Chapter 1

Introduction

1.1 Introduction

Schistosomiasis is caused by a genus of platyhelminths called *Schistosoma;* an estimated 200 million people require treatment for the disease annually, approximately 90% of whom live in sub-Saharan Africa (WHO, 2018). According to the Global Burden of Disease Study the death of 5500 people was directly caused by schistosomiasis in 2013 (Naghavi *et al.*, 2015). However the precise number is difficult to estimate because of the difficulties of measuring the contribution of chronic schistosomiasis to conditions such as kidney failure and bladder cancer and estimates are as high as 200,000 deaths per year globally WHO (2018).

1.2 Biology of schistosomes

Schistosoma mansoni as well as several other species of schistosomes, notably *S. japonicum* and *Schistosoma hematobium*, are responsible for schistosomiasis. Schistosomes belong to the class Trematoda, which usually have a flattened, oval body shape ranging in length from around 0.1 cm to 7 cm with a ventral as well as an oral sucker (Barnes and Robert, 1982). Schistosomes are around 0.6-1.1 cm long and 0.1 cm wide, with a long, slim body (Machado-Silva *et al.*, 1997). The morphology of male and female worms differs considerably (Loker & Brant, 2006). Male worms maintain the typical flattened body shape of trematodes but curled up to form the gynecophoral canal, whereas the female worms have a cylindrical profile (Figure 1.1).

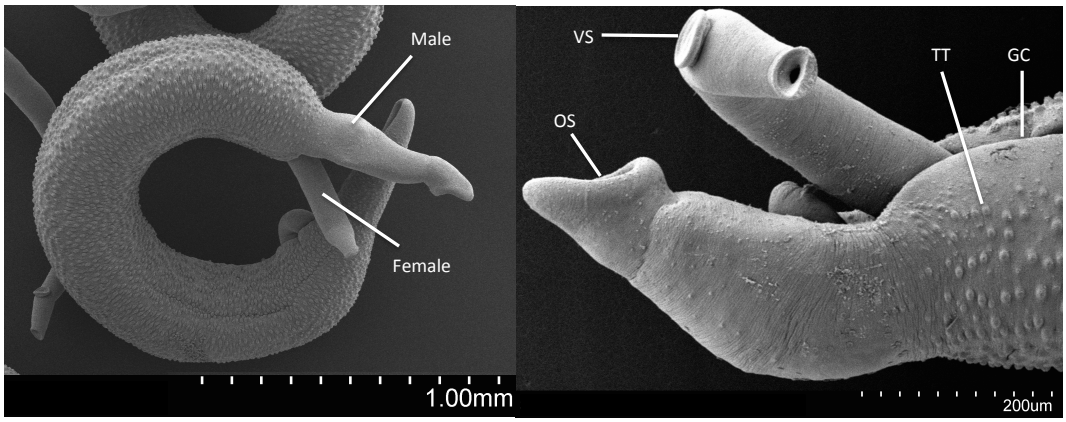

Figure 1.1: Scanning electron microscopy of adult *Schistosoma mansoni* couples. After pairing, the female worm resides inside the gynecophoral canal (GC) of the male. Both worms have an oral (OS) as well as a ventral sucker (VS). The tegument of male worms is covered with tegumental tubercles (TT). These images were taken in collaboration with Dave Goulding, WTSI, Hinxton.

1.2.1 Life cycle

The *S. mansoni* life cycle (see Figure 1.2) is similar to that of other trematodes and involves a mollusc and a mammalian host (Tucker *et al.*, 2013). However, unlike most trematodes, the members of the Schistosomatidae, the family *S. mansoni* belongs to, are dioecious (Loker & Brant, 2006). This means that there are distinct male and female worms instead of hermaphrodites. The life cycle of *S. mansoni* features two hosts, a molluscan intermediate host, in which the parasites undergo asexual reproduction, and the vertebrate definitive host, in which the adult worms reproduce sexually and lay eggs (Tucker *et al.*, 2013). When the eggs are placed in fresh water, ciliated larvae called miracidia (136 ± 3 µm in length and 54 ± 1 µm in width (Eklu-Natey *et al.*, 1985) are released. *S. mansoni* miracidia invade the intermediate host – fresh water snails of the genus *Biomphalaria* – by burrowing through the skin of its foot (Cosseau *et al.*, 2009). Inside the snail, the miracidia develop into primary (or mother) sporocysts,

which in turn produce secondary (or daughter) sporocycsts by asexual reproduction (Cosseau *et al.*, 2009).

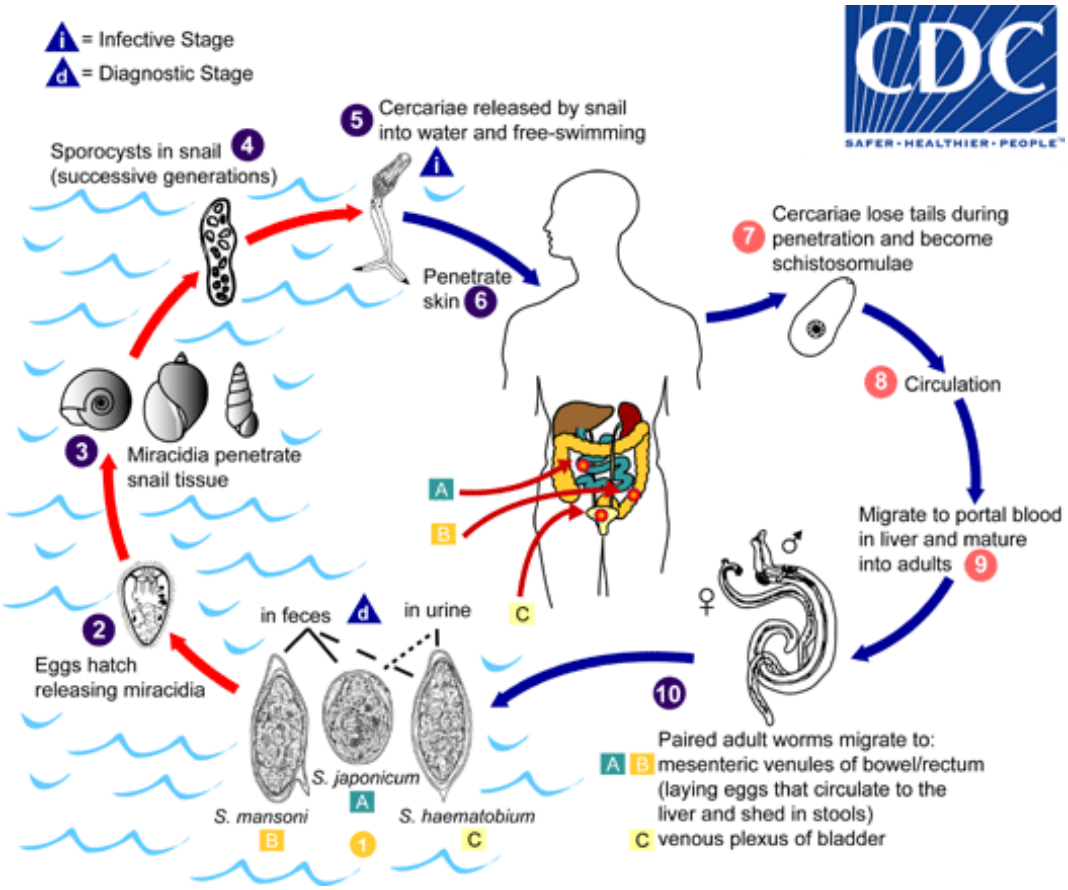

Figure 1.2: Life cycle of *Schistosoma* species. 1) Eggs are excreted via faeces or in the case of *S. haematobium* in the urine and reach fresh water. 2) Miracidia hatch from the eggs and seek a snail host. 3) The miracidia penetrate a snail host (in the case of *S. mansoni* a snail of the genus *Biomphalaria*) and 4) two generations of sporocysts develop from the miracidia inside the host. 5) Cercariae are produced and released into the water. 6) Cercariae penetrate the human skin, 7) shedding their tail and transforming into schistosomulae. 8) After moving through the circulation and lungs for several weeks, 9) male and female adult worms pair in the liver and 10) relocate to the mesenteric system of the bowel (or the venous plexus of the bladder in the case of *S. haematobium*). The couples then start egg-laying, thereby continuing the life cycle. **Reproduced from CDC (2012); https://www.cdc.gov/parasites/schistosomiasis/biology.html**

Each daughter sporocyst produces a clonal population of cercariae, the life cycle stage infecting humans (McKerrow & Salter, 2002). By shedding during daylight hours, *S. mansoni* maximise the chances of the cercariae being released when a suitable mammalian host might be nearby (Fingerut *et al.*, 2003). The cercariae use their forked tail for locomotion and use changes in light and the presence of medium-chain fatty acids in the skin to detect a host (McKerrow & Salter, 2002). When a cercaria comes into contact with a suitable host, it attaches to the skin and uses proteases to break into the skin, migrating into the subcutaneous tissue, losing its tail in the process (Salter *et al.*, 2000).

