UNIVERSITY OF LIVERPOOL

Bioinformatic analysis of *Fasciola hepatica* genome

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

By

Olukayode Olugbenga Daramola

March 2023

ABSTRACT

Fasciolosis is caused by liver flukes: *F. hepatica*, and a sister species – *F. gigantica*. A growing concern with controlling the disease is resistance to triclabendazole (TCBZ), the only drug shown to kill both adult and immature liver flukes. Currently, *F. hepatica* mechanism of resistance to TCBZ is not clearly understood and there is no effective commercially available vaccine. Previous work proposed three mechanisms associated with TCBZ mode of action and resistance: tubulin binding activity, drug uptake mechanisms, and drug metabolism mechanism. Exploring evolutionary forces acting on *F. hepatica* genes associated with TCBZ mode of action and resistance could explain how the parasite develops resistance to the drug, enable identification of potential drug targets, and facilitate development of new drugs.

A re-annotation of the current *F. hepatica* genome was done using an updated version of the published *F. hepatica* draft genome (assembly GCA_000947175.1, BioProject PRJEB6687). Subsequently, the current annotation (*Fasciola*_10x_pilon, GCA_900302435.1 WormBase Parasite Version 15) was compared and critically assessed with the newly re-annotated version. Using coding sequences (CDS) of three well-described annotated gene families, manual validation of the annotation was done. A total of 15,879 *F. hepatica* genes were identified in this project compared to the 9,401 genes in the current annotation, while differences noticed in both annotations include gene fragmentation, missing exons, and missing genes.

F. hepatica gene family members belonging to each of the three proposed mechanism of action of TCBZ action and resistance, and their trematode orthologous sequences were compiled. The gene families studied include tubulins, ATP-binding cassette transporters (ABC), AC, RAS, ADP ribosylation factor, cytochrome P450 (CYP450), GSTs, and Fatty Acid Binding Proteins (FABPs). Signals indicative of positive selection was identified using Phylogenetic Analysis by Maximum Likelihood (PAML) and McDonald and Kreitman test (MKtest). PAML branch-site model testing identified 1 alpha tubulin, 1 delta tubulin, 5 ABC genes, 9 RAS genes, and 4 ADP ribosylation factor genes with statistically significant sites under positive selection. While the MKtest analysis identified 2 RAS genes and 1 AC genes under positive selection.

The expression profile of the genes associated with TCBZ mode of action was assessed across *F. hepatica* life stages. Findings indicate that tubulin gene expression was elevated in metacercariae and newly excysted juveniles (NEJs), with a peak expression pattern noticed in NEJs 1 hour post excystment, with levels reducing in flukes 21 days post excystment. Similarly, in genes associated with TCBZ uptake, expression was predominantly raised in metacercariae and NEJs, while gene expression gradually reduced towards fluke maturity.

The effect of TCBZ on *F. hepatica* was investigated in experimentally infected sheep. Parasite response to the drug in TCBZ resistant and susceptible *F. hepatica* isolates was compared in sheep infected and treated with the drug. TCBZ treatment induced gene expression patterns were noticed in 72% (90 out 125 genes, $P < 0.05$) of all the genes assessed (excluding unexpressed genes and constitutively expressed genes). Findings in this study indicate TCBZ administration affects multiple mechanisms in the parasite. Therefore, this confirms that all the three proposed TCBZ mode of action and resistance mechanisms in *F. hepatica* could be implicated in drug TCBZ resistance.

ACKNOWLEDGMENTS

I would like to my sincere gratitude to my supervisors, Prof. Steve Paterson, and Prof. Jane Hodgkinson for their guidance and support throughout my PhD project. Their patience, understanding, and constructive comments during my research is impressive. Without my supervisors, it would be impossible to produce this thesis.

Special thanks to my internal advisers, Professor Alistair Darby, Dr Seth Barribeau, and Professor Ben Makepeace for their time, help and positive words during my year progress evaluation meetings.

Dr Nicola Beesley deserves to be mentioned for her guidance at the start, and during my PhD research. I am particularly grateful for the timely advice and assistance she always give whenever I reach out to her for explanation about datasets used in this project.

I would like to thank the Institute of Integrative Biology (now Institute of Infection, Veterinary and Ecological Sciences) for awarding me the studentship that facilitated my PhD studies.

Finally, I would like to appreciate my colleagues, friends and my family for their support and understanding, especially my wife Dr Boluwatife Daramola for her companionship and editorial assistance.

DECLARATION

I declare that this thesis was composed by myself, that work contained herein is my own except were explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

In chapter 4, RNA Sequence dataset used for investigating response to triclabendazole treatment in drug resistant and susceptible Fasciola hepatica isolates were from an experiment carried out by Nicola Beesley.

In chapter 2, *RNAseq data from 3 samples of 21 days old juveniles referred to as RNAseq dataset 2 (also referred to as erins_reads), were from Erin Mccammick of the Queen's University Belfast.*

Olukayode Daramola

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

cm	Centimetre
US	United States of America
$	Dollars
~	approximately
Kg	kilogram
%	Percentage
UK	United Kingdom
TCBZ	Triclabendazole
β	Beta
α	Alpha
k	thousand
IgG	Immunoglobulin G
AC	Adenylate Cyclase
FABP	Fatty Acid Binding Protein
GST	Glutathione S Transferase
PGP	P-glycoprotein
PKA	Protein Kinase A
FMO	Flavin mono-oxygenase
NEJ	Newly excysted juvenile
ω	Omega
Seq	Sequence
ELISA	Enzyme-linked immunosorbent assay
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
DNA	Deoxyribonucleic acid
ORF	Open reading frames
CDS	Coding sequences
mRNA	Messenger RNA
etc	*Et cetera*
i.e	that is
et al.	and others
spp	species
Gb	Giga base
NEJs	Newly excysted juveniles
NCBI	National Centre for Biotechnology Information
EBI	European Bioinformatics Institute
WP	Worm Base Parasite
ESTs	Expressed sequence tags
PacBio	Pacific Biosciences
ID	Identifier

LTR	Long terminal repeats
LINEs	Long interspersed elements
SINEs	Short interspersed elements
GC	Guanine-cytosine
Orthogroup	orthologous group
rpkm	Reads Per Kilobase Million
ND	No data
syn	synonym
gff	General Feature Format

CHAPTER 1

Literature Review on *Fasciola hepatica*

1.1. Platyhelminths overview

The phylum Platyhelminthes (commonly referred to as flatworms) are a wide group of dorsoventrally flattened and bilaterally flattened, unsegmented and soft-bodied invertebrates. These worms do not have a coelom (body cavity), skeleton, or anus, and do not possess specific respiratory and circulatory systems. They do however have a blind gut which allows food to enter and exit, while nutrients and oxygen go through their bodies via diffusion (Collins, 2017). Platyhelminthes are broadly classified as either parasitic or non-parasitic. There are three mainly parasitic groups in the phylum; the Monogenea (predominantly ectoparasites and do not require intermediate hosts), the Cestoda, and the Trematoda (these two groups are endoparasites, more complex, possess two or three hosts, including a vertebrate host, in their life-cycle) (Kearn, 2018). There are about a total of 22,500 known platyhelminth species worldwide (Dettner, 2010), most of which go through various reproduction forms. However, hermaphroditism is predominant, in which case each individual possesses a fully complementary male and female reproductive system (Caira and Littlewood, 2013). The phylum Turbellaria is generally regarded as non-parasitic "free living" flatworms (or planarians). Most planarians live in fresh water, have elongated leaf-shaped bodies, and are ciliated. They are bilaterally symmetrical, lack coelom and anus, and are hermaphrodites (El-Bawab, 2020). A common example is *Schmidtea mediterranea* – a model organism commonly used to investigate epigenetic germ cell specification, due to their rapid regenerative ability (Zayas et al., 2005).

1.2. Trematodes

Trematodes are flatworms commonly known as flukes (Table 1.1). Their incomplete digestive tract begins at the oral sucker and ends in a blind intestine. There is no anus, thus wastes are regurgitated after digestion, while liquid wastes are passed out through flame cells (a specialised excretory cell). Structurally they have a protective tegument that protects them (for example from drugs, digestive enzymes and from host immunity), and aids absorption of nutrients and gaseous exchange (Bungiro and Cappello, 2004, Jones and Cappello, 2004). A trematode possesses a male and a female reproductive system in the same individual (Saari et al., 2019). Thus, self-fertilization can occur (this is common in *Fasciola hepatica, Fasciola gigantica*, *Fasciolopsis buski*, *Clornorchis sinensis*), or cross-fertilization between two worms (such as in *Paragonimus westermani*). However, schistosome species are dioecious, *i.e*. they have separate sexes (Bungiro and Cappello, 2004). Generally, flukes have a complex life cycle which includes at least two hosts, an intermediate host mollusc (such as snails) where asexual reproduction occurs, and a definitive host - typically a vertebrate, where sexual reproduction occurs (Singh et al., 2019), although the presence of more than one intermediate host is common in flukes (Mas and Bargues, 1997). In hermaphrodites such as *Fasciola* spp., male sexual maturity is achieved before the female, while developmental rates are hugely influenced by temperature (Borgsteede, 2011). Adult *F. gigantica* are the approximately 7.5 cm by 1.5 cm, *F. hepatica* is approximately 3.5 cm by 1.5 cm, *P. westermani* is approximately 1 cm by 0.5 cm,

Paramphistomum gracile is approximately 15 mm by 7 mm, while *Heterophyes heterophyes* is about 1.7 mm by 0.4 mm, while (Bungiro and Cappello, 2004, Panyarachun et al., 2013).

Trematodes include liver flukes (common ones are *F. hepatica*, *F. gigantica, Fascioloides magna*, *C. sinensis*, and *Opisthorchis* spp.) – these cause hepatic-related diseases, blood flukes – these invade the circulatory system, intestinal flukes (common ones are *Fp. buski*, *Echinostoma* spp., and *Paramphistomum cervi*) – they attack the gastrointestinal tract, and lung flukes (such as *P. westermani*) – these target the respiratory organs (Sharma and Anand, 1997).

Table 1.1: Summary of common trematodes and diseases caused

Trematode	Habitat	Disease caused	Common hosts
Fasciola hepatica	Liver	fasciolosis	Ruminants, humans
Fasciola gigantica	Liver	fasciolosis	Ruminants, humans
Fasciolopsis buski	Intestine	fasciolopsis	Humans, pigs
Fascioloides magna	Liver		Ruminants
Dicrocoelium dendriticum	Liver	dicrocoeliosis	Ruminants, Humans, pigs, rabbits
Clonorchis sinensis	liver	clonorchiasis	Humans
Schistosoma mansoni	Veins of intestine	schistosomiasis	Humans
Schistosoma japonicum	Veins of small Intestine, liver	schistosomiasis	Humans
Schistosoma haematobium	Veins of urinary bladder	schistosomiasis	Humans
Paragonimus westermani	Lung	paragonimiasis	Humans
Heterophyes heterophyes	Intestine	heterophyiasis	Humans, dog, cat
Echinostoma spp.	Intestine	echinostomiasis	Humans, rodents, birds, carnivores
Opisthorchis viverrini	liver	opisthorchiasis	Humans, cats, dogs

1.3. Liver flukes

Parasitic trematodes such as liver flukes are very important to a wide range of mammals. The definitive hosts include humans, livestock animals, rodents, etc, with parasites causing varying levels of infection in these hosts. Infections are worldwide in distribution. Liver flukes of huge economic importance include *F. hepatica* and *F. gigantica*; these cause fascioliases (syn. fasciolosis) in livestock, wildlife, and humans (Alvarez Rojas et al., 2014). *F. hepatica* for example, is well known for its high veterinary economic impact, zoonotic potential, and is classified as a neglected tropical disease by the World Health Organization (WHO, 2020). An estimated global economic loss of at least US $ 3.2 billion annually is associated with the disease (Spithill et al., 1999), with more than 600 million animals affected by the parasite (de Waal, 2016). Interestingly, *Fasciola* was identified as far back as 1547 in sheep and goat livers; Jean de Brie mentioned *Fasciola* in 1379, while its first record in a human was in 1760 (Rokni, 2014).

F. hepatica and *F. gigantica* are zoonotic and worldwide in distribution. *F. hepatica* appears to be prominent in temperate zones, while *F. gigantica* is important in Africa and Asia. Co-infection between both species is present (Amer et al., 2011, Waikagul et al., 2015, Beesley et al., 2015). Infection within *Fasciola* spp. could occur alone, while co-infection with helminths or other organisms (such as *Mycobacterium bovis*, the causative agent of tuberculosis) could occur. Co-infections have an important role in the *Fasciola*'s host immunomodulatory abilities, leading to higher risks in cattle (Byrne et al., 2019, Howell et al., 2020). Morphologically, *F. hepatica* and *F. gigantica* are quite similar, having similar life cycles involving lymnaeid snails, and have ruminants as primary definitive hosts. The only key difference is in intermediate host; while *F. hepatica* infects *Galba truncatula*, *F. gigantica* infects *Radix rubiginosa* and *R. natalensis*, although preference is largely due to the distribution of these intermediate hosts and adult flukes. *Galba truncatula* is common in temperate regions while *Radix rubiginosa* and *R. natalensis* are predominant in the tropics (Alvarez Rojas et al., 2014).

1.4. Economic importance of liver flukes

Liver fluke infection (predominantly *F. hepatica*) has a huge impact on livestock production, while the exact estimates are difficult to assess (Skuce and Zadoks, 2013). In cattle, the disease has been associated with decrease in milk production, with losses as high as 2 kg per day (Mezo et al., 2011), and marked illness in weaned calves. The disease also causes poor carcass quality in cattle. A rejection rate of 30% of cattle livers was reported in 2012 (Skuce and Zadoks, 2013). In Scotland, abattoir data revealed that slaughtered animals from 2005 to 2015 with history of liver fluke infection had lower cold weight, lower price, lower carcass conformity scores, reduced fat levels, and an overall drop in carcass value, thus impacting farmers pricing negatively (Sanchez-Vazquez and Lewis, 2013).

Liver fluke infection is arguably the most important parasitic disease among grazing livestock; unfortunately, most infections are inapparent, thus making production losses almost inevitable. Losses in sheep can occur 10 – 12 weeks post initial infection (before eggs

are noticeable in faeces during diagnosis), thus making management challenging (Love, 2017, Mitchell, 2002). In the UK, sheep are highly predisposed to acute infections, in which case the ingestion of many metacercaria mostly in early autumn leads to migratory juvenile tissue boring flukes, while cattle predominantly suffer chronic fasciolosis due to adult fluke (Skuce and Zadoks, 2013). Fasciolosis can cause blood loss, liver damage, and sudden death of infected sheep (Williams, 2020). Acute infections affect animal welfare and profitability negatively, however accurate assessment of negative impact is challenging due to poor records and reporting in farms (Skuce and Zadoks, 2013). Estimating the accurate economic impact of liver flukes is challenging. There is a lack of specific tools to assess these economic impacts at farm level to facilitate appropriate decisions to be made by farmers and the government (Beesley et al., 2017a).

1.5. *F. hepatica* life cycle

F. hepatica life cycle involves five key phases: egg, miracidium, cercaria, metacercaria, and adult. The life cycle is largely influenced by temperature and moisture which affect propagation of snails. Favourable weather (wet summer and mild winters) facilitates the increase of the intermediate host and flukes (Skuce and Zadoks, 2013). The fluke life cycle is well studied and described (Figure 1.1). Adult flukes live in the host bile ducts where they lay eggs that are passed out in faeces. Eggs embryonate in water hatching to produce miracidia. These miracidia are motile, thus finding the intermediate host snail (*Galba truncatula*) is favoured by the presence of wet conditions. In the snail, there is clonal expansion of parasite from sporocyst to rediae, finally to cercariae. Snails have been shown to shed cercariae up to 4 months, although shedding peaks in the first 4 weeks (Dreyfuss and Rondelaud, 1994). Cercariae have tails, which support movement in wet environment till they find vegetation after which they lose their tail and encyst to metacercaria, the infective stage. Digestion of fresh pasturage with viable metacercariae increases the severity of disease. Upon ingestion by a definitive host (typically ruminants or occasionally humans) the parasite excyst in the duodenum, penetrate the small intestine and migrate towards the liver. These newly excysted juveniles cause tissue damage as they travel and can interact with the host via their tegument. The glycocalyx in their tegument (which is shed as they mature into adults) is an immune defensive mechanism deployed by the parasite to prevent host immune cells from binding (González-Miguel et al., 2021). They eventually reach the bile ducts where they mature, although occasionally ectopic flukes occur, where other tissues such as lungs can be infected. Adult flukes can live in the bile duct in untreated hosts for years leading to a chronic disease state, especially in cattle (Love, 2017).

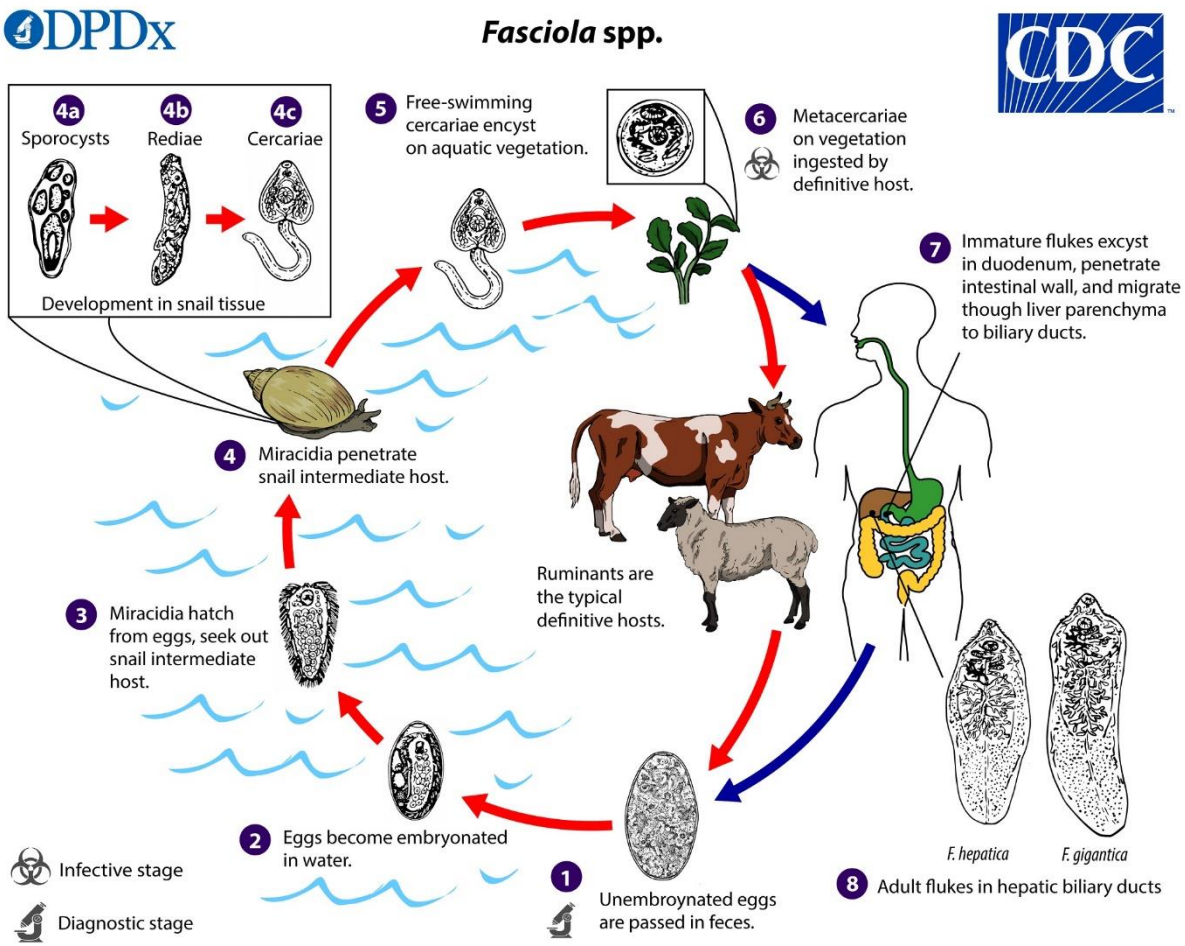

Figure 1.1: *F. hepatica* life cycle. *F. hepatica* eggs are passed in the faeces (1), embryonate in water and hatch to miracidia (2) which then infects the intermediate host (*Galba truncatula*) (4). In the snail, development occurs from sporocysts (4a), to rediae (4b), then to cercariae (4c); these are released from the snail (5), and subsequently encyst on vegetation as metacercariae – the infective stage (6). Metacercariae are ingested by the definitive host (7), the parasite excysts and these juveniles migrate through the small intestine towards liver, and finally mature in the bile ducts (8) (CDC, 2018).

1.6. Diagnosis and associated challenges

In the diagnosis of fasciolosis some factors are important, such as, early diagnosis to facilitate prompt treatment, accuracy of detection technique (sensitivity and specificity), cost effectiveness, time to get results, personnel required, ability to identify stages of infection, and ease of access to testing. It is important to point out that each diagnostic technique does comes with its disadvantages. Generally, diagnosis is by faecal egg count (FEC), however this is only beneficial about 8 weeks after infection due to the pre-patency period (Beesley et al., 2017a). This technique is challenging when there are low levels of eggs and can require repeated examinations. FEC however is relatively cheap and useful in developing countries. FEC technique is not reliable in chronic infections as egg shedding reduces in chronic infections due to low fluke burden as most flukes have died (Caravedo and Cabada, 2020) as it only detects patent infections (eggs can be detected due to high fluke burden on the host) (Tolan, 2011). Other techniques include quantitative microscopy (such Kato Katz test) and coprological sedimentation test (such as FLOTAC) (Cringoli et al., 2010); these require less training, are routine techniques and sensitive. However, these can be expensive and scarce when massive testing is needed (Caravedo and Cabada, 2020, Lopez et al., 2016, Cringoli et al., 2017). The use of diagnostic imaging approaches in humans has been reported. For example, computerised tomography was used to identify an abscess in a *F. hepatica* infected liver (Dusak et al., 2012, Behzad et al., 2014), but its use in livestock farming is not realistic. Diagnosis is also possible on physical examination at necropsy, to identify adult flukes in tissues.

In recent times immunological and molecular techniques are used, these require a higher level of personnel training. The research on various types of immunodiagnostic techniques is extensive (Alvarez Rojas et al., 2014); however enzyme-linked immunosorbent assay (ELISA) is generally favoured. Immunodiagnostic approaches rely on assaying *Fasciola* antigens (such as excretory/secretory (ES) antigens, cysteine proteases, somatic antigens, tegument antigens, etc in the blood, bile, and faeces) and antibodies stimulated in the hosts such as parasite specific IgG in serum. A major challenge with antibody testing is cross-reactivity. However, molecular methods are highly specific and sensitive, able to identify patent infection, and are increasingly affordable although still lack standardised protocols across laboratories which reduces their reproducibility. The methods are generally inaccessible to an average farmer, especially where there is limited access to laboratories providing these services (Beesley et al., 2017a).

1.7. Management of *F. hepatica*

A combination of various structured measures is often utilized to adequately manage liver fluke infections. These often include non-therapeutic and therapeutic methods. The former includes management of grazing, this involves ensuring heavily contaminated pastures are not grazed on, and rotation of grazing areas to break the parasite life cycle. Other measures include farm practices such as improved biosecurity practices, quarantine of animals introduced into the farm, adequate testing, and isolation of animals. Control of the

intermediate snail host habitat has been suggested although this is quite difficult to achieve (Williams, 2020). Therapeutic approach involves effective use of flukicides either prophylactically or in active infections. While there are many flukicides currently in use, challenges range from variable treatment coverage, high re-infections levels, and variable effectiveness of these drugs (McManus, 2020). Drug effectiveness vary with respect to the type of drug (or combination of drugs) used and the stage of parasite targeted. Despite the increasing research centred on understanding the disease pathogenesis and management, reports indicate an increase in the prevalence of the disease in recent times. A developing challenge with controlling the disease is the resistance to triclabendazole (TCBZ), the only drug shown to kill adult and immature liver flukes (Boray et al., 1983). Increasing incidences of drug resistance poses a great threat to livestock production worldwide and its control in humans (Kelley et al., 2016, Fairweather et al., 2020). Drug resistance is mostly attributed to over-dependence on TCBZ; findings indicate the parasite is able to evolve rapidly (Cwiklinski et al., 2015). Thus, evolutionary changes in the parasite could also play a key role in drug resistance. Therefore, research in the last decade has focused on finding effective alternative measures to control the parasite.

1.8. Flukicides and their usage in *F. hepatica* control

The management of fluke infections using chemical agents is generally accepted in farms, although use of these various drugs presents some limitations. Depending on various factors such as stage of parasite targeted, host species, phase of infection, and availability; flukicides or a combination of these drugs are strategically used. Drugs used include TCBZ, albendazole, ricobendazole, nitroxynil, closantel, and clorsulon. TCBZ is the only drug capable of killing the juvenile parasites; with up to 92 – 98% effectiveness against 1 week old and 100% effectiveness against 6 weeks old juveniles (Boray et al., 1983). It is important to be able to kill juveniles because, post excystment in the duodenum, their migratory movement through the intestines and liver to the bile duct damage host tissues and cause acute infection (González-Miguel et al., 2021). Also, killing adult flukes is important because, their hard outer tegument surface spines are damage host vasculature during feeding, leading to additional tissue damage (Lalor et al., 2021). The growing prevalence of resistance to TCBZ has raised serious concerns to livestock farmers. TCBZ resistance has been documented in at least 30 locations scattered across the world, 4 of which were in humans (Kelley et al., 2016, Fairweather et al., 2020). Reports of resistance in humans further highlight the zoonotic concerns particularly in countries with high incidences. Although TCBZ resistance is widespread, few reports of resistance to other flukicides such as closantel (Novobilský and Höglund, 2015), nitroxanide (Martínez-Valladares et al., 2010), clorsulon (Robles-Pérez et al., 2013), and albendazole (Ceballos et al., 2019) have been reported. However field reports indicate a predominance of TCBZ resistance in *F. hepatica*, especially in sheep, and to a large extent in cattle (Fairweather et al., 2020). This has led to use of various drug combinations for effective results (Love, 2017).

Drug combinations in liver fluke control are commonly used, although effectiveness can be reduced in some cases. There is inadequate scientific data to validate drug combinations use in liver fluke control (Skuce and Zadoks, 2013). Drug combinations vary in form, some are administered orally, some injectables; these variations pose a form of limitation to their usage. Other factors to be considered in drug combinations include the safety margin of these drugs singly and when combined, availability, and resistance. In recent times there has been reports of resistance to some of these drug combinations scattered globally (Fairweather et al., 2020). However, TCBZ resistance is predominantly of interest due to overreliance on the drug. This is understandably so considering the success attributed to the drug in controlling *Fasciola* spp. Beyond selection of appropriate choice of drug (or combination of drugs) depending on which stage of infection is targeted, veterinarians advise farmers is consider correct dosing of drugs to avoid treatment failure, rotation of anthelmintics and strategic timing of treatment with respect to weather, and repeated testing of animals. These when appropriately used can reduce development of drug resistance (Fairweather and Boray, 1999). Currently, the research community has focused on understanding the mechanism of drug resistance with the aim to control the parasite.

1.9. Mechanism of drug action and resistance

TCBZ mode of action is complex in the parasite and in the host, so pinpointing TCBZ mechanism of action is challenging. This is because currently at least three mechanisms of TCBZ action have been identified (Fairweather et al., 2020). TCBZ is a benzimidazole derivative, which is believed to be ingested by flukes or absorbed into flukes by diffusion across fluke tegmental syncytium. The drug was previously believed to bind to tubulins (particularly β-tubulins) based on drug effects such as parasite morphological integrity, feeding and digestion processes. Evidence indicates that TCBZ administration causes damage and loss of parasite tegument, loss of immunostaining ability in the tegument syncytium and disruption of mitosis in the vitelline and spermatogenic cells (McConville et al., 2006, Robinson et al., 2004). However, TCBZ binding to β-tubulins is questionable because in nematodes, benzimidazole resistance is associated with F200Y/E198A or F167Y mutations in β-tubulins and this not the case in liver flukes (Wolstenholme et al., 2004). Despite the usage of TCBZ extensively, its mode of action still eludes researchers, thus making understanding how flukes develop resistance more complicated.

Recent findings have identified genes that play a role in the mechanism of action and potentially resistance to TCBZ. These findings, some of which influence biological systems in the parasite either directly or indirectly are changing the way we look at TCBZ action, resistance, and potential identification of new drug targets. Genes investigated include (Table 1.2) adenylate cyclase (AC), Ras, tubulins, cytochrome P450 enzyme, P-glycoprotein (PGP), glutathione S-transferase (GST), and Fatty acid binding proteins (FABP) (Radio et al., 2018, Chemale et al., 2006, Ryan et al., 2008, Morphew et al., 2016). TCBZ for example was found to inhibit AC activity and/or inhibit the relationship between Ras and AC activity, leading to activation of stress-related response (Lee et al., 2013). This becomes relevant

because Ras-AC-protein kinase A (PKA) pathway (a nutrient sensory system designed to regulate metabolism, cell division, stress response, etc), when inactivated increases resistance to a stress factor (Lee et al., 2013); in this case the drug - TCBZ.

Further studies in *F. hepatica* TCBZ susceptible and resistant isolates, indicates that AC activity is reduced in TCBZ resistant isolates (Radio et al., 2018), thus highlighting how TCBZ affect the AC function via the cAMP pathways. There is an increasing interest in the role of various metabolism processes in TCBZ action. PGP activity, flavin mono-oxygenase and cytochrome enzyme pathways have been recently implicated in TCBZ resistance. Findings indicate that PGP inhibitors such as ivermectin and R (+)-verapamil affects uptake and activity of TCBZ in both susceptible and resistant isolates, with drug uptake decreased in TCBZ susceptible flukes, while resistance was reversed in presence of R (+)-verapamil. Inhibitors of PGP, flavin mono-oxygenases (FMO) such as methimazole, and cytochrome P450s such as ketoconazole; initiate an increased damage of tegument in TCBZ resistant flukes, although the FMO enzyme system is thought to be a more important system in the metabolism of TCBZ when compared to the cytochrome P450 (Fairweather et al., 2020, Radio et al., 2018, Meaney et al., 2013, Peachey et al., 2017, Devine et al., 2012). Similarly, fatty acid binding proteins (FABPs) have also been implicated in liver fluke biology. Recombinant FABPs (such as Fh12 and Fh15) have been identified as potential vaccine candidates and has been demonstrated to provide protection against *F. hepatica* infection (Muro et al., 2007, Ramos-Benítez et al., 2017, Casanueva et al., 2001). FABPs are localised in tegument and play a key role as a carrier of lipids, transporting them within the parasite (Ramos-Benítez et al., 2017). Findings indicate that FABP activity is significantly increased in TCBZ resistant adult liver flukes and has been associated with praziquantel drug binding in the trematode - *Schistosoma japonicum*. FABPs thus are important as for drug sequestration; however, despite the increasing evidence that support the importance of FABPs in *F. hepatica* biology, the knowledge of this small group of genes is little (Morphew et al., 2016). Another well studied group of genes are the glutathione-S-transferases (GSTs) These play a major role in the phase II detoxification in parasitic worms, particularly xenobiotic detoxification, ligand binding and transport, with TCBZ resistant flukes significantly exhibiting higher expression levels of glutathione-S-transferases compared to TCBZ susceptible clones (Scarcella et al., 2012). This has led to a belief that conjugation of TCBZ metabolites to GSTs is involved in TCBZ resistance via alteration of its metabolism (Fairweather et al., 2020).

Understanding TCBZ resistance has mostly focused on three areas (Table 1.2). This includes tubulin binding, altered drug uptake and drug metabolism (Fairweather et al., 2020); particularly investigating gene families of interest that play important roles in *F. hepatica* biology and TCBZ action. Understanding TCBZ resistance mechanism could facilitate understanding how the parasite responds to exposure to TCBZ and other anthelmintics. A better understanding of parasite response to various anthelmintics could enable understanding drug combination options and vaccine failures. This could potentially aid

identification of new drugs, drug combinations, and hopefully develop vaccines that are more effective. Pinpointing the mechanism of TCBZ is challenging, as it is unclear if resistance due to multiple co-interacting pathways or a single pathway.

Table 1.2: *Fasciola hepatica* candidate gene families implicated in TCBZ resistance

Mechanism of Drug Action	Gene Families	References
Altered tubulin binding	Tubulins	(Ryan et al., 2008, Robinson et al., 2002, Robinson et al., 2001, Stitt et al., 1992, Robinson et al., 2004)
Altered drug uptake	ATP-binding cassette transporters (ABCs)	(Peachey et al., 2017, Savage et al., 2013)
	Adenylate Cyclase (ACs)	(Radio et al., 2018)
	Ras	(Lee et al., 2013)
	ADP Ribosylation Factors	(Davis et al., 2019)
Altered drug metabolism	Cytochrome P450 (CYP450)	(Devine et al., 2012)
	Glutathione S-transferases (GSTs)	(Scarcella et al., 2012, LaCourse et al., 2012, Miller et al., 1993, Wijffels et al., 1992)
	Fatty Acid Binding Proteins (FABPs)	(Ramos-Benítez et al., 2017)

1.10. Prospect of Vaccine development

The effort to identify a vaccine for the control of fasciolosis is dated as far as three decades, but the consensus is that the level of efficacy for commercial production has never been achieved. This is because no vaccine, either a single antigen or a combination has produced results reliable enough to overcome the immunomodulatory activity of liver flukes, which limits development of protective immunity in infected animals (Zafra et al., 2021, Dalton et al., 2013, Molina-Hernández et al., 2015). The primary challenge is the fact that how liver flukes invade and migrate within the host and rapidly control host immune responses is not well understood. By 4 – 6 days post-infection, *Fasciola* NEJs have already reached the liver and caused obvious damage in the liver parenchyma, while initiation of immune activity could be early as 24 hour post infection experimentally (Molina-Hernández et al., 2015). Research suggests that lack of development of a protective T helper type 1 immune response in the hosts experimentally and naturally infected by fasciolids, in favour of a strong regulatory type 2 immune response (as infection advances towards chronicity) which represses type 1, is the reason *F. hepatica* vaccines have not been effective (Dalton et al., 2013). Thus, a vaccine is designed to be able to initiate the protective type 1 response. Despite these challenges, there has been huge effort centred around identifying vaccine targets, assessing efficacy, reproducibility, increasing protection percentages using adjuvants, reducing fluke burdens, etc. These have largely identified antigens important in fluke biology, orthologous genes that have been identified in closely related species as antigens of interest, and antigens that cross-react with sera of other trematodes (Toet et al., 2014). For example, genes such as GSTs, FABPs, and cathepsins have been extensively studied as vaccine candidates in liver flukes (particularly *Fasciola* spp., *C. sinensis*, and *Opisthorchis* spp.) and blood flukes (particularly *Schistosoma* spp.), however the variable level of results seen could be due to genetic variations in host and/or variations in homology of sequences of these orthologs (McManus, 2020).

Although there are currently no commercially available vaccines against liver flukes, increasing use of various multi-omic approaches in helminth research could be critical in developing an effective vaccine against helminth diseases (Daga et al., 2022). While the efficacy of the various previously investigated *Fasciola* vaccine targets have produced varying levels of protection against the parasite (Toet et al., 2014), success level in other helminths have been generally limited as well (Drurey et al., 2020). Despite these challenges, three vaccines are commercially available in helminths. The Barbervax vaccine was developed against the nematode - *Haemonchus contortus*. The vaccine was derived from parasite gut membrane glycoproteins (de Matos et al., 2017). Another vaccine, the Bovilis Huskvac was developed against the nematode - *Dictyocaulus viviparus*. The vaccine was developed using live irradiated third larvae stage of the parasite (Jarrett et al., 1958). Also, a Providean Hidatil EG95 vaccine was developed against the cestode - *Echinococcus granulosus*. The recombinant oil-based vaccine was developed in Argentina and tested in sheep and Llamas (Poggio et al., 2016). While various recombinant proteins (such as Glutathione-s-transferase, Aspartic protease, Tetraspanin, etc) have been used in clinical

trials to develop helminth vaccines in humans (Perera and Ndao, 2021), a commercially available and effective vaccine is yet to be available.

1.11. Vaccine Candidates in liver fluke research

To develop a commercially available vaccine to control *F. hepatica* and *F. gigantica* in livestock, various genes have been targeted and used either singly or in combination. Most vaccines are either in native or recombinant form. A vaccine is generally assessed on its protective ability in the host, the issue here is, these proportions are variable and not repeatable across trials. An ideal vaccine is expected to initiate about 80% protection, this is thought to reduce mean fluke burden below 30 – 53 flukes, a level where livestock productivity is not impacted (Toet et al., 2014). Many candidates have been explored; the top ones are GST, cathepsins, FABP and others include leucine aminopeptidase, haemoglobin, paramyosin, peroxiredoxin, and *Fasciola* tegmental proteins. While the protection from these various *Fasciola* vaccine candidates can be as low as 0%, trials from other candidates suggests that about 98% protection could be achieved. Findings also indicate that adult fluke extracts using alum or Freund's adjuvant could initiate about 85-96% protection, while irradiated *F. hepatica* metacercaria, despite been able to excyst, cause little tissue damage; however, high doses of these do not seem to confer sufficient protection (Haroun and Hillyer, 1986). Similarly, tegument proteins, considering they serve as the main site of interaction between liver flukes and the host, could potentially be promising vaccine candidates. In addition to these are other fluke excretory-secretory (ES) products, which are thought to be able to suppress type 1 responses and are thus are immunomodulatory as well (Donnelly et al., 2011). While there is no standardised vaccine now, it is important to ensure replicability of trials, determine performance of vaccines across herds in different locations, assess the dose of vaccine and its impact on tissues, to control fasciolosis especially as TCBZ resistant populations are increasing. As "omics" data gradually becomes more readily available and accessible, structural annotation of candidate gene families becomes easier. The availability of different high quality omics datasets could provide a wealth of information in fluke biology that can be harnessed to better understand the complex organism.

1.12. Genomic approach in Parasite Studies

In the last couple of decades, genetic studies have gradually shifted from amplifying and sequencing gene fragments to sequencing the entire genome. More recently, various 'omics' related experiments have generated huge datasets and interpreting these efficiently can provide massive biological insights. While designing sequencing experiments can be challenging, analysing data generated can present computational issues as well. Genomics for example builds on the concept of genetics. Genomics broadly refers to the science of genomes. It involves a combination of high throughput sequencing techniques and bioinformatics to sequence, assemble and interpret genomic findings. Genomics covers the structural and functional annotation of a genome. Structural annotation involves identifying genes and transcripts, non-coding RNAs, repetitive elements and regulatory elements, etc. A

detailed structural annotation is important to describe the features of coding and noncoding elements in the genome and to facilitate functional annotation. In functional annotation, putative function of genomic features (such genes, repeats, non-coding RNAs, and regulatory regions) are done by assigning biological information such as functional sites, orthologs, pathways associated etc (Stein, 2001). For example databases such as KEGG can be used for orthologous and pathway assignments, while databases like NCBI, InterPro, Pfam, Uniprot, Swissprot, Blast, etc can be used to predict domains, putative names, and family assignment, etc. (Del Angel et al., 2018).

Using various software algorithms, genomic features can be identified, while those of interest can be further investigated. For example, there are several gene prediction tools designed to identify gene structure and location, introns, exons, alternative splicing, open reading frames (ORFs), etc in various organisms. These tools are designed to analyse genome assemblies of different kinds of organism, such as prokaryotes, eukaryotes, etc. Popular gene prediction software such as Augustus (Stanke and Morgenstern, 2005), Braker (Hoff et al., 2015), GeneMark (Besemer et al., 2001), Fgnesh (Solovyev et al., 2006), GeneScan (Burge and Karlin, 1997) and Maker (Holt and Yandell, 2011) are example of popular tools designed for eukaryote genome annotation. Some of these work as stand-alone tools or as part of a pipeline to improve genome annotation. A vital aspect of genomics is quality assembly and annotation. This is dependent on the type of sequencing technique (whether long read or short read), the annotation pipeline, and other species-related factors such as genome size and repeat content. Genome annotation is important to allow correct biological inferences in the organism of interest and for comparative genomic studies. Comparative genomic studies have a resolving power to identify similarities and distinguish difference between multiple organisms (Hardison, 2003). Comparative studies provide insights into function and evolutionary related issues across multiple species (Miller et al., 2004). For example, a comparative study of parasitic nematodes and platyhelminths reported various gene births and expanded genes that are key to parasitism (International Helminth Genomes, 2019).

As of October 2022 there are 202 parasite genome assemblies on WormBase Parasite (Version: WBPS16), a resource for helminth genomes (Howe et al., 2017). The database holds genomes from 163 species: 45 platyhelminth genome assemblies, and 157 nematode genome assemblies. There are currently 20 fluke genome assemblies on the database (Table 1.3), 9 of which are *Schistosoma* spp. The database is updated regularly with new assembles and species when available. Availability of these fluke genomes have contributed to the wealth of information on the biology of these trematodes and has shaped how researchers approach them. Using a combination of traditional Sanger capillary sequencing and Illumina sequencing, 81% of the *S. mansoni* genome was assembled into chromosomes. This identified gene transcript alternative splicing and profiled the parasite's transcriptome across all stages of the parasite, observing that as the parasite progresses from the infective cercarial stage into adulthood, there is a shift from an increase in expression of genes

associated with glycolysis, translation, to transcription to genes required for parasite development and signalling pathways (Protasio et al., 2012). Upon improving the *S. mansoni* genome, the assembly size increased from 364.5 Mb to 391.4 Mb, while the number of genes identified reduced from 10,116 to 9,794. The improved assembly provided insights into the sex chromosomes of schistosomes identifying a sex-linked variation in gen copies and expression (Buddenborg et al., 2021). The *C. sinensis* genome annotation and transcriptome analysis found that genes important to the fluke's response to stimuli and muscle-associated growth are more expressed in its oral sucker, while the organism had higher expression of fatty acid, glucose, oxygen and amino acid transporter genes in comparison to its closely related species (Huang et al., 2013). In a similar genomic study, transcriptome study of *O. viverrini* revealed how the fluke survives in the host bile duct, its adaptability to rich lipid diet in the host, and ability of the parasite to modulate cell proliferation in the host via secreted proteins (Young et al., 2014). Annotation of 84% of the estimated *P. westermani* genome revealed the parasite had a proteome with 82% homology to *O. viverrine* and *C. sinensis* as expected; however phylogenetic studies suggested the *P. westermani* has considerably diverged from both nearest relatives (about 28 – 59 million years ago (Mya)) leading to expansion of long interspersed nuclear elements (LINE) and long terminal repeat (LTR) retrotransposons in *P. westermani* (Oey et al., 2018).

Fasciola has one of the largest parasite genomes ever sequenced, with a genome almost 3 times the size of *S. mansoni*. Despite them both being trematodes, there are huge differences in biology between them, such as, location in host, sexual reproduction, and life span. *F. hepatica* (1.20 Gb) is similar in genome size with its sister species, *F. gigantica* (1.13 Gb), while *Fp. buski* (0.74 Gb) is the next to both species in terms of size when compared to other trematodes (Table 1.3). In the Liverpool *F. hepatica* assembly, N50 was 204 Kbp, higher than the 161 kb reported in the Oregon assembly. Also, while a total of 15,740 RNAseq supported genes were annotated in the Liverpool assembly, a reduced number of genes (14,642 gene) were predicted in the Oregon annotation. While the total number of annotated genes vary due to various reasons such as assembly contiguity, gene fragmentation, etc. Interestingly, the average number of exons per gene was higher in the Liverpool genome project than in the Oregon genome project (5.3 versus 3.3 respectively), suggesting gene models predicted in the later were either shorter or fragmented due to the increased fragmentation of the genome due to sequencing (as indicated by a lower N50). Despite the various limitations, the Liverpool genome project noted that *F. hepatica* genes had more non-synonymous polymorphisms than *Schistosoma* spp, *Clonorchis sinensis*, and *Opisthorchis viverrini* (Cwiklinski et al., 2015). While the Oregon project reported the first observation of a *Neorickettsia* endobacterium in liver flukes and a vertical transmission potential of the organism in *F. hepatica (McNulty et al., 2017b)*. Interestingly after improving the *F. hepatica* draft genome was using a 10X Chromium platform, with gaps in scaffolds filled, polished using Pilon (Walker et al., 2014), the N50 was increased to 1.9 Mb and a total of 9,732 coding genes were predicted (*Fasciola*_10x_pilon, GCA_900302435.1). After reannotating the Oregon assembly using an improved methodology, additional

RNAseq dataset, and protein homology databases, the number of genes predicted was reduced to 11,218. These variations highlight the complexities associated to liver fluke genome projects.

These findings are somewhat similar in the *F. gigantica* and *Fp. buski* genome projects. In the *F. gigantica* genome project from the US (~1.128 Gb genome size), the assembly N50 was 181.8 kb, mean exon number per gene was 5.2, and a total of 12,647 genes were annotated (Choi et al., 2020). Comparatively, in the *F. gigantica* genome project from India (~1.04 Gb genome size), the assembly N50 was 129 kb, mean exon number per gene was 3, while a total of 20,858 genes were annotated (Pandey et al., 2020). This statistic suggests that the higher number of genes in the later project could be due to gene fragmentations as suggested by the lower N50. The *Fp. buski* however, has a smaller genome size of 0.748 Gb. Despite the smaller genome size, a total number of 11,747 genes were annotated (N50 was 190.8 kb, while mean exon per gene was 5.2) (Choi et al., 2020).

Genome examination of these species suggested the divergence between *Fasciola* and *Fasciolopsis* was about 90 Mya, while around 50 – 65 Mya there was a shift in choice of intermediate host from planorbid to lymnaeid snails and a change from intestinal to hepatic habitation in the hosts. However, *F. hepatica* and *F. gigantica* diverged much more recently around 5 Mya (Choi et al., 2020). Despite *F. hepatica* having a big genome, it appears the genome size is not due to genome duplication. Adaptive measures adopted by the parasite seem to facilitate choice and adaptation to its intermediate host and parasite migration through the definitive host, although how this happens is still unclear. Studies on the *F. hepatica* genome found considerable levels of polymorphisms, despite its ability to adopt inbreeding as an hermaphroditic organism, pointing out its rapid evolutionary adaptability to host, weather, drugs and vaccine changes (Cwiklinski et al., 2015). Interestingly, both *F. hepatica* genomes (Liverpool, UK and Oregon, USA) shared KEGG orthologous groupings, but gene models from both annotations seemed to overlap poorly possibly due to gene fragmentation or missing genes. The USA assembly reporting a higher repeat content of 55.29% compared to the 32% in the UK assembly (McNulty et al., 2017a). The annotation differences highlighted are further explored in this project and will be commented on later. Genome statistics of sequence trematodes suggests that the *Fasciola* family seem to have bigger genomes when compared to other trematodes (Table 1.3), an evolutionary trait thought to be adopted as the family emerged. Findings suggest repetitive elements of the *Fasciola* genome could have played some role in *Fasciola* evolution. While the nonrepeat aspect of trematode genomes appear similar in sizes (except in *Schistosoma* spp with a much smaller genomes), interspersed repetitive elements in *Fasciola* spp are more than twice in length when compared to *Fp. buski* (658–707Mb and 318Mb respectively), and longer in fasciolids than other trematodes (Choi et al., 2020). These corroborate observations from a big parasite genome analysis of nearly 81 parasitic and non-parasitic worms, a study which identified and ranked the key drivers of genome size variations. These include long terminal repeat transposons, simple repeats, the quality of assembly, DNA

transposable elements, intronic length and low complexity sequence, confirming observations that variations in genome sizes are predominantly due to non-coding elements (International Helminth Genomes, 2019). Thus, understanding these "junk" repetitive elements in *Fasciola* genome could help unlock the genome and provide a clearer picture of the parasite.

1.13. *F. hepatica* can adapt quickly

Various plausible factors could influence the genetic diversity of a fluke population. As previously mentioned, the fact that *Fasciola* spp are hermaphrodites means inbreeding is possible, thus influencing the spread of certain alleles in the fluke population. Thus, self-fertilisation could facilitate spread of traits such as resistance to anthelmintics, especially if it is a recessively inherited trait (Beesley et al., 2017b). Similarly, in the life cycle of the parasite, clonal expansion exists in the snail intermediate host, leading to loss of some gene variants and potentially causing a fluke population bottleneck. According to previous findings, *F. hepatica* undergo a genetic clonal expansion in the intermediate host (Hodgkinson et al., 2018), a lack of a genetic clonal expansion could lead to reduced variation in genetic variation in metacercariae and the eventual reduction in fluke population genetic diversity. Finally, the parasite has a diverse definitive host range and able to adapt to ensure survival. Findings indicates that the parasite is highly diverse genetically and capable of rapid gene flow between populations (Beesley et al., 2017b). This could be associated with the high level of polymorphisms found in the genome (Cwiklinski et al., 2015), and potentially makes drug and vaccine interventions challenging. The diversity in the fluke population and cross fertilization between *F. hepatica* and *F. gigantica* makes a multigenic drug resistance pathway highly possible. Thus posing various research questions such as, how will various fluke populations (even within a geographical location) respond to new drugs (Molina-Hernández et al., 2015)? How quickly will a fluke population develop resistance to an effective drug? How can free flow of animals across various geographical locations influence fluke population structure? Providing answers to these questions would be key to controlling the parasite. With growing application of omics technologies to fluke research, it is only a matter of time before these key research questions are answered.

1.14. Fluke evolutionary studies

Evaluation of evolutionary relationship between two or more species is common research practice. Evolutionary biology as a field measures selective pressures on protein coding sequences, a process utilized to identify genes under positive selection. The idea is that gene variants of benefit to an organism become fixed in the population over time. It is a common practice to assess evolutionary relationship among closely related organisms using multiple sequence alignments and phylogenetic analysis. With the aid of phylogenetic trees, ancestral relationships between organisms and potential speciation events can be depicted (Baum, 2008). A commonly used tool such as ClustalW (Thompson et al., 1994) facilitates assessing evolutionary studies using amino acids and DNA sequences separately. However in recent times, codon-based alignments have been used in evolutionary studies to preserve

encoded information for proteins, particularly because insertion of gaps during alignment can alter the reading frame (Steinway et al., 2010). Using homologous codon aligned genes from species of interest, the ratio of the number of non-synonymous and synonymous substitutions (dN/dS, also known as omega (ω)) is used as an indicator of selective pressure. An $\omega > 1$ indicates diversifying positive selection, $\omega < 1$ indicates purifying selection, while $\omega = 1$ indicating a neutral selection.

The Codeml software, from the Phylogenetic Analysis by Maximum Likelihood (PAML) package enables comparison of various nested statistical models which can be done by conducting likelihood ratio tests (LRTs) of various evolutionary hypothesis (Yang, 2007). These models include site, branch, and clade models. These studies have been demonstrated to detect signals that suggest a selective pressure in various genes such as the vertebrate β-globin gene, the human influenza virus A hemagglutinin gene, and a some HIV-1 genes (Yang et al., 2000). Despite importance of the technique for assessment of selective pressure, it is highly susceptible to false positives (Mallick et al., 2009). These are due to various reasons such as poor gene annotations, ortholog assignment, poor alignments, splice variations, etc making error identification, correction of errors, and automation of the methods challenging. Various packages have been developed to address various computational issues, however replication of analysis on multiple datasets is a recurrent problem. Common packages include JCoDA (Steinway et al., 2010), PosiGene (Sahm et al., 2017), FUBAR (Murrell et al., 2013), VESPA (Webb et al., 2017), FUSTr (Cole and Brewer, 2018), Datamonkey (Delport et al., 2010), and Etetookit; most of which are built on Codeml in PAML (Yang, 1997), and varying in data input format, processing and result output. While these techniques have been used in human genetics, their use in helminth evolutionary biology is somewhat limited. Considering flukes are polymorphic and rapidly adaptable, assessing adaptive pressure in fluke populations is inevitable, although doing this on genomic scale does not lack its challenges.

1.15. Omics application in *F. hepatica* research

In the last decade there has been a rapid shift from the conventional Sanger sequencing method to rapidly developing next generation sequencing technologies (NGS). In Sanger sequencing method, a DNA sequence of interest is used as template to sequence a piece of DNA via a chain termination method. While this method is cost effective and considered "gold standard" for validating gene sequencing projects, this approach is not realistic on genomic scale. This led to the development of various NGS methods such as the illumina sequencing, pyrosequencing, Nanopore, Single-molecule real-time (SMRT) sequencing, etc (Tost and Gut, 2007, Cox et al., 2010, Mikheyev and Tin, 2014, Eid et al., 2009). NGS methods are capable of sequencing multiple genes and entire genomes with a low amount of input DNA. Various NGS platforms are now available commercially and while these are relatively expensive, costs are gradually becoming affordable. These technologies generally involve sequencing DNA or RNA nucleotides in genomes or regions of a genome (Sari et al., 2022). NGS platforms are a huge basis for -omics, a term used to describe the collective

characterization and quantification of biological molecules, usually involving high-throughput assays and data analysis. While omic technologies are not exclusive to NGS technologies alone. Omic technologies span across genomics, transcriptomics, proteomics, metabolomics, nutrigenomics, epigenomics, pharmacogenomics, phenomics, etc, (Schneider and Orchard, 2011) have been applied to biomedical research to facilitate understanding diseases (Hasin et al., 2017). Transcriptomics for example involves the study of all RNA transcripts from an organism. With the aid of high-throughput methods differential expression of genes can be explored in an organism. Sequencing platforms such as illumina are commonly referred to as short read sequencers (75-300bp) while Pacific Biosciences (PacBio) and Nanopore technologies are commonly referred to as long-read sequencers (>10,000bp). Although the long-read isoform sequencing (Iso-Seq) from PacBio could be more expensive than the short-read sequencing approach, it's able to identify full transcripts better, identify splice variants of a gene, and reduce mapping ambiguities (Cho et al., 2014). Previous studies indicate the technology can be used to identify novel transcripts, map unannotated genes, and explore transcript diversities in an organism (Leung et al., 2021). While long-read sequencing alone might not be able to quantify differential gene expression adequately (Huang et al., 2021), a combination of both technologies can proved adequate information on full information on gene splice sites and expression profiles (Tilgner et al., 2014).

