Chapter One

Introduction

Malaria is caused by protozoan parasites of the *Plasmodium* genus belonging to the Apicomplexa phylum. Over 200 species of *Plasmodium* have been identified, capable of infecting a wide range of vertebrate hosts and being transmitted by adult female mosquitoes of the *Anopheles* genus. Currently found in over half the countries of the world, malaria was responsible for 247 million cases and 863,000 deaths in 2008, 85% of the deaths occurring in children under five (W.H.O. 2009). There are five species of *Plasmodium* that commonly cause malaria in humans. Infections are most frequently caused by *Plasmodium falciparum* and *Plasmodium vivax,* while cases due to *Plasmodium ovale*, *Plasmodium malariae* and *Plasmodium knowlesi* are less common. As a cause of mortality, *Plasmodium falciparum* is by far the most virulent, accounting for 93% of deaths from malaria (W.H.O. 2009; Carter et al. 2002).

History of Malaria

The origins of *Plasmodium* have been traced back at least half a billion years. Three of the *Plasmodia* species capable of infecting humans, namely *P. vivax, P. ovale* and *P. malariae,* diverged from a single clade into their own distinct lineage over 100 million years ago. All known species of *Plasmodium* that infect mammals fall into one of these three lineages with the exception of *P. falciparum* (Carter et al. 2002). Until recently it was thought that the only close relative of *P. falciparum* is *P. reichenowi*, a parasite identified as infecting chimpanzees and gorillas early in the 19th century. It is now becoming apparent that multiple species related to *P. falciparum* infect chimpanzees and gorillas, and several competing claims as to the precise origin of human *P. falciparum* have been published (Krief et al., 2010; Rich *et al.* 2009; Prugnolle 2010). While the details are still emerging, what is clear is that *P. falciparum* has its origins in a cluster of great ape *Plasmodia*, and crossed into humans at some time after the divergence of the human and ape lineages.

The end of the last glacial period ca. 10,000 years ago led to the beginning of agriculture in the "Fertile Crescent". The spread of agriculture into central and western Africa brought radical changes to the environment that bore particular importance for the spread of malaria. Humans had shifted from being largely hunter-gatherers living at very low population densities to an agricultural lifestyle that meant living in settlements close to a water supply. For the *Anopheles* mosquito this shift towards agriculture meant that a large

number of blood meals and stagnant water, its ideal breeding ground, had been brought together (Carter et al. 2002).

From beginnings in Africa, historical records suggest that *P. falciparum* had reached India by 1,000 B.C. and was referred to as "the king of diseases". Hippocrates (460 – 370 B.C.) described quartan and tertian fevers in *Book of Epidemics*, suggesting that *P. vivax* and *P. malariae* were present in Greece during that period, but made no mention of severe, malignant fever, suggesting that *P. falciparum* was not present. Hippocrates observed that symptoms were more common in late summer and autumn and tertian fever had the lower mortality rate of the two (Sherman 1998). Hippocrates also attributed the sign of splenomegaly to the proximity of the patient to stagnant marshes and this connection was one of the driving forces behind the construction of vast drainage systems built by the Roman Empire from 50 B.C. – 400 A.D. However, as the Empire fell into ruin, malaria developed as a prominent problem, especially in the Campagna marshes surrounding Rome and it was the Romans who first gave malaria its name – the literal translation being "bad air" (Sherman 1998, Carter et al. 2002). Malaria spread across Europe and, while little is recorded about the disease in this period, from the fourteenth century onwards seasonal fevers known as "agues" were described frequently in England (Sherman 1998). Voyages to the Americas by Europeans and West Africans in the fifteenth century introduced malaria to the New World, with *P. falciparum* most likely being spread by the African slaves brought to the Americas (Carter et al. 2002). These voyages also led to the discovery that native Peruvian Indians were using powdered *Cinchona* bark, containing the active ingredient quinine, to treat the disease successfully. By the eighteenth and nineteenth century malaria was endemic across the North American colonies, and by the early twentieth century 500,000 cases of malaria were being reported in the United States each year, mainly in the south (Oaks Jr. et al. 1991). Globally, malaria reached a peak sometime in the nineteenth century, when at least half the world's population was at risk of contracting malaria and those that did develop the disease had a 1 in 10 chance of dying from it.

Epidemiology

Despite Hippocrates' early observation about stagnant water, it was not until the end of the nineteenth century that the organisms responsible for malaria were identified.

Ronald Ross, working in India, developed the hypothesis of Sir Patrick Manson and Alphonse Laveran that mosquitoes were responsible for the propagation of the organisms causing malaria by witnessing parasites in humans suffering from malaria and also oocysts in the midgut of the vector *Anopheles.* The identification of a causative organism in the vector led to Ross being awarded the first of four Nobel Prizes for malaria research, and also meant that the issue of controlling the disease could be addressed. At the start of the twentieth century around 10% of all deaths worldwide were due to malaria, amounting to three million annually (Carter et al. 2002). This figure remained relatively constant until after the Second World War, when deaths from malaria plummeted in all areas of the world except sub-Saharan Africa (where they fell only slightly). Up until the Second World War, malaria control was reliant upon treatment with quinine, the use of bed nets and drainage of stagnant water where the mosquitoes preferred to lay their eggs. While these methods did have some effect, the greatest impact on malaria morbidity came with the availability of DDT, an insecticide that acted as a contact poison to several arthropods, including *Anopheles.* The discovery of DDT as a poison led to the fourth Nobel Prize in the field of malaria research being awarded to Paul Hermann Müller in 1948. The spraying of DDT meant that malaria was subsequently eradicated from Brazil and Egypt by 1949.

The WHO began the Global Malaria Eradication Programme in 1955 with an emphasis on DDT spraying and disease surveillance, which contributed to the eradication of malaria from Europe by 1975. Although the programme had notable initial successes, in some regions it began to fail. *Anopheles* started to develop resistance to the insecticides used, as did the parasites to drugs used to treat the disease. A lack of funding and participation led to the programme being abandoned by 1969 in favour of one of control, despite having reduced the number of people at risk of malaria by about 700 million. This coincided with controversy about the environmental impact of DDT, which led to its use being banned in many countries. The consequent reduction in implemented control measures led to a two- to three-fold global rise in malaria morbidity between 1972 and 1976, as well as the disease returning to countries that had previously declared eradication. The situation deteriorated throughout the 1980s leading to a new strategy for malaria control being drawn up by the WHO in 1992. This focussed on the early diagnosis and

treatment of cases, implementation of sustainable preventative measures, containment of epidemics, and local involvement in research (Schlagenhauf-Lawlor 2008; Malaria Web Site 2010).

The Control of Malaria Today

The control of malaria transmission today falls into three categories: vector control, use of prophylactic and curative drugs, and vaccine development.