Following host invasion, the parasites transform into schistosomules. They spend around days traveling through the skin and circulation, migrate through the lungs (peaking here around 5-6 d.p.i.), then continuing over the course of the next two weeks until they enter the portal vessel of the liver (P. Miller & Wilson, 1980). There, they begin to feed on blood as the gut develops and they begin to grow. At this stage the male worms become visibly larger than the females (Basch, 1991) From around 21 day *post* infection (d.p.i.) onward the parasites mature and males and females commence pairing around 28 d.p.i. Basch, 1991). It is well established that female worms require pairing with a male partner to develop mature gonads and vitellaria to become sexually mature (Kunz, 2001). Without pairing, female worms remain stunted and infertile (Collins & Newmark, 2013). Around 35 d.p.i., *S. mansoni* pairs have usually migrated into the mesenteric blood vessels and maturing females produce the first eggs (Basch, 1991). However, there is considerable variation in the time it takes worms to migrate through different parts of their hosts' bodies, depending on the species

of worm in question as well as the host species – mice, rats, hamsters and humans – varying by as much as 20 days (Yolles *et al.,* 1949; Wilson, 1978; Crabtree & Wilson, 1986; Miller & Wilson, 1980).

After *S. mansoni* pairs have migrated to the mesentery and commenced egg production, they will often continue to produce eggs for years if left untreated (Harris *et al.*, 1984) Approximately half of the eggs get trapped in various organs (LoVerde *et al.*, 2005), where they cause considerable damage (see Chapter 1.3.2). The remaining eggs cross through the wall of the intestine into the lumen of the gut from where they are excreted with faeces (deWalick *et al.*, 2011) If eggs wash into fresh water, miracidia hatch, ready to invade the snail intermediate host, and the cycle repeats.

Aside from the migration through the host, another critical aspect of worm development is the morphological changes that take place during the first 6-7 weeks after infection. Cort (1921) describes the development of male and female schistosomes from mixed sex (MS) infections beginning at the lung stage (around 7 d.p.i) until the worms have matured fully and commence oviposition (around 35 d.p.i.). The author notes that at 7 d.p.i. the young worms remain at approximately the same size as schistosomules with the gut starting to develop. The first differences between male and female worms are described at 15 d.p.i., when the worms reach the portal vessel of the liver, commence feeding on blood and begin to grow. Males are larger and have more developed oral and ventral suckers at this time. At around 21 d.p.i. organogeny begins: male worms develop testes and in females a uterus becomes visible. After another week of maturation,

at around 28 d.p.i., gametogeny is thought to take place . Males have been shown to possess 8 testes at 28 d.p.i., some of which contain sperm, and females to have developed ovaries. At 28 d.p.i., Cort (1921) first observed pairing of worms, but also showed that maturation of the vitellarium in particular requires a further week. At 30 d.p.i., the author demonstrated the presence of vitelline globes, but found that oviposition generally begins to occur at 35 d.p.i. At this point, females have well-developed vitelline glands to provide a steady supply of egg shell proteins, whilst male worms produce large quantities of sperm for fertilisation. Very similar observations were made by Biolchini *et al.* (2006), but technological improvements over the 85 years that separated these two studies, most notably confocal microscopy, allowed them to study the development of the worm organs in more detail. Indeed the latter authors noticed that at week three 47-59% of males had developed the gynecophoral canal, 5% of females had developed an ootype, and 33% of females had developed ovaries, whereas Cort (1921) did not observe ovaries at this time and made no mention of the gynecophoral canal development. At week four, Biolchini *et al.* (2006) found that 15-26% of male worms began to develop tegumental tubercles (Figure 1.1). Also, 69% of females were found to have ovaries and 12-75% an ootype, with a single female presenting with a fertilised egg. In addition, Biolchini *et al.* (2006) found the oral suckers of male worms to be significantly larger in size than that of females at 21-28 d.p.i., an important difference at 21 d.p.i. for distinguishing worms morphologically. Both papers noted the significance of pairing in the development of the females, which, unlike male worms, do not mature sexually until paired. In particular, development of the vitellarium and maturation of the ovaries was observed only following pairing (Biolchini *et al.*, 2006; Cort, 1921).

Notably, the development of *S. japonicum* appears to occur at a slightly faster pace, as most worms were found to be paired by 22 d.p.i. and the first sperm and mature oocytes are first observed at the same time, presumably in worms that paired particularly early, and sexually mature worms are found at 28 d.p.i. (Wang *et al.*, 2017).

The induced maturation of the reproductive organs of a paired female worm culminates in the production of eggs. The function of the ovary is to produce viable, fertilised oocytes, the vitellarium, on the other hand, produces eggshell precursor proteins and provides nutrient reserves for the developing embryo in the form of lipid droplets (LoVerde *et al.*, 2005). After oocytes have become fertilised with the male sperm inside the female oviduct, the fertilised cells move to the vitello-oviduct, where they are surrounded by approximately 38 vitellocytes (Loverde & Chen, 1991) (Figure 1.3). The maturation of vitellocytes is commonly divided into four stages (Erasmus, 1975): The first is a stem cell stage of undifferentiated, small cells near the periphery of the vitelline lobes (S1). Then the cell begins to differentiate and grow in size (S2). The cells then begin to synthesise large amounts of proteins and form vitelline droplets; while the rate of protein synthesis increases the vitelline droplets begin to form clusters. Subsequently, lipid droplets and a small number of β-glycogen particles accumulate to provide energy to the embryo (S3). The final stages of development, mature vitellocytes (S4), leave through the vitelline duct to reach the ootype (Figure 1.3) (Wang & Collins, 2016). Approximately 11,000 S4 vitellocytes are produced daily by mature females (LoVerde *et al.*, 2005) and to

satisfy this demand the vitellarium takes up a large portion of the female worm's body (Galanti *et al.*, 2012).