With genomics gradually replacing genetics, various organisms have gradually been sequenced. The application of omics approach to helminth research is relatively recent compared to other related fields, likely because helminths are complex to understand and sequence arguably because of lack of cell lines that can be used to model gene functions, mutations, and gene knockouts in parasites. This makes identification of gene-specific functions challenging in liver flukes. The first genome - *Haemophilus influenzae* was sequenced and assembled in 1995 (Fleischmann et al., 1995, Iskander et al., 2017), while in 1998, the nematode - *Caenorhabditis elegans* genome was published (C. elegans Sequencing Consortium, 1998). In 2002 the first eukaryotic genome from *Plasmodium falciparum* (the causative agent of malaria in humans) was published (Gardner et al., 2002), while in 2004 the first parasitic nematode - *Brugia malayi* (a human filarial worm) was published (Ghedin et al., 2004). The first trematode parasite - *Schistosoma mansoni* (Berriman et al., 2009) was sequenced in 2009. By 2019 when comparative analysis of parasitic and non-parasitic worms was made, a total of 81 genomes was available (International Helminth Genomes, 2019).

In 2001, *F. hepatica* mitochondrial genome was first sequenced (Le et al., 2001); however it was until 2015 that the first nuclear genome was sequenced (Cwiklinski et al., 2015). *F. hepatica* has the largest trematode genome (1.3 Gb), the reason for which is rather unclear. Sequencing the *F. hepatica* genome has opened the door for a wide range of studies. With the release of other liver fluke genomes, comparative studies such as number of genes and other genome statistics become feasible. For example, the *F. hepatica* genome from

Oregon, USA reported a colonization of a rickettsia endobacterium in the organs and tissues of the parasite (isolated from sheep); this is important because *Neorickettsia* can cause severe illness in humans and fasciolosis is zoonotic (McNulty et al., 2017a). Comparatively the *F. hepatica* genomes assemblies published revealed variation in number of repeats – 32% and 52.3% in the *F. hepatica* (UK) and *F. hepatica* (Oregon, USA) respectively. The reason for this marked variation in repeat content is unknown, although this could be due to variations in the repeat masking section of annotation pipeline, or an adaptive evolutionary importance in the parasite. The former is plausible considering variation in annotation pipelines influences the gene prediction noticeably, thus suggesting the need for a standardized pipeline.

1.16. Genome annotation and associated challenges

Annotating gene models in a genome assembly is predominantly a bioinformatic process. It generally involves using a software to predict gene models using the genome assembly fasta file. Other additional evidence datasets such as protein sequences from resources such as UniProt (The UniProt, 2021), expressed sequence tags (ESTs), and RNA seq data can also be used to improve gene model predictions. These softwares are used to identify and improve the structure and location of genes in the genome. Gene features such as the transcripts, the untranslated regions (UTRs), alternatively spliced variants, exon, and intron boundaries can also be identified. As the cost of DNA/RNA sequencing gradually becomes affordable, a new challenge presents itself, such as how to annotate these assemblies accurately and update and improve previous assemblies. This is important to researchers in that a poorly annotated genome limits further downstream analysis. Poor annotation of genes can limit the quality of biological inference that can be made from gene models. For example, omission of some exons in a gene can limit identifying orthologs of the gene in related species and assigning a putative function to the gene. Thus, providing a complete genome is imperative, however this is influenced by various factors. Until recently, most genomes were sequenced on Illumina platform; these are known for their high read depth and short reads generated. Having short reads means the contiguity of the assembly could be reduced, and gaps will be present in the genome assembly. However, with the introduction of long-read sequencing platforms such as Oxford Nanopore and PacBio sequencers, contiguity can be improved, especially in organism with many repeats such as *Fasciola*. Thus, a common approach is to utilize both technologies to improve scaffolding and reduce gaps; this was done in the *F. gigantica* and *Fp. buski* genome sequencing project (Choi et al., 2020). The need for updating previously assembled genomes by incorporating long-read sequencing is generally accepted, although every annotation project comes with its technical issues. The availability of various annotation procedures in different groups, pipelines varying input and output file formats, longer run times especially in bigger genomes (some pipelines take weeks to run), and bioinformatic expertise required, are some of the challenges facing genome annotation (Yandell and Ence, 2012).

1.17. Candidate gene studies versus Genomic approach

Candidate gene studies serve as the foundation of genetic-based projects. This relatively cheap and fast approach was primarily used to investigate genes of interest presumed to be associated with a disease (Patnala et al., 2013). Using this approach, the risk of having a disease was assessed based on genetic variations in the associated gene in case and control patients (Giri and Mohapatra, 2017). SNPs could be investigated at a single gene level, Genome-wide association studies (GWAS) are used at genomic level to test genetic association with diseases, estimate disease heritability, risk, etc (Uffelmann et al., 2021). Microarrays and quantitative trait loci (QTL) have been used extensively to investigate candidate genes of interest (Wayne and McIntyre, 2002) , while whole genome sequencing and whole exome sequencing methods are been used for GWAS studies to identify genetic variations in genes of interest in a population (Uffelmann et al., 2021).

Investigating genes of interest is a common approach in genetic studies. A typical experiment could involve designing primers to amplify and sequence fragments of genes of interest for various studies; sequences of which are available on databases such as NCBI and UniProt. Similarly, analysis of expressed sequence tags (ESTs) (fragments of mRNA sequences from sequenced clones of cDNA libraries) has been pivotal in gene discovery and studies in the 1990s (Sotillo et al., 2017, Parkinson and Blaxter, 2009). These gene-based datasets have been crucial to genomics, as these can be used as alignment evidence to improve genome annotation pipelines such as Maker. For example, use of external gene datasets aligned properly to their correct genomic position, can aid identifying gene prediction errors such as missing exons and gene fragmentations. While ESTs have been important in single gene studies, the use of RNA sequencing has significantly improved understanding of genes of interest by facilitating understanding gene splicing, expression, and mutations qualitatively an quantitatively (Ozsolak and Milos, 2011). With the increasing use of long-read sequencing platforms such as the Pacific Biosciences (PacBio) and Nanopore technologies, gene studies have informative. Investigating at gene level is generally easier compared to evaluating an organism at genome level due to higher number of genes to assess at genomic level. Thus, identifying the presumed structure and biology of a group of genes of interest could be used as basis to determine which families of interest to focus on in the genome (Zhu and Zhao, 2007).

For example, cysteine proteases (commonly called cathepsins) are important virulence factors expressed in all animals. A microarray analysis based approached was used to identify gene expression profile in the spleen of a mice and identified 820 induced differentially expressed genes including cathepsin B (Rojas-Caraballo et al., 2017, Rojas-Caraballo et al., 2014). In trematodes, cathepsin genes have been well studied and implicated for their excystment, migration and immunomodulatory roles in the parasites (Smooker et al., 2010). Despite these findings, it is arguable that most biological processes involve multiple connected pathways, thus, genomic studies could facilitate understating how each process of interest works. An integration of data from candidate studies and

omics studies is vital to interpretation of biological processes. Databases such as Kyoto Encyclopaedia of Genes and Genomes (KEGG) is an extensive curated resource with eukaryote (545 genomes), bacteria (6234 genomes), and archaea (343 genomes) data. Interestingly, only 3 of these are flatworms (*S. mansoni*, *S. haematobium*, and *O. viverrine*) as of April 2021. The database provides various systems, genomic, chemical, health and drug information for numerous species (Kanehisa et al., 2020). Despite the growing number of omics datasets in various species, there is a gap in the availability of datasets, including their biological and functional annotations, especially in trematodes. This lack of detailed annotations reflects the need to integrate candidate gene studies and omics datasets, to provide a clearer biological context for fluke biology and its system pathways. This hinges on investing in more omics projects, providing datasets to .scientific community, and ensuring biological and functional relevance are provided.

1.18. Recent omics advancements: Application in fluke research

There have been gradual advances in omics technologies such as genomics, proteomics, transcriptomics, metabolomics, vaccinomics, epigenomics, phenomics, etc in the last couple of decades, but only a handful of reports are available on liver fluke studies compared to other organisms. These and other post-genomic tools have supported new antigen identification, although this have not been very applicable in liver fluke vaccine development (McManus, 2020). Despite these limitations, there has been reported efforts to detect, characterize and explore various liver fluke host-parasite molecules with focus on the tegument proteome, extracellular vesicles, secretomes; this has aided understanding the parasite's biology (Sotillo et al., 2017, Robinson et al., 2009b, Wilson et al., 2011). Similarly, proteomic analysis has been used to highlight the roles of GSTs in the detoxification of TCBZ (Chemale et al., 2010, Chemale et al., 2006), the importance of cathepsins in liver flukes (Robinson et al., 2008), which are identified molecules important to various biological processes in NEJs (Hernández-González et al., 2010). Currently the genomes of *F. hepatica*, *F. gigantica*, *Fp. buski*, and some other trematodes have been published, these facilitate comparative genomic analysis (Cwiklinski et al., 2015, McNulty et al., 2017a, Choi et al., 2020, Pandey et al., 2020). Availability of genomes from these closely related species ensure gene orthologous relationships are better assessed, although comparative studies at genomic level require significant bioinformatic expertise. Extensive transcriptome datasets or various stages of *F. hepatica* are available, these provide a picture of expression pattern of genes each developmental stages of this parasite (Cwiklinski et al., 2015), and have been used to assess gene expression patterns in *F. hepatica* isolates sensitive and resistant to TCBZ and albendazole (Radio et al., 2018). There is no doubt that there is need for more omics studies in fluke biology as these provide unbiased understanding of parasite biology. A comprehensive interpretation and integration would be key to developing a sustainable control measure.

Despite the huge progress made in understanding trematode biology, vaccine development, and drug interactions, there is still much to be done. In order to facilitate liver fluke genomic

studies, the current annotation challenges need to be tackled. For example, using public databases to annotate new fluke sequences can be challenging if most of the genes are either not named adequately or described as 'unknown' (Haçarız and Sayers, 2016). This could range from not finding BLAST significant hits to not finding pathways to assign gene sequences to. Also, *Fasciola* genome annotations available tend to not contain much information associated with the gene identifiers. In future omics studies, detailed annotation of sequences would facilitate integration of various fields such as genomics, transcriptomics, proteomic and metabolomics. While metabolomics is a gradually growing field in biomedical research, there are currently no reports of its use in liver fluke studies. Metabolomics, a large-scale quantification of small molecules can potentially be used to assess *F. hepatica* response to TCBZ and identify which proteins TCBZ binds to (Kelley et al., 2016). The idea that anthelmintic resistance is likely to be multi-genic makes understanding the mechanism tricky (Molina-Hernández et al., 2015). Epigenomic studies could potentially facilitate understanding regulation of gene transcription in the parasite. This could explain expression patterns in response to drugs or vaccines. With the increasing use of various omics technologies in biomedical research such as phenomics and vaccinomics; to assess relationship between an organism's environmental factors and gene expression (Houle et al., 2010), and immunogenicity in relation to vaccine development (Poland and Oberg, 2010) respectively. It only a matter of time before scientists explore how fluke population genetics influences drug resistance and vaccine effectiveness.

1.19. Fasciolosis research: A farmer's perspective

To a typical farmer, fluke research is good if the herd is healthy. In other words, the primary focus of a farmer is towards improving the health of the animals, increased productivity, and reduced economic loss from diseases. Fasciolosis causes significant economic losses to a farm. Overreliance on TCBZ by farmers is a key factor in the development and spread of resistance, thus it is thought that development of economic models would enable farmers to make informed control measures. To achieve this though, there is need for development of a sustainable, accurate, affordable, and fast means of diagnosis accessible to farmers (Beesley et al., 2017a). Other farm management practices such as grazing and herd density, proper drug usage, adequate new animal quarantine systems, and surveillance are key to tackling fasciolosis in farms (Caira and Littlewood, 2013). As research efforts on vaccine developments are underway, one would expect various trials before commercial production. It is important to have standardised protocols to ensure that various trial results can be are compared (Molina-Hernández et al., 2015). This would facilitate the development of processes as well as update policy makers and stake holders on progress level. Thus control measures would clearly involve helping farmers make informed decisions on farm management, developing a new drug with an efficacy or more as TCBZ, and potentially development of an effective and commercially affordable vaccine (Fairweather et al., 2020). A proper balance of farming practices and intensive research would be required in the long term if fasciolosis would be controlled, or maybe eradicated worldwide.

1.20. Project overview

This project relies predominantly on the computational analysis of datasets, especially RNAseq datasets from within our research group, with an aim of providing insights into *F. hepatica* biology. This project will use a combination in-house dataset and datasets from various online databases such as WormBase Parasite, NCBI, KEGG, ENA, Pfam, etc and various tools and pipelines (methods are described in each project chapter) to the explore *F. hepatica* genome. Genes associated with TCBZ mode of action and resistances will particularly be focused on.

1.20.1. *F. hepatica* genome re-annotation

This study is designed to reannotate the current *F. hepatica* genome (*Fasciola*_10x_pilon) using the draft genome (Cwiklinski et al., 2015), and two RNAseq datasets using gene prediction tools. Observations are presented in chapter 2 of this thesis. Validation of predicted gene models will be done using well-described complete coding sequences (CDS) of gene families of interest.

1.20.2. Assessing selective pressure in *F. hepatica*

An evolutionary study approach will be used to explore *F. hepatica*. Questions to be addressed include; in *F. hepatica* genes associated with drug activity and resistance (Fairweather et al., 2020), are there genes undergoing a positive selection pressure? What genes are rapidly evolving? Are there adaptive forces influencing the parasite when compared with closely related species using orthologous gene sequences? How are these rapidly evolving genes likely to influence drug interactions? *F. hepatica* genes in the families associated with TCBZ activity and resistance will be compiled, and for each gene, signals indicative of positive selection pression would be assessed. Findings are presented in chapter 3 of this thesis.

1.20.3. Evaluating *F. hepatica* gene expression across various life stages

Using RNAseq datasets from various life stages of *F. hepatica* , the expression profile of gene families associated with drug action would be assessed across various stages of the parasite. Questions to be addressed include what life stages of the parasite are more active? Is there a relationship between gene expression and drug effectiveness at a specific life cycle stage? Findings are reported in chapter 4 of thesis.

1.20.4. Assessing gene expression in response to TCBZ

This project aims to understand TCBZ action in drug resistant and drug susceptible *F. hepatica* isolates using gene expression datasets. Questions to be addressed include how does the TCBZ induce the expression of *F. hepatica* genes in susceptible isolates compared with the drug resistant ones, especially in gene families previously described to be associated with TCBZ mode of action? Observations are reported in chapter 4.

Table 1.3: Comparison of parasitic fluke assemblies available on WormBase Parasite Database

Specie name	Common name	Assembly	BioProject ID	Genome size (Gb)	Genes	Gene transcripts	N50	Reference
Clonorchis sinensis	Chinese Liver fluke	C_sinensis-2.0	PRJDA72781	0.547	13,634	13,634	415,842	(Huang et al., 2013)
		ASM360417v1*	PRJNA386618	0.563	14,538	14,936	1,628,761	(Wang et al., 2018)
Echinostoma caproni	Intestinal fluke	E_caproni_Egypt_0011_upd	PRJEB1207	0.835	18,607	18,607	26,853	(International Helminth Genomes, 2019)
Fasciolopsis buski		F_buski_1.0.allpaths-lg	PRJNA284521	0.748	11,747		180,478	(Choi et al., 2020)
Fasciola gigantica	Liver fluke	F_gigantica_1.0.allpaths	PRJNA230515	1.13	12,669	13,940	178,720	(Choi et al., 2020)
				1.04	20,858		129 kb	(Pandey et al., 2020)
Fasciola hepatica		Fasciola_10x_pilon	PRJEB25283	1.20	9,732	16,830	1,901,411	(Cwiklinski et al., 2015)
		F_hepatica_1.0.allpaths.pg*	PRJNA179522	1.14	15,739	15,739	160,440	(McNulty et al., 2017a)
Opisthorchis viverrini	Southeast Asian liver fluke	OpiViv1.0	PRJNA222628	0.620	16,379	16,379	1,347,703	(Young et al., 2014)
Paragonimus westermani	Lung fluke	ASM850834v1	PRJNA454344	0.923	12,783**	12,805	134,838	(Oey et al., 2018)
Schistosoma mansoni	Blood fluke	Smansoni_v7	PRJEA36577	0.410	10,144	14,528	50,458,499	(Protasio et al., 2012)

*Genome assembly not used in this study as representative species in identifying orthologs.

**Dataset used in this project had 12,771 genes while published version had 12,783 genes

CHAPTER 2

Fasciola hepatica genome annotation and validation of gene models

2.1. Background

Genomic sequencing is crucial to biomedical research; thus, various methods and technologies have been developed. The choice of sequencing method depends on numerous factors, including cost, expertise, experimental design, organism, sequence coverage, data analysis. Rapid advancement in sequencing presents enormous prospects and challenges (Del Angel et al., 2018). As more reference genomes are generated, genome resequencing and availability of RNAseq datasets from these organisms means there is a constant need for improved and reliable re-annotations. Having a quality annotation is essential as this form the basis for downstream analysis. To achieve this, biologists use multiple strategies ranging from experimental validation of proteins to the use of computational tools. Despite the prospect of using Sanger sequencing on genes of interest to improve the annotation of genes in a genome, it is unrealistic to utilise this method considering the number of genes in a typical genome and the technical complexities involved. Thus various computational approaches become pertinent to use the information from high-throughput sequencers to provide detailed gene annotations (Furnham et al., 2012).

2.2. Annotating a genome

There is no standard methodology for annotating a genome (Jung et al., 2020). The process is hugely influenced by the organism, sequencing technology, computational suitability, human factors such as time and expertise, and research design. These directly or indirectly affect the quality of assemblies, accuracy of gene models, functional annotation, and other downstream analyses. In recent times, genome annotation can be automated or manual. Automation of annotation process web-based pipelines, such as ones offered by NCBI (Li et al., 2021) and EBI (McWilliam et al., 2013), are predominantly suitable for prokaryotes, while some are suitable for eukaryotic genomes (e.g. Companion software) especially smaller eukaryotic genomes such as protozoa (Steinbiss et al., 2016). These pipelines can be challenging to use in large genomes, especially in eukaryotic organisms. While eukaryotic genomes tend to be larger in size than prokaryotic genomes, size is not a primary differentiating factor between both. Apart for the large genome size in eukaryotes, eukaryotic genomes contain repetitive sequences and sometimes long intronic regions that complicate protein coding genes in the genome (Humann et al., 2019), while prokaryotic genomes generally lack introns and long repeats (Brown, 2002). Annotation errors in eukaryotic genomes can be challenging to spot and quickly transferred to other annotations (Salzberg, 2019). Manually annotating a genome, however, requires a lot of time, labour, and various computational processes; but facilitates improved annotations. Manual annotations allow for reviewing gene models and continuous re-annotation of a genome.

A typical eukaryote annotation project requires the use of at least one or more annotation software, many of which handle various stages of the process (Yandell and Ence, 2012, Jung et al., 2020). An annotation project starts with sequencing with a platform of choice followed by genome assembly. Genome assembly statistics include assessing contiguity and

completeness. Annotating a genome generally involves repeat masking as eukaryotic genomes are rich in repeats using tools such RepeatMasker (Smit et al., 2004). *Ab initio* predictions can be made from the genomic sequence using common computational software, such as Augustus (Stanke and Morgenstern, 2005). Evidence-based gene predictions are made using evidence such as protein sequence databases, ESTs, RNAseq data, etc., to improve the annotation quality. After which annotation statistics are assessed using the annotation pipeline, and annotation is viewed using genome visualisation tools such as Integrative Genomics Viewer, IGV (Thorvaldsdóttir et al., 2013). These analyses hinge on efficient software training and the availability of high-quality evidence from RNAseq and/or proteins from the organism or from close families of the organism of interest. The whole annotation project can take weeks, involving integrating various software output and a vast computer memory (Yandell and Ence, 2012).

2.3. *Fasciola hepatica* re-annotation project

Fasciola hepatica is a parasitic trematode, the cause of fasciolosis – a zoonotic disease with huge economic threat to livestock and humans (Love, 2017, Skuce and Zadoks, 2013, Charlier et al., 2013). Fasciolosis (caused by *F. hepatica* and *F. gigantica*) has been identified by WHO as a neglected tropical disease (WHO, 2020). It is estimated that 2.4 – 17 million people are infected (Caravedo and Cabada, 2020), with 91.1 million people at risk of the disease worldwide (Tolan, 2011). There have been extensive studies to understand the biology of the pathogen. However, this is challenging due to the complexity of the organism. The parasite can manipulate host immunity, and its development of resistance to anthelmintics especially Triclabendazole (TCBZ) is a rapidly growing concern (Dalton et al., 2013, Robinson et al., 2012). With the growing wealth of information genome sequencing offers, the first *F. hepatica* draft genome was published in 2015 from the UK.

Nearly 30 years ago, the first cellular organism - *Haemophilus influenzae* – complete genome was sequenced and assembled (Fleischmann et al., 1995, Iskander et al., 2017), while first trematode parasite - *Schistosoma mansoni* (Berriman et al., 2009) was sequenced about 14 years after. It took another 6 years before a *F. hepatica* genome was sequenced (Cwiklinski et al., 2015) . This vast difference in the number of years can be attributed to the higher sequencing cost and technology limitations. However, the number of genome sequencing projects has been increasing in recent times while cost drops and technology and computational tools are improving. For example, other *Fasciola*-related sequencing projects include another *F. hepatica* from Oregon, USA, two *F. gigantica* genomes and a *Fasciolopsis buski* genome (McNulty et al., 2017b, Choi et al., 2020, Pandey et al., 2020, International Helminth Genomes, 2019). This growing number of sequenced related species facilitate evolutionary studies by improving orthologous gene grouping.

Here, a re-annotation of the updated *F. hepatica* genome (assembly Fasciola_10x_pilon, GCA_900302435.1), an improved version of the draft *F. hepatica* genome (assembly GCA_000947175.1, BioProject PRJEB6687) (Cwiklinski et al., 2015) is done. After which the current *Fasciola*_10x_pilon, GCA_900302435.1 annotation (WP15) was compared with re-

annotated version (from this project) using well-described annotated gene families. We also utilised PacBio lso-Seq datasets to map and evaluate transcripts with their gene models predicted in both annotations. To achieve these, we used various computational tools to assess and compare gene models predicted by each tool to determine the best annotation. Due to a limited availability of well annotated *F. hepatica* gene models on various database, the three fairly well described gene families (Glutathione S-transferase (GST), Adenylyl Cyclase (AC), and Tubulins) were used to validate the current *F. hepatica* genome annotation (*Fasciola*_10x_pilon GCA_900302435.1, WP15) and the re-annotated version (Ryan et al., 2008, Radio et al., 2018, Scarcella et al., 2012, LaCourse et al., 2012). These are genes mostly amplified by polymerase chain reaction (PCR), with full coding sequences (CDS) sequenced by Sanger sequencing (Sanger et al., 1977).

2.4. Materials and Methods

2.4.1. Datasets

The previously reported *F. hepatica* draft genome assembly was enhanced using a 10X Chromium platform. Using *F. hepatica* genomic DNA, the 10x-Genomics technology was used to generate ultra-long reads (up to 1Mb) to map to gaps in illumina reads (Ma et al., 2019). Using these long reads, gaps in the *F. hepatica* draft genome assembly scaffolds were filled and polished using Pilon (Walker et al., 2014) (*assembly gap filling was done by Steve Paterson, University of Liverpool*). The updated *F. hepatica* genome (*Fasciola*_10x_pilon, GCA_900302435.1) assembly was used as the reference for mapping RNAseq data reads and genome annotation. RNA sequencing data (Illumina TruSeq libraries prepared from 3 biological replicates of metacercariae, two replicates of newly encysted juveniles (NEJs) 1-hour, two replicates of NEJs 3-hours, two replicates of NEJs 24-hours, one replicate of juveniles 21 days, and one replicate of adult) referred to as RNAseq dataset 1 in this report was used for mapping to the *F. hepatica* genome. These datasets (also referred to as Canda_lib reads) are previously described (Cwiklinski et al., 2015) (Figure 2.1). This dataset had ~646 million mapped reads, while ~94 million reads were unmapped to the refence genome. Another RNAseq data from 3 samples of 21 days old juveniles referred to as RNAseq dataset 2 (also referred to as erins_reads, courtesy of Erin Mccammick of the Queen's University Belfast) was mapped to the *F. hepatica* genome. This dataset had ~775 million mapped reads, while ~108 million reads were unmapped to the refence genome. These two RNAseq datasets were used as extrinsic evidence to facilitate gene model predictions.

2.4.2. Ab initio Predictions with Braker

Repeatmasker (Smit et al., 2004) was used to screen for interspersed repeats and low complexity in the *F. hepatica* updated 10x genome (Table 2.1) using the Dfam_Consensus-20170127 (available at http://repeatmasker.org/libraries/RepeatMaskerMetaData-20170127.tar.gz) and RepBase-20170127 (https://www.girinst.org/) database as the repeat library (Hubley et al., 2016, Bao et al., 2015). Reads from each RNAseq library were mapped to the genome using HISAT separately (Kim et al., 2015). SAMtools was used for sorting

mapped reads and conversion from (Sequence Alignment Map) SAM format to (Binary Alignment Map) BAM format (Li et al., 2009). After which, Braker (Hoff et al., 2015), a tool for unsupervised RNAseq based genome annotation, was used for gene prediction in each mapped library. Braker utilises GeneMark-ET (Besemer et al., 2001) and Augustus (Stanke and Morgenstern, 2005) as part of its pipeline to facilitate *ab initio* annotation. Braker analysis was done separately for each library.

2.4.3. Evidence-based annotations with MAKER

To facilitate evidence-based annotations, Stringtie was used to assemble the RNAseq alignments from both libraries into potential transcripts and eliminate redundancy (Figure 2.2). Using preliminary gene predictions, a merged RNA transcript file of both libraries, and a trematode NCBI protein homology sequence (with a total of 23,951 trematode proteins), MAKER annotation pipeline was used for re-annotation of the *F. hepatica* genome (Holt and Yandell, 2011). This was followed by a series of SNAP training of MAKER's gene predictions to filter and improve the final gene models (Johnson et al., 2008). Finally, tRNAs predicted with tRNAscan were filtered (Lowe and Eddy, 1997). Using the Awk command in linux, the "trna" features in the genome annotation gff file (in third column of the gff file) was filtered to remove tRNAa in the annotation. A total of 7,098 tRNAs were predicted and filtered. Genome Annotation Generator (GAG) software (version 2.0) was used to generate annotation statistics (Geib et al., 2018). Genome annotation files (in gff format) and the *F. hepatica* (*Fasciola*_10x_pilon, GCA_900302435.1) genome (in fasta format) was used as input files in GAG software using default settings to generate annotation summary statistics (Tables 2.2 & 2.4).

2.4.4. Other annotation tools

In addition to MAKER annotation, Transdecoder was used to identify coding regions in these datasets. The objective was to compare and evaluate Braker and MAKER predicted gene models with Transdecoder. To achieve this objective, the merged RNA transcript file of both libraries was passed into Transdecoder (https://github.com/TransDecoder/TransDecoder/wiki to identify the candidate coding regions (Figure 2.2). Portcullis was used to predict splice junctions (Mapleson et al., 2018).

2.4.5. Assessing Genome Completeness with BUSCO

The completeness of annotations was assessed using BUSCO (Benchmarking Universal Single-Copy Ortholog), which is a quality control tool that measures a genome's completeness in terms of its gene content. It can identify complete, fragmented, duplicated, and missing genes by comparing genes in the dataset with orthologs. BUSCO gene sets for Eukaryota (lineage dataset: eukaryota_odb10, contains 255 BUSCOs, BUSCO version 5.1.3) was used to assess the completeness of annotations (Simão et al., 2015).

2.4.6. Annotation Visualisation and Manual Validation of Gene models

Minimap2 was used to perform splice-aware alignment of the PacBio Iso-Seq reads to the *F. hepatica* updated 10x genome, after which samtools was used for sorting and indexing the mapped reads (Li, 2018). Iso-Seq read data was used to assess gene transcripts, explore

gene splicing, and validate exon structure in predicted gene models. Manual validation of annotation also included using well-described *F. hepatica* genes such as tubulins, GSTs, and Adenyl cyclase genes as a template to compare gene transcripts (identified by using PacBio data) to annotated gene modes predicted by MAKER. To achieve this, a literature search and compilation of genes of interest in these families (preferably genes amplified by PCR, with full CDS sequenced by Sangers sequencing) were done, and each gene was mapped to the *F. hepatica* genome using Exonerate (Slater and Birney, 2005). After which gene alignments, mapped RNAseq reads, PacBio reads, genome annotations (Braker, MAKER, Transdecoder, etc) was manually uploaded and viewed with the Integrative Genomics Viewer (IGV) (Thorvaldsdóttir et al., 2013). In the previously mentioned gene families, all members of the family were identified. Identified genes were assigned to their family based on the presence of motifs particular to the family using Pfam 34.0 (Mistry et al., 2020, Finn et al., 2016).

2.4.6. Functional annotation and Orthologous Grouping

The current *F. hepatica* annotation (*Fasciola*_10x_pilon, GCA_900302435.1, WP15) and the re-annotated annotation (version described in this project) were subjected to an automated functional annotation KEGG database - GhostKoala - to assign KEGG Orthology (KO) numbers (Figure 2.5) to annotated genes (Kanehisa et al., 2016). Gene Ontology (GO) assignments was done using PANNZER2 (Törönen et al., 2018). Orthomcl was also used to group annotations based on orthologous species using representative genomes of closely related species (Li et al., 2003). Species used included *Clonorchis sinensis* (PRJDA72781), *Echinostoma caproni* (PRJEB1207), *Fasciola gigantica* (PRJNA230515), *Fasciolopsis buski* (PRJNA284521), *Opisthorchis viverrine* (PRJNA222628), *Paragonimus westermani* (PRJNA454344), and *Schistosoma mansoni* (PRJEA36577).

2.5. Results

2.5.1. Results of Ab initio predictions

Prior to *ab initio* using Braker, analysis to mask repetitive elements in the *Fasciola hepatica* genome suggested a repetitive content of about 10% (Table 2.1) using a combined database of Dfam Consensus-20170127 and RepBase-20170127 (Storer et al., 2021, Jurka et al., 2005), whereas the initially published annotation identified a repetitive content of 32% using an in-house repeat library. Initial Braker analysis predicted 97,045 and 49,102 genes for the repeat unmasked and repeat-masked RNAseq dataset 2 analysis, respectively, similarly 100,527 and 52,257 genes in RNAseq dataset 1. Manual visualisation of these gene models was quite challenging as some observed gene models appear fragmented (Figure 2.1).

2.5.2. Results of Evidence-based MAKER Predictions

After four iterative MAKER analyses, with training data for each round generated with SNAP, a total number of 15,879 genes and 4,586 tRNAs were predicted, compared to the 9, 709 genes in the current *F. hepatica* annotation on Wormbase Parasite (*Fasciola*_10x_pilon, GCA_900302435.1). A mean gene length of 30.5 Kbp (186 bp – 564 Kbp), a mean exon length of 263 bp, and 46.3% complete BUSCOs. Comparatively, current *F. hepatica*

annotation on Wormbase Parasite (*Fasciola*_10x_pilon, GCA_900302435.1) had a mean gene length of 46 Kbp (203 bp – 1.02 Gbp), a mean exon length of 417 bp, and 75.7% complete BUSCOs. There was a 40.2% gene coverage in the genome compared to 37.1% in the initially published annotations (Tables 2.2 & 2.3).

2.5.3. Results of Transdecoder Annotation and BUSCO analysis

Using Transdecoder to predict genes based on coding regions and open reading frames (ORF), a total of 9,401 genes and 15,886 mRNAs were predicted (Table 2.2). BUSCO analysis on the *ab initio* predictions revealed the lowest levels of annotation completeness, 18% and 31%, in the RNAseq dataset 1 and 2, respectively. However, annotation predicted by Transdecoder had 82.4% completeness, while the annotation predicted by MAKER in the current annotation (*Fasciola*_10x_pilon, GCA_900302435.1, WP15) and in this re-annotation project were 75.7% and 46.9%, respectively (Table 2.3).

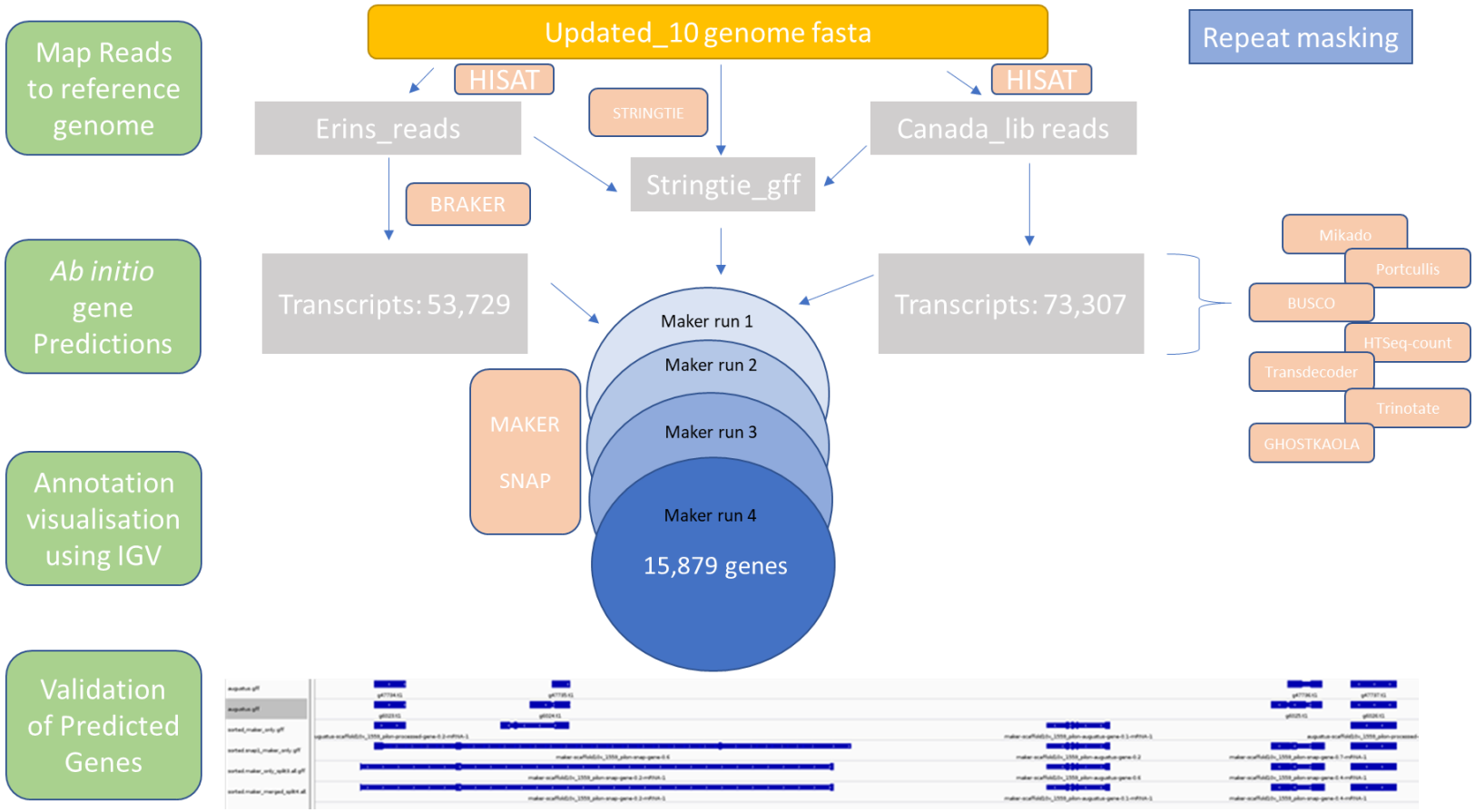

Figure.2.1: Overview of *Fasciola hepatica* genome re-annotation methods. A total of 15,879 genes were predicted, manual visualisation and validation of gene model was done using tubulin genes, GST genes and AC genes as these genes were previously described and annotated. This re-annotation recorded a higher number of genes, compared to the 9,709 genes previously published (See WormBase Parasite version 15 (WP15)), however this number is close to the 15,739 genes predicted in the *Fasciola hepatica* genome (BioProject PRJNA179522) predicted by the Mitreva laboratory, USA (annotations available on WP15 database).

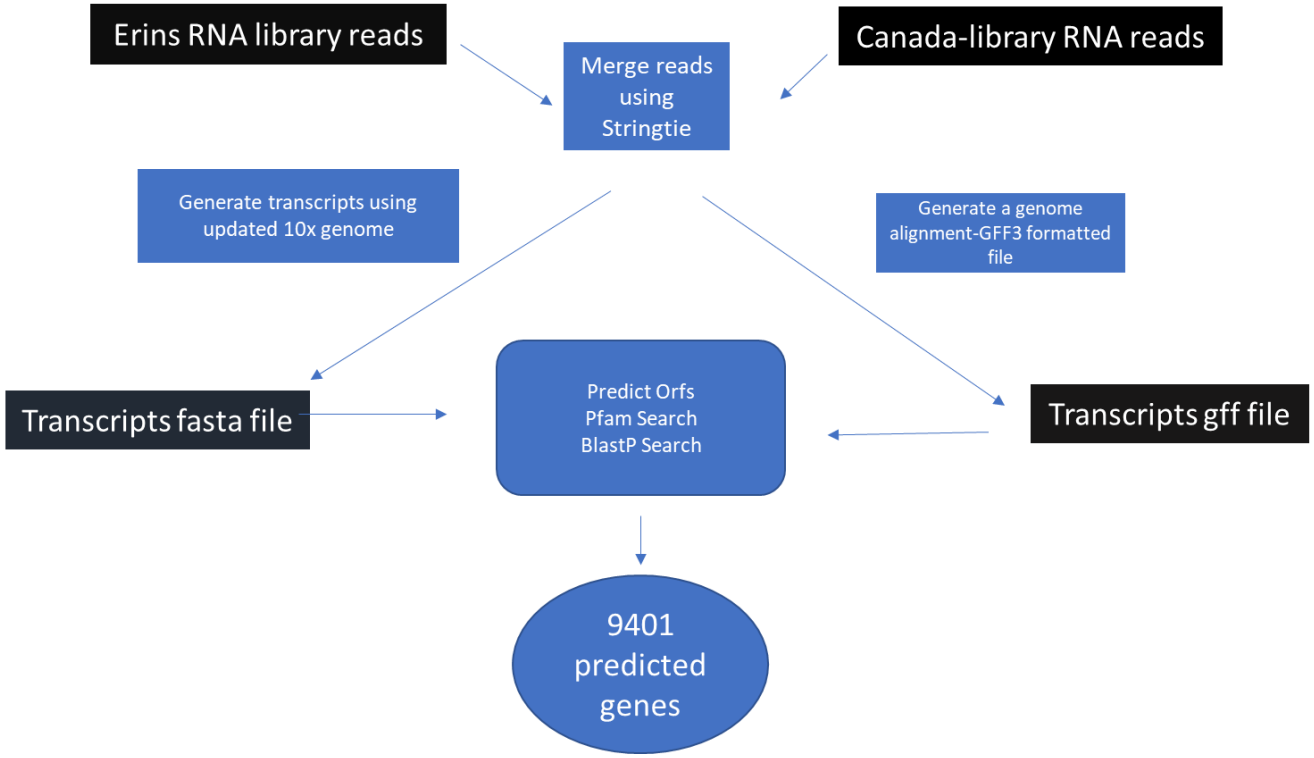

Figure 2.2: Overview of *Fasciola hepatica* genome coding regions prediction using Transdecoder. Gene models do not seem to match Braker *ab initio* predictions and MAKER SNAP trained annotations. We observed that small genes seemed to have been filtered out.

2.5.4. Observations on Manual Assessment of Annotation

For insights into the quality of the re-annotated *F. hepatica* assembly (*Fasciola*_10x_pilon, GCA_900302435.1), the compiled genes from the three relatively well described *F. hepatica* gene families (Tubulin, GST, and AC genes) were aligned to the *F. hepatica* genome manually (Figure 2.3 & 2.4). Using IGV, the current *Fasciola*_10x_pilon, GCA_900302435.1 annotation on WormBase Parasite (WP15) was compared with the re-annotated version described in this project using genes from these three gene families (see Chapter2_files appendix - F10 for screenshots of manually verified gene models).

2.5.4.1. Tubulin Genes

Sequences of the 5 α-tubulins and 6 β-tubulins previously described (Ryan et al., 2008) were each aligned to the *F. hepatica* genome and used as guide to identify other tubulins in genome. Each gene was first aligned to the *F. hepatica* annotation on wormbase parasite to find all possible hits. Genes identified as hits were compiled and checked to confirm if they have motifs common to tubulins. Also, these previously described tubulins were each aligned to the *F. hepatica* genome (using the *--model protein2genome* option to specify protein sequence to genomic sequence alignment) to identify best hits and manually visualise alignments. These α-tubulins (GenBank accession numbers AM933580–AM933584) were mapped to both annotations (the current *Fasciola*_10x_pilon, GCA_900302435.1 (WP15 version) and re-annotated version in this project) using IGV. In 4 of these α-tubulins, gene models predicted matched the CDS aligned to them, although one of the genes (*F. hepatica* α-tub 4 -AM933583) did not seem to match any predicted gene model in the *F. hepatica* annotation. This failure to identify the gene could be due to failure of annotation tools to identify the gene. Similarly, 6 β-tubulins previously described (AM933585–AM933590) were all identified in both *F. hepatica* genome annotations, with their gene structure matching those predicted in both annotations.

Using the presence of motifs to assign and gene alignment to the genome, additional tubulin gene were identified (gene identifiers used in this project refer to the gene names in the current *Fasciola*_10x_pilon, GCA_900302435.1 (WP15 version)). A total of 34 tubulin genes were identified in the *Fasciola hepatica* genome, 19 of which are α-tubulins, 11 are β-tubulins, 1 is a δ-tubulin (MAKER-scaffold10x_703_pilon-snap-gene-0.88), 1 is a γ-tubulin (MAKER-scaffold10x_1160_pilon-snap-gene-0.20), while 2 are unassigned (MAKER-scaffold10x_13_pilon-snap-gene-2.128 and MAKER-scaffold10x_500_pilon-snap-gene-0.52) tubulins (meaning it is unclear what type of tubulin they were).

2.5.4.2. Glutathione S-transferase (GST) Genes

Sequences of GSTs genes previously described (Morphew et al., 2012, Chemale et al., 2006) were used to manually assess both *F. hepatica* annotations. An NCBI search for complete CDS was carried out, with sequences aligned to the *F. hepatica* genome annotations. There was a similarity in gene structure in GST sigma (DQ974116.1) and GST mu (M77682.1) in both annotations (the current *Fasciola*_10x_pilon, GCA_900302435.1 annotation (WP15 version) and re-annotated version in this project). However, some exon variations were

noticed in GST Omega (JX157880.1) in *F. hepatica re-annotated* version suggesting missing exons were probably lost during the annotation process, as these were present in the current *Fasciola*_10x_pilon, GCA_900302435.1 annotation (WP15 version). A total of 17 GST genes were identified in the *F. hepatica* genome.

2.5.4.3. Adenylate Cyclase (AC) Genes

AC genes are reported to be differentially expressed in TCBZ resistant isolates when compared with TCBZ sensitive isolates (Radio et al., 2018). A total of 10 AC genes were identified (these gene were mapped and named based on an original draft *F. hepatica* annotation (assembly GCA_000947175.1, BioProject PRJEB6687) (Cwiklinski et al., 2015) which is no longer available on WormBase Parasite). In this re-annotation project, we used the European Nucleotide Archive (ENA) to extract these genes, mapped them to the current annotation and provide updated gene names. Interestingly only 2 genes (MAKER-scaffold10x_609_pilon-snap-gene-0.9 and MAKER-scaffold10x_102_pilon-augustus-gene-0.94) were identified, out of the 10 AC genes previous described (Radio et al., 2018) in the current *F. hepatica* annotation (*Fasciola*_10x_pilon, GCA_900302435.1). The RNAseq data also confirmed the expression of these two genes in various stages of the parasite, with higher expression levels in metacercaria and NEJs. Manual visualisation of gene - MAKER-scaffold10x_102_pilon-augustus-gene-0.94 in both annotations (the current *Fasciola*_10x_pilon, GCA_900302435.1 (WP15 version) and the re-annotated version in this project) revealed a single exon variation supported by Iso-Seq data transcript data (Figure 2.3), while in gene - MAKER-scaffold10x_609_pilon-snap-gene-0.9, gene models in both annotations match. However, Iso-Seq data suggests a probable longer gene or presence of alternative splicing (Figure 2.4). Apart from the 2 AC genes identified, 6 other AC genes were identified (although RNAseq data did not seem to support their expression) in the current *F. hepatica* assembly (*Fasciola*_10x_pilon, GCA_900302435.1) annotation. While the 6 identified AC genes could have been identified due to improving the *Fasciola*_10x_pilon, GCA_900302435.1 assembly, it is unclear why 8 previously identified AC genes were unidentified. This could mean gene annotation in the current *F. hepatica* assembly (*Fasciola*_10x_pilon, GCA_900302435.1) annotation was too strict, thus filtering out these 8 genes or they were wrongly assigned as genes previously. Thus, a of total of 8 AC genes were identified based on motifs present.

2.5.5. Result of Functional annotation and Orthologous Grouping

KEGG's GhostKoala, a web-based tool, was used to annotate genes and assign KO numbers to genes. This facilitated the identification of likely pathways and biological functions of annotated genes. It was noticed that gene models predicted in this project had a reasonably poor assignment of putative function. Only 20.2% of the predicted genes were annotated, compared to the 55.9% annotated in *Fasciola*_10x_pilon annotation (GCA_900302435.1). This considerable difference in percentages could be due to the poor annotation quality of predicted gene models in this re-annotation project. Despite these low annotation values, it was noticed that the genes annotated in both annotations play essential roles in genetic

information processing, signalling and cellular processes (Figure 2.5). Ontology assignment of predicted genes to assign putative molecular, biological, and cellular functions to predicted genes was done using PANNZER2 (see supplementary information).

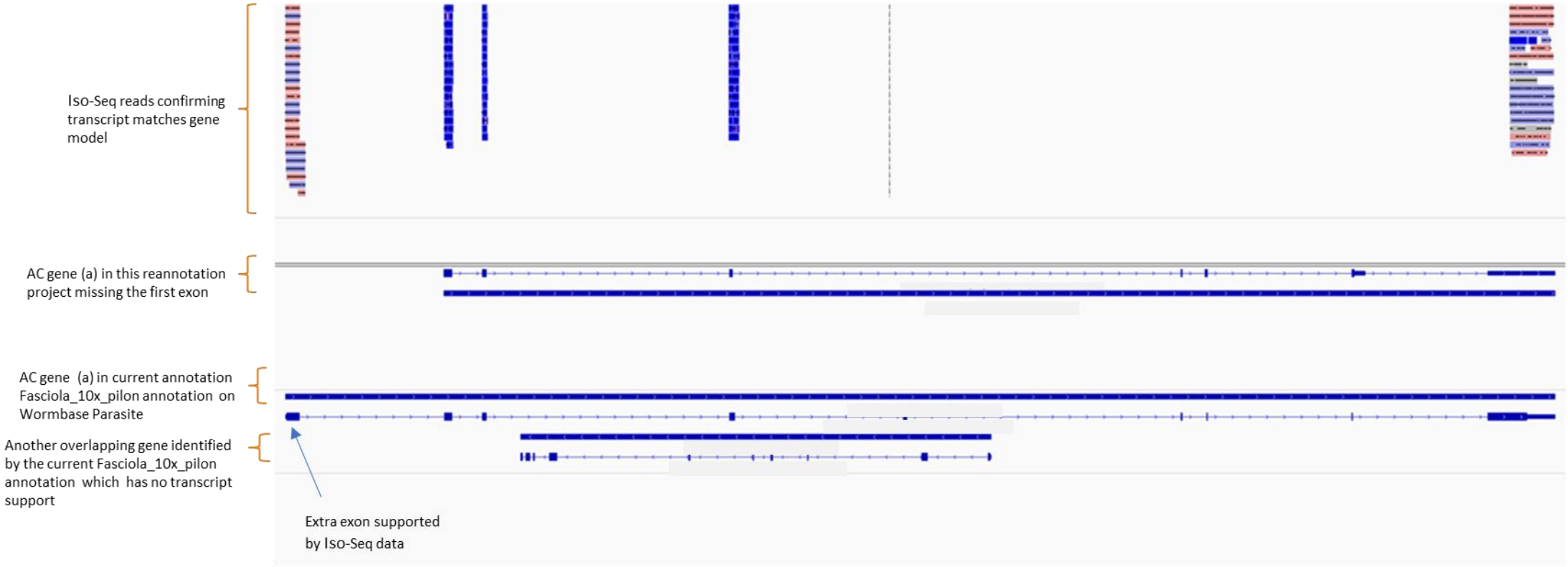

Figure 2.3: Validation of gene models. Structure of AC gene (a) - *MAKER-scaffold10x_102_pilon-augustus-gene-0.94**, an Adenyl Cyclase gene showing variation in both annotations (arrow showing omitted exon). Iso-Seq data confirming exon structure in the currently available *Fasciola*_10x_pilon annotation (GCA_900302435.1), while the new re-annotated version described in this project is missing the first exon. Also, there is no support for the small overlapping gene model and some of the exons, suggests either the annotation is wrong, or the gene is alternatively spliced. *Gene ID correspond to gene names in the current *F. hepatica Fasciola*_10x_pilon annotation GCA_900302435.1 (WP15).

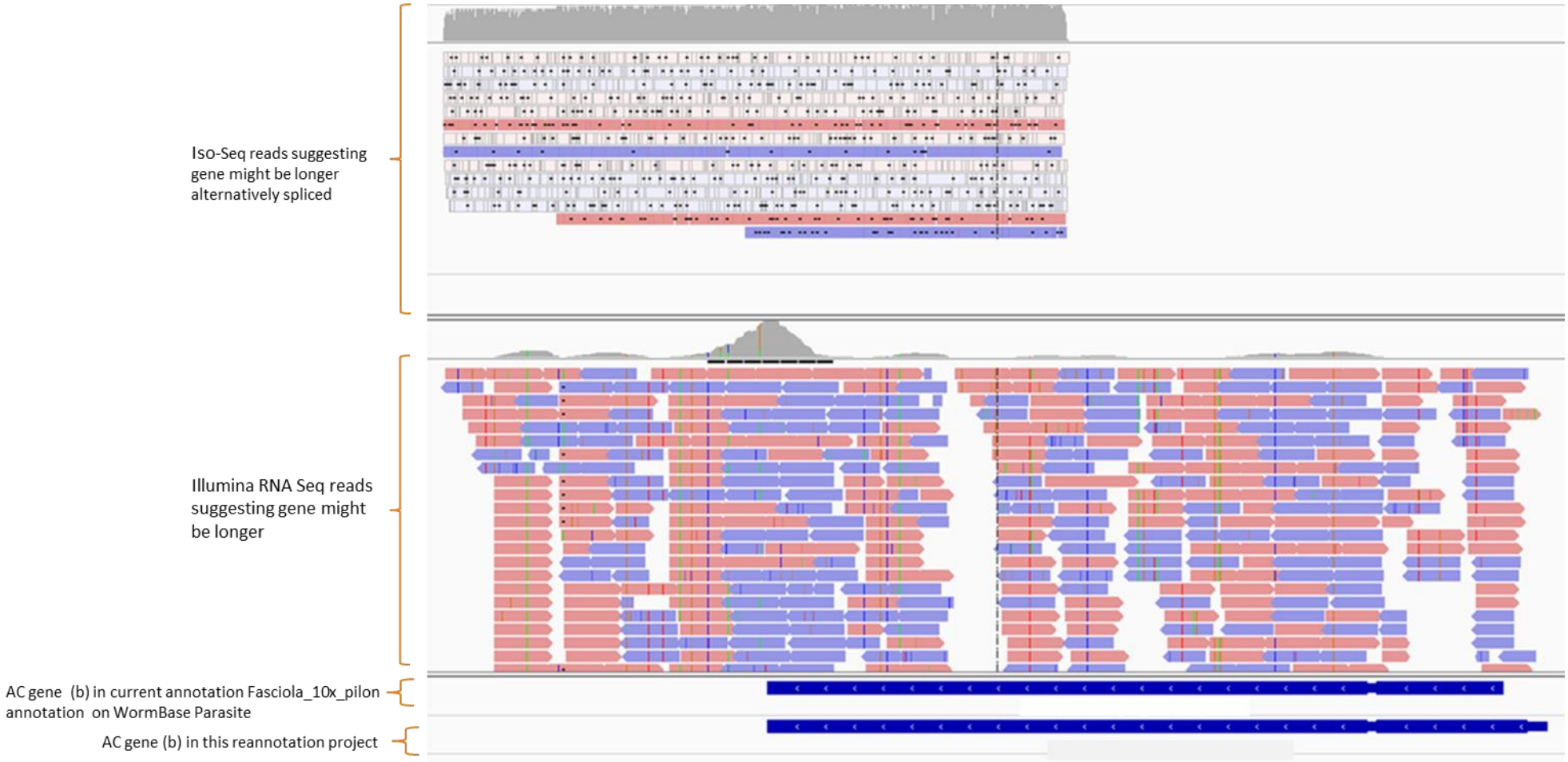

Figure 2.4: Validation of gene models. Gene structure of AC gene (b) - *MAKER-scaffold10x_609_pilon-snap-gene-0.9**, showing a similarity in both gene models from both annotations (Iso-seq and Illumina reads are presented in same scale). However, while gene models mapped to RNA-reads better (initially suggesting the possibility of two different genes), the Iso-Seq data suggests the gene is probably longer with some missing exons in the early parts of the gene or alternatively spliced. Alignment of the gene's *S. mansoni* ortholog (Smp_102340.1) which has two mapped transcripts of different length (alignment not shown) suggests gene is alternatively spliced in *F. hepatica*. *Gene ID correspond to gene names in the current *F. hepatica Fasciola*_10x_pilon annotation GCA_900302435.1 (WP15), while the different tracks (in pink and light blue colour) show RNA-reads (forward and reverse respectively).