Vector control makes use of the spraying of walls with insecticides (Indoor Residual Spraying – IRS), and twelve insecticides are currently recommended for IRS. Vector susceptibility to these insecticides should be constantly monitored and where resistance develops the insecticide in use should be changed (W.H.O. 2006). Insecticide-treated nets and materials are the other primary means of vector control currently in use. Nets impregnated with permethrin or deltamethrin can reduce malaria by up to 20% in endemic regions (W.H.O. 2007). Larvicides are the other form of vector control deployed: typically using temephos which is toxic to *Anopheles* larvae and is placed at the surface of stagnant water – the site of *Anopheles* larval breeding sites (Guarda et al. 1999).

A number of drugs, known collectively as anti-malarials, are available for the prophylaxis and treatment of malaria. However, drug resistance has reduced the effectiveness of these agents. Chloroquine was the first antimalarial to which resistance was found. Chloroquine inhibits the biocrystallisation of haemozoin from haem that occurs in the intra-erythrocytic stage of the parasite, leading to a build up of haem that is toxic to the parasite. However, *Plasmodium* has evolved resistance to chloroquine by mutating the *Plasmodium falciparum* chloroquine resistance transporter (*Pf*CRT) that is able to transport chloroquine out of the food vacuole and allows haemozoin biocrystallisation to continue (Martin et al. 2009). This pattern of implementation of new drugs and subsequent development of parasite resistance has been a constant feature of the last 50 years, with resistance having developed to both pyrimethamine-sulfadoxine and mefloquine. Recently parasites in Cambodia have been identified that have reduced clearance time in response to artemisinins, suggesting that resistance to artemisinin, the cornerstone of many anti-malarial therapies, may be evolving (Noedl 2008).

The rapid evolution of drug resistance in parasites has intensified the need for a malarial vaccine. Repeated infection with *Plasmodium falciparum* leads to development of natural immunity, exemplified by the build up of immunity throughout childhood in individuals living in malaria-endemic areas. Naturally acquired immunity to malaria largely reduces the severity of disease, resulting in the vast majority of deaths caused by malaria occurring in young children (Girard et al. 2007).

P. falciparum sporozoites attenuated by gamma-irradiation were shown to elicit an antibody response to domains of the circumsporozoite surface protein (CSP), which subsequently protected against an heterologous *P. falciparum* challenge (Egan et al. 1993). Consequently vaccines developed to target the pre-erythrocytic stage aim to convey immunity by eliciting a humoral response, preventing the invasion of the liver by sporozoites or inducing a cell-mediated response to inhibit sporozoite maturation within invaded hepatocytes. The most promising pre-erythrocytic vaccine is RTS,S which is currently in phase III trials. Targeting CSP expressed on the surface of sporozoites and infected hepatocytes, RTS,S has been shown to reduce the risk of clinical malaria by 35% and the risk of severe malaria by nearly half over a period of 18 months, with no evidence of waning efficacy (Alonso et al. 2005).

The development of a vaccine to the erythrocytic stages of the parasite life cycle is appealing as it has the potential to reduce the burden of disease and reduce parasite transmission, although it does not prevent infection (Kappe et al. 2010). Inoculation of patients with very low doses of live *P. falciparum* with concurrent anti-malarial treatment led to the prevention of detectable parasitaemia in three out of four patients when re-challenged (Pombo et al. 2002). However the inoculation of patients with live parasites has significant associated safety concerns. Therefore vaccines are being developed targeting specific merozoite antigens, avoiding the problems of possible disease development after inoculation with live parasites. Antibodies targeting a number of merozoite surface proteins and apical proteins have been shown to inhibit erythrocyte invasion *in vitro* (Woehlbier et al. 2006; Richards & Beeson 2009) and there are currently numerous phase one and two trials of vaccines targeting various combinations of merozoite proteins, including MSP-1, -2 and -3 and AMA-1. Other possible merozoite targets include the DBL family of erythrocyte

binding antigens, including EBA-175 and the variant surface antigens (VSAs); however, the large antigenic variation present in the VSA family is a drawback to developing an effective vaccine. Any successful erythrocytic vaccine will probably have to target a number of parasite antigens to avoid the problems of antigenic diversity and the search for potential targets and combinations is ongoing (Richards & Beeson 2009).

In addition to investigation of vaccines that prevent disease, considerable research has investigated blocking parasite transmission. There are several vaccines that have been moderately successful in trials. In addition to vaccines, genetic modification of *Anopheles* mosquitoes by upregulation of AKT signalling in the midgut has demonstrated successful inhibition of the *Plasmodium* life cycle (Corby-Harris et al. 2010).

Pathology of Malaria

The symptoms and pathology of malaria present during the erythrocytic stage of the *Plasmodium* life cycle (outlined below) occurs at least a week after the initial exposure to an infected mosquito. The signs and symptoms of the uncomplicated disease are non-specific; fever, rigors, myalgia, sweating and headaches are common. Fever fluctuates in malaria, worsening at the time when a large number of merozoites egress from erythrocytes (Miller et al. 1994). Schizont rupture also results in release of substances termed malaria toxins that activate mononuclear cells to produce tumour necrosis factor α, a potent endogenous pyrogen (Bate et al. 1994). Hepatomegaly and splenomegaly develop in about a third of patients and haemolytic anaemia can arise after a few days. Typically symptoms worsen as parasitaemia rises to a high level followed by a period where parasitaemia falls and symptoms lessen, before parasitaemia starts to rise again (Kyes et al. 2001). The lack of specific signs and symptoms means that a clinical diagnosis is not sufficient for cases of malaria and the disease should be diagnosed by microscopy of a thick blood film (Molyneux et al. 1993).

Pathogenesis of Severe Malaria

Severe or complicated malaria caused by *Plasmodium falciparum* infection should be regarded as such if any of the features in Table 1.1 are present (Molyneux et al. 1993).

Convulsion	Severe normochromic anaemia
Coma (cerebral malaria)	Shock
Renal failure	Haemoglobinuria
Acute Respiratory Distress Syndrome (ARDS)	Hypoglycaemia
Disseminated intravascular coagulation (DIC)	

Table 1.1. Signs of severe malaria. Adapted from Molyneux et al. 1993.

