CHAPTER 1

INTRODUCTION

1.1 Introduction

Schistosomiasis is a parasitic disease caused by platyhelminths of the genus *Schistosoma*. It has been estimated that ~780 million people live at risk of infection and ~200 million are infected (Steinmann *et al.*, 2006). The global distribution and infection rates of schistosomiasis (**Figure 1.1**) has varied little in the last 20 years (Davis, 2002; WHO, 2011). Only in Africa, where the greatest prevalence of infection occurs (WHO, 2011), it has been estimated that ~150,000 people die each year due to schistosomiasis-related causes (van der Werf *et al.*, 2003).

Symptoms of infection can be quite mild, which leads to long lasting infections often left undiagnosed. Continuous accumulation of parasite eggs in the liver causes hepatomegalia and liver failure. Infected patients are treated with praziquantel, which kills adult parasites and stops egg lying. This is orally administered, highly tolerated and cheap drug. However, this treatment does not prevent re-infection – very common in endemic areas and principally among young children - leading to long-term schistosomiasis infections with the concomitant establishment of chronic inflammation (Pearce *et al.*, 2002). This has a direct effect on morbidity, which contributes to the further impoverishment of the affected populations (King, 2010).

Without a vaccine, mechanisms of prophylaxis rely on tackling transmission through the reduction in the number of infected individuals, distribution of information regarding water contact habits and improvement of sanitary conditions in endemic areas (Davis, 2002). Although efforts have been implemented in these areas, the number of infected people has changed little over the last decades (Steinmann *et al.*, 2006).

In order to develop mechanisms of intervention against schistosomiasis infections, it is important that the process of infection is well characterised. The infectious agent for the human host is the cercariae. These microscopic free-living larvae are released by snail hosts and can infect humans by penetrating the skin and transforming into schistosomula. This transformation is characterised by phenotypic, metabolic and physiological changes that make the schistosomula a vulnerable stage in the parasite's life cycle.

The work presented in this thesis focuses on the characterisation of changes in gene expression occurring to the parasites during the transformation from the free-living cercariae to the parasitic schistosomula after 24 hours of infection.

The first step in the characterisation of such changes is the generation of reliable gene annotation (Chapter 3). Previous annotation has relied on *in silico* predictions and a limited dataset of experimental data. In this work, four time points of the parasite life cycle transcriptome were sampled and sequenced using high-throughput technology, which allowed an unprecedented level of accuracy to be obtained in the gene annotation of this organism (Protasio *et al.,* 2012). Additionally, it was possible to map trans-splicing events in a genome-wide scale and the existence of polycistronic transcripts in *S. mansoni* demonstrated for the first time.

Schistosomula from *in vivo* infections are usually recovered in very small numbers often not sufficient for high-throughput studies. The vast majority of available data generated from high-throughput gene expression studies have been gathered using a mechanical transformation approach of the schistosomula (which does not include invasion of host skin). Available comparisons of skin-transformed and mechanically transformed schistosomula are limited to changes occurring to the parasites surface (Brink *et al.*, 1977) and changes observed through electron microscopy (Cousin *et al.*, 1981). The only available gene expression study featuring a comparison from the gene expression perspective was done in *S. japonicum* and was not targeted to the skin stage of the parasites (Chai *et al.*, 2006). Because the skin stage is regarded as one of the vulnerable stages, it is important to understand what are the effects of the mechanical transformation in the parasites transcriptome. To that end, skin-transformed and mechanically transformed schistosomula were study of the bases of their transcriptome differences and similarities (Chapter 4). These results would help validate previous and future studies.

Finally, a time course analysis of the gene expression during the first 24 hours of infection is presented (Chapter 5). Close time points were selected for this study guaranteeing the maximum possible resolution. The combination of the powerful RNA-seq approach as the cutting edge technique for gene expression measurement and the newly improved genome assembly and annotation, led to the identification of developmentally regulated processes and genes that would help improve the knowledge of this critical stage in the parasites' development.

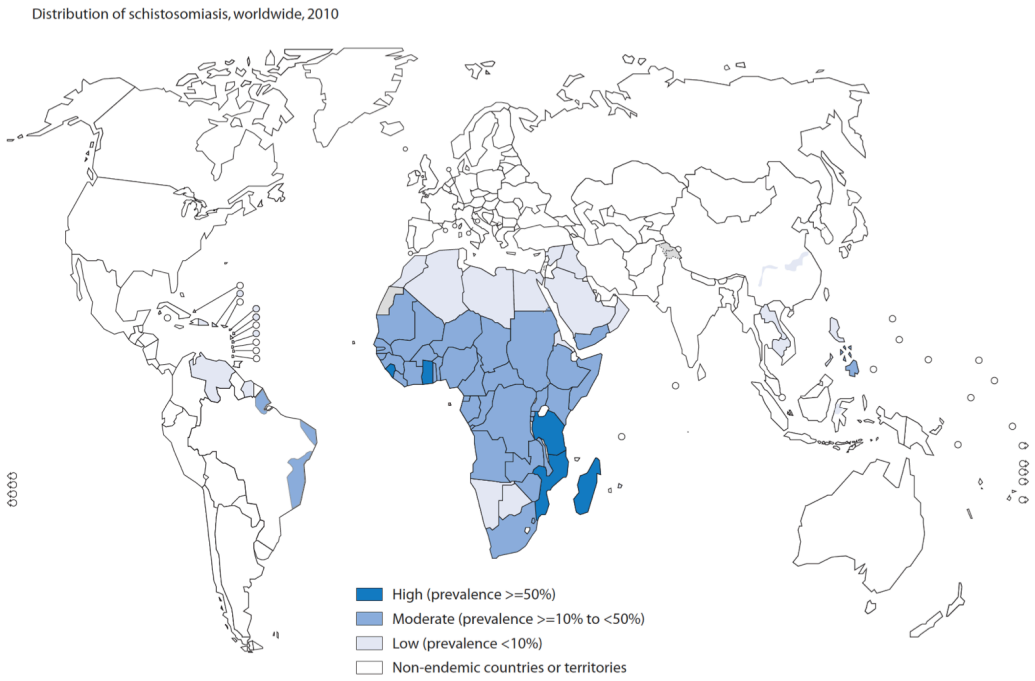

Figure 1.1 – Global distribution of schistosomiasis estimated as for the year 2009 [reproduced from (WHO, 2011)]. Shades of blue represent prevalence[1] of infection among individuals.

[1] Prevalence is the percentage of infected individuals within the population

1.2 Biology of schistosomes

Schistosome parasites are platyhelminths, blood dwelling trematodes belonging to the subclass digenea. Unusually for platyhelminths, they have separate male and female individuals (dioecious). They infect vertebrates (i.e., mammals, birds) and use them as their definitive host while aquatic or amphibious snails serve as intermediate hosts. Human schistosomiasis is caused mainly by five *Schistosoma* spp.: *S. mansoni, S. japonicum, S. haematobium, S. intercalatum* and *S. mekongi*, which are found in sub-Saharan Africa, Southeast Asia, regions of South America and the Caribbean (**Figure 1.1**). *S. haematobium* and *S. mansoni* are distributed across all sub-Saharan Africa and are the main agents for schistosomiasis in this region with some areas showing infections with *S. intercalatum* parasites. Only *S. mansoni* is found in South America and the Caribbean as a result of the slave trade between the 15th and 19th century (Morgan *et al.*, 2005). *S. japonicum* and *S. mekongi* are found in Southeast Asia with the former being the subject of a massive public health programme in China (Davis, 2002).