Table 2.1: Summary statistics of hard repeat masking. Repeat masking of *Fasciola hepatica* 10x reference genome using RepeatMasker Combined Database Dfam_Consensus-20170127, and RepBase-20170127, run with rmblastn version 2.2.23+.

Feature	Number of elements	Length occupied (Kbp)	Percentage of sequence
Retroelements	304,585	109	8.88
Sines	8,566	1.0	0.08
Lines	207,916	72	6.02
LTR elements	88,103	33	2.78
DNA transposons	56,780	5	0.42
Unclassified	5,284	1.0	0.07
Total interspersed repeats		113	9.37
Small RNA	3,552	0.3	0.03
Satellites	9,476	0.8	0.07
Simple repeats	154,473	8	0.69
Low complexity	5,704	0.2	0.02
Total length	1,203,652,875		
GC level	44.09%		
Bases masked	121 Kbp (10.09%)		

Table 2.2: Summary statistics contrasting *Fasciola hepatica* re-annotation processes.

	*Fasciola hepatica Fasciola*_10x_pilon, GCA_900302435.1	*Fasciola hepatica* re-annotation using MAKER	Coding region prediction using Transdecoder
Total sequence length	1,203,652,875	1,203,652,875	1,203,652,875
Number of genes	9,709	15,879	9,401
Number of mRNAs	9,709	15,879	15,886
Number of exons	78,852	115,479	131,860
Number of introns	69,143	99,600	115,974
Number of CDS	9,709	15,879	15,886
Overlapping genes	1764	4,690	1,166
Contained genes	465	1,493	327
CDS: no stop, no start	9,709	15,879	15,886
Total gene length	446,630,547	484,448,487	346,258,040
Total mRNA length	446,630,547	484,448,487	613,944,962
Total exon length	32,889,994	30,353,722	50,536,915
Total intron length	413,878,839	454,293,965	563,639,995
Total CDS length	16,627,596	26,587,626	26,417,715
Shortest gene	203	186	298
Shortest mRNA	203	186	298
Shortest exon	3	3	1
Shortest intron	4	4	20
Shortest CDS	21	9	258
Longest gene	1,019,584	564,059	545,213
Longest mRNA	1,019,584	564,059	545,212
Longest exon	415,522	23,504	23,504
Longest intron	510,797	486,413	471,137
Longest CDS	31,164	25,272	25,194
mean gene length	46,002	30,509	36,832
mean mRNA length	46,002	30,509	38,647
mean exon length	417	263	383
mean intron length	5,986	4,561	860
mean CDS length	1,713	1,674	1,663
% of genome covered by genes	37.1	40.2	28.8
% of genome covered by CDS	1.4	2.2	2.2
mean mRNAs per gene	1	1	2
mean exons per mRNA	8	7	8
mean introns per mRNA	7	6	7

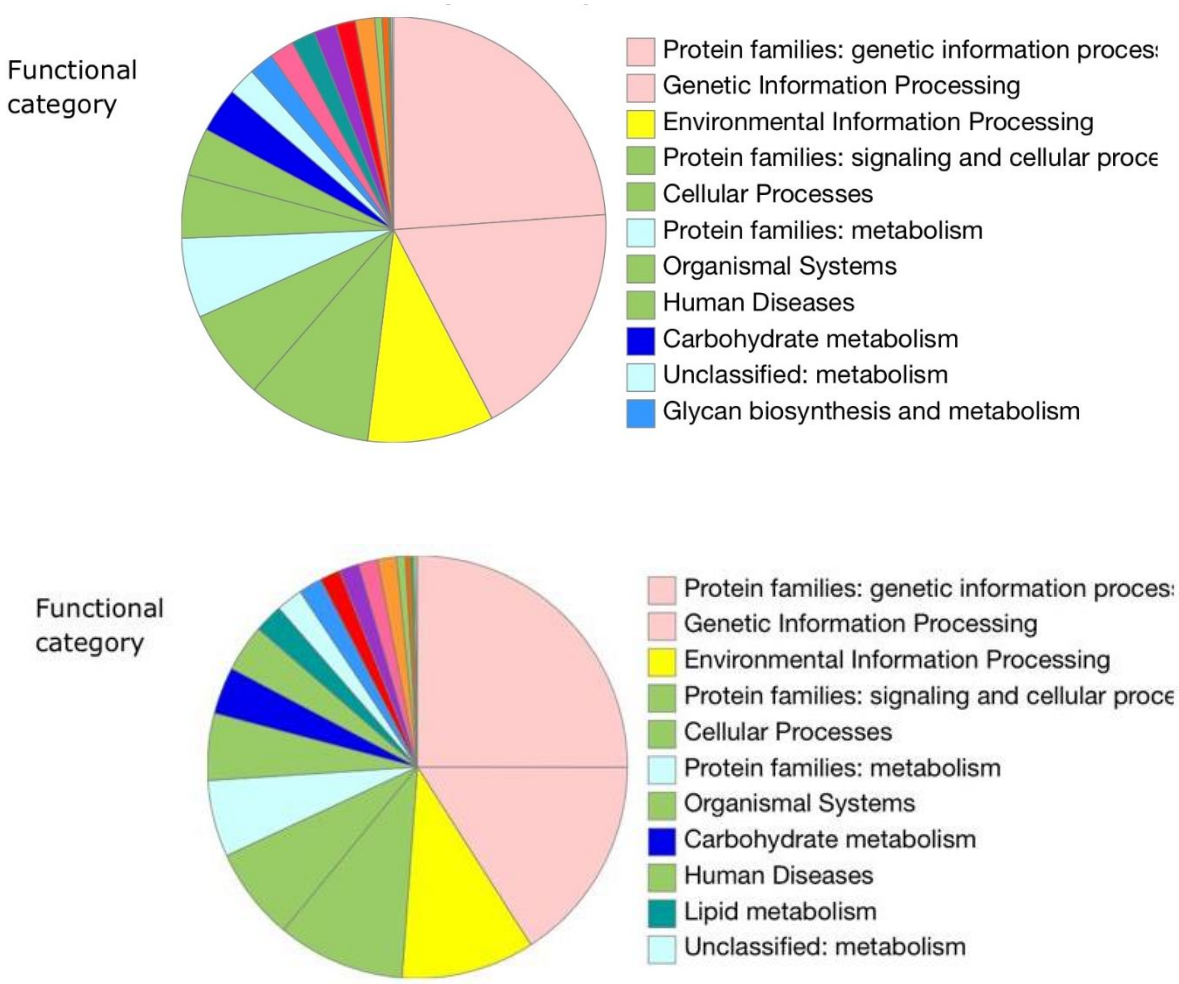

Figure 2.5: Summary of functional annotation of predicted gene models using KEGG's GhostKoala software. Out of the 15,879 predicted genes in this re-annotation project, only 3,210 genes were annotated (20.2%) - *top*. However, in the currently available *Fasciola*_10x_pilon annotation (GCA_900302435.1, WP15), 5,430 of the total 9,708 genes were annotated (55.9%) - *bottom*. The reduced annotation percentage seen in this re-annotation project could be due to models' poor quality. Despite this, the percentage distribution of annotated genes in both annotations was similar. Database accessed on June 15, 2021

Table 2.3: Summary statistics of *Fasciola hepatica* re-annotation completeness. Assessment of genome completeness against sets of 255 Benchmarking Universal Single-Copy Orthologs (BUSCO) eukaryotic database revealing variations in assembly completeness with respect to annotation pipelines.

BUSCO Input Description	Source	No of genes/transcripts	Mode	Complete BUSCOs	Complete and single-copy BUSCOs	Complete and duplicated BUSCOs	Fragmented BUSCOs	Missing BUSCOs	Total BUSCO groups searched
Final re-annotated assembly*	MAKER (SNAP trained)	15,879	Protein	118 (46.3%)	117 (45.9%)	1 (0.4%)	22 (8.6%)	115 (45.1%)	255
RNAseq transcript assembly of RNAseq libraries 1 and 2 to the *Fasciola hepatica* reference draft genome**	Transdecoder	15,886	Protein	210 (82.4%)	140 (54.9%)	70 (27.5%)	13 (5.1%)	32 (12.5%)	255
Current annotation available on WormBase Parasite	MAKER (SNAP trained)	9,709	Protein	193 (75.7%)	191 (74.9%)	2 (0.8%)	27(10.6%)	35 (13.7%)	255
Ab initio annotation of RNAseq dataset 1 (Erins-library)	Braker	53,729	transcriptome	46 (18.0%)	45 (17.6%)	1 (0.4%)	66 (25.9%)	143 (56.1%)	255
Ab initio annotation of RNAseq dataset 2 (Canada-library)	Braker	74,307	transcriptome	81 (31.7%)	74 (29.0%)	7 (2.7%)	89 (34.9%)	85 (33.4%)	255

** *Fasciola*_10x_pilon annotation (GCA_900302435.1) available on WP15 database, used as a reference in the results presented.

*See methods section for a detailed description of annotation methods used.

2.6. Discussion

2.6.1. Comparing the initial annotation and this re-annotation project

In this project, we re-annotated the *Fasciola*_10x_pilon assembly (GCA_900302435.1) using two RNAseq datasets previously described. A total of 15,879 *F. hepatica* genes were predicted compared to the 9,732 genes and 16,830 gene transcripts previously reported and available on WP15. While there seems to be marked gap in the number of genes called in both annotations, the number of genes in trematode annotations on WP16 ranged from 9,314 in *Schistosoma haematobium* to 26,189 in *Schistosoma margrebowiei*. Although improving the assembly quality improves quality of annotation. For example, in the recently improved *S. mansoni* genome project, using PacBio long-read and Illumina short-read sequencing, scaffold N50 improved from 32.1 Mb (version 5) to 52.8 Mb (Version 9). Also, the number of genes annotated reduced from 10,116 genes to 9,794 genes, with a 95.3% complete BUSCOs reported in the improved annotation (Buddenborg et al., 2021). The reduction in number of genes annotated in the improved assembly could be due to the increase in N50, as a higher N50 indicates increased average scaffold lengths. Interestingly, despite the reduction genome size in *S. mansoni* (391.4 Mb in *S. mansoni* versus 1.2 Gb in *F. hepatica*), number of genes predicted in both genomes was similar (9,794 genes in *S. mansoni* versus 9,732 genes in *F. hepatica Fasciola*_10x_pilon, GCA_900302435.1). It appears genome size does not influence number of genes predicted as previously suggested (International Helminth Genomes, 2019), although it appears there is a relationship between BUSCO annotation quality and number of genes predicted. For example, *F. hepatica* (*Fasciola*_10x_pilon assembly) with 9,732 genes had a higher BUSCO annotation quality than the *F. hepatica* (F_hepatica_1.0.allpaths.pg assembly) with 15,739 genes. Findings in this re-annotation project seems to confirm this (Table 2.3). The re-annotated project reported here with 15,879 genes had a poor BUSCO annotation (46.3%) compared to the *Fasciola*_10x_pilon assembly available on WP15 (with 75.7% complete BUSCOs)). Despite this, the Transdecoder annotation seemed to have performed better in annotating the assembly with 82.4% complete BUSCOs annotated, although it appears to have more duplicated genes (Table 2.3). Given the 95.3% complete BUSCOs identified in the newly improved *S. mansoni* annotation (Buddenborg et al., 2021), *F. hepatica* genome projects could benefit from a combination of long-read and short-read sequencing, although it would be interesting if this increases the current 75.7% complete BUSCOs identified in the current *Fasciola*_10x_pilon, GCA_900302435.1 annotation, improve *F. hepatica* annotation quality, and potentially reduce the number of genes annotated.

Using genes from three well-annotated families, validation of the re-annotated genome was done. With the aid of PacBio Iso-Seq transcript data, predicted gene models were manually explored using their transcripts to confirm gene boundaries, exon structure and identify alternative splicing using the Integrative Genomics Viewer (IGV) for visual exploration (Thorvaldsdóttir et al., 2013). Results (Figure 2.3 & 2.4) have shown that having full transcript lengths can facilitate validating genome annotations, enable comparison of

multiple annotations, and identify splicing events. This improves the reliability of gene models for downstream analyses. In this study, variations in gene structure such as missing exons, gene fragmentations, gene start, and end positions were noticed. Despite the variations in both annotations, there were matching gene models in both annotations (i.e., the current *Fasciola*_10x_pilon, GCA_900302435.1 (WP15 version) and re-annotated version in this project). Manual validation of gene models can be challenge due to the high number of genes in the genome (*as discussed below*); however, it can improve the reliability of an annotation project and facilitate understanding the organism's biology as shown in this project.

2.6.2. Repeat annotation is a crucial component of *F. hepatica* genome annotation

Eukaryotic genomes are well known for their high repeat levels. In practice, various stand-alone or a combination of tools are used to generate repeat libraries or use consensus libraries from repeat databases, after which these repeats are masked (Yandell and Ence, 2012). This study demonstrates that RepBase and consensus RepeatMasker Dfam Combined repeat database does not adequately identify repeats in the *Fasciola*_10x_pilon annotation (GCA_900302435.1) as shown by the low 10.9% repeats masked. This low level is similar to observations in the *P. westermani* annotation project where RepBase library was used (Oey et al., 2018). Interestingly, using the in-house *F. hepatica* repeat library from the initial annotation project previously described (Cwiklinski et al., 2015), the updated *F. hepatica* 10X genome's repeats level went up to 62.96% (*results not shown in this report*), compared to the 32% previously reported (the repeat masking results used in this project masked only 10.09% (see Table 2.1). This is markedly higher than the previously reported levels in closely related trematode annotation projects. In these reports repeat amount was 55.29% in *F. hepatica* (McNulty et al., 2017b), 46.85% in *F. gigantica* (Pandey et al., 2020), 45.2% in *P. westermani* (Oey et al., 2018), 30.6% in *O. viverrine* (Young et al., 2014), and 32% in *C. sinensis* (Huang et al., 2013). Interestingly, when we used another tool, Redmask, a de-novo repeat masking tool (Girgis, 2015), to identify the repeats in the same 10X updated *F. hepatica* genome, a repeat level of 36.19% was recorded. These findings highlight differences in repeats levels due to variation in the tools used, the robustness of the library, and genome assembly quality.

To confirm to what extent repeats contributed to this re-annotation project, repeats identified from the in-house *F. hepatica* repeat library from the initial annotation project previously described (Cwiklinski et al., 2015) which identified 62.96% genomic repeats level in the updated *F. hepatica* 10x, were mapped to re-annotated genome version in this project. Removal of genes that overlap repeats reduced the number of genes to 13,633 from the initial 15,879. Similarly, 36.19% repeats identified by Redmask (described above) was mapped to this genome re-annotation version, this reduced the genes to 13,970 from the initial 15,879 (Table 2.4). While the presence of genes overlapping repeats indicate repeats contributed to the higher number of genes identified in this re-annotation project to

an extent, this does not explain much of the variation in the number of genes in these annotations.

2.6.3. Gene models from ab initio annotation tools are not reliable

The term "*ab initio*" gene predictors are somewhat loosely used to describe genome annotation tools early in the field. A vital feature of software used is mathematical models such as Hidden Markov Models (HMMs) to detect specific genomic signals. Training algorithms in these tools can detect genomic features such as spicing sites, exon and intron boundaries, ORFs, promotor and terminator sequences in the genome (Scalzitti et al., 2020). Ideally, these tools are able to predict genes using a genomic reference sequence, for example, without any external evidence (Yandell and Ence, 2012). In this project, we used Augustus (Stanke and Morgenstern, 2005) as a stand-alone *ab initio* tool using the *Fasciola*_10x_pilon annotation (GCA_900302435.1) alone. This predicted 80,750 genes, which we find rather high. We also noticed gene models predicted were poor. Interestingly Augustus and Genescan have been shown to have the best accuracy scores when compared with three other *ab initio* predictors – GlimmerHMM, GeneID, and SNAP (Scalzitti et al., 2020). We find that Augustus alone is not good enough to predict genes in the *F. hepatica* genome.

It is known that although these algorithms often make errors, it has been proposed that the provision of adequate training data, RNAseq data, protein data, etc. can improve the sensitivity of *ab initio* tools to nearly 100% (Yandell and Ence, 2012); we find that this more ambitious than realistic. We find that using Braker (a genome annotation tool designed to work with Augustus and Genemark-ET) on the same updated 10X *F. hepatica* reference genome using RNAseq datasets 1 and 2 as evidence, a total of 100,527 and 52,257 genes were predicted. The high number of predicted genes and the poor quality of these gene models could be due to various reasons, such as fragmentation of genes, repeat regions predicted as genes, pseudogenes, RNA genes, etc. We do, however, find that when using these preliminary gene models with other evidence such as protein sequences from trematodes, we can improve annotation using the MAKER tool (Holt and Yandell, 2011).

2.6.4. Improving *F. hepatica* annotation with MAKER pipeline

In recent times the MAKER pipeline designed for annotating eukaryotic and prokaryotic genomes (Holt and Yandell, 2011), and adapted for plant genomes (Campbell et al., 2013) has been used for annotation projects across various species. Here we present a comparison of two annotation versions of *F. hepatica* genome, both of which was annotated with MAKER pipeline. Using the *Fasciola*_10x_pilon assembly (GCA_900302435.1), RNAseq dataset 2 previously described (Cwiklinski et al., 2015), and RNAseq dataset 1 as additional evidence, we have re-annotated the *F. hepatica* genome using the MAKER pipeline. Here we identified 15,879 genes compared to the 9,709 genes (with a total of 16,830 transcripts) previously identified (BioProject PRJEB25283, Assembly: *Fasciola*_10x_pilon; GCA_900302435.1, WBPS15). The marked variation in the number of genes predicted (i.e 15,879 versus 9,709) in both annotations is unclear. This could be due to variations in

software versions or software prediction cut-offs, with the *Fasciola*_10x_pilon GCA_900302435.1, WBPS15 annotation more stringent in gene predictions. The total number of genes predicted in the re-annotated version described here is similar to the 15,739 genes (11,218 genes after re-annotation) reported in *F. hepatica* annotation project in Oregon, USA (McNulty et al., 2017b, Choi et al., 2020). These variations in the number of genes could be due to various computational reasons such as differences in dataset handling and filtering, changes in the software version, and variation in training parameters. These factors influence the specificity and sensitivity of annotation pipelines. A key contributing factor to the variation in the number of predicted genes in the *F. hepatica* annotation project in Oregon, USA, and the *F. hepatica* (*Fasciola*_10x_pilon GCA_900302435.1) annotation on Wormbase parasite could be due to the marked variation both assembly's N50. While the former has an N50 of 161 Kb, the later has an N50 of 1.9 Mb. A lower N50 could account for the higher number of genes due to increased possibility of gene fragmentation.

Despite these differences in annotation statistics, a comparative genomic study of 45 nematode and platyhelminth species estimated a range of 9K – 17K genes across species studied, all primarily annotated with the MAKER pipeline, while genome size ranges from ~0.1Gb to 1.3Gb in platyhelminths (International Helminth Genomes, 2019), with *F. hepatica* having the largest genome of ~1.3 Gb (Cwiklinski et al., 2015). To achieve reliable annotation, continuous re-annotation is key. However, this is challenging considering time and computational requirements (Jung et al., 2020), especially if all annotation phases are to be repeated. We observed that a combined exploration of the old and new annotations from MAKER pipeline, the PacBio Iso-Seq gene transcript data, coupled with other outputs from other tools such as Exonerate and Transdecoder, provided a better understanding of the genome.

2.6.5. Manual validation using candidate gene families facilitates identifying more family members

A significant aspect of our re-annotation project included extensive manual validation of gene models using Sanger sequenced CDS of genes previously well described and Iso-Seq reads (Table 2.3 & 2.4). Using a combination of sequence mining and blast, we aligned sequences for each gene of interest to the *F. hepatica* 10X genome to assess the mapping of the gene to the genome. Our findings indicate that this approach, despite being time demanding, effectively assesses the quality of gene models from the annotation pipelines. However, the lack of detailed description of most *F. hepatica* genes on NCBI, incomplete CDS, and the unavailability of sequences for most genes limit the validation phase's scope. Generally, *F. hepatica* genes on NCBI are either fragmented or have putative names, which do not fully describe a gene of interest. For example, cathepsin genes have been extensively studied in liver flukes and have been associated with parasite's penetration of host tissues as well as invasion of the host's immune system (Dalton et al., 2003, Robinson et al., 2008, Smooker et al., 2010, Sansri et al., 2015). We attempted to validate our annotations with

cathepsin genes but observed inconclusive findings. Generally, while gene models predicted by MAKER pipeline were generally better than ones by Braker, we noticed MAKER did not identify some genes captured by Braker. Typically, when reference genes pulled from NCBI were compared with MAKER predicted models, our predicted models had extra exons, meaning longer transcripts. This suggests these reference CDS on NCBI were either incomplete or our models were wrong.

In this project, we used previously described genes in the tubulin, GST, and AC genes. We selected these gene families because they are well described, and complete CDS are available. We have identified a total of 34 tubulin genes, 17 GSTs, and 8 AC genes in *F. hepatica*. The number of tubulins and GSTs genes identified was more than those previously described in each family (Ryan et al., 2008, Chemale et al., 2006, Morphew et al., 2012, LaCourse et al., 2012). This could be because genomic sequencing and computational tools offer an opportunity to capture these genes. Here, 8 AC genes were reported compared to the 10 AC genes previously described (Radio et al., 2018). The inability to access these 10 AC previously identified genes due to their inaccessibility on databases such as WP15 highlights challenges associated with updating assemblies and annotations as this can lead to loss or gain of additional genes in the updated annotation. Identifying all gene members in a family is important to the biological investigation of gene expression profile and other evolutionary investigations such as gene expansion events and gain or loss of gene functions. Although we have solely used these three gene families to validate the genome, using more gene families will be beneficial if complete CDS are available. However, the lack of full CDS in most *F. hepatica* available in most databases, poor gene description and challenges associated with manually validating all the genes in the genome makes validation of the annotation a difficult task.

2.7. Is parasitism in liver flukes related to big genome size?

F. hepatica genome sequencing projects have shown the parasite has a big genome size (of about 1.3 Gb) compared to other trematodes (see chapter 1, Table 1.3). In liver fluke research its widely thought that increased genome size is a trait that emerged as the *Fasciola* emerged; and could play a key role in parasite adaptability and host invasion. Comparatively, in free-living planarian - *Schmidtea mediterranea* genome project, a moderately large genome size of 781.7 Mb was reported. Also, a large repetitive content of 61.7 % was reported in the genome (Grohme et al., 2018). Considering the genome size and number of repeats is similar in *S. mediterranea* and liver flukes, parasitism in *Fasciola* may not be linked to its big genome. Also, the lack of a suitable outgroup makes testing whether *Fasciola* genome size can be particularly termed large.

2.8. Conclusion

The *F. hepatica* genome was re-annotated using two RNAseq datasets as evidence. A total of 15,879 genes were identified. We compared the new annotation with the version currently on WormBase Parasite database (*Fasciola*_10x_pilon, GCA_900302435.1, WP15). Using PacBio Iso-Seq transcript dataset and complete CDS sequences of tubulins, GSTs and AC

genes, gene models predicted were manually validated. Key limiting factors include the computational time required for running annotation tools considering the genome is relatively large, lack of representative complete CDS for many *F. hepatica* genes, laborious nature of manual validation of genomes, and annotation error in pipelines. Despite these factors, the benefit of this project is the improved reliability of gene models from the gene annotation pipeline and the availability of a comprehensive gene list from families of interest for downstream studies. Findings have facilitated functional annotation, gene expression and evolutionary studies (see chapters 3 and 4 of this thesis), and an improved understanding of the parasite's biology.

Table 2.4: Comparison genome annotations statistics showing the impact of repeats on predicted gene number.

	Fasciola hepatica Re-annotation using MAKER	Re-annotation filtered by overlapping repeats from the in-house library	Re-annotation filtered by overlapping Redmask identified repeats
Total sequence length	1,203,652,875	1203652875	1203652875
Number of genes	15,879	13633	13970
Number of mRNAs	15,879	13633	13970
Number of exons	115,479	57346	72182
Number of introns	99,600	47011	59803
Number of CDS	15,879	10272	12320
Overlapping genes	4,690	172	230
Contained genes	1,493	25	57
CDS: no stop, no start	15,879	13633	13970
Total gene length	484,448,487	6529343	12297952
Total mRNA length	484,448,487	6529343	12297952
Total exon length	30,353,722	12920587	15174891
Total intron length	454,293,965	268077496	337973224
Total CDS length	26,587,626	10267965	12116723
Shortest gene	186	1	1
Shortest mRNA	186	1	1
Shortest exon	3	1	1
Shortest intron	4	4	4
Shortest CDS	9	1	1
Longest gene	564,059	12712	22074
Longest mRNA	564,059	12712	22074
Longest exon	23,504	8823	8823
Longest intron	486,413	486413	512823
Longest CDS	25,272	22971	23108
mean gene length	30,509	479	880
mean mRNA length	30,509	479	880
mean exon length	263	225	210
mean intron length	4,561	5702	5651
mean CDS length	1,674	1000	984
Percentage of genome covered by genes	40.2	0.5	1.0
Percentage of genome covered by CDS	2.2	0.9	1.0
mean mRNAs per gene	1	1	1
mean exons per mRNA	7	4	5
mean introns per mRNA	6	3	4

CHAPTER 3

Assessment of selective pressure in selected *Fasciola hepatica* gene families

3.1. Background

Molecular evolution mechanisms are due to positive selection, negative selection, and neutral processes such as random genetic drift (Haller and Messer, 2017, Smith and Eyre-Walker, 2002). While negative or purifying selection can lead to removal of new genetic variants that decrease fitness, variants that increases fitness can become fixed in the population via positive selection. Identifying positive selection is a common way of identifying adaptive changes in an organism. Positive selection is the process whereby new advantageous genetic variants become fixed in a population (Zhang et al., 2005). Most genetic mutations are selectively neutral, meaning they have no effect on fitness (ability to survive and reproduce) (Duret, 2008, Kimura, 1983). Thus, only a small fraction of these mutations affects the fitness in a population. This fraction, i.e., adaptive mutations can spread quickly in the population because they provide a fitness advantage in the organism. New advantageous variants increase in frequency, become fixed in the population in a process referred to as directional (or positive) selection, while those of no advantage are lost in the population via purifying (or negative) selection (Duret, 2008, Harris, 2018). Positive selection in protein coding genes is commonly evaluated by analysing nucleotide changes in an organism (for example using the McDonald and Kreitman test) and/or by doing molecular sequence analysis (such as the PAML tests) (Andolfatto, 2005, Murrell et al., 2013, Yang, 1997). By identifying genes evolving under positive selection, organism adaptation in the environment is better understood (Cole and Brewer, 2018, Murrell et al., 2013). The challenge is, identifying signals indicative of positive selection in protein-coding genes can be masked by negative selection (Zhang et al., 2005), but when successful, it can indicate how protein domains can be changed by sites under positive selection (Wagner, 2007).

A commonly method used in identifying positive selection is to compare synonymous and nonsynonymous changes in aligned genes. The ratio of nonsynonymous substitutions to synonymous substitution ($\omega = dN/dS$) per site is indicative of selection pressure, where $\omega > 1$, $\omega < 1$, and $\omega = 1$, indicates positive selection, negative (purifying) selection, and neutral selection respectively in the species examined (Yang et al., 2000). These methods have been used to assess positive selection pressure in genes such as the β-globin in vertebrates, hemagglutinin gene from the human influenza virus A, etc. (Yang et al., 2000). Practically, detecting a significant $\omega > 1$ value would assume that a high proportion of amino acids are under positive selection. However, $\omega > 1$ is stringent and not easily achieved because averaging substitution rate across sites in the sequence reduces the power to identify positive selection. To address this issue, various codon site-specific, lineage-specific and branch-site models are used to detect positive selection signals (Yang and Nielsen, 2002). The lineage-specific models allow variation of ω across lineages and not among sites. The site-specific models allow variation of ω across sites not in lineages. In contrast, the branch-site models allow for variation in sites and lineages and detect positive selection on specific lineages of interest, commonly called the foreground branch (Yang, 2007). Despite the advantage of these various models, their computational complexities are enormous.

3.2. Challenges of Assessing Positive Selection and available tools

Assessing dN/dS is commonly done using the codeml software in PAML (Yang, 1997). The software requires strict input files, nucleotide sequences and the phylogenetic tree of a group of genes of interest, and a complex set-up file where model parameters are stated. Due to these issues, analysis are generally realistic on a single-gene basis, thus making genomic studies challenging (Webb et al., 2017). Other challenges include data preparation, orthologous grouping, alternatively spliced gene variants, alignment and phylogenetic tree errors, interpretation, and presentation of results. Thus, these issues tend to cause low reproducibility of published findings (Huerta-Cepas et al., 2016, Sahm et al., 2017). For example, stop codons and indels in aligned sequences limit the codeml package from working. Minimisation of these computational issues requires automation. Thus, various tools have been developed mainly through automating the codeml package to test multiple models of interest. Some packages designed to assess positive selection pressure include VESPA (Webb et al., 2017), PosiGene (Sahm et al., 2017), POTION (Hongo et al., 2015), FUSTr (Cole and Brewer, 2018), ETEtookit (Huerta-Cepas et al., 2016), JCoDA (Steinway et al., 2010), DnaSP (Librado and Rozas, 2009), and web-based ones such as Selection (Stern et al., 2007) and Datamonkey (Delport et al., 2010). Generally, the likelihood ratio test (LRT) is used to detect positive selection by comparing a null model (a relaxed model which does not allow $\omega > 1$) with the alternative model (a positive selection model which allows $\omega > 1$) (Yang and Nielsen, 2002, Yang, 2007).

3.3. Assessing Positive Selection using McDonald and Kreitman test

As previously described, ω = dN/dS ratio is used to assess selection selection pressure using divergence data among related species (Yang, 1997). A $\omega = 1$ indicative of neutral evolution can be evaluated, using a combination of polymorphism data ("within species") of a species of interest and a divergence data of the species of interest with other related species ("between species"). A McDonald and Kreitman test (MKtest) can be used to identify a recent positive selection signal, an indicator of adaptive evolutionary change (McDonald and Kreitman, 1991). In the MKtest, the number of divergent nonsynonymous sites, Dn (or Fixed Nonsynonymous sites, Fn), divergent synonymous sites, Ds (or Fixed Synonymous sites, Fs), polymorphic nonsynonymous sites (Pn), and polymorphic synonymous sites (Ps) are used as an indicator of selection pressure. A positive selection pressure signal is indicated mathematically by $(Fn/Fs) > (Pn/Ps)$ (Parsch et al., 2009), while the neutrality index (NI) is indicated by the mathematical value of $(Pn/Ps)/(Fn/Fs)$ for a gene of interest. The MKtest is a powerful but conservative tool used to test the proportion of substitutions in sequence driven by positive adaptive evolution (Eyre-Walker, 2002). In recent times various web-based tools have been developed based on the principle of MKtest, such as asymptoticMK (Haller and Messer, 2017), iMKT (Murga-Moreno et al., 2019), MKT-website (Egea et al., 2008). Despite the conservative nature of the test observed as undervaluing positive selection pressure due to slightly deleterious mutations, the test can provide a reasonable estimation of the proportion of mutations driven by positive selection at protein-coding sites (Messer and Petrov, 2013, Parsch et al., 2009).

3.4. Evolutionary Biology and Liver Flukes

Previous genomic studies on liver fluke – *F. hepatica* have stressed the parasite's genetic variability and highlighted the capability for rapid adaptation to host, weather, and control measures such as drugs and vaccines (Cwiklinski et al., 2015). The hermaphrodite nature of the parasite means self-fertilisation (a form of in-breeding) can occur, potentially leading to rapid spread of drug resistance genes in a population (Beesley et al., 2017b, Wolstenholme et al., 2004, Hurtrez-Boussès et al., 2001). Despite this, substantial genotypic diversity has been reported in the *F. hepatica* population (Beesley et al., 2017b). Expansion of some gene families has also been reported, such as cathepsins (McNulty et al., 2017b), a group of cysteine proteases known for their roles in host immune invasion (Dalton et al., 2003). Other expanded gene families include fatty-acid-binding proteins, asparaginyl endopeptidases (legumain), protein disulfide-isomerases and molecular chaperones. These are genes that play essential excretory and secretory roles in the parasite, facilitate host invasion and immune modulation, and are important to parasite adaptation to host immunity (Choi et al., 2020, Jefferies et al., 2001).

3.5. Drug Resistance in *F. hepatica*

The growing spread of resistance to anthelmintic drugs (predominantly triclabendazole, TCBZ) and the lack of availability of an effective vaccine is a significant concern to livestock production. Anthelmintic administration eliminates susceptible liver flukes; however, the inability to kill resistant parasites means resistant genes are quickly passed to the next generation (McManus et al., 2014). Coupled with the self-fertilisation potential and lack of population structure in flukes (Beesley et al., 2017b), the risk of the spread of resistant genes is high. Due to over-reliance on TCBZ, there is a growing worldwide threat of resistance to TCBZ, the only drug able to kill the adult flukes and the early immature ones (these cause damage to the host organs due to their migratory actions) (Kelley et al., 2016). There is a need for more research into parasite biology, understanding TCBZ resistance, and identifying potential new drug targets.

Biochemically, resistance to a drug can occur through an alteration in its target, such that it becomes ineffective since it cannot bind to the target. An alteration in the metabolism of a drug could lead to early inactivation or excretion, while a change in its delivery could limit drug accessing the acting site. Similarly, gene targets could be amplified to prevent or overcome drug action (Wolstenholme et al., 2004). Unfortunately, the mechanism of resistance of *F. hepatica* to TCBZ is not clearly understood. Various genes have been implicated and investigated due to their roles in the parasite's biology, while various drug resistance pathways have been proposed. Recently the resistance to TCBZ has been grouped into three likely mechanisms. These include; alteration in TCBZ target, an alteration in TCBZ uptake, and TCBZ metabolism (Fairweather et al., 2020). Various *F. hepatica* genes have been implicated and investigated to understand their role in the biology of the parasite and elucidate TCBZ resistance (Table 3.1).

3.6. Importance of Assessing Positive Selection Pressure in *F. hepatica*

Various genes have been explored to understand *F. hepatica* resistance to drugs (especially TCBZ) and identify potential drug targets (Table 3.1). The role of these *F. hepatica* gene families in drug resistance and vaccine prospects has been researched and reviewed to varying degrees (Ryan et al., 2008, Radio et al., 2018, Kelley et al., 2016, Toet et al., 2014). There is need to explore genes associated with TCBZ resistance from an evolutionary perspective. Positive selection signals have already been identified in *F. hepatica* Cathepsin, FhCatL5 (Irving et al., 2003) and G-protein-coupled receptors (GCPRs) (Choi et al., 2020) using the dN/dS and MKtest, respectively. The known roles of *F. hepatica* cathepsin L proteases in parasite migration in the host via tissue degradation and host immune modulation (Mulcahy and Dalton, 2001), as well as GCPRs' vital role in membrane signal transduction and helminth neuromuscular activity (McVeigh et al., 2017), make using them as drug targets challenging, especially when these gene targets are under adaptive pressures. Positive selection pressure on sites on a gene implies alleles advantageous to the organism are passed down the population (vertically). For example, these could be genes responsible for drug resistance; thus, a species could be resistant to a drug while a related species is susceptible to the drug. Thus identifying *F. hepatica* gene families under positive selection is essential to understand how the organism adapts rapidly, explain liver fluke resistance to drugs such as TCBZ, identify new drug targets, and explain differences in dug susceptibility.

This study assessed positive selection pressure in *F. hepatica* gene families that are of interest in recent research using a candidate gene approach. Genomic studies are now possible since the two *F. hepatica* genomes (Cwiklinski et al., 2015, McNulty et al., 2017b) were published. However, comprehensive annotation of genes to their families and their respective function roles is lacking. Focusing on various gene families that have been investigated for their importance in *F. hepatica* biology and its resistance to drugs (Fairweather et al., 2020), we have compiled a list of genes for each gene family and their orthologous sequences from other trematodes with available genomes. Using this information, we have analysed each gene of interest for signals indicative of positive selection using codeml site models (Yang, 1997) and MKtest (McDonald and Kreitman, 1991) in the coding sequences of these genes.

3.7. Materials and Methods

For each gene family of interest (Table 3.1), *F. hepatica* genes were compiled using previously described gene sequences on the NCBI database and literature. After which an alignment of each gene to the *F. hepatica* genome was done using WormBase Parasite (WP15) and Exonerate (version 2.2.0, to manually verify alignment quality) (Slater and Birney, 2005) was performed to identify the best hit using default settings. All gene family members were compiled (when possible, gene names are indexed to the *F. hepatica Fasciola*_10x_pilon GCA_900302435.1 assembly, 2020-05-WormBase Annotation version), with each gene included or eliminated based on the presence of motifs unique to the family

using Pfam 34.0 (Finn et al., 2016, Mistry et al., 2020). Using the genome proteins from selected representative trematodes, orthologous grouping was done using OrthoFinder (version 2.3.12) using default settings (Emms and Kelly, 2019). Genomes used for orthologous grouping (Table 3.2 & 3.3) are *F. hepatica* - PRJEB25283 (Cwiklinski et al., 2015), *Clonorchis sinensis* - PRJDA72781 (Huang et al., 2013), *Echinostoma caproni* - PRJEB1207 (International Helminth Genomes, 2019), *Fasciola gigantica* - PRJNA230515 (Choi et al., 2020), *Fasciolopsis buski* - PRJNA284521 (Choi et al., 2020), *Opisthorchis viverrine* - PRJNA222628 (Young et al., 2014), *Paragonimus westermani* - PRJNA454344 (Oey et al., 2018), and *Schistosoma mansoni* - PRJEA36577 (Protasio et al., 2012). After the orthofinder analysis, the orthologous grouping for each *F. hepatica* gene of interest was extracted. In other to confirm reliability of gene ortholog assignment results, the motifs in each gene in an orthogroup was checked using Pfam against previously identified motifs in the gene family.

3.7.1. Assessing Positive Selection using PAML

For a *F. hepatica* gene of interest in an orthologous group (orthogroup), the orthogroup protein sequences and corresponding nucleotide coding sequences were used for codon-based alignment using Pal2nal (version 12) (Suyama et al., 2006). To reduce analysis errors, Pal2nal analysis command was set up with *-nogap* option used to remove alignment gaps. After which phylogenetic trees analysis and positive selection analysis was done using ETE3 – the Etetookit pipeline (Huerta-Cepas et al., 2016). Phylogenetic trees were constructed using a standard RAxML workflow (Stamatakis, 2014), after which positive selection analysis was done using codeml tool in PAML for site model testing in ETE3 (Yang and Nielsen, 2002, Yang, 2007, Huerta-Cepas et al., 2016). Models M7 for relaxation (beta) and M8 for positive selection (beta & ω) were tested, and LRTs compared ($P < 0.05$ was considered statistically significant). Sites under positive selection pressure were identified using the Naive Empirical Bayes (NEB) analysis and Bayes Empirical Bayes (BEB) analysis. Sites identified by NEB analysis that are statistically significant ($P < 0.05$) are generally regarded to be under positive selection.

In other to confirm if sites identified to be positive selection using the site models indicate positive selection pressure is in the *F. hepatica* lineage, branch and branch-site models were carried out. branch-site (bsA - positive-selection, bsA1 – relaxation) and branch (b_free-positive-selection, b_neut – relaxation) models were carried out by labelling *F. hepatica* as the foreground. Similarly to the site models, only sites identified by the branch-site model NEB analysis that are statistically significant ($P < 0.05$) are generally regarded to be under positive selection. Given it has been shown that the NEB method has a higher possibility of reporting false positives compared to the BEB analysis (Yang et al., 2005), sites identified on selection by NEB method was generally compared with results presented by BEB analysis.

3.7.1. Protein model prediction

In other to predict the effect of sites under positive selection in *F. hepatica* genes on statistically significant positive selection. SWISS-MODEL tool was used to build the protein

model using default settings (Waterhouse et al., 2018). After which FTSite (https://ftsite.bu.edu/) was used to predict the protein active sites. Phymol (version 2.0.5) was used to make graphical presentation of protein models and label sites under positive selection (Ngan et al., 2012, DeLano, 2002), and evaluate proximity of sites under positive selection to the predicted active sites.

3.7.2. Assessing Positive Selection using MKtest

For each *F. hepatica* gene of interest, the direct *F. gigantica* ortholog was determined using Orthofinder (direct ortholog relationship between *F. hepatica* and *F. gigantica* was carried out as part of the orthofinder analysis previously described). Using the nucleotide coding sequences and corresponding protein sequences from each gene of interest, DnaSP (version 6.12.01) was used to determine the number of nonsynonymous and synonymous polymorphisms between both *Fasciola* species (Librado and Rozas, 2009).

In other to estimate polymorphisms within *F. hepatica*, the previously described Single Nucleotide Polymorphism (SNPs) dataset from the 5 different sequenced parasite isolates - FhepLivSP, FhepLivS1, FhepLivR1, FhepLivR2 and FhepLivR3 was used (Cwiklinski et al., 2015). Using the SNPs dataset (in VCF format) from the five isolates, the *F. hepatica* reference genome sequence (in Fasta format) and the genome annotation file (in GFF format), bedtools (version 2.22.1-13-g2d836be) was used to extract the polymorphisms in coding regions in each gene of interest. The effect of each polymorphism (synonymous or nonsynonymous) was determined using an R script (See Appendix - Chapter3_files/ MKTest-Analysis/F1).

Thereafter, MK test was performed by assessing the neutrality index (NI) (Figure 3.1), where NI = (Pn /Ps)/ (Fn /Fs); where a Fn/Fs > Pn/Ps (i.e., NI < 1) indicates a positive selection pressure. Statistical testing was done using the Fisher's Exact Test (to test if genes differ from neutral expectations statistically) in R.

Table 3.1: Proposed mechanism of TCBZ resistance in liver flukes and gene families implicated.

Mechanism of Drug Action	Gene Families	Reference
Altered tubulin-binding	β-tubulin α- tubulins	(Ryan et al., 2008)
Altered drug uptake	ATP-binding cassette transporters (ABC) genes	(Kudlacek et al., 2012)
	Adenylate Cyclase (AC) genes	(Radio et al., 2018)
	RAS	(Lee et al., 2013)
	ADP ribosylation factor	(Meaney et al., 2013)
Altered drug metabolism	Cytochrome P450 (CYP450)	
	Glutathione S-transferases (GSTs)	(LaCourse et al., 2012, Stuart et al., 2021)
	Fatty Acid Binding Proteins (FABPs)	(Ramos-Benítez et al., 2017)

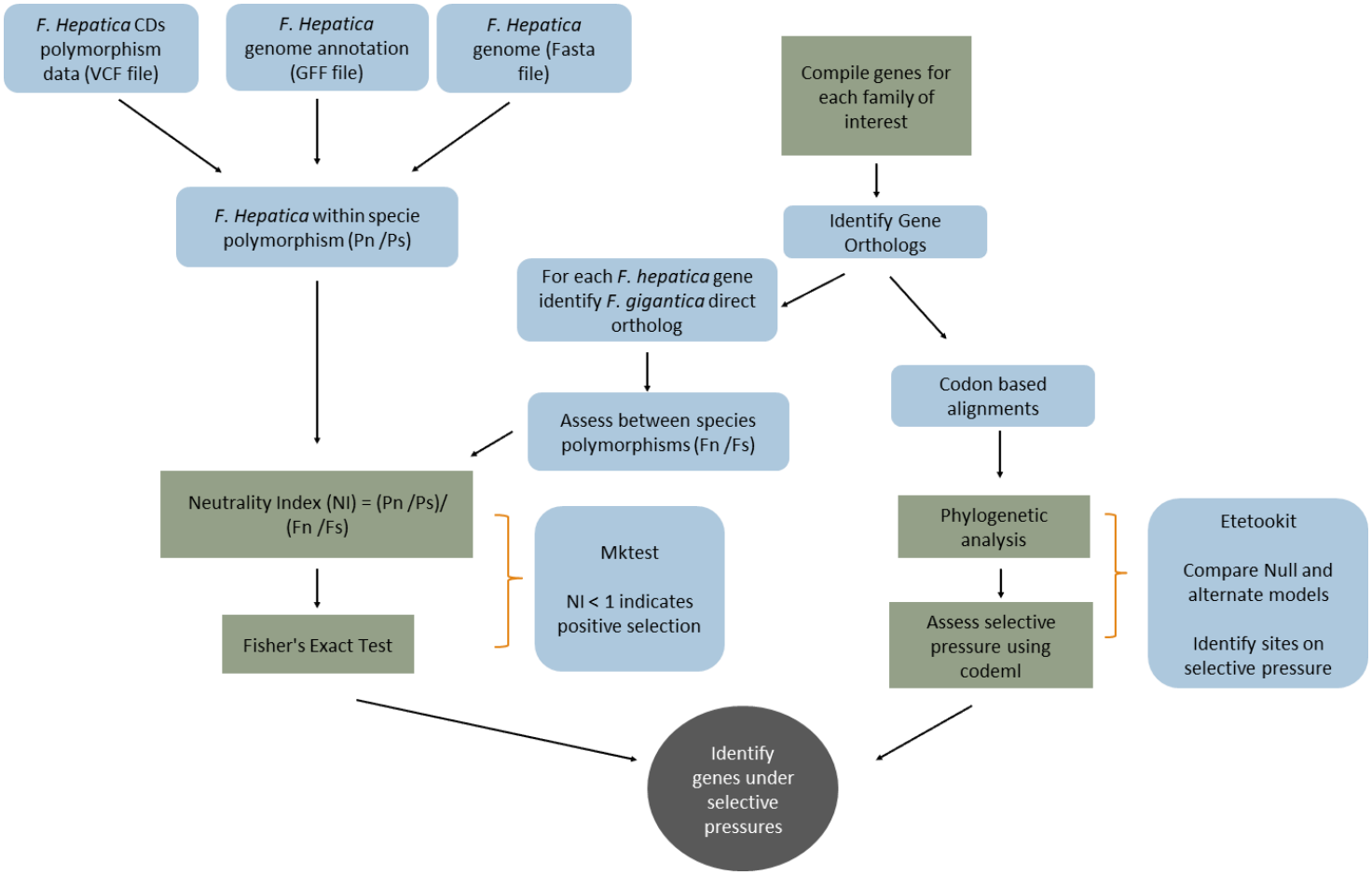

Figure 3.1: Overview of methods used in assessing positive selection pressure in genes of selected families of interest using codeml. PAML analysis using *F. hepatica* and McDonald-Kreitman (MK) test, (where Pn = number of Polymorphic Nonsynonymous sites, Ps = number of Polymorphic Synonymous sites, Fn = number of Fixed Nonsynonymous sites, and Fs = number of Fixed Synonymous sites)

3.8. Results

3.8.1. Summary of positive selection pressure analysis (PAML and MK tests)

3.8.1.1. PAML Test Results

Orthologous grouping (Orthogrouping) of the eight trematode genomes identified a total of 11,976 orthogroups (Table 3.2). A total of 9,114 out of the 9,708 *F. hepatica* genes had orthogroups, while 594 genes were not assigned to any orthogroup (Table 3.3). For an orthogroup of interest, positive selection on sites was tested using PAML (model M7 for relaxation and M8 for positive selection). The first group of genes studied were the tubulins. In the α-tubulins, a total of 21 genes were compiled, while there were 7 orthogroups. In the models tested, *F. hepatica* genes across 2 of these orthogroups had *F. hepatica* genes with sites under positive selection, out of which 1 orthogroup (OG0006908) had *F. hepatica* genes with statistically significant sites under positive selection. In the β-tubulins with 11 genes with 7 identified orthogroups, no orthogroup had *F. hepatica* genes with sites under positive selection. In addition to these, other types of tubulins identified and analysed include one delta-tubulin, one epsilon-tubulin and one gamma-tubulin, each belonging to different orthogroups. Of these orthogroups, the delta tubulins, orthogroup (OG0000431) had statistically significant sites under positive selection in the *F. hepatica* - maker-scaffold10x_703_pilon-snap-gene-0.88 gene (Table 3.4).

The second group of genes studied were those associated with altering drug uptake. These include the ABC genes, RAS genes, Adenylyl Cyclase (AC) genes, and ADP ribosylation factor genes. In the ATP-binding cassette transporters (ABC) genes, 21 genes were assessed. Out of the 19 ABC gene orthogroups, 15 orthogroups had *F. hepatica* genes with sites under positive selection, 5 orthogroups of which (OG0002392, OG0003131, OG0003738, OG0004307, OG0004483) had *F. hepatica genes* with sites under positive selection that were statistically significant (Table 3.4). Similarly, in the RAS genes, a total of 42 genes were assessed. There were 37 orthogroups, 22 of which had *F. hepatica* genes with sites under positive selection. A total of 9 orthogroups (OG0000484, OG0002938, OG0003926, OG0004010, OG0004466, OG0005115, OG0005250, OG0006392, OG0006405) had *F. hepatica* genes with statistically significant sites under positive selection (Table 3.4). Similarly, a total of 14 ADP ribosylation genes were assessed. Out of the 11 ADP gene orthogroups analysed, 4 orthogroups had *F. hepatica* genes predicted to be under positive selection. Out of these 4 ADP ribosylation orthogroups, 2 orthogroups (OG0004152 and OG0006927) had *F. hepatica* genes with statistically significant sites under positive selection. In addition to these, a total 9 adenyl cyclase (AC) genes were assessed (grouped into 7 orthogroups), out of which none of the genes had statistically significant sites under positive selection (Table 3.4).

The last group of genes is associated with altered drug metabolism. Three families were studied: the CYP450, GST, and FABP. A total of 9 Cytochrome P450 genes were compiled and grouped into 7 orthogroups. Four of these orthogroups had *F. hepatica* genes with sites predicted to be under positive selection (none of which was statistically significant).

Similarly, 17 GST genes were assessed. There was a total of 9 GST orthogroups, only 1 of which had *F. hepatica* gene with sites under positive selection (which was not statistically significant). Finally, a total of 6 Fatty Acid Binding (FABP) genes were assessed. These were grouped into 3 orthogroups, and only 2 had *F. hepatica* genes with sites under positive selection, none of which was statistically significant (Table 3.4).

Site models were used to test for evidence of positive selection in the orthogroups of interest for each gene family. By labelling the *F. hepatica* branch as the foreground, branch and site models were used to validate results observed in the site models (Table 3.6). Branch-site models confirmed that *F. hepatica* RAS genes (maker-scaffold10x_317_pilon-snap-gene-0.52 and snap_masked-scaffold10x_418_pilon-processed-gene-0.133), ADP ribosylation factor gene - snap_masked-scaffold10x_492_pilon-processed-gene-0.0, and Adenylyl cyclase gene - maker-scaffold10x_609_pilon-snap-gene-0.9 has statistically significant site under positive selection (Figures 3.3 & 3.4).

3.8.1.2. McDonald-Kreitman (MK) Test Results

All the genes assessed for evidence of positive selection using PAML test were also assessed using the MK test. However, evidence of a recent positive selection pressure was only found in 2 gene families – the RAS and AC genes. Although there were other genes likely to be under a positive selection pressure (i.e., had Neutrality index < 1), most of these were not significant statistically (Table 3.5). The RAS genes identified to be under positive selection pressure (neutrality index < 1 and P-value < 0.05) were maker-scaffold10x_1566_pilon-snap-gene-0.19 and maker-scaffold10x_1568_pilon-snap-gene-0.0, while maker-scaffold10x_239_pilon-augustus-gene-0.114 was the only AC gene (Table 3.5). Interestingly, maker-scaffold10x_1568_pilon-snap-gene-0.0, a RAS gene, was the only gene identified to be on a statistically significant positive selection pressure by the PAML and MKtest approach.

Table 3.2: Summary statistics of the orthologous grouping. Orthologous grouping of the *F. hepatica* genome (PRJEB25283) and seven other trematode species.

Number of species	8
Number of genes	110081
Number of genes in orthogroups	99199
Number of unassigned genes	10882
Percentage of genes in orthogroups	90.1
Percentage of unassigned genes	9.9
Number of orthogroups	11976
Number of species-specific orthogroups	1054
Number of genes in species-specific orthogroups	4123
Percentage of genes in species-specific orthogroups	3.7
Mean orthogroup size	8.3
Median orthogroup size	8
G50 (assigned genes) *	10
G50 (all genes)	9
O50 (assigned genes) •	3263
O50 (all genes)	3827
Number of orthogroups with all species present	5007
Number of single-copy orthogroups	1144

*G50: number of genes in an orthogroup where 50% of genes are in orthogroups of that size or larger; •O50: smallest number of genes in an orthogroup where 50% of genes are in orthogroups of that size or larger. An assigned gene refers to a gene that has been allocated in an orthogroup, while an unassigned gene is a gene that has not been allocated to any orthogroup.

Table 3.3: Summary Statistics per species used for Orthologous grouping.

Species	*C. sinensis*	*E. caproni*	*F. gigantica*	*F. hepatica*	*Fp. buski*	*O. viverrini*	*P. westermani*	*S. mansoni*
Number of genes	13634	18607	12669	9708	11837	16356	12771	14499
Number of genes in orthogroups	12077	15631	11821	9114	11394	13606	11883	13673
Number of unassigned genes	1557	2976	848	594	443	2750	888	826
Percentage of genes in orthogroups	88.6	84	93.3	93.9	96.3	83.2	93	94.3
Percentage of unassigned genes	11.4	16	6.7	6.1	3.7	16.8	7	5.7
Number of orthogroups containing species	8435	8929	8976	7753	8793	8393	8190	7749
Percentage of orthogroups containing species	70.4	74.6	74.9	64.7	73.4	70.1	68.4	64.7
Number of species-specific orthogroups	80	159	63	17	18	167	143	407
Number of genes in species-specific orthogroups	251	1149	205	46	38	490	454	1490
Percentage of genes in species-specific orthogroups	1.8	6.2	1.6	0.5	0.3	3	3.6	10.3

Table 3.4: Results of the assessment of sites exhibiting evidence of positive selection pressure in selected gene families of 8 trematode species. Evolutionary site models comparing null and alternative hypotheses were used to estimate the nonsynonymous/synonymous rate ratio (ω = dN/dS), a measure of positive selection.