Plasmodium species modify the erythrocyte membrane greatly once they have invaded the cell. The membrane becomes less flexible and specific pores are created to transport nutrients required by the parasite into the cell. The parasite also inserts several of its own proteins into the erythrocyte membrane. One major *P. falciparum*-specific surface antigen is *Plasmodium falciparum* erythrocyte membrane protein-1 (*Pf*EMP-1), a large (200-400 kDa) trypsin-sensitive polypeptide which is encoded by between fifty to sixty *var* genes per haploid genome (Miller et al. 1994; Keyes et al. 2001; Chen et al. 2000). Each *var* gene comprises a large 5' exon that is highly variable and encodes the extracellular domain, and a highly conserved 3' exon encoding the transmembrane and intracellular domains. The two exons are separated by a conserved intron. The extracellular domain is highly variable, consisting of three building blocks: a globular N-terminal segment (NTS), followed by a cysteine-rich interdomain region (CIDR) that is novel to *Pf*EMP-1, and a Duffy binding-like antigen (DBL) domain also found in the erythrocyte binding antigen (EBA) family involved in erythrocyte invasion. DBL domains cluster as five types (α-ε) while the CIDR domains fall into three groups (α-γ). A single *Pf*EMP-1 protein may have several DBL and CIDR domains, although the first DBL domain (DBL1) to follow the NTS is always DBL1α and is followed by a CIDRα domain (Smith et al. 2001). Critically, at any one time primarily only one *var* gene will be expressed, with an imperfectly understood mechanism causing switching between *var* genes at rates that may be as high as 2% per generation in the absence of immune selection pressures (Keyes et al. 2001). The huge repertoire of variants with regular switches between them makes host recognition complicated, and leads to waves of antigenically distinct infections within a given patient, referred to as antigenic variation (Lavstsen et al. 2005).

Despite this variability, *Pf*EMP1 proteins share the same core function, mediating binding of the infected erythrocyte to one of multiple different receptors located on vascular endothelial cell surfaces, with different types of DBL and CIDR domains producing a diverse affinity for receptor families. Receptors that have been identified as binding

parasitised erythrocytes include CD36, ICAM-1, thrombospondin (TSP), E-selectin and VCAM-1 (Kyes et al, 2001; Smith et al. 2001; Chakravorty et al. 2005; Miller et al. 1994). Binding of parasitised erythrocytes to endothelial cells removes them from the circulation so they evade splenic clearance of erythrocytes (Keyes et al. 2001). While this sequestration is beneficial for the parasite, it can have significant pathogenic results, as sequestration of large numbers of infected erythrocytes can occlude the vessel and induce hypoxia, resulting in tissue death (Chen et al. 2000). Sequestration can occur in any tissue but the microvasculature of the brain is a prime target and this causes cerebral malaria, a complication of severe malaria that has mortality estimates of around 15%, even when treatment is initiated promptly (Snow et al. 1999; Carme et al. 1993; Sowunmi 1997).

The occlusive outcome of cytoadherence may be multiplied by the phenomenon of rosetting, the spontaneous binding of normal erythrocytes to malaria-infected erythrocytes, which has been associated with severe malaria (Chen et al. 2000). Like sequestration, rosetting seems to rely on *Pf*EMP-1 expressed by parasitized erythrocytes binding to multiple receptors on other uninfected erythrocytes. Five receptors on erythrocyte surfaces have been shown to be involved: blood group antigens A and B, complement receptor 1, heparin sulphate-like glycosaminoglycans (HS-like GAGs), and CD36 (Chen et al. 1998).

The Life Cycle of *Plasmodium*

Despite the connection between stagnant water and malaria transmission being identified in ancient times, the mystery surrounding the cause of malaria has only been unravelled in the last one hundred and fifty years. The reasons for this are, in part, due to the complexity of the life cycle of the organism *Plasmodium,* which undergoes stages in both the vertebrate host and the *Anopheles* vector.

Male *Anopheles* mosquitoes feed on plants, but females require a full blood meal to facilitate egg development. When a female *Anopheles* carrying *Plasmodium* parasites bites a vertebrate host, she injects an anticoagulant that inhibits coagulation factor Xa to ensure an even flow of blood from the vertebrate host (Stark et al. 1996). In addition to the anticoagulant, parasite sporozoites present in the female mosquito salivary gland are injected into the host. After transfer by circulation to the liver, sporozoites

specifically invade hepatocytes via interactions between sporozoite and hepatocyte cell surface molecules. Interactions between TSP domains on both the circumsporozoite protein (CSP) and on thrombospondin-related adhesive protein (TRAP) and heparin sulphate proteoglycans on hepatocyte surfaces are known to play a key role in facilitating invasion (Miller et al. 2002). Once the parasite is internalised in the hepatocyte, exoerythrocytic schizogony takes place (Fig. 1.1 Part A). This is an asexual replication that takes between five and fifteen days, depending on the species, to produce a pre-erythrocytic schizont containing between ten and thirty thousand merozoites (Fujioka et al. 2002). *Plasmodium vivax* and *Plasmodium ovale* may also produce hypnozoites, which are able to persist in the liver for many years. Reactivation of hypnozoites results in the production of merozoites and symptomatic relapses, possibly many years after inoculation (Imwong et al. 2007). *Plasmodium falciparum* and *Plasmodium malariae* lack the hypnozoite stage and consequently do not relapse. There are no symptoms or pathology during the liver stages of development, resulting in a lag between initial infection and the appearance of symptoms, which are specific to the blood stages.

Once liver development is complete, merozoites are released from pre-erythrocytic schizonts into the circulation where they invade erythrocytes, initiating the process of intra-erythrocytic development (Fig. 1.1 Part B). Intra-erythrocytic development is divided into three stages defined according to parasite morphology. The immature trophozoite or ring stage immediately follows merozoite invasion and lasts for approximately twenty four hours. Rings develop into mature trophozoites, which in turn develop into schizonts. The intra-erythrocytic cycle of the different human-infecting species of *Plasmodium* vary in duration. *Plasmodium falciparum, Plasmodium ovale* and *Plasmodium vivax* all have a forty eight hour cycle resulting in tertian fever. *Plasmodium malariae* has a seventy two hour life cycle leading to a quartan fever (Carter et al. 2002). As *P. falciparum* is the focus of this project, the molecular details of the intra-erythrocytic cycle discussed below are confined to this species.

