1 INTRODUCTION

1.1 *Plasmodium vivax*: a major and neglected human pathogen

Plasmodium vivax malaria afflicts millions of people throughout the world each year, particularly in Asia, the Pacific, and Latin America, and continues to be recognized as a neglected tropical disease. The global burden of *P. vivax* costs an estimated US$1,400,000,000–$4,000,000,000 per year (Price et al., 2007), yet *P. vivax* research receives only a small fraction of the overall malaria research and development funding, estimated at around 3% of the total malaria research funds from 2006-2009 (PATH, 2011). While there is widespread agreement that an effective vaccine would be a critical component of any global malaria eradication effort, development of a *P. vivax* vaccine has been hindered by a lack of understanding of basic *P. vivax* biology. This thesis aims to begin filling in this significant knowledge gap by identifying novel *P. vivax* vaccine candidates.

1.1.1 Global burden of malaria historically

Malaria is caused by infection from single-celled eukaryotic parasites from the genus *Plasmodium*. A wide variety of vertebrate species from mammals, birds, and reptiles are naturally infected with *Plasmodium* parasites, and there are hundreds of *Plasmodium* species described to date (Levine, 1988). Six species are known to cause human malaria: *P. falciparum, P. vivax,* two distinct species of *P. ovale* (Sutherland et al., 2010), *P. malariae,* and *P. knowlesi,* with the latter primarily infecting Southeast Asian macaques, but increasingly detected as a zoonotic source of human infections (Millar and Cox-

Singh, 2015, Singh et al., 2004). However, *P. falciparum* and *P. vivax* cause the vast majority of infections, and will be the focus of this introduction.

The transmission dynamics of *P. falciparum* and *P. vivax* have changed significantly over the last century. The essential role of mosquitoes to the *Plasmodium* life cycle (Figure 1.2) was discovered by Ronald Ross in 1897 and led to significant vector control measures, such as the draining of swamps and the widespread of use of insecticides. Cheap and effective drugs, with chloroquine approved for use in 1946, being the exemplar, added a further major weapon in malaria control. Chloroquine, the insecticide DDT, and active case surveillance together underpinned a major world eradication campaign from 1955-1964 by the World Health Organization (WHO) termed the "Global Malaria Eradication Programme" (GMEP), which completely eliminated malaria from some areas (WHO, 1999). Measures such as these have ultimately eliminated malaria in over 50 countries since 1900, shrinking the area for risk of malaria infection by half and saving countless lives (Hay et al., 2004).

While some countries maintained gains and approached elimination, dozens of countries experienced a resurgence in transmission due to a combination of factors (WHO, 1999, Smith et al., 2013). In particular, and as detailed further below, the development of drug resistance in parasites and insecticide resistance in mosquitoes has consistently plagued malaria control efforts worldwide. These factors, coupled with insufficient funding and chronic health infrastructure problems, contributed to the failure of the GMEP, which was discontinued in favor of control programs. Global funding for malaria control declined in the 1970s and 1980s, due in part to the perceived failure of the GMEP and the global economic crises.

The funding landscape has improved significantly over the last 10 years as malaria has, once again, found a place on the global agenda. The Bill and Melinda Gates Foundation committed US$2.5 billion in basic research and control measures since 2000, and the US President's Malaria Initiative (PMI) started in 2005 distributes hundreds of millions in control-related funding each year (US$618 million in 2014). Additional commitments by the Global Fund and World Bank have further contributed to the improvement in funding for both research and control measures, which now nears US$2 billion annually (Greenwood and Targett, 2011). Malaria mortality declined 32% from 2004-2010 in line with the surge in funding (Murray et al., 2012). A variety of factors contributed to this

decline, including increased use of insecticide-treated bed nets (ITNs), changing transmission dynamics, the use of highly effective treatments such as artemesinin combination therapy (ACT), and improved health infrastructure (Murray et al., 2012). However, recent gains are under threat from the emergence of artemisinin-resistant *P. falciparum* parasites in Southeast Asia and the rise in insecticide resistance in *Plasmodium*-spreading mosquito populations (Greenwood and Targett, 2011). *P. vivax* malaria is also proving difficult to control with initial evidence suggesting that as *P. falciparum* transmission declines, *P. vivax* often assumes a more dominant role in disease incidence (Maude et al., 2014). This is supported by previous eradication campaigns where *P. vivax* transmission persisted long after *P. falciparum* was eliminated or reduced (Yekutiel, 1980). These challenges underline the need for an efficacious vaccine targeting both *P. vivax* and *P. falciparum* to ultimately control this global disease threat.

1.1.2 *P. vivax* epidemiology

At present, more than 200 million cases of malaria occur annually, and up to 45% of the world population is at risk of *Plasmodium* infection (WHO, 2014). *P. falciparum* causes the vast majority of malaria mortality with over 500,000 deaths each year. Over 90% of these deaths occur in Africa, and the majority of deaths occur in children under 5 years of age, although the age spectrum of mortality is changing with changing transmission patterns (WHO, 2014). Malaria deaths disproportionately affect populations in extreme poverty (those living on less than US$1.25 per day) (WHO, 2014). *P. vivax* malaria, in contrast, has been described as "benign" because it causes relatively few deaths, although a growing body of work is challenging that description due to the large economic and social repercussions of *P. vivax* malaria morbidity, as well as evidence of severe cases (Kochar et al., 2005, Williams et al., 1997, Sharma and Khanduri, 2009, Beg et al., 2002, Kumar et al., 2007, Price et al., 2009). Recent reviews of case reports showed that while *P. vivax* causes far fewer cases of severe disease than *P. falciparum*, the outcomes of severe *P. vivax* and *P. falciparum* cases are similarly poor (Rogerson and Carter, 2008, Tjitra et al., 2008, Nurleila et al., 2012).

P. vivax causes the majority of malaria cases outside Africa; approximately 2.5 billion people are at risk of *P. vivax* infection annually, 80% of whom live in South and Southeast Asia (Gething et al., 2012). Illness caused by *P. vivax* is estimated to affect 80–390 million people each year throughout the tropics, primarily in Southeast Asia and the

Western Pacific (Price et al., 2007, Mendis et al., 2001, Hay et al., 2004). *P. vivax* has a much larger geographical spread than *P. falciparum*, affecting 95 countries across 4 continents. This is partly due to *P. vivax* parasites' ability to survive more readily in seasonal and temperate transmission environments; specifically their dormant liver forms (described below) act as a reservoir, enabling *P. vivax* to relapse into the bloodstream months or even years after the initial infection (Gething et al., 2012, Krotoski et al., 1982). *P. vivax* also has a lower temperature limit in its mosquito host (14.5°C) compared to *P. falciparum* (16°C), contributing to *P. vivax*'s expanded range into more temperate climates (Sinden, 2002, Guerra et al., 2006).

P. vivax maintains stable transmission in areas with much lower endemicity than *P. falciparum*, and *P. vivax* infection prevalence rates rarely exceed 7% (Figure 1.1). This is in stark contrast to *P. falciparum*, which can have prevalence rates as high as 70% across parts of Africa (Gething et al., 2011). It is important to note that most surveys employed microscopy or RDT to determine *P. vivax* malaria incidence, which may significantly underestimate the *P. vivax* burden due to the generally lower parasite densities achieved by this species (Mueller et al., 2009b, Gething et al., 2012). *P. vivax* is noticeably absent throughout most of Africa due to the high rate of the Duffy negative allele, a blood group central to *P. vivax* infection, although recent data from Madagascar and other African countries suggest that the barrier in Duffy-negative individuals is not as absolute as previously thought (Menard et al., 2010, Ngassa Mbenda and Das, 2014, Woldearegai et al., 2013). Nevertheless, selective pressure from *P. vivax* malaria appears to have driven this polymorphism to fixation in the majority of African populations. *P. falciparum* malaria has similarly selected for several other erythrocyte polymorphisms that protect against this disease, such as sickle-cell trait (hemoglobin S, HbS), hemoglobin C (HbC), and α-thalassemia [reviewed in (Taylor et al., 2013)].