Mechanism of drug action	Gene families		No of the genes identified	Total no of orthogroups	No of orthogroups with sites predicted to be under positive selection pressure	Orthogroup	P-value (*P<0.05; **P<0.001)	*F. hepatica* genes in the orthogroup
Altered tubulin binding	Tubulins	Alpha	21	7	3	OG0006908	0.000849**	maker-scaffold10x_2152_pilon-augustus-gene-0.28-mRNA-1
						OG0007608	1	maker-scaffold10x_1444_pilon-snap-gene-0.40-mRNA-1
		Beta	11	7	0			
		Delta	1	1	1	OG0000431	0.024773*	maker-scaffold10x_703_pilon-snap-gene-0.88•
		Epsilon	1	1	1	OG0003519	1	maker-scaffold10x_500_pilon-snap-gene-0.52
		Gamma	1	1	0	OG0006408	1	maker-scaffold10x_1160_pilon-snap-gene-0.20
Altered drug uptake	ABC genes		21	19	15	OG0000323	0.516036	maker-scaffold10x_59_pilon-augustus-gene-0.5-mRNA-1 (ABCD) maker-scaffold10x_363_pilon-snap-gene-0.0-mRNA-1 (ABCD)
						OG0000521	0.296718	maker-scaffold10x_598_pilon-snap-gene-1.129-mRNA-1 (ABCC)
						OG0000672	0.172436	maker-scaffold10x_383_pilon-snap-gene-1.6-mRNA-1 (ABCG)

					OG0001497	0.99675	maker-scaffold10x_83_pilon-snap-gene-0.200 (ABCC)
					OG0001765	0.655818	maker-scaffold10x_83_pilon-snap-gene-0.200-mRNA-1 (ABCC)
					OG0002152	0.111761	maker-scaffold10x_923_pilon-snap-gene-0.11-mRNA-1 (ABCA)
					OG0002392	0.047823*	maker-scaffold10x_52_pilon-snap-gene-0.38-mRNA-1 (ABCA)
					OG0002435	0.10302	snap_masked-scaffold10x_317_pilon-processed-gene-0.41 (ABCA)
					OG0003131	0.000206**	maker-scaffold10x_1995_pilon-augustus-gene-0.59 (ABCF)
					OG0003232	0.282104	maker-scaffold10x_604_pilon-snap-gene-0.72-mRNA-1 (ABCF)
					OG0003738	0.020683*	maker-scaffold10x_380_pilon-snap-gene-0.4-mRNA-1 (ABCB)•
					OG0003968	0.088729	maker-scaffold10x_173_pilon-snap-gene-0.13-mRNA-1 (ABCE)
					OG0004307	0.032107*	maker-scaffold10x_1_pilon-snap-gene-0.157-mRNA-1 (ABCA)
					OG0004483	0.003980**	maker-scaffold10x_90_pilon-snap-gene-0.86-mRNA-1 (ABCA)
					OG0004768	0.145234	maker-scaffold10x_149_pilon-snap-gene-0.91 (ABCC)
	RAS genes	42	37	22	OG0000484	0.000022**	maker-scaffold10x_680_pilon-snap-gene-0.22-mRNA-1•
					OG0000928	1	maker-scaffold10x_157_pilon-snap-gene-0.182-mRNA-1
					OG0001028	0.106037	maker-scaffold10x_951_pilon-augustus-gene-0.24-mRNA-1•
					OG0002938	0.000000**	snap_masked-scaffold10x_2595_pilon-

				processed-gene-0.0-mRNA-1••
			OG0003469	0.22523	maker-scaffold10x_443_pilon-snap-gene-0.32-mRNA-1
			OG0003744	0.051358	maker-scaffold10x_715_pilon-snap-gene-0.26-mRNA-1
			OG0003926	0.000001**	maker-scaffold10x_206_pilon-snap-gene-0.59-mRNA-1•
			OG0004010	0.003053**	maker-scaffold10x_537_pilon-snap-gene-0.69-mRNA-1•
			OG0004062	0.912533	maker-scaffold10x_646_pilon-snap-gene-0.17-mRNA-1
			OG0004182	0.33151	maker-scaffold10x_1272_pilon-snap-gene-0.12-mRNA-1
			OG0004466	0.000000**	maker-scaffold10x_317_pilon-snap-gene-0.52-mRNA-1• maker-scaffold10x_142_pilon-snap-gene-0.82-mRNA-1
			OG0004747	1	maker-scaffold10x_1825_pilon-snap-gene-0.20-mRNA-1
			OG0004920	0.115777	maker-scaffold10x_768_pilon-snap-gene-0.5-mRNA-1
			OG0005115	0.029092*	maker-scaffold10x_1568_pilon-snap-gene-0.0-mRNA-1•‡
			OG0005240	0.229099	maker-scaffold10x_1165_pilon-augustus-gene-0.1-mRNA-1
			OG0005250	0.035085*	maker-scaffold10x_1332_pilon-pred_gff_StringTie-gene-0.67-mRNA-1
			OG0005747	0.20448	maker-scaffold10x_2148_pilon-snap-gene-0.0-mRNA-1•
			OG0006392	0.041945*	maker-scaffold10x_109_pilon-snap-gene-0.3-mRNA-1

					OG0006405	0.000000**	snap_masked-scaffold10x_418_pilon-processed-gene-0.133-mRNA-1•
					OG0006406	1	maker-scaffold10x_64_pilon-augustus-gene-0.16-mRNA-1
					OG0007158	0.126699	maker-scaffold10x_1566_pilon-snap-gene-0.19-mRNA-1•
					OG0008497	0.171531	maker-scaffold10x_532_pilon-snap-gene-0.4-mRNA-1
	ADP ribosylation factor	14	11	4	OG0004152	0.004895**	maker-scaffold10x_217_pilon-snap-gene-1.113-mRNA-1
					OG0004291	0.520714	maker-scaffold10x_419_pilon-snap-gene-0.64-mRNA-1
					OG0006174	0.615967	maker-scaffold10x_1010_pilon-pred_gff_StringTie-gene-0.79-mRNA-1
					OG0006927	0.000000**	snap_masked-scaffold10x_492_pilon-processed-gene-0.0-mRNA-1•
	Adenylyl cyclase	9	7	2	OG0000839	0.000000**	maker-scaffold10x_609_pilon-snap-gene-0.9-mRNA-1• maker-scaffold10x_102_pilon-augustus-gene-0.94-mRNA-1
					OG0002578	0.265107	maker-scaffold10x_245_pilon-augustus-gene-0.96-mRNA-1
Altered drug metabolism	Cytochrome P450 (CYP450)	9	7	4	OG0002059	0.599423	maker-scaffold10x_2113_pilon-augustus-gene-0.4-mRNA-1
					OG0002918	0.517665	maker-scaffold10x_1257_pilon-snap-gene-0.67-mRNA-1
					OG0003646	0.883156	maker-scaffold10x_1478_pilon-snap-gene-0.30-mRNA-1
					OG0006916	1	maker-scaffold10x_1_pilon-augustus-gene-0.132-mRNA-1

				OG0000157	0.117074	maker-scaffold10x_1043_pilon-snap-gene-0.18-mRNA-1
	Glutathione S-transferase (GST)	17	9	1	OG0000157	0.117074	maker-scaffold10x_1043_pilon-snap-gene-0.18-mRNA-1
	FABP	6	3	2	OG0001256	0.122201	maker-scaffold10x_2403_pilon-snap-gene-0.20-mRNA-1 maker-scaffold10x_2403_pilon-snap-gene-0.19-mRNA-1
					OG0001842	0.839282	maker-scaffold10x_331_pilon-snap-gene-0.56-mRNA-1 maker-scaffold10x_2403_pilon-snap-gene-0.11-mRNA-1 maker-scaffold10x_331_pilon-snap-gene-0.57-mRNA-1

● Refer to reference gene in the orthogroup with a statistically significant site(s). Sites under positive selection are not shown (Additional information on each gene assessed is available in supplementary information)

●● *Fasciola gigantica* gene TPP49991.1 as reference

‡ Gene statistically significantly identified to be under positive selection pressure by using the PAML-codeml (site models) and MK test techniques

Table 3.5: Results of assessing a recent positive selection pressure between *F. hepatica* and *F. gigantica* in gene families of interest.

Mechanism of Drug Action	Gene families	No of genes with NI < 1	No of genes with P value < 0.05	No of genes with NI < 1 & P value < 0.05	*F. hepatica* genes predicted to be under positive selection (Genes with NI < 1 & P value < 0.05)
Altered tubulin-binding	Alpha tubulins	6	1		-
	Beta tubulins	3	0		-
Altered drug uptake	ABC genes	11	0		-
	RAS genes	10	5	2	maker-scaffold10x_1566_pilon-snap-gene-0.19-mRNA-1 (NI - 0.1028, P - 0.038, NNS - 552) maker-scaffold10x_1568_pilon-snap-gene-0.0-mRNA-1 (NI – 0.3201, P - 0.0413, NNS – 1167) ‡
	ADP ribosylation factor	0	0	0	
	Adenylyl cyclase	6	1	1	maker-scaffold10x_239_pilon-augustus-gene-0.114-mRNA-1 (NI - 0.1385, P - 0.0006, NNS - 1140)
Altered drug metabolism	Cytochrome P450 (CYP450)	3	0	0	-
	Glutathione S-transferase (GST)	3	0	0	-
	FABP	3	0	0	-

Additional information on each gene assessed is available in the supplementary information

NNS - Number of Nucleotide sites

‡ Gene statistically significantly identified to be under positive selection pressure by using the PAML-codeml (site models) and MK test technique

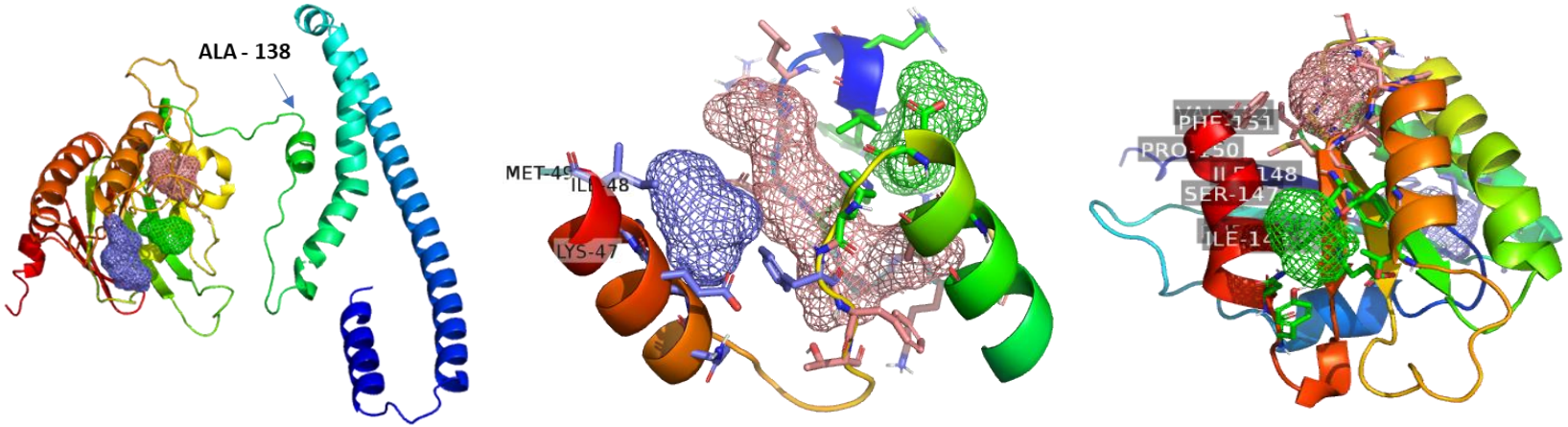

Figure 3.2: Predicted protein models of the *F. hepatica* genes. Protein models of *F. hepatica* genes (ABCB gene - scaffold10x_380_pilon-snap-gene-0.4 (left), RAS genes - maker-scaffold10x_680_pilon-snap-gene-0.22 (middle) and maker-scaffold10x_317_pilon-snap-gene-0.52 (right)) showing the sites under positive selection in the predicted protein binding sites (shown as the mesh structures in each model). In the ABCB gene model (left) positively selected site is distant from the protein binding sites, suggesting protein activity will not be influenced, while positively selected sites are in proximity with protein binding sites in the RAS genes (middle and right).

Table 3.6: *F. hepatica* sites predicted to be under positive selection identified by labelling *F. hepatica* as the foreground branch in the branch-site model

Mechanism of Drug Action	Gene families	Orthogroup	*F. hepatica* predicted to have sites under positive selection using site models (gene labelled as foreground in the branch and branch-site models)	Branch model P-value (*P<0.05; **P<0.001)	Branch-site model P-value (*P<0.05; **P<0.001)	*F. hepatica* sites under positive selection identified by branch-site model. Naive Empirical Bayes (NEB) analysis (*P<0.05; **P<0.001)	*F. hepatica* sites under positive selection identified by branch-site model. Bayes Empirical Bayes (BEB) analysis (*P<0.05; **P<0.001)
Altered tubulin binding	Delta tubulin	OG0000431	maker-scaffold10x_703_pilon-snap-gene-0.88	0.032325*	1.0000		
Altered drug uptake	ABC genes	OG0003738	maker-scaffold10x_380_pilon-snap-gene-0.4-mRNA-1 (ABCB)	0.553318	1.0000	309 Q 0.732	
	RAS genes	OG0000484	maker-scaffold10x_680_pilon-snap-gene-0.22-mRNA-1	1.0000	1.0000		
		OG0001028	maker-scaffold10x_951_pilon-augustus-gene-0.24-mRNA-1	0.000000**	0.071461	100 K 0.535 158 L 0.771 310 N 0.550 363 Q 0.720 395 S 0.767 398 S 0.521	158 L 0.612 363 Q 0.619 395 S 0.675
		OG0002938	snap_masked-scaffold10x_2595_pilon-processed-gene-0.0-mRNA-1●●	0.016372*	0.028735*	60 V 0.789 61 S 0.997** 62 K 0.690 63 V 0.552 65 T 0.737 68 A 0.997** 69 F 0.546 70 I 0.550	60 V 0.719 61 S 0.994** 62 K 0.609 63 V 0.500 65 T 0.660 68 A 0.994** 71 R 0.988* 72 P 0.737

				71 R 0.995**	73 V 0.996**
				72 P 0.806	74 Q 0.997**
				73 V 0.998**	75 Y 0.993**
				74 Q 0.999**	78 V 0.991**
				75 Y 0.997**	79 P 0.999**
				78 V 0.996**	80 L 0.771
				79 P 1.000**	81 L 0.997**
				80 L 0.831	82 T 0.998**
				81 L 0.998**	83 H 0.980*
				82 T 0.999**	84 M 0.990**
				83 H 0.988*	85 S 0.746
				84 M 0.996**	86 L 0.996**
				85 S 0.813	87 L 0.810
				86 L 0.998**	89 K 0.791
				87 L 0.863	91 S 0.669
				89 K 0.808	92 S 0.507
				91 S 0.685	
				92 S 0.525	
OG0003926	maker-scaffold10x_206_pilon-snap-gene-0.59-mRNA-1	0.579310	1.0000		
OG0004010	maker-scaffold10x_537_pilon-snap-gene-0.69-mRNA-1	0.574487	1.000000	85 I 0.687	
OG0004466	maker-scaffold10x_317_pilon-snap-gene-0.52-mRNA-1	0.830963	0.000323**	144 S 0.509	147 S 0.967*
				147 S 0.999**	156 S 0.913
				153 P 0.822	

					156 S 0.995**	
	OG0005115	maker-scaffold10x_1568_pilon-snap-gene-0.0-mRNA-1●‡	0.595059	1.0000		
	OG0005250	maker-scaffold10x_1332_pilon-pred_gff_StringTie-gene-0.67-mRNA-1	0.408698	1.0000		
	OG0005747	maker-scaffold10x_2148_pilon-snap-gene-0.0-mRNA-1	1.000000	0.737046		
	OG0006405	snap_masked-scaffold10x_418_pilon-processed-gene-0.133-mRNA-1	0.778270	0.000130**	111 N 0.622	111 N 0.853
					125 A 0.997**	125 A 0.960*
					126 G 0.997**	126 G 0.954*
					127 H 0.834	127 H 0.712
					128 L 0.997**	128 L 0.896
	OG0007158	maker-scaffold10x_1566_pilon-snap-gene-0.19-mRNA-1	0.757015	0.981188	29 N 0.730	
ADP ribosylation factor	OG0006927	snap_masked-scaffold10x_492_pilon-processed-gene-0.0-mRNA-1	0.644383	0.000000**	181 M 0.986*	181 M 0.982*
					182 I 0.995**	182 I 0.986*
					184 M 1.000**	183 D 0.553
					185 T 0.994**	184 M 0.979*
					187 N 0.932	185 T 0.992**
					188 E 0.842	186 E 0.930
					189 Q 1.000**	187 N 0.816
					190 A 0.996**	188 E 0.871
					191 T 0.735	189 Q 0.983*
						190 A 0.948
						191 T 0.782

	OG0004152	maker-scaffold10x_217_pilon-snap-gene-1.113-mRNA-1	0.772638	1.000000		
Adenylyl cyclase	OG0000839	maker-scaffold10x_609_pilon-snap-gene-0.9-mRNA-1	0.332026	0.007710**	31 R 0.741 33 A 0.973* 35 T 0.830	32 F 0.521 35 T 0.574

●● *Fasciola* gigantica gene TPP49991.1 as reference

‡ Gene statistically significantly identified to be under positive selection pressure by using the PAML-codeml (site models) and MK test techniques

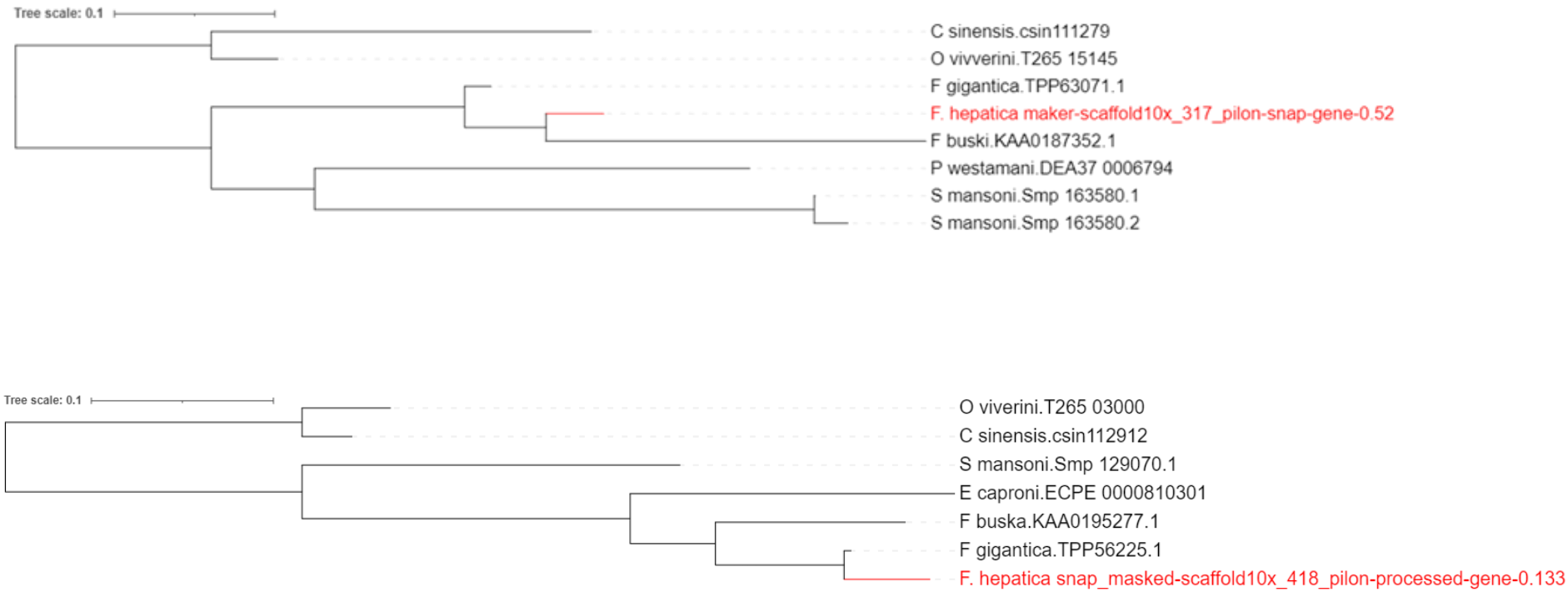

Figure 3.3: Phylogenetic tree of *F. hepatica* genes with statistically significant sites under positive selection using branch-site model. RAS genes in orthogroups OG0004466 – top (maker-scaffold10x_317_pilon-snap-gene-0.52) and OG0006405 – bottom (snap_masked-scaffold10x_418_pilon-processed-gene-0.133) had sites under positive selection identified using branch-site models in PAML . In each orthogroup, *F. hepatica* gene branch is labelled as the foreground (in red). Phylogenetic trees were generated using RAxML in ETEtookkit, and displayed using iTOL (Letunic and Bork, 2021, Stamatakis, 2014, Huerta-Cepas et al., 2016).

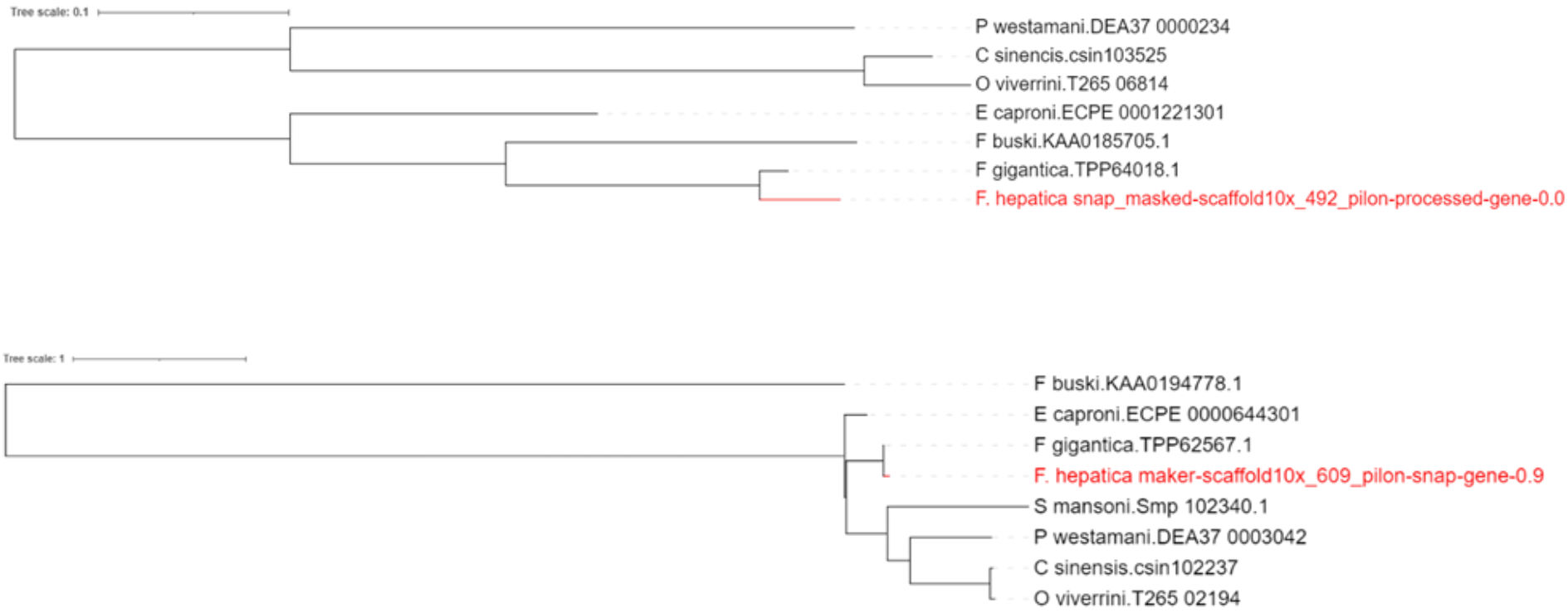

Figure 3.4: Phylogenetic tree of *F. hepatica* genes with statistically significant sites under positive selection using branch-site model. ADP ribosylation factor genes in orthogroup OG0006927 – top (snap_masked-scaffold10x_492_pilon-processed-gene) and Adenylyl cyclase genes in orthogroup OG0000839 – bottom (maker-scaffold10x_609_pilon-snap-gene-0.9) had sites under positive selection identified using branch-site models in PAML . In each orthogroup, *F. hepatica* gene branch is labelled as the foreground (in red). Phylogenetic trees were generated using RAxML in ETEtookkit, and displayed using iTOL (Letunic and Bork, 2021, Stamatakis, 2014, Huerta-Cepas et al., 2016).

3.9. Discussion

3.9.1. Tubulin genes are mostly not under positive selection

A total of 35 tubulin genes were identified in the *F. hepatica* genome. This is more than the five α-tubulin and six β-tubulin genes previously described (Ryan et al., 2008). Using these previously described tubulin genes as references, the tubulins in the *F. hepatica* genome were compiled (*Fasciola*_10x_pilon, GCA_900302435.1) using the presence of gene motifs. Tubulins are highly conserved genes (Sackett et al., 2010). For example, orthogroup - OG0000010 has ten α-tubulin genes while OG0000033 has nine β-tubulin genes. There was no evidence of sites under positive selection in most *F. hepatica* tubulins. However, in gene maker-scaffold10x_703_pilon-snap-gene-0.88, delta tubulin, two statistically significant sites under positive selection pressure were identified (Table 3.4). The gene was not identified by MKtest to be under positive selection. This could either be because it was not a recent adaptive change, or the site could not be identified due to the conservative nature of the MKtest approach. The result could also be a false positive. The latter is reasonably plausible, considering the PAML method can be limited by introducing an alignment error. Using SWISS-MODEL to build the protein model (Waterhouse et al., 2018), FTSite and Phymol for active site prediction (Ngan et al., 2012, DeLano, 2002), the sites under positive selection pressure were not in the binding sites of the protein. The absence of the sites under positive selection pressure in the protein's binding site could suggest that the binding of the protein to its substrate would not be altered. Delta tubulins have been predicted to regulate centriole function in various organisms (Stathatos et al., 2021). However, its specific role in *F. hepatica* has never been described.

3.9.2. Genes with sites under positive selection pressure are associated with altering drug uptake

Four gene families have been associated with altering the TCBZ uptake as a proposed drug resistance mechanism (Table 3.1). In these four gene families, results identified signals suggesting the presence of adaptive changes in some genes in each family (Tables 3.4b, 3.4c, & 3.4d). For example, in gene maker-scaffold10x_380_pilon-snap-gene-0.4, an ATP-binding cassette transporter (ABCB) gene, three sites of statistical significance were under positive selection pressure. Interestingly, previous reports claim that inhibitors of the ABCB1 gene can cause resistance to TCBZ in previously susceptible fluke (El-Kasaby et al., 2011); however, the mechanism is not clearly understood. It is thought that these glycoproteins play essential roles in immature flukes, such as removal of metabolites, considering expression levels of the gene is higher in immature flukes and decreases in matured ones (Reed et al., 1998). But then, precisely how the parasite alters the update of TCBZ is unclear. A likely explanation is that the parasite protein binds to the drugs but can change its binding site to prevent binding to the drug; hence resistance develops. It is also plausible that expression levels of the gene are reduced as fluke maturity is achieved, however, this does not explain resistance to TCBZ. It is known that expression levels of membrane transporters are elevated in TCBZ-resistant flukes. However, it is not clear if the reduced potency of TCBZ in resistant flukes is directly due drug binding failures in membrane transporter protein

active sites (Fairweather et al., 2020). A further study of the ABCB gene - scaffold10x_380_pilon-snap-gene-0.4 protein model suggests that binding of the protein would not be altered due to the positive selection pressure on the site identified. The site under positive selection was distant from the predicted binding sites of the protein model (Figure 3.2). While comparing the null and the alternate site models suggest that adaptive changes exist in the ABC genes, there was no substantial reason to conclude that sites identified to be under positive selection play critical roles in the parasite's biology.

In the RAS genes, however, more genes with statistically significant sites under positive selection were identified. Out of the 42 genes compiled, ten genes were found with varying numbers of statistically significant sites under positive selection pressure. The RAS genes had the highest number of sites under positive selection pressure compared to the other families studied. RAS and Adenyl cyclases play essential roles in converting ATP to the second messenger cyclic AMP (cAMP). This nutrient-sensing mechanism plays a vital role in protein kinases, glucose uptake and has been shown to be inhibited by TCBZ (Lee et al., 2013). Although the interaction of *F. hepatica* RAS and AC genes with TCBZ is unclear, results here indicate the presence of adaptive changes in these genes in the parasite. In two RAS genes with sites under positive selection pressure, maker-scaffold10x_680_pilon-snap-gene-0.22 (UniProtKB - A0A4E0RHR7) and maker-scaffold10x_317_pilon-snap-gene-0.52, both of which have 3 and 12 statistically significant sites under positive selection, protein modelling of potential active sites suggests these sites under positive selection could be of biological interest. With the closeness of these positively selected sites to the protein binding site, it is likely protein binding to its substrate is impacted (Figure 3.2).

Interestingly, despite the conservativeness of the MKTest used to identify a recent adaptive mutation between *F. hepatica* and *F. gigantica*, RAS gene - maker-scaffold10x_1568_pilon-snap-gene-0.0 (described as a Rad and Gem GTP binding protein 1 on UNIPROT, A0A4E0RHY2) was identified to be under positive selection by both MKtest approach and the PAML site model approach (Tables 3.4c & 3.5). However, the protein model and active site predictions suggest that the site under positive selection is not in the protein's active site. This could indicate the site under positive selection does not influence protein binding to the substrate, or there is a need for better structural modelling. One of the challenges noticed when predicting the effects of a site under positive selection is that in most of the RAS genes modelled, the sites under positive selection were in the unstructured part of the protein model. Although, arguably, lack of reliable homologs reduces the accuracy of homology-based protein structure modellers such as the SWISS-MODEL (Cringoli et al., 2017, Rodrigues et al., 2013), other factors such as poor gene models and alignment errors can influence protein modelling. Thus, the inability to identify sites under positive selection based on predicted protein structure could be attributed the poor quality of these protein models.

In contrast to the results observed in the RAS genes, none of the *F. hepatica* AC genes had statistically significant sites under positive selection. However, findings suggest some form

of adaptive changes in orthogroup - OG0000839, compared to the null and alternate model hypothesis. The orthogroup OG0000839 is interesting because it contains the predominantly expressed adenylyl cyclase genes in *F. hepatica*; maker-scaffold10x_102_pilon-augustus-gene-0.94 and maker-scaffold10x_609_pilon-snap-gene-0.9 (discussed in chapter 4). Previous findings indicate that AC gene activity is decreased in flukes resistant to TCBZ (Radio et al., 2018), based on the previous gene identified. Here, while we complied and provided an updated list of AC genes in the *F. hepatica* genome, there was no substantial evidence to conclude the presence of signals indicative of positive selection pressure in most of these gene families. However, in the ADP ribosylation factor genes, sites under positive selection were identified in only one gene – snap_masked-scaffold10x_492_pilon-processed-gene-0.0.

Despite the modelling of the protein structure and active sites, it is unclear what the biological importance of these sites might be. Despite these findings, additional studies on genes associated with altering the uptake of TCBZ could provide a better understanding of TCBZ resistance in liver flukes. The evidence of positive selection identified in them potentially questions their suitability as drug targets.

3.9.3. No substantial evidence of positive selection in genes associated with altering TCBZ metabolism

In the three gene families implicated in the metabolism of TCBZ (Table 3.4); the cytochrome P450, glutathione S-transferase (Table 3.4), and the fatty acid-binding genes, there was no statistically significant gene under positive selection. Although some genes in these families had sites under positive selection, the sites identified were not statistically significant. This is somewhat unexpected considering previous studies on GST and CYP450 genes in *F. hepatica* suggesting that alteration in TCBZ uptake and metabolism of the drug is more vital to the parasite compared to altering tubulin activity (Devine et al., 2009, Devine et al., 2008, Meaney et al., 2013). One would have expected a higher level of adaptive changes in the parasite to deal with an active form of the drug effectively. Thus, it would be helpful to determine which mechanism is more vital or if a combination of mechanisms is required for TCBZ mode of action and resistance in liver flukes.

Furthermore, the small soluble FABP gene families, which has been of interest as vaccine candidates, only seemed to be undergoing some adaptive changes of no statistical significance. Using previously described sequences (Morphew et al., 2016), we identified six FABP genes in the *F. hepatica* genome (*Fasciola*_10x_pilon, GCA_900302435.1), which is relatively low. Detailed identification and description of this family could facilitate a better understanding of their role in liver fluke biology, particularly investigating what evolutionary forces are acting on the family.

3.9.4. The MKtest: a conservative approach

The MKtest compares the ratio of polymorphism within a species (i.e. Pn/Ps) and the ratio of divergence between two closely related species, Dn/Ds (or Fn/Fs), as an indicator of adaptive evolutionary change by selection (McDonald and Kreitman, 1991). Positive

selection increases divergence relative to polymorphism. However, negatively selected polymorphisms can reduce the ability to detect adaptive evolution signals (Andolfatto, 2005, Messer and Petrov, 2013). A Pn/Ps value < 1 or equals zero limits the ability to detect selection pressure since ratio of polymorphism within a species needs to always be greater than the ratio of divergence between species studied. This was a challenge in most of the genes assessed for positive selection pressure between *F. hepatica* and *F. gigantica*. Out of all the nine families of genes investigated, only three genes (Table 3.5) had a neutrality index < 1 (indicating a positive selection signal) of statistical significance ($P < 0.05$). Interestingly, three genes, two RAS genes and one AC gene belong to the gene families believed to be associated with altering TCBZ update in *F. hepatica* were identified to be under positive selection using the MKtest method. These MKtest results corroborate results of identifying sites under positive selection pressure using PAML in these gene families; although fewer genes were identified by the MKtest approach, further evolutionary study across the genome might be more beneficial.

3.9.5. Limitations of Assessing Positive Selection Pressure using PAML

The codeml software in PAML (Yang, 1997) is a widely used tool for studying evolutionary questions using molecular sequence data. Despite its effectiveness, the approach is limited primarily by its computational intensiveness. With the increasing rate of genomic studies, the PAML package struggles to cope because it is designed to run on a gene by gene basis (Macías et al., 2020). Thus, assessing on a genomic scale can be less practicable. Even for evaluating a single gene, the package requires a strict data workflow. A typical single gene analysis would require compilation of orthologous sequences, sequence alignment, phylogenetic tree construction, filtering data to be suitable for codeml input format, configuration file set up, detection of positive selection and interpretation of data. Each stage of the analysis can introduce an error, which could lead to failure of the software or false-positive results. Various pipelines have been developed to reduce these computational errors, each designed to address one or more stages of the codeml workflow (Hongo et al., 2015, Cole and Brewer, 2018, Delport et al., 2010, Macías et al., 2020, Huerta-Cepas et al., 2016). Despite the development of these various pipelines, their use requires a good understanding how they work; although this can be challenging especially when software dependencies change. Although the usage of these pipelines reduces computational challenges, in this study, it was observed that a careful check of results is needed to ensure errors and false-positive results are minimised.

3.10. Conclusion

This study assessed positive selection pressure using site models in PAML (Yang et al., 2000) in nine *F. hepatica* gene families. These have been associated with TCBZ metabolism and resistance in the liver fluke (Fairweather et al., 2020). The sequence of each gene of interest and orthologous sequences from eight trematode species was used to identify sites under positive selection. MK tests were also carried out in these nine gene families in *F. hepatica*, using direct orthologs from a sister species – *F. gigantica*, both of which cause Fasciolosis.

Results presented in this study identified sites under positive selection, predominantly in genes associated with altering TCBZ uptake in *F. hepatica*. Of note is the RAS genes, which had the most genes with sites under positive selection. The identification of sites under positive selection in genes associated with altering the uptake of the TCBZ supports previous findings that suggest this mechanism and the drug's metabolism are of more importance in the parasite than the widely investigated tubulin role (Brennan et al., 2007, Kelley et al., 2016).

Despite using a candidate gene approach, this study provided insights into the presence of adaptive changes in these gene families (Table 3.1), considering that a positive selection pressure could influence their binding to drug or vaccine targets, potentially making them inactive. Despite the interesting observations in this study, a genomic approach would facilitate a better understanding of adaptive changes happening across the parasite's genome. However, to achieve this, the availability of high-quality *F. hepatica* gene models across the genome and in trematode orthologs and significant automation of workflow to reduce computational challenges would be vital.

CHAPTER 4

Gene Expression Studies in Selected *F. hepatica* Gene Families Implicated in Triclabendazole Resistance

4.1. Background

The flatworm - *Fasciola hepatica* is the most common liver fluke. The trematode is the predominant cause of fasciolosis, a zoonotic disease predominantly affecting cattle and sheep. The disease has been described as a neglected tropical disease (WHO, 2020). *F. hepatica* is a complex and highly adaptive parasite. Despite being a self-fertilising hermaphrodite, genetic variability in parasites and gene expansions have facilitated the evolution of mechanisms needed to invade the vertebrate hosts, adapt to environmental factors, and develop drug resistance (Cwiklinski et al., 2015). The newly excysted juveniles (NEJs), upon early infection of a host, with the aid of their outer glycocalyx tegument are able to modulate the host immune systems by secreting various proteins (González-Miguel et al., 2021). Understanding NEJs biology could unravel relationships between liver flukes and their hosts. To understand liver fluke biology, various gene families (as discussed more fully in chapter 1) have been studied to assess their role in multiple stages of the parasite. Excretory-secretory proteins, ESPs (predominantly cathepsins) play important roles in the ability of *F. hepatica* to invade hosts. Prominent divergence and expansion of cathepsins highlight their key role in parasite adaptation to their host (McNulty et al., 2017b). Other expanded gene families associated with excretory and secretory functions in the parasite include fatty acid-binding proteins, protein disulfide-isomerases, and molecular chaperones (Choi et al., 2020). Similarly, in *F. gigantica*, a sister-specie, recent genomic studies have identified ESPs, highlighting their importance in host-parasite interactions. Studies suggest more gene expansions and contractions in *F. gigantica* than in *F. hepatica* (Luo et al., 2021). These gene duplications and contractions play important adaptive roles in amplifying gene function or gaining new functions (Näsvall et al., 2012). Understanding parasite-host interactions by understanding adaptive forces shaping gene function are vital to explaining parasitism.

4.2. Gene Expression Profiling of *F. hepatica* Life Cycle Stages

Investigating the expression profile of crucial liver fluke gene families has provided a better understanding of host-parasite interactions. One study (McNulty et al., 2017) found overexpression of genes such as cysteine proteases in the infective stage of *F. hepatica* compared in adults and their eggs. The study noticed that genes associated with signal transduction and the development of the parasite were generally more expressed in the infective stage. In contrast, those related to various metabolic processes and microtubule-related activities were more expressed in adults (McNulty et al., 2017b). The elevated expression of genes associated with parasite development in metacecariae could be related to survival of developmental stages of the parasite in the vegetation. Whereas the elevated expression of genes associated with metabolic and microtubule-related activities could be associated with migratory movement of juveniles in the host and establishment of disease by adult flukes. Like these findings, a *F. hepatica* study noticed increased gene expression in the infective stage and the NEJs, noticing a shift in expression patterns in adult flukes. However, downregulation of various biological processes in maturing and adult fluke compared to in metacercariae was reported (Cwiklinski et al., 2015). These studies explored

the expression of critical biological processes and associated genes such as tubulins, cathepsins, and ABC genes and especially how these differences across the life stages drive parasite adaptation of developmental stages to climatic changes, vegetation, intermediate and definitive host selection, and their ability to invade and adapt to host immunity. Thus, exploring the differences in gene expression levels across liver fluke life cycle stages aid understanding parasite survival within and outside the host.

Exploring differential expression of gene families associated with critical biological processes can help understand their activity in the parasite and their response to parasite control strategies such as to vaccines and drugs. For example, triclabendazole (TCBZ) and albendazole are benzimidazoles (Fairweather and Boray, 1999). Benzimidazoles are well known for disrupting tubulin-based operations in the parasite. In terms of effects on the parasite, both drugs disrupt fluke secretory processes and damage reproductive systems (Kouadio et al., 2021). However, albendazole is only effective against adult flukes older than 12 weeks, while TCBZ is effective against all parasite stages (Babják et al., 2021, McKellar and Scott, 1990). Exploring the expression profile of the tubulin family could explain why *F. hepatica* respond to both drugs in different stages of the parasite. The differences in the expression pattern of tubulins across the various stages of the parasite (Cwiklinski et al., 2015) could explain why both drugs produce varying effects in the parasite and suggest the role of another (either a combination or an entirely different) mechanism of drug action.

4.3. TCBZ Mechanism of Action and Resistance

Triclabendazole (TCBZ), an anthelmintic, has been a drug of choice used for treating infections caused by *F. hepatica* and *F. gigantica* for more than 30 years due to its effectiveness against adults and immature (also known as juvenile) flukes (Laird and Boray, 1992). However, overreliance on TCBZ has led to the development of an alarming global spread of resistance to the drug (Brennan et al., 2007). TCBZ resistance was reported in 30 properties across different countries in the world (Kelley et al., 2016). This number of reported farms is low possibly due to poor reporting. TCBZ's effect on liver flukes has been investigated, identifying its ability to alter parasite tegument (Halferty et al., 2008, Toner et al., 2010a, Toner et al., 2010b), but the TCBZ resistance mechanism is generally unclear. To understand TCBZ mode of action, tubulin gene activity has been investigated, considering it's a benzimidazole derivative and binds to β – tubulins (Stitt et al., 1992, Robinson et al., 2004, Robinson et al., 2002). Interestingly, multiple tubulin genes have been identified and explored in *F. hepatica* (Ryan et al., 2008). However, it's unclear how tubulins interact with TCBZ and why other benzimidazoles have a reduced effect on juvenile *F. hepatica*. To understand these questions, other mechanisms of TCBZ action has been suggested. These include alteration in TCBZ uptake by the parasite and metabolism of the drug (Fairweather et al., 2020).

It has been proposed that altering TCBZ uptake or TCBZ metabolism (or a combination of both) play an essential role in the TCBZ mechanism of action in liver flukes and the development of resistance to the drug; however, this is inconclusive at the moment (Kelley

et al., 2016). Currently various *F. hepatica* gene families have been implicated and are being investigated (see Table 4.1). The predominantly implicated genes include the adenylate cyclase (AC), Ras, ADP Ribosylation factors, cytochrome P450 enzyme, ATP-binding cassette transporters (ABCs), glutathione S-transferase (GST), and fatty acid-binding proteins (FABP). However, it is unclear how TCBZ resistance develops in liver fluke populations. Without a solid understanding of the TCBZ mode of action, understanding how liver flukes develop resistance to the drug will always be challenging.

4.4. Chapter Rationale and Objectives

Previous studies have assessed the differential expression profile of certain gene families key to various biological processes in *F. hepatica* (McNulty et al., 2017b, Cwiklinski et al., 2015), particularly ESPs, cysteine proteases such as cathepsins, and tubulins. However, given the fact that TCBZ is the only drug that targets all life stages of the parasite in the host, particularly the highly pathogenic juveniles; the increasing spread of TCBZ resistance possess a serious problem. In this project, compiling candidate gene families involved in TCBZ resistance based on the proposed mechanism of TCBZ action (Fairweather et al., 2020), and exploration of their stage specific expression was done. Also, expression patterns of these gene families in adult parasites exposed to TCBZ in vivo was explored. Exploring differential expression patterns of these implicated genes could improve understanding the importance of these genes to different parasite stages and facilitate understanding of drug mechanisms of action.

The complexity of liver fluke biology and the multi-genic nature of anthelmintic resistance (Molina-Hernández et al., 2015) makes understanding TCBZ activity challenging. Despite findings suggesting TCBZ mode of action involves multiple pathways, the targets of the drug are unknown. TCBZ has complicated effects on liver flukes, affecting various systems. The impact of TCBZ on tubulin activity is well described (Robinson et al., 2004, Robinson et al., 2002). However, albendazole (a benzimidazole like TCBZ) is effective against flukes resistant to TCBZ, while flukes resistant to albendazole has been shown to be susceptible to TCBZ (Sanabria et al., 2013), thus suggesting the presence of another mode of drug action. Studies have also shown that parasite detoxifying systems are affected by TCBZ, highlighting the role of metabolism pathways. Glutathione S Transferase enzyme (a key component of phase II detoxification) activity has been found to be elevated in TCBZ resistant liver flukes (Scarcella et al., 2012), suggesting metabolism of the drug could be key to explaining resistance.

Similarly, genes associated with drug uptake in the parasite have been explored, with results indicating TCBZ limits RAS and AC gene activity (Lee et al., 2013). Also, AC gene function was limited in flukes resistant to TCBZ (Radio et al., 2018). Despite these findings supporting the idea of multiple TCBZ mechanisms of action and resistance pathways, the relationship between them is unclear. While some of these results were observed, for example, in yeast (Lee et al., 2013), in TCBZ susceptible and resistant fluke isolates (Radio et al., 2018,

Scarcella et al., 2012), there is need to assess the impact of TCBZ treatment on resistant and susceptible *F. hepatica* isolates in livestock.

Here, to facilitate understanding parasite response to TCBZ in susceptible and resistant isolates, adult flukes from experimentally infected sheep were used for gene expression studies. The *F. hepatica* TCBZ resistant and susceptible isolates used for the experiment in this study are previously described (Hodgkinson et al., 2018). Focusing on gene families previously associated with TCBZ resistance (Table 4.1), the response to TCBZ in these resistant flukes was assessed.

4.5. Materials and Methods

4.5.1. Gene Families Assessed

This study used *Fasciola hepatica* genes implicated in TCBZ resistance. Using a combination of literature search, WormBase Parasite and NCBI database, *F. hepatica* genes associated with TCBZ resistance were compiled (gene names are indexed to the *F. hepatica Fasciola*_10x_pilon GCA_900302435.1 assembly, 2020-05-WormBase Annotation version). Exonerate alignment (version 2.2.0) and the presence of motifs particular to each family (Slater and Birney, 2005, Finn et al., 2016, Mistry et al., 2020) was used to confirm gene family relationship (the methods for compiling genes was previously described in Chapter 3 Materials and Methods).

4.5.2. Stage Specific Expression

The previously described RNAseq data (Cwiklinski et al., 2015) from multiple life stages of *F. hepatica* was used to assess expression patterns of the selected gene families. Parasite life stages included *F. hepatica* eggs, snails infected with *F. hepatica* parasites, metacercariae (3 replicates), newly excysted juveniles (NEJs) 1 hour (2 replicates), 3 hours NEJs (2 replicates), 21-day old juveniles (1 replicate) and adult *F. hepatica* (1 replicate) flukes from the bile duct of cattle naturally infected in Uruguay (Cwiklinski et al., 2015). The paired-end reads for each stage (life stages with replicates were mapped as a single parasite life stage as we are interested in parasite life stage not individual parasites) were mapped to the updated 10x genome (*Fasciola*_10x_pilon, GCA_900302435.1) using Hisat2 (Kim et al., 2015). FeatureCounts (v2.0.0) was used to count mapped reads for exons in genes of interest (Liao et al., 2014), using the *F. hepatica* annotation (2020-05-WormBase version). Read counts were normalised to account variation in total read number within and between samples (Mortazavi et al., 2008), and expression patterns were presented as Reads Per Kilobase Million (rpkm) using EdgeR (Robinson et al., 2009a). The expression pattern of each gene in a family of interest was plotted across each lifecycle stage of the *F. hepatica* parasite.

4.5.3. Expression Patterns in TCBZ resistant and Susceptible Isolates

4.5.3.1. Experimental Design

Data was used from an experiment conducted by Dr Nicola Beesley, University of Liverpool. The experiment was designed to assess effect of TCBZ in sheep infected with TCBZ resistant flukes and those infected with TCBZ susceptible flukes. Gene expression patterns in a total

of 35 sheep experimentally infected with *F. hepatica* was assessed. Three TCBZ resistant and three TCBZ susceptible isolates (Table 4.2) previously described (Hodgkinson et al., 2018) were used to infect the sheep. For each isolate, a total of 6 sheep were infected, out of which three sheep were treated with TCBZ and the other three untreated. The untreated sheep were used as control for comparison of effect of TCBZ treatment. Therefore, for each isolate (either susceptible or resistant), there were 3 untreated sheep used as controls. From each sheep, three liver fluke parasites were used for RNA sequencing (Table 4.4). RNA library preparation and sequencings were done at the Centre for Genomic Research, University of Liverpool.

4.5.3.2. RNA Datasets

RNAseq data generated from using the Illumina NovaSeq platform, generating ~23million paired end reads per library (samples were randomised prior to library preparation and sequencing). Three parasites from each sheep were used for RNA sequencing (a total of 105 parasite samples were sequenced). Library preparation and sequencing was done at the Centre for Genomic Research, University of Liverpool. Paired-end reads for each parasite was mapped the updated 10x genome (*Fasciola*_10x_pilon, GCA_900302435.1) using Hisat2 (Kim et al., 2015). FeatureCounts (Liao et al., 2014) was used to count mapped exonic reads in *F. hepatica* annotated genes (WP15). Read counts were normalised, and expression patterns were presented as Reads Per Kilobase Million (rpkm) using EdgeR (Robinson et al., 2009a).

4.5.3.3. Statistical Model tests

Mixed effect model testing was done using lme4 in R to test statistical hypotheses (Paterson and Lello, 2003, Bates et al., 2014). Two linear mixed effects statistical models were tested (Table 4.4). The first model (A) assesses if genes of interest are always expressed (constitutive expression) at a different level in the TCBZ resistant relative to susceptible isolates (this test does not consider treatment with TCBZ), versus the null hypothesis that there no difference between expression between susceptible and resistant isolates. The second model (B) assesses if differential expression is induced in response to TCBZ treatment, in the resistant isolates compared with the susceptible isolate, versus the null hypothesis that difference in expression within resistant and susceptible isolates is not associated with the drug TCBZ. The difference in log-likelihood comparing the full model with a model omitting the interaction term was compared to a chi-square distribution, and a $P < 0.05$ was considered statistically significant. Effect sizes were calculated using emmeans package in R (Version 1.8.1.1, https://github.com/rvlenth/emmeans), where effect sizes 0.2 are considered small effect, 0.5 considered a medium effect, 0.8 considered a large effect, 1.2 considered a very large effect, and 2.2 considered a huge effect based on Cohen's d approach (Sullivan and Feinn, 2012).

Table 4.1: Proposed TCBZ mechanism of action and *F. hepatica* genes associated.

Mechanism of Drug Action	Gene Families	Number of Genes compiled
Altered tubulin-binding	Alpha tubulin	19
	Beta tubulin	11
	Delta tubulin	2
	Epsilon tubulin	1
	Gamma tubulin	1
	Unassigned tubulin	1
Altered drug uptake	ATP-binding cassette transporter (ABC) genes	27
	Adenylate Cyclase (AC) genes	8
	RAS	41
	ADP ribosylation factor	14
Altered drug metabolism	Cytochrome P450 (CYP450)	6
	Glutathione S-transferases (GSTs)	17
	Fatty Acid Binding Proteins (FABPs)	6

Table 4.2: Description of clonal isolates used for gene expression studies to compare expression in TCBZ resistant and susceptible *F. hepatica*

Triclabendazole Status	Clonal isolate	Identifier
Resistant	FhLivR1	A
	FhLivR2	B
	FhLivR3	C
Susceptible	FhLivS1	S
	FhLivS2	T
	FhLivS3	N

Table 4.3: Summary of the experimental design

Susceptible Isolates	Resistant Isolates	Paired sheep infection groups (sex, age matched, and randomly assigned)	

FhLivS1 (S)	FhLivR1 (A)	*Fh*LivS1 + *Fh*LivR1	• 5 sheep per group
			  ○ 2 infected with TCBZ susceptible isolate
			  ○ 2 infected with TCBZ resistant isolate
			  ○ 1 control sheep
			• Sheep 4 – 5 months at infection time
			• At infection time, sheep screened to be negative by ELISA (using *F. hepatica* ES antigens).
			• Metacecariae used for infecting sheep were ~ 59 days old. Oral infections using ~200 metacecariae per sheep
			• Sheep housed in-doors
			• Total of 36 sheep (24 female and 12 male)
			• One sheep (infected with FhLivS3 died)
			- At day 102 post-infection, for each infected group, 1 TCBZ susceptible infected sheep is treated while 1 TCBZ resistant infected is randomly chosen and treated with 10mg/kg TCBZ. The remaining 2 sheep in the group remained as controls
			- After 24 hours, sheep are euthanised, 3 adult parasites are retrieved from bile duct from each sheep
			- A total of 105 samples (3 parasites from 35 sheep) was used for RNA extraction

		FhLivS1 + FhLivR2	
		*Fh*LivS1 + *Fh*LivR3	
*Fh*LivS2 (T)	FhLivR2 (B)	*Fh*LivS2 + *Fh*LivR1	
		*Fh*LivS2 + *Fh*LivR2	
		*Fh*LivS2 + *Fh*LivR3	
FhLivS3 (N)	FhLivR3 (C)	*Fh*LivS3 + *Fh*LivR1	
		*Fh*LivS3 + *Fh*LivR2	
		*Fh*LivS3 + *Fh*LivR3	

Experimental design summary provided by Dr. Nicola Beesley (S, T, N, A, B, and C are identifiers used to identify each isolate), who conducted the experiment and provided the datasets for analysis.

Table 4.4: Summary of statistical models tested to explore expression in TCBZ susceptible and resistant *F. hepatica* isolates.