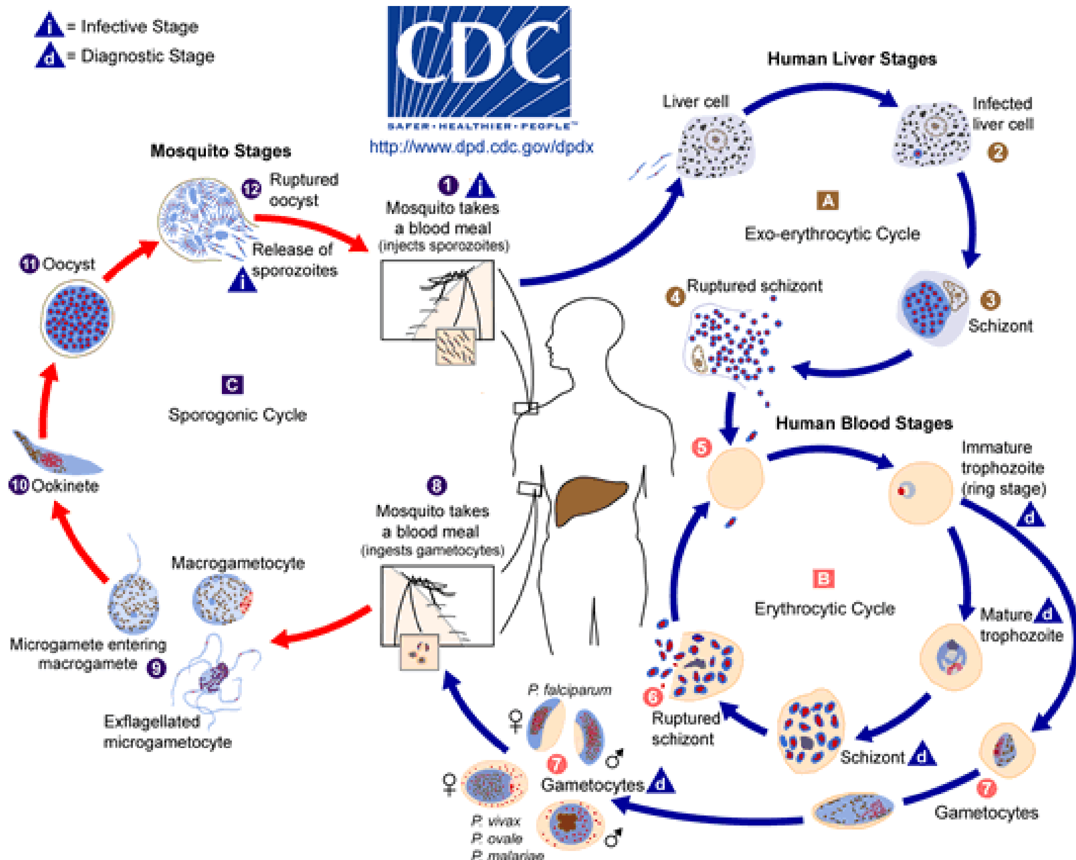

Fig. 1.1. The Life Cycle of *Plasmodium*. Reproduced from the Center for Disease Control & Prevention, Division of Parasitic Disease (DPDx) 2008.

During invasion the merozoite becomes surrounded by a membrane partly derived from the erythrocyte surface and partly from lipids expelled from its own secretory organelles; this becomes the parasitophorous vacuole membrane (PVM) (Ward et al. 1993). The PVM acts as the interface between the parasite and the erythrocyte, exchanging nutrients required from the host cell cytoplasm by the parasite with parasite-synthesised surface proteins being exported to the erythrocyte cell surface. As the parasite matures it develops a specialised organelle for the uptake of host cytoplasm called the cytostome. Double membrane vesicles pinch off the cytostome and fuse with a food vacuole. The food vacuole is acidic and contains a number of proteases, including members of the plasmepsin and falcipain families. Haemoglobin provides the primary source of amino acids for the parasite, with 60 – 80% of the erythrocytic haemoglobin being digested. However the parasite cannot catabolise the by-product of haemoglobin breakdown, haem. The toxic haem is therefore detoxified by a number of different mechanisms, the main one of which is conversion to haemozoin and storage as an inert crystalline structure (Egan et al. 2007; Fujioka et al. 2002).

During the trophozoite stage the parasite is highly metabolically active, and somewhere between twenty eight and thirty one hours DNA synthesis commences and continues for eight to ten hours (Arnot et al. 1998; Bozdech et al. 2003). In the schizont stage the parasite replicates and divides to form between sixteen and thirty two daughter merozoites. The erythrocyte eventually ruptures and releases the newly formed merozoites that then invade new erythrocytes (Fujioka et al. 2002).

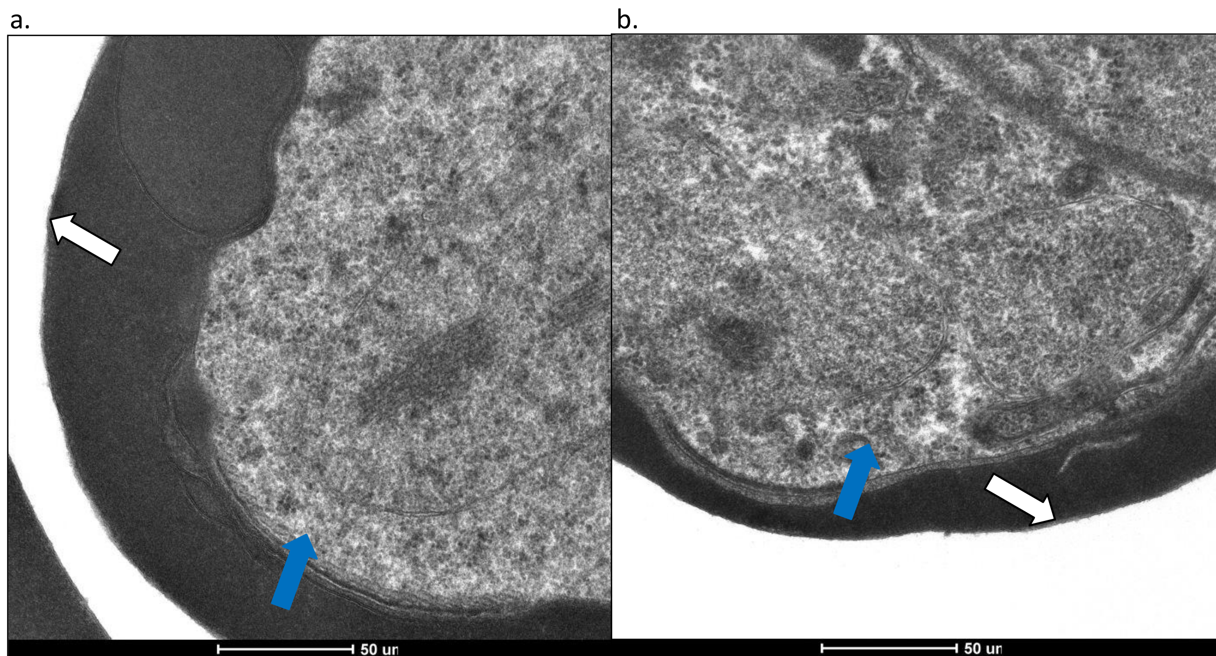

Fig. 1.2. Transmission electron micrographs of two Peruvian *P. falciparum* isolates in the erythrocytic stage. White arrows indicate the erythrocyte membrane and Blue arrows indicate the parasite. (a) Sample 3106. (b) Sample 3769.

This basic cycle of growth and multiplication is responsible for the increased parasite burden and symptoms of malaria, but it is not these asexual stages that are transmitted back to mosquitoes. Transmission depends on gametocytogenesis, whereby a small proportion of parasites do not develop into trophozoites post-invasion, but instead switch to develop into microgametocytes (male) and macrogametocytes (female). The trigger for gametocytogenesis is unknown, although a variety of changes *in vitro* have been shown to increase gametocyte production, including the addition of lymphocytes and red cell lysate, suggesting that it may be in part a parasite density-sensing mechanism. Different lab strains of *Plasmodium falciparum* also produce varying levels of gametocytes, with 3D7 and NF54 strains being reliable gametocyte producers, although their capacity to produce gametocytes diminishes with time in culture (Baker 2010).