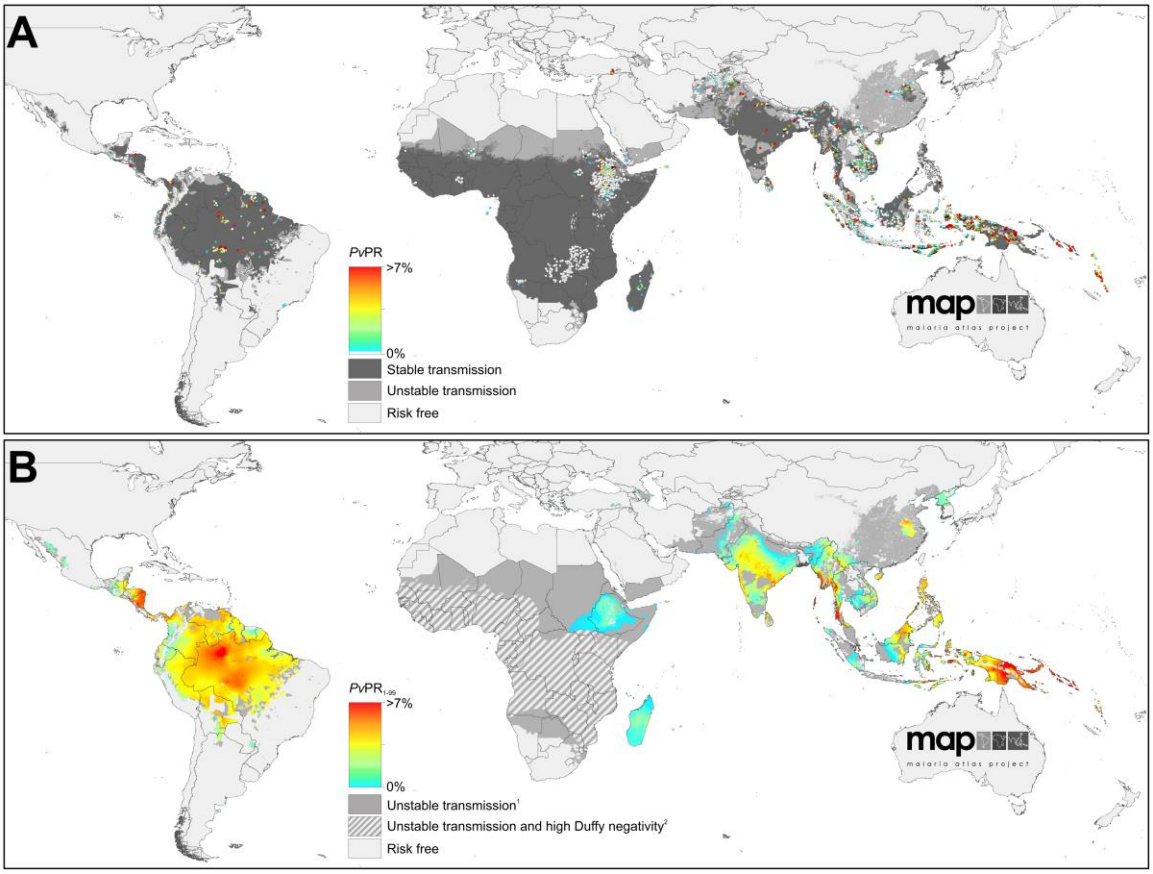

Figure 1.1: *P. vivax* endemicity in 2010

(A) Spatial limits of *P. vivax* malaria based on temperature and aridity data, as well as community surveys (from 1985 to 2010) for annual *P. vivax* parasite incidence (PvAPI). Individual survey data are presented as dots with a color continuum from green to red for a > 0% to > 7% *P. vivax* parasite rate (*Pv*PR*) and white for surveys with zero *P. vivax* cases (see map legend). Transmission areas were defined as stable (dark grey, where PvAPI ≥ 0.1 per 1,000 pa), unstable (medium grey, where PvAPI < 0.1 per 1,000 pa), or having no risk (light grey, where PvAPI = 0 per 1,000 pa). (B) Model-based geostatistics for *Pv*PR, for people aged 1-99 years using the color scale from (A). Hatched areas represent areas with Duffy negativity gene frequency exceeding 90%. Reprinted from (Gething et al., 2012) under the Creative Commons Attribution (CC BY) license. **Pv*PR, the proportion of randomly sampled individuals in a surveyed population with patent parasitemia in their peripheral blood, as detected by microscopy or rapid diagnostic test (RDT)

1.1.3 *Plasmodium* life cycle

Plasmodium species share similar life cycles and develop through several morphologically distinct stages across their vertebrate and mosquito hosts (Figure 1.2) (Mueller et al., 2009a). When an infected female *Anopheles* mosquito takes a blood meal from a human, haploid, motile sporozoites in the mosquito saliva are injected into the skin (Prudencio et al., 2006). The sporozoites burrow through the vascular tissue and

enter the bloodstream about 15 minutes later (Vanderberg and Frevert, 2004). Those parasites that reach the liver and traverse the sinusoidal cell layer to infect hepatocytes, remain largely undetected by the human immune system (Tavares et al., 2013). Once inside hepatocytes, the parasites differentiate and proliferate for 2 days to 3 weeks (depending on the species) into multi-nucleated hepatic schizonts, which each contain many thousands of infectious forms called merozoites packed into vesicles called merosomes (Tarun et al., 2006, Prudencio et al., 2006). The infected hepatocyte detaches, and merosomes begin budding from it into the bloodstream, where they ultimately rupture (Sturm et al., 2006). The infective merozoites are then released, whereupon they attach to and invade erythrocytes (often of specific maturities) in an extremely rapid process that only takes about 1 minute (Weiss et al., 2015). This pattern of multiplication in the liver does not universally occur; in the cases of *P. vivax* and *P. ovale*, for example, some liver-stage parasites convert to dormant forms called hypnozoites, which can remain in situ for months or even years (Krotoski, 1985). These dormant forms are one of the unique challenges of *P. vivax* control, as they enable disease to relapse even in the absence of active transmission.

Upon erythrocyte invasion parasites are described as "ring-stage" (~0-24 hours for *P. falciparum* and *P. vivax*), so called for their appearance when stained with Giemsa and viewed by light microscopy (LM). This is followed by the most metabolically active trophozoite stage, in which the parasite digests host erythrocyte hemoglobin, resulting in crystals of hemozoin in the parasite food vacuole. Giemsa-stained trophozoites appear dark and circular by LM (~24-36 hours for *P. falciparum* and *P. vivax*). The parasites then undergo replication and segmentation into schizonts containing multiple nuclei (~36-48 hours for *P. falciparum* and *P. vivax*). A single schizont can contain 16-32 daughter merozoites in a range of sizes depending on the species. Upon fully maturing, proteases break down the parasitophorous vacuole membrane followed by the host erythrocyte membrane resulting in the daughter cells' "egress" into the bloodstream (Blackman, 2008, Blackman and Carruthers, 2013), where they invade new erythrocytes (Gilson and Crabb, 2009). This cycle can continue indefinitely unless the infection is treated or cleared by the host's immune system.

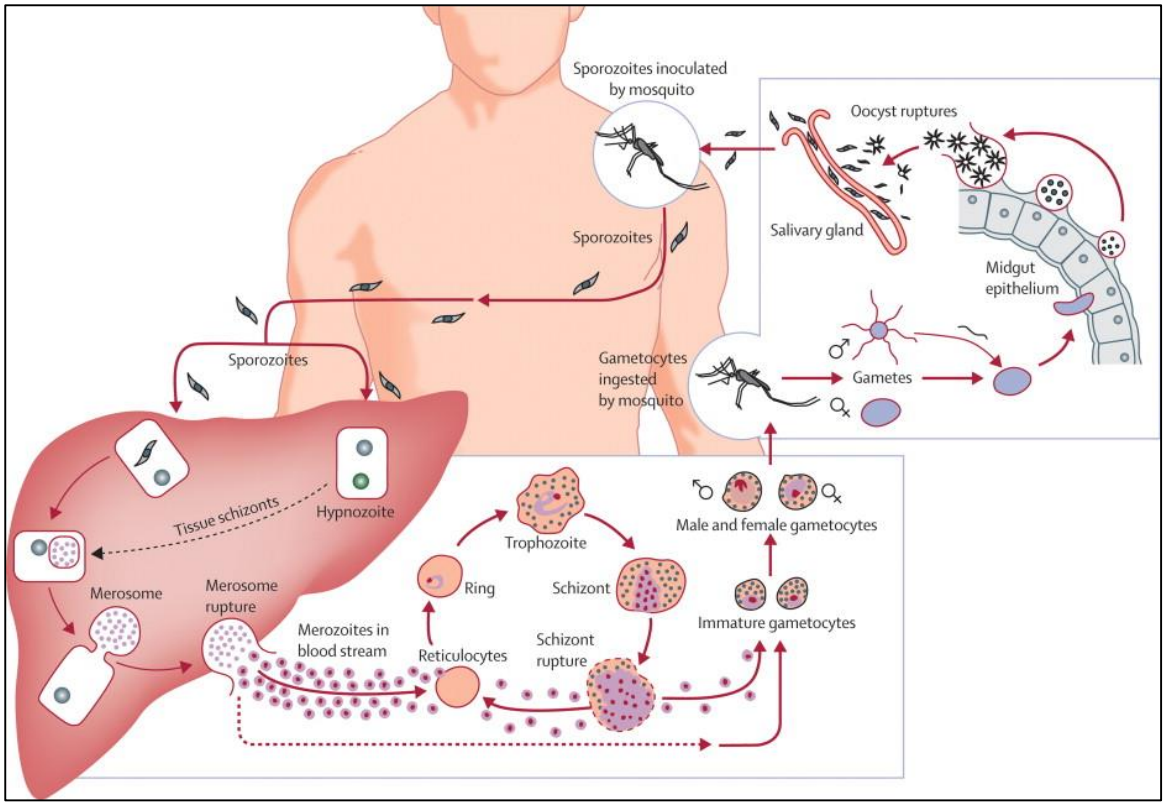

Figure 1.2: *P. vivax* life cycle

P. vivax sporozoites injected during the biting of a female *Anopheles* mosquito, enter the bloodstream and infect hepatocytes, a portion of which lie dormant as hypnozoites. Other parasites mature into schizonts that ultimately rupture and release infective merozoites into the bloodstream. Merozoites attach to and invade reticulocytes, in which they undergo a 48-hour asexual replication cycle to produce daughter merozoites, which subsequently invade new reticulocytes and cause the clinical symptoms of malaria. Some infected cells differentiate into gametocytes, which are ingested by mosquitoes and undergo fertilization in the mosquito midgut. Ookinetes traverse the midgut wall, develop into oocysts, and release sporozoites, which migrate to the salivary glands. Reprinted from (Mueller et al., 2009a), © 2009, with permission from Elsevier.