Model	Test Description	R code	
A	In the untreated isolates, is there a difference in constitutive gene expression between TCBZ Resistant and the Susceptible isolates?	*lmer (gene_rpkm ~ parasite_status + (1| clonename) + (1|sheep_no_kay), data = temp2[temp2$treat == "untreated" ,], REML = FALSE)*	parasite_status - fixed effect *clonename* - random effect *sheep_no_kay* - random effect
B	Are expression patterns induced by TCBZ treatment?	*lmer(gene_rpkm ~ parasite_status + treat + parasite_status:treat + (1| clonename), data = temp2, REML = FALSE)*	parasite_status - fixed effect treat – fixed effect *clonename* - random effect

gene_rpkm - reads per kilobase per million mapped reads for CDS of each gene
parasite_status - TCBZ status: resistant or susceptible
clonename – isolates A, B, C, S, T, N
sheep_no_kay – Assigned sheep number from which parasites were collected
treat – Treatment status of sheep (Sheep treated after infection or untreated)
lmer - Fit Linear Mixed-Effects Models Function

4.6. Results

4.6.1. Gene Expression profile across developmental stages

4.6.1.1. Elevated Tubulin-Binding Associated Genes in Developmental Stages

Generally, tubulin gene expression was elevated in metacercariae and NEJs, with a peak expression pattern noticed in NEJs 1 hour post excystment. As flukes approach maturity in the host (21 days post excystment), tubulin gene expression is reduced, although some tubulin gene expression was noticed in the adult flukes. For example, 3 alpha tubulin genes had peak expression levels in NEJs 1 hour post excystment but reduced levels in juveniles. Despite these results, in 2 alpha tubulin genes, expression levels were low in metacecariae, NEJs (1 hour and 3 hours) post excystment, peaked in juveniles, but reduced in adult flukes. Comparatively expression levels peaked in eggs in 1 alpha tubulin but reduced in a stable level in the other stages of live cycle (Figures 4.1). In 2 beta tubulins peak expression levels was noticed in NEJs 1 hour post excystment, but reduced levels in snails and juveniles. Peak expression level in juveniles was noticed in 1 beta tubulin gene, with reduced expression levels in other stages of the parasite (Figure 4.2).

In a gamma tubulin, peak expression level was noticed in NEJs 1 hour post excystment, while expression levels reduced in juveniles, but increased in adults. In a delta tubulin gene, expression level was stable in all stages of the parasite but peaked in juvenile flukes. In the epsilon tubulin peak expression level was noticed in metacecariae, but gradually reduced towards fluke maturity (Figure 4.10).

4.6.1.2. Increased Expression of Genes Associated with Drug Uptake in Developmental Stages

In most of the genes in these groups, there was elevated expression in metacercariae and NEJs, except in ABC genes. Adenyl cyclase (Figure 4.4), RAS (Figure 4.5), and ADP ribosylation factor (Figure 4.6) gene expression was slightly raised in metacercariae, NEJs (1 hour & 24 hours post excystment), while gene expression gradually reduced towards maturity. There were relatively lower expression levels in NEJs (21 days old), adult flukes, eggs, and snails than the levels noticed in the metacercariae and NEJs (1 hour & 24 hours post excystment). Findings suggest a general low expression level of ATP-binding cassette transporters (ABC) expression in all the parasite life stages (Figure 4.3). However, it appears there was slightly higher levels in some ABC genes in the NEJs (21 days old) and adult flukes.

4.6.1.3. Variable Expression of Genes Associated with TCBZ Metabolism Across Life-Cycle Stages

In the three families assessed in these groups, expression patterns varied among genes studied, within each life cycle stage and across each family. Two differential expression patterns were noticed in the cytochrome P450 genes (Figure 4.7). Out of the total 6 genes in the family, while expression gradually increased in 2 of these genes in the developmental stages and peaked in NEJs 1 hour post excystment, expression levels reduced towards maturity of the flukes. Contrary observations were seen in the remaining genes. In 3 cytochrome P450 genes, expression was unchanged in the developmental stages. However,

as flukes mature (in 21 days old NEJs and adults), cytochrome P450 gene expression gradually increased. In 1 cytochrome P450 gene, gene expression gradually increased from eggs through all the life stages and peaking and adult liver flukes. In most of the GST genes (Figure 4.8), there was a reduced expression in developmental stages. Results suggest these drug detoxification enzymes have slightly elevated expression in adult liver flukes and their eggs, except in a single GST gene predominantly active in metacercariae and NEJs (1 hour & 24 hours post excystment). Finally, FABP gene expression (Figure 4.9) was low in metacercariae and NEJs (1 hour & 24 hours post excystment) but becomes elevated in immature flukes (21 days old NEJs) and adult flukes. While expression of FABP genes were low in snails, except in gene - maker-scaffold10x_2403_pilon-snap-gene-0.20 with elevated expression level in snails.

Figure 4.1: Expression profile levels (in reads per kilo base per million mapped reads) of alpha tubulin genes in *F. hepatica* across various stages of the parasite

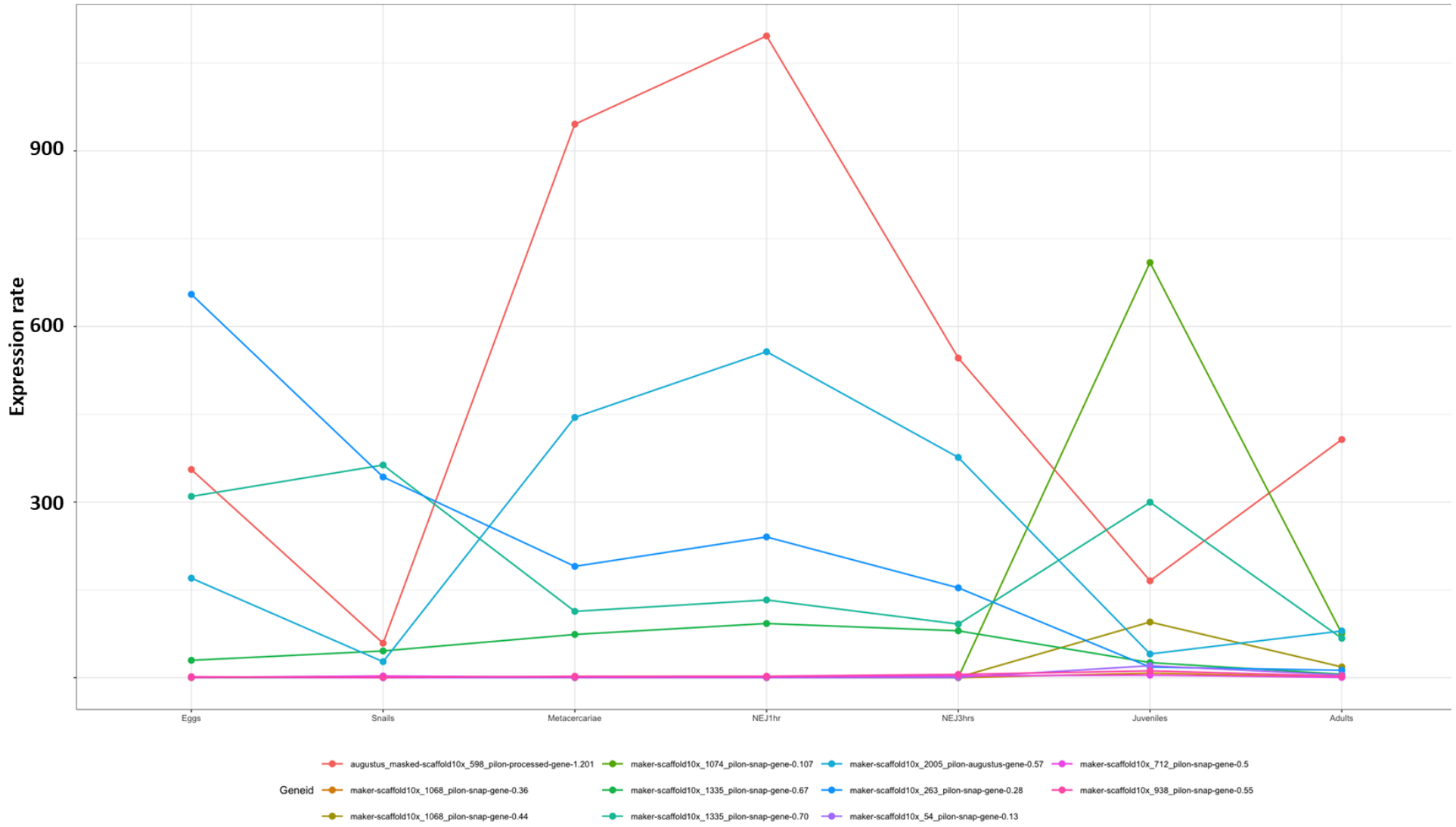

Figure 4.2: Expression profile levels (in reads per kilo base per million mapped reads) of beta tubulin genes in *F. hepatica* across various stages of the parasite

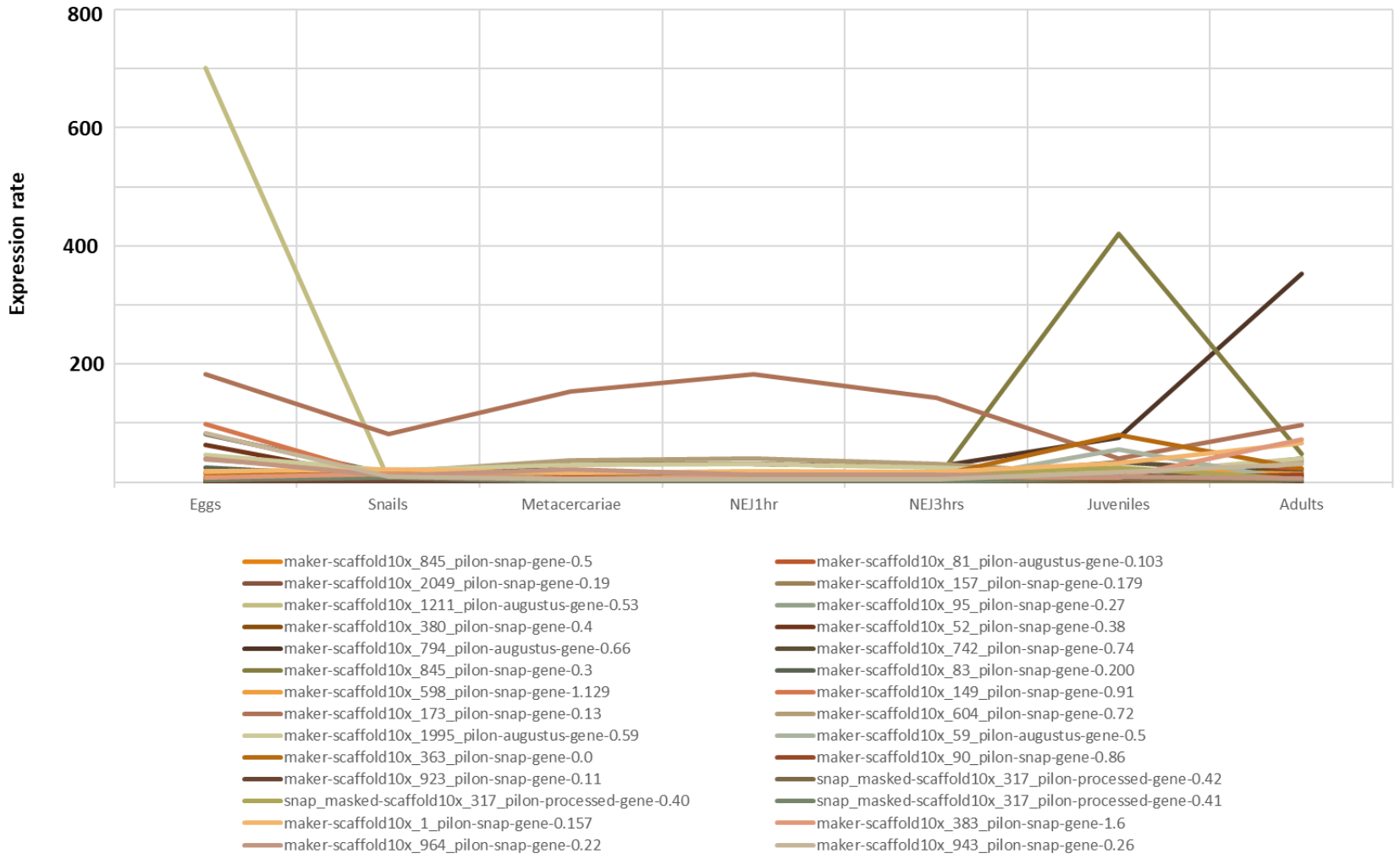

Figure 4.3: Expression profile levels (in reads per kilo base per million mapped reads) of ATP-binding cassette transporters (ABC) genes in *F. hepatica* across various stages of the parasite

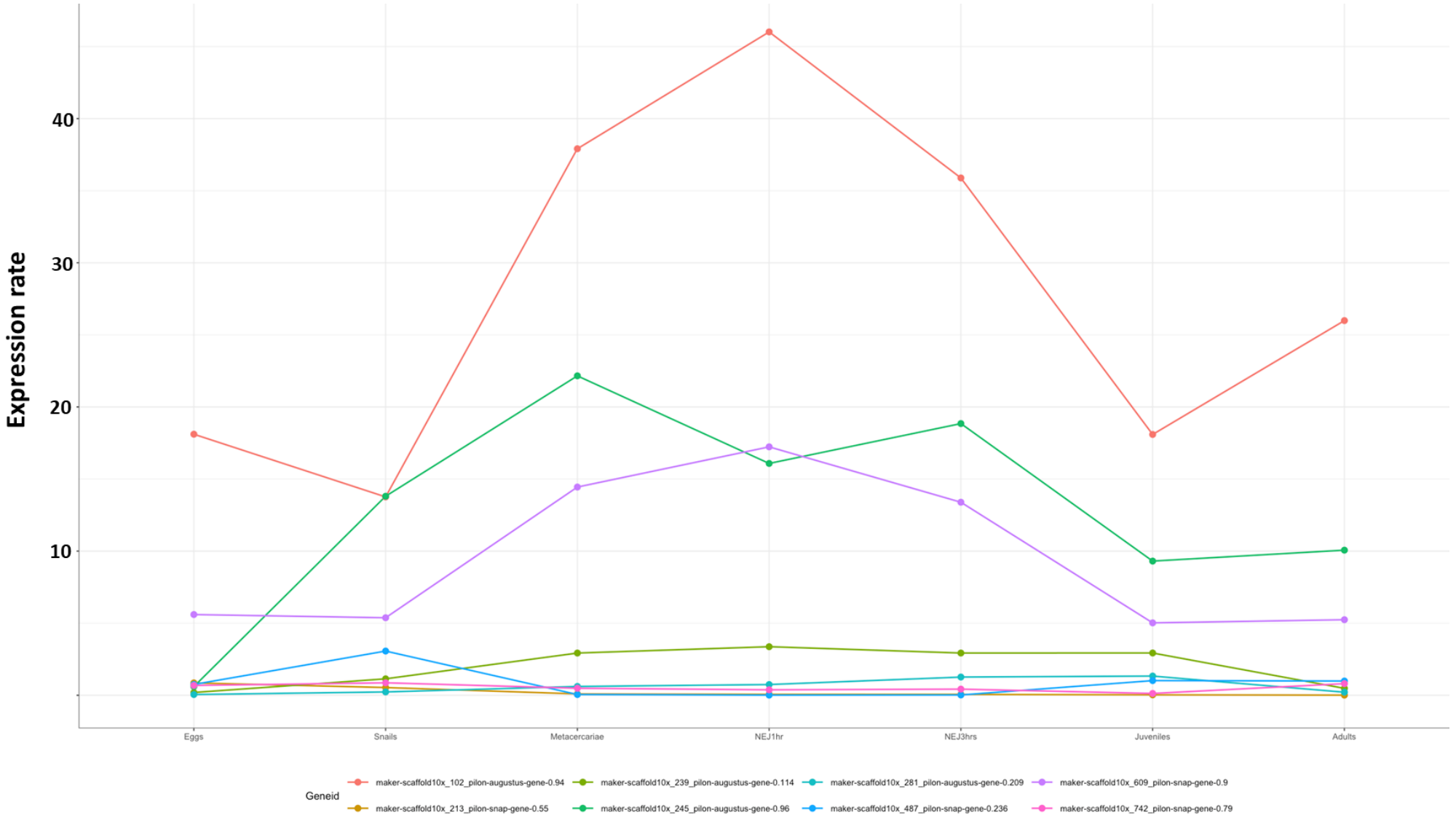

Figure 4.4: Expression profile levels (in reads per kilo base per million mapped reads) of adenylate cyclase genes in *F. hepatica* across various stages of the parasite

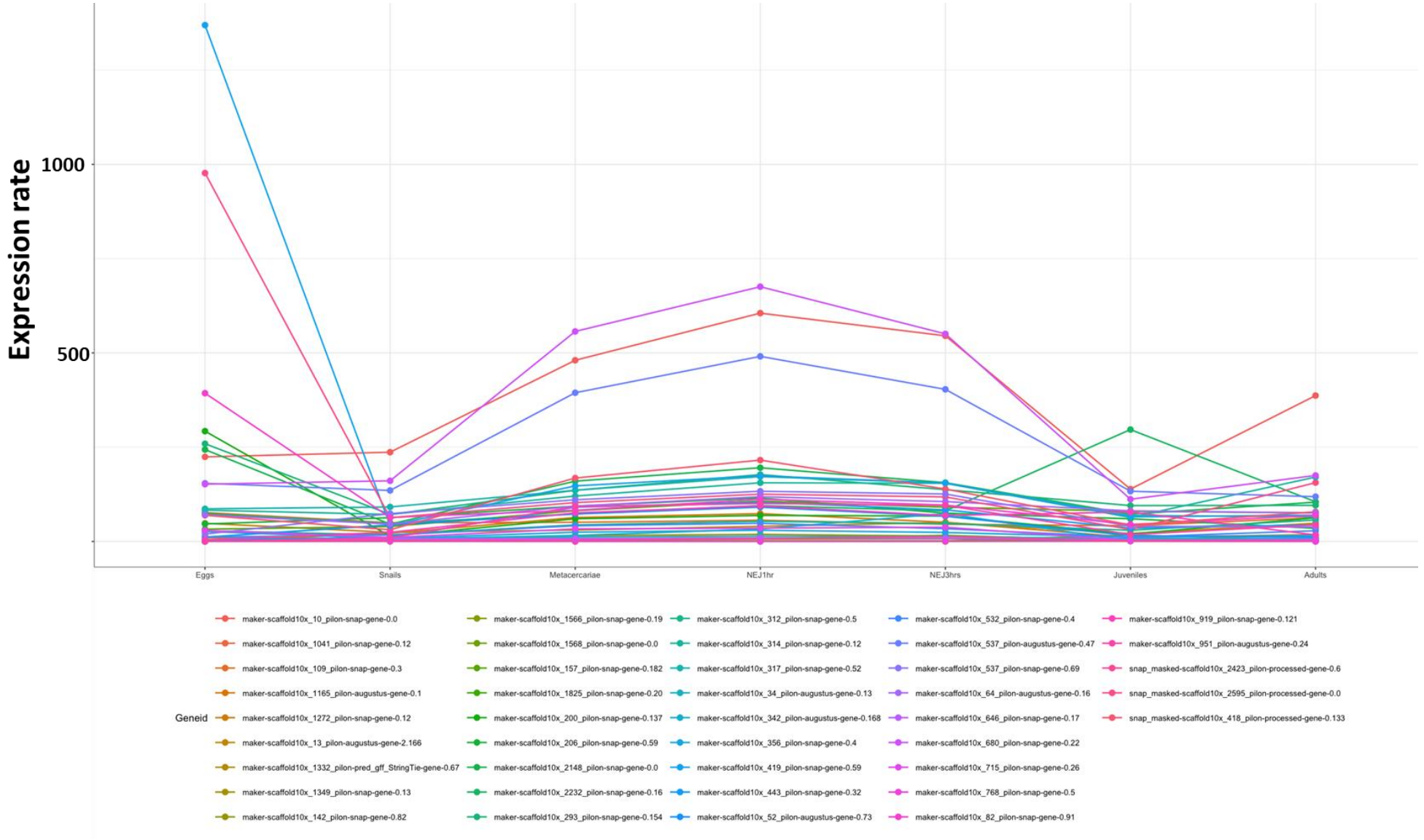

Figure 4.5: Expression profile levels (in reads per kilo base per million mapped reads) of alpha RAS genes in *F. hepatica* across various stages of the parasite

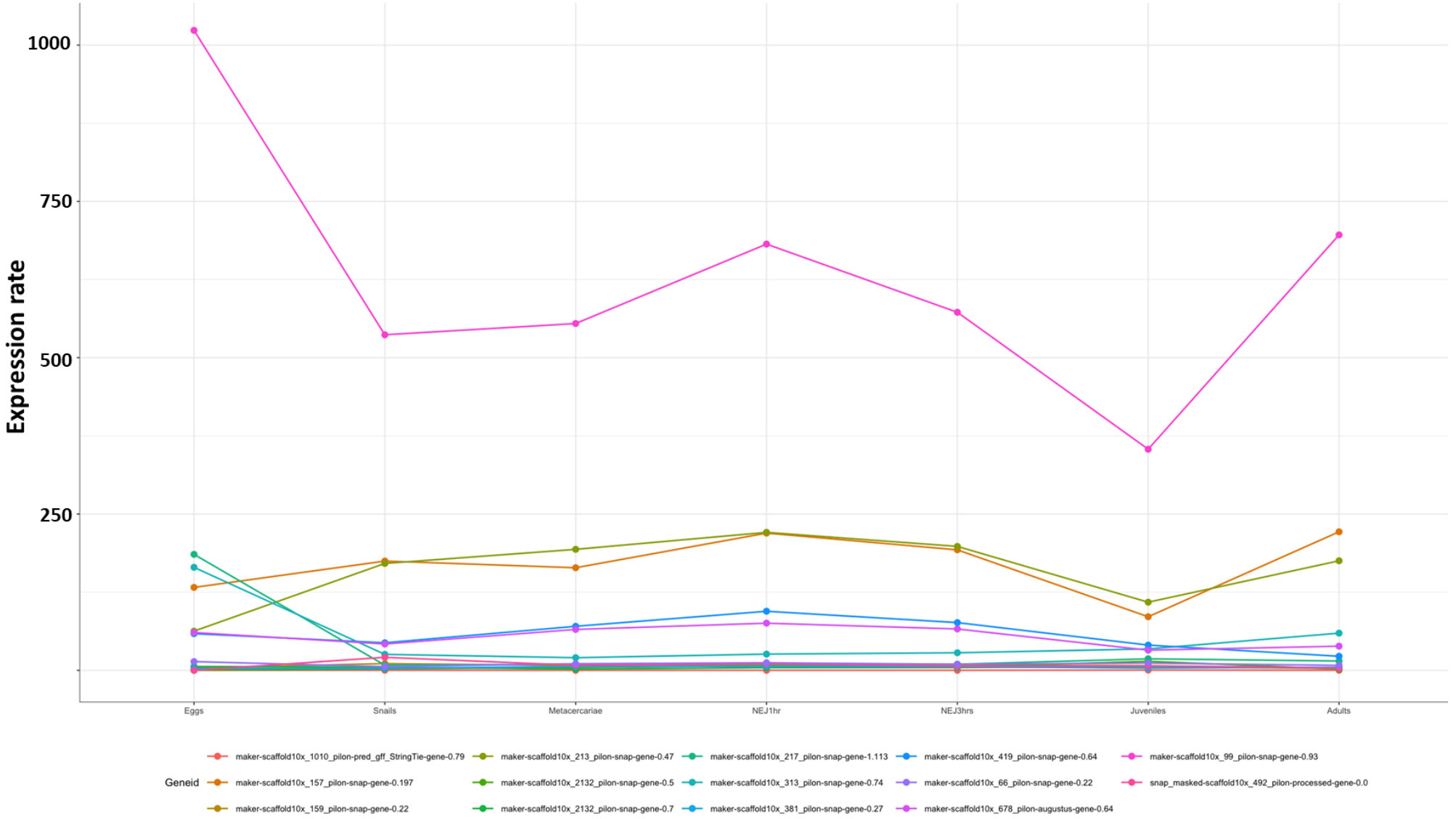

Figure 4.6: Expression profile levels (in reads per kilo base per million mapped reads) of ADP Ribosylation Factor genes in *F. hepatica* across various stages of the parasite

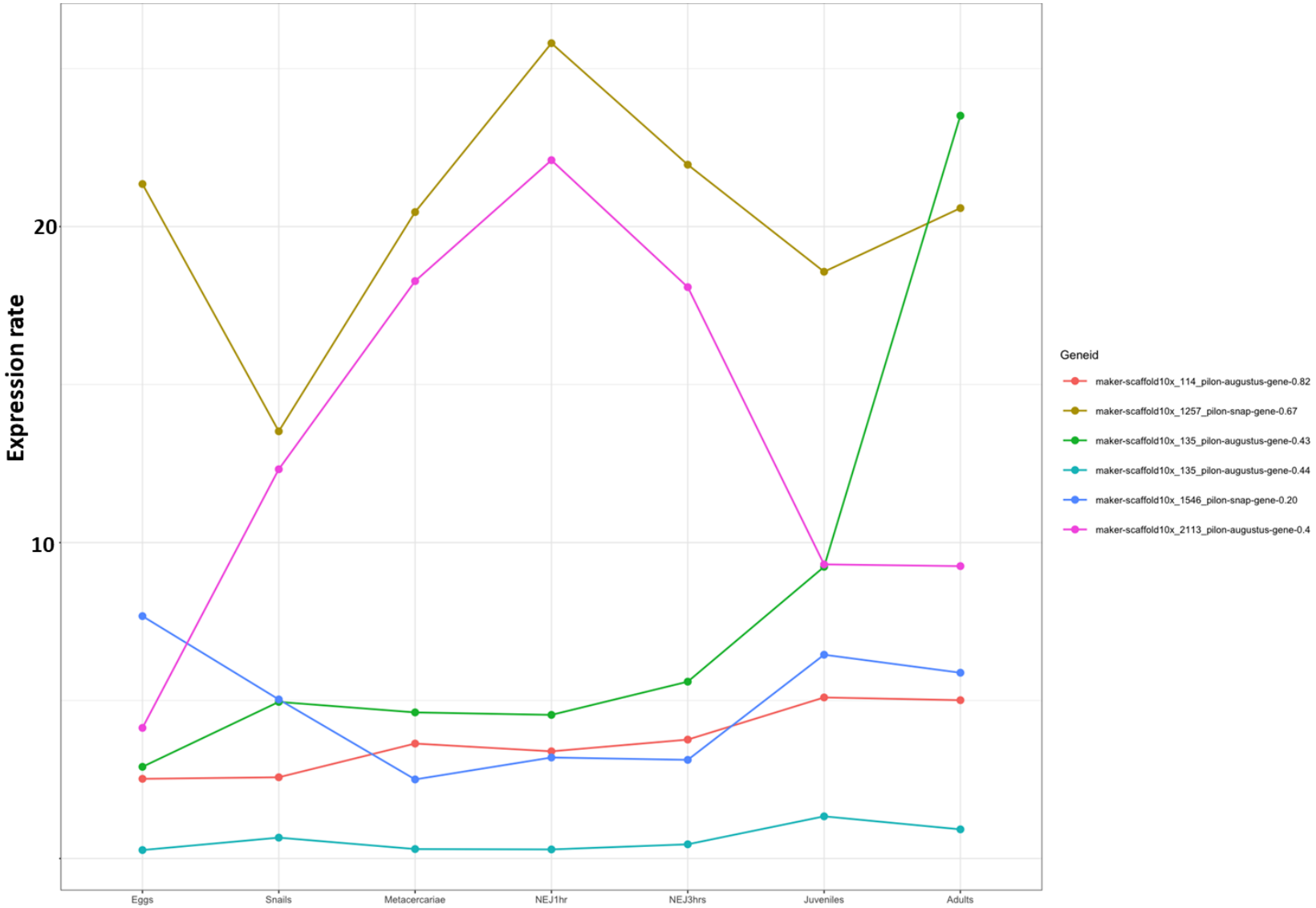

Figure 4.7: Expression profile levels (in reads per kilo base per million mapped reads) of cytochrome P450 genes in *F. hepatica* across various stages of the parasite

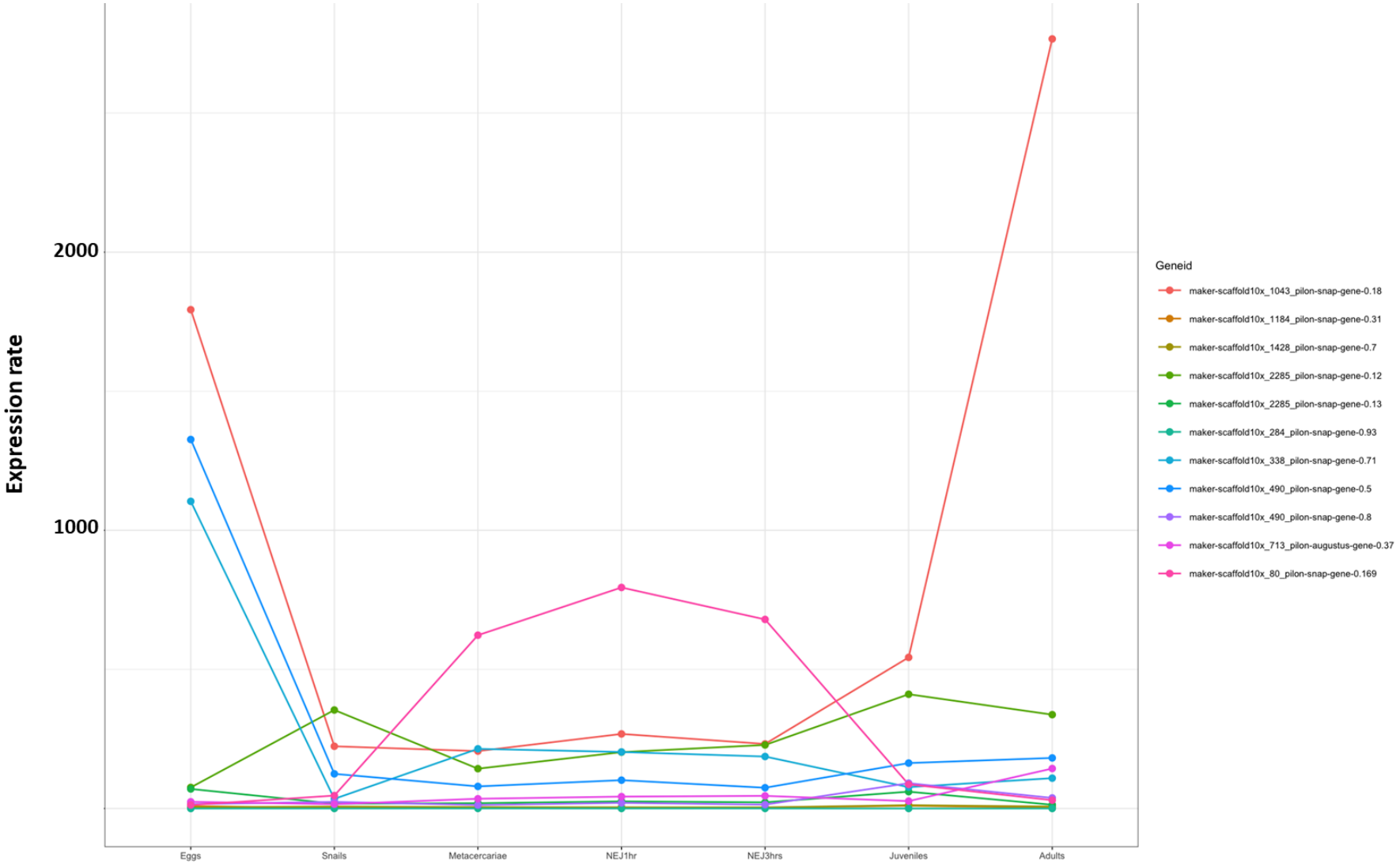

Figure 4.8: Expression profile levels (in reads per kilo base per million mapped reads) of glutathione S-transferases (GST) genes in *F. hepatica* across various stages of the parasite

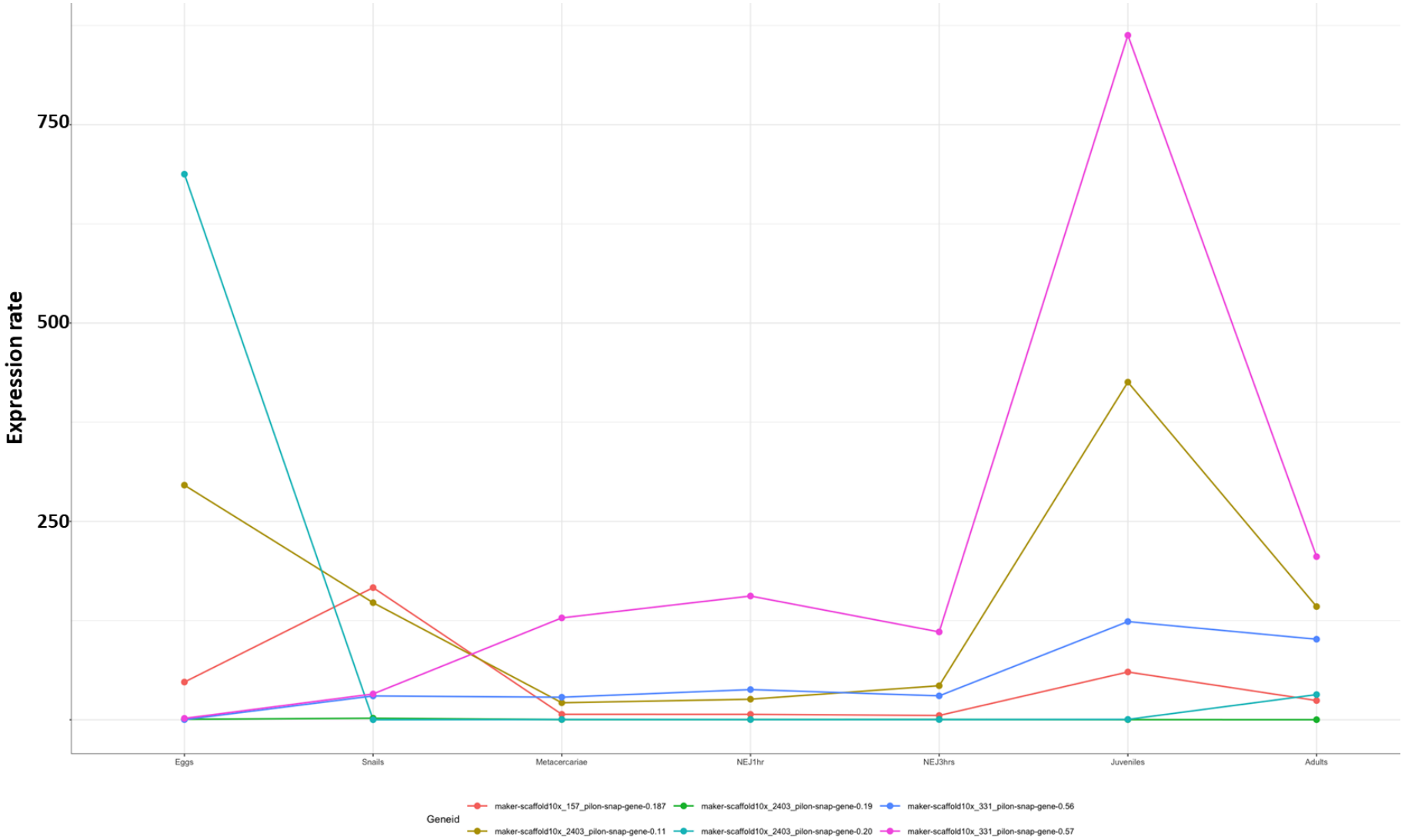

Figure 4.9: Expression profile levels (in reads per kilo base per million mapped reads) of fatty acid binding protein (FABP) genes in *F. hepatica* across various stages of the parasite

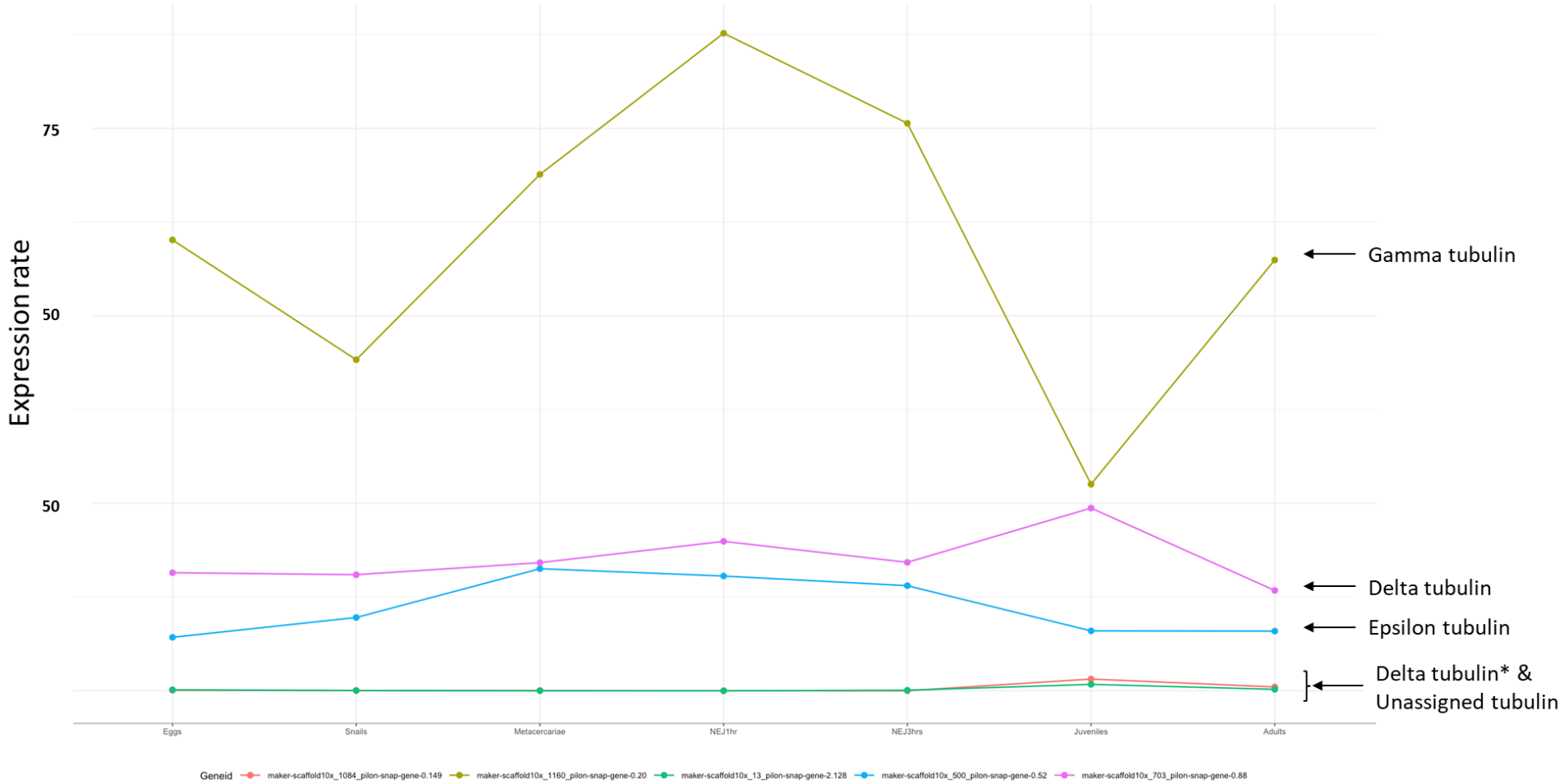

Figure 4.10: Expression profile levels (in reads per kilo base per million mapped reads) of Gamma, Delta, and Epsilon, and an unassigned (maker-scaffold10x_13_pilon-snap-gene-2.128) tubulin genes in *F. hepatica* across various stages of the parasite.

**Delta tubulin - maker-scaffold10x_1084_pilon-snap-gene-0.149 (orange colour)*

4.6.2. TCBZ changes expression of multiple genes in susceptible flukes compared to resistant ones

Statistical modelling to assess the impact of TCBZ in flukes in various gene families associated with resistance indicated that TCBZ treatment induced differences in expression levels of multiple genes in the resistant versus susceptible fluke isolates.

4.6.2.1. Expression Profile in Untreated Resistant versus Susceptible Isolates

Results of the first modelling analysis (Model A) suggest no constitutive differences in gene expression levels between susceptible and resistant isolates in almost all the families investigated. Out of the total 154 genes assessed, constitutive differences in gene expression levels between susceptible and resistant isolates was noticed in only 16 genes, i.e., one alpha tubulin and one gamma tubulin (Table 4.5), 1 ABCB (Table 4.6), 4 RASs (Table 4.7), 2 ACs and 1 ADP ribosylation factor (Table 4.8), 2 GSTs (Table 4.9), 3 Cytochrome P450s and 1 FABP gene (Table 4.10). However, only 9 of these genes showed statistically significant differences in expression levels between untreated susceptible and untreated resistant isolates (Table 4.5 – 4.10, Figure 4.12).

4.6.2.2. Effect of TCBZ treatment in Resistant and Susceptible Isolates

Administration of TCBZ to resistant and susceptible isolates changed the expression of multiple genes in each of the families studied, especially in the susceptible isolates in response to the drug. Generally, induced gene expression patterns were noticed in 52% ($P < 0.05$) of all the families assessed (excluding unexpressed genes and constitutively expressed genes). Induced gene expression in response to TCBZ was noticed in 88.9% of the tubulin genes (24 out of 27), 27.5% of the ABC genes (21 out of 24), 62.5% of the RAS genes (20 out of 32), 83.3% of the AC genes (5 out of 6), 58.3% of the ADP ribosylation factor genes (7 out of 12), 46.7% of the GST genes (7 out of 15), 50% of the Cytochrome P450 genes (2 out of 4), and 80% of the FABP genes (4 out of 5) (Tables 4.5 – 4.10).

Visual presentation of expression profiles suggests a similar expression pattern in resistant isolates treated with TCBZ and untreated ones, suggesting resistant fluke isolates were unresponsive to the drug as expected. In the susceptible isolates, TCBZ treatment initiated a statistically significant response of most genes to the drug in the treated sheep compared to the untreated ones (Figure 4.11). While the size of these responses to TCBZ varied across gene families, ranging from upregulating expression in some genes, and downregulation in other genes, statistical modelling indicated a significant impact of treatment on susceptible isolates compared to resistant ones in most of the genes (Tables 4.5 – 4.10).

Apart from these families showing induced responses to TCBZ in susceptible isolates, there were another 9 genes (across the families studied) that showed significant statistical differences in constitutive expression between resistant and susceptible flukes, that also seemed to show statistical evidence of induced expression response to the drug (Figure 4.12). It is unclear if this is a genuine drug induced expression or not.

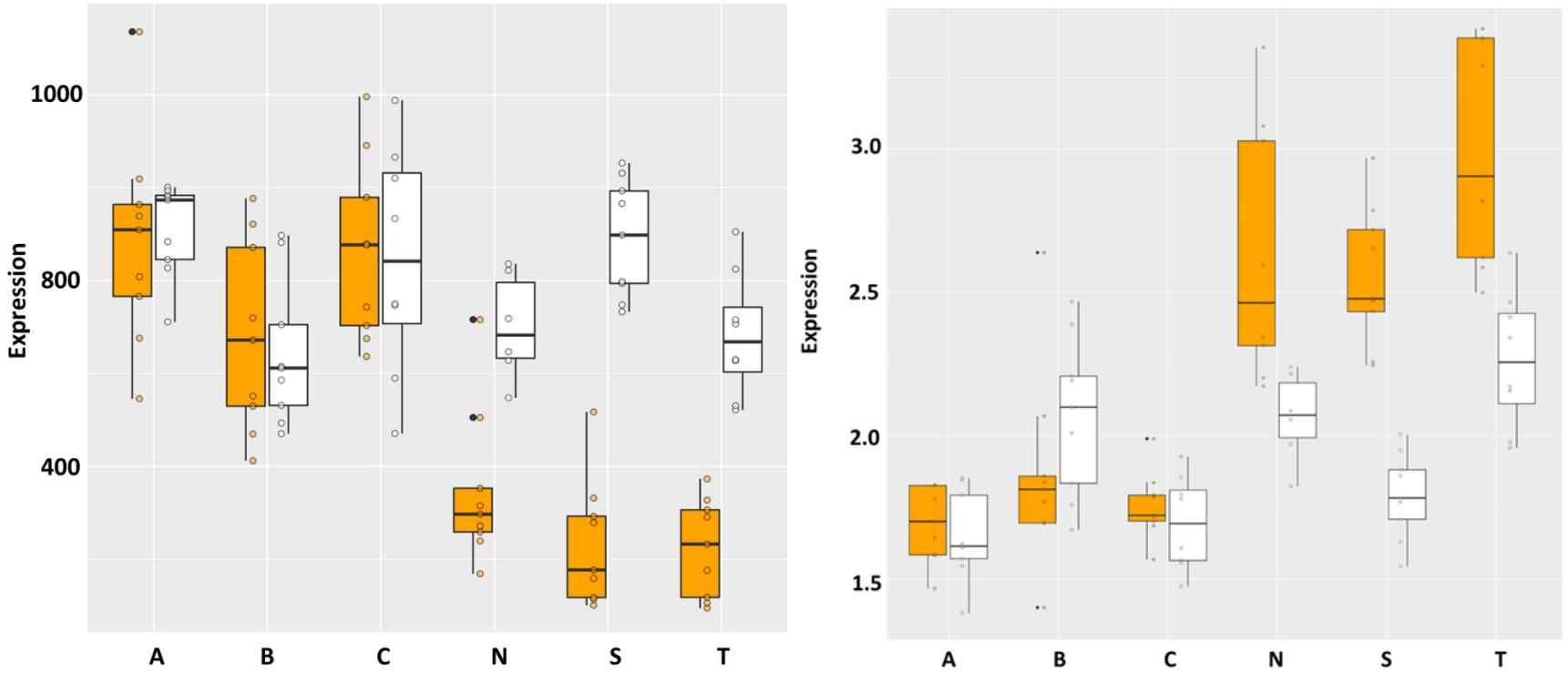

Figure 4.11: Plot showing gene expression levels. Gene expression levels (in reads per kilo base per million mapped reads) of α – Tubulin - maker-scaffold10x_313_pilon-snap-gene-0.79 (left) and GST - maker-scaffold10x_1189_pilon-snap-gene-0.107 (right) across (TCBZ resistant – A, B, & C) and (TCBZ susceptible – N, S, & T) isolates treated (Orange Colour) and untreated (White Colour). Drug administration in experimentally infected sheep induced gene expression of significance ($P < 0.05$) in *F. hepatica* TCBZ susceptible isolates compared to the resistant ones.

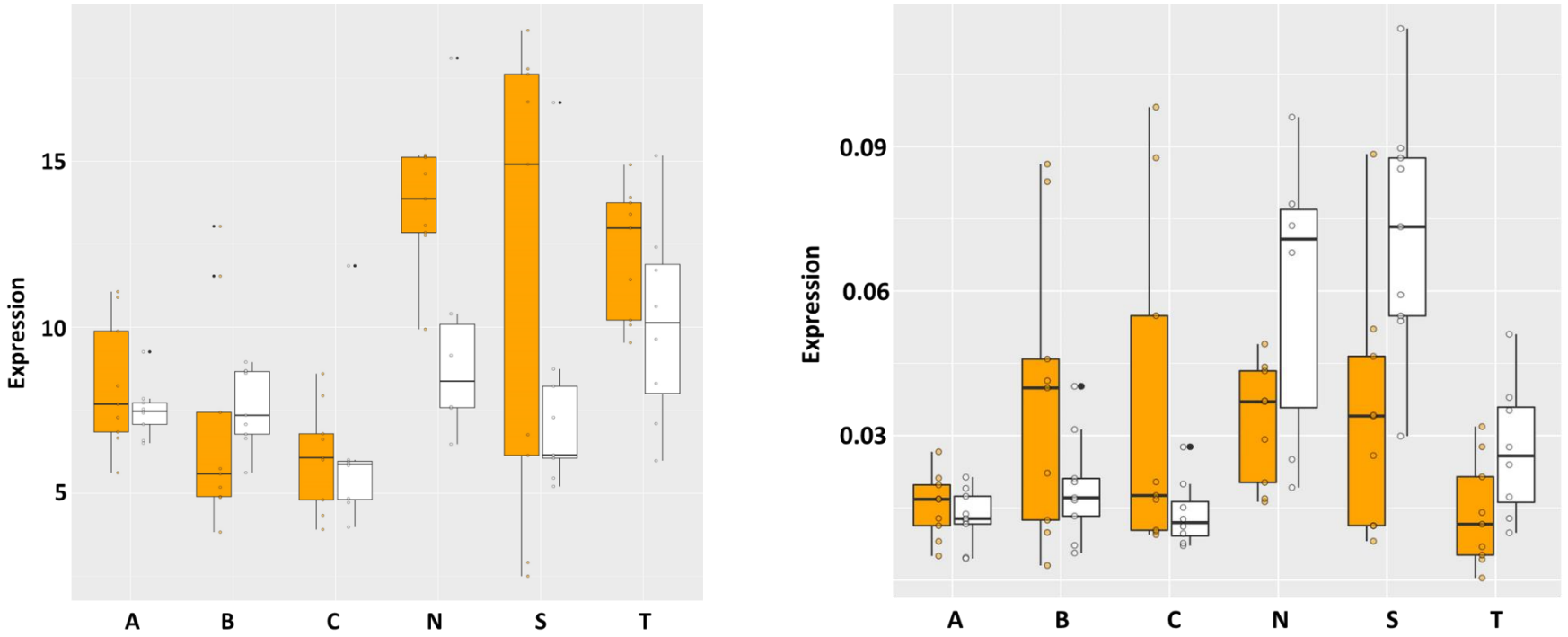

Figure 4.12: Plot showing expression levels. Gene expression levels (in reads per kilobase per million mapped reads) of ABCB gene - maker-scaffold10x_794_pilon-augustus-gene-0.66 (left) and Adenyl cyclase maker-scaffold10x_742_pilon-snap-gene-0.79- (right) across (TCBZ resistant – A, B, & C) and (TCBZ susceptible – N, S, & T) isolates treated (Orange Colour) and untreated (White Colour). In both genes, findings identified statistically significant constitutive gene expression (in untreated sheep infected with *F. hepatica* TCBZ resistant and susceptible isolates) and induced gene expression in susceptible isolates compared to resistant isolates in response to TCBZ treatment, suggesting TCBZ also affects the expression of some fluke basal genes.