When an *Anopheles* mosquito takes a blood meal from an infected individual it ingests erythrocytes containing gametocytes (Fig. 1.1 Part C). Within seconds of ingestion gametogenesis begins, triggered by a drop in temperature, an increase in pH and exposure to xanthurenic acid, and the gametocytes egress from the erythrocytes to form gametes (Billker et al. 1998). With three rounds of DNA replication, microgametogenesis produces eight separate male gametes. In the process, which lasts no more than twenty minutes, each microgamete also undergoes exflagellation (Guinet et al. 1996). Macrogametogenesis produces a single large, female gamete. Fertilisation takes place in the midgut lumen of the *Anopheles* mosquito to produce a diploid zygote which develops into an ookinete. The ookinete attaches to, and then migrates through the midgut epithelium prior to contact with the basal lamina and developing as an oocyst. The oocyst develops on the basal lamina for at least ten days, during which multiple nuclear divisions take place to produce haploid sporozoites. The mosquitoes provide essential nutrition to the oocyst during this time through haemolymph (Beier 1998).

Upon reaching maturity, sporozoites are released from the oocyst and migrate to the salivary glands via the mosquito haemolymph system. Sporozoites remain alive and infective in the salivary glands and are then ready to re-infect a host the next time the mosquito takes a blood meal (Beier 1998).

Molecular Basis of Erythrocyte Invasion

In addition to *Plasmodium,* the Apicomplexa phylum consists of mostly parasitic protists, including *Babesia*, *Cryptosporidium* and *Toxoplasma*. Members of the Apicomplexa phylum share a common apical complex, the function of which is to allow the parasite to invade host cells. Perhaps the most studied invasion of host cells by an Apicomplexan organism is that of *Plasmodium falciparum* merozoites into erythrocytes, primarily because this step of the *P. falciparum* life cycle, responsible for the symptoms and the majority of mortality associated with malaria, utilises many parasitic ligands that are potential vaccine targets (Cowman et al. 2006; Baum et al. 2002). Invasion of erythrocytes by merozoites can be broken down into four phases: primary contact, reorientation, secondary ligand interactions and host cell invasion (Cowman et al. 2006). The whole process, from exiting the schizont to erythrocyte recognition, attachment and invasion, is very rapid and can take

less than sixty seconds. This short time frame facilitates immune evasion, the merozoite surface being highly antigenic (Cowman et al. 2006).

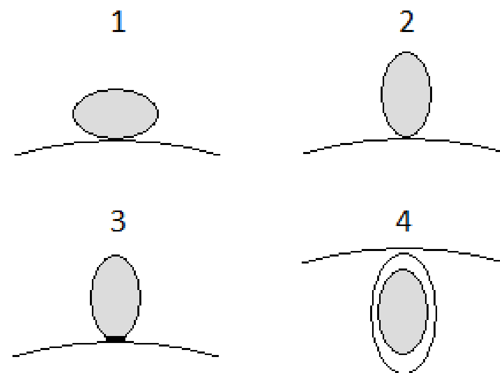

Fig. 1.3. The Stages of Erythrocyte Invasion. 1. Primary contact; 2. Reorientation; 3. Secondary ligand interactions; 4. Invasion.

1) Primary Contact

The extracellular surface of the merozoite is coated with glycosylphosphatidylinositol (GPI)-anchored proteins. Many of these GPI-anchored proteins also contain cysteine rich domains, suggesting a potential role in cytoadherence. The most abundant GPI-anchored protein is merozoite surface protein-1 (MSP-1) which is considered the most likely candidate for primary contact between merozoite and erythrocyte although a specific function is yet to be defined (Holder 1994; Cowman 2006). On the merozoite surface MSP-1 consists of 83-, 30-, 38- and 42 kDa (MSP-1-42) fragments. During invasion *Pf*SUB2 processes MSP-1-42 into MSP-1-19 and MSP-1-33 fragments. The MSP-1-19 fragment is attached to the merozoite surface via the GPI-anchor and upon invasion is taken into the erythrocyte. The remaining fragments are shed from the merozoite and not retained after invasion. A recent study has reported that MSP-1-42 is capable of binding heparin-like molecules that are present on the surface of the erythrocyte (Boyle et al. 2010). In addition to MSP-1, other members of the MSP family (-2, -4, -5, -10), *Pf*12, *Pf*38 and *Pf*92 are GPI-anchored proteins located on the merozoite surface and potentially involved in primary contact with erythrocytes (Gilson et al. 2006). Peripheral proteins secreted to the parasitophorous vacuole of schizonts and then bound to merozoite surface proteins may also initiate primary contact. Of these the likely candidates include MSP-6 and MSP-7, which form a complex with MSP-1 on the merozoite

surface (Trucco et al, 2001; Kadekoppala & Holder 2010). The difficulty of defining ligands responsible for primary contact suggests that the initial adhesion to erythrocytes is of low affinity (Cowman et al. 2006).

2) Reorientation

The merozoite is polarised and the organelles of invasion are concentrated at its apical end. Three separate secretory organelles have been identified at the apical end: micronemes, rhoptries and dense granules. There are up to forty micronemes per merozoite, and each one is about 160 nm long, shaped like a long-necked bottle (Bannister et al. 2003). The two rhoptries are larger pear shaped organelles around 570 nm in length, while the dense granules are spheroid vesicles around 120 nm in diameter.

After primary contact has been established at any point on the merozoite surface, for invasion to take place the merozoite must reorientate itself so that the apical end is juxtaposed to the erythrocyte surface. The mechanism of reorientation has not been defined, however a micronemal protein, *Pf*AMA-1, may be involved. Blocking the translocation of *Pf*AMA-1 from the microneme to the merozoite surface results in the inhibition of invasion after initial attachment. Interaction between *Pf*AMA-1 and rhoptry proteins is also required for activation of rhoptry secretion in *Toxoplasma gondii* and *P. falciparum* (Mital et al. 2005; Richard et al. 2010).

3) Secondary Ligand Interactions

After reorientation has taken place the merozoite must activate the invasion process and this is most probably achieved through specific interactions between merozoite ligands and erythrocyte receptors. Each interaction between a different pair of ligands is known as an invasion pathway. *Plasmodium vivax* merozoites rely primarily on binding of the Duffy-binding protein to the erythrocyte Duffy blood group antigen, meaning that Duffy-negative humans are refractory to *P. vivax* infection (VanBuskirk et al. 2004; Miller et al. 1976). Conversely in *Plasmodium falciparum* two merozoite protein families have been identified in this process, namely the DBL protein family and the reticulocyte binding protein homolog (*Pf*Rh or *Pf*RBL) that are able to interact with a number of different erythrocyte receptors. The upshot of multiple invasion pathways is that there is significant redundancy and, unlike

P. vivax, no single human genotype confers resistance to *P. falciparum* (Cowman et al. 2006; Rayner 2009). The redundancy of invasion pathways maximises the probability of invasion by each merozoite by limiting the effects of host receptor polymorphisms, variations in receptor expression and immune responses. If, by one of these methods, one invasion pathway becomes blocked, there remains the possibility of utilising a different pathway (Cowman et al. 2006).