To complete the life cycle, a small proportion of parasites commit to sexual forms through the expression of transcription factor AP2-G (Sinha et al., 2014, Kafsack et al., 2014). All daughter merozoites from these committed cells will all go on to differentiate into either male (micro-) or female (macro-) gametocytes upon invasion of the next erythrocyte. After 3-6 days of sexual development, mature gametocytes then circulate in the bloodstream, to be ingested by mosquitoes taking a blood meal (Bousema and Drakeley, 2011). Once inside the mosquito midgut, several cues trigger gametocyte 'activation' including a drop in temperature to around 18°C and the presence of mosquito-made xanthurenic acid (Billker et al., 1998), resulting in male gametocyte

exflagellation. The motile male and immotile female gametes fuse to form diploid zygotes that mature into ookinetes, a process during which genetic recombination can occur. Ookinetes then traverse the midgut wall and develop into oocysts, which divide into thousands of sporozoites (Rosenberg and Rungsiwongse, 1991). Upon release, the haploid, motile sporozoites migrate through mosquito tissues to the salivary glands, ending a ~2-week process inside mosquitoes, after which the *Plasmodium* parasites can infect new human hosts.

The life cycle is similar for all mammalian *Plasmodium* parasites but with some crucial differences in human-infecting malaria parasites. The period of asexual replication in human blood for *P. vivax*, *P. falciparum*, and *P. ovale* follows a cycle of 48 hours, while this period is 24 hours and 72 hours for *P. knowlesi* and *P. malariae*, respectively. As described above, *P. vivax* and *P. ovale* can cause relapsing malaria through the activation of hypnozoites lying dormant in hepatocytes months to years after initial infection (Krotoski, 1985), which has significant implications for treatment. Both of these relapsing species, as well as initial *P. knowlesi* infections, also display a preference for invading reticulocytes (Mons et al., 1988, Kitchen, 1938, Lim et al., 2013, Collins and Jeffery, 2005). Since reticulocytes account for approximately 1% of the overall erythrocyte count, *P. vivax* achieves much lower parasitemias than *P. falciparum*, which invades erythrocytes of all ages. This host cellular tropism has made *P. vivax* extremely difficult to maintain in continuous *in vitro* culture, as there is no readily available source for reticulocyte-enriched blood; this is a major contributing factor to the relative paucity of our understanding of *P. vivax* basic biology. Low parasitemias also make *P. vivax* more difficult to detect by microscopy and RDTs, which hinders surveillance and control efforts.

P. vivax-infected reticulocytes also display a very different appearance than *P. falciparum*-infected erythrocytes, the former being larger, more deformable, and decorated on their surface with caveolae-vesicle complexes or Schüffner's dots (Aikawa et al., 1975, Barnwell, 1990, Suwanarusk et al., 2004). Deformable *P. vivax*-infected reticulocytes are not cleared by the spleen and thus all developmental stages circulate in the peripheral blood (Suwanarusk et al., 2004). These characteristics contrast with the smaller, rigid, and knobbed morphology of *P. falciparum* trophozoite-infected erythrocytes, which escape splenic clearance by sequestering in microvessels (Luse and Miller, 1971, Cranston et al., 1984). Sequestration is thought to be unique to *P.*

falciparum and along with higher parasitemias contributes to severe disease and mortality in *P. falciparum* infections.

P. vivax gametocytes are frequently present in the blood at the onset of symptoms, in contrast to *P. falciparum*, and consequently these parasite forms can be transmitted to mosquitoes before treatment begins (Douglas et al., 2010). This feature combined with low parasitemias and a potentially large reservoir of asymptomatic carriers further adds to the complexity in interrupting transmission in the many low-transmission *P. vivax* settings.

1.1.4 *P. vivax* pathology

All the symptoms and pathology of malaria are caused by blood-stage parasites, and classically include periodic attacks of fever, chills, and sweats called "paroxysms" (Evans and Wellems, 2002). While periodicity of attacks is common, most patients with uncomplicated *P. vivax* malaria will have a combination of chills, sweats, headaches, nausea, vomiting, body aches, and general malaise (cdc.gov) approximately 7-14 days after a bite from an infectious mosquito that can be mistaken for influenza or other infections in areas where malaria is uncommon. Additional symptoms may include enlarged spleen and/or liver, mild jaundice and increased respiratory rates (cdc.gov).

Severe cases of *P. vivax* malaria have also been reported, with growing agreement that their overall impact has been largely underestimated (Price et al., 2007). Nearly all of the severe complications caused by *P. falciparum* have also been reported for *P. vivax* infections, although coma appears to be rare for *P. vivax*. Severe disease from *P. vivax*-associated anemia, respiratory distress, renal failure, spleen enlargement, and low birth weight in infants have been reported in Pakistan, India, Southeast Asia, and Oceania [reviewed in (Price et al., 2009, White et al., 2014)]. Many aspects of *P. falciparum*-associated severe disease relates to either infected erythrocytes sequestering in organs or high parasitemias. As described above, sequestration of the parasite biomass in organs such as the brain and lungs can cause severe complications, such as coma and pulmonary edema. *P. falciparum* infects erythrocytes of all ages leading to the potential complication of severe anemia as the number of parasites exponentially increases in the blood. The WHO lists parasite counts >250,000/μl as a risk for severe disease in *P. falciparum* infections. In contrast, erythrocytes infected with *P. vivax* parasites do not sequester and parasitemias remain far lower at under 1%, yet still have the potential for severe disease.

A study in Brazil reported admissions to intensive care for *P. vivax* infections with parasite counts >500/μl (Lanca et al., 2012), indicating that a much lower parasite density can lead to severe disease in *P. vivax* compared to *P. falciparum*. The basis of this has not been fully determined, but relapsing or chronic infections with *P. vivax* may play a role (White et al., 2014). Several reviews have found that similar to *P. falciparum*, 10-20% of hospitalizations for *P. vivax* are classified as severe, and 5-15% of severe cases will lead to death (Price et al., 2007, Baird, 2013). Severe disease from *P. vivax* is often confined to populations in lower-resourced, tropical settings with high rates of hypnozoite-based relapse and comorbidities (nutritional, genetic, and immunological disorders), which may make them particularly vulnerable to severe disease (Price et al., 2009). This burden of chronic relapse and its potential contributions to severe pathology underline the need for effective treatments for the dormant liver stage, which is discussed further below.

1.1.5 Current *P. vivax* treatment and control measures

Several aspects of *P. vivax* biology make it challenging for treatment and control measures. As discussed previously, *P. vivax* infections generally reach very low parasitemias (under 1%), and a study in the Solomon Islands demonstrated that traditional measures of detection, such as LM and RDTs, significantly underestimate parasite prevalence in endemic populations (Waltmann et al., 2015). This same study also demonstrated that the vast majority of cases were asymptomatic. These factors make surveillance for *P. vivax* particularly difficult and contribute to the persistence of this pathogen, even in regions where significant progress is made in controlling *P. falciparum* infections.

Vector biting habits also pose challenges for control. The main vector control measures including long-lasting insecticide-treated nets (LLINs) and indoor residual spraying (IRS) aim to prevent biting of human hosts indoors at night and have proven highly effective in reducing the incidence of *P. falciparum* malaria (Goodman et al., 1999). The dominant malaria-transmitting anopheline species in *P. vivax*-endemic Southeast Asia and South America, such as *Anopheles dirus*, *Anopheles minimus*, and *Anopheles maculatus* tend to bite outdoors at dusk, rendering indoor strategies like IRS and LLINs less effective (Trung et al., 2005, Trung et al., 2004). In addition to this, LLIN usage varies significantly in populations and is additionally challenged by increasing mosquito resistance to pyrethroids (Alonso and Tanner, 2013).

Drug treatment challenges pose the final and perhaps greatest threat to *P. vivax* control. The foremost of these problems is the seemingly inevitable development of parasite resistance to frontline treatments. While chloroquine resistance was widespread in *P. falciparum* parasites by the 1980s, it remained the first-line treatment for *P. vivax* through 2010 in all but three malaria-endemic countries (WHO, 2010). The WHO now lists chloroquine as the recommended treatment only in "chloroquine-sensitive areas" (WHO, 2014), and there is ample evidence that chloroquine resistance in *P. vivax* parasites is spreading throughout Southeast Asia, though the mechanism remains unclear (Douglas et al., 2010). This delay in the acquisition of chloroquine resistance in *P. vivax* compared to *P. falciparum* is potentially caused by differences in parasite biology, as *P. vivax* gametocytes present at the time of symptoms while they are delayed by several days during *P. falciparum* infections. Thus, *P. vivax* parasites can be transmitted to mosquitoes before any drug pressure is experienced (Mendis et al., 2001).

Front-line treatment in most *P. vivax*-endemic areas is now ACT for both *P. vivax* and *P. falciparum*. These therapies include 3-day regimens using a short-acting artemisinin derivative as the potent effector of rapid parasite clearance, while a longer-acting partner drug such as mefloquine, lumefantrine, or piperaquine eliminates the remaining parasite burden (WHO, 2015). ACTs have contributed to the 31% reduction in malaria mortality over the last 10 years; however, their continued use is under threat with the development of "slow-clearing" *P. falciparum* parasites in western Cambodia (Dondorp et al., 2009) that were recently shown to be capable of infecting African vectors in laboratory-based experiments (St Laurent et al., 2016). Significant resistance to ACTs in Africa would have devastating consequences for malaria mortality. This adds additional urgency to the need for new treatments and vaccine candidates. ACTs are currently highly effective against *P. vivax* with the exception of artesunate + sulfadoxine-pyrimethamine (SP), due to significant *P. vivax* resistance to SP (WHO, 2015, Young and Burgess, 1959).