Table 4.5: Tubulin genes assessed and their respective statistical significance for each of the mixed models tested

Tubulins Class	*F. hepatica* Genes	Model A						Model B					
		P-values	Effect Size	SE	dF	Upper CI	Lower CI	P-values	Effect Size	SE	dF	Upper CI	Lower CI
Alpha	augustus_masked-scaffold10x_474_pilon-processed-gene-0.52				ND					ND			
	maker-scaffold10x_13_pilon-snap-gene-2.125				ND					ND			
	maker-scaffold10x_13_pilon-snap-gene-2.129	0.2607	1.22 ‡‡‡‡	1.22	8.78	-1.55	3.99	0.0132 *	1.61 ‡‡‡‡	0.836	8.94	-0.286	3.5
	maker-scaffold10x_1444_pilon-snap-gene-0.40	0.2645	1.19 ‡‡‡	1.19	8.76	1.52	3.89	0.0110 *	1.55 ‡‡‡‡	0.76	8.94	-0.169	3.27
	maker-scaffold10x_1781_pilon-snap-gene-0.3				ND					ND			
	maker-scaffold10x_1781_pilon-snap-gene-0.4	0.4138	1.03 ‡‡‡	1.44	8.9	-2.23	4.28	0.01487 *	1.59 ‡‡‡‡	1.39	8.99	-1.55	4.73
	maker-scaffold10x_199_pilon-snap-gene-0.41	0.2591	1.26 ‡‡‡‡	1.25	8.79	-1.58	4.1	0.02175 *	1.6 ‡‡‡‡	0.829	8.94	-0.279	3.48
	maker-scaffold10x_2152_pilon-augustus-gene-0.28	0.2507	1.3 ‡‡‡‡	1.26	8.74	-1.57	4.17	0.02124 *	1.6 ‡‡‡‡	0.797	8.94	-0.206	3.4
	maker-scaffold10x_313_pilon-snap-gene-0.79	0.6921	0.274 ‡	0.805	8.51	-1.56	2.11	1.289e-11 ***	1.71 ‡‡‡‡	0.538	8.84	0.484	2.93
	maker-scaffold10x_313_pilon-snap-gene-0.80	0.04085 *	-1.8 ‡‡‡‡	0.892	0.892	-3.83	0.22	0.09831	-1.08 ‡‡‡	0.69	8.99	-2.64	0.486
	maker-scaffold10x_313_pilon-snap-gene-0.98	0.2618	-1.27 ‡‡‡‡	1.27	8.88	-4.14	1.6	0.1728	-1.37 ‡‡‡‡	0.799	8.95	-3.17	0.442
	maker-scaffold10x_45_pilon-snap-gene-0.45	0.2594	1.21 ‡‡‡‡	1.2	8.74	-1.52	3.94	0.01233 *	1.55 ‡‡‡‡	0.753	8.93	-0.151	3.26

	maker-scaffold10x_592_pilon-snap-gene-0.21	0.2356	1.32 ‡‡‡‡	1.24	8.77	-1.5	4.14	0.02116 *	1.64 ‡‡‡‡	0.828	8.94	-0.231	3.52
	maker-scaffold10x_62_pilon-snap-gene-0.31	0.3109	0.813 ‡‡‡	0.908	8.93	-1.24	2.87	0.001942 **	1.46 ‡‡‡‡	0.779	8.97	-0.305	3.22
	maker-scaffold10x_680_pilon-snap-gene-0.21	0.2850	1.17 ‡‡‡	1.23	8.76	-1.62	3.96	0.01875 *	1.5 ‡‡‡‡	0.793	8.94	-0.297	3.29
	maker-scaffold10x_809_pilon-snap-gene-0.10	0.2827	1.19 ‡‡‡	1.24	8.78	-1.63	4	0.01454 *	1.54 ‡‡‡‡	0.793	8.94	-0.26	3.33
	maker-scaffold10x_944_pilon-snap-gene-0.44	0.3097	1.15 ‡‡‡	1.27	8.74	-1.75	4.04	0.00999 **	1.48 ‡‡‡‡	0.764	8.93	-0.253	3.21
	snap_masked-scaffold10x_1189_pilon-processed-gene-0.77	0.3821	0.904 ‡‡‡	1.18	8.7	-1.78	3.59	0.01052 *	1.25 ‡‡‡‡	0.646	8.91	-0.214	2.72
	maker-scaffold10x_908_pilon-snap-gene-1.157				ND					ND			
Beta	augustus_masked-scaffold10x_598_pilon-processed-gene-1.201	0.4255	-0.714 ‡‡	1.03	8.87	-3.04	1.62	1.525e-06 ***	-1.53 ‡‡‡‡	0.52	8.9	-2.7	-0.349
	maker-scaffold10x_1068_pilon-snap-gene-0.36	0.1448	2.44 ‡‡‡‡‡	1.82	8.9	-1.68	6.56	0.2159	2.57 ‡‡‡‡‡	1.5	8.98	-0.816	5.95
	maker-scaffold10x_1068_pilon-snap-gene-0.44	0.2157	1.35 ‡‡‡‡	1.21	8.76	-1.4	4.11	0.03079 *	1.65 ‡‡‡‡	0.79	8.94	-0.14	3.44
	maker-scaffold10x_1074_pilon-snap-gene-0.107	0.3199	1.08 ‡‡‡	1.23	8.75	-1.72	3.88	0.005526 **	1.47 ‡‡‡‡	0.749	8.93	-0.227	3.16
	maker-scaffold10x_1335_pilon-snap-gene-0.67	0.07666	-1.14 ‡‡‡	0.673	9.42	-2.65	0.372	0.005376 **	-0.389 ‡	0.518	9.1	-1.56	0.781
	maker-scaffold10x_1335_pilon-snap-gene-0.70	0.4506	0.681 ‡‡	1.04	8.81	-1.67	3.04	1.053e-07 ***	1.69 ‡‡‡‡	0.629	8.92	0.263	3.11
	maker-	0.3635	-0.845 ‡‡‡	1.06	8.91	-3.25	1.56	0.04218 *	-0.579 ‡‡	0.814	8.96	-2.42	1.26

	scaffold10x_2005_pilon-augustus-gene-0.57				ND						ND		
	maker-scaffold10x_263_pilon-snap-gene-0.28	0.2944	1.09 ‡‡‡	1.17	8.76	-1.57	3.75	0.0216 *	1.43 ‡‡‡‡	0.732	8.92	-0.231	3.08
	maker-scaffold10x_54_pilon-snap-gene-0.13	0.6153	-0.292 ‡	0.681	7.32	-1.89	1.3	0.04384 *	0.165	0.61	8.9	-1.22	1.55
	maker-scaffold10x_712_pilon-snap-gene-0.5	0.5064	0.682 ‡‡	1.18	8.77	-2.01	3.37	0.01763 *	1.09 ‡‡‡	0.782	8.95	-0.682	2.86
Delta	maker-scaffold10x_938_pilon-snap-gene-0.55	0.3236	0.885 ‡‡‡	1.01	8.76	-1.42	3.19	0.008093 **	1.37 ‡‡‡‡	0.718	8.94	-0.261	2.99
	maker-scaffold10x_1084_pilon-snap-gene-0.149				ND						ND		
	maker-scaffold10x_703_pilon-snap-gene-0.88												
Epsilon	maker-scaffold10x_500_pilon-snap-gene-0.52	0.05404	1.63 ‡‡‡‡	0.869	8.79	-0.346	3.6	3.467e-06 ***	0.301 ‡	0.605	9.09	-1.07	1.67
Gamma	maker-scaffold10x_1160_pilon-snap-gene-0.20	0.003352 **	-0.883 ‡‡‡	0.383	9.37	-1.74	-0.0219	0.2046	-0.985 ‡‡‡	0.263	5.77	-1.63	-0.335
Unassigned	maker-scaffold10x_13_pilon-snap-gene-2.128	0.1264	1.74 ‡‡‡‡	1.23	8.73	-1.05	4.53	0.2057	1.79 ‡‡‡‡	0.831	8.94	-0.0956	3.67

*** P < 0.001, ** P < 0.01, * P < 0.05, ND – No Data (No Evidence of Gene Expression). Effect Sizes in resistant versus susceptible Isolates (‡small effect, ‡‡medium effect, ‡‡‡large effect, ‡‡‡‡very large effect, ‡‡‡‡‡huge effect)

Table 4.6: ATP-binding cassette transporters (ABC) genes assessed and their respective statistical significance for each of the mixed models tested

Class	*F. hepatica* ABC genes	Model A						Model B					
		P-values	Effect Size	SE	dF	Upper CI	Lower CI	P-values	Effect Size	SE	dF	Upper CI	Lower CI
ABCA	maker-scaffold10x_1_pilon-snap-gene-0.157	0.4184	-0.813 ‡‡‡	1.15	8.77	-3.43	1.8	1.77e-05 ***	0.155 ‡‡‡‡	0.939	8.97	-1.97	2.28
	snap_masked-scaffold10x_317_pilon-processed-gene-0.40	0.704	-0.184	0.511	4.54	-1.54	1.17	6.642e-14 ***	1.5 ‡‡‡‡	0.424	9.43	0.546	2.45
	snap_masked-scaffold10x_317_pilon-processed-gene-0.41	0.6246	-0.23 ‡	0.546	8.24	-1.48	1.02	6.002e-05 ***	0.569 ‡‡	0.433	8.88	-0.412	1.55
	snap_masked-scaffold10x_317_pilon-processed-gene-0.42	0.2161	0.591 ‡‡	0.532	8.28	-0.628	1.81	0.9306	0.565 ‡‡	0.403	8.88	-0.349	1.48
	maker-scaffold10x_90_pilon-snap-gene-0.86	0.2164	-0.824 ‡‡‡	0.741	8.78	-2.51	0.857	2.092e-05 ***	-1.61 ‡‡‡‡	0.613	8.92	-2.99	-0.218
	maker-scaffold10x_923_pilon-snap-gene-0.11	0.2898	1.24 ‡‡‡	1.32	8.76	-1.76	4.24	0.01751 *	1.51 ‡‡‡‡	0.792	8.93	-0.281	3.31
ABCB	maker-scaffold10x_1211_pilon-augustus-gene-0.53	0.3175	0.632 ‡‡	0.715	8.78	-0.993	2.26	0.006258 **	1.14 ‡‡‡	0.523	8.9	-0.0471	2.32
	maker-scaffold10x_157_pilon-snap-gene-0.179	0.2984	0.946 ‡‡‡	1.02	8.67	-1.39	3.28	0.004157 **	1.43 ‡‡‡‡	0.667	8.93	-0.0818	2.94
	maker-scaffold10x_2049_pilon-snap-gene-0.19	0.736	-0.259 ‡	0.897	8.65	-2.3	1.78	0.00163 **	-0.846 ‡‡‡	0.58	9.06	-2.16	0.466
	maker-scaffold10x_380_pilon-snap-gene-0.4	0.5783	-0.412 ‡	0.861	8.86	-2.36	1.54	0.001933 **	-0.93 ‡‡‡	0.789	8.96	-2.72	0.856
	maker-scaffold10x_52_pilon-snap-gene-0.38	0.5008	0.531 ‡‡	0.909	8.61	-1.54	2.6	0.2553	0.634 ‡‡	0.682	8.97	-0.91	2.18
	maker-scaffold10x_742_pilon-snap-gene-0.74	0.7991	-0.209 ‡	0.966	8.53	-2.41	2	0.000143 ***	0.533 ‡‡‡	0.659	8.93	-0.96	2.03
	maker-scaffold10x_794_pilon-	0.01096 *	-0.757 ‡‡	0.355	7.01	-1.6	0.083	0.006373 **	-1.23 ‡‡‡‡	0.321	8.94	-1.96	-0.502

	augustus-gene-0.66												
	maker-scaffold10x_845_pilon-snap-gene-0.3	0.9061	0.0608	0.606	8.57	-1.32	1.44	0.002312 **	0.655 ‡‡	0.416	8.79	-0.29	1.6
	maker-scaffold10x_845_pilon-snap-gene-0.5	0.3572	0.951 ‡‡‡	1.18	8.81	-1.72	3.62	0.001142 **	1.32 ‡‡‡‡	0.57	8.92	0.0291	2.61
	maker-scaffold10x_95_pilon-snap-gene-0.27			ND						ND			
ABCC	maker-scaffold10x_149_pilon-snap-gene-0.91	0.05221	2.08 ‡‡‡‡‡	1.1	8.77	-0.419	4.58	0.0005044 ***	2.27 ‡‡‡‡‡	0.619	8.86	0.867	3.67
	maker-scaffold10x_598_pilon-snap-gene-1.129	0.4816	-0.443 ‡	0.728	8.56	-2.1	1.22	8.794e-05 ***	0.318 ‡	0.69	8.98	-1.24	1.88
	maker-scaffold10x_83_pilon-snap-gene-0.200	0.08483	0.946 ‡‡‡	0.581	8.3	-0.385	2.28	0.7635	0.901 ‡‡‡	0.58	8.87	-0.414	2.22
ABCD	maker-scaffold10x_363_pilon-snap-gene-0.0	0.649	-0.434 ‡	1.13	9.54	-2.96	2.09	0.003995 **	0.255 ‡	0.858	9.04	-1.69	2.19
	maker-scaffold10x_59_pilon-augustus-gene-0.5	0.6827	-0.31 ‡	0.881	8.59	-2.32	1.7	0.0004677 ***	0.341 ‡	0.616	8.91	-1.05	1.74
ABCE	maker-scaffold10x_173_pilon-snap-gene-0.13	0.1484	-0.42 ‡	0.322	5.22	-1.24	0.397	1.066e-11 ***	-1.88 ‡‡‡‡	0.499	8.97	-3.01	-0.748
ABCF	maker-scaffold10x_1995_pilon-augustus-gene-0.59	0.8737	0.118	0.868	8.46	-1.86	2.1	0.0004134 ***	-0.657 ‡‡‡	0.378	8.69	-1.52	0.202
	maker-scaffold10x_604_pilon-snap-gene-0.72	0.7354	-0.135	0.465	7.98	-1.21	0.937	2.2e-16 ***	-2.07 ‡‡‡‡‡	0.489	8.79	-3.18	-0.958
ABCG	maker-scaffold10x_383_pilon-snap-gene-1.6	0.2062	1.5 ‡‡‡‡	1.32	8.87	-1.48	4.49	2.2e-16 ***	-2.62 ‡‡‡‡‡	0.631	9.22	-4.04	-1.19
	maker-scaffold10x_943_pilon-snap-gene-0.26	0.3881	-0.434 ‡	0.573	8.46	-1.74	0.876	0.01437 *	-0.826 ‡‡‡	0.37	8.9	-1.67	0.0134
	maker-scaffold10x_964_pilon-snap-gene-0.22			ND						ND			

*** P < 0.001, ** P < 0.01, * P < 0.05, ND – No Data (No Evidence of Gene Expression). Effect Sizes in resistant versus susceptible Isolates (‡small effect, ‡‡medium effect, ‡‡‡large effect, ‡‡‡‡very large effect, ‡‡‡‡‡huge effect)

Table 4.7: RAS genes assessed and their respective statistical significance for each of the mixed models tested

F. hepatica RAS Genes	Model A						Model B					
	P-values	Effect Size	SE	dF	Upper CI	Lower CI	P-values	Effect Size	SE	dF	Upper CI	Lower CI
maker-scaffold10x_314_pilon-snap-gene-0.12	0.5007	-0.522 ‡‡	0.892	8.62	-2.55	1.51	0.0001544 ***	-1.13 ‡‡‡‡	0.379	8.75	-1.99	-0.268
maker-scaffold10x_206_pilon-snap-gene-0.59	0.5206	-0.508 ‡‡	0.912	8.52	-2.59	1.57	7.228e-05 ***	0.424 ‡	0.448	8.87	-0.593	1.44
maker-scaffold10x_532_pilon-snap-gene-0.4	0.002646 **	-2.66 ‡‡‡‡‡	0.773	8.77	-4.41	-0.905	0.5713	-1.94 ‡‡‡	0.546	8.83	-3.18	-0.7
maker-scaffold10x_951_pilon-augustus-gene-0.24			ND						ND			
maker-scaffold10x_1272_pilon-snap-gene-0.12	0.4783	-0.658 ‡‡	1.07	8.73	-3.09	1.77	0.8972	-0.702 ‡‡	1.18	8.99	-3.38	1.98
maker-scaffold10x_1165_pilon-augustus-gene-0.1	0.09977	-1.46 ‡‡‡‡	0.947	8.86	-3.61	0.684	0.03229 *	-0.862 ‡‡‡	0.639	8.97	-2.31	0.583
maker-scaffold10x_1566_pilon-snap-gene-0.19			ND						ND			
maker-scaffold10x_142_pilon-snap-gene-0.82	0.7162	-0.257 ‡	0.834	9.06	-2.14	1.63	1.211e-06 ***	0.76 ‡‡	0.618	9	-0.637	2.16
maker-scaffold10x_317_pilon-snap-gene-0.52	0.8041	0.429 ‡	2.02	8.68	-4.17	5.03	0.4027	0.183	1.22	8.98	-2.59	2.95
maker-scaffold10x_34_pilon-augustus-gene-0.13	0.4143	0.814 ‡‡‡	1.14	8.55	-1.79	3.41	4.332e-05 ***	1.52 ‡‡	0.832	8.96	-0.367	3.4
maker-scaffold10x_109_pilon-snap-gene-0.3	0.07122	-1.39 ‡‡‡‡	0.808	8.66	-3.23	0.445	0.01063 *	-0.4 ‡	0.361	8.8	-1.22	0.42
maker-scaffold10x_419_pilon-snap-gene-0.59	0.1441	1.53 ‡‡‡‡	1.14	8.57	-1.07	4.13	0.06998	1.04 ‡‡‡	0.796	8.93	-0.764	2.84
maker-scaffold10x_13_pilon-augustus-gene-2.166	0.5744	0.307 ‡	0.636	8.91	-1.13	1.75	0.1473	-0.0213	0.349	8.96	-0.811	0.769
maker-scaffold10x_52_pilon-augustus-gene-0.73	0.202	1.56 ‡‡‡‡	1.35	8.85	-1.51	4.63	0.01315 *	1.92 ‡‡‡‡	1.09	8.98	-0.546	4.38
maker-scaffold10x_64_pilon-augustus-gene-0.16	0.1436	-0.987 ‡‡‡	0.734	8.7	-2.66	0.682	0.05063	-0.477 ‡	0.533	8.96	-1.68	0.729
maker-scaffold10x_10_pilon-snap-gene-0.0	0.1263	-0.853 ‡‡‡	0.602	8.68	-2.22	0.518	0.1081	-1.02 ‡‡‡	0.421	8.99	-1.97	-0.0715
maker-scaffold10x_680_pilon-snap-gene-0.22	0.1551	-0.865 ‡‡‡	0.664	8.46	-2.38	0.652	0.1989	-0.462 ‡	0.394	8.81	-1.36	0.431
maker-scaffold10x_2148_pilon-snap-gene-0.0	0.6569	0.452 ‡	1.18	8.54	-2.24	3.15	0.5125	0.499 ‡	0.808	8.94	-1.33	2.33
maker-scaffold10x_919_pilon-snap-gene-0.121	0.5551	-0.187	0.336	4.6	-1.07	0.7	7.297e-14 ***	1.49 ‡‡‡‡	0.46	8.87	0.443	2.53
maker-scaffold10x_200_pilon-snap-gene-0.137	0.004319 **	1.94 ‡‡‡‡	0.61	8.5	0.548	3.33	0.01002 *	2.12 ‡‡‡‡‡	0.412	8.92	1.18	3.05
maker-scaffold10x_537_pilon-snap-gene-0.69	0.9541	-0.0431	0.877	8.78	-2.03	1.95	0.002714 **	-0.583 ‡‡	0.873	8.99	-2.56	1.39
maker-scaffold10x_537_pilon-augustus-gene-0.47	0.5217	-0.299 ‡	0.493	4.54	-1.6	1.01	0.2212	0.0327	0.209	4.74	-0.514	0.579
maker-scaffold10x_356_pilon-snap-gene-0.4	0.3077	-0.822 ‡‡‡	0.91	8.52	-2.9	1.25	1.86e-09 ***	-1.6 ‡‡	0.505	9.13	-2.74	-0.461

maker-scaffold10x_312_pilon-snap-gene-0.5	0.06421	-1.02 ‡‡‡	0.592	5.55	-2.49	0.462	0.005496 **	-1.23 ‡‡‡‡	0.391	8.82	-2.11	-0.341
maker-scaffold10x_768_pilon-snap-gene-0.5	0.253	1.06 ‡‡‡	1.04	8.76	-1.3	3.42	0.009506 **	1.48 ‡‡‡‡	0.656	8.94	-0.00359	2.97
snap_masked-scaffold10x_2595_pilon-processed-gene-0.0	0.2829	-0.635 ‡‡	0.665	8.68	-2.15	0.878	1.115e-05 ***	-1.51 ‡‡‡‡	0.427	8.83	-2.48	-0.541
maker-scaffold10x_1041_pilon-snap-gene-0.12			ND						ND			
maker-scaffold10x_82_pilon-snap-gene-0.91	0.8336	0.21 ‡	1.17	8.66	-2.45	2.87	0.5006	0.389 ‡	0.764	8.95	-1.34	2.12
maker-scaffold10x_1332_pilon-pred_gff_StringTie-gene-0.67			ND						ND			
snap_masked-scaffold10x_2423_pilon-processed-gene-0.6	0.08546	-1.21 ‡‡‡‡	0.745	8.79	-2.91	0.478	0.5055	-0.889 ‡‡‡	0.738	8.95	-2.56	0.782
maker-scaffold10x_646_pilon-snap-gene-0.17	0.4347	-0.64 ‡‡	0.939	8.61	-2.78	1.5	0.6338	-0.593 ‡‡	0.698	8.91	-2.17	0.988
maker-scaffold10x_2232_pilon-snap-gene-0.16	0.1586	-0.805 ‡‡‡	0.625	8.59	-2.23	0.62	0.0003124 ***	-1.44 ‡	0.523	9.08	-2.63	-0.262
maker-scaffold10x_443_pilon-snap-gene-0.32	0.5965	-0.495 ‡	1.09	8.73	-2.97	1.98	1.98e-08 ***	-1.57 ‡‡‡‡	0.493	8.79	-2.69	-0.448
maker-scaffold10x_293_pilon-snap-gene-0.154	0.02689 *	-1.73 ‡‡‡‡	0.771	8.76	-3.48	0.0262	0.3445	-1.44 ‡‡‡‡	0.577	8.93	-2.75	-0.136
snap_masked-scaffold10x_418_pilon-processed-gene-0.133	0.1144	-1.18 ‡‡‡	0.805	8.75	-3.01	0.645	6.792e-09 ***	-2.26 ‡‡‡‡‡	0.604	8.97	-3.62	-0.89
maker-scaffold10x_342_pilon-augustus-gene-0.168	0.1846	-0.545 ‡‡	0.453	8.11	-1.59	0.496	0.01324 *	-0.782 ‡‡	0.382	8.99	-1.65	0.083
maker-scaffold10x_1349_pilon-snap-gene-0.13	0.5791	0.262 ‡	0.546	8.25	-0.991	1.52	2.641e-05 ***	-0.649 ‡‡	0.565	8.93	-1.93	0.632
maker-scaffold10x_1825_pilon-snap-gene-0.20			ND						ND			
maker-scaffold10x_715_pilon-snap-gene-0.26	0.03032 *	1.52 ‡‡‡‡	0.692	9.15	-0.0437	3.08	0.04747 *	0.76 ‡‡	0.534	8.96	-0.45	1.97
maker-scaffold10x_1568_pilon-snap-gene-0.0	0.9644	-0.0195	0.511	8.56	-1.18	1.14	0.01892 *	0.445 ‡	0.373	8.82	-0.4	1.29
maker-scaffold10x_157_pilon-snap-gene-0.182	0.1161	-1.11 ‡‡‡	0.762	8.85	-2.84	0.613	0.03694 *	-1.24 ‡‡‡‡	0.317	8.7	-1.96	-0.52

*** P < 0.001, ** P < 0.01, * P < 0.05, ND – No Data (No Evidence of Gene Expression). Effect Sizes in resistant versus susceptible Isolates (‡small effect, ‡‡medium effect, ‡‡‡large effect, ‡‡‡‡very large effect, ‡‡‡‡‡huge effect)

Table 4.8: Adenylate Cyclase (AC) and ADP ribosylation factor genes assessed and their respective statistical significance for each of the mixed models tested

Gene family	*F. hepatica* Genes	Model A						Model B					
		P-values	Effect Size	SE	dF	Upper CI	Lower CI	P-values	Effect Size	SE	dF	Upper CI	Lower CI
Adenylate Cyclase	maker-scaffold10x_609_pilon-snap-gene-0.9	0.4456	-0.766 ‡‡	1.15	8.89	-3.38	1.85	0.03913 *	-1.15 ‡‡‡	1.07	8.99	-3.57	1.27
	maker-scaffold10x_102_pilon-augustus-gene-0.94	0.6352	0.236 ‡	0.579	8.6	-1.08	1.56	7.316e-05 ***	-0.61 ‡‡	0.432	8.9	-1.59	0.368
	maker-scaffold10x_213_pilon-snap-gene-0.55	0.001102 **	-0.993 ‡‡‡	0.33	3.83	-1.92	-0.0608	0.4073	-0.711 ‡‡	0.25	6.73	-1.31	-0.114
	maker-scaffold10x_742_pilon-snap-gene-0.79	0.01488 *	-2.18 ‡‡‡‡‡	0.854	8.86	-4.12	-0.242	2.127e-06 ***	-0.854 ‡‡‡	0.548	8.94	-2.09	0.386
	maker-scaffold10x_245_pilon-augustus-gene-0.96	0.3213	-0.694 ‡‡	0.793	8.8	-2.49	1.11	3.16e-05 ***	-1.28 ‡‡‡‡	0.637	8.98	-2.73	0.157
	maker-scaffold10x_281_pilon-augustus-gene-0.209	0.8207	0.147	0.757	8.56	-1.58	1.87	0.02382 *	0.584 ‡‡	0.552	8.95	-0.666	1.83
	maker-scaffold10x_487_pilon-snap-gene-0.236	0.06765	0.923 ‡‡‡	0.508	11.3	-0.191	2.04	8.792e-06 ***	1.72 ‡‡‡‡	0.585	9.03	0.395	3.04
	maker-scaffold10x_239_pilon-augustus-gene-0.114	0.6814	0.288 ‡	0.818	8.69	-1.57	2.15	0.07886	0.574 ‡‡	0.581	8.9	-0.743	1.89
ADP ribosylation factor	maker-scaffold10x_157_pilon-snap-gene-0.197	0.3851	-0.738 ‡‡	0.97	8.76	-2.94	1.47	0.0008621 ***	-1 ‡‡‡	0.45	9.12	-2.02	0.0148
	maker-scaffold10x_159_pilon-snap-gene-0.22	0.1132	1.75 ‡‡‡‡	1.19	8.63	-0.951	4.46	0.8513	1.48 ‡‡‡‡	0.726	8.94	-0.166	3.12
	maker-scaffold10x_313_pilon-snap-gene-0.74	0.4157	-0.307 ‡	0.432	8.45	-1.29	0.68	0.005186 **	0.289 ‡	0.288	8.61	-0.367	0.946
	maker-scaffold10x_2132_pilon-snap-gene-0.5	0.03703 *	-1.26 ‡‡‡‡	0.609	8.58	-2.65	0.129	0.0009905 ***	-0.579 ‡‡	0.511	8.92	-1.74	0.579
	maker-scaffold10x_2132_pilon-snap-gene-0.7	0.1362	3.86 ‡‡‡‡‡	2.81	8.87	-2.5	10.2	0.8991	2.85 ‡‡‡‡‡	1.8	8.99	-1.23	6.92
	maker-scaffold10x_217_pilon-snap-gene-1.113	0.2363	1.5 ‡‡‡‡	1.41	8.86	-1.7	4.7	0.008978 **	1.77 ‡‡‡‡	0.858	8.94	-0.173	3.71
	maker-scaffold10x_1010_pilon-pred_gff_StringTie-gene-0.79	0.132	1.88 ‡‡‡‡	1.35	8.82	-1.18	4.94	0.1655	1.94 ‡‡‡‡	0.911	8.95	-0.12	4.01
	maker-scaffold10x_678_pilon-augustus-gene-0.64	0.2059	-0.988	0.865	8.79	-2.95	0.977	0.7095	-0.699 ‡‡	0.647	9.01	-2.16	0.765
	maker-scaffold10x_99_pilon-snap-gene-0.93			ND						ND			
	snap_masked-scaffold10x_492_pilon-processed-gene-0.0	0.7572	0.207 ‡	0.797	9.19	-1.59	2.01	0.0002484 ***	0.87 ‡‡‡	0.577	8.96	-0.437	2.18
	maker-scaffold10x_381_pilon-snap-gene-0.27	0.5277	0.528 ‡‡	0.97	9.06	-1.66	2.72	5.325e-08 ***	-0.62 ‡‡‡	1.02	9.04	-2.93	1.69
	maker-scaffold10x_419_pilon-snap-gene-0.64	0.165	1.32 ‡‡‡‡	1.04	8.46	-1.06	3.69	1.697e-07 ***	2.13 ‡‡‡‡‡	0.547	8.89	0.89	3.37
	maker-scaffold10x_213_pilon-snap-gene-0.47	0.1905	1.36 ‡‡‡‡	1.14	8.92	-1.23	3.94	0.03049 *	1.83 ‡‡‡‡	1.15	8.98	-0.768	4.42
	maker-scaffold10x_66_pilon-snap-gene-0.22	0.4219	-0.909 ‡‡‡	1.3	8.81	-3.85	2.03	0.0001446 ***	-0.135	1.24	8.98	-2.95	2.68

*** P < 0.001, ** P < 0.01, * P < 0.05, ND – No Data (No Evidence of Gene Expression). Effect Sizes in resistant versus susceptible Isolates (‡small effect, ‡‡medium effect, ‡‡‡large effect, ‡‡‡‡very large effect, ‡‡‡‡‡huge effect)

Table 4.9: GST genes assessed and their respective statistical significance for each of the mixed models tested

F. hepatica GST Genes	Class	Model A						Model B					
		P-values	Effect Size	SE	dF	Upper CI	Lower CI	P-values	Effect Size	SE	dF	Upper CI	Lower CI
maker-scaffold10x_338_pilon-snap-gene-0.71	omega	0.01215 *	2.38 ‡‡‡‡‡	0.896	8.81	0.344	4.41	0.0005833 ***	2.71 ‡‡‡‡‡	0.598	8.88	1.36	4.07
maker-scaffold10x_713_pilon-augustus-gene-0.37		0.09203	-1.46 ‡‡‡‡	0.922	8.68	-3.56	0.633	0.3085	-1.4 ‡‡‡‡	0.6	8.97	-2.76	-0.0416
maker-scaffold10x_1428_pilon-snap-gene-0.7	Zeta	0.1498	-1.63 ‡‡‡‡	1.23	8.85	-4.42	1.17	6.033e-08 ***	-0.583 ‡‡	1.33	8.99	-3.6	2.44
maker-scaffold10x_80_pilon-snap-gene-0.169	mu	0.1196	1.52 ‡‡‡‡	1.06	9.26	-0.857	3.9	0.8374	1.06 ‡‡‡	0.642	9	-0.39	2.51
maker-scaffold10x_1043_pilon-snap-gene-0.18	Sigma	0.6077	0.598 ‡‡	1.35	8.81	-2.47	3.67	0.01131 *	0.971 ‡‡‡	0.809	8.95	-0.861	2.8
maker-scaffold10x_1184_pilon-snap-gene-0.31		0.06891	-2.06 ‡‡‡‡‡	1.18	9.32	4.71	0.588	0.02656 *	-1.9 ‡‡‡‡	1.45	9.02	-5.23	1.32
maker-scaffold10x_490_pilon-snap-gene-0.8		0.4057	1.11 ‡‡‡	1.52	8.99	-2.34	4.55	0.3491	1.22 ‡‡‡‡	1.14	9.01	-1.37	3.81
maker-scaffold10x_284_pilon-snap-gene-0.93		0.8566	-0.052	0.295	3.83	-0.887	0.783	0.8006	0.026	0.231	7.17	-0.517	0.569
maker-scaffold10x_2285_pilon-snap-gene-0.13		0.2445	-1.04 ‡‡‡	1	9.23	-3.3	1.22	0.06253	-1.08 ‡‡‡‡	1.33	9.02	-4.09	1.94
maker-scaffold10x_2285_pilon-snap-gene-0.12		0.3065	-0.993 ‡‡‡	1.1	8.85	-3.48	1.5	0.005469 **	-0.487	0.86	8.97	-2.43	1.46
maker-scaffold10x_490_pilon-snap-gene-0.5		0.02725 *	-2.73 ‡‡‡‡‡	1.22	8.91	-5.5	0.0378	0.03128 *	-2.11 ‡‡‡‡‡	0.938	8.97	-4.23	0.00972
augustus_masked-scaffold10x_1115_pilon-processed-gene-0.3		0.5122	-0.541 ‡‡	0.955	8.92	-2.7	1.62	0.0003415 ***	-1.26 ‡‡‡‡	0.691	8.96	-2.82	0.304
maker-scaffold10x_1189_pilon-snap-gene-0.107		0.1914	-1.47 ‡‡‡‡	1.24	8.67	-4.29	1.35	1.298e-10 ***	-2.31 ‡‡‡‡‡	0.721	8.93	-3.94	-0.676
maker-scaffold10x_938_pilon-snap-gene-0.52		0.3885	0.691 ‡‡	0.915	8.56	-1.39	2.78	2.074e-07 ***	1.64 ‡‡‡‡	0.653	8.91	0.158	3.12
maker-scaffold10x_284_pilon-augustus-gene-0.126		0.9284	-0.0833	1.09	8.96	-2.55	2.38	0.8721	-0.0783	1.08	8.99	-2.52	2.36
maker-scaffold10x_381_pilon-augustus-gene-0.37		0.7802	0.231 ‡	0.969	8.83	-1.97	2.43	0.4864	0.0729	0.806	8.97	-1.75	1.9
maker-scaffold10x_436_pilon-snap-gene-0.22		0.4874	-0.286 ‡	0.477	8.71	-1.37	0.798	0.8766	-0.208 ‡	0.34	8.87	-0.98	0.564

*** P < 0.001, ** P < 0.01, * P < 0.05, ND – No Data (No Evidence of Gene Expression). Effect Sizes in resistant versus susceptible Isolates (‡small effect, ‡‡medium effect, ‡‡‡large effect, ‡‡‡‡very large effect, ‡‡‡‡‡huge effect)

Table 4.10: cytochrome P450 (CYP450) and FABP genes assessed and their respective statistical significance for each of the mixed models tested

Gene Family	*F. hepatica* Genes	Model A						Model B					
		P-values	Effect Size	SE	dF	Upper CI	Lower CI	P-values	Effect Size	SE	dF	Upper CI	Lower CI
Cytochrome P450	maker-scaffold10x_1257_pilon-snap-gene-0.67	0.2895	-0.365 ‡	0.39	8.56	-1.25	0.524	0.5636	-0.225 ‡	0.307	8.98	-0.921	0.47
	maker-scaffold10x_135_pilon-augustus-gene-0.43	0.005773 **	-1.45 ‡‡‡‡	0.475	8.72	-2.53	-0.37	0.4828	-1.51 ‡‡‡‡	0.541	9.18	-2.73	-0.29
	maker-scaffold10x_135_pilon-augustus-gene-0.44	0.02659 *	-1.72 ‡‡‡‡	0.761	9.26	-3.44	-0.0116	0.0004358 ***	-0.733 ‡‡	0.659	9.07	-2.22	0.756
	maker-scaffold10x_114_pilon-augustus-gene-0.82	0.03032 *	3.36 ‡‡‡‡‡	1.54	8.85	-0.141	6.85	0.001371 **	1.5 ‡‡‡‡	1.16	8.99	-1.13	4.13
	maker-scaffold10x_2113_pilon-augustus-gene-0.4	0.1107	1.65 ‡‡‡‡	1.11	8.64	-0.874	4.17	2.2e-16 ***	-0.889 ‡‡‡	0.707	8.93	-2.49	0.711
	maker-scaffold10x_1546_pilon-snap-gene-0.20	0.1095	-1.35 ‡‡‡‡	0.901	8.39	-3.41	0.716	0.001286 **	-0.463 ‡	0.626	8.92	-1.88	0.954
Fatty Acid Binding Proteins	maker-scaffold10x_157_pilon-snap-gene-0.187	0.3709	-0.431 ‡	0.549	8.64	-1.68	0.818	6.183e-14 ***	1.28 ‡‡‡‡	0.358	8.88	0.469	2.09
	maker-scaffold10x_2403_pilon-snap-gene-0.11	0.9754	-0.0286	1.09	8.83	-2.5	2.44	0.003237 **	0.538 ‡‡	0.857	8.96	-1.4	2.48
	maker-scaffold10x_2403_pilon-snap-gene-0.19	0.258	0.568 ‡‡	0.562	8.53	-0.713	1.85	0.6758	0.638 ‡‡	0.525	8.85	-0.553	1.83
	maker-scaffold10x_2403_pilon-snap-gene-0.20	0.02252 *	1.71 ‡‡‡‡	0.734	8.81	0.0494	3.38	0.1366	1.76 ‡‡‡‡	0.412	8.75	0.822	2.69
	maker-scaffold10x_331_pilon-snap-gene-0.56	0.3361	-1.13 ‡‡‡	1.33	8.89	-4.15	1.89	0.0006166 ***	-0.485 ‡	1.06	8.98	-2.89	1.92
	maker-scaffold10x_331_pilon-snap-gene-0.57	0.2931	-1.35 ‡‡‡‡	1.45	8.95	-4.63	1.93	0.03877 *	-1.14 ‡‡‡	1.39	8.99	-4.29	2.01

*** P < 0.001, ** P < 0.01, * P < 0.05, ND – No Data (No Evidence of Gene Expression). Effect Sizes in resistant versus susceptible Isolates (‡small effect, ‡‡medium effect, ‡‡‡large effect, ‡‡‡‡very large effect, ‡‡‡‡‡huge effect)

4.7. Discussion

4.7.1. Differential Gene expression in Developmental Stages: Importance to Parasite Survival

Understanding the shift in gene activity across various stages of the *F. hepatica* life cycle is vital to understand the parasite stage to target in drug administration, especially when TCBZ is the only drug effective against immature and adult flukes. This could help explain why other benzimidazoles (such as albendazole) are effective against adult flukes and not against juvenile flukes (Mottier et al., 2006). The efficacy of TCBZ against juvenile and mature flukes could be linked to expression of the drug target genes in both stages of the parasite. While the inefficacy of other anthelmintics against juvenile flukes could be due to variations in gene expression pattern of the drug target genes of these anthelmintics in adult versus juvenile flukes. Gene families associated with TCBZ uptake (Fairweather et al., 2020) were noticed to be more transcriptionally active in metacercariae and day-old NEJs than in the adult. This could contribute to the increased effectiveness of TCBZ against NEJs. However, uptake of TCBZ has been shown to be reduced in TCBZ resistant flukes compared to susceptible ones (Alvarez et al., 2005). It is unclear if TCBZ acts by altering tubulin binding, drug uptake, or metabolism mechanisms in the parasite (Fairweather et al., 2020); understanding drug action in NEJs could improve understanding. In this study, differences among developmental in expression was noticed among gene families associated with TCBZ metabolism. Therefore, Identifying the impact of individual genes in each family and gene interaction pathways could explain their role in drug metabolism. One clear thing is that these TCBZ metabolism-related genes are active in all parasite life stages, but the specific role of each family member is unclear.

F. hepatica survival within the host, intermediate host, and the environment is underpinned by regulating gene expression levels of genes associated with various biological processes required for each parasite life cycle stage. Results in this study support findings in previous stage-specific gene expression studies, identifying elevated gene expression predominantly in metacercariae and the early NEJs. Elevated activity of various biological processes has been reported in metacercariae and NEJs, particularly those associated with fluke structure, migration, and invading host immunity (Cwiklinski et al., 2015). Elevated expression of tubulin in these immature flukes is understandable, considering their essential roles in maintaining the cell cytoskeleton (Stitt et al., 1992). Similarly, cysteine peptidases (predominantly cathepsins L and B) play crucial roles in metacercariae excystment, juvenile fluke feeding and migration through the host abdomen to reach the liver (McNulty et al., 2017b). Thus immature flukes are transcriptionally active, particularly in genes associated with the production of various proteins and signal transduction to establish infection in the hosts (Cwiklinski et al., 2021, Cwiklinski et al., 2018). While cathepsins are not investigated here as drug targets due to their generally poor annotation, they have been extensively studied as potential vaccine targets (Molina-Hernández et al., 2015, Dalton et al., 2003).

Interestingly, a marked elevated gene transcriptional expression of biological processes associated with snail intermediate host invasion, and increased expression of peptidases such as cathepsins in fluke in snail developmental stages has been noticed in *F. gigantica* eggs (Zhang et al., 2019), highlighting the importance of snail life cycle stage to parasite development. It is unconfirmed if observations would be consistently similar in *F. hepatica*, but it is likely the case. However, most gene families assessed in this study had moderate expression in snails and in *F. hepatica* eggs. Additional studies to compare expression profiles among *Fasciola* species would be fundamental to understanding how these could influence fluke development from eggs to matured flukes. Additionally, it might be essential to compare the stage-specific expression profile of these *Fasciola* species in response to TCBZ to understand further which parasite stage to target for effective control of parasites.

4.7.2. TCBZ Induces Expression in *F. hepatica* multiple Gene Families

In all the families assessed, a TCBZ induced response was observed in most family members, particularly in susceptible isolates. This indicates TCBZ affects multiple mechanisms in the parasite associated with altering tubulin related activities, drug uptake and metabolism. These findings corroborate previous observations suggesting the anthelmintics initiate complex effects on parasite metabolism, function and structure (Radio et al., 2018). While this study only focuses on TCBZ, it has not been determined if other anthelmintics produce similar complex effects on the parasite. This highlights the importance of exploring TCBZ and other anthelmintics' effects on *F. hepatica* and *F. gigantica* to identify which mechanisms have biological significance. Comparing the impact of TCBZ on resistant and susceptible *F. hepatica* isolates enabled an assessment of parasite responses in both groups. As expected, descriptive plots (Figure 4.9 & 4.10) revealed TCBZ resistant isolates (A, B, and C) were not responsive to the drug. While effect sizes in resistant versus susceptible isolates were mostly large in genes upon which TCBZ induced statistically significant activities, the drug also initiated small effects on other genes. This pattern was noticed across all the gene families evaluated (effect sizes are presented tables 4.8 – 4.10 for each gene).

In the resistant isolates, gene expression levels appeared to be similar in treated and untreated worms. This suggests that the drug had little effect on resistant flukes. Although this should be expected, this observation in genes associated with the three proposed mechanisms of action of TCBZ could all be important in TCBZ resistance. This could mean resistant isolates have mechanisms in place that either limits drug effects on tubulin structure or does not allow uptake of the drug and limit its metabolism. Assuming this is true, spread of drug-resistant *F. hepatica* isolates in the population could further increase incidences of treatment failure in farm. The increasing inability to treat herds due to drug resistance could economically impact livestock production.

Despite the findings in the drug resistant isolates, in the TCBZ susceptible isolates (N, S, and T), response to TCBZ was seen across gene families associated with the drug in *F. hepatica*. TCBZ induced interaction between TCBZ treatment and TCBZ resistance across multiple gene families assessed suggests TCBZ initiates multiple actions in susceptible isolates compared

to the resistant ones. Noticeable differential gene expression levels were noticed between resistant (referred to as the Sligo Strain) and susceptible (referred to as the Cullompton Strain) isolates, in which an increased expression of GST *mu* was noticed in TCBZ resistant isolate compared to the susceptible isolate (Scarcella et al., 2012). However, in this study, most GST genes assessed had reduced expression in the TCBZ resistant isolates compared to the susceptible ones (Figure 4.11). While only one GST *mu* (maker-scaffold10x_80_pilon-snap-gene-0.169) was identified here (*Predicted name: Mu Class glutathione transferase, UniProtKB/TrEMBL accession no : A0A890CT21*), findings suggest the action of TCBZ may vary among GST genes. In the GST *mu* identified, TCBZ had large effect on the gene, although the effect noticed was not statistically significant. Failure to identify more GST *mu* is unclear. This could be due to reduced presence in *Fasciola* or lack of representative gene models in other closely related species to use reference in *Fasciola*.

Similarly, adenylate cyclase gene expression has been shown to be reduced in the TCBZ resistant isolates (Radio et al., 2018); these findings were similar to findings in this study. However, expression levels between resistant and susceptible isolates in some AC genes did not vary. These differences could be attributed to various reasons, such as differences in *F. hepatica* gene expression concerning population differences or lack of detailed gene sequences, thus limiting family studies. While the latter is more logical, other reasons could be responsible. For instance, the AC genes reported previously (Radio et al., 2018) were mapped and indexed to the draft *F. hepatica* genome assembly (Cwiklinski et al., 2015). However, this assembly is no longer accessible on the WormBase Parasite database (ENA Project: PRJEB6687, WGS CCMX01000001-CCMX01195709, Scaffolds: LN627018-LN647175, Sample: SAMEA2629804). The lack of gene name consistency presents an important challenge to *F. hepatica* candidate gene study.

Apart from TCBZ inducing a response in these families explored, it was noticed that TCBZ also seemed to influence the expression of some of the constitutively active genes. While it is unclear if TCBZ is directly responsible for this expression or not. Constitutive gene expression can benefit environmental adaptiveness in bacteria (Geisel, 2011). However, in majority of the gene families studied (across the three proposed drug action mechanisms), there was a TCBZ induced gene expression in the susceptible isolates when compared with resistant ones (52% of the total genes assesses indicated statistically significant effect of the drug across all families assessed). These findings could suggest the drug dysregulates gene expression causing transcriptome-wide changes in the parasite, with the parasite deploying various mechanisms to get rid of the drug. Expression levels can be induced in response to stress factors (Davis and Moyle, 2020), such as drugs. It has been shown that following helminth infections, host tissue-derived cytokines levels are upregulated, initiating development of Type 2 immunity in host to the helminth (Hepworth et al., 2012). Previous studies on *Schistosoma mansoni*, a trematode, indicate that specific constitutively expressed genes play roles in suppressing host type 2 immune responses (Everts et al., 2016). *F. hepatica* has immunomodulatory ability (Dalton et al., 2013, Robinson et al., 2012, Ryan et

al., 2020, Corral-Ruiz and Sánchez-Torres, 2020); the parasite is a able to influence host immune response to infection. It is unclear what role constitutively expressed genes play in *F. hepatica* infection and how parasite respond to exposure drugs and other stressors. Thus, findings in this study indicate TCBZ administration affects multiple systems on the parasite, initiating the expression of various genes and increasing the expression of constitutively expressed genes. Therefore, confirming that all the three proposed TCBZ resistance mechanisms (Fairweather et al., 2020) could be important in *F. hepatica* infection. While it is unclear the relationship between the three mechanism, additional investigation is needed to identify the possibility of other unidentified mechanisms involved in TCBZ resistance. There may also be a need to investigate which mechanism of drug can be most exploited to control the parasite.

Chapter 5

General Discussion

5.1. Thesis Overview

F. hepatica is an economically important livestock zoonotic flatworm. The parasite, the predominant cause of fasciolosis, is well known for its global prevalence, devastating impact on livestock productivity, and complexities associated with its control in farms (Singh et al., 2021). A review of the parasite life cycle, farm management options available for disease treatment, lack of effective commercial vaccines, and, more importantly, the growing concerns relating to drug resistance has been presented (see chapter 1). Fasciolosis control has been an issue because of the lack of a clear understanding of the drug resistance mechanism, coupled with challenges associated with explaining triclabendazole (TCBZ) mode of action. *F. hepatica* is a genetically complex parasite capable of rapid adaptation to host and environment (Cwiklinski et al., 2015). The increasing spread of resistance to TCBZ, the only drug capable of killing all stages of the parasite, has necessitated the need to understand the drug mechanism and identify other potential drug targets in the parasite. It is known that benzimidazoles such as TCBZ alter parasite microtubule-based processes by binding to tubulins (Robinson et al., 2004); however, recent findings indicate other mechanisms are involved in TCBZ action and its resistance in liver flukes (Fairweather et al., 2020).

Given the growing concerns of TCBZ resistance, understanding the genetic basis of TCBZ resistance in *F. hepatica* is vital to developing new flukicides and vaccines. As at 2016, TCBZ resistance was reported in at least 30 farms worldwide (Kelley et al., 2016), although this number might be markedly higher due to poor reporting and unidentified resistant incidences on farm. In a study between Autumn 2013 to Spring 2015 in England and Wales, reduced efficacy of TCBZ was observed in at least 21 sheep farms out of the 26 farms investigated, while TCBZ had no therapeutic effect 6 of these farms (Kamaludeen et al., 2019). Another study in South-eastern Australia observed a 39 % prevalence of *F. hepatica* from 83 cattle herds, 3 of which had confirmed cases of TCBZ resistance (Kelley et al., 2020), while a study in Ireland observed a prevalence of 50.6 % from a total of 305 sheep flocks (Munita et al., 2019). In South America, a report of 25 % prevalence of TCBZ resistance (out of 462 liver flukes studied) observed in *F. hepatica* parasites from naturally infected livestock in Peru (Fernandez-Baca et al., 2022). TCBZ inefficacy was also observed in 11.6 % children (17 out of 146 children) with chronic *Fasciola* infections in Peru (Morales et al., 2021). The actual prevalence of TCBZ resistance is unknown, sometimes due to lack of reporting. For example, a survey of farmers in the UK suggest TCBZ is the most frequently used drug for liver fluke treatment, although most farmers seemed confused about their diagnosis and treatment (Hoyle et al., 2022). Thus, understanding the genetic basis of TCBZ could be critical in controlling the parasite.

5.1.1. *F. hepatica* Genome Re-annotation: Improving *F. hepatica* gene models

The publication of the first *F. hepatica* draft genome (Cwiklinski et al., 2015) has facilitated liver fluke research at a genomic scale. There are currently two published *F. hepatica* genomes (McNulty et al., 2017b), three *F. gigantica* genomes (Choi et al., 2020, Pandey et

al., 2020, Luo et al., 2021), and one *Fasciolopsis buski* genome (Choi et al., 2020). However, there are noticeable differences between each of these genomes (see Chapter 1, Table 1.3), especially in terms of gene content. While differences in genome assembly statistics could be due to various reasons such as varying genome size, sequencing platform used, genome interspersed repeats, etc., detailed identification and description of genomic features is key to inferring biological information (Jung et al., 2020). A comparative genomic study of major parasitic worms (a total of 14 platyhelminths were assessed) indicated that genome size varies from 104 to 1,259 Mb in platyhelminths while predicted genes ranged from 9,132–17,274 genes per species (International Helminth Genomes, 2019). This comparative study indicates the similar genome sizes in *Fasciola species* was not much different from other trematodes, although *Fasciola* genome is around the upper limit (~ 1.2 Gb) in terms of genome size when compared with other platyhelminths. It is thought that the differences in genomic sizes could be due to non-coding elements (International Helminth Genomes, 2019).

While non-coding elements could contribute to genome sizes, they also influence ability of genome annotation tools to effectively predict gene models. In this project, it was noticed that different repeat masking tools produced varying results, and strict genome repeats masking influenced genome annotation quality and number of genes predicted. It was also noticed that while reannotating the *F. hepatica* genome, evaluation of annotation completeness against universal single copy orthologs using BUSCO provided varying results (Chapter 2, See Table 2.3). For example, Transdecoder predicted 15,886 genes, the re-annotated version identified 15,879 genes, while the *Fasciola*_10x_pilon, GCA_900302435.1 annotation had 9,709 annotated genes; however, complete BUSCOs identified were 46.3%, 83.4%, and 75.7% respectively. While the reason for these varying BUSCO percentages for each annotation is unclear, it does appear that fragmentation and omission of exons in predicted gene models considerably reduces annotation completeness. Thus, these computational issues should be considered when annotating trematode genomes to ensure annotations reflect the actual gene number and features.

In reannotating of the updated *F. hepatica* draft genome (*Fasciola*_10x_pilon, GCA_900302435.1), predicted gene models were validated using tubulins, GST, and AC genes. The disparity in the number of genes identified, variations in gene models (such as insertion and omission of exons), gene fragmentation, etc., highlight the challenges of annotating a typically large and complex genome. Although there are tools that provide an overview of annotation statistics, manual verification and validation of predicted genes is encouraged. The lack of representative *F. hepatica* complete CDS was also a concern in this project; this limited gene model validation studies to some extent. Despite these challenges, the project provided additional insight into the *F. hepatica* genome and improved the ability to benchmark the current *F. hepatica* annotation available on WormBase Parasite (*Fasciola*_10x_pilon, GCA_900302435.1, WBPS15), thus, facilitating downstream analysis.

5.1.2. Gene Families Implicated in TCBZ mode of action and resistance

Studies exploring the biochemical mechanism of TCBZ action have suggested that multiple processes are involved. These proposed mechanisms (altered tubulin binding, altering drug uptake, efflux, and altered drug metabolism) are poorly understood, especially in TCBZ resistant isolates. Also, it is unclear if there is a combination of mechanisms at play in drug mode of action (Kelley et al., 2016). Concerning the three potential mechanisms of TCBZ action recently reviewed (Fairweather et al., 2020), gene families associated with each of these mechanisms compiled using the *F. hepatica* genome annotation (see Chapter 1, Table 1.3) were used for evolutionary and gene expression studies.

Although three mechanisms of actions been proposed in *F. hepatica* TCBZ action and resistance, it is unclear which mechanism plays a preeminent role in drug resistance. This challenge is similar in schistosomes, an extensively studied trematode. Praziquantel is the drug of choice against *Schistosoma*, however overreliance on the drug is leading to resistance (Gönnert and Andrews, 1977), while oxamniquine, a species-specific drug is effective against *S. mansoni* (Valentim et al., 2013). In schistosomes, multidrug resistance (MDR) has been proposed, in which case, a parasite is resistant to a single drug and resistant to structurally unrelated compounds (Greenberg, 2013). While it is unclear if MDR occurs in *Fasciola* considering TCBZ and albendazole, despite both been benzimidazoles, each drug provide different effects on liver flukes (Kouadio et al., 2021). However, liver flukes resistant to albendazole and Clorsulon (a benzenesulphonamide) have been reported (Martínez-Valladares et al., 2014), suggesting presence of MDR. Given these findings, it is possible that MDR occurs in *Fasciola* and *Schistosoma* and is likely present in other trematodes. A key discovery in *S. mansoni* research identified a transient receptor potential (TRP) channel as the target for praziquantel (Le Clec'h and Chevalier, 2021). Findings indicate praziquantel causes calcium influx and paralysis of these blood flukes. Interestingly, praziquantel does not produce similar effects in *F. hepatica*, and reasons remain unclear (Park and Friedrich, 2021). While the importance of the TRP gene in natural infections have not been ascertained (Cotton and Doyle, 2022), it is known that praziquantel causes vacuoles in the parasite tegument, ultimately causing permanent damages to the tegument (Park and Marchant, 2020). Similarly, TCBZ damages parasite tegument in *F. hepatica* and *F. gigantica* (Devine et al., 2011, Savage et al., 2013).

Understanding anthelmintic resistance is critical to understanding how to control parasites of interest. Identifying genes associated with drug resistance has been vital in predicting emergence and spread of resistance across parasite population (Fissiha and Kinde, 2021). Various gene targets are being identified in various parasites, however, understanding the mechanism of anthelmintic resistance is challenging. For example, resistance to ivermectin, a macrocyclic lactone has been associated with mutations in ligand-gated chloride channels in *Haemonchus contortus* especially the glutamate-gated channels (Kotze et al., 2014). Also, genes P-glycoprotein genes have been associated with ivermectin resistance in the parasite, while another research suggested presence of a multidrug resistance in the parasite.

Similarly, resistance to levamisole and pyrantel have been linked to L-type nicotinic acetylcholine receptors in *H. contortus* (Beech et al., 2011, Fissiha and Kinde, 2021).

Identifying TCBZ drug target could facilitate understanding resistance to the drug. In *S. mansoni* for example, resistance to Oxamniquine was associated to a sulfotransferase gene. Using the genome sequence of the parasite and a genetic map, the gene was identified, providing insights into specie-specific efficacy of the drug (Valentim et al., 2013). Also, in nematode – *H. contortus*, there was an upregulation of a pharyngeal-expressed transcription factor gene (HCON_00155390:cky-1) in ivermectin resistant populations (Laing et al., 2021). While this gene does not appear to have an orthologue in liver flukes, identification of the gene has been crucial to understanding resistance in the helminth. Presently there is no published report of a single gene responsible for TCBZ resistance, identifying regions of the *F. hepatica* genome responsible for drug resistance will be vital in understanding resistance. However, given the genetic diversity and lack of population structure in the parasite, understanding the mechanism of TCBZ could be very challenging (Beesley et al., 2017b, Hodgkinson et al., 2013).

5.1.3. Detecting Signals indicative of Positive Selection Pressure in Gene Families

Considering three proposed mechanisms have been implicated in TCBZ action, associated gene families were investigated as potential drug targets. Evolutionary forces acting on each candidate gene was assessed using site models in PAML (Yang and Nielsen, 2002, Yang, 1997) and MKtest (McDonald and Kreitman, 1991) to identify signals indicative of positive selection (see chapter 3). This is important considering the rapid evolutionary adaptivity of *F. hepatica* to the environment, host, drugs, etc. Thus, positive selection pressures on genes associated with resistance and potential drug targets could mean the rapid spread of resistant genes across the fluke population (Beesley et al., 2017b, Wolstenholme et al., 2004) and a continual need to develop new drugs. This study noted that gene families associated with altering drug uptake had sites exhibiting positive selection, predominantly in the RAS genes (see Chapter 3, Table 3.4). Similarly, signals indicative of positive selection pressure was noticed in some other genes associated with drug uptake (one ABC gene, one ADP ribosylation factor gene (ADP), and one AC gene) (see Chapter 3, Table 3.4). The presence of sites exhibiting positive selection could be of biological relevance, especially because uptake of TCBZ in liver flukes could be potentially influenced.