The DBL family of erythrocyte binding antigens (EBA) includes EBA-175, EBA-140 and EBA-181. In EBA-175, cysteine rich, dual DBL domains mediate binding specifically to glycophorin A (Sim et al. 1994). EBA-140 binds glycophorin C and it appears that, in Melanesian populations, individuals with the Gerbich negative blood group (lacking glycophorin C) arose through natural selection against severe malaria due to the ablation of this *P. falciparum* invasion pathway (Maier et al. 2002). EBA-181 binds an unknown receptor.

The *Pf*Rh family of proteins originating from the rhoptry neck are also capable of mediating invasion pathways. Six *Pf*Rh proteins have been identified; *Pf*Rh 1-5 including *Pf*Rh2a and *Pf*Rh2b. While no evidence exists for the binding of *Pf*Rh2a or *Pf*Rh2b to erythrocytes, disruption of their genes leads to a switch in the erythrocyte receptor used for invasion (Duraisingh et al. 2003). *Pf*Rh1 and *Pf*Rh5 bind to unknown erythrocyte receptors while *Pf*Rh4 has recently been shown to bind to complement receptor 1 (CR1) (Rayner et al. 2001; Tham et al. 2009; Rodriguez et al. 2008; Tham et al. 2010).

4) Host Cell Invasion

Subsequent to apical ligand interactions, the parasite employs microneme and rhoptry proteins to create a tight junction around the apical end of the parasite in a calcium-dependent process (Moskes et al. 2004). The invasion of the merozoite is driven by an actin-myosin motor complex that is conserved across the Apicomplexa phylum (Green et al. 2006; Baum et al, 2006; Jones et al 2006). Once an invasion ligand has established contact with an erythrocyte receptor, a protein links the ligand through the merozoite surface membrane to a molecule capable of interacting with an F actin chain. In *Plasmodium* sporozoites these roles are performed by TRAP and aldolase, respectively. Myosin A heads that are anchored

to the inner membrane complex of the merozoite by a complex of MTIP, GAP45 and GAP50 proteins can bind to the filamentous actin chain and generate force, pulling the invasion ligand towards the posterior end of the merozoite and having the overall effect of driving the merozoite forward into the erythrocyte.

Variation in Invasion Pathways

The ligand interactions of the DBL and *Pf*Rh families are candidates for the activation of erythrocyte invasion and have been studied in depth. While the erythrocyte receptors for some ligands have been identified, other interactions have remained elusive. The study of invasion pathways has been carried out in several ways: (1) erythrocyte binding assays have been used to show binding of parasite ligands to erythrocytes (Gaur et al. 2007; Mayer et al. 2002). (2) antibodies against parasite ligands have been used to block binding to erythrocytes (Gaur et al. 2007) or invasion (Narum et al. 2000) and (3) antibodies against erythrocyte receptors have been used to block invasion (Holt et al. 1989).

Ligand interactions can also be studied when erythrocytes are deficient in particular surface molecules involved in merozoite invasion. There are certain blood groups where this happens naturally, e.g. Duffy negative populations among black West Africans; Gerbich negative (glycophorin C deficient), and these blood groups have been used to study invasion. As an alternative to using rare blood types, erythrocytes can be artificially depleted of receptors by treating with enzymes that modify surface molecules. The enzymes most commonly used for this purpose are neuraminidase, trypsin and more recently chymotrypsin.

As a surface antigen of the influenza virus, neuraminidase is commonly encountered as a viral antigenic determinant, along with its other major surface antigen, haemagglutinin. Neuraminidase hydrolyses sialic acid residues from cell surface glycoproteins. For bacteria such as *Vibrio cholerae* this reveals the binding site required for cell invasion (Moustafa et al. 2004). However, the sialic acid residues of certain erythrocyte receptors are essential to merozoite binding and therefore their removal can inhibit invasion; the best established example is the interaction between EBA-175 and glycophorin A, which can be abolished by neuraminidase treatment (Kain et al. 1993). Trypsin and chymotrypsin are endopeptidases

and cleave the peptide backbone of several surface proteins (Baum et al. 2002). The sensitivities of receptors to these enzymes are listed in Table 1.2.

Merozoite ligand	EBA175	EBL-1	EBA140	EBA181	*Pf*Rh1	*Pf*Rh2b	*Pf*Rh4	*Pf*Rh5
RBC receptor	GYP A	GYP B	GYP C	E	Y	Z	CR1	?
Neuraminidase	S	S	S	S	S	R	R	R
Trypsin	S	R	S	R	R	R	S	S
Chymotrypsin	R	S	R	S	R	S	S	R

Table 1.2. A summary of merozoite – erythrocyte ligand interactions and their enzyme sensitivity. S = sensitive; R = resistant; GYP = glycophorin. EBA = erythrocyte binding antigen. EBL = erythrocyte binding ligand; CR1 = complement receptor 1. (Rodriguez et al. 2008; Mayer et al. 2004; Tham et al. 2009; Baum et al. 2002; Cowman et al. 2006; Maier et al. 2002; Duraisingh et al. 2003; Tham et al. 2010).

P. falciparum lab strains are known to vary in their ability to utilise different invasion pathways; for example HB3 and 7G8 both use sialic acid-independent pathways and invade neuraminidase-treated cells, whereas Dd2 is completely dependent on receptors bearing sialic acid and can be completely inhibited by neuraminidase treatment. The variation seen in lab isolates led to the first studies of field isolates, which were set up to determine whether *P. falciparum* used alternative invasion pathways or whether field isolates were entirely dependent upon EBA-175 and glycophorin A.

Field studies of erythrocyte invasion have been undertaken in sub-Saharan Africa (Baum et al. 2003; Gomez Escobar et al. 2010; Deans et al. 2007; Bei et al. 2007; Jennings et al. 2007), India (Okoyeh et al. 1999) and Brazil (Lobo et al. 2004). All used neuraminidase and trypsin treatment of erythrocytes to study invasion, while chymotrypsin was also used as a treatment in the studies performed since 2005. All studies found that invasion pathway diversity in field isolates is comparable to that seen in lab strains and all studies found multiple invasion pathways utilised across the cohort of isolates. The predominant invasion phenotype present in The Gambia was determined as neuraminidase- and trypsin-sensitive, suggesting that the majority of isolates are dependent on glycophorin A and C for erythrocyte entry. However, in the 2010 study from The Gambia, most isolates were also unable to invade chymotrypsin-treated cells (Baum et al. 2003; Gomez Escobar et al. 2010). In contrast, most Kenyan field isolates displayed resistance to neuraminidase and sensitivity to trypsin and chymotrypsin, a phenotype consistent with *Pf*Rh4 utilisation Deans et al. 2007). The major phenotype present in Brazil was neuraminidase- and trypsin-sensitive but

chymotrypsin-resistant, more suggestive of glycophorin A and C utilisation (Lobo et al. 2004).