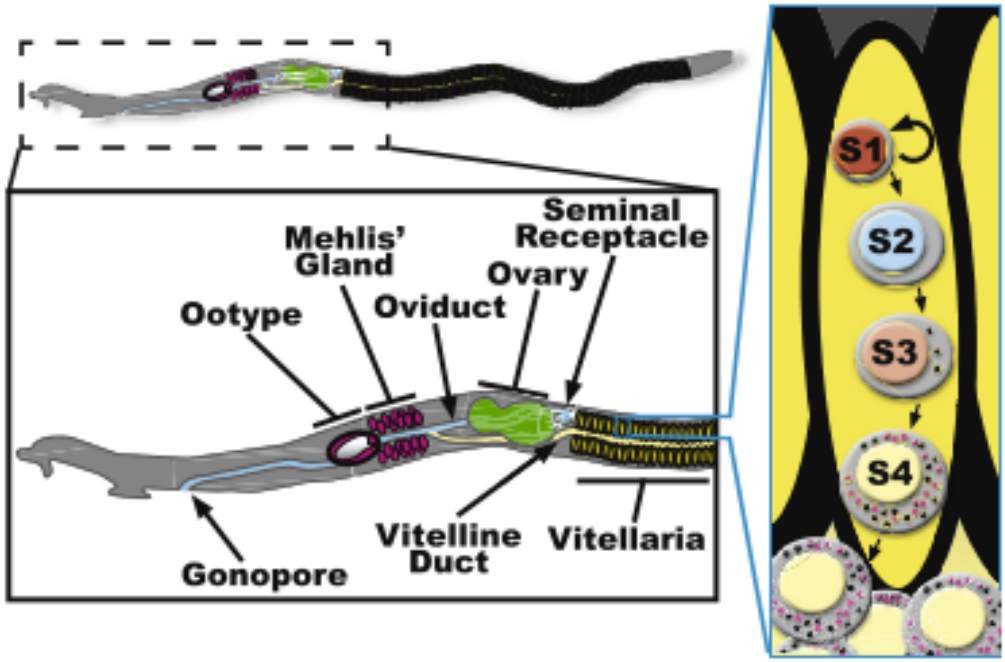

Figure 1.3: Diagram of the female reproductive organs, reproduced from Wang & Collins (2016). Left: The relative positions are shown for the gonopore, ootype, Mehlis' gland, oviduct, vitelline duct, ovary, vitellaria and seminal receptacle. Rigth: vitellocytes move towards the central vitelline duct as they mature through a sequence of stages (S1-S4); stem cell-like S1 cells, proliferating S2 cells, larger S3 vitellocytes accumulating granules, and finally mature S4 vitellocytes.

Each paired female has been estimated to lay around 300 eggs per day for several years (Loverde & Chen, 1991). Schistosome are estimated to live for 2.7-4.5 years (3.3 on average) (Goddard & Jordan, 1980) although there are reports of worms surviving and continuing to produce viable eggs for up to 47 years (Harris *et al.*, 1984; Markel *et al.*, 1978; Warren *et al.*, 1974). Adult *S. mansoni* cannot reproduce asexually and female reproductive maturity is critically dependent on pairing with a male worm. Not only does transmission rely on

efficient egg production but also the pathology caused by schistosomes is in large parts driven by the host immune reaction against the eggs of these parasites and the resulting inflammation and formation of granulomas in the liver and spleen (Gryseels *et al.*, 2006). For these reasons, the process of pairing-induced maturation is an attractive target for intervention and an intensely studied aspect of schistosome biology (Loverde & Chen, 1991).

1.3 Impact on human welfare

According to the WHO (2018) Approximately 200 million people required treatment for schistosomiasis in 2016, of which only about 90 million receive it. Furthermore the WHO (2018) estimates that approximately 90% of those infected live in Africa, with limited access to medical care. Hence the number of schistosomiasis-related deaths is considerable, but much of it impact is related to chronic morbidity (WHO, 2016). To accurately reflect the impact of a disease not just in terms of lifes lost but also taking into account the morbidity caused by diseases such as schistosomiasis a disability weight is used. This is a weight factor reflecting the magnitude of the health loss associated with the5 disease on a scale from 0-1, representing full health and death respectively (Salomon *et al.*, 2015). As part of the Global Burden of Disease Study, schistosomiasis was attributed a disability weight of 0.6% for the year 1990 (Murray & Lopez, 1997), although other studies put the disability weight as high as 2-15% (King *et al.*, 2005). In the most recent Global Burden of Disease Study, the estimate of total disability-adjusted life years (DALYs) lost due to schistosomiasis was raised from 2.125 million to 3.309 million DALYs (Murray *et al.*, 2012), reflecting the higher

estimates of disease prevalence as well as revised estimates of mortality and morbidity as measured by DALYs. Schistosomiasis transmission has been reported in 78 countries, including European countries such as Portugal and more recently France (Berry *et al.*, 2014). Schistosomiasis is considered to be endemic in 52 out of 78 countries (WHO, 2018). Although sub-Saharan Africa is most affected, schistosomiasis is also found in countries along the Atlantic coast of South America and the Caribbean as well as several Southeast Asian countries (Gryseels *et al.*, 2006).

1.3.1 Chemotherapy and control of schistosomiasis

Currently the only drug available to fight schistosomiasis effectively is praziquantel, which is used heavily in large parts of Africa with the declared aim of morbidity control (WHO, 2018). That means that rather than aiming to eradicate the disease in these areas, the aim is to reduce the intensity of the infection in the population and thereby lower the frequency of complications associated with schistosomiasis. The WHO considers schoolchildren to be target populations for these mass treatment programs, as well as at-risk adults such as fishermen or farmers who work in irrigation ditches and communities who heavily depend on water in endemic regions (WHO, 2018). Evidence has shown that pre-school children can also carry high intensity infections. For example Dabo *et al.* (2011) found between 15.7% of preschool children in different villages in Mali to have high intensity infections, determined by an egg count of ≥50eggs/ml in their urine. However, there are concerns about the safety of praziquantel in very young children due to a lack of reliable safety data (Stothard

et al., 2013). This mass treatment strategy, although representing the most cost-effective method in high-prevalence settings (WHO, 2018), has major limitations: Firstly, it does not eradicate schistosomiasis in the regions covered but, as stated above, aims to control morbidity. Secondly, it leaves the vector population unaffected, including infected snails. For this reason there is a risk of reinfection and so mass treatment needs to be repeated regularly to provide long-term benefit, which puts further strain on limited resources (WHO, 2018). Thirdly, due to mass drug administration, there is an increased pressure on parasite populations to become resistant to praziquantel - indeed reports of reduced sensitivity to praziquantel have been published in recent years (Crellen *et al.*, 2016; Wang *et al.*, 2014).

Unfortunately, despite decades of research into the development of a schistosomiasis vaccine, no successful formulation has been yet found (Hagan & Sharaf, 2003; McManus & Loukas, 2008). There is evidence that protective immunity against percutaneous infection with schistosomes can be raised with radiation-attenuated cercariae (Bickle, 2009; El Ridi & Tallima, 2015). Reductions in worm burdens up to 40-60% in treated mice were observed (Bickle *et al.*, 1979). However, no single- or multi-valent vaccine tested to date has been able to advance from clinical trials. As a result, the search for viable vaccine targets continues including "–omics" approaches such as transcriptomics, proteomics but also glycomics and immunomics for candidate generation (Loukas *et al.*, 2011) and several promising candidates are being investigated (Merrifield *et al.*, 2016). These include a fatty acid binding protein (Sm14), tetraspanin-2 (SmTSP-2), glutathione S-transferase (Sh28GST) and

calpain (Sm80) (Tebeje *et al*., 2016). While vaccines using Sm14, Sh28GST and SmTSP-2 have already been tested in humans (Bourke *et al.*, 2014; Merrifield *et al.*, 2016; Santini-Oliveira *et al.*, 2016) the vaccine using calpain has been tested in mice and baboons (Ahmad *et al.*, 2010; Ahmad *et al.*, 2011) and is now undergoing preparations for human trials (Rojo *et al.*, 2017).

1.3.2 Pathology of schistosomiasis

Acute pathology

Individuals who are exposed to schistosomal cercariae for the first time, such as travellers or migrants are prone to developing acute symptoms in response to infection (Paris & Caumes, 2010; Ross *et al.*, 2007). It has been suggested that the low rate of acute schistosomiasis in exposed populations, especially children, is due to a combination of underdiagnosis and *in-utero* sensitization (Gryseels *et al.*, 2006). Within the first hours or days after penetration of the skin by cercariae, some may develop a temporary rash or lesions at the site of penetration (Gryseels *et al.*, 2006). Over the weeks or sometimes months following the infection, some individuals may experience an acute systemic hypersensitivity to the developing worms also known as Katayama fever (Ross *et al.*, 2007); symptoms include fever, fatigue, myalgia, malaise, cough, eosinophilia, and lung lesions (Gryseels *et al.*, 2006). Although most patients recover within 2-10 weeks, some go on to develop abdominal symptoms as well, including weight loss, diarrhoea, abdominal pain, hepatosplenomegaly, and rashes (Bottieau *et al.*, 2006).