P. vivax (along with *P. ovale*) is capable of relapse of disease due to activation of dormant, liver hypnozoites. This stage will make *P. vivax* a much more difficult parasite to eradicate than *P. falciparum*, as the hypnozoites represent a pool of parasites that can re-infect the erythrocytes of individuals that have been previously drug cured. This is readily observed in patients treated by ACTs, where *P. vivax* relapses occur at different intervals depending on the half-lives of the long-acting partner drug used (WHO, 2014). Currently, primaquine is the only drug approved and available to eliminate hypnozoites.

This reliance on a single drug poses a serious risk to public health as parasites in Oceania have already shown resistance to primaquine treatment (Price et al., 2009). Primaquine is also contraindicated in pregnant women and individuals with G6PD deficiency, which affects up to 25% of people in *P. vivax*-endemic areas, as use of primaquine in such patients can cause severe hemolytic anemia (Nkhoma et al., 2009). This problem is complicated by the lack of inexpensive, point-of-care tests for G6PD deficiency, meaning that primaquine is simply not prescribed in many *P. vivax* cases.

There has been some progress at developing alternative drugs for eliminating liver stage parasites. The most developed drug, tafenoquine, is in the same class (8-aminoquinoline) as primaquine and has been shown to prevent relapse in 89% of patients up to 6 months after treatment (Llanos-Cuentas et al., 2014). It is administered as a single dose, a large improvement over the 14-day regimen for primaquine; however, its use is still contraindicated in individuals with G6PD deficiency. This remains a significant limitation in areas where *P. vivax* is most prevalent.

1.2 *P. falciparum* and *P. vivax* genomics

1.2.1 *P. falciparum* and *P. vivax* whole genome sequencing

Several innovations have been truly transformative in the study of *Plasmodium*. One of the first breakthroughs came through the successful adaptation of *P. falciparum* to *in vitro* culture (Trager and Jensen, 1976), providing worldwide access to a continuous supply of parasites for experiments not previously possible. The development impacted nearly all aspects of basic *Plasmodium* research (Trager and Jensen, 1997), and led to the stepwise characterization of a host of individual genes. In contrast, *P. vivax* has remained firmly recalcitrant to reliable *in vitro* culture, in part due to this parasite's preference for invading reticulocytes, and thus it is very difficult to study. Characterizing gene function in *P. vivax* has primarily relied on parasites from primate infections, clinical isolates from *P. vivax*-endemic areas, or both.

Another transformative innovation in *Plasmodium* research has come through the development of the genomics field, beginning with the completion of the entire genome sequence of the *P. falciparum* 3D7 strain in 2002 (Gardner et al., 2002) and the *P. vivax* Sal 1 strain in 2008 (Carlton et al., 2008). The sequencing of both reference genomes

involved multi-year, globally collaborative efforts and relied on the capillary sequencing "Sanger" method, including significant manual genome finishing.

The *P. falciparum* genome produced several important results, and provided a valuable reference for the subsequent *P. vivax* reference genome. The *P. falciparum* 3D7 reference genome led to the annotation (including significant manual annotation) of over 5300 gene models at the time of publication (now near 5400). Manual curation still occurs and is actively maintained by GeneDB (www.genedb.org) and PlasmoDB (www.plasmodb.org). Nuclear genes are organized on 14 chromosomes ranging in size from 0.6 Mb (chromosome 1) to 3.3 Mb (chromosome 14), with a total genome size of ~23 Mb. The majority (53%) of the genome contains protein-coding content. *P. falciparum* also harbors a 6-kb mitochondrial genome and a 35-kb apicoplast genome. The apicoplast, a plastid homologous to chloroplasts (McFadden et al., 1996, Waller and McFadden, 2005), appears to be essential for parasite survival in the asexual blood stage (He et al., 2001) through the biosynthesis of an isoprenoid precursor (Yeh and DeRisi, 2011). The majority of *P. falciparum* genes appeared to be unique to *Plasmodium*; 60% could not be assigned functions, as they were not sufficiently similar to other sequenced and annotated genes in other genomes.

The *P. falciparum* 3D7 reference genome contained three highly variable gene families, including *vars* (variable), *rifs* (repetitive interspersed family) and *stevors* (sub-telomeric variable open reading frame). The 60-member, primarily subtelomeric, *var* gene family is only found in *P. falciparum* and is responsible for encoding *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) (Baruch et al., 1995, Smith et al., 1995, Su et al., 1995). PfEMP1 proteins are trafficked to the surface of infected erythrocytes (Leech et al., 1984), where they mediate adherence to host endothelial receptors (Kyes et al., 2001), enabling parasites to sequester in microvessels and avoid clearance by the spleen. IgG responses directed at PfEMP1 are a primary component of the host protective antibody response in *P. falciparum* infections (Bull et al., 1998). Sequestration appears not to occur in any other human-infecting *Plasmodium* species. The two remaining gene families, *rifs* and *stevors*, are also primarily sub-telomeric with 149 and 28 members, respectively, and are members of the Pir superfamily (Janssen et al., 2004) found in all *Plasmodium* species sequenced to date. Little is known about the proteins produced by these families, rifins and stevors, but they also appear to be trafficked to the surface of infected erythrocytes and may also be involved in evading the host's immune system.

The *P. vivax* Sal 1 genome was published several years after the *P. falciparum* 3D7 genome (in 2008) from an isolate from a malaria patient in El Salvador in 1972, and subsequently propagated through *Saimiri boliviensis boliviensis* monkeys. *P. vivax* Sal 1 showed a similar nuclear genome organization to *P. falciparum* 3D7 with ~27 Mb, including over 5400 annotated genes spread over 14 chromosomes, albeit with a much lower AT-content (57.7% vs. 80.6% in *P. falciparum*). Synteny was well conserved and 77% of genes had orthologs with *P. falciparum, P. knowlesi,* and *P. yoelii*. No *var* gene equivalents were detected, but there were 346 Pir family representatives, referred to as *virs* in *P. vivax,* primarily in telomeric regions. The genome sequencing did have some limitations, however, with missing or incomplete assemblies in the highly repetitive telomeric regions and around ~4.3 Mb of contigs that could not be linked to specific chromosomes. Gene models were largely predicted through automated gene-model prediction software, and gene names and functions were often inferred from *P. falciparum* annotation. The annotation contained far less manual curation of gene models, making for a less complete data set with missing transcripts and erroneous gene boundaries possible. Nevertheless, it was a huge multi-year achievement and an enormous push forward for the study of *P. vivax* biology.

Genomic sequencing technology has developed rapidly over the last 25 years, beginning with gel- and then capillary-based "Sanger" sequencing technology developed by Fred Sanger (Sanger et al., 1977). This technology produced long (up to 900 bases/read) reads which dominated the reference genome sequencing projects of the late 1990s and early 2000s including most notably, the human genome in 2001 (Lander et al., 2001). Both the *P. falciparum* 3D7 and *P. vivax* Sal 1 reference genomes were sequenced using this expensive sequencing technology. Several sequencing technologies gained prominence as cheaper, higher-throughput options in the 2000s [Illumina/Solexa (Bentley et al., 2008), 454 (Droege and Hill, 2008), SOLiD (Shendure et al., 2005)], with additional new technologies being developed and more commonly used [Pacific Biosciences (Roberts et al., 2013)] or emerging on the horizon (Oxford Nanopore). Sequencing technology today is dominated by Illumina, which generates gigabases (Gb) of data per run (compared to kb/run for capillary sequencing) through medium length, relatively inexpensive reads (up to 150 bases/read). The material required to produce libraries has declined 100-fold from the 5 μg required for a capillary sequencing library to the 50 ng required for an Illumina HiSeq library today (leading to Gb more data). The reductions in sequencing cost and

material requirements and increases in data output have greatly facilitated the development of genome-wide approaches to study *Plasmodium* over the last several years including studies aimed at large-scale population genetics (Manske et al., 2012) and large-scale gene-knockout experiments (Schwach et al., 2015).

P. vivax research has also benefitted from this technological shift and several additional genomes have been fully sequenced since *P. vivax* Sal 1, including monkey-adapted laboratory isolates from India, North Korea, Mauritania, Brazil, and Peru (Neafsey et al., 2012) and field isolates from Madagascar, Cambodia, Peru, and Thailand (Menard et al., 2013, Chan et al., 2012, Auburn et al., 2013, Dharia et al., 2010, Bright et al., 2012). More parasite material was available for the monkey-adapted samples, which therefore generated deeper sequence coverage and the ability for *de novo* assembly and annotation (though assemblies were still more fragmented than the *P. vivax* Sal 1 reference genome). Sequencing from the field isolates was primarily suitable for SNP calling after mapping to the *P. vivax* Sal 1 reference genome. Comparisons between the genomes demonstrated that the genetic diversity in *P. vivax* is twice that of *P. falciparum*, indicative of the long global history of *P. vivax* colonization (Neafsey et al., 2012, Carlton et al., 2013). The data also suggested the "capacity for greater functional variation," which may add to the challenge of global *P. vivax* malaria elimination (Carlton et al., 2013).

1.2.2 *P. falciparum* and *P. vivax* transcriptome studies

The publication of the *P. falciparum* and *P. vivax* reference genomes laid the foundation for the study of transcription in both organisms. Studies of transcription involve probing for the repertoire of RNA molecules within cells, which may include protein-coding mRNA, non-coding RNA, and regulatory RNA. Transcription studies highlight the dynamic changes within cells, which drive the development of organisms and assist with connecting genes with functions. This is particularly important for *Plasmodium*, as over half the genes have unknown functions. Some early RNA studies connected to reference genome sequencing projects included the single-sided sequencing of cDNA derived from parasite RNA transcripts, called expressed sequence tags (ESTs). These data trained or validated automated gene-model prediction software. However, it represented only a short fragment of a transcript, and was therefore limited in detecting gene boundaries and actual abundance of particular RNA molecules within *Plasmodium* cells. Two major

technologies subsequently dominated genome-wide RNA studies in *Plasmodium*: microarrays and RNA-Seq.