The following sections introduce the life cycle of schistosomes, principal aspects of the pathology and the current mechanisms of intervention.

1.2.1 Life cycle of schistosomes

The life cycle of schistosomes comprises two parasitic and two free-living phases. During the parasitic phases schistosomes colonize an intermediate snail host or a definitive vertebrate host. The free-living stages occur in fresh water environments and provide a link between the two parasitic stages (**Figure 1.2A**) breaching the physical gap between the two hosts. Each schistosome species shows a preference for a particular genus of snail host: *S. mansoni* infects snails of the genus *Biomphalaria*, while *S. japonicum* and *S. haematobium* infect snails of the genus *Oncomelania* and *Bulinus* respectively. However, not all the *Biomphalaria* species are susceptible to infection by *S. mansoni* (Davis, 2002). The geographic distribution of the susceptible snail population strongly influences the epidemiology of schistosomiasis (Steinmann *et al.*, 2006). Vertebrate hosts can be from a variety of classes including mammals and birds. Because of their clinical relevance, human-infectious *Schistosoma* spp. are the most widely studied. In the laboratory, the life cycle of *S. mansoni* can be maintained by using small rodents such as mice and hamsters as definitive hosts.

Adult worms of different species have different preferences for their final location in the definitive host: *S. mansoni* and *S. japonicum* stay in the inferior and superior

mesenteric vessels respectively while *S. haematobium* prefers the small venules around the bladder and the ureter (Cook *et al.*, 2003). In the case of *S. japonicum* and *S. mansoni*, once male and female have paired they migrate against the blood flow through the hepatic portal vein and towards the mesenteric branches around the intestine. Once they have reached sexual maturity, male worms are 6-13 mm long and 1 mm wide while females are typically between 10-20 mm long and 0.16 mm wide. Female worms are easily recognisable even to the naked eye as they appear more slender than the males and show a noticeable dark pigmentation in their gut (**Figure 1.2B**). Other distinctive phenotypic characteristics are the oral (for feeding) and ventral (for attachment) suckers in both genders and the gynaecophoric canal in the male, where the female resides. The female lays eggs continuously and in close proximity to the intestine's endothelia to facilitate their migration through the gut wall into the intestinal lumen through which the eggs finally reach the exterior in the excreta. *S. mansoni* and *S. haematobium* females produce up to ~300 eggs a day while *S. japonicum* can produce up to 10 times more (~3,500). Eggs are about 100-150 um long, with each species presenting a characteristic shape (**Figure 1.2A**) commonly used in the diagnostics of the disease [reviewed in (Davis, 2002)]. Approximately 50% of the eggs are not excreted and are retained within the host tissues. Eggs are passively taken by the blood flow towards the liver (*S. japonicum* and *S. mansoni*) or bladder (*S. haematobium*) where granulomas are formed as a consequence of the host's immune response against egg secreted antigens. The formation of a granuloma around the egg in the host tissues is the cause of pathology in all *Schistosoma* infections. Granulomas are organized agglomerations of cells (eosinophils, macrophages, CD4+ T cells) and collagen fibres, whose main objective is to isolate the egg (source of antigen) from the host. The egg eventually dies and the granuloma resolves itself leaving a fibrotic plaque. As infection progresses, more and more granulomas are formed and resolved and the affected tissue becomes fibrotic causing the pathology [reviewed in (Pearce *et al.*, 2002)].

After contact with fresh water, eggs hatch into the second free-living stage called miracidium. Miracidia can actively swim and are infective to snails for 8-12 hours after hatching. Once in the snail, miracidia transform into mother sporocysts and after eight days, germ cells bud off from them and develop into daughter sporocysts. Another round of germinal-cell production generates the cercariae, which are released into the fresh-water in response to environmental cues [reviewed in (Davis, 2002)]. The stage in the snail host represents cycles of asexual reproduction.

As with miracidia, cercariae are short lived and need to find a suitable host within hours. Host encounter is promoted by the active and intermittent swimming of the cercariae [reviewed in (Curwen *et al.*, 2003)]. Once the larvae have found a suitable host, probably aided by the presence of host fatty acids (Haeberlein *et al.*, 2008), proteases are secreted to assist the penetration of the cercarial heads in the host skin (Salter *et al.*, 2000; McKerrow *et al.*, 2002; Curwen *et al.*, 2006; Hansell *et al.*, 2008). At this point, the tails are cast off and the cercarial heads transform into schistosomula. The process of transformation involves mainly the remodelling of the surface membrane, which has consequences both in the parasites' anatomy (i.e., generation of a double outer bilayer) and physiology (i.e., schistosomula are water-intolerant). The passage through the skin can take from few minutes and ends when schistosomula find and penetrate into a venule.

Between four and seven days after infection parasites are found in the vasculature of the lungs where they stay for at least two to three days (Miller *et al.*, 1980). After the lung stage, parasites migrate to the hepatic portal system where they reduce their size and are now phenotypically more similar to the schistosomula at day 0 post transformation. The first parasites to arrive at this location do it around the tenth day after infection. From the moment of host invasion until day 10 to 11, parasites show a decrease of many metabolic markers/parameters, such as wet weight and oxygen consumption (Lawson *et al.*, 1980). At this point, male and female worms are morphologically distinguishable but not yet sexually matured. Worms of opposite sex pair up and migrate together against the blood flow towards the mesenteric veins where they attach and reside.

