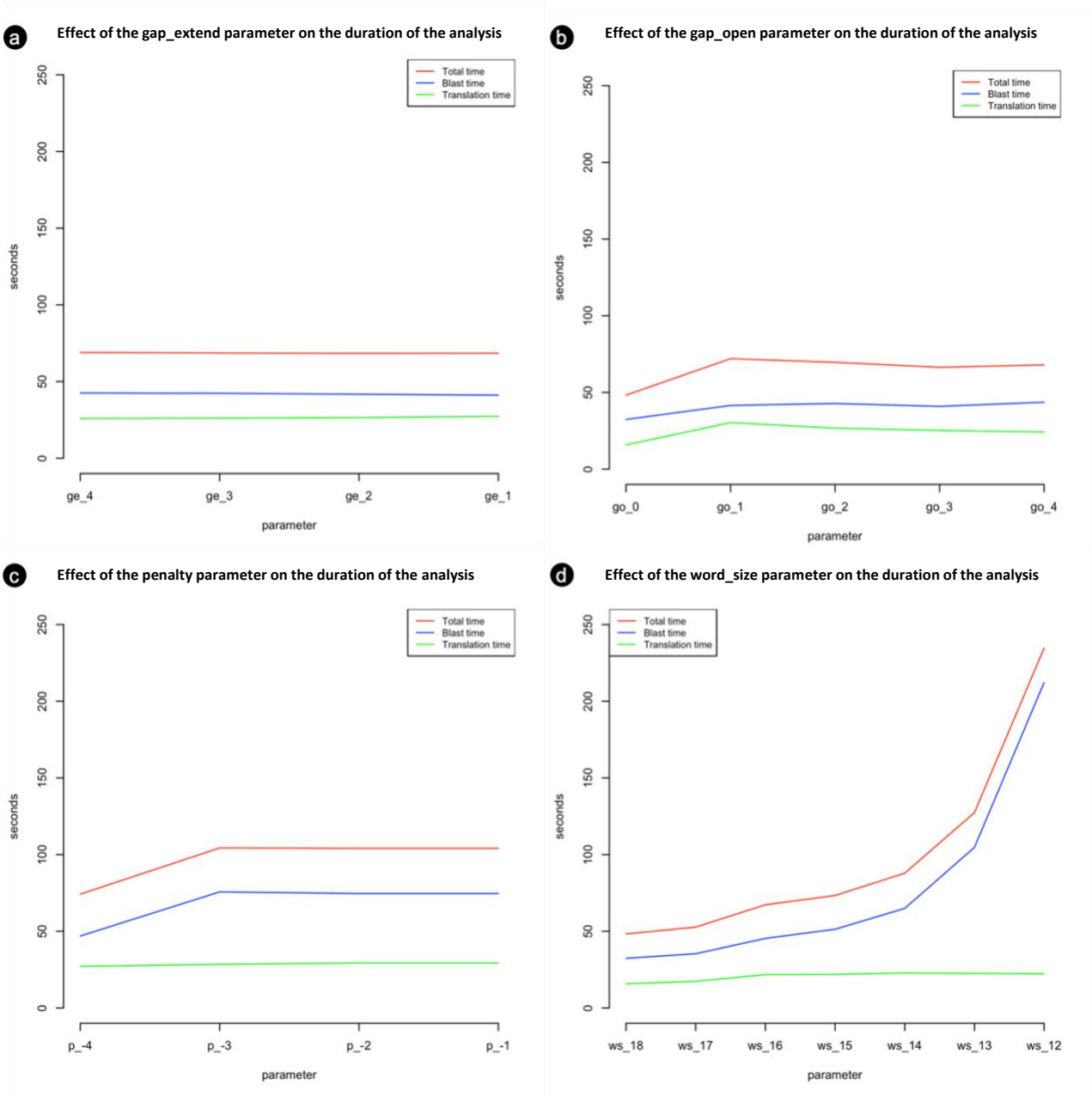

Figure 15. The effect of relaxing the parameters on duration of the analysis. While **a)** gap_extend (ge), **b)** gap_open (go) and **c)** penalty (p) have almost no effect on time running of the analyses, **d)** decreasing the word size (ws) significantly increases the duration of the MAGIC-BLAST mapping and thus the whole analysis.

3.3.2 Altering MAGIC-BLAST parameters – specificity

The effect of changing one MAGIC-BLAST parameter was also evaluated with respect to the specificity of the analysis, i.e. how many reads are mapped to individual taxa in particular setting. While changing **word_size**, **gap_open** and **gap_extend** did not have any effect on the read mapping, decreasing the value of the **penalty** parameter not only significantly increases number of hits mapping to the Mollusca taxon (which the testing oyster sample belongs to) but also causes incorrect mapping of the reads to Annelida (**Figure 16.**).

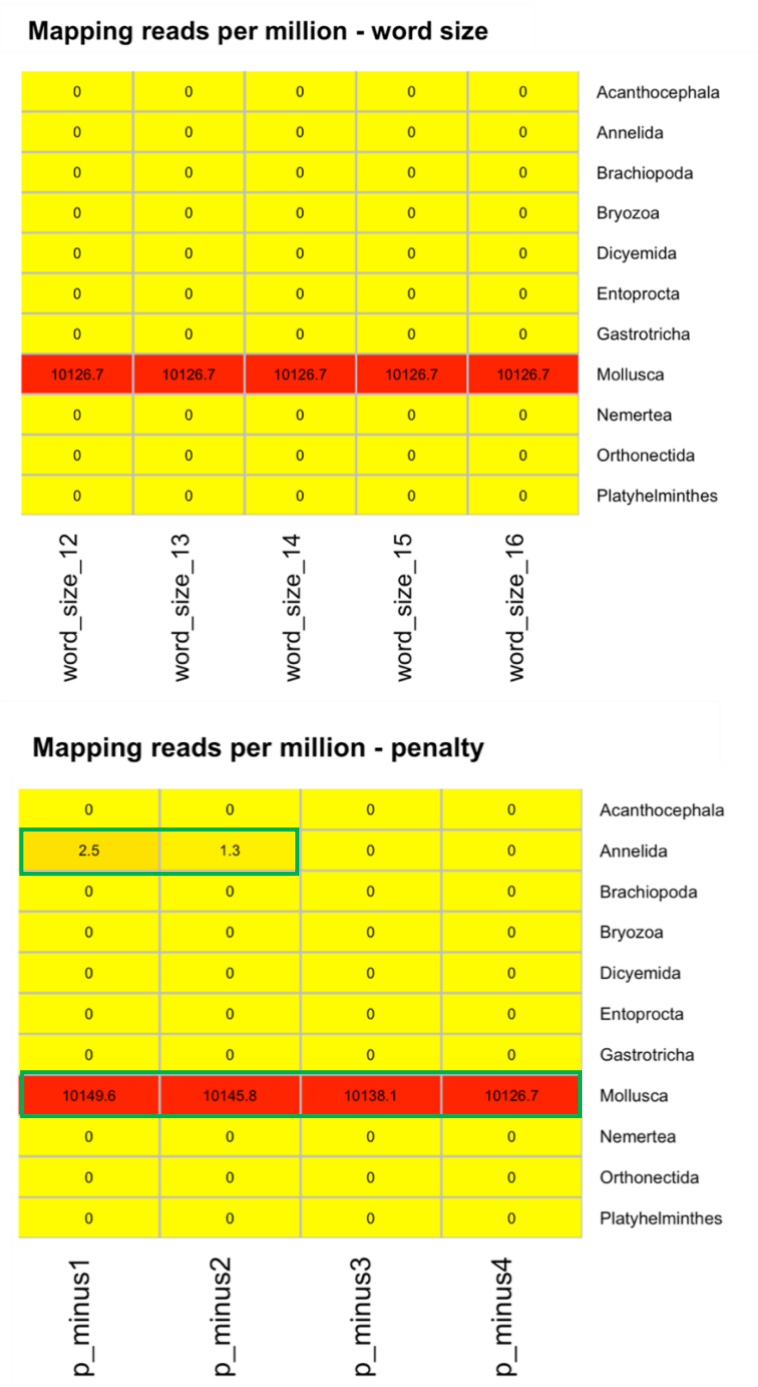

Figure 16. Number of the reads mapping to different phyla under various parameters. Changing value of word_size has no effect, while decreasing the penalty value leads to increasing number of reads mapping to Mollusca itself, but also different phyla.

All six combinations of changing two parameters at once -{go,ge} {go,ws} {go,p} {ge,ws} {ge,p} {ws,p}- and evaluation of their specificity using the ratio of the total number of hits mapping to the database to the number of hits mapping to a specific target organism (*Crassostrea gigas*) showed that the only combination that has a specificity effect is the penalty in combination with gap_extend or in combination with word_size. Since the time simulations showed that decreasing the value of word_size is inappropriate because of the significant increase in analysis time, the combination of the lowest **penalty** with **gap_extend** was chosen as the "relaxed parameters" for further experiments **(Figure 17.).**

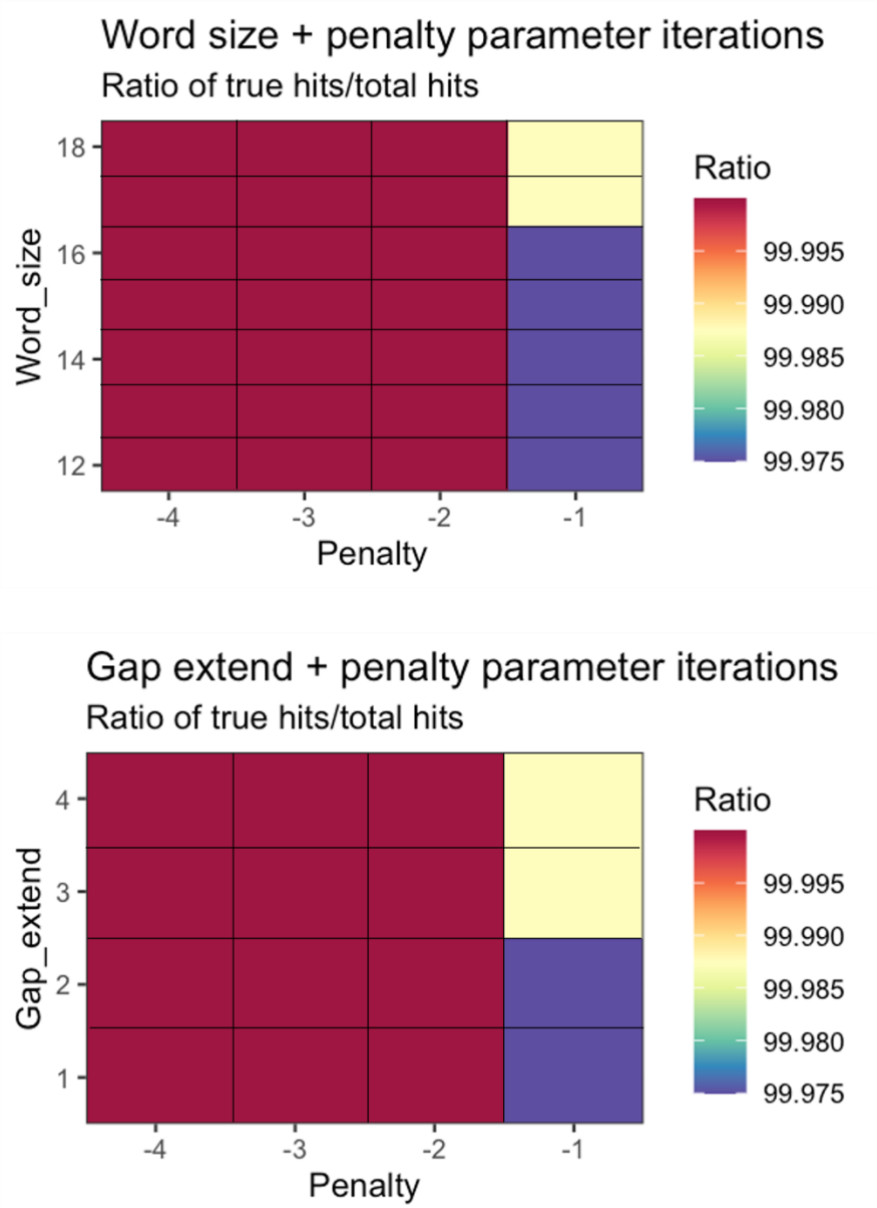

Figure 17. Changing two parameters of MAGIC-BLAST at once in relation to ratio of "true reads" mapping to specific target testing organism and total mapping reads. The lowest values of word_size or gap_extend in combination with penalty leads to the most reads mapping to different species in the provided database.