1.2.2.1 Microarrays

Microarray technology enables the abundance of transcripts to be measured simultaneously based on a set of predefined probes often produced from known or predicted transcripts for a given reference genome. The principle behind microarrays involves measuring the light emitted from hybridization events between a set of chemically-labelled DNA probes with known sequences (reference genome sequences, for instance) to a set of unknown DNA or RNA molecules (e.g., RNA from different time points in a life cycle). The proof of principle for DNA microarray technology was established by Fodor et al. from the Affymax Research Institute in 1991 (Fodor et al., 1991). The first use for transcriptome study came in 1995 where binding events between 45 arrayed *Arabidopsis thaliana* cDNAs and fluorescently-labeled *A. thaliana* mRNA were measured. The florescence intensity measured was proportional to the hybridization between the cDNA and mRNA, and correlated with the initial mRNA concentration (Schena et al., 1995). This opened the door to assessing the transcription of multiple genes in parallel.

Transcriptional microarray technology was first applied to *Plasmodium* in the early 2000s with the publication of the transcriptional profile of the asexual blood stages for laboratory-adapted *P. falciparum* HB3 (Bozdech et al., 2003). The blood stages cause the symptomatic portion of malaria and were the most tractable for study, given the availability of large volumes of highly synchronous *in vitro P. falciparum* HB3 cultures. Bozdech and colleagues measured RNA every hour during blood-stage development using microarrays comprised of long oligonucleotides synthesized based on the *P. falciparum* 3D7 reference genome and spotted onto glass slides. The data showed transcriptional profiles for over 60% of genes and found that the majority of expressed genes (80%) appeared to have single peak expression during the 48-hour cycle (Figure 1.3). The transcription of genes appeared to be a highly-ordered cascade with genes of related function being highly synchronized; for example, genes relating to pyrimidine ribonucleotide synthesis had peak abundance at the same time (18-22 hours post erythrocyte invasion) as genes involved in purine salvage pathways. Similarly, deoxyribonucleotide synthesis reached peak abundance (~32 hours post erythrocyte

invasion) just as DNA replication to produce daughter merozoites is required. Genes relating to erythrocyte invasion, such as AMA1 and EBA175 (discussed further below) reached peak abundance in the schizont stage. This pattern of genes being transcribed maximally when their encoded proteins function was termed the "just in time" model of gene expression. *Plasmodium* appeared to possess this feature much more than other eukaryotes such as yeast (Spellman et al., 1998) or mammalian cells (Whitfield et al., 2002), which exhibit this pattern in only 15% of genes. The profile appeared most similar to developmental processes, such as early development in *Drosophila melanogaster* (Arbeitman et al., 2002). This cyclical transcriptional profile mirroring developmental processes led to the renaming of the asexual blood stages as the intraerythrocytic development cycle or IDC.

A second microarray study of 7 *P. falciparum* IDC time points, as well as sporozoites and gametocyte stages was performed and published in parallel (Le Roch et al., 2003). It confirmed expression in at least 1 stage for 4557 genes (88% of the genome) with up to 5-fold variation in abundance. The work showed single peak abundance for 43% of genes (varying at least 1.5 fold between stages). The ring and trophozoite stages appeared enriched for peak abundance of genes relating to metabolism and cellular processes, such as transcription, translation, and energy metabolism. Genes relating to parasite-host interactions, such as merozoite surface proteins (MSPs) or circumsporozoite protein (CSP), reached peak abundance during schizont and sporozoite stages, respectively. This suggested that parasite transcription is closely linked with the functional requirements of the immediate stage.

A microarray study in 2006 compared the transcriptional profiles of 3 laboratory strains of *P. falciparum*: 3D7, HB3, and DD2 (Llinas et al., 2006). The data showed similar transcriptional profiles for the vast majority of genes in each isolate, suggesting that transcriptional timing is "hard-wired" for nearly all genes. A small group of 69 genes (1.3% of the genome), however, displayed transcriptional shifts of 12 or more hours and was enriched for genes relating to parasite-host interactions and antigenic variation. This pattern was also conserved in newly-adapted field isolates (Mackinnon et al., 2009).

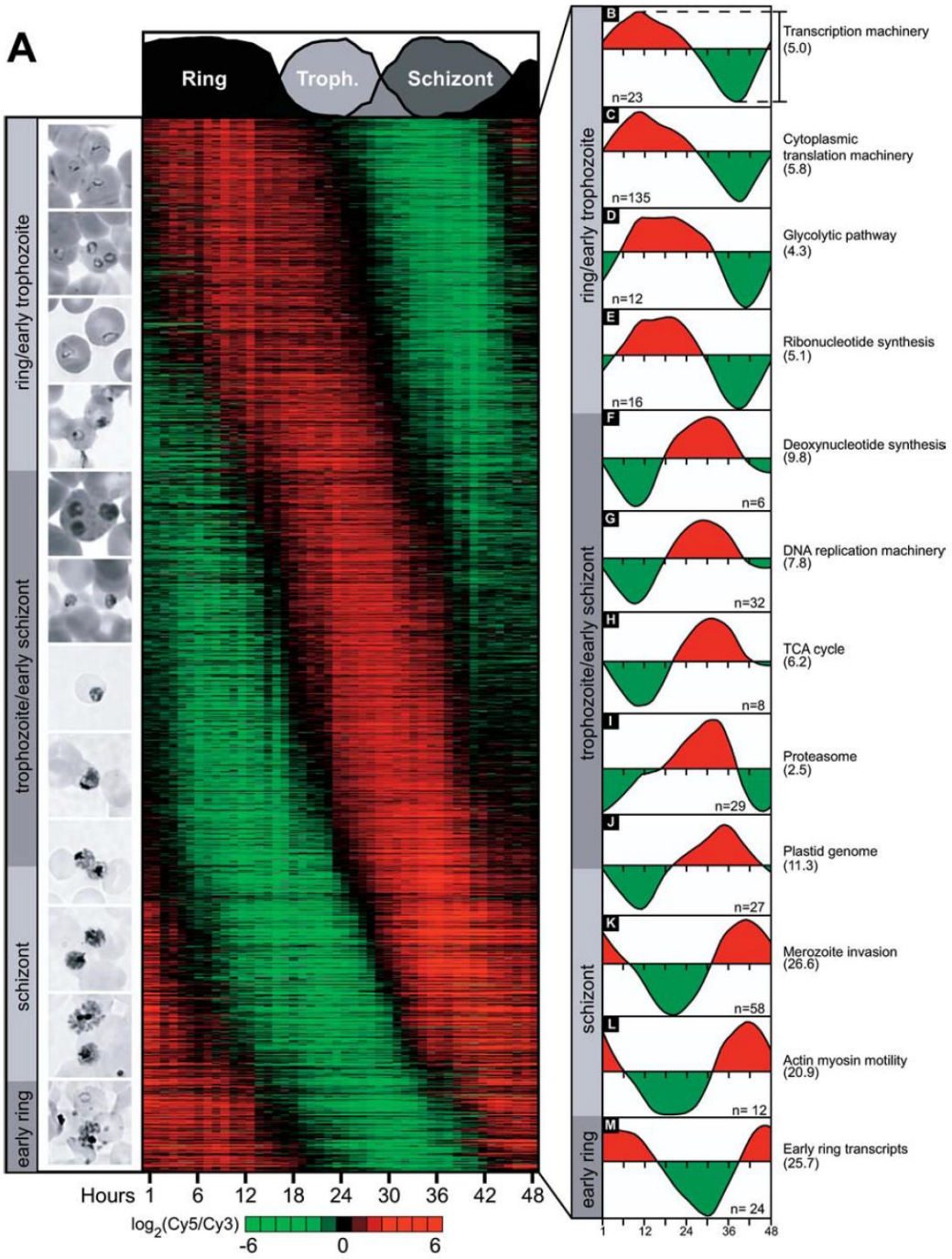

Figure 1.3: Intraerythrocytic development cycle (IDC) overview for *P. falciparum*

(A) Phaseogram for an ordered set of 2712 genes (rows) showing transcription every hour post invasion (columns), with morphological stages shown on the left. (B-M) Average expression profiles for genes with similar biochemical processes and functions at the various stages with the mean peak-to-trough amplitude shown in parentheses. Figure reprinted from (Bozdech et al., 2003) under the Creative Commons Attribution (CC BY) license.