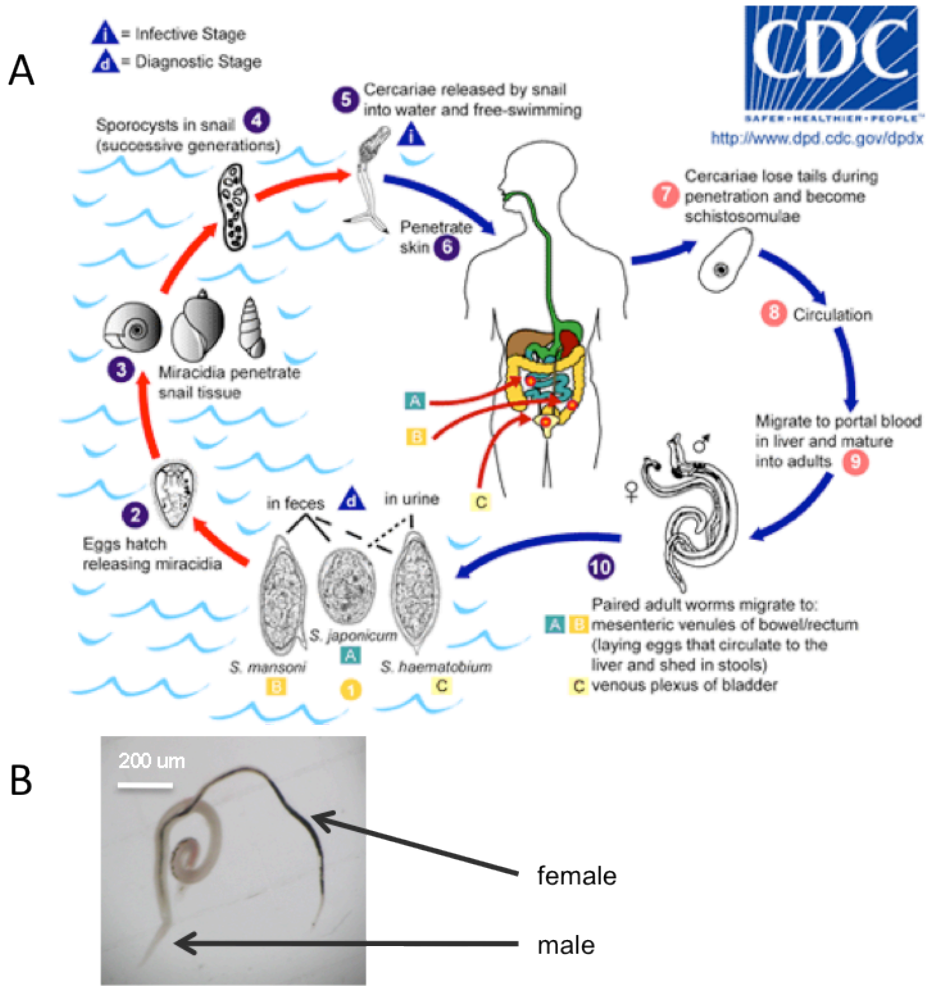

Figure 1.2 – Life cycle of *Schistosoma* spp. A - Life cycle of *S. mansoni, S. japonicum* and *S. haematobium* [reproduced from (CDC, 2011)]. These three parasites have a very similar life cycle. Eggs are released in the urine or faeces of infected humans (1). Upon contact with fresh water, the miracidia are released (2). Once they have infected a susceptible snail host (3), start a series of asexual reproduction stages occur (4) finally generating the human-infective larvae called cercariae (5). These are released by the snail into the freshwater and infect humans and other species by penetrating intact skin (6). The cercariae lose their tails and transform into schistosomula (7). After a short passage through the skin, the parasites reach the circulation (8) migrating to the lungs (9) and later the portal system where they develop. Finally, male and female worms pair up and lodge themselves (10) in the mesenteric veins of the portal system (*S. japonicum* and *S. mansoni*) or the bladder tributaries (*S. haematobium*). Females produce hundreds (*S. mansoni* and *S. haematobium*) to thousands (*S. japonicum*) of eggs. B - *S. mansoni* male and female worms partially attached. The female worm is more slender and appears darker than the male.

1.2.2 Chemotherapy and control of schistosomiasis

The main objective of chemotherapy is to restore the patient's wellbeing and reduce transmission of the infection. Anti-schistosomiasis chemotherapy introduced in the late 1960s fulfilled these criteria. Current methods of treatment of infected individuals produce the death of adult worms and the concomitant reduction in eggs deposited in tissues and hence reduction of pathology. At the same time, reduction in egg laying generates a break in the parasites' life cycle reducing community infections and therefore improving the health of the population. However, in real-life treatment programmes the situation is rarely this idyllic: in mass treatments usually only 50-60% of the population is cured – partly due to non-compliance - and this situation commonly leads to re-infection which is further favoured by poor sanitary conditions (Davis, 2002).

Praziquantel is the drug of choice because it is very effective against the adult worms of the three main *Schistosoma* spp., is also cheap and well tolerated. Patients may suffer from side effects, but these are temporary and have no long-term consequences. There is a common fear amongst *Schistosoma* researchers and public health officers that resistance to praziquantel (PZQ) may develop in the near future (Davis, 2002). Reduced susceptibility of *S. mansoni* worms to PZQ has been reported in the field (Ismail *et al.*, 1996; Melman *et al.*, 2009) and it has been proven that PZQ resistance can be induced in experimental conditions (Ismail *et al.*, 1994); raising the possibility that a similar situation could be also seen in the field.

Antioxidant pathways are known to play an important role in the survival of schistosomes in the oxidative environment of the blood stream where they reside (Loverde, 1998). Research into these pathways led to the identification of one chokepoint in the parasites' antioxidant portfolio, the thioredoxin glutathione reductase, which has been proven a lethal target for all intra-mammalian stages of the life cycle of *S. mansoni* (Sayed *et al.*, 2008). Further studies into essential pathways may lead to the discovery of new targets for intervention.

1.2.3 Insights into the cercariae and skin schistosomula stages

The following sections describe in detail the anatomy and physiology of the cercariae and the schistosomula. These stages are the focus of this work.

1.2.3.1 Cercariae

Cercariae, the human infective stage, emerge from the infected snail in response to light stimulus. They are typically 500 um long but can vary due to their great capacity of contraction/extension (Dorsey *et al.*, 2002). Once in the water, the cercariae move with sudden upwards motions followed by passive sinking (Graefe *et al.*, 1967; Cook *et al.*, 2003). The body of a cercaria consists of a head and a tail (**Figure 1.3**) and is well adapted for the task of invasion: the head can penetrate the host skin whereas the tail is lost.

Anatomically, the head can be divided into three parts: oral, middle and aboral (**Figure 1.3**). The oral part contains the oral sucker, the mouth and a strong musculature structure that is thought to assist migration throughout the skin. A ventral sucker (commonly referred to as acetabulum) is located towards the aboral part [reviewed in (Stirewalt, 1974)]. Pre- and post- unicellular acetabular glands are found anterior and posterior from the acetabulum with their cytoplasmatic processes extend towards the oral sucker. These are used to release peptidase-containing vesicles to the exterior [(Fishelson *et al.*, 1992), see section 1.2.3.1.1].

The tail is a highly specialized organ, which provides motility to the parasite during its free-living stage in the water. This organ is packed with myocytes, neurons, osmoregulatory cells and supporting cells. Myocytes are organised in an inner longitudinal, a subtegumental and three outer circular muscle layers that form the tail musculature structure. The tail must fulfil high-energy demands and is packed with large mitochondria, large numbers of ribosomes and glycogen (Dorsey *et al.*, 2002).

1.2.3.1.1 Cercariae secretions during skin penetration

During cercariae development in the snail, the most prominent differentiating cells are the ones that would give origin to the pre- and post-acetabular glands. These are unicellular structures that have their cell body located either posterior or anterior to the acetabulum or ventral sucker and whose cytoplasmatic processes extend and open at the apical end of the cercarial head (**Figure 1.3B**). Secretory products are packed in vesicles and reach the exterior through the cytoplasmatic processes. These vesicles burst upon contact with the hosts skin and expose proteases, which diffuse through the dermal extracellular matrix. The main function of these secretions is to aid the penetration of the cercariae across the different layers of the skin barrier. Many of the secretion products have been found to elicit an immune response [e.g, the *S. japonicum* paramyosin (Gobert *et al.*, 1997)] and therefore have received much attention from the research community

(Curwen *et al.*, 2003; McKerrow, 2003; Curwen *et al.*, 2006; Hansell *et al.*, 2008) due to their potential as vaccine targets.