Chronic pathology

As is the case for many infectious diseases, the severity of symptoms is determined by a combination of several factors: the intensity of infection, the host immune response and environmental factors. The sympthoms of chronic schistosomiasis include abdominal pain, diarrhea and rectal bleeding (or haematuria in the case of an *S. haematobium* infection). Over the course of the infection the patients can develop anaemia, hepatosplenomegaly, an enlargement of liver and spleen as the result of fibrosis, and periportal fibrosis.

Most of the chronic pathology caused by the infection is not inflicted by the adult worms. Instead it results from the host immune response against the eggs during their perivesicular or periintestinal migration or once they have become trapped in organs such as liver, spleen and lung (Gryseels *et al.*, 2006). The human immune system responds to *S. mansoni* infections in a T helper 2-associated manner (Colley & Secor, 2014). This includes the use of parasite-specific antibodies, eosinophils and alternatively activated macrophages (Barron & Wynn, 2011; Colley & Secor, 2014), driven by the release of interleukins 4 and 13 (Pagán & Ramakrishnan, 2018). These interleukins help to modulate the immune repose to a state of chronic inflammation and the formation of granulomas around *S. mansoni* eggs (Colley & Secor, 2013). The eggs play an active role in this process through the secretion of proteolytic enzymes that provoke an eosinophilic pro-inflammatory and granulomatous response (Cass *et al.*, 2007; Jenkins *et al.*, 2007; Everts *et al.*, 2009). In the case of *S. haematobium*, the deposition of eggs around the bladder can also result in squamous cell carcinoma of the bladder (Nesi *et al.*, 2015). In the liver, this immune response can lead to

fibrosis and later cirrhosis (Wynn, 2015). This, in turn, can result in potentially lethal pulmonary hypertension and oesophageal varices (Colley & Secor, 2013). In other parts of the body this immune response can cause genital, urinary, intestinal and pulmonary schistosomiasis as well as neuroschistosomiasis (Gryseels *et al.*, 2006). Individuals with genital schistosomiasis may have symptoms including inflammatory lesions in reproductive organs that can lead to infertility as well as increasing the transmission of sexually transmitted infections, such as HIV (Gryseels, 2012).

Apart from the potentially irreversible damage to the internal organs, schistosomiasis can have a range of more subtle effects. These effects include, but are not limited to, fatigue, stunted physical and cognitive development of young children, as well as anaemia and exacerbation of malnutrition (Mcgarvey, 2000; Sousa-Figueiredo *et al.*, 2012). As these conditions affect large numbers of people in endemic areas, it becomes clear that successful intervention would be of great benefit to these communities (Gryseels *et al.*, 2006).

1.4 Male-female interaction

The dependence of female *S. mansoni* on pairing with a male for sexual maturation is long established (Kunz, 2001). Mating usually occurs in a parallel fashion (anterior to anterior), but about 12% of *S. mansoni* pairs are found to be paired in a reverse orientation (Faust *et al.*, 1934). Although it is usual for *S. mansoni* females to pair with one male only, it has been observed, both with freshly perfused worms as well as *in vitro*, that multiple males will clasp onto a

single female. (Basch, 1991). In other members of the family *Schistosomatidae*, the number of females paired with a single male can be greater. For example, Chu & Cutress (1954) found male *Austrobilharzia variglandis* to be paired with 3 to 8 females at once. Furthermore, Armstrong (1965) found a male *Heterobilharzia americana* to be paired simultaneously with 29 female *S. mansoni*, all of which appeared to be sexually mature. Male *S. mansoni* have a propensity to pair and in the absence of females will clasp onto objects of similar shape and size to female worms, *e.g.* made from cotton or agar (Basch & Nicolas, 1989). Even *in vitro,* worms pair or re-pair quickly: Michaels & Prata (1968) observed that when keeping male and a female worms together *in vitro*, the proportion of paired worms increased steadily. Out of a total of 140 potential pairs, after 1 day in culture 20.7% of worms were paired, 84.2% after 2 days, 94.5% after 3 days and 100% after 5 days .

When trying to determine the nature of the stimulus that leads to female sexual maturation, the first processes to be examined were insemination as well as the transfer of metabolites, such as carbohydrates, cholesterol, lipids, amino acids and polypeptides, to the female (Basch, 1991). Shaw (1977) demonstrated that pairing alone is enough to induce some development of vitelline tissue and occasionally even formation of an egg, without insemination. To do so, he allowed females from single sex (SS) infection and males from MS infections to pair, noting changes in female morphology within 5-6 days, whilst showing that ~85% of females remained unfertilised up to this point. Michaels (1969) also showed that males irradiated with X-rays readily paired and induced female maturation. Even without intact testes, this lead to successful female sexual

development and egg laying. The male stimulus was also shown not to be under the control of the cerebral ganglia of the nervous system, as decapitated males are still able to induce female maturation (Popiel & Basch, 1984); however, as the authors stated, this does not unequivocally rule out involvement of the central nervous system, as parts of the nervous system in the rest of the body were found to remain functional (Popiel & Basch, 1984). Furthermore, the stimulus is also not species-specific as females paired with males of other species (heterospecific pairing) will produce eggs, albeit at a reduced rate (Kunz, 2001).

Although the nature of the stimulation of female worms is unclear, it requires contact, occurs along the whole gynecophoral canal and is local, *i.e.* is not propagated longitudinally through the body of the female worm (Basch, 1991). The nature of the stimulus has been speculated to be physical, for example tactile, or involve transfer of hormones, nutrients and other signalling molecules (LoVerde *et al.*, 2005). Several studies have provided evidence for nutrients playing a key role in supporting female fertility (Pearce & Huang, 2015), especially the availability of red blood cells (Wang *et al.*, 2015) and fatty acids (Huang *et al*., 2012). *S. mansoni* cannot synthesise fatty acids, relying instead on their host to provide them (Meyer *et al.*, 1970). Metabolites such as glucose (Cornford & Huot, 1981) and cholesterol (Popiel & Basch, 1986) can be transferred through the gynecophoral canal to the female which may help to fulfil the female nutritional demands.

1.5 Molecular biology of male-female interaction

Several signalling pathways thought to be involved in stimulating the growth and maturation of female *S. mansoni* have been characterised, including the transforming growth factor (TGF)-β/SMAD and the MAPK pathway (Beckmann *et al.*, 2010). In the context of female maturation they serve to transmit growth-inducing stimuli from the male worm to the nuclei of cells in the vitellarium and ovaries (Beckmann *et al.*, 2010).

The TGF-β pathway plays an important role in the development of multicellular organisms (Osman *et al.*, 2006). TGF-β is a polypeptide and a cytokine that helps to regulate cell proliferation, differentiation and apoptosis (Massagué, 2012). Signalling is initiated by transmembrane serine/threonine kinase receptors, the TGF-β receptors I & II (TβRI & TβRII) (Osman *et al.*, 2006). TGF-β signalling also relies on SMAD proteins, which were first detected in *Drosophila melanogaster* (Sekelsky *et al.*, 1995) and *Caenorhabditis elegans* (Savage *et al.*, 1996) where they were termed "mothers against decapentaplegic" (MAD) and "small body size" (SMA), respectively. Proteins found to be homologues of these two are now called SMAD by merging the two names. SMADs form trimers of two R-SMAD and one co-SMAD molecule. These act as transcription factors to activate expression of downstream genes (Shi & Massagué, 2003). Homologues of TβRs, SMADs, as well as of several other proteins that form part of the intracellular TGF-β signalling pathway, have been discovered in schistosomes (Loverde *et al.*, 2007). These include the TβRI-regulating protein FKBP12 as well as the proteins SMAD2, a receptor regulated SMAD (R-SMAD) (Wu *et al.*, 2001) and SMAD4, a

common-mediator-SMAD (co-SMAD) (Shi *et al.*, 1997). Many of these genes have been shown to be expressed in the ovaries and vitellarium of female and in the testis of male worms (Beckmann *et al.*, 2010).