3.3.3 Effect of the absence of different taxonomic units in the database

We wanted to see if enough reads would still be mapped to allow the tool still to assemble a COX1 sequence that is not known to us (if there is a species present not represented the BOLD database) and thus identify an unknown species present in a SRA dataset. MAGIC-BLAST uses the database we provide for mapping of the reads. We wanted to find out what happens if COX1 sequence of the species or its higher taxonomic group is not present in this database. To test this, we removed individual COX1 sequences of species, genus and finally the whole family of the species present in the sample from the database and assessed how this absence affects the results of the analysis and the assembled COX1 sequences.

Experiments showed, that while the tool assembled the whole COX1 sequences of the host and the parasite when provided the complete database, both host and, surprisingly, also the parasite sequences were affected by removal of the COX1 sequences of the host species, genus and family.

Removing a COX1 sequence of a specific species and genus from the database has a strong effect on the resulting sequence using both the default and the relaxed parameters. In the absence of species sequence, the analysis using default MAGIC-BLAST parameters results in a significant truncation of the resulting COX1 sequence which paradoxically led to more accurate sequence similarity match in BOLD identification (see Discussion. In case of removing the COX1 sequences of the whole genus, the truncation of the resulting sequence is even more pronounced. However, in neither case did these changes prevent correct species identification. Analysis using relaxed MAGIC-BLAST parameters performed better, with both the species and genus removal resulting in a longer assembled COX1 sequence than in case of the analysis using default settings. When removing the entire family of the host, the tool was unable to assemble the sequence in either case (see Discussion).

However, the actual effect of removing host taxa sequences from the database on the resulting parasite sequence was surprising. Also, in this case the analysis using relaxed MAGIC-BLAST parameters worked significantly better, where the resulting COX1 sequence was always longer. All results shown in **Table 7**..

Table 7. The effect on assembled sequences of the host and the parasite of removing COX1 sequences of the host different taxonomic units from the database. Relaxed parameters of MAGIC-BLAST allowed more accurate and longer assemblies in comparison to the default settings.

Taxa removal experiments		default MAGIC-BLAST parameters				relaxed MAGIC-BLAST parameters			
		Identity %	Alignment length	Query cover %	BOLD Identification %	Identity %	Alignment length	Query cover %	BOLD Identification%
Molluscan assembled sequences	full database	100	1548	100	98,62	100	1548	100	98,62
	w/o species	100	534	34,5	100	100	913	59	98,62
	w/o genus	99,7	393	25,3	98,73	100	774	50	98,62
	w/o family	0	0	0	0	0	0	0	0
Playhelminthes assembled sequences	full database	100	1830	100	98,62	100	1830	100	98,62
	w/o species	96,4	1248	68,2	98,62	100	1830	100	98,62
	w/o genus	96,4	1248	68,2	98,62	96,3	1578	86,2	98,62
	w/o family	96,4	1248	68,2	98,62	96,3	1578	86,2	98,62

3.3.4 Identified host and parasite COX1 sequences in analysed Gastropoda SRA datasets

The ultimate goal of this project was, after developing and testing a tool to detect COX1 sequences in existing datasets, to use it to analyse all public Gastropoda RNA datasets and thus reveal hidden, known and unknown nematode or platyhelminth parasite species. After filtering the MIDORI 2 and BOLD databases to contain only COX1 sequences of Mollusca, Platyhelminthes and Nematoda phyla and removing duplications, the resulting database used for the final analysis counted total 132,616 records with a huge preponderance of Mollusca representatives (**Figure 18.**).

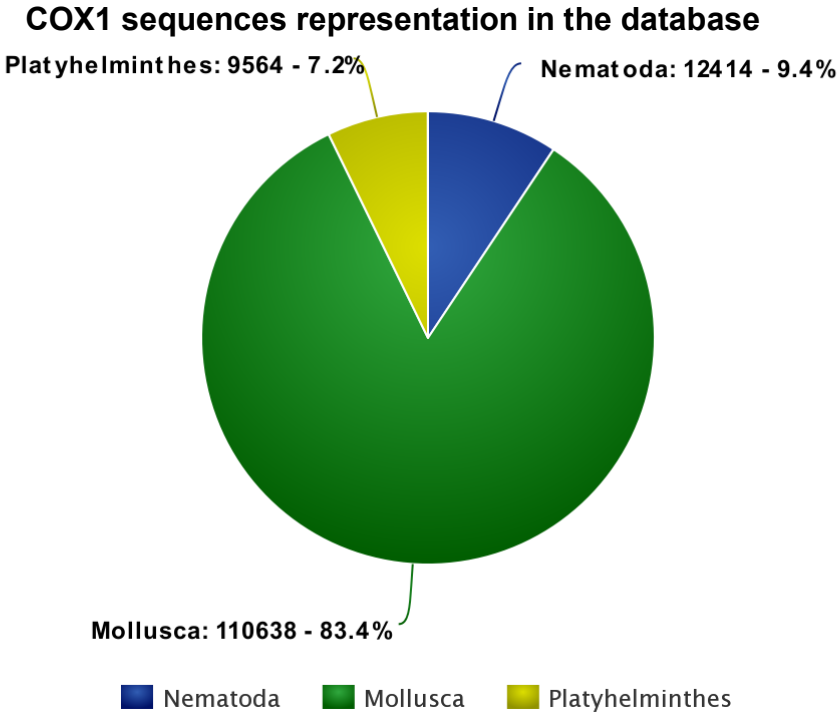

Figure 18. Proportional representation of COX1 sequences of individual target groups after merging the BOLD and MIDORI 2 databases and removing duplications. The largest part of the database is represented by species belonging to the Mollusca group. Nematoda and Platyhelminthes represented less than a tenth of the sequences represented in the database.

A total of 2150 Gastropoda RNA-seq datasets were analysed, of which all molluscan representatives were identified as expected by the tool (data not shown). These identifications were tabulated and later matched to the identified Nematoda or Platyhelminthes species using the SRA accession number under the column “host”.

In total, 67 representatives of the potentially parasitic groups were identified in the samples, of which 39 belonged to Platyhelminthes and 24 to Nematoda. Within the Platyhelminthes, as expected, flukes predominated with 37 representatives over the two tapeworms found (Table 8. Platyhelminthes species identified in 2150 Gastropoda SRA datasets from NCBI SRA archive. Similarity in % is calculated by BOLD identification tool based on Kimura 2-parameter model. The expectancy indicates whether the identified species was on purpose in the dataset or has been identified previously.**Table .**). For nematodes, then,

only two representatives were from the Enoplea group, and the rest of the species found belonged to the Rhabditida (**Table .**)

Identified Platyhelminthes

Representatives of Cestoda and Trematoda were found from the Platyhelminthes group. The species belonging to these groups are all obligately parasitic. Out of 39 identified platyhelminths, only 14 were expected to be present in a sample (i. e. the infection with the parasite was known to the researchers that generated the specific datasets). This was primarily the case for all identified schistosomes in the samples - unfortunately no schistosomes were discovered, for example, in the context of a prepatent infection in the analysed datasets of seemingly healthy snails. The closest to this was the discovery of a fluke from the genus *Calicophoron sp*. in hosts that were thought to be infected only with schistosomes, and whose infection was thus unknown to the researchers when processing the sample but was identified in the sequencing data within their publication (Buddenborg et al., 2017). Another important parasite that was discovered in the samples was *Opisthorchis viverrini* in *Bithynia siamensis goniomphalos* where its presence was known.

Another significant finding in the results is that more than two-thirds (16 out of 25) of the hidden and discovered platyhelminths came from the host sample *Batillaria attramentaria*. These were representatives of various groups of medically and veterinary important parasites (*Fasciola sp*., *Opisthorchis sp*., *Calicophoron sp*.) as well as less known and important representatives (*Cercaria batillariae*, *Pygidiopsis sp*.). It is important to mention that all datasets of this host where these flukes were found came from one Bioproject (PRJNA415312) and these results and their relevance will be discussed more thoroughly in the discussion.

Probably the most shocking result was the discovery of two tapeworms from the important genera *Hymenolepis* and *Taenia.* Both parasites were identified in marine gastropods, namely in abalone *Haliotis diversicolor* in the case of *Taenia pisiformis* and in intertidal whelk *Reisha clavigera* in the case of the tapeworm *Hymenolepis microstoma*. Considering that no case of any tapeworm using gastropods as paratenic hosts, intermediate hosts or even definitive hosts has been described so far, this is a truly unique find. Perhaps sensible explanations of these finds will be proposed again in the discussion.

A rather strange result was obtained when analysing transcriptome of the assassin snail *Anentome helena,* where COX1 sequence of lancet liver fluke *Dicrocoelium dendriticum* were detected.

The remaining unexpected parasites in this analysis are a fluke of the genus *Zoogonus* found in the Southeast Asian gastropod *Pila ampullacea*; *Telorchis sp*. a common fluke of amphibians and reptiles in the gastropod *Calliostoma zizyphinum* and the avian schistosome *Trichobillharzia stagnicolae* in the marine abalone *Haliotis discus*.

The similarity of the input sequence to the barcoding database as reported by the BOLD identification tool ranged from one hundred percent to less than 80 percent, which indicates a solid confidence for species identification. An exception was the similarity of only 50.67% for *C. batillariae* identified in the snail *Ba. attramentaria*, but even so, the bold-identification tool recognized it as this parasite. All identifications are shown in **Table 8.** below.

Table 8. Platyhelminthes species identified in 2150 Gastropoda SRA datasets from NCBI SRA archive. Similarity in % is calculated by BOLD identification tool based on Kimura 2-parameter model. The expectancy indicates whether the identified species was on purpose in the dataset or has been identified previously.