Two *P. vivax* microarray studies have been published to date (Bozdech et al., 2008, Westenberger et al., 2010). The Bozdech *et al.* study focused on the *P. vivax* IDC using 3 highly synchronous patient isolates from Thailand that were cultured *ex vivo* and sampled at 9 time points over 48 hours. The isolates displayed a similar transcriptional cascade to *P. falciparum* with over 3500 genes (70% of the genome) showing peaks in abundance during 1 time point. Transcription profiles were evenly spaced throughout the IDC for syntenic orthologs between *P. falciparum* and *P. vivax*, while non-syntenic orthologs tended to have transcription shifted to the late IDC/schizont-to-ring transition. This implies that genes related to invasion and early establishment of the parasite in new host cells are largely responsible for the key differences between the species. The majority of orthologous, syntenic genes (68%) appeared to have identical expression profiles between *P. vivax* and *P. falciparum* orthologs, and a further 22% (including *msp*8) showed only slight shifts. The remaining 10% appeared to have dramatic shifts between the species (Pearson correlation coefficient between -0.2 and -0.1). This set is likely to contain genes underlying the differences between *P. vivax* and *P. falciparum*; for example, it includes the 2 *P. vivax* orthologs for the knob-associated histidine-rich protein (KAHRP), which have peak expression during the schizont stage in *P. vivax* and the early ring stage in *P. falciparum*. The single-copy KAHRP in *P. falciparum* has been associated with the formation of "knobs" on the surface of infected erythrocytes, which are essential for cytoadherence (Rug et al., 2006). This feature appears to be unique to *P. falciparum*, and KAHRP appears to serve a different purpose in *P. vivax*. In line with the above example, antigenic presentation in general appeared to differ between *P. falciparum* and *P. vivax* with *P. vivax* virs presenting in two waves: just after invasion and later in the schizont stages, when surface antigen transcription in *P. falciparum* is silent.

The Westenberger *et al.* study expanded the transcriptional profile of *P. vivax* to include additional life stages, including sporozoites, gametes, zygotes, and ookinetes, and human asexual blood stages, reporting differential expression in nearly 4326 genes (80% of the genome) (Westenberger et al., 2010). The study included transcriptional data for nearly 200 additional genes not included in the Bozdech *et al. P. vivax* microarray study, and established that the general processes for growth, metabolism, and host-parasite interactions are shared between *Plasmodium* species. It also provided a means for prioritizing genes with potential transmission-blocking or pre-erythrocytic stage vaccine development.

The majority of transcriptional studies in *Plasmodium* have focused on the IDC, although there have been some exceptions. Gametocyte stages of *P. falciparum* have been the subject of several studies (Le Roch et al., 2003, Silvestrini et al., 2005, Young et al., 2005). Rodent models have also been used as a model for gametocyte study, including a proteomics study published in 2005 (Khan et al., 2005). Parasite material from outside of the asexual blood-stage is easier to obtain from model rodent malarias, which have also been used to study sporozoite stages (Lasonder et al., 2008, Mikolajczak et al., 2008) and liver stages (Tarun et al., 2008).

In summary, microarrays have proven to be a powerful technology for describing the genome-wide transcriptional profiles of *Plasmodium*. Most studies have focused on the IDC of *P. falciparum* and *P. vivax* due to the limited availability of material from other life stages. These studies establish a highly-ordered periodic transcriptional profile for the majority of *Plasmodium* proteins, and similar transcriptional profiles in laboratory-adapted and clinical isolates. Microarray technology is inherently limited in several ways, however, as transcriptional profiles can only be measured for probes included in the arrays and probe content is contingent on high-quality reference genomes. In the case of *P. vivax*, this made for an incomplete gene list in the Bozdech *et al.* study, as the reference genome annotation was incomplete at the time of the study. Microarrays are also limited in sensitivity, such that genes with low transcription are unlikely to be detected. At the opposite end, highly abundant transcripts will saturate the available probes, thus limiting the ability to describe true abundance. Using pre-designed probes also limits the detection of alternatively-spliced transcripts and 5' and 3' UTR regions.

1.2.2.2 RNA-Seq

A number of the technical limitations of microarrays are addressed by the direct sequencing of cDNA libraries made from RNA, called RNA-Seq. As previously discussed, the technological improvements in sequencing technologies over the last decade have greatly reduced the quantities of starting material needed and the cost for generating whole-genome sequencing data. Transcriptome sequencing in particular benefits from "deep sequencing" and is most commonly generated using Illumina sequencing technology, which currently provides the lowest cost per base. Early RNA-Seq studies focused on model organisms, such as yeast (Wilhelm et al., 2008), mouse (Mortazavi et al., 2008), and human (Pan et al., 2008, Wang et al., 2008), and were

reviewed by (Wang et al., 2009). In addition to producing comparable transcript abundance data with microarrays, additional alternative splicing data were often detected; alternative splicing was estimated to occur in 95% of multi-exon human genes (Pan et al., 2008). Early comparisons to microarray data, for instance by Marioni *et al*., found RNA-Seq data to be highly reproducible and comparable to microarray data in identifying differential gene expression, with the added benefits of detecting transcripts with low abundance, alternative-splice variants, and novel transcripts (Marioni et al., 2008).

RNA-Seq was first applied to *Plasmodium* in 2010 through the high-throughput sequencing of cDNA made from RNA from a highly synchronous *P. falciparum* 3D7 *in vitro* culture (Otto et al., 2010). The study measured RNA abundance from 7 regularly-spaced time points throughout the 48-hour IDC. The study led to the improvement of over 10% of the ~5400 gene models and identified 121 novel transcripts. It uncovered 84 cases of alternative splicing and confirmed 75% of predicted splice sites. The RNA-Seq data showed good correlation with microarray experiments (Pearson correlation coefficients from 0.7 to 0.8). It also expanded the list of genes exhibiting some expression during the IDC to 4871 genes (~ 90% of the genome), suggesting that the majority of the genome is transcriptionally active during the blood stages. Overall, RNA sequencing proved to validate and expand on microarray results.

At the time of initiating this work, no RNA-Seq data have been published for *P. vivax*. This is primarily due to the inability to culture *P. vivax in vitro*. But recent reductions in the RNA requirements for Illumina libraries have made sequencing from field isolates more feasible. We aimed to address this need using clinical isolates from Cambodia as discussed in Chapter 3. RNA-Seq using *P. vivax* clinical isolates is a particularly pressing need as the initial *P. vivax* microarray data set lacked probes for at least 250 genes now included in the most up-to-date genome annotation. No alternative splicing events are yet known for *P. vivax.* Identifying new blood-stage genes abundant during the schizont stage may expand the list of genes to consider as potential merozoite development or invasion-related genes to prioritize for further functional characterization as vaccine candidates.

1.3 Erythrocyte invasion

1.3.1 Overview of invasion

The process by which infective merozoites invade new host erythrocytes has long been the focus of study and vaccine development for several reasons. Blood-stage parasites cause the symptoms of malaria and vaccines directed at merozoite invasion would therefore halt disease progression. This also represents one of the few points during which free parasite forms are fully exposed to the host immune system, and may therefore represent a "weak link" that both therapeutics and vaccines could target. A full understanding of blood-stage parasites is thus essential for rational prioritization of genes for further consideration as drug or vaccine targets.

Merozoites that egress from schizonts represent one of the smallest eukaryotic cells known, at 1.2 um in length or less than 1/5 the size of the erythrocytes they invade (Figure 1.4) (Garcia et al., 2008). This tiny form has a polar structure with a pointed apical end and a wider posterior end. The apical end of merozoites contains 3 sets of secretory organelles known as the rhoptries, micronemes, and dense granules, all known to be important for invasion of new erythrocytes. The wider posterior end of the merozoite houses the genetic and metabolic machinery including the nucleus, mitochondrion, and apicoplast (Bannister and Mitchell, 2003, Garcia et al., 2008). These organelles are contained within a double membrane structure called the inner membrane complex (IMC), which is connected to the outer plasma membrane via actin filaments (Farrow et al., 2011). An adhesive coat 15 nm thick covers the outer surface in clumps of narrow protruding bristles (Garcia et al., 2008). The structure of merozoites is supported by 3 cytoskeletal polar rings, 1 of which connects to microtubules linked to the IMC at the posterior end. These structures help to orient the invasive secretory organelles during merozoite assembly.

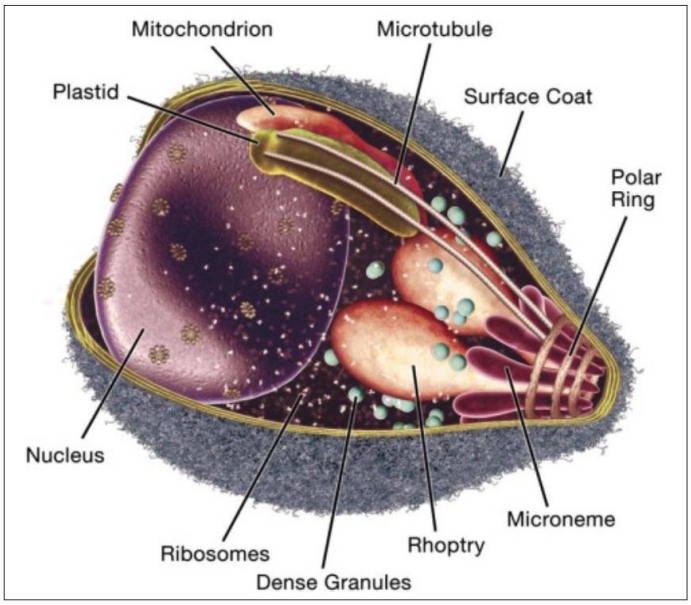

Figure 1.4: *Plasmodium* merozoite structure

Schematic of merozoite structure and organelles, including the genetic material of the posterior (left) end (nucleus, mitochondrion, plastid/apicoplast) and the invasive secretory organelles of the apical (right) end (rhoptries, micronemes, dense granules). Figure reprinted from (Cowman & Crabb, 2006), © 2006, with permission from Elsevier.