Several of these studies have combined invasion phenotyping with either parasite genotyping or analysis of ligand expression. Genotyping was primarily used to determine whether samples contained multiple parasite infections by typing of MSP-1 and MSP-2 polymorphic loci, although one study also typed three known SNPs of *eba*-175 (Baum et al. 2003). Ligand expression investigation has only involved analysis of selected ligands, for example the expression levels of eight members of the DBL and *Pf*Rh families (Gomez Escobar et al. 2010). Other phenotypic associations have been made between parasite multiplication rate (PMR) and severity of disease. In Africa two studies have found no such associations, while in Thailand PMR was found to be positively correlated with disease severity (Deans et al. 2007; Deans et al. 2006; Chotivanich et al. 2000).

A limitation of studies to date is that between 11 and 38 field isolates have been used in each, until the latest study from The Gambia which involved 163 field isolates (Gomez Escobar et al. 2010). The primary reason for low sample numbers is the time required to phenotype samples using microscopy-based techniques, and the low culture adaptation rates of field isolates (generally around 50%). The higher numbers of samples in the second Gambian study was made possible by the use of flow cytometry to generate data from invasion assays, greatly reducing the time taken to count cells. The use of this method also increases the accuracy of the assay, with many more cells counted and human bias eliminated (Gomez Escobar et al. 2010).

Despite the benefits of redundancy in *P. falciparum* invasion pathways being clear, the mechanisms of variation are still not fully understood. Single nucleotide polymorphisms in Region II of EBA-175 have been observed to direct binding to different erythrocyte receptors, suggesting that an underlying mechanism of genome sequence polymorphism was responsible for variations in invasion pathways (Mayer et al. 2002). Additionally, *Pf*Rh5 polymorphism has been observed to cause switching of host receptor recognition and consequently affect the ability of human *P. falciparum* strains to infect *A. nancymaae* monkeys (Hayton et al. 2008). As well as the presence of sequence polymorphism,

differential expression of ligands has also been detected across the *Pf*Rh family in different laboratory strains and is an alternative mechanism by which variation in invasion pathways may occur (Taylor et al. 2002). Variation in the expression of the *Pf*Rh and DBL family of ligands has been shown to correlate with different invasion pathways utilised, with high levels of *eba*-175 expression corresponding to sialic acid-dependent invasion, while sialic acid-independent invasion is more likely to occur when there are high levels of *Pf*Rh2b and *Pf*Rh4 expression (Nery et al. 2006).

Rationale for This Study

Most of the studies of invasion phenotypes in field samples mentioned previously have combined the technique with an investigation of polymorphism or expression of ligands with the goal of identifying genotypes that might explain the observed phenotypic variation. However, in each case this has involved investigation of a handful of ligands, for example the Gambian study (Gomez-Escobar et al. 2010) investigated the expression of eight proteins from the DBL and *Pf*Rh families. It is clear that to identify genotype-phenotype associations in an unbiased manner, whole-genome approaches are needed to assess variation across all genes. Furthermore, phenotypes vary by two mechanisms: sequence polymorphism and/or differential expression of ligands, suggesting that analysis of the transcriptome is also needed. This study is a pilot step towards whole genome genotype-phenotype association studies for erythrocyte invasion, using *Plasmodium falciparum* samples from Peru. It is powered by three fundamental considerations: the development of high throughput cost-effective DNA sequencing technologies (so-called next generation, or second generation sequencing) as an approach to measure genetic variation at a whole genome and transcriptome level, the development of a high-throughput flow cytometry-based assay to phenotype erythrocyte invasion in a highly reproducible and sensitive manner, and sourcing parasites from low transmission regions in Peru where genetic diversity is limited and linkage disequilibrium is high.

1) The Sequencing Revolution

DNA sequencing has undergone a revolution since Fred Sanger announced that the bacteriophage ϕX174 had been sequenced in 1977 using a method based on synthesis of

random shotgun sub-cloned DNA fragments and chain termination with dideoxynucleotides (Sanger et al. 1977). The Human Genome Project, started in 1990, was completed in a little over a decade, predominantly using semi-automated shotgun sequencing with fluorescently labelled dideoxynucleotide chemistry, based on the Sanger method (Shendure & Ji 2008). Improvements in the technology meant that by the conclusion of the Human Genome Project read lengths of up to 1,000 base pairs (bp) at accuracy rates of 99.999% were achieved on capillary sequencers capable of sequencing 96 samples in parallel and up to 0.5 million bases per day. Since then sequencing technology has been transformed by the introduction of three highly parallelised sequencing platforms using novel sequencing chemistries: Illumina's (formerly Solexa) polymerase-based sequencing-by-synthesis, pyrosequencing by Roche (454), and ligation-based sequencing from Applied Biosystems (AB SOLiD). The output on all platforms has been increasing steadily and in mid-2010 a single run on an Illumina Genome Analyzer II platform yields approximately 50,000 megabases and takes twelve days to complete. Consequently, it is now possible to sequence a human genome, which initially took years, to a depth of 15-fold coverage on a single machine in less than two weeks (Mardis 2008). However there are drawbacks; the short read length produced by Illumina (36 – 100 bp) and SOLiD (50 bp) platforms means that genome assembly is more difficult, error rates have risen, and accurate assignment of sequence to repeat-rich non-coding regions is often not possible. Some of these issues are reduced by the Roche 454 platform, which generates sequence reads up to 500 bp. However a 454 run is 40 times more expensive than Illumina and high error rates are observed when sequencing homopolymers (e.g. AAA or TTT sequences), making it of limited use for the highly AT-rich *P. falciparum* genome.

While these second generation sequencing platforms are undoubtedly a huge leap forward, technology is being developed that would see sequencing carried out in real-time and generate sequence read lengths of tens of thousands of bases. Pacific Biosciences have successfully tested a method using DNA polymerase incorporation of bases with fluorescent labelled phosphate groups (Shendure & Ji 2008), which will be released this year and other groups are looking at the potential of nanopore-based sequencing.