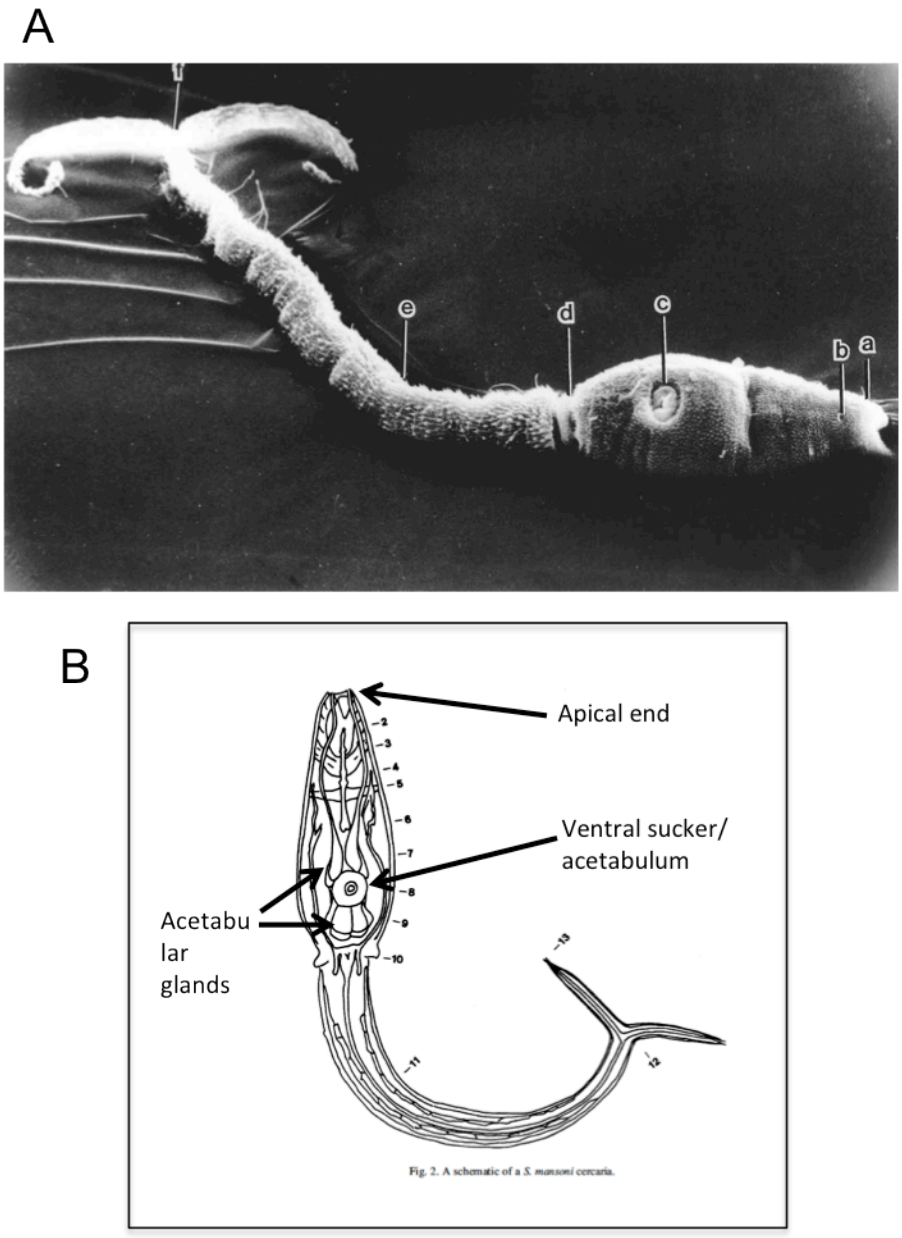

Figure 1.3 – Body organization of the cercariae. A - Scanning electron microscopy [reproduced from (Dorsey *et al.*, 2002)]. a – oral sucker; b – mouth; c – ventral sucker or acetabulum; d - body/tail junctions; e – tail; f – bifurcated tail. B - Schematic representation of the cercariae. Note how the cell bodies of the acetabular glands are located next to the acetabulum but their cytoplasmatic processes extend towards the apical end of the cercarial head [modified from (Dorsey *et al.*, 2002)].

Because the secretions are found to cover the penetration tunnel, they are also found in association with the parasite's surface and they might contribute towards the surface transformation of the parasite (Fishelson *et al.*, 1992).

1.2.3.1.2 Cercarial tegument

The general structure of the tegument (**Figure 1.4**) is similar in the cercariae and the adult worms (Stirewalt, 1974) and covers the entire surface of the parasites. Much of the research done regarding the tegument has been focused around the host-parasite interaction in the adult worms [reviewed in (Skelly *et al.*, 2006)], but most of the general characteristics are also valid for the cercariae. The tegument is located immediately underneath the apical plasma membrane (either cercariae, schistosomula or adult worm) and above the basal membrane. It is a syncytium (a continuum of cytoplasmatic material) with its cell bodies located deep under the muscle layers (Dorsey *et al.*, 2002). The tegument contains lipids, carbohydrates (glycoproteins and glycolipids) and many different proteins including enzymes and receptors. However, there is no evidence of DNA or RNA molecules found in the syncytium, suggesting that protein synthesis occurs exclusively in the cell bodies. Thin cytoplasmatic processes communicate the nucleated cell bodies with the tegument (Skelly *et al.*, 2006).

There are five types of tegumental cells, named I to IV plus the head gland. They connect to the syncytium through their cytoplasmatic processes, which are lined with microtubules. The different types of cells have different collections of vesicles and inclusion bodies. For example, one of the cell types contains multi-laminated vesicles thought responsible for the generation of the double bilayer, while others are packed with biogenic amines or ribosomes (Dorsey *et al.*, 2002).

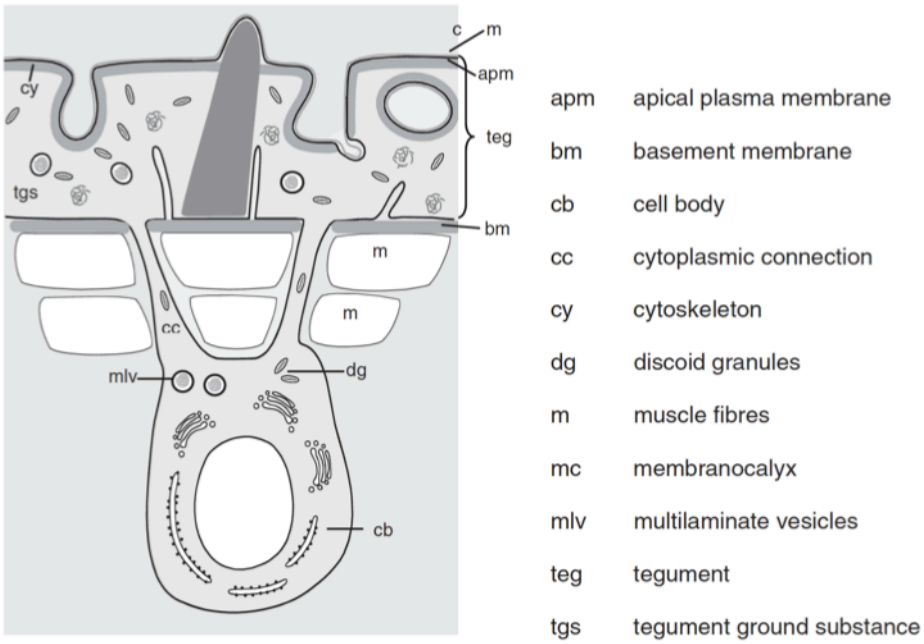

Figure 1.4 – Schematic organization of the tegument in a transversal cut. The cell bodies of the tegumental cells are located underneath the muscle layers and connected to the syncytium through their cytoplasmatic processes The tegument is located between the basal membrane and the apical membrane. The glycocalyx (or membranocalyx) is the outermost layer. [reproduced from (Skelly *et al.*, 2006)]