To demonstrate the importance of TGF-β signalling for female reproductive activity, Knobloch *et al.* (2007) exposed *S. mansoni* pairs *in vitro* to the synthetic TβR-I inhibitor TRIKI and demonstrated a reduction in mitotic activity the vitelline cells of females by 46%, as well as a 28% reduction of laid eggs. However, as higher concentrations of TRIKI did not lead to a greater reduction in mitosis and egg laying, the authors concluded that alternative signalling pathways must be involved in the development of the vitellarium (Knobloch *et al.*, 2007).

Using another inhibitor, the Src-kinase specific herbimycin A, Knobloch *et al.* (2007, 2006) found egg laying *in vitro* to be reduced by 40-60% and mitotic activity in females (but not males) reduced by 25-75% in a dose-dependent manner. Src kinases are cytoplasmic tyrosine kinases (CTK) and hence belong to the superfamily of protein tyrosine kinases (Beckmann *et al.*, 2010). CTKs are divided into eight subcategories based on the presence or absence of domains as well as their structural arrangements and several members of the CTKs have been discovered and found to be expressed in schistosome reproductive tissues (Beckmann *et al.*, 2010).

The Src kinases SmTK3 (Kapp *et al.*, 2004) and SmTK6 (Beckmann *et al.*, 2011) have been localised to the gonads of *S. mansoni*, however, SmTK6 expression was

found to be absent from the vitellarian tissue (Beckmann *et al.*, 2011). The Fyn-type Src kinase SmTK5 is expressed in the gonads and vitellarium, and SmTK4, which belongs to the class of Syk kinases, is also expressed in the gonads but not the vitellarium (Knobloch *et al.*, 2002).. SmTK3, SmTK4 and SmTK6 weree shown to interact directly with one another to regulate downstream signalling using yeast-2-hybrid screening (Beckmann *et al.*, 2010; Kapp *et al.*, 2004). Syk kinases regulate cell proliferation and progression through the cell cycle by activating expression of mitotic genes (Beckmann *et al.*, 2010) but also by regulating cytokinesis and the cytoskeleton rearrangements (Mócsai *et al.*, 2010). Syk kinases are activated by receptor tyrosine kinases (RTK), such as the epidermal growth factor receptor (EGFR) (Vicogne *et al.*, 2004) or venus kinase receptors (Vicogne *et al.*, 2003). Sky kinase activation leads to the activation of a cascade of mitogen activating protein kinases, known as the MAPK signalling pathway (Qi & Elion, 2005). At least 15 RTKs have been identified in the *S. mansoni* genome: This includes four EGFRs, two insulin receptors, two fibroblast growth factor receptors (FGFR), an Ephrin receptor, a RTK-like orphan receptor, a muscle-specific kinase receptor, a cholecystokinin tetrapeptide receptor families as well as an unknown receptor (Andrade *et al.*, 2011; Avelar *et al.*, 2011). A variety of extracellular stimuli can activate the MAPK pathway through RTKs in the cell membrane. Through a series of kinases activating one another, a MAPK such as the extracellular responsive kinase (ERK) is activated (Qi, 2005). ERK translocates into the nucleus where it phosphorylates the transcription factor ELK-1, thereby allowing it to form a heterodimer with the serum response factor and activate transcription by binding to the promoter of target genes such as c-Fos in humans (Cavigelli *et al.*, 1995).

Several components of the MAPK signalling cascade were discovered in the *S. mansoni* genom by *in silico* analysis (Andrade *et al.*, 2011) and their biological role has been further characterised since then: In planarians, ERK is now known to play a central role in the regeneration of tissue by stem cells (Tasaki *et al.*, 2011), a process thought to be similar to the growth of the female reproductive organs in *S. mansoni* in response to pairing (Collins & Newmark, 2013), whereas in *C. elegans* the MAPK pathway is known to be important in several developmental events as well as establishing meiosis in the germ line (Sundaram, 2006). Andrade *et al.* (2014) demonstrated that RNAi using long dsRNA molecules can be used *in vivo* in S*. mansoni* to investigate the role of the MAPK pathway. They successfully knocked down ERK and found no effect on worm survival but a 44% reduction in hepatic egg count relative to a negative control, suggesting that MAPK may play a role in female fertility.

Other proteins involved in regulation of the MAPK pathway for which homologues have been found in *S. mansoni* include the adapter proteins SHC and growth factor receptor-bound protein 2, which act by binding to the cytoplasmic portion of RTKs, and control recruitment of "Son of sevenless" (SOS), a guanine nucleotide exchange factor protein (Knobloch *et al.*, 2007). This in turn activates the membrane associated GTPase called Ras by catalysing the exchange of GDP with GTP, thereby activating Ras which goes on to activate the MAPK cascade (Beckmann *et al.*, 2010). Another protein, the GTPase-activating protein (GAP), promotes the exchange of GTP with GDP, thereby leading to deactivation of Ras, causing deactivation of the MAPK pathway (Beckmann *et al.*, 2010). However,

activated Src kinases can phosphorylate GAP to down-regulate its inhibitory effect on the MAPK pathway (Downward, 1997; Qi, 2005).

In addition to RTKs, the integrin receptors α and β have been found to activate focal adhesion kinase (FAK), another CTK (Guan, 1997; Mitra & Schlaepfer, 2006). Integrins are composed of two non-covalently associated transmembrane glycoprotein subunits, α and β, and primarily serve to attach the cytoskeleton to the extracellular matrix (Alberts *et al.*, 2007). They also play an important role in signalling by passing on extracellular stimuli to the FAK and often cooperate to promote cell growth, survival and proliferation (Alberts *et al.*, 2007). *S. mansoni* integrin receptors and their role in reproduction were described by Beckmann *et al.* (2012) who found that integrin-β-1 was expressed in the reproductive organs of *S. mansoni*, co-localising with SmTK3, SmTK4 and SmTK6. Using a yeast-2-hybrid screen the same authors showed that integrin-β-1 interacts with the SH2 domain of SmTK4 and with the SH3 domain of SmTK3 and SmTK6 (Beckmann *et al.*, 2012). Recently, the interaction between integrins and the Venus Kinase Receptor 1, a RTK, was shown to regulate proliferation of oocytes by apoptosis in *S. mansoni* (Gelmedin *et al.*, 2017).

Integrins often occur in complexes with tetraspanins (Tugues *et al.*, 2013). These are a group of proteins characterised by four transmembrane domains, which form so called tetraspanin enriched microdomains which allow other molecules to assemble into signalling complexes (Tugues *et al.*, 2013). Several tetraspanin homologues have been identified in *S. mansoni*, including members of the CD63 family of tetraspanins (Cogswell *et al.*, 2012; Tran *et al.*, 2010). One of the CD63

homologues is known to be expressed in a female-specific manner (Cogswell *et al.*, 2012). In humans, CD63 has been shown to down-regulate apoptosis by interacting with the protein "tissue inhibitor of metalloprotease-1" (Jung *et al.*, 2006).

Another group of kinases found in *S. mansoni* are polo like kinases (Long *et al.*, 2010), also known as Cdc5, which have been shown to regulate cell cycle progression in a wide range of organisms such as *Saccharomyces cerevisiae*, *D. melanogaster*, *C. elegans* and humans (Chase *et al.*, 2000; Llamazares *et al.*, 1991; Toczyski *et al.*, 1997). Polo-like kinases have a serine/threonine kinase domain as well as two signature motifs, known as polo boxes (Barr *et al.*, 2004). In model organisms they have been found to tightly control the G2/M transition in the cell cycle by phosphorylation of Cdc25 which in turn activates CDKs by dephosphorylation (Roshak *et al.*, 2000). This process drives the activation of mitotic events including centrosome maturation, bipolar spindle formation, activation of the anaphase promoting complex, chromosome segregation and actin ring formation in preparation for cytokinesis (Van De Weerdt & Medema, 2006). Polo-like kinases themselves are regulated by a group called polo-like kinase kinases, one member of which has been found in *S. mansoni* and is a homologue of human protein Ste20 (Yan *et al.*, 2007). Ste20-like kinases belong to the germinal center kinase subfamily and their function, like that of the integrin receptors, relates to the regulation of actin-mediated processes including cell shape and cytoskeletal rearrangements (Belkina *et al.*, 2009).