It was also noticed that MKtest was quite conservative in identifying positive selection in genes of interest as previously noticed (Eyre-Walker, 2002). Using this MKtest approach, statistically significant signals indicative of positive selection pressure was only detected in three genes (two RAS genes and one AC gene) (see Chapter 3, Table 3.5). Also, this work highlighted the computational challenges associated with assessing positive selection pressure in *F. hepatica* using a candidate gene approach. One thing is clear, the availability of well-annotated *F. hepatica* gene models improved the ability to evaluate adaptive evolutionary signals. Poor gene models resulting from missing exonic regions and gene fragmentations limit orthologous assignment and alignment of gene coding regions.

Furthermore, the need for detailed manual verification of gene codon-based alignments and the complexities of executing most available pipelines due to their use of various software dependency-related issues make assessing positive selection pressure at a genomic level challenging. While identifying positive selection at genomic level can be computationally challenging, in reported cases its use has mostly involved parallel analysis of single genes on a large (genomic) scale (Cole and Brewer, 2018, Hongo et al., 2015, Sahm et al., 2017). In practice, analysis of positive selection has been found effective predominantly in gene families such as expanded genes of biological interest, as was done in this study. For example, in the trematode *Atriophallophorus winterbourni* genomic study, out of the 11,499 annotated genes, only 2 orthologous groups of interest (with at least 20 gene copies in each orthologous group) were studied for signatures of positive selection using branch site models in codeml (Zajac et al., 2021). Also, in a *S. mansoni* genome sequencing project, identification signatures indicative of positive selection was noticed in gene families previously associated with praziquantel action in the parasite (Berger et al., 2021), and in some genes associated with host-parasite relationships (Crellen et al., 2016). Currently there no other reports identifying signatures of positive selection at genomic level in any trematode. Observations in this study provide a benchmark approach that can be used to explore evolutionary studies in *Fasciola* and other parasites on a candidate gene level, and potentially across the genome.

5.1.4. Expression profile of selected gene families in *F. hepatica* life stages

In other to further understand why TCBZ is effective against all stages of *Fasciola,* the expression levels of gene families associated with the three proposed mechanisms of TCBZ action and resistance across *F. hepatica* life stages was explored. Previous *F. hepatica* stage-specific expression profile studies indicate proteins associated with host-parasite interactions had elevated activities as parasites develop from infective stages into migratory juveniles in the host (Cwiklinski et al., 2015, McNulty et al., 2017b). These findings corroborate observations in this study in that elevated gene expression levels were observed predominantly in 24-hours old NEJs across most genes in the three proposed mechanisms, suggesting multiple activity of all the TCBZ action mechanisms. In *Schistosoma japonicum*, differential expression levels of genes associated with host immune response, parasite defence, and other key biological pathways were found across different life stages of the parasite (Gobert et al., 2009). Similarly, in the nematode parasite *Brugia malayi*, stage-specific studies revealed differential expression patterns in genes critical to parasite establishment of infection and survival (Li et al., 2012). Recent liver fluke findings indicate *F. hepatica* NEJs express unique miRNAs, while matured liver flukes also express unique miRNAs (Ricafrente et al., 2022), suggesting the parasite is able to regulate gene expression post-transcriptionally in different life stages. Given NEJs and adult liver flukes exhibit metabolic differences, for example use of oxygen in the host, parasite ability to regulate gene expression and adapt to host will be important (González-Miguel et al., 2021). Considering TCBZ is effective against all stages of the parasites, especially in juvenile flukes (Fairweather, 2005, Brennan et al., 2007), understanding drug activity in NEJs would be

critical to control measures. However, it is unclear if the elevated expression of these gene families associated with TCBZ mode of action in juvenile fluke is linked to the high efficacy of the drug in this life stage. While this is possible, there is a need to compare these findings with expression patterns in gene families not effective against early juvenile flukes such as albendazole, a benzimidazole like TCBZ (Love, 2017, Fairweather and Boray, 1999), as well as in *F. gigantica*.

5.1.5. TCBZ affects expression of multiple gene Families in *F. hepatica* Isolates

To assess the impact of TCBZ on *F. hepatica* liver flukes, expression profile of genes associated with the three proposed mechanisms of TCBZ action was assessed using RNAseq data. In this project linear mixed models was used to explore gene expression in response to TCBZ treatment instead of conventional differential expression tools such as EdgeR or DESeq2 (Robinson et al., 2009a, Love et al., 2014) In this project, a focused analyses on specific genes was done based on predefined hypothesis. Mixed models are robust in handling the "noise" in parasitological data such as grouping factors, data independence, randomness of effect (Paterson and Lello, 2003). Given our data structure has a mix of fixed effect (such as parasite status (resistant or susceptible) and treatment of sheep with TCBZ or not), and random effects such as (biological variations due to variation in sheep used for experiment and variation in clones (refer to Table 4.4)) we decided to use linear mixed models. While EdgeR and DESeq2 have power to detect differential gene expression at genomic level with an assumption that most genes are not responding to TCBZ treatment, here we are interested in a subset gene of interest we think are responding to treatment. However, despite the advantages of these models, they can be complex and subject to errors, if model formulas are wrong. Wrong formulas especially from poor data structure or understanding of biological questions can make analysis prone to errors.

RNAseq data from flukes extracted from the sheep experimentally infected with *F. hepatica* metacercariae of drug-resistant and susceptible isolates (Hodgkinson et al., 2018) was analysed using lmer in R. Findings indicate TCBZ initiates a response to the drug in multiple genes in each family across the three proposed mechanisms of drug action. These observations change the approach to research TCBZ's mode of action. Previous studies have implicated the role of tubulins, while recent studies have explored alteration of drug uptake and drug metabolism as possible mechanisms (Brennan et al., 2007). It was unclear if a single or combined mechanism was involved in TCBZ biochemical activity and its resistance (Fairweather, 2005, Kelley et al., 2016, Fairweather et al., 2020). However, the results in this study indicate all three proposed mechanisms are affected by the drug. This is in line with previous report that anthelmintic drugs initiate complex effects on liver fluke (Radio et al., 2018). Thus, suggesting the possibility of other yet to be identified mechanisms associated to TCBZ resistance. Although this study focused on *F. hepatica*, it is likely similar patterns will be noticed in *F. gigantica*; however, this would need to be confirmed. Future studies exploring the effect on TCBZ in other *Fasciola* species would aid understanding if findings noticed here are specific to *F. hepatica* or not.

5.1.6. Is there a link between TCBZ induced gene expression and Selective Pressure?

The findings here indicate that eight *F. hepatica* genes (Table 5.1) were identified to have statistically significant sites under positive selection by the branch-site models and had drug-induced expression levels due to administration of TCBZ in the drug-susceptible isolates compared to the resistant ones. These genes (4 Ras genes, 1 ADP ribosylation factor gene, and 1 AC gene), could be significant in understanding how the parasite responds with TCBZ, although it is yet unclear how. Thus, this leads to the question, does positive selection pressure on *F. hepatica* genes influence the expression of the gene? While there is no straightforward answer, a previous report suggest gene regulatory mechanisms could play a crucial role in controlling gene expression profiles (Jovanovic et al., 2021). This idea proposes that an alteration in gene regulation is critical to an organism's adaptation ability (Romero et al., 2012). Selection pressure in regulatory genes influence gene expression, phenotypic variations, and disease susceptibility in humans (Blekhman et al., 2008). However, attributing adaptive variations within and between species to differential gene expression levels and phenotypic patterns is challenging due to various genetic and non-genetic confounding factors (Gilad et al., 2006).

Interestingly, a study on 7 *Drosophila* species orthologous genes noted that selection pressure influenced gene expression to a lesser extent, while stabilising selection noticeably reduces expression levels (Bedford and Hartl, 2009). Gene expression and positive selection studies revealed that upregulated genes in birds in response to pathogens exhibited signals indicative of positive selection (Shultz and Sackton, 2019). Moreover, some enriched genes associated with immunity and metabolic syndromes were under positive selection in human populations (Vatsiou et al., 2016). Despite identifying these eight genes with induced expression in response to TCBZ treatment and with sites under positive selection in this study, these findings were not consistent across all other genes identified to be induced by the drug. While adaptive evolutionary forces acting on *F. hepatica* could play critical roles in influencing gene expression levels, altering gene function, and binding to the drug target, it has been shown here that TCBZ induce differential expression of multiple genes in the parasite. Thus, an adaptive selective pressure in parasite genes could be an evolutionary mechanism adopted by the parasite in response to anthelmintics such as TCBZ, leading to drug resistance.

5.1.7. Exploring *F. hepatica* genes for potential further studies

The six genes of potential interest in the parasite were explored using the information from various databases. These are genes with sites under positive selection (Chapter 3) and had variations in expression levels in response to TCBZ treatment (Chapter 4). While these genes are expressed in adult liver flukes, RAS3 was the most highly expressed (Figure 5.1). The orthologous copies of RAS1, RAS4, and the ADP ribosylation factor genes have been associated with various conditions in humans (Table 5.1). Given that Rad and Gem GTP binding proteins play a crucial role in the regulation of cytoskeletal proteins (Ward et al., 2002), it is unclear if these small RAS genes influence tubulin activity. Studies indicate that a

specific Ras-related -protein (known as kir/Gem) can regulate calcium ion channels by interacting with β-subunits (Béguin et al., 2001). It is interesting to know that no evidence of positive selection was observed in liver fluke tubulins, although TCBZ influenced tubulin activity in this study. It is known that TCBZ can inhibit GTP-RAS and AC expression in yeast (Lee et al., 2013); however, it is unclear if this is similar in liver flukes. Downregulation of a β – tubulin and a Ras protein has been noticed in acute *F. hepatica* infection in cattle in response to liver fluke infection (Garcia-Campos et al., 2019).

The importance of tubulins (especially α - and β – tubulins) in helminths have been well documented. In liver flukes, tubulins genes play a vital role in the microtubule process in parasite tegument, and various tubulins have been identified in liver flukes (Ryan et al., 2008). However, there is little literature on delta, epsilon, and gamma tubulins in *F. hepatica*. It is thought that delta and epsilon tubulins have similar roles to alpha and beta tubulins, although studies suggest the act within and outside the centriole, as well as playing a critical role in the development of mammals (Stathatos et al., 2021, Chang and Stearns, 2000). Here, it was shown that TCBZ influenced the expression of delta tubulin - maker-scaffold10x_1084_pilon-snap-gene-0.149 and Epsilon tubulin - maker-scaffold10x_500_pilon-snap-gene-0.52, in resistant vs susceptible isolates with huge effect sizes. While there was no statistically significant effect in Gamma tubulin - maker-scaffold10x_1160_pilon-snap-gene-0.20, small effect sizes were noticed (see chapter 4, table 4.5). These observations could have similar effects on alpha, beta, delta, and epsilon tubulins, although the effect on gamma tubulins could be minimal. Further investigation is required to understand how the various tubulins, RAS genes, and AC genes interact in the parasite tegument in response to TCBZ.

Table 5.1: *Fasciola hepatica* genes with statistically significant sites under positive selection pressure using PAML using branch-site models and with induced expression in response to TCBZ in susceptible compared to resistant isolates

Gene family	Identifier	*Fasciola hepatica* Gene	Description	UniProt (Accession)	Predicted Function	Expression in adult flukes	*Gene length (amino acids)*	*Number of exons*	*Pfam*	InterPro Predicted Domains	*Additional comments*
RAS	RAS1	maker-scaffold10x_1568_pilon-snap-gene-0.0	Rad and Gem GTP binding protein 1	A0A4E0RHY2	GTPase activity (GO:0003924) GTP binding (GO:0005525)	Yes	560	5	Ras - PF00071.25	PS51419: small GTPase Rab1 family profile. SM00173: ras_sub_4 SM00175: rab_sub_5	C. elegans orthologue WormBase - Y52B11A.4 predicted to be in plasma membrane - Human ortholog of this gene implicated in congestive heart failure and type 2 diabetes mellitus
	RAS2	snap_masked-scaffold10x_2595_pilon-processed-gene-0.0	Ras protein Rab-5C	A0A4E0RPV3	GTPase activity (GO:0003924) GTP binding (GO:0005525)	Yes	58	1	Ras - PF00071.25 Ras of Complex, Roc, domain of DAPkinase - PF08477.16	SM00175: rab_sub_5 PS51419: small GTPase Rab1 family profile.	
	RAS3	maker-scaffold10x_317_pilon-snap-gene-0.52	Ras-related protein Rab-6A (Predicted	ND	GTPase activity (GO:0003924) GTP binding (GO:0005525)	Yes	308	8	Ras - PF00071.25	PS51419: small GTPase Rab1 family profile. cd01861: Rab6 SM00175: rab_sub_5 SM00173: ras_sub_4 SM00176: ran_sub_2 PS51420: small GTPase Rho family profile. PR00449: Transforming protein P21 ras signature	
	RAS4	snap_masked-scaffold10x_418_pilon-processed-gene-0.133	A0A4E0RV87	RAP (Vertebrate Rap GTPase family)	GTPase activity (GO:0003924) GTP binding (GO:0005525) signal transduction (GO:0007165) membrane roles (GO:0016020)	Yes	166	3	Ras - PF00071.25	PR00449: Transforming protein P21 ras signature SM00173: ras_sub_4 PS51419: small GTPase Rab1 family profile. SM00175: rab_sub_5	C. elegans orthologue WormBase - rap-2 predicted to enable GDP binding activity, located in the Plasma membrane, involved in collagen and cuticle development.
Adenylate Cyclase		maker-scaffold10x_609_pilon-snap-gene-0.9	A0A2H1CJ50 A0A4E0RFU7	Adenylate/guanylate cyclase catalytic domain protein	lyase activity (GO:0016829) cyclic nucleotide biosynthetic process (GO:0009190) intracellular signal transduction (GO:0035556) integral component of membrane (GO:0016021) membrane roles (GO:0016020)	Yes	680	2	Adenylyl cyclase class-3/4/guanylyl cyclase - PF00211	SM00044: cyc_6 cd07302: CHD Dimer interface Metal binding site Nucleotidyl binding site PS50125: Guanylate cyclase domain profile. PF00211: Adenylate and Guanylate cyclase catalytic domain	
ADP ribosylation factor		snap_masked-scaffold10x_492_pilon-processed-gene-0.0			GTPase activity (GO:0003924) GTP binding (GO:0005525)	Yes	283	4	Small GTPase superfamily, ARF/SAR type	cd00878: Arf_Ar SM00177: arf_sub_2 PS51417: small GTPase Arf family profile	C. elegans orthologue WormBase - warf-1 (Worm ADP-Ribosylation Factor homolog) predicted to facilitate GTP binding activity, located in plasma membrane, human ortholog implicated in a brain malformation of cortical development

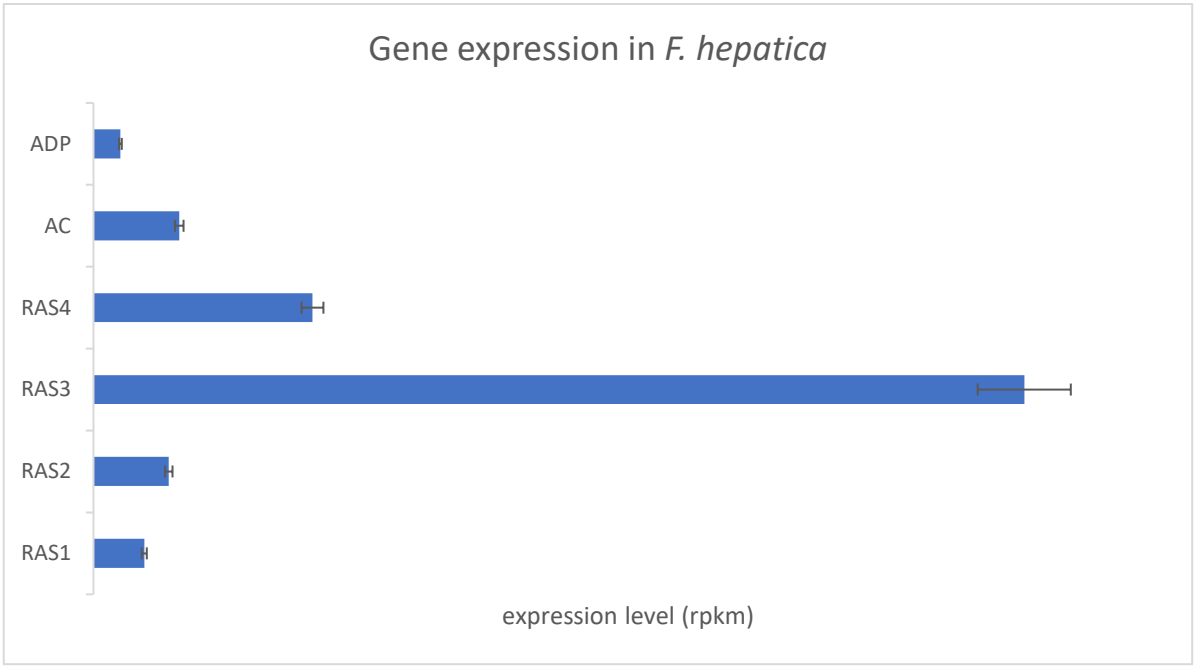

Figure 5.1: Expression profile of the six *F. hepatica* genes of further interest in adult liver flukes. Plot indicates all the five genes are expressed in adult liver flukes, although RAS3 (see table 5.1 for full gene name) was the most highly expressed gene.

5.1.8. The implication of these findings for *F. hepatica* research

The observations reported in this thesis has contributed to understanding *F. hepatica* biology, highlighting the potential impact of positive selection pressure of the fluke and how this impacts drug development. *F. hepatica* response to TCBZ - the drug of choice - was discussed and importance of improving the *F. hepatica* genome annotation was emphasised. The importance of good gene models in understanding biological processes in the parasite was also highlighted. To fully unlock the biological information in the *F. hepatica* genome, a continuous improvement of the genome is vital (Jung et al., 2020). Due to the complexity associated with large genome size (such as in the *F. hepatica* genome), computational requirements can limit re-annotation, thus, underscoring the need for an entirely automated and easy to execute annotation pipeline. However, it was noticed in this project that re-annotation of the genome and validating gene models provided a reliable benchmarking tool in downstream analysis. Thus, enabling understanding of biological studies from "reliable" gene models.

The evolutionary forces shaping *F. hepatica* adaptation provided information on the *F. hepatica* genes associated with TCBZ action. This is crucial to *F. hepatica* research mainly because recent studies have focused on explaining mechanisms of TCBZ action and resistance in *F. hepatica* (Fairweather et al., 2020, Kelley et al., 2016, Brennan et al., 2007) to develop new drugs and vaccines (Toet et al., 2014). Identifying signals indicative of positive selection in genes associated with TCBZ mode of action, predominantly genes associated with altering parasite TCBZ uptake (RAS, ABC, AC, and ADP), points out the need to investigate these mechanisms further. These findings suggest drug uptake by the parasite could be changed gradually as an adaptive measure, especially if sites under positive selection influence protein binding sites. Additional studies are needed to pinpoint TCBZ targets in the *F. hepatica* and identify if there's an evolutionary pressure on the targets, while highlighting the need to consider the effects of positive selection pressure with regards to the development of resistance via the drug target in the parasite in response to the drug (particularly in new drug targets). Furthermore, there is a need to further explore the proposed TCBZ uptake mechanism (Fairweather et al., 2020) from an evolutionary perspective, as this could explain the spread of TCBZ resistance across *F. hepatica* populations.

The studies in this project indicated that TCBZ induced the gene expression of multiple genes in the three proposed mechanisms of TCBZ action in the drug-susceptible isolates with respect to the resistant ones. While the order of importance of either or a combination of these proposed mechanisms of action is unclear (Kelley et al., 2016), findings here support a previous idea that *F. hepatica* anthelmintic resistance is possibly multigenic (Molina-Hernández et al., 2015). If indeed TCBZ impact is multigenic, then detailed identification of all associated *F. hepatica* gene families and their functional interactions would be essential to understand TCBZ mode of action. The drug induces the expression of multiple genes in response to its administration, suggesting the parasite could possibly use

either mechanism to get rid of the drug. It could therefore be vital to explore and consider parasite response to drug absorption, delivery, metabolism, and elimination into design of new anthelmintics.

5.2. Recommendations and Potential future studies

The availability of *F. hepatica* genomes (Cwiklinski et al., 2015, McNulty et al., 2017b) has substantially enabled the scope of *F. hepatica* studies at a genomic level. Also, the availability of *F. gigantica* (Choi et al., 2020, Pandey et al., 2020, Luo et al., 2021) and other closely related trematodes have facilitated comparative studies using orthologous gene relationships. However, despite this wealth of genomic data and annotations, a detailed description of every gene of these genomes is lacking. The availability of a detailed description of various gene networks and a clear understanding of their functionalities would facilitate an understanding of parasite biological processes.

Therefore, there is a need for a clear understanding of the various gene pathways involved with drug response. Although in this study, the genes associated with the three proposed mechanisms of action of TCBZ were compiled and assessed for evidence of positive selection pressure and response to treatment with TCBZ in resistant and susceptible isolates of *F. hepatica*, the scope of the study is limited to a detailed compilation of members of each family. Thus, detailed annotation and description of every *F. hepatica* gene with respect to their families and specific function will have greatly benefitted these studies. While the scope of these studies was limited to the three proposed mechanisms, a genomic level study of positive selection pressure and expression levels response to TCBZ would provide a broader understanding of *F. hepatica* adaptive and biological processes. Therefore, a robust automated pipeline would be essential to achieve these genomic scale evolutionary and biological studies in *F. hepatica* and *F. gigantica*. On a general note, the methods described here can be applied to other available trematode genomes, this could provide a broader understanding of drug resistance in parasites generally. Exploring drug resistance in other parasites could aid developing more effective helminth control measures.

It is worth mentioning that understanding the gene regulatory mechanism in liver flukes could enable understanding of how the parasite controls various systems responsible for survival in the intermediate and definite hosts. This study has already shown that differential expression levels of *F. hepatica* genes occur across the different life stages of the parasite. Other stage-specific gene expression studies also indicate that these varying levels across parasite life stages are essential for invading host immunity survival inside and outside the host (McNulty et al., 2017b, Cwiklinski et al., 2015). Despite these findings, there is little information on the regulatory mechanisms in parasites and how these could drive parasite adaptation and survival. Understanding gene regulatory mechanism in parasites, could be critical in understanding parasite biology and development of new drug antiparasitic drugs.

5.3. Concluding Remarks

Currently, the mode of action of TCBZ is unclear despite the various proposed mechanisms. Unfortunately, the spread of resistance to the drug across farms worldwide is a growing concern (Kelley et al., 2016). In this study, sites under positive selection pressure were identified in *F. hepatica* genes in the three proposed mechanisms of TCBZ action (Fairweather et al., 2020), predominantly in those associated with altering the drug uptake. These findings are key considering a positive selection signal, especially in the binding region of *F. hepatica* genes that are drug targets could influence drug efficacy. There is a need to evaluate adaptive evolutionary forces shaping *F. hepatica* biology across the parasite genome for better understanding of drug resistance.

This study also revealed that TCBZ induced the expression of multiple *F. hepatica* genes in the drug's three proposed mechanisms of action in *F. hepatica* TCBZ susceptible isolates but not the resistant ones. Although these observations indicate all three mechanisms play a role in how TCBZ acts, it does not reveal the order of importance of each mechanism to the fluke. There is a need to explore which mechanism is more crucial to the parasite, to ascertain which mechanism to prioritize in drug resistance research. There may also be a need to identify other likely mechanisms associated with the drug mechanism, a detailed description of associated genes and their functional role. Thus, additional studies are needed to understand the effect levels of each of the mechanisms to determine which mechanism can be targeted to understand drug resistance.

References

ALVAREZ, L. I., SOLANA, H. D., MOTTIER, M. L., VIRKEL, G. L., FAIRWEATHER, I. & LANUSSE, C. E. 2005. Altered drug influx/efflux and enhanced metabolic activity in triclabendazole-resistant liver flukes. *Parasitology,* 131**,** 501-510.

ALVAREZ ROJAS, C. A., JEX, A. R., GASSER, R. B. & SCHEERLINCK, J.-P. Y. 2014. Chapter Two - Techniques for the Diagnosis of Fasciola Infections in Animals: Room for Improvement. *In:* ROLLINSON, D. & STOTHARD, J. R. (eds.) *Advances in Parasitology.* Academic Press.

AMER, S., DAR, Y., ICHIKAWA, M., FUKUDA, Y., TADA, C., ITAGAKI, T. & NAKAI, Y. 2011. Identification of Fasciola species isolated from Egypt based on sequence analysis of genomic (ITS1 and ITS2) and mitochondrial (NDI and COI) gene markers. *Parasitology international,* 60**,** 5-12.

ANDOLFATTO, P. 2005. Adaptive evolution of non-coding DNA in Drosophila. *Nature,* 437**,** 1149-1152.

BABJÁK, M., KÖNIGOVÁ, A., BURCÁKOVÁ, Ľ., KOMÁROMYOVÁ, M., DOLINSKÁ, M. U. & VÁRADY, M. 2021. Assessing the Efficacy of Albendazole against Fasciola hepatica in Naturally Infected Cattle by In Vivo and In Vitro Methods. *Veterinary Sciences,* 8**,** 249.

BAO, W., KOJIMA, K. K. & KOHANY, O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile Dna,* 6**,** 1-6.

BATES, D., MÄCHLER, M., BOLKER, B. & WALKER, S. 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

BAUM, D. 2008. Reading a phylogenetic tree: the meaning of monophyletic groups. *Nature Education,* 1**,** 190.

BEDFORD, T. & HARTL, D. L. 2009. Optimization of gene expression by natural selection. *Proceedings of the National Academy of Sciences,* 106**,** 1133-1138.

BEECH, R., SKUCE, P., BARTLEY, D., MARTIN, R., PRICHARD, R. & GILLEARD, J. 2011. Anthelmintic resistance: markers for resistance, or susceptibility? *Parasitology,* 138**,** 160-174.

BEESLEY, N., CAMINADE, C., CHARLIER, J., FLYNN, R., HODGKINSON, J., MARTINEZ-MORENO, A., MARTINEZ-VALLADARES, M., PEREZ, J., RINALDI, L. & WILLIAMS, D. 2017a. Fasciola and fasciolosis in ruminants in Europe: Identifying research needs. *Transboundary and emerging diseases*.

BEESLEY, N., CWIKLINSKI, K., WILLIAMS, D. & HODGKINSON, J. 2015. Fasciola hepatica from naturally infected sheep and cattle in Great Britain are diploid. *Parasitology,* 142**,** 1196-1201.

BEESLEY, N. J., WILLIAMS, D. J., PATERSON, S. & HODGKINSON, J. 2017b. Fasciola hepatica demonstrates high levels of genetic diversity, a lack of population structure and high gene flow: possible implications for drug resistance. *International journal for parasitology,* 47**,** 11-20.

BÉGUIN, P., NAGASHIMA, K., GONOI, T., SHIBASAKI, T., TAKAHASHI, K., KASHIMA, Y., OZAKI, N., GEERING, K., IWANAGA, T. & SEINO, S. 2001. Regulation of Ca2+ channel expression at the cell surface by the small G-protein kir/Gem. *Nature,* 411**,** 701-706.

BEHZAD, C., LAHMI, F., IRANSHAHI, M. & MOHAMMAD ALIZADEH, A. H. 2014. Finding of Biliary Fascioliasis by Endoscopic Ultrasonography in a Patient with Eosinophilic Liver Abscess. *Case Reports in Gastroenterology,* 8**,** 310-318.

BERGER, D. J., CRELLEN, T., LAMBERTON, P. H. L., ALLAN, F., TRACEY, A., NOONAN, J. D., KABATEREINE, N. B., TUKAHEBWA, E. M., ADRIKO, M., HOLROYD, N., WEBSTER, J. P., BERRIMAN, M. & COTTON, J. A. 2021. Whole-genome sequencing of Schistosoma mansoni reveals extensive diversity with limited selection despite mass drug administration. *Nature Communications,* 12**,** 4776.

BERRIMAN, M., HAAS, B. J., LOVERDE, P. T., WILSON, R. A., DILLON, G. P., CERQUEIRA, G. C., MASHIYAMA, S. T., AL-LAZIKANI, B., ANDRADE, L. F., ASHTON, P. D., ASLETT, M. A., BARTHOLOMEU, D. C., BLANDIN, G., CAFFREY, C. R., COGHLAN, A., COULSON, R., DAY, T. A., DELCHER, A., DEMARCO, R., DJIKENG, A., EYRE, T., GAMBLE, J. A., GHEDIN, E., GU, Y., HERTZ-FOWLER, C., HIRAI, H., HIRAI, Y., HOUSTON, R., IVENS, A., JOHNSTON, D. A., LACERDA, D.,

MACEDO, C. D., MCVEIGH, P., NING, Z., OLIVEIRA, G., OVERINGTON, J. P., PARKHILL, J., PERTEA, M., PIERCE, R. J., PROTASIO, A. V., QUAIL, M. A., RAJANDREAM, M.-A., ROGERS, J., SAJID, M., SALZBERG, S. L., STANKE, M., TIVEY, A. R., WHITE, O., WILLIAMS, D. L., WORTMAN, J., WU, W., ZAMANIAN, M., ZERLOTINI, A., FRASER-LIGGETT, C. M., BARRELL, B. G. & EL-SAYED, N. M. 2009. The genome of the blood fluke Schistosoma mansoni. *Nature,* 460**,** 352-358.

BESEMER, J., LOMSADZE, A. & BORODOVSKY, M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic acids research,* 29**,** 2607-2618.

BLEKHMAN, R., OSHLACK, A., CHABOT, A. E., SMYTH, G. K. & GILAD, Y. 2008. Gene regulation in primates evolves under tissue-specific selection pressures. *PLoS genetics,* 4**,** e1000271.

BORAY, J., CROWFOOT, P., STRONG, M., ALLISON, J., SCHELLENBAUM, M., VON ORELLI, M. & SARASIN, G. 1983. Treatment of immature and mature Fasciola hepatica infections in sheep with triclabendazole. *The Veterinary Record,* 113**,** 315-317.

BORGSTEEDE, F. H. M. 2011. Diseases of Dairy Animals | Parasites, Internal: Liver Flukes. *In:* FUQUAY, J. W. (ed.) *Encyclopedia of Dairy Sciences (Second Edition).* San Diego: Academic Press.

BRENNAN, G., FAIRWEATHER, I., TRUDGETT, A., HOEY, E., MCCONVILLE, M., MEANEY, M., ROBINSON, M., MCFERRAN, N., RYAN, L. & LANUSSE, C. 2007. Understanding triclabendazole resistance. *Experimental and molecular pathology,* 82**,** 104-109.

BROWN, T. A. 2002. Genome anatomies. *Genomes. 2nd edition.* Wiley-Liss.

BUDDENBORG, S. K., TRACEY, A., BERGER, D. J., LU, Z., DOYLE, S. R., FU, B., YANG, F., REID, A. J., RODGERS, F. H. & RINALDI, G. 2021. Assembled chromosomes of the blood fluke Schistosoma mansoni provide insight into the evolution of its ZW sex-determination system. *bioRxiv*.

BUNGIRO, R. D. & CAPPELLO, M. 2004. Helminth Infections. *In:* JOHNSON, L. R. (ed.) *Encyclopedia of Gastroenterology.* New York: Elsevier.

BURGE, C. & KARLIN, S. 1997. Prediction of complete gene structures in human genomic DNA11Edited by F. E. Cohen. *Journal of Molecular Biology,* 268**,** 78-94.

BYRNE, A. W., GRAHAM, J., MCCONVILLE, J., MILNE, G., GUELBENZU-GONZALO, M. & MCDOWELL, S. 2019. Liver fluke (Fasciola hepatica) co-infection with bovine tuberculosis in cattle: A prospective herd-level assessment of herd bTB risk in dairy enterprises. *Transboundary and Emerging Diseases,* 66**,** 1727-1736.

C. ELEGANS SEQUENCING CONSORTIUM, C. 1998. Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science,* 282**,** 2012-8.

CAIRA, J. N. & LITTLEWOOD, D. T. J. 2013. Worms, Platyhelminthes. *In:* LEVIN, S. A. (ed.) *Encyclopedia of Biodiversity (Second Edition).* Waltham: Academic Press.

CAMPBELL, M. S., LAW, M., HOLT, C., STEIN, J. C., MOGHE, G. D., HUFNAGEL, D. E., LEI, J., ACHAWANANTAKUN, R., JIAO, D., LAWRENCE, C. J., WARE, D., SHIU, S.-H., CHILDS, K. L., SUN, Y., JIANG, N. & YANDELL, M. 2013. MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations *Plant Physiology,* 164**,** 513-524.

CARAVEDO, M. A. & CABADA, M. M. 2020. Human Fascioliasis: Current Epidemiological Status and Strategies for Diagnosis, Treatment, and Control. *Research and reports in tropical medicine,* 11**,** 149.

CASANUEVA, P., HILLYER, G., RAMAJO, V., OLEAGA, A., ESPINOZA, E. & MURO, A. 2001. Immunoprophylaxis against Fasciola hepatica in rabbits using a recombinant Fh15 fatty acid-binding protein. *Journal of Parasitology,* 87**,** 697-700.

CDC. 2018. *parasites* [Online]. Global Health, Division of Parasitic Diseases and Malaria. Available: https://www.cdc.gov/parasites/fasciola/biology.html.

CEBALLOS, L., CANTON, C., PRUZZO, C., SANABRIA, R., MORENO, L., SANCHIS, J., SUAREZ, G., ORTIZ, P., FAIRWEATHER, I. & LANUSSE, C. 2019. The egg hatch test: A useful tool for albendazole resistance diagnosis in Fasciola hepatica. *Veterinary parasitology,* 271**,** 7-13.

CHANG, P. & STEARNS, T. 2000. δ-Tubulin and ε-tubulin: two new human centrosomal tubulins reveal new aspects of centrosome structure and function. *Nature Cell Biology,* 2**,** 30-35.

CHARLIER, J., VERCRUYSSE, J., MORGAN, E., VAN DIJK, J. & WILLIAMS, D. J. L. 2013. Recent advances in the diagnosis, impact on production and prediction of Fasciola hepatica in cattle. *Parasitology,* 141**,** 326-335.

CHEMALE, G., MORPHEW, R., MOXON, J. V., MORASSUTI, A. L., LACOURSE, E. J., BARRETT, J., JOHNSTON, D. A. & BROPHY, P. M. 2006. Proteomic analysis of glutathione transferases from the liver fluke parasite, Fasciola hepatica. *Proteomics,* 6**,** 6263-6273.

CHEMALE, G., PERALLY, S., LACOURSE, E. J., PRESCOTT, M. C., JONES, L. M., WARD, D., MEANEY, M., HOEY, E., BRENNAN, G. P., FAIRWEATHER, I., TRUDGETT, A. & BROPHY, P. M. 2010. Comparative Proteomic Analysis of Triclabendazole Response in the Liver Fluke Fasciola hepatica. *Journal of Proteome Research,* 9**,** 4940-4951.

CHO, H., DAVIS, J., LI, X., SMITH, K. S., BATTLE, A. & MONTGOMERY, S. B. 2014. High-resolution transcriptome analysis with long-read RNA sequencing. *PloS one,* 9**,** e108095.

CHOI, Y.-J., FONTENLA, S., FISCHER, P. U., LE, T. H., COSTÁBILE, A., BLAIR, D., BRINDLEY, P. J., TORT, J. F., CABADA, M. M. & MITREVA, M. 2020. Adaptive Radiation of the Flukes of the Family Fasciolidae Inferred from Genome-Wide Comparisons of Key Species. *Molecular biology and evolution,* 37**,** 84-99.

COLE, T. J. & BREWER, M. S. 2018. FUSTr: a tool to find gene Families Under Selection in Transcriptomes. *PeerJ,* 6**,** e4234.

COLLINS, J. J. 2017. Platyhelminthes. *Current Biology,* 27**,** R252-R256.

CORRAL-RUIZ, G. M. & SÁNCHEZ-TORRES, L. E. 2020. Fasciola hepatica-derived molecules as potential immunomodulators. *Acta Tropica,* 210**,** 105548.

COTTON, J. A. & DOYLE, S. R. 2022. A genetic TRP down the channel to praziquantel resistance. *Trends in Parasitology,* 38**,** 351-352.

COX, M. P., PETERSON, D. A. & BIGGS, P. J. 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC bioinformatics,* 11**,** 1-6.

CRELLEN, T., ALLAN, F., DAVID, S., DURRANT, C., HUCKVALE, T., HOLROYD, N., EMERY, A. M., ROLLINSON, D., AANENSEN, D. M. & BERRIMAN, M. 2016. Whole genome resequencing of the human parasite Schistosoma mansoni reveals population history and effects of selection. *Scientific reports,* 6**,** 1-13.

CRINGOLI, G., MAURELLI, M. P., LEVECKE, B., BOSCO, A., VERCRUYSSE, J., UTZINGER, J. & RINALDI, L. 2017. The Mini-FLOTAC technique for the diagnosis of helminth and protozoan infections in humans and animals. *Nat Protoc,* 12**,** 1723-1732.

CRINGOLI, G., RINALDI, L., MAURELLI, M. P. & UTZINGER, J. 2010. FLOTAC: new multivalent techniques for qualitative and quantitative copromicroscopic diagnosis of parasites in animals and humans. *Nature Protocols,* 5**,** 503-515.

CWIKLINSKI, K., DALTON, J. P., DUFRESNE, P. J., LA COURSE, J., WILLIAMS, D. J., HODGKINSON, J. & PATERSON, S. 2015. The Fasciola hepatica genome: gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution. *Genome biology,* 16**,** 71.

CWIKLINSKI, K., JEWHURST, H., MCVEIGH, P., BARBOUR, T., MAULE, A. G., TORT, J., O'NEILL, S. M., ROBINSON, M. W., DONNELLY, S. & DALTON, J. P. 2018. Infection by the helminth parasite Fasciola hepatica requires rapid regulation of metabolic, virulence, and invasive factors to adjust to its mammalian host. *Molecular & Cellular Proteomics,* 17**,** 792-809.

CWIKLINSKI, K., ROBINSON, M. W., DONNELLY, S. & DALTON, J. P. 2021. Complementary transcriptomic and proteomic analyses reveal the cellular and molecular processes that drive growth and development of Fasciola hepatica in the host liver. *BMC genomics,* 22**,** 1-16.

DAGA, V., GREEN, E., RAVICHANDRAN, P., SHORT, M. & MAY, M. 2022. Multi-Omic Approaches to Vaccine Development against Helminth Diseases. *Parasitic Helminths and Zoonoses-From Basic to Applied Research.* IntechOpen.

DALTON, J. P., NEILL, S. O., STACK, C., COLLINS, P., WALSHE, A., SEKIYA, M., DOYLE, S., MULCAHY, G., HOYLE, D. & KHAZNADJI, E. 2003. Fasciola hepatica cathepsin L-like proteases: biology, function, and potential in the development of first generation liver fluke vaccines. *International journal for parasitology,* 33**,** 1173-1181.

DALTON, J. P., ROBINSON, M. W., MULCAHY, G., O'NEILL, S. M. & DONNELLY, S. 2013. Immunomodulatory molecules of Fasciola hepatica: candidates for both vaccine and immunotherapeutic development. *Veterinary parasitology,* 195**,** 272-285.

DAVIS, C. N., PHILLIPS, H., TOMES, J. J., SWAIN, M. T., WILKINSON, T. J., BROPHY, P. M. & MORPHEW, R. M. 2019. The importance of extracellular vesicle purification for downstream analysis: A comparison of differential centrifugation and size exclusion chromatography for helminth pathogens. *PLoS neglected tropical diseases,* 13**,** e0007191.

DAVIS, J. S. & MOYLE, L. C. 2020. Constitutive and plastic gene expression variation associated with desiccation resistance differences in the Drosophila americana species group. *Genes,* 11**,** 146.

DE MATOS, A. F. I. M., NOBRE, C. O. R., MONTEIRO, J. P., BEVILAQUA, C. M. L., SMITH, W. D. & TEIXEIRA, M. 2017. Attempt to control Haemonchus contortus in dairy goats with Barbervax®, a vaccine derived from the nematode gut membrane glycoproteins. *Small ruminant research,* 151**,** 1-4.

DE WAAL, T. 2016. Diseases of Dairy Animals: Parasites, Internal: Liver Flukes. *Reference Module in Food Science.* Elsevier.

DEL ANGEL, V. D., HJERDE, E., STERCK, L., CAPELLA-GUTIERREZ, S., NOTREDAME, C., PETTERSSON, O. V., AMSELEM, J., BOURI, L., BOCS, S. & KLOPP, C. 2018. Ten steps to get started in Genome Assembly and Annotation. *F1000Research,* 7.

DELANO, W. L. 2002. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography,* 40**,** 82-92.

DELPORT, W., POON, A. F., FROST, S. D. & KOSAKOVSKY POND, S. L. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics,* 26**,** 2455-2457.

DETTNER, K. 2010. 4.09 - Chemical Defense and Toxins of Lower Terrestrial and Freshwater Animals. *In:* LIU, H.-W. & MANDER, L. (eds.) *Comprehensive Natural Products II.* Oxford: Elsevier.

DEVINE, C., BRENNAN, G. P., LANUSSE, C. E., ALVAREZ, L. I., TRUDGETT, A., HOEY, E. & FAIRWEATHER, I. 2008. Effect of the metabolic inhibitor, methimazole on the drug susceptibility of a triclabendazole-resistant isolate of Fasciola hepatica. *Parasitology,* 136**,** 183-192.

DEVINE, C., BRENNAN, G. P., LANUSSE, C. E., ALVAREZ, L. I., TRUDGETT, A., HOEY, E. & FAIRWEATHER, I. 2009. Inhibition of cytochrome P450-mediated metabolism enhances ex vivo susceptibility of Fasciola hepatica to triclabendazole. *Parasitology,* 137**,** 871-880.

DEVINE, C., BRENNAN, G. P., LANUSSE, C. E., ALVAREZ, L. I., TRUDGETT, A., HOEY, E. & FAIRWEATHER, I. 2011. Enhancement of triclabendazole action in vivo against a triclabendazole-resistant isolate of Fasciola hepatica by co-treatment with ketoconazole. *Veterinary Parasitology,* 177**,** 305-315.

DEVINE, C., BRENNAN, G. P., LANUSSE, C. E., ALVAREZ, L. I., TRUDGETT, A., HOEY, E. & FAIRWEATHER, I. 2012. Potentiation of triclabendazole action in vivo against a triclabendazole-resistant isolate of Fasciola hepatica following its co-administration with the metabolic inhibitor, ketoconazole. *Veterinary Parasitology,* 184**,** 37-47.

DONNELLY, S., DALTON, J. P. & ROBINSON, M. W. 2011. How pathogen-derived cysteine proteases modulate host immune responses. *Advances in experimental medicine and biology,* 712**,** 192-207.

DREYFUSS, G. & RONDELAUD, D. 1994. Fasciola hepatica: a study of the shedding of cercariae from Lymnaea truncatula raised under constant conditions of temperature and photoperiod. *Parasite,* 1**,** 401-404.
DRUREY, C., COAKLEY, G. & MAIZELS, R. M. 2020. Extracellular vesicles: new targets for vaccines against helminth parasites. *International Journal for Parasitology,* 50**,** 623-633.
DURET, L. 2008. Neutral theory: the null hypothesis of molecular evolution. *Nature education,* 1**,** 803-806.
DUSAK, A., ONUR, M. R., CICEK, M., FIRAT, U., REN, T. & DOGRA, V. S. 2012. Radiological imaging features of Fasciola hepatica infection–a pictorial review. *Journal of clinical imaging science,* 2.
EGEA, R., CASILLAS, S. & BARBADILLA, A. 2008. Standard and generalized McDonald–Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic acids research,* 36**,** W157-W162.
EID, J., FEHR, A., GRAY, J., LUONG, K., LYLE, J., OTTO, G., PELUSO, P., RANK, D., BAYBAYAN, P. & BETTMAN, B. 2009. Real-time DNA sequencing from single polymerase molecules. *Science,* 323**,** 133-138.
EL-BAWAB, F. 2020. Chapter 7 - Phylum Platyhelminthes, Turbellaria. *In:* EL-BAWAB, F. (ed.) *Invertebrate Embryology and Reproduction.* Academic Press.
EL-KASABY, A., STOCKNER, T. & KUDLACEK, O. 2011. ABC transporters of Fasciola hepatica as putative drug targets. *BMC Pharmacology,* 11**,** A36.
EMMS, D. M. & KELLY, S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology,* 20**,** 238.
EVERTS, B., TUSSIWAND, R., DREESEN, L., FAIRFAX, K. C., HUANG, S. C., SMITH, A. M., O'NEILL, C. M., LAM, W. Y., EDELSON, B. T., URBAN, J. F., JR., MURPHY, K. M. & PEARCE, E. J. 2016. Migratory CD103+ dendritic cells suppress helminth-driven type 2 immunity through constitutive expression of IL-12. *J Exp Med,* 213**,** 35-51.
EYRE-WALKER, A. 2002. Changing effective population size and the McDonald-Kreitman test. *Genetics,* 162**,** 2017-2024.
FAIRWEATHER, I. 2005. Triclabendazole: new skills to unravel an old (ish) enigma. *Journal of helminthology,* 79**,** 227-234.
FAIRWEATHER, I. & BORAY, J. C. 1999. Fasciolicides: Efficacy, Actions, Resistance and its Management. *The Veterinary Journal,* 158**,** 81-112.
FAIRWEATHER, I., BRENNAN, G. P., HANNA, R. E. B., ROBINSON, M. W. & SKUCE, P. J. 2020. Drug resistance in liver flukes. *International Journal for Parasitology: Drugs and Drug Resistance,* 12**,** 39-59.
FERNANDEZ-BACA, M. V., HOBAN, C., ORE, R. A., ORTIZ, P., CHOI, Y.-J., MURGA-MORENO, C., MITREVA, M. & CABADA, M. M. 2022. The Differences in the Susceptibility Patterns to Triclabendazole Sulfoxide in Field Isolates of Fasciola hepatica Are Associated with Geographic, Seasonal, and Morphometric Variations. *Pathogens,* 11**,** 625.
FINN, R. D., COGGILL, P., EBERHARDT, R. Y., EDDY, S. R., MISTRY, J., MITCHELL, A. L., POTTER, S. C., PUNTA, M., QURESHI, M., SANGRADOR-VEGAS, A., SALAZAR, G. A., TATE, J. & BATEMAN, A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research,* 44**,** D279-D285.
FISSIHA, W. & KINDE, M. Z. 2021. Anthelmintic Resistance and Its Mechanism: A Review. *Infection and Drug Resistance,* 14**,** 5403.
FLEISCHMANN, R. D., ADAMS, M. D., WHITE, O., CLAYTON, R. A., KIRKNESS, E. F., KERLAVAGE, A. R., BULT, C. J., TOMB, J.-F., DOUGHERTY, B. A. & MERRICK, J. M. 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science,* 269**,** 496-512.
FURNHAM, N., DE BEER, T. A. & THORNTON, J. M. 2012. Current challenges in genome annotation through structural biology and bioinformatics. *Current opinion in structural biology,* 22**,** 594-601.

GARCIA-CAMPOS, A., CORREIA, C. N., NARANJO-LUCENA, A., GARZA-CUARTERO, L., FARRIES, G., BROWNE, J. A., MACHUGH, D. E. & MULCAHY, G. 2019. Fasciola hepatica infection in cattle: analyzing responses of peripheral blood mononuclear cells (PBMC) using a transcriptomics approach. *Frontiers in immunology,* 10**,** 2081.

GARDNER, M. J., HALL, N., FUNG, E., WHITE, O., BERRIMAN, M., HYMAN, R. W., CARLTON, J. M., PAIN, A., NELSON, K. E., BOWMAN, S., PAULSEN, I. T., JAMES, K., EISEN, J. A., RUTHERFORD, K., SALZBERG, S. L., CRAIG, A., KYES, S., CHAN, M.-S., NENE, V., SHALLOM, S. J., SUH, B., PETERSON, J., ANGIUOLI, S., PERTEA, M., ALLEN, J., SELENGUT, J., HAFT, D., MATHER, M. W., VAIDYA, A. B., MARTIN, D. M. A., FAIRLAMB, A. H., FRAUNHOLZ, M. J., ROOS, D. S., RALPH, S. A., MCFADDEN, G. I., CUMMINGS, L. M., SUBRAMANIAN, G. M., MUNGALL, C., VENTER, J. C., CARUCCI, D. J., HOFFMAN, S. L., NEWBOLD, C., DAVIS, R. W., FRASER, C. M. & BARRELL, B. 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature,* 419**,** 498-511.

GEIB, S. M., HALL, B., DEREGO, T., BREMER, F. T., CANNOLES, K. & SIM, S. B. 2018. Genome Annotation Generator: a simple tool for generating and correcting WGS annotation tables for NCBI submission. *Gigascience,* 7**,** giy018.

GEISEL, N. 2011. Constitutive versus responsive gene expression strategies for growth in changing environments. *PLoS One,* 6**,** e27033.

GHEDIN, E., WANG, S., FOSTER, J. M. & SLATKO, B. E. 2004. First sequenced genome of a parasitic nematode. *Trends in Parasitology,* 20**,** 151-153.

GILAD, Y., OSHLACK, A. & RIFKIN, S. A. 2006. Natural selection on gene expression. *TRENDS in Genetics,* 22**,** 456-461.

GIRGIS, H. Z. 2015. Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. *BMC Bioinformatics,* 16**,** 227.

GIRI, P. & MOHAPATRA, B. 2017. Candidate Gene. *Encyclopedia of Animal Cognition and Behavior. Cham: Springer International Publishing. Available: DOI,* 10**,** 978-3.

GOBERT, G. N., MOERTEL, L., BRINDLEY, P. J. & MCMANUS, D. P. 2009. Developmental gene expression profiles of the human pathogen Schistosoma japonicum. *BMC Genomics,* 10**,** 128.

GÖNNERT, R. & ANDREWS, P. 1977. Praziquantel, a new board-spectrum antischistosomal agent. *Z Parasitenkd,* 52**,** 129-50.

GONZÁLEZ-MIGUEL, J., BECERRO-RECIO, D. & SILES-LUCAS, M. 2021. Insights into Fasciola hepatica Juveniles: Crossing the Fasciolosis Rubicon. *Trends in Parasitology,* 37**,** 35-47.

GREENBERG, R. M. 2013. ABC multidrug transporters in schistosomes and other parasitic flatworms. *Parasitology international,* 62**,** 647-653.

GROHME, M. A., SCHLOISSNIG, S., ROZANSKI, A., PIPPEL, M., YOUNG, G. R., WINKLER, S., BRANDL, H., HENRY, I., DAHL, A. & POWELL, S. 2018. The genome of Schmidtea mediterranea and the evolution of core cellular mechanisms. *Nature,* 554**,** 56-61.

HAÇARıZ, O. & SAYERS, G. P. 2016. The omic approach to parasitic trematode research—a review of techniques and developments within the past 5 years. *Parasitology Research,* 115**,** 2523-2543.

HALFERTY, L., BRENNAN, G. P., HANNA, R. E. B., EDGAR, H. W., MEANEY, M. M., MCCONVILLE, M., TRUDGETT, A., HOEY, L. & FAIRWEATHER, I. 2008. Tegumental surface changes in juvenile Fasciola hepatica in response to treatment in vivo with triclabendazole. *Veterinary Parasitology,* 155**,** 49-58.

HALLER, B. C. & MESSER, P. W. 2017. asymptoticMK: a web-based tool for the asymptotic McDonald–Kreitman test. *G3: Genes, Genomes, Genetics,* 7**,** 1569-1575.

HARDISON, R. C. 2003. Comparative genomics. *PLoS biology,* 1**,** e58.

HAROUN, E. T. M. & HILLYER, G. V. 1986. Resistance to fascioliasis — A review. *Veterinary Parasitology,* 20**,** 63-93.

HARRIS, K. 2018. Neutral Evolution: The randomness that shapes our DNA. *Elife,* 7**,** e41491.

HASIN, Y., SELDIN, M. & LUSIS, A. 2017. Multi-omics approaches to disease. *Genome Biology,* 18**,** 83.