The DNA sequence of the 14 chromosome, 22.8 Mb genome of *Plasmodium falciparum* lab strain 3D7 was published in 2002. 5,300 protein-encoding genes have been annotated in the genome, which has a nucleotide composition of over 80% A or T, making it the most (A+T)-rich genome that has been sequenced (Gardner et al. 2002). In the past 2 years the Sanger Institute has developed Illumina sequencing of *P. falciparum* and has sequenced more than 400 isolates, primarily from samples taken directly from patients (Kwiatkowski et al., unpublished data). The availability of these high-throughput technologies at the Wellcome Trust Sanger Institute mean that *P. falciparum* parasites that are being phenotyped for erythrocyte invasion can be sequenced at the same time, allowing a whole genome assessment of genetic variation, and rapid comparison with a large number of other *P. falciparum* genomes.

2) High Throughput Assay for Phenotyping *P. falciparum* Erythrocyte Invasion

As discussed above, phenotyping invasion involves the treatment of target erythrocytes with the enzymes neuraminidase, trypsin and chymotrypsin and quantifying merozoite invasion into these target cells. In previous invasion assays parasitised donor red blood cells have been treated with trypsin and neuraminidase to prevent invasion into the donor cells so that the target erythrocytes are the only subset being reinvaded (Duraisingh et al. 2003). Alternatively, purification of schizonts on a density gradient has been performed to eliminate un-invaded erythrocytes from the donor population (Lobo et al. 2004; Deans et al. 2007; Okoyeh et al. 1999). The former has the disadvantage that neuraminidase treatment of parasitised red blood cells has been shown to affect the parasite's ability to re-invade untreated erythrocytes (Theron et al. 2010). The latter method of schizont purification is inefficient, requiring large volumes of culture, and the accuracy of the assay depends mainly on the quality of purification. Parasite handling may also have a detrimental effect on invasion.

Previous invasion assays have also counted the number of parasites by microscopy, but this is labour intensive and can decrease sensitivity, as only 1,000 – 2,000 erythrocytes are routinely counted on each slide (Okoyeh et al. 1999; Lobo et al. 2004; Deans et al. 2007; Baum et al. 2002). An alternative method of parasite counting is using fluorescent DNA dyes and flow cytometry, taking advantage of the fact that erythrocytes lack DNA (Gomez-

Escobar et al. 2010; and others). Using fluorescence detection greatly improves the resolution, while flow cytometry allows counting of thousands of cells per second increasing the throughput and the reproducibility of the assay.

Theron et al. (2010) at the Sanger Institute have developed the concept of two-colour flow cytometry invasion assays. This approach combines counting parasites labelled with fluorescent DNA dyes with the use of fluorescent cell dyes to label target erythrocytes and distinguish them from donor cells present in the starting culture. The cell dyes were chosen for their ability to label cells cytoplasmically so as to avoid possible interference with invasion and to have minimal emission overlap with the DNA dyes. In this assay target cells are labelled with 7-hydroxy-9*H*-(1,3-dichloro-9,9-dimethylacridin-2-one) succinimidyl ester (DDAO-SE), treated with one of the standard invasion-inhibitory enzymes, then added to an equal volume of unlabelled erythrocytes taken from a *P. falciparum* culture. After 48 hours of culture parasites are stained using either Hoechst 33342 or SYBR Green I, and invasion into labelled cells quantified by flow cytometry. The resulting assay is sensitive, with minimal variation between replicates, and can be performed in 96 well plates to increase throughput.

3) Peru as a Source of *Plasmodium falciparum* Samples

Venous blood samples from patients with malaria were collected in Iquitos, in the Loreto department at the heart of the Peruvian Amazon. *P. falciparum* was briefly eradicated from Loreto in the late 1980s but in the early 1990s began to make a comeback, reflecting the cessation of DDT spraying in 1988. *Anopheles* numbers also surged during the 1990s and the highly competent and anthropophilic (preferentially bites humans) *Anopheles darlingi* spread from the Brazilian Amazon and now predominates over at least five other endogenous *Anopheles* species present in Loreto. As in other areas of the world there is seasonal variation in incidence, with the greatest number of cases occuring approximately two months after the water level of the Amazon River peaks (Guarda et al. 1999; Branch et al. 2005).

P. vivax is the predominant species found in Peru. However, as the incidence of malaria rose in the 1990s to epidemic levels, the proportion of cases attributed to *P.*

falciparum increased from 1.6% in 1992 to 44.8% of a total of 158,115 cases in 1997. The number of cases fell substantially in 1998 with only 84,059 reported, a third due to *P. falciparum.* As well as increased vector numbers and termination of DDT spraying, other factors contributing to this rise include the El Niño climate phenomenon (that caused torrential rain in coastal Peru in 1997) and the emergence of drug resistance: by 1996 there was widespread resistance to pyrimethamine-sulfadoxine and chloroquine, while mefloquine and quinine remained effective (Guarda et al. 1999; Branch et al. 2005; Perekh et al. 2007). Since the 1995 – 1998 epidemic, malaria has remained hypoendemic in Peru (<0.5 infections / person / year), with incidence varying between 5 and 50 cases / 1000 persons. This low incidence, combined with the known low genetic diversity in South American *P. falciparum* samples (Mu et al. 2005), makes this an ideal setting for preliminary genotype-phenotype association studies for invasion. In such an environment, infections are genetically simple, without the multiple overlapping infections that are a hallmark of *P. falciparum* in hyperendemic regions such as much of sub-Saharan Africa. Despite significant spatial and temporal clustering of malaria cases in Peru, only 4% of individuals were found to have more than one *P. falciparum* infection (Branch et al. 2005).

The Malaria Immunity and Genetics in the Amazon (MIGIA) longitudinal epidemiological study has been underway near Iquitos since 2005, collecting parasite and serum samples from *P. falciparum* and *P. vivax* infected patients using both active (house visit) and passive (clinic patients) case detection (Branch et al. 2005). Critically, *P. falciparum* samples with high parasitaemia are placed in *in vitro* culture, and multiple samples have been grown for several cycles and frozen. For this study 46 frozen isolates were transferred from Iquitos to Sanger for invasion phenotyping and whole genome sequencing.

Study Outline

This study was set up with the aim of generating genotype–phenotype correlations for erythrocyte invasion pathways. *P. falciparum* isolates received from Peru were thawed and grown in culture and isolates that were cultured successfully were investigated as follows:

- The two-colour flow cytometry phenotyping platform was used to generate a profile of invasion for each isolate, based on its ability to invade enzyme-treated erythrocytes, and to compare the parasite multiplication rate (PMR) of each isolate.
- DNA from each isolate was extracted and sequenced using the Illumina platform. Whole genome analysis was carried out using bioinformatics tools developed at the Wellcome Trust Sanger Institute specifically for the analysis of *P. falciparum* genomes.
- Parasite schizonts were isolated and RNA was extracted. This will be reverse transcribed to a cDNA library which will be sequenced using the Illumina platform allowing analysis of the whole transcriptome to be accomplished.
- Once investigations of each isolate were complete, multiple aliquots were frozen for future investigation.

The data generated will provide in-depth correlations between variation of invasion phenotype, genomic polymorphism and parasite ligand expression.