1.2.3.1.3 The glycocalyx, glycocalyx shedding and tegument dynamics

The glycocalyx is the outer-most layer of the parasite surface and it is tightly connected to the tegument. Transmission microscopy experiments have shown that the glycocalyx forms a layer 1-2 um thick with fibrils 15-30 nm long. It is of very high molecular weight (in the order of millions) and it bears antigenic properties (Samuelson *et al.*, 1985). Different staining techniques have shown that the glycocalyx is of carbohydrate or possibly lipopolysaccharide nature (Samuelson *et al.*, 1985); with fucose and glucose being the major components of the head and tail respectively (Nanduri *et al.*, 1991). The glycocalyx protects the cercariae from the water environment and once inside the host skin, it may protect the transforming schistosomula from the attack of immune agents by trapping host molecules or acting as a physical barrier between the parasite's tegument and the host environment. The loss of the glycocalyx, which happens only after the newly acquired double bilayer membrane is formed, is necessary for the transformation of the cercariae into schistosomula. Hence, it is possible that it has a role in protecting the parasite during membrane remodelling (Skelly *et al.*, 2006).

Three main events are known to contribute towards the loss of the glycocalyx layer. These were described in detailed by Fisherson *et al.*, (1982) and a summary is presented below.

The first event is the thinning of the glycocalyx layer by the action of cercariae secretions (Samuelson *et al.*, 1985). These seem to originate from the secreted vesicles from the pre- and post-acetabular glands. Because these vesicles are released during penetration, probably triggered by the presence of specific fatty acids in the surface of the host (Stirewalt, 1978), the vesicle's contents are found covering the penetration channels and therefore can get in contact with the cercariae' surface (Hansell *et al.*, 2008). The thinning glycocalyx makes way to the exposure of spines covering the surface of the parasite. In transformed schistosomula, these spines can be clearly seen (Crabtree *et al.*, 1980). A similar effect is observed when cercariae are concentrated in aquarium water (Fishelson *et al.*, 1992). Concentration of cercariae by centrifugation is known to trigger the release of the contents of the acetabular glands producing the "artificial" thinning of the glycocalyx as cercarial proteases are found within the cercariae pellet. Second, microvilli start extending from the surface of the cercarial body and through the glycocalyx layer. These extensions may appear as soon as five minutes after transformation and are also shed from the parasite within 60 minutes. They are transient structures; the glycocalyx is anchored to them and when the microvilli are shed, part of

the glycocalyx is shed with them. The molecular structure of the glycocalyx is not damaged during this process (i.e., this is not a degradation event). Thirdly, multi-laminated vesicles originating from the tegumental cells reach the surface and discharge their contents. The migration of the vesicles occurs at the same time as the microvilli formation and it may be linked to the generation of the double bilayer that will replace the single bilayer previously found. The shedding of the glycocalyx and replacement of the membrane occurs asynchronically and 1 hour after transformation, the parasite's surface still has both single and double bilayers distributed in patches (Fishelson *et al.*, 1992). With time, the whole of the membrane transforms into a double bilayer. In the maintenance of the parasites' surface after schistosomula stage, the multi-laminated vesicles continue to provide material for the double bilayer and the glycocalyx (Skelly *et al.*, 2006).

One hour after penetration of the host skin, the transformation of the cercarial surface is almost completed and the parasites show changes in aspects of their physiology. Through close examination of transforming parasites, Stirewalt *et al.,* (1966) described the schistosomula as saline and serum adapted tail-less organisms derived from the cercarial stage. These transformed parasites are water-intolerant and their acetabular glands are emptied after skin penetration. As previously described, the surface of the schistosomula has changed into a double bilayer (Stirewalt *et al.*, 1966) that looks as if heptalaminated under the light microscope.

1.2.3.2 Schistosomula

Schistosomula reach the boundary between the dermis and epidermis sometime between 30 minutes and two and a half hours after skin penetration. By then, the glands from the apical end still contain secretory granules (Dorsey, 1976). Many parasite-derived proteases and proteins are found in the epidermal tunnels and are thought to assist the migration process. Some of these are a serine protease called cercarial elastase/protease, a serpin (serin protease inhibitor), heat shock proteins (HSP) 86 and 70 and a couple of proteins putatively involved in immune evasion [paramyosin and Sm16 – a protein found in the secretions of invading cercariae (Holmfeldt *et al.*, 2007)]. All these are released within the first half of the hour of the skin penetration process (Hansell *et al.*, 2008). Acetabular gland tubes, previously prominent in the cercariae, are no longer found at 40 hours post infection suggesting that certain secretory mechanisms disappear after invasion (Crabtree *et al.*, 1985). This is in agreement with their proposed role in transforming the surface membrane. Although passage through the skin is supposed to be a relatively rapid process, many parasites seem to take longer. Even when a population of

cercariae are let to penetrate live animal skin under experimental conditions, schistosomula were shown to still be in the epidermis after 24 or even 40 hours post infection (Crabtree *et al.*, 1985).

At 48 hours post-penetration, parasites that successfully migrated into the dermal layer of the skin still contain vesicles in the head glands. After locating a blood vessel it will take schistosomula approximately eight hours to penetrate through the blood vessel endothelia and reach the blood stream (Wilson *et al.*, 1980). The fact that head glands have not been completely emptied when the parasite arrives at the blood vessel endothelia suggests that some of the glands' contents may be used to disrupt the extracellular "cement" found in endothelial walls and make way into the venule. The parasite also displays rapid everted/inverted movements, with the head capsule showing a strong and prominent musculature that controls the protrusion of the apical end. Taken altogether these suggest that endothelium penetration is achieved by using a combination of chemical (contents of the head gland) and physical (movements resembling a "battering ram") action (Crabtree *et al.*, 1985).

1.2.3.3 Energy metabolism

In terms of energy production, free-living stages (miracidia and cercariae) use the same typical aerobic metabolism as other higher eukaryotes: glycogen is degraded to pyruvate through the classic glycolysis pathway [reviewed in (Barrett, 1981)]. Pyruvate is then transported to the mitochondria where it enters the tricarboxylic acid cycle (TCA, Kreb's cycle) releasing CO_2 and producing NADH to feed into the respiratory chain [reviewed in (Tielens, 1994)] and generate ATP through oxidative phosphorylation. These mechanisms have been proven for both miracidia and cercariae stages in *Fasciola hepatica* and *S. mansoni* (Barrett, 1981).

During host-larval stages, such as that of the schistosomula, energy metabolism is mainly anaerobic but parasites retain their potential to use aerobic metabolism. It has been shown that approximately half of the production of L-lactic acid in the newly transformed schistosomula comes from oxidative phosphorylation, hence the metabolic switch seems to be an incomplete one (Coles, 1973). The cause for the change in the metabolism is not yet clear: some authors suggest that it is due to the availability of glucose, which increases the glycolytic flux (Barrett, 1981) while others suggest that it is linked to what triggers the transformation of the parasite as a whole [i.e., host fatty acids (Coles, 1973)]. Irrespective of the mechanism, this metabolic switch only occurs in the cercarial heads and not in the tails (Horemans *et al.*, 1991). Although the whole organism

has the same capacity for using different metabolic pathways, the tails use a different one from that used in the cercarial head. Tails have more cytochrome oxidase activity (Coles, 1973; Skelly *et al.*, 1993) to meet the energy demands of this specialized organ. It is commonly understood that the up-regulation of enzymes corresponding to anaerobic metabolism (i.e. lactate dehydrogenase) corroborates "the described switch in the larval metabolism from aerobic to anaerobic pathways during transformation" [from cercariae to schistosomula] (Lawson *et al.*, 1980; Farias *et al.*, 2011).