Plasmodium invasion of erythrocytes is a multistep process that begins with the initial reversible attachment of a merozoite to an erythrocyte (Figure 1.5). Insight into this process was first described using video microscopy with *P. knowlesi* in 1975 (Dvorak et al., 1975) and confirmed by similar experiments in *P. falciparum* 34 years later (Gilson and Crabb, 2009). This initial merozoite binding can occur at any part of the merozoite surface and is presumed to involve interactions between proteins in the parasite outer coating and the host erythrocyte at somewhat long range (20-30 nm), leading to a weak distortion of the erythrocyte membrane (Cowman and Crabb, 2006, Garcia et al., 2008, Weiss et al., 2016). The merozoite then reorients so that its apical end faces the erythrocyte and the gap between the parasite and erythrocyte narrows (Aikawa et al., 1978, Gilson and Crabb, 2009), leading to a more significant distortion of the erythrocyte membrane (Weiss et al., 2015). Receptor-ligand interactions thought to mediate this and downstream invasion events often involve proteins released from apical organelles, discussed in more detail in the following section. The release of proteins from the apical organelles appears to occur in multiple (potentially calcium or potassium-dependent) stages starting from egress until after the merozoite is engulfed in the erythrocyte (Zuccala et al., 2012, Weiss et al., 2016). An irreversible commitment to invasion occurs,

and a tight junction forms at the point of attachment (Aikawa et al., 1978, Bannister et al., 1975). The tight junction moves from the apical to the posterior end of the merozoite driven by an actin-myosin motor connected to the IMC. The merozoite is engulfed into the erythrocyte or reticulocyte as an internal parasitophorous vacuole (Aikawa et al., 1978, Baum et al., 2006). Surface proteins are shed throughout the invasion process through a calcium-sensitive serine protease, SUB2, discharged from the micronemes, which translocates across the parasite surface at the tight junction (Harris et al., 2005, Withers-Martinez et al., 2012). The adhesive proteins mediating the tight junction are also shed via serine proteases as the merozoite is engulfed and the erythrocyte membrane is resealed (O'Donnell et al., 2006, Garcia et al., 2008). The parasite then begins the process of growth towards either additional asexual replication or gametocyte differentiation.

Both genomic and transcriptional data support that *P. vivax* and *P. falciparum* frequently differ in the genes relating to immune evasion and host-interaction. This underlines the fact that while the general process of merozoite invasion of erythrocytes is conserved for mammalian *Plasmodium* parasites, the ligand-receptor parasite-host interactions mediating invasion tend to be highly parasite-specific.

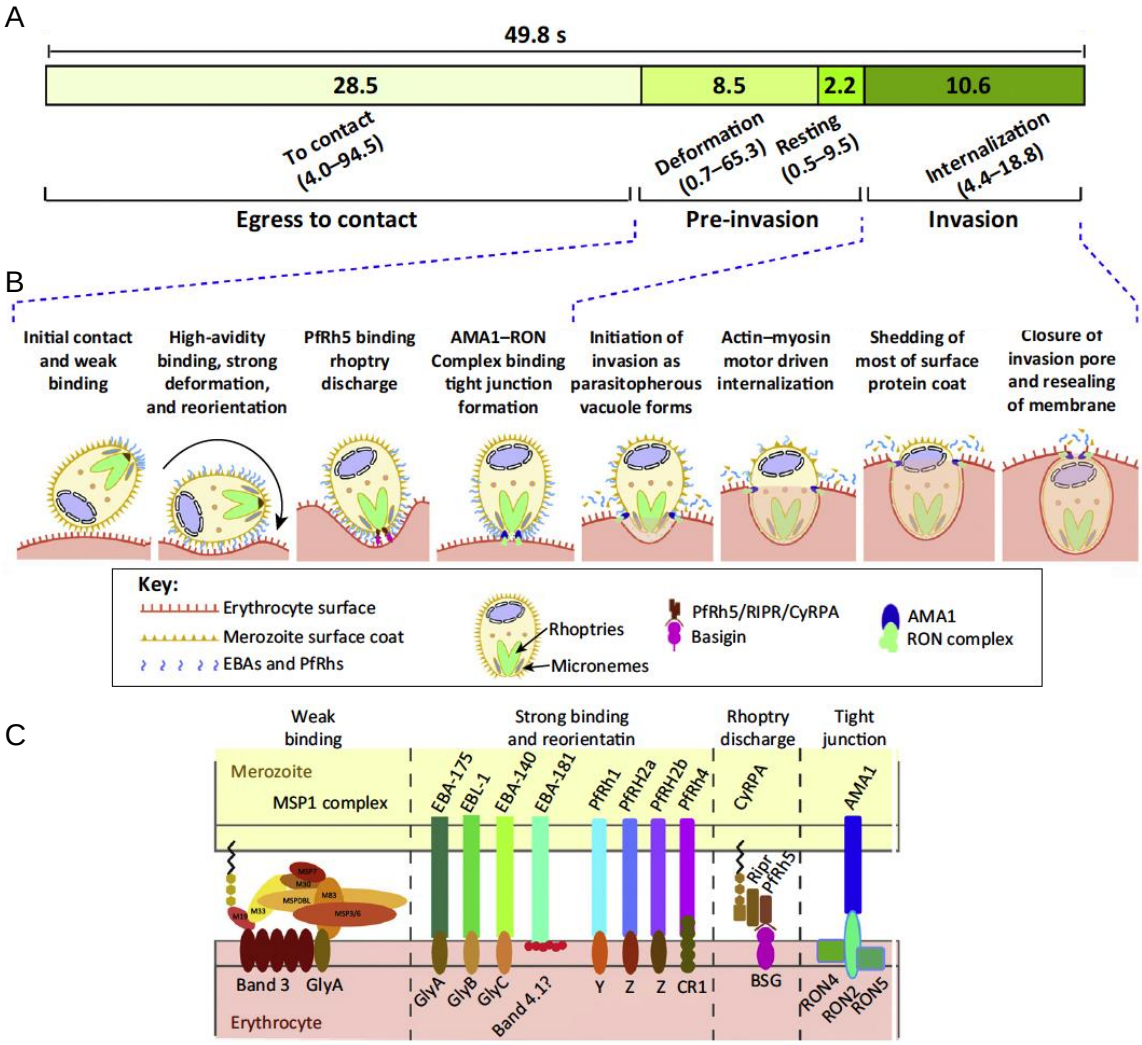

Figure 1.5: Overview of *P. falciparum* invasion of erythrocytes

P. falciparum merozoite egress through invasion is under 1 minute for 77% of *in vitro* invasion events. (A) Bar shows average times (seconds) for invasion stages. Timing for "egress to contact" is likely longer *in vivo* with flowing conditions and higher-density erythrocytes. "Pre-invasion" includes initial contact, significant deformation during merozoite reorientation, and resting, during which formation of tight junction likely occurs. "Internalization" occurs in a tightly-defined window (95% confidence interval of 9.99–11.11 s). (B) Cartoons with general description of events and several *P. falciparum* invasion ligands. (C) Known *P. falciparum* ligands with putative binding partners, including arbitrarily-named Y, Z, and Z for unknown receptors. Abbreviations for erythrocyte receptors: BSG, basigin; CR1, complement receptor 1; GlyA–C, glycophorins A–C; RON, rhoptry neck protein. Figure reprinted with slight modifications from (Weiss et al., 2016), © 2016, with permission from Elsevier.

1.3.2 *P. falciparum* erythrocyte invasion ligands

Much more is known about the specific interactions required for merozoite invasion in *P. falciparum* parasites (described in detail below) than for *P. vivax* and as a result, studies of invasion in *P. vivax* tend to focus on *P. falciparum* invasion homologs. The molecular details of the *P. falciparum* invasion process are therefore described below in detail, with the known features of *P. vivax* invasion described subsequently.

1.3.2.1 Initial reversible attachment between merozoites and erythrocytes

Plasmodium merozoite attachment to and invasion of erythrocytes is generally mediated by a series of receptor-ligand interactions between erythrocytes and merozoites [recently reviewed in (Weiss et al., 2016)]. Initial reversible binding occurs at the outer coat of the parasite, and over 30 proteins have been identified and localized to the merozoite surface (Cowman and Crabb, 2006, Garcia et al., 2008). Proteins with glycosylphosphatidylinositol (GPI)-anchors are enriched in this set and mostly cluster within detergent-resistant membrane (DRM) regions (Sanders et al., 2005, Cowman and Crabb, 2006). Binding is likely initiated by a complex of proteins including MSP1, the most abundant GPI-anchored surface protein, leading to weak deformation of the erythrocyte membrane in the area as receptors-ligand molecules coalesce in the region of contact (Weiss et al., 2015). MSP1 is processed from a 195-kDa precursor into 4 fragments that remain non-covalently linked prior to invasion. The membrane-bound fragment, $MSP1_{42}$, is further cleaved by SUB2 during invasion into $MSP1_{19}$, which remains associated with the surface throughout the invasion process (Blackman et al., 1991, Harris et al., 2005). Blocking the final MSP1-processing event has been shown to block merozoite invasion (Blackman et al., 1994). The MSP1 complex is comprised of the processed MSP1 fragments along with MSP7, MSP6, MSPDBL1, and MSPDBL2 (Stafford et al., 1996, Pachebat et al., 2001, Trucco et al., 2001, Lin et al., 2014)

Additional GPI-anchored proteins are also enriched at the merozoite surface in DRMs, including MSP2, MSP4, MSP5, P12, P92, P38, and P113. A number of these proteins (P12, P92, P38) are members of the six-cysteine (6-cys or Cys_6) protein family, members of which contain 2-7 domains (and occasionally partial domains) with highly-conserved cysteine residues (Templeton and Kaslow, 1999, Williamson, 2003). 6-cys proteins are structurally similar to SAG proteins in *Toxoplasma gondii* (Cowman and Crabb, 2006, Gerloff et al., 2005), with 3 disulfide bonds between the 6 cysteine residues supporting a

conserved structure between family members. 6-cys domain proteins act as ligands with host cell receptors during multiple life stages [i.e., P36 and P36p in sporozoites (Ishino et al., 2005, van Dijk et al., 2005); P230, P48/45 on the surface of sexual gamete stages (van Dijk et al., 2010)], making it highly likely that they serve a similar role in the asexual blood stages.