SRA dataset	Phylum	Class	Order	Family	Genus	Species	Similarity %	Host	Expected
SRR8439319	Platyhelminthes	Cestoda	Cyclophyllidea	Taeniidae	*Taenia*	*pisiformis*	100.00	*Haliotis diversicolor*	No
SRR6214989	Platyhelminthes	Trematoda	Plagiorchiida	Heterophyidae	*Cercaria*	*batilariae*	100.00	*Batillaria attramentaria*	No
SRR6214981	Platyhelminthes	Trematoda	Plagiorchiida	Heterophyidae	*Lepotrema*	*amansis*	100.00	*Batillaria attramentaria*	No
SRR6214985	Platyhelminthes	Trematoda	Plagiorchiida	Heterophyidae	*Cercaria*	*batilariae*	100.00	*Batillaria attramentaria*	No
SRR2059857	Platyhelminthes	Trematoda	Diplostomida	Schistosomatidae	*Trichobilharzia*	*stagnicolae*	99.84	*Haliotis discus hannai*	No
SRR11015439	Platyhelminthes	Trematoda	Plagiorchiida	Dicrocoeliidae	*Dicrocoelium*	*dendriticum*	99.81	*Anentome helena*	No
SRR4473781	Platyhelminthes	Trematoda	Plagiorchiida	Telorchiidae	*Telorchis*	*sp.*	99.49	*Calliostoma zizyphinum*	No
SRR6214991	Platyhelminthes	Trematoda	Opisthorchiida	Opisthorchiidae	*Opisthorchis*	*sp.*	98.77	*Batillaria attramentaria*	No
SRR6214988	Platyhelminthes	Trematoda	Plagiorchiida	Heterophyidae	*Pygidiopsis*	*sp.*	98.54	*Batillaria attramentaria*	No
SRR6214976	Platyhelminthes	Trematoda	Opisthorchiida	Opisthorchiidae	*Opisthorchis*	*sp.*	98.1	*Batillaria attramentaria*	No
SRR6395706	Platyhelminthes	Trematoda	Plagiorchiida	Zoogonidae	*Zoogonus*	*sp.*	97.92	*Pila ampullacea*	No
SRR6214991	Platyhelminthes	Trematoda	Plagiorchiida	Paramphistomidae	*Calicophoron*	*sp.*	97.92	*Batillaria attramentaria*	No
SRR6214977	Platyhelminthes	Trematoda	Plagiorchiida	Paramphistomidae	*Calicophoron*	*sp.*	97.82	*Batillaria attramentaria*	No
SRR9598643	Platyhelminthes	Trematoda	Plagiorchiida	Monorchiidae	*Hurleytrematoides*	*loi*	97.7	*Oncomelania hupensis*	No
SRR6214971	Platyhelminthes	Trematoda	Plagiorchiida	Fasciolidae	*Fasciola*	*nyanzae*	97.36	*Batillaria attramentaria*	No
SRR6214980	Platyhelminthes	Trematoda	Plagiorchiida	Fasciolidae	*Fasciola*	*nyanzae*	97.21	*Batillaria attramentaria*	No
SRR6214975	Platyhelminthes	Trematoda	Plagiorchiida	Fasciolidae	*Fasciola*	*nyanzae*	96.31	*Batillaria attramentaria:*	No
SRR6214973	Platyhelminthes	Trematoda	Plagiorchiida	Echinostomatidae	*Lyperorchis*	*lyperorchis*	96.29	*Batillaria attramentaria*	No
SRR2040660	Platyhelminthes	Cestoda	Cyclophyllidea	Hymenolepididae	*Hymenolepis*	*microstoma*	95.93	*Reishia clavigera*	No
SRR11015439	Platyhelminthes	Trematoda	Diplostomida	Clinostomidae	*Clinostomum*	*sp.*	95.93	*Anentome helena*	No
SRR5456827	Platyhelminthes	Trematoda	Diplostomida	Schistosomatidae	*Schistosoma*	*mansoni*	90.00	*Biomphalaria pfeifferi*	Yes

SRR6214971	Platyhelminthes	Trematoda	Plagiorchiida	Fasciolidae	*Fasciola*	*gigantica*	89.29	*Batillaria attramentaria:*	No
SRR5456830	Platyhelminthes	Trematoda	Plagiorchiida	Haplosplanchnidae	*Hymenocotta*	*manteri*	88.12	*Biomphalaria pfeifferi*	No
SRR9598640	Platyhelminthes	Trematoda	Diplostomida	Schistosomatidae	*Schistosoma*	*japonicum*	87.36	*Oncomelania hupensis*	Yes
SRR9598638	Platyhelminthes	Trematoda	Diplostomida	Schistosomatidae	*Schistosoma*	*japonicum*	86.23	*Oncomelania hupensis*	Yes
SRR9598637	Platyhelminthes	Trematoda	Diplostomida	Schistosomatidae	*Schistosoma*	*japonicum*	86.18	*Oncomelania hupensis*	Yes
SRR9598639	Platyhelminthes	Trematoda	Diplostomida	Schistosomatidae	*Schistosoma*	*japonicum*	85.87	*Oncomelania hupensis*	Yes
SRR5456823	Platyhelminthes	Trematoda	Diplostomida	Schistosomatidae	*Schistosoma*	*mansoni*	85.87	*Biomphalaria pfeifferi*	Yes
SRR1049094	Platyhelminthes	Trematoda	Opisthorchiida	Opisthorchiidae	*Opisthorchis*	*viverrini*	83.74	*Bithynia siamensis goniomphalos*	Yes
SRR9598645	Platyhelminthes	Trematoda	Diplostomida	Schistosomatidae	*Schistosoma*	*japonicum*	83.33	*Oncomelania hupensis*	Yes
SRR9598641	Platyhelminthes	Trematoda	Diplostomida	Schistosomatidae	*Schistosoma*	*japonicum*	82.83	*Oncomelania hupensis*	Yes
SRR5456827	Platyhelminthes	Trematoda	Diplostomida	Schistosomatidae	*Schistosoma*	*mansoni*	82.83	*Biomphalaria pfeifferi*	Yes
SRR5456826	Platyhelminthes	Trematoda	Diplostomida	Schistosomatidae	*Schistosoma*	*mansoni*	81.06	*Biomphalaria pfeifferi*	Yes
SRR11248244	Platyhelminthes	Trematoda	Diplostomida	Schistosomatidae	*Schistosoma*	*mansoni*	80.42	*Biomphalaria glabrata*	Yes
SRR6214984	Platyhelminthes	Trematoda	Plagiorchiida	Heterophyidae	*Cercaria*	*batilariae*	79.66	*Batillaria attramentaria*	No
SRR5456822	Platyhelminthes	Trematoda	Diplostomida	Schistosomatidae	*Schistosoma*	*mansoni*	79.26	*Biomphalaria pfeifferi*	Yes
SRR5456825	Platyhelminthes	Trematoda	Plagiorchiida	Paramphistomidae	*Calicophoron*	*sp.*	78.75	*Biomphalaria pfeifferi*	Yes
SRR6214989	Platyhelminthes	Trematoda	Plagiorchiida	Heterophyidae	*Cercaria*	*batilariae*	78.57	*Batillaria attramentaria*	No
SRR6214987	Platyhelminthes	Trematoda	Plagiorchiida	Heterophyidae	*Cercaria*	*batilariae*	50.67	*Batillaria attramentaria*	No

Identified nematodes

In total, only 24 representatives of the Nematoda group were found in 2150 gastropod SRA datasets, none of which were expected in specific samples (i.e. the authors did not report the infection).

Since nematodes are a huge group of ubiquitous organisms including both parasites and free-living representatives, not all representatives found necessarily imply a parasitic relationship to the host. This fact is well represented in our results, where than one quarter of revealed nematodes in SRA samples are represented by the notoriously known model organism, free-living nematode *Caenorhabditis elegans*. It has been found in freshwater, marine and terrestrial gastropods (*Pomacea canaliculata*, *Haliotis tuberculata* and *Cornu aspersum* respectively) and parasitism has never been observed in this species. Another free-living nematode found in the datasets from *B. glabrata and P. canaliculata* was *Diploscapter coronatus*, which is also exclusively bacteriovorous, although possibility of its ability to parasitise was speculated (Chandler, 1938; Kang et al., 2017).

A very important discovery is also the presence of *Angiostrongylus cantonensis* in a total of 7 SRA datasets, from which five obtained from *Achatina immaculata* and two from *P. canaliculata*. This finding shows well that this parasite, which is very dangerous for human health as already described above, can be easily overlooked even by scientists. Since both snails are frequently consumed by humans, this fact is even more striking and will be further debated in discussion.

An unusual discovery, which could be compared to the discovery of tapeworms in the SRA datasets described above, is the presence of the parasitic nematode *Trichinella pseudospiralis* in two datasets obtained from the abalone *H. diversicolor*. the life cycle of this parasite is very well described and certainly does not include a gastropod host. Possible reasons for this result will be discussed in more detail in the Discussion.

The possible discovery of a hitherto unknown parasite-host relationship may be represented by two identifications of *Rhabdias picardiae* in the planorbid snail *B. glabrata*. The genus *Rhabdias* is a globally distributed genus of parasitic nematode of reptiles and.amphibians with an occasional gastropod paratenic host. *R. picardiae*, however, has not yet been described as having an intermediate or paratenic host.

The last four parasitic nematode representatives found are rather atypical, perhaps even obscure organisms. For example, the bee parasite *Acrostichus halicti* and parasite of manatees *Cutidiplogaster manati* - both from *B. glabrata*, the intestinal nematode *Parastrongyloides trichosuri* infecting brushtail possums found in the slug *Arion vulgaris*, and *Ortleppascaris sinensis* from *Cornu aspersum* snail, a typical parasite from a digestive system of alligators.
All identification are shown in **Table 9.** below.

Table 9. Nematoda species identified in 2150 Gastropoda SRA datasets from NCBI SRA archive. Similarity in % is calculated by BOLD identification tool based on Kimura 2-parameter model. The expectancy indicates whether the identif ied species was on purpose in the dataset or has been identified previously.

SRA dataset	Phylum	Class	Order	Family	Genus	Species	Similarity %	Host	Expected
SRR2062360	Nematoda	Chromadorea	Rhabditida	Rhabditidae	*Caenorhabditis*	*elegans*	100	*Cornu aspersum*	No
SRR2581512	Nematoda	Chromadorea	Rhabditida	Rhabditidae	*Caenorhabditis*	*elegans*	100	*Haliotis tuberculata*	No
SRR2581513	Nematoda	Chromadorea	Rhabditida	Rhabditidae	*Caenorhabditis*	*elegans*	100	*Haliotis tuberculata*	No
SRR7454806	Nematoda	Chromadorea	Rhabditida	Angiostrongylidae	*Angiostrongylus*	*cantonensis*	100	*Pomacea canaliculata*	No
SRR10054003	Nematoda	Chromadorea	Rhabditida	Angiostrongylidae	*Angiostrongylus*	*cantonensis*	99.89	*Achatina immaculata*	No
SRR10054005	Nematoda	Chromadorea	Rhabditida	Angiostrongylidae	*Angiostrongylus*	*cantonensis*	99.89	*Achatina immaculata*	No
SRR10053982	Nematoda	Chromadorea	Rhabditida	Angiostrongylidae	*Angiostrongylus*	*cantonensis*	99.89	*Achatina immaculata*	No
SRR10053983	Nematoda	Chromadorea	Rhabditida	Angiostrongylidae	*Angiostrongylus*	*cantonensis*	99.88	*Achatina immaculata*	No
SRR10053981	Nematoda	Chromadorea	Rhabditida	Angiostrongylidae	*Angiostrongylus*	*cantonensis*	99.86	*Achatina immaculata*	No
SRR6429139	Nematoda	Chromadorea	Rhabditida	Rhabditidae	*Caenorhabditis*	*elegans*	99.7	*Pomacea canaliculata*	No
SRR6429163	Nematoda	Chromadorea	Rhabditida	Rhabditidae	*Caenorhabditis*	*elegans*	99.7	*Pomacea canaliculata*	No
SRR7454808	Nematoda	Chromadorea	Rhabditida	Angiostrongylidae	*Angiostrongylus*	*cantonensis*	99.68	*Pomacea canaliculata*	No
SRR2581517	Nematoda	Chromadorea	Rhabditida	Rhabditidae	*Caenorhabditis*	*elegans*	99.55	*Haliotis tuberculata*	No
SRR13700391	Nematoda	Enoplea	Trichocephalida	Trichinellidae	*Trichinella*	*pseudospiralis*	96.84	*Haliotis diversicolor*	No
SRR1617599	Nematoda	Chromadorea	Rhabditida	Rhabdiasidae	*Diploscapter*	*coronatus*	96.08	*Biomphalaria glabrata*	No
SRR1617540	Nematoda	Chromadorea	Rhabditida	Rhabdiasidae	*Rhabdias*	*picardiae*	95.4	*Biomphalaria glabrata*	No
SRR1617537	Nematoda	Chromadorea	Rhabditida	Rhabdiasidae	*Rhabdias*	*picardiae*	95.24	*Biomphalaria glabrata*	No
SRR13700389	Nematoda	Enoplea	Trichocephalida	Trichinellidae	*Trichinella*	*pseudospiralis*	90.43	*Haliotis diversicolor*	No
SRR1617536	Nematoda	Chromadorea	Rhabditida	Diplogastridae	*Acrostichus*	*halicti*	88.98	*Biomphalaria glabrata*	No

SRR1617630	Nematoda	Chromadorea	Rhabditida	Diplogastridae	*Acrostichus*	*halicti*	88.98	*Biomphalaria glabrata*	No
SRR1617541	Nematoda	Chromadorea	Rhabditida	Diplogastridae	*Cutidiplogaster*	*manati*	88.31	*Biomphalaria glabrata*	No
SRR7224659	Nematoda	Chromadorea	Rhabditida	Rhabditidae	*Diploscapter*	*coronatus*	88.3	*Pomacea canaliculata*	No
SRR7841657	Nematoda	Chromadorea	Rhabditida	Rhabdiasidae	*Parastrongyloides*	*trichosuri*	87.74	*Arion vulgaris*	No
SRR2062359	Nematoda	Chromadorea	Rhabditida	Ascarididae	*Ortleppascaris*	*sinensis*	83.91	*Cornu aspersum*	No

3.4 Discussion

When you say the word parasite, most people automatically associate it with negative emotions, believing that we should do everything we can to get rid of these organisms forever. But as it was already mentioned in the introduction of this work, the understanding of the term parasite can have many forms and in certain interpretations we humans are also parasites. Parasites are an integral part of living systems and play a major role in biodiversity, evolution, but also, for example, in the beneficial control of populations. For this reason, too, an initiative for the protection of certain species of parasites has arisen. However, in spite of the many positives that parasitism brings, it is necessary to admit that some parasites are a very important problem from the point of view of human and veterinary medicine and it is necessary to fight them intensively within the framework of our self-preservation. A fundamental predisposition in the fight against medically and veterinary important parasites is a good understanding of their life cycles, modes of transmission, but especially host specificity and geographical distribution. With many scientists around the world collecting material from nature and generating vast amounts of sequencing data for various purposes, there is a huge opportunity to expand our knowledge about host-parasite relationships from the comfort of an office chair. With these motivations in mind, the tool described above was developed for rapid identification of hidden parasites in publicly available datasets.