1.6 Genomics & Transcriptomics

1.6.1 Genome

S. mansoni is a diploid organism with seven pairs of autosomes, a pair of sex chromosomes and a mitochondrial genome (Swain *et al.*, 2011). Schistosomes have a ZW determination, meaning that the male is homogametic (ZZ) and the female heterogametic (ZW) (Short & Grossman, 1981).

Berriman *et al.* (2009) found the nuclear genome to be 30% larger than previously estimated, around 374.9 Mb, approximately one ninth of the human genome but considerably larger than the genome of *C. elegans* (~98 Mbp) (*C. elegans* Sequencing Consortium, 1998) or *D. melanogaster* (~120 Mbp) (Adams *et al.*, 2000). 40% of the *S. mansoni* genome is made up of repetitive sequences, including 72 families of long-terminal repeat (LTR) and non-LTR transposons each, comprising 15% and 5% respectively of the whole genome. In the most recently published version of the genome its size was revidsed down to 364.5 Mb and the number of genes now stands at 10,852, having been revised down from 11,807 in the original genome paper, after more than 45% of gene models were modified (Protasio *et al.*, 2012). These advances were due in part to RNA-Seq evidence that allowed for improved gene predictions and due to a less fragmented genome with fewer duplicated sequences that had resulted from polymorphism (Protasio *et al.*, 2012). The current genome assembly consists of 9203 contigs with an average length of 39.4 kb, a N50 length of 78.3 kb – a measure of assembly quality indicating that 50% of the genome sequence was

contained on contigs greater or equal to 78.3 kb – and 885 scaffolds with an average length of 411.9 kb and a N50 length of 32.1 Mb (Protasio *et al*., 2012).

Assembly of the Z and W chromosomes proved to be difficult as there are large regions of homology where the chromosomes are indistinguishable the Z and W chromosomes assembled together in 34 scaffolds with 59 Mb in total (Protasio *et al.*, 2012). The same authors used differences in coverage of male and female DNA to identify female-specific regions of the W chromosome. Overall, 30% of the Z/W chromosome sequence was unique to the Z chromosome, and 23 Z-specific genetic markers were identified in the assembled Z/W scaffolds (Protasio *et al.*, 2012). 114 unplaced scaffolds, totalling approximately 1.1 Mb, were found to be W-specific but contained approximately 90% repetitive sequence including known female-specific repeats (Portela *et al.*, 2010).

1.6.2 Transcriptomics of sex-specific schistosome biology

Due to its relevance for pathology and transmission, understanding the sexual dimorphism of schistosomes is critically important. This aspect of schistosome biology was and continues to be studied using different methods of transcriptome analysis (for example (Fitzpatrick *et al.*, 2005; Lu *et al.*, 2017)). In recent decades, several new methods of transcriptome analysis have become available, including methods based on quantitative real time PCR (qRT-PCR), expressed sequence tags (EST), serial analysis of gene expression (SAGE), cDNA microarrays and later RNA–seq as well as variations of these methods (Lowe *et al*., 2017). In the following paragraphs, I will summarise schistosome

transcriptomics, the development of methods and their use to address increasingly detailed biological questions.

Early papers to examine the differential expression of genes in male and female *S. mansoni* examined individual genes such as actin (Davis *et al.*, 1985) or small groups of genes, due to a lack of high-throughput methods. In this case, Davis *et al.* (1985) used northern blot analysis to compare mRNA levels. The authors showed that two actin transcripts were more abundant in male worms than in females, cercariae and eggs. Using the same methodology the transcriptoion of the gene coding for the egg shell protein p48 was studied (Chen *et al.*, 1992). The authors found it to be exclusively expressed in paired female worms. Furthermore Grevelding *et al.* (1997) identified three more genes, encoding egg shell protein p14, ferritin-1, and mucin-like protein A11, which they showed to only be expressed the vitelline tissue of paired female worms.

Thanks to the progress in human transcriptomics (Adams *et al.*, 1991, 1992, 1993; Hoog, 1991), increasing availability of genomic data (Franco *et al.*, 1995), programs and resources such as the WHO/UNDP/World Bank Schistosoma Genome Network (http://www.nhm.ac.uk/hosted_sites/schisto), the *Schistosoma mansoni* EST Genome Project (http://verjo18.iq.usp.br/schisto/) and the Institute for Genomic Research *Schistosoma mansoni* genome project (http://www.tigr.org/tdb/e2k1/sma1/index.shtml), it became feasible to design *S. mansoni* specific microarrays. This allowed scientists to examine hundreds of genes at the same time (Lowe *et al.*, 2017). Hoffmann *et al.* (2002) used a cDNA microarray with 576 probes and were able to double the number of genes

known to be differentially expressed between male and female schistosomes, identifying sex-specific expression for twelve female- and four new male-specific transcripts.

One year later, 16000 ESTs and 163 full-length cDNA had been sequenced, when Verjovski-Almeida *et al.* (2003) generated ten times more ESTs and provided the first sequencing-based comprehensive overview of a schistosome transcriptome. However most transcripts were still not sequenced along their full length, instead rather small, unique nucleotide sequences (tags) were sequenced, which allowed the corresponding genes to be identified (Lowe *et al.*, 2017)l.

Many of the subsequent publications that examined differences between male and female worms as well as paired and single worms used microarray or tag sequencing methods: Between 2004 and 2014 several such papers were published (Fitzpatrick *et al.*, 2004; Fitzpatrick *et al.*, 2005; Fitzpatrick & Hoffmann, 2006; Leutner *et al.*, 2013; Moertel *et al.*, 2006; Sun *et al.*, 2014) leading to a further expansion of genes known to be gender- or pairing-associated. When examining differences between male and female worms from the genus *Schistosoma*, the number of gender-associated genes discovered by such undertakings thus increased from 16 (four male- and twelve female-specific genes (Hoffmann *et al.*, 2002) to 30 (eight male- and 22 female- associated genes) in *S. japonicum* (Fitzpatrick *et al.*, 2004), then later 227 (86 male- and 141 female-specific genes) in *S. mansoni* (Fitzpatrick *et al.*, 2005) and finally 2179 (1163 male- and 1016 female-specific genes in *S. japonicum* (Moertel *et al.*, 2006) genes, an 136-fold increase of genes known to be differentially expressed

between male and female worms within four years.

Due to these investigations, we now know the transcripts most differentially expressed between sexually mature male and female worms. These differences closely reflect the biology of this parasite, with sexually mature females up-regulating transcritption of genes associated with egg production, such as the major egg antigens other eggshell proteins, p48 (chorion), as well as tyrosinases (Fitzpatrick & Hoffmann, 2006). Male-specific genes on the other hand are dominated by tegument and muscle genes (e.g. tetraspanin genes, tegumental antigen as well as actin, dynein light chain 3, myosin and tropomyosin) (Fitzpatrick *et al.*, 2005, 2004; Moertel *et al.*, 2006). These may reflect the role of the male worm in anchoring the female inside blood vessels, moving her to suitable egg deposition sites (Fitzpatrick *et al.*, 2005).