HEPWORTH, M. R., MAURER, M. & HARTMANN, S. 2012. Regulation of type 2 immunity to helminths by mast cells. *Gut microbes,* 3**,** 476-481.
HERNÁNDEZ-GONZÁLEZ, A., VALERO, M. L., PINO, M. S. D., OLEAGA, A. & SILES-LUCAS, M. 2010. Proteomic analysis of in vitro newly excysted juveniles from Fasciola hepatica. *Molecular and Biochemical Parasitology,* 172**,** 121-128.
HODGKINSON, J., CWIKLINSKI, K., BEESLEY, N., PATERSON, S. & WILLIAMS, D. 2013. Identification of putative markers of triclabendazole resistance by a genome-wide analysis of genetically recombinant Fasciola hepatica. *Parasitology,* 140**,** 1523-1533.
HODGKINSON, J. E., CWIKLINSKI, K., BEESLEY, N., HARTLEY, C., ALLEN, K. & WILLIAMS, D. J. L. 2018. Clonal amplification of Fasciola hepatica in Galba truncatula: within and between isolate variation of triclabendazole-susceptible and -resistant clones. *Parasit Vectors,* 11**,** 363.
HOFF, K. J., LANGE, S., LOMSADZE, A., BORODOVSKY, M. & STANKE, M. 2015. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics,* 32**,** 767-769.
HOLT, C. & YANDELL, M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC bioinformatics,* 12**,** 491.
HONGO, J. A., DE CASTRO, G. M., CINTRA, L. C., ZERLOTINI, A. & LOBO, F. P. 2015. POTION: an end-to-end pipeline for positive Darwinian selection detection in genome-scale data through phylogenetic comparison of protein-coding genes. *BMC Genomics,* 16**,** 567.
HOULE, D., GOVINDARAJU, D. R. & OMHOLT, S. 2010. Phenomics: the next challenge. *Nature Reviews Genetics,* 11**,** 855-866.
HOWE, K. L., BOLT, B. J., SHAFIE, M., KERSEY, P. & BERRIMAN, M. 2017. WormBase ParaSite− a comprehensive resource for helminth genomics. *Molecular and biochemical parasitology,* 215**,** 2-10.
HOWELL, A. K., MCCANN, C. M., WICKSTEAD, F. & WILLIAMS, D. J. L. 2020. Co-infection of cattle with Fasciola hepatica or F. gigantica and Mycobacterium bovis: A systematic review. *PLOS ONE,* 14**,** e0226300.
HOYLE, R. C., VINEER, H. R., DUNCAN, J. S., WILLIAMS, D. J. & HODGKINSON, J. E. 2022. A survey of sheep and/or cattle farmers in the UK shows confusion over the diagnosis and control of rumen fluke and liver fluke. *Veterinary parasitology***,** 109812.
HUANG, K. K., HUANG, J., WU, J. K. L., LEE, M., TAY, S. T., KUMAR, V., RAMNARAYANAN, K., PADMANABHAN, N., XU, C. & TAN, A. L. K. 2021. Long-read transcriptome sequencing reveals abundant promoter diversity in distinct molecular subtypes of gastric cancer. *Genome biology,* 22**,** 1-24.
HUANG, Y., CHEN, W., WANG, X., LIU, H., CHEN, Y., GUO, L., LUO, F., SUN, J., MAO, Q., LIANG, P., XIE, Z., ZHOU, C., TIAN, Y., LV, X., HUANG, L., ZHOU, J., HU, Y., LI, R., ZHANG, F., LEI, H., LI, W., HU, X., LIANG, C., XU, J., LI, X. & YU, X. 2013. The Carcinogenic Liver Fluke, Clonorchis sinensis: New Assembly, Reannotation and Analysis of the Genome and Characterization of Tissue Transcriptomes. *PLOS ONE,* 8**,** e54732.
HUBLEY, R., FINN, R. D., CLEMENTS, J., EDDY, S. R., JONES, T. A., BAO, W., SMIT, A. F. & WHEELER, T. J. 2016. The Dfam database of repetitive DNA families. *Nucleic acids research,* 44**,** D81-D89.
HUERTA-CEPAS, J., SERRA, F. & BORK, P. 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Molecular biology and evolution,* 33**,** 1635-1638.
HUMANN, J. L., LEE, T., FICKLIN, S. & MAIN, D. 2019. Structural and Functional Annotation of Eukaryotic Genomes with GenSAS. *In:* KOLLMAR, M. (ed.) *Gene Prediction: Methods and Protocols.* New York, NY: Springer New York.
HURTREZ-BOUSSÈS, S., MEUNIER, C., DURAND, P. & RENAUD, F. 2001. Dynamics of host–parasite interactions: the example of population biology of the liver fluke (Fasciola hepatica). *Microbes and Infection,* 3**,** 841-849.
INTERNATIONAL HELMINTH GENOMES, C. 2019. Comparative genomics of the major parasitic worms. *Nature genetics,* 51**,** 163-174.

IRVING, J. A., SPITHILL, T. W., PIKE, R. N., WHISSTOCK, J. C. & SMOOKER, P. M. 2003. The evolution of enzyme specificity in Fasciola spp. *Journal of Molecular Evolution,* 57**,** 1-15.

ISKANDER, M., HAYDEN, K., VAN DOMSELAAR, G. & TSANG, R. 2017. First Complete Genome Sequence of Haemophilus influenzae Serotype a. *Genome Announc,* 5.

JARRETT, W., JENNINGS, F., MARTIN, B., MCINTYRE, W., MULLIGAN, W., SHARP, N. & URQUHART, G. 1958. A field trial of a parasitic bronchitis vaccine. *Veterinary record,* 70**,** 451-454.

JEFFERIES, J. R., CAMPBELL, A. M., ROSSUM, A. J. V., BARRETT, J. & BROPHY, P. M. 2001. Proteomic analysis of Fasciola hepatica excretory-secretory products. *PROTEOMICS: International Edition,* 1**,** 1128-1132.

JOHNSON, A. D., HANDSAKER, R. E., PULIT, S. L., NIZZARI, M. M., O'DONNELL, C. J. & DE BAKKER, P. I. W. 2008. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics,* 24**,** 2938-2939.

JONES, B. F. & CAPPELLO, M. 2004. Trematodes. *In:* JOHNSON, L. R. (ed.) *Encyclopedia of Gastroenterology.* New York: Elsevier.

JOVANOVIC, V. M., SARFERT, M., REYNA-BLANCO, C. S., INDRISCHEK, H., VALDIVIA, D. I., SHELEST, E. & NOWICK, K. 2021. Positive selection in gene regulatory factors suggests adaptive pleiotropic changes during human evolution. *Frontiers in genetics,* 12**,** 753.

JUNG, H., VENTURA, T., CHUNG, J. S., KIM, W.-J., NAM, B.-H., KONG, H. J., KIM, Y.-O., JEON, M.-S. & EYUN, S.-I. 2020. Twelve quick steps for genome assembly and annotation in the classroom. *PLoS Computational Biology,* 16**,** e1008325.

JURKA, J., KAPITONOV, V. V., PAVLICEK, A., KLONOWSKI, P., KOHANY, O. & WALICHIEWICZ, J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research,* 110**,** 462-467.

KAMALUDEEN, J., GRAHAM-BROWN, J., STEPHENS, N., MILLER, J., HOWELL, A., BEESLEY, N. J., HODGKINSON, J., LEARMOUNT, J. & WILLIAMS, D. 2019. Lack of efficacy of triclabendazole against Fasciola hepatica is present on sheep farms in three regions of England, and Wales. *Veterinary Record,* 184**,** 502-502.

KANEHISA, M., FURUMICHI, M., SATO, Y., ISHIGURO-WATANABE, M. & TANABE, M. 2020. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Research,* 49**,** D545-D551.

KANEHISA, M., SATO, Y. & MORISHIMA, K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of molecular biology,* 428**,** 726-731.

KEARN, G. C. 2018. Reproduction in the Platyhelminthes (Flatworms). *In:* SKINNER, M. K. (ed.) *Encyclopedia of Reproduction (Second Edition).* Oxford: Academic Press.

KELLEY, J. M., ELLIOTT, T. P., BEDDOE, T., ANDERSON, G., SKUCE, P. & SPITHILL, T. W. 2016. Current threat of triclabendazole resistance in Fasciola hepatica. *Trends in parasitology,* 32**,** 458-469.

KELLEY, J. M., RATHINASAMY, V., ELLIOTT, T. P., RAWLIN, G., BEDDOE, T., STEVENSON, M. A. & SPITHILL, T. W. 2020. Determination of the prevalence and intensity of Fasciola hepatica infection in dairy cattle from six irrigation regions of Victoria, South-eastern Australia, further identifying significant triclabendazole resistance on three properties. *Veterinary parasitology,* 277**,** 109019.

KIM, D., LANGMEAD, B. & SALZBERG, S. L. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature methods,* 12**,** 357.

KIMURA, M. 1983. *The neutral theory of molecular evolution*, Cambridge University Press.

KOTZE, A. C., HUNT, P. W., SKUCE, P., VON SAMSON-HIMMELSTJERNA, G., MARTIN, R. J., SAGER, H., KRÜCKEN, J., HODGKINSON, J., LESPINE, A. & JEX, A. R. 2014. Recent advances in candidate-gene and whole-genome approaches to the discovery of anthelmintic resistance markers and the description of drug/receptor interactions. *International Journal for Parasitology: Drugs and Drug Resistance,* 4**,** 164-184.

KOUADIO, J. N., GIOVANOLI EVACK, J., ACHI, L. Y., BALMER, O., UTZINGER, J., N'GORAN, E. K., BONFOH, B., HATTENDORF, J. & ZINSSTAG, J. 2021. Efficacy of triclabendazole and

albendazole against Fasciola spp. infection in cattle in Côte d'Ivoire: a randomised blinded trial. *Acta Tropica,* 222**,** 106039.
KUDLACEK, O., DÖNMEZ, Y. & STOCKNER, T. 2012. ABC transporters of the liver fluke Fasciola hepatica. *BMC Pharmacology and Toxicology,* 13**,** A76.
LACOURSE, E. J., PERALLY, S., MORPHEW, R. M., MOXON, J. V., PRESCOTT, M., DOWLING, D. J., O'NEILL, S. M., KIPAR, A., HETZEL, U. & HOEY, E. 2012. The Sigma class glutathione transferase from the liver fluke Fasciola hepatica. *PLoS neglected tropical diseases,* 6**,** e1666.
LAING, R., DOYLE, S. R., MCINTYRE, J., MAITLAND, K., MORRISON, A., BARTLEY, D. J., KAPLAN, R., CHAUDHRY, U., SARGISON, N. & TAIT, A. 2021. Transcriptomic analyses implicate neuronal plasticity and chloride homeostasis in ivermectin resistance and recovery in a parasitic nematode. *bioRxiv*.
LAIRD, P. & BORAY, J. 1992. Human fascioliasis successfully treated with triclabendazole. *Australian and New Zealand journal of medicine,* 22**,** 45-47.
LALOR, R., CWIKLINSKI, K., CALVANI, N. E. D., DOREY, A., HAMON, S., CORRALES, J. L., DALTON, J. P. & DE MARCO VERISSIMO, C. 2021. Pathogenicity and virulence of the liver flukes Fasciola hepatica and Fasciola Gigantica that cause the zoonosis Fasciolosis. *Virulence,* 12**,** 2839-2867.
LE CLEC'H, W. & CHEVALIER, F. D. 2021. Genetic analysis of praziquantel response in schistosome parasites implicates a transient receptor potential channel. 13**,** eabj9114.
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**,** 609-21.
LEE, Y. J., SHI, R. & WITT, S. N. 2013. The small molecule triclabendazole decreases the intracellular level of cyclic AMP and increases resistance to stress in Saccharomyces cerevisiae. *PloS one,* 8**,** e64337.
LETUNIC, I. & BORK, P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research,* 49**,** W293-W296.
LEUNG, S. K., JEFFRIES, A. R., CASTANHO, I., JORDAN, B. T., MOORE, K., DAVIES, J. P., DEMPSTER, E. L., BRAY, N. J., O'NEILL, P. & TSENG, E. 2021. Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. *Cell reports,* 37**,** 110022.
LI, B. W., WANG, Z., RUSH, A. C., MITREVA, M. & WEIL, G. J. 2012. Transcription profiling reveals stage- and function-dependent expression patterns in the filarial nematode Brugia malayi. *BMC Genomics,* 13**,** 184.
LI, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics,* 34**,** 3094-3100.
LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-2079.
LI, L., STOECKERT, C. J. & ROOS, D. S. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome research,* 13**,** 2178-2189.
LI, W., O'NEILL, K. R., HAFT, D. H., DICUCCIO, M., CHETVERNIN, V., BADRETDIN, A., COULOURIS, G., CHITSAZ, F., DERBYSHIRE, M. K., DURKIN, A. S., GONZALES, N. R., GWADZ, M., LANCZYCKI, C. J., SONG, J. S., THANKI, N., WANG, J., YAMASHITA, R. A., YANG, M., ZHENG, C., MARCHLER-BAUER, A. & THIBAUD-NISSEN, F. 2021. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucleic Acids Res,* 49**,** D1020-d1028.
LIAO, Y., SMYTH, G. K. & SHI, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics,* 30**,** 923-930.
LIBRADO, P. & ROZAS, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* 25**,** 1451-1452.

LOPEZ, M., MORALES, M. L., KONANA, M., HOYER, P., PINEDA-REYES, R., WHITE, A. C., JR., GARCIA, H. H., LESCANO, A. G., GOTUZZO, E. & CABADA, M. M. 2016. Kato-Katz and Lumbreras rapid sedimentation test to evaluate helminth prevalence in the setting of a school-based deworming program. *Pathogens and global health,* 110**,** 130-134.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology,* 15**,** 550.

LOVE, S. 2017. Liver fluke - a review. May 2017 ed.: Primefact 813 third edition.

LOWE, T. M. & EDDY, S. R. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res,* 25**,** 955-64.

LUO, X., CUI, K., WANG, Z., LI, Z., WU, Z., HUANG, W., ZHU, X.-Q., RUAN, J., ZHANG, W. & LIU, Q. 2021. High-quality reference genome of Fasciola gigantica: Insights into the genomic signatures of transposon-mediated evolution and specific parasitic adaption in tropical regions. *PLOS Neglected Tropical Diseases,* 15**,** e0009750.

MA, Z. S., LI, L., YE, C., PENG, M. & ZHANG, Y.-P. 2019. Hybrid assembly of ultra-long Nanopore reads augmented with 10x-Genomics contigs: Demonstrated with a human genome. *Genomics,* 111**,** 1896-1901.

MACÍAS, L. G., BARRIO, E. & TOFT, C. 2020. GWideCodeML: A Python Package for Testing Evolutionary Hypotheses at the Genome-Wide Level. *G3: Genes, Genomes, Genetics,* 10**,** 4369-4372.

MALLICK, S., GNERRE, S., MULLER, P. & REICH, D. 2009. The difficulty of avoiding false positives in genome scans for natural selection. *Genome research,* 19**,** 922-933.

MAPLESON, D., VENTURINI, L., KAITHAKOTTIL, G. & SWARBRECK, D. 2018. Efficient and accurate detection of splice junctions from RNA-seq with Portcullis. *GigaScience,* 7.

MARTÍNEZ-VALLADARES, M., CORDERO-PÉREZ, C. & ROJO-VÁZQUEZ, F. A. 2014. Efficacy of an anthelmintic combination in sheep infected with Fasciola hepatica resistant to albendazole and clorsulon. *Experimental Parasitology,* 136**,** 59-62.

MARTÍNEZ-VALLADARES, M., DEL ROSARIO FAMULARO, M., FERNÁNDEZ-PATO, N., CASTAÑÓN-ORDÓÑEZ, L., CORDERO-PÉREZ, C. & ROJO-VÁZQUEZ, F. A. 2010. Efficacy of nitroxynil against Fasciola hepatica resistant to triclabendazole in a naturally infected sheep flock. *Parasitology research,* 107**,** 1205-1211.

MAS, C. S. & BARGUES, M. 1997. Human liver flukes: a review. *Res. Rev. Parasitol,* 57**,** 145-218.

MCDONALD, J. H. & KREITMAN, M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. *Nature,* 351**,** 652.

MCKELLAR, Q. & SCOTT, E. 1990. The benzimidazole anthelmintic agents-a review. *Journal of veterinary pharmacology and therapeutics,* 13**,** 223-247.

MCMANUS, C., DO PRADO PAIM, T., DE MELO, C. B., BRASIL, B. S. & PAIVA, S. R. 2014. Selection methods for resistance to and tolerance of helminths in livestock. *Parasite,* 21.

MCMANUS, D. P. 2020. Recent Progress in the Development of Liver Fluke and Blood Fluke Vaccines. *Vaccines,* 8**,** 553.

MCNULTY, S. N., TORT, J. F., RINALDI, G., FISCHER, K., ROSA, B. A., SMIRCICH, P., FONTENLA, S., CHOI, Y.-J., TYAGI, R. & HALLSWORTH-PEPIN, K. 2017a. Genomes of Fasciola hepatica from the Americas reveal colonization with Neorickettsia endobacteria related to the agents of Potomac horse and human Sennetsu fevers. *PLoS genetics,* 13**,** e1006537.

MCNULTY, S. N., TORT, J. F., RINALDI, G., FISCHER, K., ROSA, B. A., SMIRCICH, P., FONTENLA, S., CHOI, Y.-J., TYAGI, R., HALLSWORTH-PEPIN, K., MANN, V. H., KAMMILI, L., LATHAM, P. S., DELL'OCA, N., DOMINGUEZ, F., CARMONA, C., FISCHER, P. U., BRINDLEY, P. J. & MITREVA, M. 2017b. Genomes of Fasciola hepatica from the Americas Reveal Colonization with Neorickettsia Endobacteria Related to the Agents of Potomac Horse and Human Sennetsu Fevers. *PLOS Genetics,* 13**,** e1006537.

MCVEIGH, P., MCCAMMICK, E., MCCUSKER, P., WELLS, D., HODGKINSON, J., PATERSON, S., MOUSLEY, A., MARKS, N. J. & MAULE, A. G. 2017. Profiling G protein-coupled receptors of

Fasciola hepatica identifies orphan rhodopsins unique to phylum Platyhelminthes. *bioRxiv*, 207316.
MCWILLIAM, H., LI, W., ULUDAG, M., SQUIZZATO, S., PARK, Y. M., BUSO, N., COWLEY, A. P. & LOPEZ, R. 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research,* 41**,** W597-W600.
MEANEY, M., SAVAGE, J., BRENNAN, G. P., HOEY, E., TRUDGETT, A. & FAIRWEATHER, I. 2013. Increased susceptibility of a triclabendazole (TCBZ)-resistant isolate of Fasciola hepatica to TCBZ following co-incubation in vitro with the P-glycoprotein inhibitor, R(+)-verapamil. *Parasitology,* 140**,** 1287-1303.
MESSER, P. W. & PETROV, D. A. 2013. Frequent adaptation and the McDonald–Kreitman test. *Proceedings of the National Academy of Sciences,* 110**,** 8615-8620.
MEZO, M., GONZÁLEZ-WARLETA, M., CASTRO-HERMIDA, J. A., MUIÑO, L. & UBEIRA, F. M. 2011. Association between anti-F. hepatica antibody levels in milk and production losses in dairy cows. *Veterinary Parasitology,* 180**,** 237-242.
MIKHEYEV, A. S. & TIN, M. M. 2014. A first look at the Oxford Nanopore MinION sequencer. *Molecular ecology resources,* 14**,** 1097-1102.
MILLER, C. M. D., HOWELL, M. J. & BORAY, J. C. 1993. Host effects on glutathione S-transferase activity in Fasciola hepatica. *International Journal for Parasitology,* 23**,** 1073-1076.
MILLER, W., MAKOVA, K. D., NEKRUTENKO, A. & HARDISON, R. C. 2004. Comparative genomics. *Annu. Rev. Genomics Hum. Genet.,* 5**,** 15-56.
MISTRY, J., CHUGURANSKY, S., WILLIAMS, L., QURESHI, M., SALAZAR, GUSTAVO A., SONNHAMMER, E. L. L., TOSATTO, S. C. E., PALADIN, L., RAJ, S., RICHARDSON, L. J., FINN, R. D. & BATEMAN, A. 2020. Pfam: The protein families database in 2021. *Nucleic Acids Research,* 49**,** D412-D419.
MITCHELL, G. 2002. Update on fasciolosis in cattle and sheep. *In Practice,* 24**,** 378-385.
MOLINA-HERNÁNDEZ, V., MULCAHY, G., PÉREZ, J., MARTÍNEZ-MORENO, Á., DONNELLY, S., O'NEILL, S. M., DALTON, J. P. & CWIKLINSKI, K. 2015. Fasciola hepatica vaccine: we may not be there yet but we're on the right road. *Veterinary parasitology,* 208**,** 101-111.
MORALES, M. L., TANABE, M. B., WHITE JR, A. C., LOPEZ, M., BASCOPE, R. & CABADA, M. M. 2021. Triclabendazole Treatment Failure for Fasciola hepatica Infection among Preschool and School-Age Children, Cusco, Peru. *Emerging infectious diseases,* 27**,** 1850.
MORPHEW, R. M., ECCLESTON, N., WILKINSON, T. J., MCGARRY, J., PERALLY, S., PRESCOTT, M., WARD, D., WILLIAMS, D., PATERSON, S. & RAMAN, M. 2012. Proteomics and in silico approaches to extend understanding of the glutathione transferase superfamily of the tropical liver fluke Fasciola gigantica. *Journal of Proteome research,* 11**,** 5876-5889.
MORPHEW, R. M., WILKINSON, T. J., MACKINTOSH, N., JAHNDEL, V., PATERSON, S., MCVEIGH, P., ABBAS ABIDI, S. M., SAIFULLAH, K., RAMAN, M. & RAVIKUMAR, G. 2016. Exploring and expanding the fatty-acid-binding protein superfamily in Fasciola species. *Journal of proteome research,* 15**,** 3308-3321.
MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods,* 5**,** 621-628.
MOTTIER, L., ALVAREZ, L., FAIRWEATHER, I. & LANUSSE, C. 2006. Resistance-induced changes in triclabendazole transport in Fasciola hepatica: ivermectin reversal effect. *J Parasitol,* 92**,** 1355-60.
MULCAHY, G. & DALTON, J. P. 2001. Cathepsin L proteinases as vaccines against infection with Fasciola hepatica (liver fluke) in ruminants. *Res Vet Sci,* 70**,** 83-6.
MUNITA, M. P., REA, R., MARTINEZ-IBEAS, A. M., BYRNE, N., MCGRATH, G., MUNITA-CORBALAN, L. E., SEKIYA, M., MULCAHY, G. & SAYERS, R. G. 2019. Liver fluke in Irish sheep: prevalence and associations with management practices and co-infection with rumen fluke. *Parasites & vectors,* 12**,** 1-14.
MURGA-MORENO, J., CORONADO-ZAMORA, M., HERVAS, S., CASILLAS, S. & BARBADILLA, A. 2019. iMKT: the integrative McDonald and Kreitman test. *Nucleic acids research,* 47**,** W283-W288.

MURO, A., CASANUEVA, P., LÓPEZ-ABÁN, J., RAMAJO, V., MARTÍNEZ-FERNÁNDEZ, A. R. & HILLYER, G. V. 2007. IDENTIFICATION OF FASCIOLA HEPATICA RECOMBINANT 15-KDA FATTY ACID–BINDING PROTEIN T-CELL EPITOPES THAT PROTECT AGAINST EXPERIMENTAL FASCIOLIASIS IN RABBITS AND MICE. *Journal of Parasitology,* 93**,** 817-823.

MURRELL, B., MOOLA, S., MABONA, A., WEIGHILL, T., SHEWARD, D., KOSAKOVSKY POND, S. L. & SCHEFFLER, K. 2013. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molecular biology and evolution,* 30**,** 1196-1205.

NÄSVALL, J., SUN, L., ROTH, J. R. & ANDERSSON, D. I. 2012. Real-time evolution of new genes by innovation, amplification, and divergence. *Science,* 338**,** 384-7.

NGAN, C.-H., HALL, D. R., ZERBE, B., GROVE, L. E., KOZAKOV, D. & VAJDA, S. 2012. FTSite: high accuracy detection of ligand binding sites on unbound protein structures. *Bioinformatics,* 28**,** 286-287.

NOVOBILSKÝ, A. & HÖGLUND, J. 2015. First report of closantel treatment failure against Fasciola hepatica in cattle. *International Journal for Parasitology: Drugs and Drug Resistance,* 5**,** 172-177.

OEY, H., ZAKRZEWSKI, M., NARAIN, K., DEVI, K. R., AGATSUMA, T., NAWARATNA, S., GOBERT, G. N., JONES, M. K., RAGAN, M. A., MCMANUS, D. P. & KRAUSE, L. 2018. Whole-genome sequence of the oriental lung fluke Paragonimus westermani. *GigaScience,* 8.

OZSOLAK, F. & MILOS, P. M. 2011. RNA sequencing: advances, challenges and opportunities. *Nature reviews genetics,* 12**,** 87-98.

PANDEY, T., GHOSH, A., TODUR, V. N., RAJENDRAN, V., KALITA, P., KALITA, J., SHUKLA, R., CHETRI, P. B., SHUKLA, H., SONKAR, A., LYNGDOH, D. L., SINGH, R., KHAN, H., NONGKHLAW, J., DAS, K. C. & TRIPATHI, T. 2020. Draft Genome of the Liver Fluke Fasciola gigantica. *ACS Omega,* 5**,** 11084-11091.

PANYARACHUN, B., NGAMNIYOM, A., SOBHON, P. & ANURACPREEDA, P. 2013. Morphology and histology of the adult Paramphistomum gracile Fischoeder, 1901. *Journal of Veterinary Science,* 14**,** 425-432.

PARK, S. K. & FRIEDRICH, L. 2021. Mechanism of praziquantel action at a parasitic flatworm ion channel. 13**,** eabj5832.

PARK, S. K. & MARCHANT, J. S. 2020. The Journey to Discovering a Flatworm Target of Praziquantel: A Long TRP. *Trends Parasitol,* 36**,** 182-194.

PARKINSON, J. & BLAXTER, M. 2009. Expressed sequence tags: an overview. *Expressed sequence tags (ESTs)***,** 1-12.

PARSCH, J., ZHANG, Z. & BAINES, J. F. 2009. The Influence of Demography and Weak Selection on the McDonald–Kreitman Test: An Empirical Study in Drosophila. *Molecular Biology and Evolution,* 26**,** 691-698.

PATERSON, S. & LELLO, J. 2003. Mixed models: getting the best use of parasitological data. *TRENDS in Parasitology,* 19**,** 370-375.

PATNALA, R., CLEMENTS, J. & BATRA, J. 2013. Candidate gene association studies: a comprehensive guide to useful in silicotools. *BMC genetics,* 14**,** 1-11.

PEACHEY, L. E., PINCHBECK, G. L., MATTHEWS, J. B., BURDEN, F. A., LESPINE, A., VON SAMSON-HIMMELSTJERNA, G., KRÜCKEN, J. & HODGKINSON, J. E. 2017. P-glycoproteins play a role in ivermectin resistance in cyathostomins. *International Journal for Parasitology: Drugs and Drug Resistance,* 7**,** 388-398.

PERERA, D. J. & NDAO, M. 2021. Promising technologies in the field of helminth vaccines. *Frontiers in Immunology,* 12.

POGGIO, T. V., JENSEN, O., MOSSELLO, M., IRIARTE, J., AVILA, H. G., GERTISER, M. L., SERAFINO, J., ROMERO, S., ECHENIQUE, M. A. & DOMINGUEZ, D. 2016. Serology and longevity of immunity against Echinococcus granulosus in sheep and llama induced by an oil-based EG 95 vaccine. *Parasite Immunology,* 38**,** 496-502.

POLAND, G. A. & OBERG, A. L. 2010. Vaccinomics and bioinformatics: accelerants for the next golden age of vaccinology. *Vaccine,* 28**,** 3509-3510.

PROTASIO, A. V., TSAI, I. J., BABBAGE, A., NICHOL, S., HUNT, M., ASLETT, M. A., DE SILVA, N., VELARDE, G. S., ANDERSON, T. J. C., CLARK, R. C., DAVIDSON, C., DILLON, G. P., HOLROYD, N. E., LOVERDE, P. T., LLOYD, C., MCQUILLAN, J., OLIVEIRA, G., OTTO, T. D., PARKER-MANUEL, S. J., QUAIL, M. A., WILSON, R. A., ZERLOTINI, A., DUNNE, D. W. & BERRIMAN, M. 2012. A Systematically Improved High Quality Genome and Transcriptome of the Human Blood Fluke Schistosoma mansoni. *PLOS Neglected Tropical Diseases,* 6**,** e1455.

RADIO, S., FONTENLA, S., SOLANA, V., SALIM, A. C. M., ARAÚJO, F. M. G., ORTIZ, P., HOBAN, C., MIRANDA, E., GAYO, V. & PAIS, F. S.-M. 2018. Pleiotropic alterations in gene expression in Latin American Fasciola hepatica isolates with different susceptibility to drugs. *Parasites & vectors,* 11**,** 56.

RAMOS-BENÍTEZ, M. J., RUIZ-JIMÉNEZ, C., AGUAYO, V. & ESPINO, A. M. 2017. Recombinant Fasciola hepatica fatty acid binding protein suppresses toll-like receptor stimulation in response to multiple bacterial ligands. *Scientific reports,* 7**,** 1-14.

REED, M. B., PANACCIO, M., STRUGNELL, R. A. & SPITHILL, T. W. 1998. Developmental expression of a Fasciola hepatica sequence homologous to ABC transporters1. *International journal for parasitology,* 28**,** 1375-1381.

RICAFRENTE, A., CWIKLINSKI, K., NGUYEN, H., DALTON, J. P., TRAN, N. & DONNELLY, S. 2022. Stage-specific miRNAs regulate gene expression associated with growth, development and parasite-host interaction during the intra-mammalian migration of the zoonotic helminth parasite Fasciola hepatica. *BMC Genomics,* 23**,** 419.

ROBINSON, M., TRUDGETT, A., HOEY, E. & FAIRWEATHER, I. 2002. Triclabendazole-resistant Fasciola hepatica: β-tubulin and response to in vitro treatment with triclabendazole. *Parasitology,* 124**,** 325-338.

ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2009a. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics,* 26**,** 139-140.

ROBINSON, M. W., DALTON, J. P., O'NEILL, S. M. & DONNELLY, S. M. 2012. Mechanisms of Immune Modulation by Fasciola hepatica: Importance for Vaccine Development and for Novel Immunotherapeutics. *Parasitic Helminths.*

ROBINSON, M. W., HOEY, E. M., FAIRWEATHER, I., DALTON, J. P., MCGONIGLE, S. & TRUDGETT, A. 2001. Characterisation of a β-tubulin gene from the liver fluke, Fasciolahepatica. *International Journal for Parasitology,* 31**,** 1264-1268.

ROBINSON, M. W., MCFERRAN, N., TRUDGETT, A., HOEY, L. & FAIRWEATHER, I. 2004. A possible model of benzimidazole binding to β-tubulin disclosed by invoking an inter-domain movement. *Journal of Molecular Graphics and Modelling,* 23**,** 275-284.

ROBINSON, M. W., MENON, R., DONNELLY, S. M., DALTON, J. P. & RANGANATHAN, S. 2009b. An integrated transcriptomics and proteomics analysis of the secretome of the helminth pathogen Fasciola hepatica proteins associated with invasion and infection of the mammalian host. *Molecular & Cellular Proteomics,* 8**,** 1891-1907.

ROBINSON, M. W., TORT, J. F., LOWTHER, J., DONNELLY, S. M., WONG, E., XU, W., STACK, C. M., PADULA, M., HERBERT, B. & DALTON, J. P. 2008. Proteomics and Phylogenetic Analysis of the Cathepsin L Protease Family of the Helminth Pathogen Fasciola hepatica: Expansion of a Repertoire of Virulence-associated Factors. *Molecular & Cellular Proteomics,* 7**,** 1111-1123.

ROBLES-PÉREZ, D., MARTÍNEZ-PÉREZ, J. M., ROJO-VÁZQUEZ, F. A. & MARTÍNEZ-VALLADARES, M. 2013. The diagnosis of fasciolosis in feces of sheep by means of a PCR and its application in the detection of anthelmintic resistance in sheep flocks naturally infected. *Veterinary Parasitology,* 197**,** 277-282.

RODRIGUES, J. P., MELQUIOND, A. S., KARACA, E., TRELLET, M., VAN DIJK, M., VAN ZUNDERT, G. C., SCHMITZ, C., DE VRIES, S. J., BORDOGNA, A., BONATI, L., KASTRITIS, P. L. & BONVIN, A. M.

2013. Defining the limits of homology modeling in information-driven protein docking. *Proteins,* 81**,** 2119-28.

ROJAS-CARABALLO, J., LÓPEZ-ABÁN, J., MORENO-PÉREZ, D. A., VICENTE, B., FERNÁNDEZ-SOTO, P., DEL OLMO, E., PATARROYO, M. A. & MURO, A. 2017. Transcriptome profiling of gene expression during immunisation trial against Fasciola hepatica: identification of genes and pathways involved in conferring immunoprotection in a murine model. *BMC infectious diseases,* 17**,** 1-14.

ROJAS-CARABALLO, J., LÓPEZ-ABÁN, J., PÉREZ DEL VILLAR, L., VIZCAÍNO, C., VICENTE, B., FERNÁNDEZ-SOTO, P., OLMO, E. D., PATARROYO, M. A. & MURO, A. 2014. In vitro and in vivo studies for assessing the immune response and protection-inducing ability conferred by Fasciola hepatica-derived synthetic peptides containing B-and T-cell epitopes. *PloS one,* 9**,** e105323.

ROKNI, M. B. 2014. Helminth-Trematode: Fasciola hepatica and Fasciola gigantica. *In:* MOTARJEMI, Y. (ed.) *Encyclopedia of Food Safety.* Waltham: Academic Press.

ROMERO, I. G., RUVINSKY, I. & GILAD, Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics,* 13**,** 505-516.

RYAN, L. A., HOEY, E., TRUDGETT, A., FAIRWEATHER, I., FUCHS, M., ROBINSON, M. W., CHAMBERS, E., TIMSON, D. J., RYAN, E. & FELTWELL, T. 2008. Fasciola hepatica expresses multiple α-and β-tubulin isotypes. *Molecular and biochemical parasitology,* 159**,** 73-78.

RYAN, S., SHIELS, J., TAGGART, C. C., DALTON, J. P. & WELDON, S. 2020. Fasciola hepatica-Derived Molecules as Regulators of the Host Immune Response. *Frontiers in immunology,* 11**,** 2182-2182.

SAARI, S., NÄREAHO, A. & NIKANDER, S. 2019. Chapter 3 - Trematoda (Flukes). *In:* SAARI, S., NÄREAHO, A. & NIKANDER, S. (eds.) *Canine Parasites and Parasitic Diseases.* Academic Press.

SACKETT, D. L., WERBOVETZ, K. A. & MORRISSETTE, N. S. 2010. Chapter 2 - Isolating Tubulin from Nonneural Sources. *In:* WILSON, L. & CORREIA, J. J. (eds.) *Methods in Cell Biology.* Academic Press.

SAHM, A., BENS, M., PLATZER, M. & SZAFRANSKI, K. 2017. PosiGene: automated and easy-to-use pipeline for genome-wide detection of positively selected genes. *Nucleic Acids Research,* 45**,** e100-e100.

SALZBERG, S. L. 2019. Next-generation genome annotation: we still struggle to get it right. BioMed Central.

SANABRIA, R., CEBALLOS, L., MORENO, L., ROMERO, J., LANUSSE, C. & ALVAREZ, L. 2013. Identification of a field isolate of Fasciola hepatica resistant to albendazole and susceptible to triclabendazole. *Veterinary Parasitology,* 193**,** 105-110.

SANCHEZ-VAZQUEZ, M. J. & LEWIS, F. I. 2013. Investigating the impact of fasciolosis on cattle carcase performance. *Veterinary Parasitology,* 193**,** 307-311.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences,* 74**,** 5463-5467.

SANSRI, V., MEEMON, K., CHANGKLUNGMOA, N., KUEAKHAI, P., CHANTREE, P., CHAICHANASAK, P., LORSUWANNARAT, N., ITAGAKI, T. & SOBHON, P. 2015. Protection against Fasciola gigantica infection in mice by vaccination with recombinant juvenile-specific cathepsin L. *Vaccine,* 33**,** 1596-1601.

SARI, B., ISIK, M., EYLEM, C. C., KILIC, C., OKESOLA, B. O., KARAKAYA, E., EMREGUL, E., NEMUTLU, E. & DERKUS, B. 2022. Omics Technologies for High-Throughput-Screening of Cell-Biomaterial Interactions. *Molecular Omics*.

SAVAGE, J., MEANEY, M., BRENNAN, G. P., HOEY, E., TRUDGETT, A. & FAIRWEATHER, I. 2013. Increased action of triclabendazole (TCBZ) in vitro against a TCBZ-resistant isolate of Fasciola hepatica following its co-incubation with the P-glycoprotein inhibitor, R(+)-verapamil. *Experimental Parasitology,* 135**,** 642-653.

SCALZITTI, N., JEANNIN-GIRARDON, A., COLLET, P., POCH, O. & THOMPSON, J. D. 2020. A benchmark study of ab initio gene prediction methods in diverse eukaryotic organisms. *BMC Genomics,* 21**,** 293.

SCARCELLA, S., LAMENZA, P., VIRKEL, G. & SOLANA, H. 2012. Expression differential of microsomal and cytosolic glutathione-S-transferases in Fasciola hepatica resistant at triclabendazole. *Molecular and Biochemical Parasitology,* 181**,** 37-39.

SCHNEIDER, M. V. & ORCHARD, S. 2011. Omics technologies, data and bioinformatics principles. *Bioinformatics for omics Data***,** 3-30.

SHARMA, S. & ANAND, N. 1997. Chapter 1 - Parasitic Diseases : An Overview. *In:* SHARMA, S. & ANAND, N. (eds.) *Pharmacochemistry Library.* Elsevier.

SHULTZ, A. J. & SACKTON, T. B. 2019. Immune genes are hotspots of shared positive selection across birds and mammals. *Elife,* 8**,** e41815.

SIMÃO, F. A., WATERHOUSE, R. M., IOANNIDIS, P., KRIVENTSEVA, E. V. & ZDOBNOV, E. M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics,* 31**,** 3210-3212.

SINGH, D. K., SINGH, V. K., SINGH, R. N. & KUMAR, P. 2021. Fasciolosis. *Fasciolosis: Causes, Challenges and Controls.* Springer.

SINGH, P. K., SINGH, R. P., SINGH, P. & SINGH, R. L. 2019. Chapter 2 - Food Hazards: Physical, Chemical, and Biological. *In:* SINGH, R. L. & MONDAL, S. (eds.) *Food Safety and Human Health.* Academic Press.

SKUCE, P. & ZADOKS, R. 2013. Liver fluke–a growing threat to UK livestock production. *Cattle Pract,* 21**,** 138-149.

SLATER, G. S. C. & BIRNEY, E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics,* 6**,** 31.

SMIT, A., HUBLEY, R. & GREEN, P. 2004. RepeatMasker Open-3.0. 2004. *Seattle (WA): Institute for Systems Biology*.

SMITH, N. G. & EYRE-WALKER, A. 2002. Adaptive protein evolution in Drosophila. *Nature,* 415**,** 1022-1024.

SMOOKER, P. M., JAYARAJ, R., PIKE, R. N. & SPITHILL, T. W. 2010. Cathepsin B proteases of flukes: the key to facilitating parasite control? *Trends in Parasitology,* 26**,** 506-514.

SOLOVYEV, V., KOSAREV, P., SELEDSOV, I. & VOROBYEV, D. 2006. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol,* 7 Suppl 1**,** S10.1-12.

SOTILLO, J., TOLEDO, R., MULVENNA, J. & LOUKAS, A. 2017. Exploiting helminth–host interactomes through big data. *Trends in parasitology,* 33**,** 875-888.

SPITHILL, T., SMOOKER, P. & COPEMAN, D. 1999. Fasciola gigantica: epidemiology, control, immunology and molecular biology In: Dalton JP (Ed.), 1999. Fasciolosis. CABI Publishing, Oxon, UK.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30**,** 1312-1313.

STANKE, M. & MORGENSTERN, B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic acids research,* 33**,** W465-W467.

STATHATOS, G. G., DUNLEAVY, J. E. M., ZENKER, J. & O'BRYAN, M. K. 2021. Delta and epsilon tubulin in mammalian development. *Trends in Cell Biology,* 31**,** 774-787.

STEIN, L. 2001. Genome annotation: from sequence to biology. *Nature reviews genetics,* 2**,** 493-503.

STEINBISS, S., SILVA-FRANCO, F., BRUNK, B., FOTH, B., HERTZ-FOWLER, C., BERRIMAN, M. & OTTO, T. D. 2016. Companion: a web server for annotation and analysis of parasite genomes. *Nucleic Acids Res,* 44**,** W29-34.

STEINWAY, S. N., DANNENFELSER, R., LAUCIUS, C. D., HAYES, J. E. & NAYAK, S. 2010. JCoDA: a tool for detecting evolutionary selection. *BMC Bioinformatics,* 11**,** 284.

STERN, A., DORON-FAIGENBOIM, A., EREZ, E., MARTZ, E., BACHARACH, E. & PUPKO, T. 2007. Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Research,* 35**,** W506-W511.

STITT, A., FAIRWEATHER, I., TRUDGETT, A. & JOHNSTON, C. 1992. Fasciola hepatica: localization and partial characterization of tubulin. *Parasitology research,* 78**,** 103-107.

STORER, J., HUBLEY, R., ROSEN, J., WHEELER, T. J. & SMIT, A. F. 2021. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA,* 12**,** 2.

STUART, R. B., ZWAANSWIJK, S., MACKINTOSH, N. D., WITIKORNKUL, B., BROPHY, P. M. & MORPHEW, R. M. 2021. The soluble glutathione transferase superfamily: role of Mu class in triclabendazole sulphoxide challenge in Fasciola hepatica. 120**,** 979-991.

SULLIVAN, G. M. & FEINN, R. 2012. Using effect size—or why the P value is not enough. *Journal of graduate medical education,* 4**,** 279-282.

SUYAMA, M., TORRENTS, D. & BORK, P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research,* 34**,** W609-W612.

THE UNIPROT, C. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research,* 49**,** D480-D489.

THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research,* 22**,** 4673-4680.

THORVALDSDÓTTIR, H., ROBINSON, J. T. & MESIROV, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics,* 14**,** 178-192.

TILGNER, H., GRUBERT, F., SHARON, D. & SNYDER, M. P. 2014. Defining a personal, allele-specific, and single-molecule long-read transcriptome. *Proceedings of the National Academy of Sciences,* 111**,** 9869-9874.

TOET, H., PIEDRAFITA, D. M. & SPITHILL, T. W. 2014. Liver fluke vaccines in ruminants: strategies, progress and future opportunities. *International journal for parasitology,* 44**,** 915-927.

TOLAN, R. W., JR. 2011. Fascioliasis Due to Fasciola hepatica and Fasciola gigantica Infection: An Update on This 'Neglected' Neglected Tropical Disease. *Laboratory Medicine,* 42**,** 107-116.

TONER, E., BRENNAN, G. P., HANNA, R. E., EDGAR, H. W. & FAIRWEATHER, I. 2010a. Tegumental surface changes in adult Fasciola hepatica in response to treatment in vivo with triclabendazole in the sheep host. *Vet Parasitol,* 172**,** 238-48.

TONER, E., BRENNAN, G. P., HANNA, R. E., EDGAR, H. W. & FAIRWEATHER, I. 2010b. Time-dependent changes to the tegumental system and gastrodermis of adult Fasciola hepatica following treatment in vivo with triclabendazole in the sheep host. *Vet Parasitol,* 174**,** 218-27.

TÖRÖNEN, P., MEDLAR, A. & HOLM, L. 2018. PANNZER2: a rapid functional annotation web server. *Nucleic Acids Research,* 46**,** W84-W88.

TOST, J. & GUT, I. G. 2007. DNA methylation analysis by pyrosequencing. *Nature protocols,* 2**,** 2265-2275.

UFFELMANN, E., HUANG, Q. Q., MUNUNG, N. S., DE VRIES, J., OKADA, Y., MARTIN, A. R., MARTIN, H. C., LAPPALAINEN, T. & POSTHUMA, D. 2021. Genome-wide association studies. *Nature Reviews Methods Primers,* 1**,** 1-21.

VALENTIM, C. L., CIOLI, D., CHEVALIER, F. D., CAO, X., TAYLOR, A. B., HOLLOWAY, S. P., PICA-MATTOCCIA, L., GUIDI, A., BASSO, A. & TSAI, I. J. 2013. Genetic and molecular basis of drug resistance and species-specific drug action in schistosome parasites. *Science,* 342**,** 1385-1389.

VATSIOU, A. I., BAZIN, E. & GAGGIOTTI, O. E. 2016. Changes in selective pressures associated with human population expansion may explain metabolic and immune related pathways enriched for signatures of positive selection. *BMC genomics,* 17**,** 1-11.

WAGNER, A. 2007. Rapid detection of positive selection in genes and genomes through variation clusters. *Genetics,* 176**,** 2451-2463.
WAIKAGUL, J., SATO, M. & SATO, M. O. 2015. 10 - Foodborne trematodes. *In:* GAJADHAR, A. A. (ed.) *Foodborne Parasites in the Food Supply Web.* Oxford: Woodhead Publishing.
WALKER, B. J., ABEEL, T., SHEA, T., PRIEST, M., ABOUELLIEL, A., SAKTHIKUMAR, S., CUOMO, C. A., ZENG, Q., WORTMAN, J., YOUNG, S. K. & EARL, A. M. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE,* 9**,** e112963.
WANG, D., KORHONEN, P. K., GASSER, R. B. & YOUNG, N. D. 2018. Improved genomic resources and new bioinformatic workflow for the carcinogenic parasite Clonorchis sinensis: Biotechnological implications. *Biotechnol Adv,* 36**,** 894-904.
WARD, Y., YAP, S.-F., RAVICHANDRAN, V., MATSUMURA, F., ITO, M., SPINELLI, B. & KELLY, K. 2002. The GTP binding proteins Gem and Rad are negative regulators of the Rho–Rho kinase pathway. *The Journal of cell biology,* 157**,** 291-302.
WATERHOUSE, A., BERTONI, M., BIENERT, S., STUDER, G., TAURIELLO, G., GUMIENNY, R., HEER, F. T., DE BEER, T. A. P., REMPFER, C. & BORDOLI, L. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research,* 46**,** W296-W303.
WAYNE, M. L. & MCINTYRE, L. M. 2002. Combining mapping and arraying: an approach to candidate gene identification. *Proceedings of the National Academy of Sciences,* 99**,** 14903-14906.
WEBB, A. E., WALSH, T. A. & O'CONNELL, M. J. 2017. VESPA: very large-scale evolutionary and selective pressure analyses. *PeerJ Computer Science,* 3**,** e118.
WHO. 2020. *Neglected tropical diseases: Fascioliasis* [Online]. World Health Organization. Available: https://www.who.int/news-room/questions-and-answers/item/q-a-on-fascioliasis.
WIJFFELS, G. L., SEXTON, J. L., SALVATORE, L., PETTITT, J. M., HUMPHRIS, D. C., PANACCIO, M. & SPITHILL, T. W. 1992. Primary sequence heterogeneity and tissue expression of glutathione S-transferases of Fasciola hepatica. *Exp Parasitol,* 74**,** 87-99.
WILLIAMS, D. 2020. Update on liver fluke in sheep. *In Practice,* 42**,** 341-347.
WILSON, R. A., WRIGHT, J. M., DE CASTRO-BORGES, W., PARKER-MANUEL, S. J., DOWLE, A. A., ASHTON, P. D., YOUNG, N. D., GASSER, R. B. & SPITHILL, T. W. 2011. Exploring the Fasciola hepatica tegument proteome. *International journal for parasitology,* 41**,** 1347-1359.
WOLSTENHOLME, A. J., FAIRWEATHER, I., PRICHARD, R., VON SAMSON-HIMMELSTJERNA, G. & SANGSTER, N. C. 2004. Drug resistance in veterinary helminths. *Trends in Parasitology,* 20**,** 469-476.
YANDELL, M. & ENCE, D. 2012. A beginner's guide to eukaryotic genome annotation. *Nature Reviews Genetics,* 13**,** 329.
YANG, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics,* 13**,** 555-556.
YANG, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution,* 24**,** 1586-1591.
YANG, Z. & NIELSEN, R. 2002. Codon-Substitution Models for Detecting Molecular Adaptation at Individual Sites Along Specific Lineages. *Molecular Biology and Evolution,* 19**,** 908-917.
YANG, Z., NIELSEN, R., GOLDMAN, N. & PEDERSEN, A.-M. K. 2000. Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites. *Genetics,* 155**,** 431.
YANG, Z., WONG, W. S. W. & NIELSEN, R. 2005. Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. *Molecular Biology and Evolution,* 22**,** 1107-1118.
YOUNG, N. D., NAGARAJAN, N., LIN, S. J., KORHONEN, P. K., JEX, A. R., HALL, R. S., SAFAVI-HEMAMI, H., KAEWKONG, W., BERTRAND, D., GAO, S., SEET, Q., WONGKHAM, S., TEH, B. T., WONGKHAM, C., INTAPAN, P. M., MALEEWONG, W., YANG, X., HU, M., WANG, Z., HOFMANN, A., STERNBERG, P. W., TAN, P., WANG, J. & GASSER, R. B. 2014. The Opisthorchis viverrini genome provides insights into life in the bile duct. *Nature Communications,* 5**,** 4378.

ZAFRA, R., BUFFONI, L., PÉREZ-CABALLERO, R., MOLINA-HERNÁNDEZ, V., RUIZ-CAMPILLO, M. T., PÉREZ, J., MARTÍNEZ-MORENO, Á. & MARTÍNEZ MORENO, F. J. 2021. Efficacy of a multivalent vaccine against Fasciola hepatica infection in sheep. *Veterinary Research,* 52**,** 1-9.

ZAJAC, N., ZOLLER, S., SEPPÄLÄ, K., MOI, D., DESSIMOZ, C., JOKELA, J., HARTIKAINEN, H. & GLOVER, N. 2021. Gene duplication and gain in the trematode Atriophallophorus winterbourni contributes to adaptation to parasitism. *Genome biology and evolution,* 13**,** evab010.

ZAYAS, R. M., HERNÁNDEZ, A., HABERMANN, B., WANG, Y., STARY, J. M. & NEWMARK, P. A. 2005. The planarian Schmidtea mediterranea as a model for epigenetic germ cell specification: analysis of ESTs from the hermaphroditic strain. *Proceedings of the National Academy of Sciences,* 102**,** 18491-18496.

ZHANG, J., NIELSEN, R. & YANG, Z. 2005. Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level. *Molecular Biology and Evolution,* 22**,** 2472-2479.

ZHANG, X.-X., CWIKLINSKI, K., HU, R.-S., ZHENG, W.-B., SHENG, Z.-A., ZHANG, F.-K., ELSHEIKHA, H. M., DALTON, J. P. & ZHU, X.-Q. 2019. Complex and dynamic transcriptional changes allow the helminth Fasciola gigantica to adjust to its intermediate snail and definitive mammalian hosts. *BMC genomics,* 20**,** 1-18.

ZHU, M. & ZHAO, S. 2007. Candidate gene identification approach: progress and challenges. *International journal of biological sciences,* 3**,** 420.

Appendix

The **Appendix files** folder includes selected output files from the various analysis done (words in "**bold**" depict folders) .

Chapter2_files

F1 – Braker *ab initio* annotation output files, including the 74,307 predicted gene models from the RNAseq dataset 1.

F2 – Braker *ab initio* annotation output files including the 53,729 predicted gene models from the RNAseq dataset 2

F3 – This folder includes output files of the Maker annotation (after the 4th SNAP training of gene models). A total of 15,879 genes were predicted.

F4 – This folder includes output files of Transdecoder annotation to predict genes. A total of 9,401 genes were predicted.

F5 – This folder includes a "draft" annotation of the *F. gigantica* genome from the North-Eastern Hill University, India (ASM286751v3, BioProject: PRJNA339660, BioSample: SAMN05601579). Using Augustus, a total of 11,947 *F. gigantica* genes were predicted.

F6 – Files include the output of the RepeatMasking process of the analysis. Files describe repetitive elements identified and their genome coordinates.

F7 – Orthomcl output file of the orthologous grouping of the *F. hepatica* annotation described earlier (See F3), *F. gigantica* "draft annotation (See F4), *Paragonimus westermani* (ASM850834v1), *Echinostoma caproni* (E_caproni_Egypt_0011_upd), *Schistosoma mansoni* (Smansoni_v7), *Opisthorchis viverrini* (OpiViv1.0), and *Clonorchis sinensis* (C_sinensis-2.0).

F8 – Files describe the functional annotation of the predicted gene models. These output files were generated using PANNZER and include each predicted gene's functional description (DE) and Gene Ontology (GO).

F9 – GhostKOALA annotation output of the predicted gene models (See F3) against nonredundant set of KEGG GENES to assign K numbers to the *F. hepatica* query gene models.

F10 – Screenshots from IGV showing the various genes aligned with the *F. hepatica* genome.

Chapter3_files

Gene-List – This folder contains the compiled list of genes (gene names refer to the *F. hepatica Fasciola*_10x_pilon, GCA_900302435.1, WBPS16 annotation).

MKTest-Analysis – In folder F1, there is an R script used to identify the effect of the *F. hepatica* intra-specie SNPs from 5 isolates used for the *F. hepatica* genome sequencing. The genome "fasta" file, genome annotation "gff" file, and the SNPs "vcf" files was used determine the effect of each SNP (R code adapted from https://eacooper400.github.io/gen8900/exercises/mk.html). The folder also contains the

list of genes and their SNPs effect. The folders F2 – F10 contain the output file of the MKtest analysis and their respective R codes for each of the *F. hepatica* candidate genes assessed.

Orthofinder-Results – This folder contains the orthologous grouping of eight trematode species. These species include *F. hepatica* (*Fasciola*_10x_pilon, GCA_900302435.1), *F. gigantica* (F_gigantica_1.0.allpaths), *Fasciolopsis buski* (F_buski_1.0.allpaths-lg), *Paragonimus westermani* (ASM850834v1), *Echinostoma caproni* (E_caproni_Egypt_0011_upd), *Schistosoma mansoni* (Smansoni_v7), *Opisthorchis viverrini* (OpiViv1.0), and *Clonorchis sinensis* (C_sinensis-2.0). A total of 11,976 orthogroups were identified (see the Statistics_Overall.tsv file)

PAML-Analysis – This folder has sub-folders (F1 – F9) – each including the output of the evolutionary hypothesis tested. Site models M8a (relaxation) and M8 (positive selection) were tested using PAML software (see chapter 3, Table 3.4). For each candidate gene family assessed, results are presented (such as identified gene duplication events, orthologous grouping, gene annotations, sites on selection pressure).

Chapter4_files

StageSpecificExpressionStudy – This folder includes R codes used to analyse the RNAseq data from the various stages of the *F. hepatica* life stages.

TCBZ-GeneExpressionStudy – This folder has sub-folders (F1 – F10) – each folder includes the output of the models tested (see chapter 4, Table 4.3) to assess the impact of TCBZ on gene activity and the R codes for each of the candidate families analysed. The expression levels in each gene assessed across the TCBZ susceptible versus resistant isolates were described using boxplots (see the **gene_plots** sub-folders).