The main condition that this tool had to meet is speed, because it had to be designed to work with a huge amount of data. After initial attempts with lengthy downloading of specific SRA datasets and their mapping using STAR or Hisat2, MAGIC-BLAST was chosen. This tool allows mapping SRA datasets to the database without the need to download them. When it became evident after pilot analyses and time measurements that MAGIC-BLAST is the most time-consuming part of the analysis and also the most influential on its specificity, we decided to iterate through its adjustable parameters to achieve the most time efficient analysis with the specificity we define. The word_size parameter was the only one that slowed down the analysis very significantly when lowered, of all four of the adjustable parameters. This was probably due to the fact that the smaller the word (which in this case stands for the number of exactly mapping bases in a row), the larger the number of words MAGIC-BLAST has to evaluate. After it was found in the following tests that word_size was the only one besides gap_extend that, in combination with penalty parameter reduction, significantly changed the tool's specificity in our analysis, it was necessary to choose which of these parameter combinations to choose as our "relaxed parameters" for further experiments. Apart from the duration of the analysis itself, which played a huge role in our decision, word_size was eliminated as a changeable parameter because its reduction also significantly increases the risk of seeding to ambiguous positions. Penalty and gap_extend parameters on the other hand does not affect the computational process itself, it just changes the score of relevance for every mismatch (Boratyn et al., 2018).

Removing certain taxonomic units from the input COX1 database also had an interesting effect? when testing the tool. The first at first sight peculiarity was a more precise identification of the species present in the sample when the COX1 sequence of this species itself was missing in the database, but all other sequences of higher taxonomic units such as

genus or family were preserved. This phenomenon has a simple explanation, namely that the exact complete COX1 sequence present in the database may not always be the same as the exact complete COX1 sequence in the sample. This is due to the genetic variation of individuals of the same species within or between populations. Therefore, in the case where the identification is limited to only a part of the sequence that is well conserved and therefore not varies between individuals, the identification itself may be higher in percentage similarity.

Another and probably the biggest peculiarity that appeared in the framework of these tests was the impairment of the parasite's COX1 sequence assembly present in the sample (*S. mansoni* from phylum Platyhelminthes) during the successive removal of COX1 sequences of different taxonomic units of the host (*B. glabrata* from phylum Mollusca). When the COX1 sequences of the genus and family of the host were removed, the assembled COX1 sequence of the parasite was significantly truncated, and even though the parasite was still always correctly identified, the explanation for this phenomenon is difficult to find. The mapping of the found reads was always done using a database that contained the complete sequence of the parasite, including the sequences of other related species of its genus or family. The only possible explanation could be that due to the high level of conservation of COX1 sequences, certain reads of highly conserved regions that map in complete database to the COX1 sequence of the snail are involved in the assembly sequence of the parasite in the next steps of the analysis.

An absence of newly discovered schistosome infections in public SRa datasets

After the tool was optimised, the main hope was to identify schistosomes in specimens of snails collected in the wild and thus discover potential new locations, species or even new hosts of these most important helminthic agents of human disease with it. This has unfortunately not been achieved, and probably for several reasons. First of all, only a small fraction of all datasets tested were performed on freshwater gastropods, and if so, these were often laboratory breeds. Secondly, although in some cases the snails were collected in the wild, it can be assumed that experienced scientists, especially malacologists, who are directly involved in, for example, genotyping of snails or studying specific gene expression in these organisms, are well aware of the possible parasitic infections and thus probably carry out inspections before collecting the snails. This is often done, for example, by exposing them to direct light, which is generally a signal for cercariae of schistosomes and other trematodes to leave the intermediate host. In this way, the researcher can easily ascertain whether the gastropod is really healthy and thus suitable for further experiments (Ouma et al., 1989). A prepatent period that would prevent the differentiation of a healthy from an infected snail could lead to a collection of trematode-infected individuals, but the subsequent RNA isolation for sequencing is usually collected only from a small part of the body, for example the leg, where the parasite is typically absent. The specific tissue from which the RNA originates is of course not the reason for the absence of only schistosomes in the samples, but in general of most parasites that could otherwise be theoretically discovered.

Identifications of hidden trematodes – interesting and dubious?

Even though we did not find any schistosomes, we still managed to detect many interesting parasitic infections in apparently healthy gastropods thanks to our tool. The highest number of different parasite species was identified in the snail *Batillaria attramentia*. All analyses of this host come from one laboratory and were produced in the framework of one Bioproject PRJNA415312, which produced a total of 24 datasets - 16 of them positive for parasites. Apart from the fluke *Cercaria batilariae*, which is a specific trematode infecting this snail and is not important for human or veterinary medicine, parasites causing very serious diseases in humans and animals were also identified. In this case, it is very important to critically evaluate whether the findings can be reliable and whether it is not, for example, a laboratory contamination of the sample by the researchers or the servicing sequencing company.

The most interesting and potentially very important discovery in this dataset that must be critically assessed was the identification of the *Opisthorchis sp*. parasite in this analysis. This medically very important genus of parasites has been studied intensively for many years and only snails of the genus *Cordiella* sp. or *Bithynia* sp. are known to serve as intermediate hosts (CDC, 2018). This infection and its detection would seem realistic because the known geographic distribution of these parasites is intertwined with the sampling locations where the authors of this study collected snails, i.e. Southeast and East Asia (Ho et al., 2021). However, a fact that casts doubt on this theory is that the known intermediate hosts of these parasites are small freshwater snails, and the paratenic hosts of this parasite are freshwater fish whereas *Batillaria* is a genus of salt marsh or mudflat snails and the parasites were never detected in marine fish. On the other hand, if we take into account that the tool did not identify a specific parasite species, but only the genus *Opisthorchis sp*., it is likely that this is a previously undescribed species whose COX1 sequence is already in the BOLD database. If this is the case, intermediate hosts, paratenic hosts and definitive hosts of this parasite are not known and this finding in the *Batillaria* snail is not impossible. \

Other potentially very important discovery is the identification of medically important parasite *Fasciola gigantica*. This liver fluke causes the very serious disease fasciolosis in humans and ruminants and in some regions the prevalence is between staggering 80 and 100%. For this fluke, only one susceptible intermediate host species is known to this day - *Lymnaea Auricularia*. Based on geographical location, the intermediate hosts differ only in a subspecies. In Asia, where this fluke is also abundant, it is *L. auricularia rubiginosa* (Spithill et al., 1999). If it could indeed be empirically demonstrated that *B. atramentia* can also serve in the life cycle of this parasite, it would be a discovery of great importance for human and veterinary medicine. On the other hand, the identification itself was determined on the basis of a similarity of 89.29%, which may indicate that it is a different, previously unknown species. Given that all species of the genus *Fasciola* cause serious diseases in animals and some in humans, even the identification of a new species from this genus would be a major contribution.

Another discovery that unfortunately might reduce the relevance of the identifications from this whole Bioproject is the detection of the *Fasciola nyanzee* which is a specific parasite of hippopotamuses, whose territory is exclusively limited to Africa and thus not found

throughout Asia (Dinnik & Dinnik, 1961). On the contrary, the identification of this parasite by the BOLD-identification tool was made in three different samples with similarity 96.31%, 97.21%, 97.36%, which may indicate a previously undiscovered evolutionarily closely related species that is not in the BOLD database yet. This potentially unknown species may thus not be limited to hippopotamuses and its distribution in Asia where the samples were collected is then entirely possible.

Another curious identification, but one that is unmistakable in its degree of similarity to another parasite or unknown species, is the finding of the notoriously known *Dicrocoelium dendriticum* in the popular aquarium snail *Anentome helena*. Unfortunately, the publication that produced these RNAseq data does not indicate where the snails came from - whether from the wild or from laboratory or other breeding facilities (Dvorak et al., 2020). Although the areas of distribution of the two species overlap, the *D. dendriticum* fluke is exclusively restricted to terrestrial environments. Typical first intermediate hosts are terrestrial snails of the genera *Zebrina sp*. and *Helicella sp.*, which become infected by ingesting eggs with the faeces of definitive hosts, primarily ruminants. The infected snails then produce slime balls full of infective metacercariae ingested by secondary intermediate hosts, ants of the genus *Formica sp*. It is evident from this life cycle, that the specific participation of *A. helena*, which is strictly aquatic, in it is highly unlikely. What is possible, however, that this snail may have ingested the eggs of this parasite with the faeces that entered the aquatic environment. This would be supported by the fact that the transcriptome in study was generated from the midgut gland, which is in primary contact with the snail's gut. However, since *A. helena* is a popular aquarium snail that is commonly bred all over the world, it is more likely that the authors of the study obtained it from such breeds and the identification of *D. dendriticum* is again only contamination from the laboratory.

Identifications of hidden cestodes – only dubious

Probably the most surprising identifications, whose validity can hardly be explained, are the finding of two tapeworm species in marine gastropods, namely *Hymenolepis microstoma* in the intertidal whelk *Reishia clavigera* and *Taenia pisiformis* in the abalone *Haliotis diversicolor*. Although both of these tapeworm species occur worldwide, both have very well described life cycles, and in neither case is the parasite dependent on the aquatic environment nor the gastropod intermediate hosts - *H. microstoma* uses beetles and other arthropods as intermediate hosts and rodents as definitive hosts, and *T. pisiformis* uses rabbits and carnivores in its life cycle (Dvorak et al., 1961; D. Yang et al., 2012). While for the identification of *H. microstoma* no other explanation is offered than laboratory contamination either during the RNA isolation or during sample processing for sequencing, for the identification of *T. pisiformis* one, admittedly very unlikely, explanation is offered. The meat of *H. diversicolor* is valued as a delicacy, which has led to a decline in its population in some areas. To ensure a supply of this animal for these purposes, it is often bred on the farms from which the abalones in this study also come. Although traditionally fed by grazing on algae, these abalones are often also fed with artificial food to increase the protein in the diet and improve growth (Fitzguerald, 2008). The sources of this protein are up to half of animal origin, which is traditionally produced in rendering plants where dead animals end up, including

roadkills, euthanized animals from animal shelters and the like. Since *T. pisiformis* forms huge tissue cysts in the intermediate rabbit host, the genetic material of the parasite is found in extensive quantities in the animal. If such a rabbit could end up in a rendering plant and then subsequently in the artificial food on these abalone farms, there is a certain probability that the sequences of the parasite could be transferred to the abalone which fed on it. The idea to reconstruct trophic networks or food chains through metabarcoding of gut contents or stool samples was already proven possible and useful (Casey et al., 2019; Rytkönen et al., 2019).