The role of pairing on gene expression and the resulting sexual maturation has also generated considerable interest (Fitzpatrick & Hoffmann, 2006; Leutner *et al.*, 2013; Sun *et al.*, 2014). A number of different comparisons have been made: first Fitzpatrick & Hoffmann (2006) used a cDNA microarray to examine differential gene expression in seven week old adult males and sexually mature female from mixed sex (MS) and females from single sex (SS) infections. This allowed the author to identify 245 genes differentially expressed between females from MS infections (64 genes) and SS infections (181 genes), and 138 138 genes differentially expressed between males from equivalent infections (104 genes up-regulated in males from SS and 34 in males from MS infections). In 2013, Leutner *et al.* used a combination of oligonucleotide microarray and

SuperSAGE (a form of Serial Analysis of Gene Expression using high-throughput sequencing (Lowe *et al.*, 2017)) to examine males from MS and SS infections (using the terms pairing-experienced and inexperienced). Their microarray analysis allowed them to discover a total of 526 genes, which were differentially expressed in MS (229 genes) and SS males (297 genes) whereas their SuperSAGE analysis only detected 288 genes (253 genes in MS males and 35 in SS males). The discrepancy was explained partially by the fact that expression for over 4000 (out of a total 6326 examined) genes was only detected by one of the two methods as well as the use of stringent significance thresholds (Leutner *et al.*, 2013). The intersection of those two data sets contains a total of 29 genes, 21 of which had functional annotation at the time, from which 16 genes were up-regulated in males from SS, and 5 in males from MS infections (Leutner *et al.*, 2013). Sun *et al.* (2014) examined *S. japonicum* at 18 and 23 d.p.i. to get an insight into the early changes induced by pairing, as well as the changes taking place in females from SS and MS infections between 18 and 23 d.p.i. At 18 d.p.i. they found a total of 318 differentially expressed genes (241 in females from SS and 77 in females from MS infections). However, at 23 d.p.i. the number of differentially expressed genes had risen to 3446 genes (533 in females from MS and 2913 in females from SS infections) (Sun *et al.*, 2014). The numbers of genes differentially expressed between females from 18 and 23 d.p.i. are also large: for females from SS infections, 2545 genes were found to be differentially expressed (2193 up-regulated in females from 18 d.p.i. and 352 up-regulated in females from 23 d.p.i.) and for females from MS infections, 1998 genes were found to be differentially expressed (740 up-regulated in females at 18 d.p.i. and 1258 up-regulated at 23 d.p.i.) (Sun *et al.*, 2014). Due to an apparent lack of replicates in

this study, it is hard to estimate how much biological variability there is, but a general trend can be seen clearly: The transcriptome of paired female worms (MS infections) becomes increasingly divergent from females that remain unpaired (SS infections) (Sun *et al*., 2014). As paired females mature sexually the differences between them and females from single sex females become more noticeable both morphologically (Fitzpatrick & Hoffmann, 2006), but also at the transcriptome level and these differences increase as pairing continues until the paired female reaches complete sexual maturity around 5 weeks p.i. (Basch, 1991).

Recently, Wang *et al.* (2017) published a large RNA-Seq study in which they presented evidence that biogenic amines serve as the male stimulus for female sexual maturation. Having sequenced 48 *S. japonicum* transcriptomes from eight time points in the intra-mammalian stage (between 14-28 d.p.i.) and both sexes, with three biological replicates each, these authors found the expression of an aromatic-L-amino acid decarboxylase (AADC) to increase leading up to pairing, and showed the expression of AADC to be located in the gynecophoral canal by using whole mount *in situ* hybridisation (Wang *et al.*, 2017). Furthermore the authors found an allatostatin-A receptor-like gene (AlstR), coding for an insect-like hormone receptor, to be involved in the maturation of female reproductive organs (Wang *et al.*, 2017). Indeen, when AlstR was knocked down using RNAi, ovaries and vitelline glands failed to mature, in contrast to control females (Wang *et al*., 2017). Another study found evidence that histone modifications (H3K27me3) are different between male and female *S. mansoni* (Picard *et al.*, 2016). The authors found these epigenetic modifications to correlate with gene

expression showed there to be a sex-biased dynamic of histone methylation, in particular a depletion of H3K27me3 along the transcription unit in male worms (Picard *et al.*, 2016). However, as all RNA-Seq data in this study was derived from worms of SS infections, it remains to be seen if these results hold true for sexually mature females as well.

Publication	Method	Species	Number of probes/read	Comparison	Results
Hoffmann *et al.*, 2002	cDNA microarray	*S. mansoni*	576 ESTs	Adult male vs female	23 DEGs*
Verjovski-Almeida *et al.*, 2003	cDNA EST sequencing	*S. mansoni*	163000 ESTs sequenced	N/A	N/A
Fitzpatrick *et al.*, 2004	cDNA microarray	*S. japonicum*	457 probes	Adult male vs female	30 DEGs
Fitzpatrick *et al.*, 2005	Oligonucleotide microarray	*S. mansoni*	7638 probes	Adult male vs female	197 DEGs
Moertel *et al.*, 2006	Oligonucleotide microarray	*S. japonicum*	19221 probes	Adult male vs female	647/700 DEGs**
Fitzpatrick & Hoffmann, 2006	DNA microarray	*S. mansoni*	7638 probes	Single vs paired adult female	248 DEGs
				Single vs paired adult male	138 DEGs
Leutner *et al.*, 2013	Oligonucleotide microarray & SuperSAGE	*S. mansoni*	19197 probes; Number of sequenced tags not stated	Single vs paired adult male	29 DEGs
Sun *et al.*, 2014* **	Tag-Seq	*S. japonicum*	> 3 million tags sequenced	Paired vs single female, 21 d.p.i.	3446 DEGs
Picard *et al.*, 2016	RNA-seq	*S. mansoni*	65.8 million reads per sample	Male vs female: Cercariae, Schistosomule, Adult worms	7168 DEGs
Wang *et al.*, 2017	RNA-seq	*S. japonicum*	Number of sequenced reads not stated	Time course - Male	1934 DEGs
				Time course - Female	6535 DEGs

*) Including positive controls; **) Results for Phillipine and Chinese Strains respectively; ***) Not all comparisons summarised here

Table 1.1: Summary of the published literature on schistosome transcriptomic discussed in this chapter. The table provides the name of the first author and year of publication, the methods of transcirptomics used, the Schistosoma species examined in a given publication, the number of probes used (in the case of microarrays) or DNA reads sequenced (in the case of RNA-seq), the samples which were compared and finally the number of differentially expressed genes (DEGs) observed.

1.6.3 Methods of transcriptomics

This thesis relies heavily on gene expression data, including RNA sequencing (RNA-Seq) first and foremost, but also used quantitative reverse transcription polymerase chain reaction (qRT-PCR) data, which is often considered the gold standard for comparing gene expression, for validation of RNA-seq data. Microarray data is only used to briefly, by correlating previously published microarray data to RNA-seq data. The following section introduces these methods for measuring changes in the transcriptome as well as their strengths and weaknesses.

Quantitative reverse transcription PCR

qRT-PCR is used frequently to compare expression of small numbers of genes between different samples. After reverse transcription of mRNA, the cDNA is used as template for a PCR reaction where, over the course of several rounds of amplification, the quantity of target cDNA is measured. This can for example be done using SybrGreen, a compound that fluoresces when bound to DNA. Over successive rounds of PCR this results in stronger fluorescence as more DNA has been produced. The quantities of DNA in the sample and control can then be compared to determine the relative abundances of the target sequence in the

cDNA. The method most commonly used for inferring differences in expression between samples is called ΔΔCt Calculation Method, and it makes use of one or more internal reference (IR) genes thought to be expressed at identical levels across all samples, to control for differences in the quantity of input cDNA (Kubista *et al.*, 2006). The IR is used for normalisation of the amount of input cDNA between different samples, so that any differences between target gene expression observed after normalisation represents the treatment effect (see Chapter 2.2.5). qRT-PCR has been widely used to measure gene expression in schistosomes. For example it has been used to identify or validate gender-specific gene expression by comparing male and female samples (Hoffmann *et al.*, 2002), pairing-induced gene expression by comparing cDNA samples from paired to those of virginal females (Fitzpatrick & Hoffmann, 2006). Another example is the use of qRT-PCR to compare gene expression in different organs (Hahnel *et al.*, 2014) and in RNAi knock-down experiments (Tran *et al.*, 2010).