1.2.4 Pathology and Immunology of schistosomiasis

General characteristics of the pathology of schistosomiasis are directly related to the life cycle of the parasite in the human host (Table 1.1). These are different in non-immune individuals in comparison to individuals leaving in endemic areas where their exposure to continuous infection and re-infection episodes grants them certain level of protection (Cook *et al.*, 2003).

The Katayama fever syndrome is common to all schistosome infections affecting humans. It is most marked in the primary infections of individuals living in non-endemic areas. The period between infection and set off fever varies: for *S. japonicum* it ranges from two to six weeks while in *S. mansoni* is generally from three to seven weeks. Symptoms are basically those of an acute fever episode: continuing high body temperature, shivering/trembling, sweating, general muscles pain, headaches and in less percentage of cases: anorexia, nausea and abdominal discomfort (Cook *et al.*, 2003). These are the characteristics of a dominating T-helper 1 (Th1) response. This pro-inflammatory response features high levels of circulating tumour-necrosis factor (TNF), interleukin-1 (IL-1) and IL-6 produced by peripheral blood mononuclear cells. Approximately six to eight weeks after infection, adult worms are fully matured and egg laying starts. Eggs lodge in the liver where they produce secretions rich in carbohydrates (also known as schistosome egg antigen or SEA). These secretions induce a T-helper 2 (Th2) response which in turn down regulates the effector function of the Th1 pro-inflammatory response. This switch from Th1 to Th2 is vital for the survival of the host. However, long lasting Th2 responses, as occurs in endemic areas of schistosomiasis infections, are also cause of morbidity (Pearce *et al.*, 2002). Eggs trapped in the liver (*S. mansoni* and *S. japonicum*) or bladder (*S. haematobium*) also produce SEA causing the formation of granulomas (Cook *et al.*, 2003). As previously mentioned, these are organized agglomerations of cells such as eosinophils, macrophages, CD4+ T cells and collagen fibres. As infection progresses, more and more granulomas are formed and resolved leaving a significant number of fibrotic

plaques and a fibrotic liver. This leads to the blockage of the portal tracts causing portal hypertension. To compensate, liver capillaries are enlarged and branched leading to enlargement of the liver (hepatomegaly). Other complications arising from hypertension in the portal venous system include the spleen becoming tough, fibrotic and enlarged (splegnomegaly). The spleen may also become hyperactive (hyperslenism) causing the reduction of red and white cells, platellets and anemia [reviewed in (Pearce *et al.*, 2002 and Cook *et al.*, 2003)].

Table 1.1 - Summary of clinical manifestations in a *S. mansoni* infection.

DESCRIPTION	APPEARANCE CLINICAL MANIFESTATIONS (Time post-infection)	CHARACTERISTICS	IMMUNE EFFECTORS
Cercariae infection and schistosomula migration	24 to 48 hours	Cercarial allergic dermatitis	Eosinophilia and antibody-dependent cell-mediated cytotoxic response involving IgG.
Schistosomula maturation and establishment of paired adult worms	2 to 16 weeks	Febrile illness (Katayama syndrome)	Initiation of Th1, TNF, IL-1, IL-6.
Egg laying	From 2 months onwards	Granulomas	Start of Th2 response, down regulation of Th1.
Late staged of infection	Years	Portal hypertension and hepatosplegnomegaly (Sm)*	

In endemic areas, re-infection is more the rule than the exception. Children in school age and until puberty are the most heavily infected subgroup. Older people are usually less susceptible to re-infection and this protection to re-infection has been associated with high levels of immunoglobulin-E (IgE) directed against the adult worm. Interestingly, adult worms are not susceptible to the immune response of the host and therefore the slow development of resistance to re-infection agrees with the long life span of the adult parasites (Pearce *et al.*, 2002).

1.2.5 Vaccine development and vulnerability of schistosomes.

The search for a vaccine against schistosomiasis is based on the need to decrease the disease burden, which has not been reduced by chemotherapy. Additionally, high rates of post-treatment re-infection and the lack of an alternative to praziquantel raise the issue of the inadequacy of the intervention programmes so far used. The realisation of a successful vaccine is further encouraged by evidence of partial immunity against schistosome infections acquired by adults leaving in endemic areas [reviewed in (Hotez *et al.*, 2010)].

Previous efforts using irradiated cercariae have elicited immunity against a subsequent challenge infection in laboratory animals. However, the administration of a live vaccine represents many logistic inconveniences making it not easily viable in the field. Other approaches include the use of recombinant proteins such as the *S. haematobium* 28 kDa GST against urinary schistosomiasis, the *S. mansoni* 14 kDa fatty acid-binding protein, DNA vaccine Sm-p80 as well as several *S. mansoni* tetraspanins, all of which have been or are in early stages of clinical trials. However, these have had either limited success or are awaiting clinical trials that will assay their safety and efficacy [reviewed in (Hotez *et al.*, 2010)].

The debate about which of the life cycle stages is the most vulnerable to intervention has received some attention (Curwen *et al.*, 2003; McKerrow, 2003). However, it has been suggested that the schistosome's weakest point in the life cycle might reside in the early encounter of the parasite with its mammalian host; that is the skin schistosomula stage (Wilson *et al.*, 2009). This was based in the following principles. First, it has been shown that the schistosomula stage is susceptible to the attack of oxidative species rendering certain level of vulnerability (Loverde, 1998). Second, a series of studies reviewed by Capron *et al.,* (1986) showed that schistosomula killing in the rat model is mediated by antibody-dependent cell cytotoxicity involving IgE (Capron *et al.*, 1986). In summary, developmental changes undergone by the migrating schistosomula leave open

opportunities for the interaction with the host defences enabling both recognition and killing (Wilson *et al.*, 2009).

A prophylactic vaccine would have the capacity of priming the host's immune system against a challenging *Schistosoma* infection by introducing one or many substances that resemble structures found in the invading schistosomula. The host's immune system detects and attacks such structures but also remembers them. In the event of a challenge infection, the immune response is ready to act defending the host (Abbas *et al.*, 2000).

One of the current strategies for finding the magical bullet(s) that would lead to a successful vaccine is to look for molecules that would comply with certain characteristics. Apart from being expressed in the skin stage of the parasites, such potential antigens would have to be secreted or exposed in the parasites surface where the host's immune system can detect them. In the last five years, our understanding of which genes are being expressed upon infection has been improved by the many microarray studies targeting the cercariae and schistosomula stages (Chai *et al.*, 2006; Dillon *et al.*, 2006; Fitzpatrick *et al.*, 2009; Gobert *et al.*, 2010; Parker-Manuel *et al.*, 2011). Although these studies achieved good resolution in identifying differential expression of hundreds and thousands of genes within the first days of infection (3 to 7-10 days old schistosomes), only one study focused entirely in the first 3 hours post infection (Gobert *et al.*, 2010) when important changes in the surface of the parasite take place. If the early schistosomula stage is the one chosen to be targeted for intervention, a more thorough and comprehensive study of its gene expression portfolio would need to be achieved.