Identifications of hidden nematodes – interesting and credible

The situation is bit different for nematodes. Because parasitic nematodes do not have typical larval stages that could be easily distinguished in the intermediate host, such as sporocysts and cercariae of trematodes, their findings in snails are rather rare. Although this would theoretically favour our analysis to detect these hidden infections, nematodes that normally use gastropods as intermediate hosts are very few. One of them, however, are the representatives of the genus *Angiostrongylus sp*., which are extremely dangerous parasites to humans precisely because of their difficult recognition in snails. These parasites were identified in seven different datasets from two different Bioprojects that we analysed. We consider all of our identifications of this parasite in *Achatina immaculata* and *Pomacea canaliculata* to be very reliable in this case. The *A. immaculata* individuals for this study were obtained from a farm in Guangdong province, China, which is a hotspot for the *A. cantonensis* parasite and where the first human case in the country was reported in 1984 (C. Liu et al., 2021; He et al, 1984 as cited in Chen et al., 2011). Unfortunately, we could not find the source of *P. canaliculata* for the Bioproject PRJNA476647, because it was not linked to any published work, but the data were submitted by the Institute of Zoology from Chinese Academy of Sciences, so it can be assumed that the analysed snails came from China, which strengthens our confidence in the results. An epidemiological study from 2011 that aimed to master the epidemic situation of the parasite in mentioned Chinese province examined 3,184 closely related *Achatina* snails of which 797 were infected (over 25%) and 3,723 *P. canaliculata* snails with infection rate of 6,5% (Chen et al., 2011). Such a huge prevalence is obviously alarming, given that both species of these snails are often consumed by humans. Although these identifications do not provide new information about the hosts and their distribution, they serve as another good "sanity check" (in addition to the previously known infected samples) for our tool, and thus increase the credibility of the other identifications in this work.

One of these potentially credible revelations is the identification of *Rhabdias picardiae* in *B. glabrata*. *Rhabdias sp*. lungworms are worldwide distributed parasites of reptiles and amphibians, and some species are known to use gastropods as paratenic hosts (Mehlhorn, 2008). *R. picardiae* is a relatively recently discovered species, so its life cycle is not well described yet and this discovery probably reveals important part of it (Junker et al., 2010).

As with the analysis of platyhelminthes, the analysis of nematodes does not lack an engimatic result, which is only difficult to explain and is again from a dataset obtained from the abalone *H. diversicolor*. As in the case of the strange discovery of the tapeworm *T. pisiformis* discussed above, which forms tissue cysts, for the identification of which we offered an explanation through the feeding of abalones with artificial food, no other explanation is

offered in the case of the identification of *Trichinella pseudospiralis*. Nematodes from the genus *Trichinella sp.* are biohelminths, with no exogenous phase in their life cycle (i.e. they do not occur in the external environment outside their hosts, which are only mammals and birds), which means that their occurrence in the aquatic environment and in *H. diversicolor* itself cannot have a natural cause. As an alternative explanation, contamination from the laboratory is again suggested, but this would also be very strange considering that both of these strange findings come from the same species of abalone, two different Bioprojects and two absolutely different tissue parasites.

Limitations of the tool

Although the tool for recognition and identification of arbitrary COX1 sequences in publicly available RNAseq datasets that we developed in this project produced interesting results, there are still many limitations and question marks that need to be taken into account. As is obvious from the results, the biggest limitation of this tool is the input database it works with. As described in chapter **3.3.3**, in the absence of sequences of a congenic species, genus or even family, the tool is very weak in the assembly of COX1 sequences, and of course the identification of the present species suffers. This limitation is of course solvable by the consistent compilation of as many reliable public COX1 databases as possible to maximize the amounts of mapped reads and thus possible identifications. Another limitation is the identification of individual species, which relies on the BOLD identification tool, which of course identifies species according to their presence in the BOLD database. This means that sequences that were assembled, for example, by mapping to sequences obtained from the MIDORI database, are not found in the BOLD database and their identification may be inaccurate, erroneous or even missing. This problem could be solved by implementing a local BLAST in the pipeline, where the database used for the MAGIC-BLAST mapping itself would serve as the database for subsequent identification. The optimization of this way of identification should not be very demanding, but its testing was unfortunately no longer within the time possibilities of this project.

Another limitation, or rather a prospect for future work, are the identifications that were made based on imperfect similarities, such as in case of *F. nyanzee*, which almost certainly represents another species unknown to the database, as discussed above. In this case, it would be necessary to either proceed or to implement to the pipeline a phylogenetic analysis of the COX1 sequences of the most closely related species, which would allow us to better reveal whether it is an as yet unknown species and alternatively to which species the identified species is closest to, and thus draw more accurate conclusions about the possible biology of the hosts and their respective parasites.

Development of gene-engineering based strategies against schistosomiasis

Besides this basic research, which includes the development of the bioinformatic tool, which was created in the framework of this thesis and thanks to which we can better understand the relationships between parasites and hosts, there is also a need for applied research in field of parasitology. This research is directly based on such basic research, but which has the power to shape the world around us not only for our benefit and perhaps also for the benefit of nature.

With the increasing incidence of resistance of various parasites to established drugs and the long-standing ineffectiveness in education and prevention against these infections in the endemic areas, there is an increased demand for new, effective strategies to combat many of these debilitating diseases.

With our better understanding and the rapid progress in development of modern molecular and genetic methods, entirely new avenues are opening for the control and eradication, especially of diseases caused by vector-borne parasites. Even though these approaches bring with them several ethical problems that must be considered, it is now becoming clear that genetic modification of vectors so that they cannot transmit disease, and their introduction into the wild, where they should eventually replace the original population can be an effective solution. Following the successful proof of principle and gradual implementation of this strategy in the case of malaria and the *Anopheles sp.* mosquitoes that transmit the disease, scientists working on other diseases have begun to focus on the use of these technologies (Kyrou et al., 2018; Pare Toe et al., 2021). One of these diseases is of course schistosomiasis, which is still the most medically important helminthiasis.

As in the case of the development of CRISPR/Cas9, as the method of first choice for irreversible targeted changes in genetic information, in other models, also in the case of *B. glabrata*, which is an intermediate host of *S. mansoni*, it is necessary to select suitable genes on which to test the method and to choose a suitable way of transferring the respective constructs into the cells of the organism (Maier et al., 2019). Since the development of this *de novo* protocol was originally the primary goal of this work, these tasks were also the first ones to start my research with.

The *in vitro* cultivation of *B. glabrata* embryos

The main and most important characteristic of the CRISPR/Cas9 is that the genetic information is changed permanently. If this is done in the 1-cell stage, this changed genetic information is present in all other cells derived from it and this means also the cells of the sex organs and ultimately the gametes. The offspring of such altered individuals then carries the same altered information. However, in order to do this in the case of *B. glabrata*, we must first decapsulate the 1-cell embryo from the egg envelope and the egg fluid in which it is contained, modify it, and then ensure its survival and development into an adult fertile individual. Even though this task may seem trivial, our experiments have shown that it is an extremely challenging objective. An indication that it will not be easy was also the fact that, according to the available literature, only one laboratory, which has been dealing with the manipulation of snail embryos for nearly 15 years, has succeeded in doing so in *Lymnaea stagnalis* (Abe et al., 2009, 2014; Kuroda et al., 2009, 2016; Abe & Kuroda, 2019). For this reason, we decided to follow their procedures, where the successful cultivation of decapsulated embryos is done by placing these cells in glass capillaries filled with egg fluid, in which the embryo was originally surrounded in an egg capsule. The main problem turned out to be the size and number of egg capsules of our snail. While *L. stagnalis* lays 50-120 eggs within one egg mass, in *B. glabrata* this number is around one third. The individual egg capsules themselves are then 2-3 times smaller, which means that they also contain significantly less egg fluid, which is needed for the cultivation of eggs after decapsulation (Kuroda & Abe, 2020). The *B. glabrata* zygote itself

is only about 60 um in size, and is very fragile, so that during the collection of the egg fluid, individual early embryos are often damaged. Without a suitable micromanipulator, which unfortunately I did not have at my disposal during these experiments, the successful collection of this fluid, as well as the manipulation and transfer of the embryo to the glass capillary without damage was almost impossible. For these reasons I then proceeded to experiment with dilution of egg fluid with Bge culture medium in the hope of reducing the need and consumption of egg fluid. The Bge medium is based on long-term experiments on the 1818 embryonic Bge cell line, which is derived from the embryonic cells of *B. glabrata* and is the only existing cell line originating from any molluscan species (Hansen, 1979; Yoshino et al., 2013). The logic so far has been that if this medium is sufficient for this cell line to survive and divide, it could at least partially replace egg fluid in the cultivation.

However, as is evident from the results, although this medium probably can nutritionally satisfy molluscan embryotic cells for short-term survival, it cannot replace the various developmental factors contained in natural egg fluid. Moreover, statistical analysis of the results showed that the concentration of egg fluid in the medium significantly increases the ability of embryos to survive in culture. These empirical findings are supported by older literature where *L. stagnalis* embryos were cultured in different ratios of different culture media with distilled water and egg fluid in glass capillaries with open ends. As in my case, the embryos survived and developed relatively normally only in pure undiluted egg fluid. It is not clear from the publication whether the authors were able to raise the snails to adulthood, but they declare that 90% of the snails cultured in pure egg fluid survived to the stage where they were able to survive in artificial pond water (Dickinson & Croll, 2001). The authors even tried to culture the media with different growth factors, which have been established in the past as important for the proper development of *L. stagnalis* embryos (Nagle et al., 1999; Hermann et al 2000, both in Dickinson & Croll, 2001), but even in this case they did not succeed. However, as they themselves acknowledge in the discussion, they used human growth factors, which are sequentially only very slightly similar to the snail ones. The absence of these growth factors may also be a probable explanation for the deformed embryo described in this thesis, which was cultured only in Bge medium. Although the cells had sufficient nutrients for growth and cell division, the absence of these growth factors may have led to non-specific development and subsequent death.

If one would continue in the efforts to raise *B. glabrata* embryos effectively, several possibilities are offered based on the presented literature. The basic predisposition is the aforementioned suitable micromanipulator, which would simplify the work considerably and would probably allow efficient collection of necessary egg fluid, even in such small quantities. Alternatively, such a micromanipulator could be used to transplant the removed and manipulated embryo into a new ”host” egg capsule, as has already been demonstrated on larger molluscan species (Kuang et al., 2002). A potentially interesting route to take, even in terms of developmental biology, would be to use the egg fluid of another snail species, which has significantly larger egg capsules and thus easier to extract, to culture our model. Although it has never been tried, there are many species of larger gastropods in the family Planorbidae (where *B. glabrata* also belongs), where the sequences of the necessary growth factors are probably well conserved and might be thus very similar.

Cho osing the appropriate gene for the future CRISPR/Cas9 testing

In parallel with these experiments, we also wanted to identify a suitable gene for future testing, at a time when *in vitro* embryo culture and manipulation will be optimized. Such a gene should have a clear phenotypic but non-lethal expression to make it easy to evaluate the success of the experiment. Although a focus on pigmentation in *B. glabrata* would be suggested, since individuals of this species have a frequent albino form, the genes responsible for this have still not been determined. For this reason, as in the case of *in vitro* culture, we decided to follow a published study that successfully applied CRISPR/Cas9 to *L. stagnalis*, where a knockout of the *LsDia*1 gene altered the left-right symmetry of this gastropod. The aim of the bioinformatic analysis was to determine whether, naturally sinistral, *B. glabrata* also has two orthologous copies of the *Dia* gene as in *L. stagnalis* and whether these genes are therefore a suitable target in our case.

When it became apparent from pilot analyses that our model has only one copy of this gene, we hypothesized that the sinistral symmetry of our snail is due to the secondary loss of one copy in the course of evolution. We attempted to strengthen this hypothesis by analysing the duplication of the *Dia* gene across Gastropoda in relation to the symmetry of the shells of the representatives of this group. Unfortunately, this analysis did not yield clear results that would make it obvious that this relationship is indeed present and therefore it was not possible to determine whether this gene would be suitable for verification of CRISPR/Cas9 success. The phylogenetic tree of Gastropoda with assigned number of *Dia* genes possesed by individual species indicated that the presence of only one *Dia* gene may not directly correlate with the shell orientation of the respective species. However, it is also important to note that the *Diaphanous* gene sequences were extracted from the transcriptomes and not from the snail genomes (for many species not available). These transcriptomes were often produced from different parts of the bodies of different developmental stages of these snails. This may mean that some *Diaphanous* genes may not necessarily have been discovered, as they may not have been expressed in the samples from which the transcriptomes were derived.