Microarray

cDNA microarrays are a collection of DNA probes which are immobilised on a small chip to allow sample RNA (after reverse transcription to cDNA) to be bound by hybridisation with the complementary probe sequences. When applying sample cDNA to the microarray, the relative abundance of sample cDNA molecules is usually measured using fluorophore- or chemiluminescence-labelled probes (M. B. Miller & Tang, 2009). It is worth noting that reference sequences have to be known prior to the production of the microarray probes (Lowe *et al.*, 2017). This allows comparisons of transcript abundances of two or more experimental conditions. As in the case of qRT-PCR, stringent

normalisation is necessary to obtain accurate results. This includes measuring technical as well as biological variation across the transcriptome. Microarrays can be performed at very high throughput, due to the relatively low labour intensity (compared to RNA-seq), but they require a relatively large amount of mRNA input (about 1 µg of mRNA) (Lowe *et al.*, 2017). This can limit the application of microarray, if the biological samples are difficult to obtain and very small in size.

RNA sequencing

RNA-Seq relies on high-throughtput sequencing technology that can provide the sequences of vast numbers of short cDNA fragments to determine the abundance of cDNA molecules in each sample. Depending on the precise method used to make the cDNA library to be sequenced, the data may include information on the expression of not just messenger RNA (mRNA) but also non-coding RNAs such as ribosomal RNA (rRNA), transfer RNA (tRNA) and others. Careful normalisation is required for accurate results. The normalisation needs to correct for within-sample biases, caused by factors such as transcript length and GC-content, as well as between-sample biases, especially amount of sequenced cDNA. To correct within-sample biases RPKM (reads per kilobase of transcript per million mapped reads) or FPKM (fragments per kilobase of transcript per million mapped reads) are often chosen. This method scales the number of reads for a given gene to the length the given transcript as well as the overall number of reads across the whole transcriptome (Verk *et al.*, 2013). However, it has been demonstrated that this method skews calls for differential expression in favour of longer transcripts as the number of reads mapping to a gene will be proportional to its length and

higher read counts give more statistical power to determine significant differences (Oshlack & Wakefield, 2009). In cases where expression of a gene is compared across different samples (rather than different genes within one sample), only between-sample normalisation is necessary (Verk *et al.*, 2013). In such cases sequencing yield is the most important variable to take into account. The sequencing yield, i.e. the total number of transcript reads generated during sequencing, has to be considered during the normalisation as it often differs between replicates and treatment groups. Rather than dividing the read count per gene by the total number of reads for a given sample, more robust methods are used that can take extreme expression values into account (Love *et al.*, 2014). Such methods include the median-of-ratios (Love *et al.*, 2014) and the trimmed mean of M-values method (Robinson *et al.*, 2009). However in this context it is also important to consider the effect of relative abundances of mRNAs in a transcriptome. If the real expression of a target gene is identical in two samples, but its relative abundance is higher in one sample (for example due to a large number of other genes becoming down-regulated) the target gene will appear to be differentially expressed. Technical variation in transcriptome analysis is the measured differences introduced by the method or transcriptome analysis and sample handling, rather than differences in the underlying biology. Technical variation is thought to be smaller in RNA-Seq than in microarray technology (Marioni *et al.*, 2008), biological variation has to be measured carefully to allow statistical inference of differential expression of a gene between two samples (Love *et al.*, 2014). Furthermore, during the final steps of RNA-Seq library production, the cDNA is usually amplified using PCR. During this process some biases may be introduced as some GC-rich cDNA molecules don't become

amplified at the same rate as other molecules (Aird *et al.*, 2011). However, this effect should be negligible when comparing expression of the same gene across different samples, as these samples should experience the same amplification biases at that locus.

RNA-Seq offers several advantages over microarrays. Firstly, microarrays rely on a library of probes that is pre-determined and based on existing genome annotations (Lowe *et al.*, 2017). Thus genes that have not yet been discovered will be excluded from the analysis. Unlike microarrays, RNA-Seq data can be used to discover new transcripts and genes. This information can then be used to validate predicted gene models in the genome and contribute to improving the quality of genome annotation (Lowe *et al.*, 2017). Secondly, RNA-Seq allows to detect splice variants, which can remain undiscovered in microarrays as, in many cases, different variants will hybridise with the same probe. Another advantage of RNA-Seq over microarray data is a larger dynamic range, which means accurate quantification of genes with very low or high expression (Zhao *et al.*, 2014). In the absence of expression microarrays will often measure non-zero expression levels due to background noise such as non-specific hybridisation. Conversely at the high expression levels micro arrays can become saturated and unable resolve different expression levels anymore (Zhao *et al.*, 2014). This allows for a more quantitative measurement of low and very high abundance cDNA using RNA-Seq (Wang *et al.*, 2009).

Comparison of RNA-Seq and proteomics

A discrepancy between mRNA and protein levels is to be expected in any organism as there are many mechanisms regulating the abundance of proteins in a cell (Maier *et al.*, 2009). These include different efficiencies of mRNA cap binding proteins, the rate at which ribosomes are recruited to internal ribosome binding sites, codon usage, regulation by microRNAs, binding to RNA-binding proteins and mRNA stability which is mainly controlled by the mRNA poly-A tail (Vogel & Marcotte, 2012). Protein abundance is also regulated further down-stream of translation by degradation signals leading to ubiquitination and proteolysis (Vogel & Marcotte, 2012). All these factors combined mean that there is not a perfect correlation between mRNA and protein levels (Maier *et al.*, 2009). Schwanhaeusser *et al.* (2011) examined the correlation of steady-state mRNA and protein abundance in mouse fibroblasts, finding that, across 5028 genes, mRNA levels only explain between one and two thirds of the variation in protein levels observed (R^2=0.41). In human medulloblastoma cells, Vogel *et al.* (2010) reported that 66% of variation in protein levels could be explained by factors including mRNA concentration, UTR sequences and the coding sequence. Howver, these authors showed that mRNA concentration only explained 27% of the variance, 3' UTRs 8% and that 31% was explained by the coding sequence. Nonetheless, Vogel & Marcotte (2012) conclude that mRNA is a suitable proxy for the presence of a protein as mRNA levels detected by RNA-Seq directly correlated to the corresponding proteins in a proteomics experiment. However, work by Li *et al.* (2014) showed that due to problems with the normalisation of proteomics data in the aforementioned studies, mRNA concentrations predict around 84% of protein abundances (instead of 27%). Furthermore, the mRNA

and protein abundance were shown to correlate more strongly after when cells responded to a stimulus (as opposed to a being in a steady state) (Koussounadis *et al.* 2015). In such situations the mRNA abundance of differentially expressed genes is even better correlated to their protein products (Koussounadis *et al.* 2015).

1.7 Aims of my project

The aim of this thesis is to explore and characterise the differences between male and female worms during their intra-mammalian development at the transcriptome level. Using the vast quantities of expression data that can be generated using RNA-Seq and the unprecedented detail it affords, what factors involved in the process of female maturation can be identified from the transcriptome? How does this change our understanding of the role that the male-female interaction plays in the transmission and pathology of the disease? In particular, this thesis addresses the differential expression of genes between male and female schistosomes during their intra-mammalian development including the changes in gene expression induced by the male-female interaction. Three main topics are examined in this thesis.

The first is the development of male and female worms inside the mammalian host. Particularly, when and how do male and female worms start to diverge at the transcriptome level? And similarly, what are the early differences of female worms in MS infections compared to SS infections at the transcriptome level?

Furthermore, what roles do some of the genes identified play in the female fertility? These results are presented in Chapter 3.

Next I wanted to discover the effect of cell proliferation and apoptosis on the maintenance of sexual maturity as well as regression of female reproductive tissues. Furthermore, how does pairing regulate these processes? Using an *in vitro* culture system of female worms in the presence or absence of male worms, the changes in expression of fertility-related genes were examined as well as genes that control apoptosis and cell proliferation. Results from this work comprise Chapter 4.

In Chapter 5, the transcriptomes of *S. mansoni* testes and ovaries were examined. This work was performed to learn how gene expression changed in the gonads in response to pairing. As pairing induces proliferation of the ovaries, it was of particular interest to us which transcriptome changes allowed for this proliferation to take place. These changes are obscured in the whole worms transcriptome by the larger vitellarium, but can be better resolved by examining the gonad transcriptomes.

Finally, Chapter 6 provides a summary of the findings in this thesis and places them in the context of the current understanding of the biology of schistosomes.