Current efforts in the development of intervention strategies base their platforms in combining the genome and the transcriptome information together with high-throughput expression studies (TheSchistoVac, 2009) to identify potential drug targets and vaccine candidates. To this end, the accuracy of both genome and transcriptome data is a fundamental aspect of the drug and vaccine development pipeline.

1.3 Genome biology of S. mansoni and current status

The *S. mansoni* karyotype comprises 7 pairs of autosomes and a pair of sex chromosomes. All chromosomes are distinguishable by their unique chromosomal banding patterns (Grossman *et al.*, 1981). Females are the heterogametic sex with both Z and W chromosomes while males have two copies of the Z chromosome.

The haploid genome size had been estimated to be approximately 270 Mb (million DNA base-pairs) (Simpson *et al.*, 1982), much larger than the nematode *Caenorhabditis elegans* [~100 Mb (*C. elegans* Sequencing Consortium, 1998)] and the fruitfly *Drosophila melanogaster* [~117 Mb (Adams *et al.*, 2000)]. The genome has a GC content of 35% and is 40% repetitive (Berriman *et al.*, 2009), with retrotransposons being the most commonly repeated element found in the genome (Simpson *et al.*, 1982).

The quest for large-scale gene discovery started in mid 1990s with the first publication of an Expression Sequence Tags (ESTs) study (Franco *et al.*, 1997). In the early studies, and limited by the technology available at that time, only a restricted number (466 unique genes) of genes could be discovered (Franco *et al.*, 1997). Other ESTs projects both in *S. mansoni* [i.e., (Franco *et al.*, 1997; Santos *et al.*, 1999)] and *S. japonicum* [i.e., (Fan *et al.*, 1998; Fung *et al.*, 2002)] followed and in 2003 probably the most comprehensive EST study in *S. mansoni* was published (Verjovski-Almeida *et al.*, 2003). In this publication, the authors generated ~160,000 ESTs with 31,000 assembled sequences spanning six different life cycle stages (cercariae, 7-day old *in vitro* cultured schistosomula, adult worms, egg, miracidia and germ-balls). With this work it was possible to estimate that the gene complement of *S. mansoni* would be around ~14,000 genes. However, the success of using ESTs for gene finding is limited mainly by the under-representation of low expressed transcripts (specially in normalized libraries such as those used in Verjovski-Almeida *et al.,* 2003) and lack of coverage over the full length of the transcript (specially towards the 5'-end). In addition, some life cycle stages are still poorly represented in existing EST resources. Furthermore, EST sequencing using conventional methods, such as the capillary Sanger method, is nowadays relatively expensive and laborious

The first published draft of the *S. mansoni* genome was assembled into 19,022 scaffolds with a gene complement of 11,809 genes and 13,197 transcripts (Berriman *et al.*, 2009), a similar figure to that obtained by a comprehensive study of *S. mansoni*'s ESTs (Verjovski-Almeida *et al.*, 2003). More recently the genome has been systematically improved. This was done using the original draft data, new capillary sequencing data and

second-generation genomic data produced from DNA obtained from single miracidial-derived worms, resulting in a much less fragmented assembly with only ~885 scaffolds (Protasio *et al.*, 2012). Furthermore, 86% of the improved assembly can now be allocated into physical chromosomes thanks to linkage markers (Criscione *et al.*, 2009) and fluorescence *in situ* hybridization of mapped BACs that had been previously generated (Berriman *et al.*, 2009). A summary of statistics from both assemblies is presented in **Table 1.2**.

Table 1.2 – Characteristics of the old and improved *S. mansoni* genome assemblies [reproduced from (Protasio *et al.*, 2012)].

	Old version[a]	**New version[b]**
Assembly size (Mb)	374.9	364.5
Proportion assigned to chromosome (%)	43	86
Contig statistics		
Number	50,292	9,203
Average length (kb)	7.5	39.4
N50 length (kb)	16.3	78.3
Largest contig (kb)	139.4	460
Scaffold statistics		
Number	19,022	885
Average length (kb)	20	411.9
N50 length (Mb)	0.8	32.1
Largest scaffold (Mb)	4.2	65.5 [c]

[a] Version 4.0 of the *S. mansoni* genome was the published draft genome (Berriman *et al.*, 2009). [b] Version 5.0 (Protasio *et al.*, 2012).

Both EST databases and the genome assembly provided a good tool kit for the investigation of gene expression profiles. The study of the transcriptome originated from EST projects and greatly advanced by the availability of microarrays quickly led to high-throughput studies that looked at genes expressed at different stages of the parasite's life cycle. These advances would help the study of the organism's basic biology and in the identification of pathways that could represent points of weakness for intervention (Hoffmann *et al.*, 2003; Gobert, 2010).

Microarrays have been used to achieve a systematic and quantitative approach to gene expression in *S. mansoni* (Fitzpatrick *et al.*, 2005; Dillon *et al.*, 2006; Fitzpatrick *et al.*, 2006; Vermeire *et al.*, 2006; Jolly *et al.*, 2007; Verjovski-Almeida *et al.*, 2007; Fitzpatrick *et al.*, 2009; Gobert *et al.*, 2010; Parker-Manuel *et al.*, 2011). They have some clear advantages - such as being able to study a large number of genes simultaneously, and for a larger number of individuals than possible with previous techniques. However, they also have certain disadvantages, when compared to high-throughput transcriptome sequencing: given that microarrays are (typically) designed based on known gene models or known genomic sequences, by definition, this type of survey is biased and it does not allow the identification of new transcripts. In addition, the study of related sequences and transcripts derived from alternative splicing is difficult to interrogate with microarrays due to cross-hybridization. Furthermore, microarrays rely on detection of continuous analogue signals that are difficult to quantify and often saturable; they depend on the inclusion of internal standards, and background fluorescence prevents the measurement of transcripts with low expression. Results are usually difficult to normalize between platforms and laboratories because each of them often uses different experimental designs (e.g, array design, glass slide, etc) (Shendure, 2008).

The advent of second generation sequencing technologies (also referred as "new" or "next generation" sequencing technologies) developed by several companies (Illumina/Solexa, 454-pyrosequencing/Roche, ABI/SOLiD and Helicos) has given new horizons to the study of functional genomics [reviewed in (Mardis, 2008; Morozova *et al.*, 2008)]. The year 2008, the same year this thesis work begun, saw the first reports on Illumina sequencing technology applied to gene expression quantification. These reports introduced the direct sequencing of the transcriptome through "RNA-seq" (Marioni *et al.*, 2008; Mortazavi *et al.*, 2008), which involves sampling the whole transcriptome of an organism at a given time point (or several) and subjecting it to high-throughput deep-sequencing [reviewed in (Wang *et al.*, 2009)]. To date, hundreds of publications have

featured the use of RNA-seq for gene calling, gene structure refinement, *de novo* transcriptome assembly and full characterisation of all kind of organisms' transcriptomes [i.e., (Otto *et al.*, 2010; Severin *et al.*, 2010; Chaudhuri *et al.*, 2011; Xia *et al.*, 2011)] including some describing improvement to probably some of the most highly characterised genomes and transcriptomes such as *Caenorhabditis elegans* (Hillier *et al.*, 2009) or *Drosophila melanogaster* (Daines *et al.*, 2011). Even some parasitic worms have been subjected to this type of studies (Laing *et al.*, 2011) including *S. mansoni* (Almeida *et al.*, 2011; Protasio *et al.*, 2012).