However, if it was a matter of validation alone, almost any gene whose modification does not lead to the death of the snail could be used for these experiments. The validation of successful modification, deletion or insertion would not necessarily depend on phenotypic expression, but only on the basis of molecular-genetic methods. These include for example routine PCR for presence of the certain gene followed by sequencing of the PCR product in case of gene modification, qPCR and *in situ* hybridization to evaluate gene expression, or even evaluation of protein expression by immunological methods such as western blot or immunohistochemistry.

3.5 Concluding remarks

Even though my PhD project was severely disrupted by the ongoing COVID-19 pandemic, the resulting work has produced interesting results that are a good motivation to continue these efforts. Probably the most important contribution of this work is the bioinformatics tool that has proven to be functional in the detection and identification of parasites present in available free gastropod datasets. Although individual results must be critically evaluated and cannot automatically be taken as fact, they can serve as a guide for further research. The identification of two liver flukes of the genera *Opisthorchis* and *Fasciola* may serve as an example for all. Considering that in both cases these are genera which are of great importance for human and veterinary medicine, it would certainly be worthwhile to verify on the basis of this identification whether it is correct. This could be done, for example, by re-collecting *Ba*. *atramentia* at the locations from which the data were obtained and subsequent detailed examination of the snails by parasitological dissection. If intramolluscan stages of parasites are found, which in these cases are sporocysts, rediae and cercariae, it could then be determined whether these species are present by morphological key, simple PCR or direct sequencing. Such a discovery has had far-reaching consequences in the fight against these tropical diseases in endemic countries. In spite of the mentioned limitations, this tool can be useful not only in the detection of parasitic sequences, but also in the study of symbiotic organisms, trophic networks or even in the description of the microbiome of any group of organisms with a suitably selected database.

In the framework of this work, the development of a protocol for CRISPR/Cas9 on *B. glabrata*, which is an intermediate host of the important blood fluke *S. mansoni*, was also included. In spite of difficult conditions in pilot experiments in embryo culture of this snail, the lack of necessary equipment and *in situ* supervision resulted in interesting results and laid the foundations for future work on this project. The findings, such as the impossibility of culturing these embryos without the presence of egg fluid, but also the proof of concept of the suitability of culturing these embryos in glass capillaries for our model will be very useful for further experiments. The analysis of the *diphanous* gene in this case only showed that this gene is an unsuitable target for optimization of CRISPR/Cas9 in our model, but this is also useful information that needs to be handled accordingly. I believe that even though this work is a relatively disparate result of a disrupted PhD project, it contains useful information and unique and interesting data have been produced within it, thus contributing to our better understanding of living systems.

3.6 References

Abe M., Kuroda R. (2019). The development of CRISPR for a mollusc establishes the formin Lsdia1 as the long-sought gene for snail dextral/sinistral coiling. *Development (Cambridge)*, *146*, 1–7.

Abe M., Kuroda R. (2020). Erratum: Correction: The development of CRISPR for a mollusc establishes the formin Lsdia1 as the long-sought gene for snail dextral/sinistral coiling (doi:10.1242/dev.175976) (Development (Cambridge, England) (2019) 146 9 PII: dev180778). *Development (Cambridge, England)*, *147*.

Abe M., Shimizu M., Kuroda R. (2009). Expression of exogenous fluorescent proteins in early freshwater pond snail embryos. *Development Genes and Evolution*, *219*, 167–173.

Abe M., Takahashi H., Kuroda R. (2014). Spiral cleavages determine the left-right body plan by regulating nodal pathway in monomorphic gastropods, *Physa acuta*. *International Journal of Developmental Biology*, *58*, 513–520.

Acosta D., Cancela M., Piacenza L., Roche L., Carmona C., Tort J. F. (2008). *Fasciola hepatica* leucine aminopeptidase, a promising candidate for vaccination against ruminant fasciolosis. *Molecular and biochemical parasitology*, *158*, 52–64.

Adli M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature Communications*, *9*, 1911.

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

Anthony R. M., Rutitzky L. I., Urban J. F. J., Stadecker M. J., Gause W. C. (2007). Protective immune mechanisms in helminth infection. *Nature reviews. Immunology*, *7*, 975–987.

Ardila-Garcia A. M., Umphrey G. J., Gregory T. R. (2010). An expansion of the genome size dataset for the insect order Hymenoptera, with a first test of parasitism and eusociality as possible constraints. *Insect molecular biology*, *19*, 337–346.

Augusto R. de C., Tetreau G., Chan P., Walet-Balieu M. L., Mello-Silva C. C., Santos C. P., Grunau C. (2017). Double impact: natural molluscicide for schistosomiasis vector control also impedes development of *Schistosoma mansoni* cercariae into adult parasites. *PLoS Neglected Tropical Diseases*, *11*, 1–19.

Benchimol M. (2000). Ultrastructural characterization of theisolated hydrogenosome in *Tritrichomonas foetus*. *Tissue and Cell*, *32*, 518–526.

Berg C. O. (1973). Biological control of Snail-Borne diseases: A review. *Experimental Parasitology*, *33*, 318–330.

Bolger A. M., Lohse M., Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, *30*, 2114–2120.

Boratyn G. M., Thierry-Mieg J., Thierry-Mieg D., Busby B., Madden T. L. (2018). Magic-BLAST, an accurate DNA and RNA-seq aligner for long and short reads. *bioRxiv*, 1–19.

Buchfink B., Xie C., Huson D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature methods*, *12*, 59–60.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*, 421.

Capella-Gutiérrez S., Silla-Martínez J. M., Gabaldón T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)*, *25*, 1972–1973.

Casey J. M., Meyer C. P., Morat F., Brandl S. J., Casey J. M. (2019). Reconstructing hyperdiverse food webs : Gut content metabarcoding as a tool to disentangle trophic interactions on coral reefs, *2019*, 1157–1170.

Chandler A. C. (1938). *Diploscapter coronata* as a facultative parasite of man, with a general review of vertebrate parasitism by rhabditoid worms. *Parasitology*, *30*, 44–55.

Chen D., Zhang Y., Shen H., Wei Y., Huang D., Tan Q., Lan X., Li Q., Chen Z., Li Z., Ou L., Suen H., Ding X., Luo X., Li X., Zhan X. (2011). Epidemiological survey of *Angiostrongylus cantonensis* in the west-central region of Guangdong Province, China. *Parasitology Research*, *109*, 305–314.

Correa A. C., Escobar J. S., Durand P., Renaud F., David P., Jarne P., Pointier J. P., Hurtrez-Boussès S. (2010). Bridging gaps in the molecular phylogeny of the Lymnaeidae (Gastropoda: Pulmonata), vectors of Fascioliasis. *BMC Evolutionary Biology*, *10*, 381.

Cowie R. H. (2013). Biology, systematics, life cycle, and distribution of *Angiostrongylus cantonensis*, the cause of rat lungworm disease. *Hawai'i journal of medicine & public health : a journal of Asia Pacific Medicine & Public Health*, *72*, 6–9.

Dai J. R., Li Y. Z., Wang W., Xing Y. T., Qu G. L., Liang Y. S. (2015). Resistance to niclosamide in *Oncomelania hupensis*, the intermediate host of *Schistosoma japonicum*: Should we be worried? *Parasitology*, *142*, 332–340.

Danna K., Nathans D. (1971). Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Hemophilus influenzae*. *Proceedings of the National Academy of Sciences of the United States of America*, *68*, 2913–2917.

Dickinson A. J. G., Croll R. P. (2001). A culture technique for experimental studies of embryonic development in the pond snail *Lymnaea stagnalis*. *Invertebrate Reproduction and Development*, *40*, 39–48.

Diemert D. J., Pinto A. G., Freire J., Jariwala A., Santiago H., Hamilton R. G., Periago M. V., Loukas A., Tribolet L., Mulvenna J., Correa-Oliveira R., Hotez P. J., Bethony J. M. (2012). Generalized urticaria induced by the Na-ASP-2 hookworm vaccine: implications for the development of vaccines against helminths. *The Journal of allergy and clinical immunology*, *130*, 169–76.e6.

Dinnik J. A., Dinnik N. N. (1961). On the Morphology and Life History of *Fasciola nyanzae* Leiper, 1910 from the *Hippopotamus*. *Journal of Helminthology*, *35*, 53–62.

Donald K. M., Kennedy M., Poulin R., Spencer H. G. (2004). Host specificity and molecular phylogeny of larval Digenea isolated from New Zealand and Australian topshells (Gastropoda: Trochidae). *International journal for parasitology*, *34*, 557–568.

Dong Y., Simões M. L., Marois E., Dimopoulos G. (2018). CRISPR/Cas9 -mediated gene knockout of *Anopheles gambiae* FREP1 suppresses malaria parasite infection. *PLOS Pathogens*, *14*, e1006898.

Dvorak J. A., Jones A. W., KuhlmanH. H. (1961). Studies on the biology of *Hymenolepis microstoma* (Dujardin, 1845). *The Journal of parasitology*, *47*, 833–838.

Dvorak M., Schnegg R., Niederwanger M., Pedrini-Martha V., Ladurner P., Lindner H., Kremser L., Lackner R., Dallinger R. (2020). Cadmium pathways in snails follow a

complementary strategy between metallothionein detoxification and auxiliary inactivation by phytochelatins. *International Journal of Molecular Sciences*, *21*, 1–14.

El-Emam M. A., Madsen H. (1982). The effect of temperature, darkness, starvation and various food types on growth, survival and reproduction of *Helisoma duryi*, *Biomphalaria alexandrina* and *Bulinus truncatus* (Gastropoda: Planorbidae). *Hydrobiologia*, *88*, 265–275.

Elbaz T., Esmat G. (2013). Hepatic and Intestinal Schistosomiasis: Review. *Journal of Advanced Research*, *4*, 445–452.

Fitzguerald A. (2008). Abalone feed requirements. *Final Report for SEAFISH, South west abalone growers association*, 34.

Fu L., Niu B., Zhu Z., Wu S., Li W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics (Oxford, England)*, *28*, 3150–3152.

Gantz V. M., Bier E. (2015). The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science*, *348*, 442–444.

Gantz V. M., Jasinskiene N., Tatarenkova O., Fazekas A., Macias V. M., Bier E., James A. A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, E6736–E6743.

Giannelli A., Cantacessi C., Colella V., Dantas-Torres F., Otranto D. (2016). Gastropod-Borne Helminths: A Look at the Snail-Parasite Interplay. *Trends in Parasitology*, *32*, 255–264.

Giraldo-Calderón G. I., Emrich S. J., MacCallum R. M., Maslen G., Dialynas E., Topalis P., Ho N., Gesing S., Madey G., Collins F. H., Lawson D. (2015). VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic acids research*, *43*, D707-13.

Glöckner G., Noegel A. A. (2013). Comparative genomics in the Amoebozoa clade. *Biological reviews of the Cambridge Philosophical Society*, *88*, 215–225.

Grabherr M. G., Haas B. J., Yassour M., Levin J. Z., Thompson D. A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B. W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, *29*, 644–652.

Graf I. (1989). Plant Molluscicides. Edt. by K. E. Mott, WHO 1987. *Pharmazie in Unserer Zeit*, *18*, 28.

Grimes J. E., Croll D., Harrison W. E., Utzinger J., Freeman M. C., Templeton M. R. (2015). The roles of water, sanitation and hygiene in reducing schistosomiasis: A review. *Parasites and Vectors*, *8*, 1–16.

Guo L., Sui Z., Zhang S., Ren Y., Liu Y. (2015). Comparison of potential diatom 'barcode' genes (The 18S rRNA gene and ITS, COI, rbcL) and their effectiveness in discriminating and determining species taxonomy in the Bacillariophyta. *International Journal of Systematic and Evolutionary Microbiology*, *65*, 1369–1380.

Hansen E. L. (1979). Initiating a cell line from embryos of the snail *Biomphalaria glabrata*. *TCA manual / Tissue Culture Association*, *5*, 1009–1014.

Hewitson J. P., Maizels R. M. (2014). Vaccination against helminth parasite infections. *Expert Review of Vaccines*, *13*, 473–487.

Ho P. T., Nguyen H. Q., Kern E. M. A., Won Y. J. (2021). Locomotor responses to salt stress in native and invasive mud-tidal gastropod populations (Batillaria). *Ecology and Evolution*, *11*, 458–470.

Horii T., Arai Y., Yamazaki M., Morita S., Kimura M., Itoh M., Abe Y., Hatada I. (2014). Validation of microinjection methods for generating knockout mice by CRISPR/Cas-mediated genome engineering. *Scientific reports*, *4*, 4513.

Hota-Mitchell S., Siddiqui A. A., Dekaban G. A., Smith J., Tognon C., Podesta R. B. (1997). Protection against *Schistosoma mansoni* infection with a recombinant baculovirus-expressed subunit of calpain. *Vaccine*, *15*, 1631–1640.

Hu M., Chilton N. B., Gasser R. B. (2003). The mitochondrial genome of *Strongyloides stercoralis* (Nematoda) - idiosyncratic gene order and evolutionary implications. *International journal for parasitology*, *33*, 1393–1408.

Isaacs A. T., Li F., Jasinskiene N., Chen X., Nirmala X., Marinotti O., Vinetz J. M., James A. A. (2011). Engineered resistance to *Plasmodium falciparum* development in transgenic *Anopheles stephensi*. *PLoS Pathogens*, *7*, e1002017.

Ishino Y., Shinagawa H., Makino K., Amemura M., Nakatura A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology*, *169*, 5429–5433.

James A. A., Beerntsen B. T., De Lara Capurro M., Coates C. J., Coleman J., Jasinskiene N., Krettli A. U. (1999). Controlling malaria transmission with genetically-engineered, *Plasmodium*-resistant mosquitoes: Milestones in a model system. *Parassitologia*, *41*, 461–471.

Jansen R., Embden J. D. A. van, Gaastra W., Schouls L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology*, *43*, 1565–1575.

Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, *337*, 816–821.

Jobin W. R. (1970). Control of *Biomphalaria glabrata* in a small reservoir by fluctuation of the water level. *The American journal of tropical medicine and hygiene*, *19*, 1049–1054.

Jones, H. (1993). Water velocity as a control of aquatic snails in concrete canal systems for irrigation. Loughborough University. Thesis. https://hdl.handle.net/2134/11708

Joubert P. H., Pretorius S. J. (1985). The effect of constant temperatures on the hatching of eggs and survival of the freshwater snail *Biomphalaria glabrata* (Say). *Journal of the Limnological Society of Southern Africa*, *11*, 75–77.

Junker K., Lhermitte-Vallarino N., Barbuto M., Ineich I., Wanji S., Bain O. (2010). New species of *Rhabdias* (Nematoda: Rhabdiasidae) from Afrotropical anurans, including molecular evidence and notes on biology. *Folia Parasitologica*, *57*, 47–61.

Kang H., Eun G., Ha J., Park N., Choi I., Kim D. (2017). First report of a bacteriovorus nematode , *Diploscapter coronatus* (Cobb , 1893) Cobb , 1913 (Rhabditida : Diploscapteridae), in Korea, *6*, 177–180.

Kappagoda S., Ioannidis J. P. A. (2014). Prevention and control of neglected tropical diseases: overview of randomized trials, systematic reviews and meta-analyses. *Bulletin of*

the World Health Organization, *92*, 356-366C.

Katoh K., Standley D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, *30*, 772–780.

Khallaayoune K., Madsen H., Laamrani H. (1998). Evaluation of three methods to control *Bulinus truncatus*, the intermediate host of *Schistosoma haematobium* in an irrigation scheme, Tessaout-Amont, Morocco. *Acta Tropica*, *69*, 51–63.

Kikuchi T., Cotton J. A., Dalzell J. J., Hasegawa K., Kanzaki N., McVeigh P., Takanashi T., Tsai I. J., Assefa S. A., Cock P. J. A., Otto T. D., Hunt M., Reid A. J., Sanchez-Flores A., Tsuchihara K., Yokoi T., Larsson M. C., Miwa J., Maule A. G., Sahashi N., Jones J. T., Berriman M. (2011). Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS pathogens*, *7*, e1002219.

Kim J. R., Hayes K. A., Yeung N. W., Cowie R. H. (2014). Diverse Gastropod Hosts of *Angiostrongylus cantonensis*, the Rat Lungworm, Globally and with a Focus on the Hawaiian Islands. *PLOS ONE*, *9*, e94969.

Klug A., Rhodes D. (1987). Zinc fingers: a novel protein fold for nucleic acid recognition. *Cold Spring Harbor symposia on quantitative biology*, *52*, 473–482.

Kuang S., Regnier M., Goldberg J. I. (2002). Long-term culture of decapsulated gastropod embryos: a transplantation study. *The Biological bulletin*, *203*, 278–288.

Kuroda R., Abe M. (2020). The pond snail *Lymnaea stagnalis*. *EvoDevo*, *11*, 1–10.

Kuroda R., Endo B., Abe M., Shimizu M. (2009). Chiral blastomere arrangement dictates zygotic left-right asymmetry pathway in snails. *Nature*, *462*, 790–794.

Kuroda R., Fujikura K., Abe M., Hosoiri Y., Asakawa S., Shimizu M., Umeda S., Ichikawa F., Takahashi H. (2016). Diaphanous gene mutation affects spiral cleavage and chirality in snails. *Scientific Reports*, *6*, 1–15.

Kyrou K., Hammond A. M., Galizi R., Kranjc N., Burt A., Beaghton A. K., Nolan T., Crisanti A. (2018). A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nature Biotechnology*, *36*, 1062–1066.

Lardans V., Dissous C. (1998). Snail Control Strategies for Reduction of Schistosomiasis Transmission. *Parasitology Today*, *14*, 413–417.

Le T H, Blair D., McManus D. P. (2001). Complete DNA sequence and gene organization of the mitochondrial genome of the liverfluke, *Fasciola hepatica* L. (Platyhelminthes; Trematoda). *Parasitology*, *123*, 609–621.

Le Thanh H, Blair D., McManus D. P. (2002). Mitochondrial genomes of parasitic flatworms. *Trends in parasitology*, *18*, 206–213.

Lebonah D. E., Dileep A., Chandrasekhar K., Sreevani S., Sreedevi B., Pramoda Kumari J. (2014). DNA Barcoding on Bacteria: A Review. *Advances in Biology*, *2014*, 541787.

Leger M. M., Kolisko M., Kamikawa R., Stairs C. W., Kume K., Čepička I., Silberman J. D., Andersson J. O., Xu F., Yabuki A., Eme L., Zhang Q., Takishita K., Inagaki Y., Simpson A. G. B., Hashimoto T., Roger A. J. (2017). Organelles that illuminate the origins of *Trichomonas* hydrogenosomes and *Giardia* mitosomes. *Nature Ecology & Evolution*, *1*, 92.

Leinonen R., Sugawara H., Shumway M., Collaboration I. N. S. D. (2011). The sequence read archive. *Nucleic acids research*, *39*, D19–D21.

Leray M., Ho S.-L., Lin I.-J., Machida R. J. (2018). MIDORI server: a webserver for taxonomic assignment of unknown metazoan mitochondrial-encoded sequences using a curated database. *Bioinformatics*, *34*, 3753–3754.

Li D., Liu C.-M., Luo R., Sadakane K., Lam T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*, 1674–1676.

Lima M. G., Augusto R. de C., De Vasconcellos M. C., Mello-Silva C. C., Pinheiro J. (2012). Metabolic changes in *Biomphalaria glabrata* infected with *Schistosoma mansoni* exposed to latex of *Euphorbia milii* solution versus times of preparation. *Journal of Natural Products*, *5*, 222–232.

Lino C. A., Harper J. C., Carney J. P., Timlin J. A. (2018). Delivering crispr: A review of the challenges and approaches. *Drug Delivery*, *25*, 1234–1257.

Littlewood D. T. J., Bray R. A. (2000). "The Digenea". Interrelationships of the Platyhelminthes. Systematics Association Special Volume. Vol. 60 (1 ed.). CRC. pp. 168–185. ISBN 978-0-7484-0903-7

Liu C., Ren Y., Li Z., Hu Q., Yin L., Wang H., Qiao X., Zhang Y., Xing L., Xi Y., Jiang F., Wang S., Huang C., Liu B., Liu H., Wan F., Qian W., Fan W. (2021). Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic-terrestrial transition. *Molecular ecology resources*, *21*, 478–494.

Liu G.-H., Wang Y., Song H.-Q., Li M.-W., Ai L., Yu X.-L., Zhu X.-Q. (2013). Characterization of the complete mitochondrial genome of *Spirocerca lupi*: sequence, gene organization and phylogenetic implications. *Parasites & vectors*, *6*, 45.

Liu Q., Wei F., Liu W., Yang S., Zhang X. (2008). Paragonimiasis: an important food-borne zoonosis in China. *Trends in parasitology*, *24*, 318–323.

Liu S., Roellig D. M., Guo Y., Li N., Frace M. A., Tang K., Zhang L., Feng Y., Xiao L. (2016). Evolution of mitosome metabolism and invasion-related proteins in Cryptosporidium. *BMC genomics*, *17*, 1006.

Lo N. C., Gurarie D., Yoon N., Coulibaly J. T., Bendavid E., Andrews J. R., King C. H. (2018). Impact and cost-effectiveness of snail control to achieve disease control targets for schistosomiasis. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E584–E591.

Loreau M., Baluku B. (1991). Shade as a means of ecological control of *Biomphalaria pfeifferi*. *Annals of tropical medicine and parasitology*, *85*, 443–446.

Louradour I., Ghosh K., Inbar E., Sacks D. L. (2019). CRISPR/Cas9 Mutagenesis in *Phlebotomus papatasi*: the Immune Deficiency Pathway Impacts Vector Competence for *Leishmania major*. *mBio*, *10*.

Lu X.-T., Gu Q.-Y., Limpanont Y., Song L.-G., Wu Z.-D., Okanurak K., Lv Z.-Y. (2018). Snail-borne parasitic diseases: an update on global epidemiological distribution, transmission interruption and control methods. *Infectious diseases of poverty*, *7*, 28.

Maier T., Wheeler N. J., Namigai E. K. O., Tycko J., Grewelle R. E., Woldeamanuel Y., Klohe K., Perez-Saez J., Sokolow S. H., De Leo G. A., Yoshino T. P., Zamanian M., Reinhard-Rupp J. (2019). Gene drives for schistosomiasis transmission control. *PLoS*

neglected tropical diseases, *13*, e0007833.

Martin-Alonso A., Abreu-Yanes E., Feliu C., Mas-Coma S., Bargues M. D., Valladares B., Foronda P. (2015). Intermediate Hosts of *Angiostrongylus cantonensis* in Tenerife, Spain. *PLOS ONE*, *10*, e0120686.

McCullough F. S., Gayral P., Duncan J., Christie J. D. (1980). Molluscicides in schistosomiasis control. *Bulletin of the World Health Organization*, *58*, 681–689.

Mehlhorn H. (Ed.). (2008). *Rhabdias bufonis*. In *Encyclopedia of Parasitology*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1243–1244.

Miller J. C., Holmes M. C., Wang J., Guschin D. Y., Lee Y.-L., Rupniewski I., Beausejour C. M., Waite A. J., Wang N. S., Kim K. A., Gregory P. D., Pabo C. O., Rebar E. J. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature biotechnology*, *25*, 778–785.