Briefly, Illumina uses massive parallel sequencing through the sequencing-by-synthesis technology (Bennett *et al.*, 2005) capable of great depth of coverage and ultra high-throughput. The basis of the reaction chemistry is shown and explained in **Figure 1.5**. Sequencing yield and quality has been dramatically improved over the recent years due to the ongoing research and development that is put into this technology. In January 2008 the state of the art in production pipeline at the Wellcome Trust Sanger Institute was of 3 million 37 bases single end reads per lane. By July 2011, the state of the art sequencing in the same facilities was of 120 million 108 bases paired end reads per lanes.

As previously mentioned, RNA samples can be sequenced using this approach. In the case of eukaryotic samples extracted RNA, usually performed by standard methods such as extraction with TRIzol® or column based methods, is subjected to a selection process that enrich the samples in polyadenylated molecules, which will contain mainly protein-encoding mRNAs. These are later fragmented and double-stranded DNA is generated based on the RNA sequences (Mortazavi *et al.*, 2008). After ligation of the appropriate adapters the DNA molecules are ready to be sequenced as previously described (**Figure 1.5**).

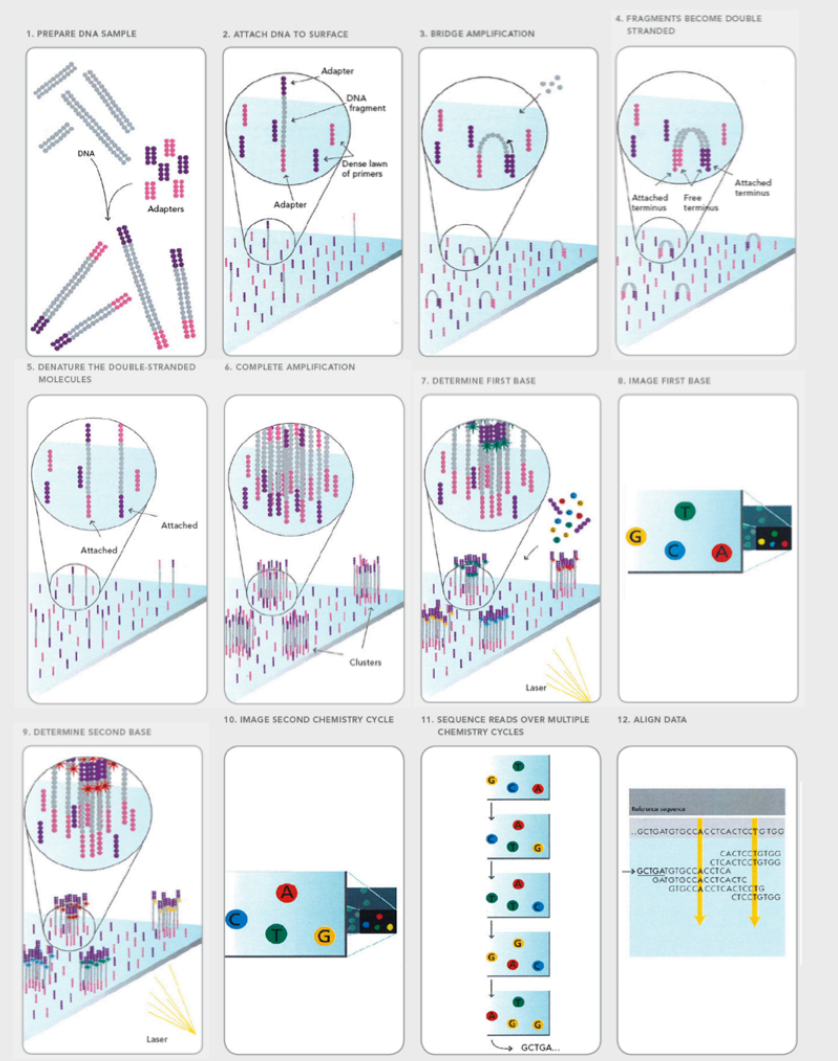

Figure 1.5 – DNA sequencing using Illumina technology. A – This series of panels explains the steps taken to sequence DNA using high-throughput Illumina technology. Double stranded DNA is randomly fragmented, size selected and adaptors are ligated to each end of the molecules (1). These molecules are bound (single-stranded) to a flow cell containing a dense lawn of previously bound primers (with same sequence as the adapters - 2). The amplification reaction occurs through cycles of "bridging" and PCR reactions (usually referred to as bPCR) with unlabeled nucleotides (3-5). Repeated cycles produce a "polony" (cluster or colony of molecules), one for every molecule that was initially fixed to the flow cell (6). Then, the actual sequencing begins: labelled reversible terminators (nucleotides) are added together with primers and all reagents need for an amplification reaction. Given the nature of the terminators, only one base can be added at a time and the rest of the reagents are washed off. The laser excites the labelled molecules unblocking them and a camera captures the emitted fluorescence and coordinates of the emission in the slide. Then, a second round of sequencing can take place (7-10). After a number of cycles, a sequence of "colours" is recorded for a particular position in the slide representing a read (11). Alignment algorithms can be used to map these reads to a reference (12). For paired-end sequencing, a second round of sequencing takes place using the opposite adaptor as primer and therefore sequencing the other end of the same molecule (not shown). Images were reproduced from (Illumina, 2011).

1.4 What this thesis is about

The work of this thesis concentrates on three main topics. The first topic is the application of *S. mansoni* RNA-seq data to assist and improve the structural annotation of genes in the recently upgraded genome assembly. Four time points in the life cycle of the parasite were sampled and sequence. These data was used to resolve gene structures to a single-base resolution. Results obtained from this work are featured in Chapter 3 and have been recently accepted for publication in a peer-reviewed international journal (Protasio *et al.*, 2012).

The second topic that this thesis deals with is the equivalency of skin-transformed and mechanically transformed schistosomula. As previously introduced, cercariae transforms into schistosomula as it penetrates the skin barrier of the definitive host. This transformation can be mimicked in the laboratory by application of shear pressure to a cercarial sample. Because most downstream application use mechanically transformed parasites, it is important to understand what are the differences between these and more naturally transformed parasites that would better resemble a natural infection. Often very low numbers of parasites are obtained from skin samples *in situ*. Hence, the samples analysed here were obtained using a modified skin transformation method that allows recovery of a significant number of parasites that have been transformed by penetrating through host skin. These results are presented in Chapter 4.

The third topic focuses on RNA-seq transcriptome analysis applied to the study of a short time course involving cercariae, 3-hours old and 24-hours old schistosomula life cycle time points. The aim of this analysis was to identify processes that had been previously missed by less sensitive techniques, such as microarrays. Focus is done in aspects of the transcriptome that may have a role in assisting the adaptation of the parasites to the new environment in the mammalian host. These results are presented in Chapter 5.

Finally, Chapter 6 reviews the contribution of this thesis work and placed its main findings in the context of current knowledge of schistosome biology.