Miller J. C., Tan S., Qiao G., Barlow K. A., Wang J., Xia D. F., Meng X., Paschon D. E., Leung E., Hinkley S. J., Dulay G. P., Hua K. L., Ankoudinova I., Cost G. J., Urnov F. D., Zhang H. S., Holmes M. C., Zhang L., Gregory P. D., Rebar E. J. (2011). A TALE nuclease architecture for efficient genome editing. *Nature biotechnology*, *29*, 143–148.

Moazeni M., Ahmadi A. (2016). Controversial aspects of the life cycle of *Fasciola hepatica*. *Experimental parasitology*, *169*, 81–89.

Moscou M. J., Bogdanove A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science (New York, N.Y.)*, *326*, 1501.

Mounsey K. E., Willis C., Burgess S. T. G., Holt D. C., McCarthy J., Fischer K. (2012). Quantitative PCR-based genome size estimation of the astigmatid mites *Sarcoptes scabiei*, Psoroptes ovis and Dermatophagoides pteronyssinus. *Parasites & vectors*, *5*, 3.

Nguyen L.-T., Schmidt H. A., von Haeseler A., Minh B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, *32*, 268–274.

Ofoezie I. E., Asaolu S. O. (1997). Water level regulation and control of schistosomiasis transmission: a case study in Oyan Reservoir, Ogun State, Nigeria. *Bulletin of the World Health Organization*, *75*, 435–441.

Ouma J. H., Sturrock R. F., Klumpp R. K., Kariuki H. C. (1989). A comparative evaluation of snail sampling and cercariometry to detect *Schistosoma mansoni* transmission in a large-scale, longitudinal field-study in Machakos, Kenya. *Parasitology*, *99 Pt 3*, 349–355.

Pacheco G., Jansen J. (1951). Destruição de cercárias de *Schistosoma mansoni* com sabões associados a materias corantes. *Brasil-médico*, *65*, 301–304.

Pare Toe L., Barry N., Ky A. D., Kekele S., Meda W., Bayala K., Drabo M., Thizy D., Diabate A. (2021). Small-scale release of non-gene drive mosquitoes in Burkina Faso: from engagement implementation to assessment, a learning journey. *Malaria journal*, *20*, 395.

Park J.-K., Kim K.-H., Kang S., Jeon H. K., Kim J.-H., Littlewood D. T. J., Eom K. S. (2007). Characterization of the mitochondrial genome of *Diphyllobothrium latum* (Cestoda: Pseudophyllidea) - implications for the phylogeny of eucestodes. *Parasitology*, *134*, 749–759.

Perry K. J., Henry J. Q. (2015). CRISPR/Cas9-mediated genome modification in the mollusc, *Crepidula fornicata*. *Genesis*, *53*, 237–244.

Piedrafita D., Preston S., Kemp J., de Veer M., Sherrard J., Kraska T., Elhay M., Meeusen E. (2013). The Effect of Different Adjuvants on Immune Parameters and Protection following Vaccination of Sheep with a Larval-Specific Antigen of the Gastrointestinal Nematode, *Haemonchus contortus*. *PLoS ONE, 8*.

Poulin R., Morand S. (2000). The diversity of parasites. *Quarterly Review of Biology*, *75*, 277–293.

Pulendran B., Ahmed R. (2011). Immunological mechanisms of vaccination. *Nature immunology*, *12*, 509–517.

Raffaele S., Kamoun S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature reviews. Microbiology*, *10*, 417–430.

Ratnasingham S., Hebert P. D. N. (2007). bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular ecology notes*, *7*, 355–364.

Richards C. S., Knight M., Lewis F. A. (1992). Genetics of *Biomphalaria glabrata* and its effect on the outcome of *Schistosoma mansoni* infection. *Parasitology today (Personal ed.)*, *8*, 171–174.

Rickard M. D., Williams J. F. (1982). Hydatidosis/cysticercosis: immune mechanisms and immunization against infection. *Advances in parasitology*, *21*, 229–296.

Robinson J. T., Thorvaldsdóttir H., Winckler W., Guttman M., Lander E. S., Getz G., Mesirov J. P. (2011). Integrative genomics viewer. *Nature biotechnology*, *29*, 24–26.

Rödelsperger C., Streit A. and Sommer R. J. (2013). Structure, function and evolution of the nematode genome In eLS (Encyclopedia of Life Sciences). John Wiley and Sons, Chichester, UK. doi: 10.1002/978047001 5902.a0024603

Rohde K. (1994). The Minor Groups of Parasitic Platyhelminthes. In J. R. Baker & R. B. T.-A. in P. Muller (Eds.), (Vol. 33). Academic Press, 145–234.

Rollinson D., Knopp S., Levitz S., Stothard J. R., Tchuem Tchuenté L.-A., Garba A., Mohammed K. A., Schur N., Person B., Colley D. G., Utzinger J. (2013). Time to set the agenda for schistosomiasis elimination. *Acta tropica*, *128*, 423–440.

Rytkönen S., Vesterinen E. J., Leviäkangas T., Vatka E., Mutanen M., Välimäki P., Hukkanen M., Suokas M., Orell M. (2019). From feces to data : A metabarcoding method for analyzing consumed and available prey in a bird - insect food web, 631–639.

Schall V. T., Vasconcellos M. C., Rocha R. S., Souza C. P., Mendes N. M. (2001). The control of the schistosome-transmitting snail *Biomphalaria glabrata* by the plant Molluscicide *Euphorbia splendens var. hislopii* (syn milli Des. Moul): a longitudinal field study in an endemic area in Brazil. *Acta tropica*, *79*, 165–170.

Schiffer P. H., Robertson H. E., Telford M. J. (2018). Orthonectids Are Highly Degenerate Annelid Worms. *Current Biology*, *28*, 1970-1974.e3.

Schoch C. L., Seifert K. A., Huhndorf S., Robert V., Spouge J. L., Levesque C. A., Chen W. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 6241–6246.

Seligman L. M., Chisholm K. M., Chevalier B. S., Chadsey M. S., Edwards S. T., Savage J. H., Veillet A. L. (2002). Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic acids research*, *30*, 3870–3879.

Slyusarev G. S., Starunov V. V, Bondarenko A. S., Zorina N. A., Bondarenko N. I. (2020). Extreme Genome and Nervous System Streamlining in the Invertebrate Parasite *Intoshia variabili*. *Current biology : CB*, *30*, 1292-1298.e3.

Sokolow S. H., Wood C. L., Jones I. J., Swartz S. J., Lopez M., Hsieh M. H., Lafferty K. D., Kuris A. M., Rickards C., De Leo G. A. (2016). Global Assessment of Schistosomiasis Control Over the Past Century Shows Targeting the Snail Intermediate Host Works Best. *PLoS Neglected Tropical Diseases*, *10*, 1–19.

Spithill T. W., Smooker P. M., Copeman D. B. (1999). "*Fasciola gigantica*": Epidemiology, control, immunology and molecular biology.

Stockdale-Walden H. D., Slapcinsky J., Qvarnstrom Y., McIntosh A., Bishop H. S., Rosseland B. (2015). *Angiostrongylus cantonensis* in introduced gastropods in southern Florida. *Journal of Parasitology*, *101*, 156–159.

Thiengo S. C., Simões R. de O., Fernandez M. A., Maldonado A. (2013). *Angiostrongylus cantonensis* and rat lungworm disease in Brazil. *Hawai'i journal of medicine & public health : a journal of Asia Pacific Medicine & Public Health*, *72*, 18–22.

Tovar J., Fischer A., Clark C. G. (1999). The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Molecular microbiology*, *32*, 1013–1021.

Tsai I. J., Zarowiecki M., Holroyd N., Garciarrubio A., Sánchez-Flores A., Brooks K. L., Tracey A., Bobes R. J., Fragoso G., Sciutto E., Aslett M., Beasley H., Bennett H. M., Cai X., Camicia F., Clark R., Cucher M., De Silva N., Day T. A., Deplazes P., Estrada K., Fernández C., Holland P. W. H., Hou J., Hu S., Huckvale T., Hung S. S., Kamenetzky L., Keane J. A., Kiss F., Koziol U., Lambert O., Liu K., Luo X., Luo Y., Macchiaroli N., Nichol S., Paps J., Parkinson J., Pouchkina-Stantcheva N., Riddiford N., Rosenzvit M., Salinas G., Wasmuth J. D., Zamanian M., Zheng Y., Cai J., Soberón X., Olson P. D., Laclette J. P., Brehm K., Berriman M. (2013). The genomes of four tapeworm species reveal adaptations to parasitism. *Nature*, *496*, 57–63.

Urnov F. D., Rebar E. J., Holmes M. C., Zhang H. S., Gregory P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics*, *11*, 636–646.

Vale N., Gouveia M. J., Rinaldi G., Brindley P. J., Gärtner F., da Costa J. M. C. (2017). Praziquantel for Schistosomiasis: Single-Drug Metabolism Revisited, Mode of Action, and Resistance. *Antimicrobial Agents and Chemotherapy*, *61*, e02582-16.

Valente R., Robles M. D. R., Diaz J. I. (2020). Gastropods as intermediate hosts of *Angiostrongylus spp*. in the Americas: bioecological characteristics and geographical distribution. *Memorias do Instituto Oswaldo Cruz*, *115*, e200236.

van Emden H. M., Kroon C. C. M., Schoeman E. N., van Seventer H. A. (1974). The toxicity of some detergents tested on *Aëdes aegypti L.*, *Lebistes reticulatus peters*, and *Biomphalaria glabrata* (Say). *Environmental Pollution (1970)*, *6*, 297–308.

Wang Q. P., Wu Z. D., Wei J., Owen R. L., Lun Z. R. (2012). Human *Angiostrongylus cantonensis*: An update. *European Journal of Clinical Microbiology and Infectious Diseases*, *31*, 389–395.

Weinstein S. B., Kuris A. M. (2016). Independent origins of parasitism in Animalia. *Biology letters*, *12*, 20160324.

WHO (2002). World Malaria Report 2020. 30 November 2020. www.who.int/teams/global-

malaria-programme/reports/world- malaria-report-2020.

WHO (2002). Prevention and control of schistosomiasis and soil-transmitted helminthiasis. Geneva: World Health Organization

WHO (2006). Preventive chemotherapy in human helminthiasis: coordinated use of anthelmintic drugs in control interventions: a manual for health professionals and programme managers. Geneva: World Health Organization.

WHO (2012). Accelarating work to overcome the global impact of neglected tropical diseases: a roadmap for implementation. Geneva: World Health Organization.

WHO (2017). Guidelines: alternative mass drug administration regimens to eliminate lymphatic filariasis. Geneva: World Health Organization.

WHO (2018). Meeting report of the WHO Evidence Review Group on mass drug administration for malaria: 11–13, Geneva, Switzerland. Geneva: World Health Organization; p. 2019.

WHO (2021). Wash, sanitation and hygiene. https://www.who.int/news-room/fact-sheets/detail/sanitation.

Yang D., Fu Y., Wu X., Xie Y., Nie H., Chen L., Nong X., Gu X., Wang S., Peng X., Yan N., Zhang R., Zheng W., Yang G. (2012). Annotation of the Transcriptome from *Taenia pisiformis* and Its Comparative Analysis with Three Taeniidae Species. *PLOS ONE*, *7*, e32283.

Yang H., Wang H., Shivalila C. S., Cheng A. W., Shi L., Jaenisch R. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell*, *154*, 1370–1379.

Yoshino T. P., Bickham U., Bayne C. J. (2013). Molluscan cells in culture: Primary cell cultures and cell lines. *Canadian Journal of Zoology*, *91*, 391–404.

Zapata F., Wilson N. G., Howison M., Andrade S. C. S., Jörger K. M., Schrödl M., Goetz F. E., Giribet G., Dunn C. W. (2014). Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proceedings of the Royal Society B: Biological Sciences*, *281*.

Zheng S., Zhu Y., Zhao Z., Wu Z., Okanurak K., Lv Z. (2017). Liver fluke infection and cholangiocarcinoma: a review. *Parasitology research*, *116*, 11–19.

Zuber J. A., Takala-Harrison S. (2018). Multidrug-resistant malaria and the impact of mass drug administration. *Infection and drug resistance*, *11*, 299–306.