A number of additional proteins peripherally associate with GPI-anchored merozoite surface proteins during merozoite assembly (Sanders et al., 2005, Cowman and Crabb, 2006, Garcia et al., 2008). This includes the 6-cys protein P41, which lacks a GPI anchor and has been shown to interact with membrane-bound 6-cys protein P12 (Taechalertpaisarn et al., 2012). Members of the SERA family also associate with the merozoite membrane, with SERA5 being the most abundant (Sanders et al., 2005). The proteins contain a central protease domain with an active site cysteine or serine (Hodder et al., 2009). SERA5, while essential in the asexual blood stages, may lack a working protease domain and its function remains unclear (Hodder et al., 2009).

1.3.2.2 Merozoite EBL and RH proteins interact with specific erythrocyte receptors.

After initial attachment, the merozoite reorients, causing significant erythrocyte membrane deformation, so that its apical end interacts more closely with the erythrocyte membrane (Weiss et al., 2015). This requires interactions with proteins that span the plasma membrane and link directly or indirectly to the merozoite cytoskeleton (Weiss et al., 2016). A number of receptor-ligand interactions during this period have been well studied in *P. falciparum*, dominated by 2 protein families: the erythrocyte binding antigens (EBAs and EBLs) and the reticulocyte-binding protein homologues (RHs). Proteins from both groups are secreted from the apical organelles during the invasion process, facilitating it after the initial reversible contact (Harvey et al., 2012, Riglar et al., 2011). The signaling which precipitates the release of the proteins is discussed further below. It is hypothesized that a concentration gradient of EBL and RH proteins concentrated at the apical end of the merozoite may facilitate merozoite reorientation, though the function of the erythrocyte membrane deformation, which is independent of the actin-myosin motor, remains unclear (Farrow et al., 2011, Weiss et al., 2016).

The EBL family contains type-I transmembrane proteins with 2 cysteine-rich regions [Region II (RII) and Region VI (RVI)] in the extracellular region, and include EBA140 (PF3D7_1301600), EBA175 (PF3D7_0731500), EBA181 (PF3D7_0102500), EBL-1

(PF3D7_1371600) and a likely pseudogene with missense mutations, EBA165 (PF3D7_0424300) (Adams et al., 2001, Cowman and Crabb, 2006, Tham et al., 2012, Triglia et al., 2001). The RII contains two tandem Duffy binding-like (DBL) domains, which are homologous to the single DBL domain found in the *P. vivax* DBP (Sim et al., 1994). RH family proteins were first identified as the *P. yoelii* 235-kDa rhoptry protein (Py235) (Holder and Freeman, 1981) followed by the *P. vivax* reticulocyte-binding proteins (PvRBP-1 and 2) 11 years later (Galinski et al., 1992). The superfamily is common to *Plasmodium,* and in *P. falciparum* the group includes 5 members with transmembrane domains [RH1 (PF3D7_0402300), RH2a (PF3D7_1335400), RH2b (PF3D7_1335300), RH4 (PF3D7_0424200)] and a single smaller member lacking a transmembrane domain [RH5 (PF3D7_0424100)] (Rayner et al., 2000, Rayner et al., 2001, Triglia et al., 2001, Taylor et al., 2002, Rodriguez et al., 2008). A sixth member contains missense mutations and is likely encoded by a pseudogene [RH3, PF3D7_1252400)] (Taylor et al., 2001).

At least 5 receptor-ligand interactions occur between the *P. falciparum* EBLs or RHs and erythrocytes during invasion (Figure 1.5): Glycophorin A-EBA175, Glycophorin B-EBL-1, Glycophorin C-EBA140, Complement receptor 1-RH4, and Basigin-RH5 (Camus and Hadley, 1985, Sim, 1995, Mayer et al., 2009, Maier et al., 2002, Crosnier et al., 2011). The EBLs and RHs have been described as "alternative pathway" ligands because many of these interactions appear to be functionally redundant; all EBLs and RHs with the exception of RH5 can be disrupted individually without affecting parasite survival (Duraisingh et al., 2003, Baum et al., 2009, Lopaticki et al., 2011). Knockout experiments targeting several members suggest that a minimum number of interactions is essential for survival (Duraisingh et al., 2003). This built-in functional redundancy may aid the parasite in overcoming host receptor polymorphism and phenotypic variation in ligands arising from immune pressure (Harvey et al., 2012, Tham et al., 2012).

The order of and signaling required for microneme and rhoptry protein release is the subject of continued research and debate. Singh, *et al.* found that exposure to low potassium ion concentrations (as in found in blood plasma) led to increased calcium levels in merozoite cytosol and the release of micronemal proteins such as EBA175 and AMA1 to the merozoite surface. Binding of EBA175 to erythrocyte receptor Glycophorin A restored calcium concentrations to basal levels and led to the release of rhoptry proteins CLAG3.1 and RH2b (Singh et al., 2010). The low potassium trigger for invasion was

challenged by Pillai *et al*. who found that asexual growth *in vitro* was not affected by high potassium ion concentrations in the growth media (Pillai et al., 2013). Gao *et al.* showed that monoclonal antibodies against rhoptry protein RH1 blocked invasion by disrupting calcium signalling, and that monoclonal antibodies against EBA175 had no effect on calcium signalling. They further showed that disrupting the calcium signalling via RH1 monoclonal antibodies prevented the release of EBA175 to the merozoite surface (Gao et al., 2013). Together these results suggest that the release of micronemal and rhoptry proteins may not follow a simple two-step process, and additional work is needed to clarify these portions of the invasion process.

Contrary to other receptor-ligand interactions, the basigin (BSG) interaction with RH5 is essential in all isolates studied to date. Recent work supports that the RH5-BSG interaction may be downstream of other EBL and RH receptor-ligand interactions (Weiss et al., 2015). RH5 forms a complex with both RIPR and GPI-anchored CyRPA, and antibodies against any of these 3 proteins block erythrocyte invasion, though neither RIPR nor CyRPA appear to interact with erythrocytes directly (Chen et al., 2011, Reddy et al., 2015). A recent paper, however, contradicts that CyRPA has a GPI-anchor, thus leaving open the question of how the complex is attached to the merozoite surface (Volz et al., 2016). Live imaging of invasion events shows that antibodies blocking the interaction between the RH5 complex and BSG have normal pre-invasion, deformation, and reorientation, indicating the interaction may play a role in the release of rhoptry proteins that are needed to form the tight junction (Weiss et al., 2015).

1.3.2.3 Formation of the tight junction and the molecular motor of invasion

After rhoptry discharge the formation of a tight junction (also called the moving junction) is initiated. This process is thought to hinge on the interaction between the rhoptry neck (RON) proteins and AMA1. The current model suggests that the RON complex is translocated to the surface of erythrocytes (Riglar et al., 2011) where RON2 then binds to AMA1 on the parasite surface to initiate the formation of the tight junction (Bai et al., 2005, Pizarro et al., 2005, Vulliez-Le Normand et al., 2012) and recently reviewed in (Weiss et al., 2016). Micronemal AMA1 binds the an extracellular loop of RON2 in both *P. falciparum* and the related apicomplexan parasite, *Toxoplasma gondii* (Lamarque et al., 2014, Srinivasan et al., 2011, Tonkin et al., 2011, Tyler et al., 2011). This interaction has led to the proposal that the AMA1-RON2 binding mediates the tight junction, which

links either directly or indirectly to the actin-myosin motor connected to the IMC. Recent high resolution imaging work supports that AMA1 is localized to the tight junction, which is at or directly adjacent to the actomyosin force propelling invasion into the host cell (Riglar et al., 2016, Bichet et al., 2014). However, work in both *Plasmodium* and *T. gondii* demonstrates that parasites with knockouts or conditional knockouts of AMA1 were still invasive though with greatly reduced efficiency, suggesting that AMA1 may play a greater role in host cell binding rather than linkage to the tight junction and actin-myosin motor (Bargieri et al., 2013, Giovannini et al., 2011). Component genes of the motor complex itself have been successfully disrupted in *T. gondii* while still maintaining successful invasion (Egarter et al., 2014). This calls into question whether the motor is absolutely required for invasion in other apicomplexans (Bargieri et al., 2014) and supports further, careful study of the potentially versatile invasion process.

Several additional proteins are also linked to invasion. In sporozoites, this process is aided by the membrane-spanning TRAP protein, which contains a thrombospondin type I repeats (TSR) domain and connects extracellular adhesins to the motor. TRAP is linked to the parasite cytoskeleton by aldolase (Buscaglia et al., 2003). In merozoites, it is proposed that MTRAP performs a similar function, as it also contains the TSR domain, is essential to the asexual blood stages, and contains an aldolase-binding cytoplasmic tail (Baum et al., 2006, Morahan et al., 2009). The extracellular region of MTRAP binds to the erythrocyte receptor Semaphorin 7A, although the biological implications for this are not fully understood (Uchime et al., 2012, Bartholdson et al., 2012). Several other TSR domain-containing proteins (PTRAMP, SPATR, TLP) also have peak expression during the asexual blood stage, though their function also remains unclear (Baum et al., 2006, Heiss et al., 2008, Morahan et al., 2009).

1.3.3 *P. vivax* reticulocyte invasion

In stark contrast to *P. falciparum,* current knowledge of *P. vivax* invasion of reticulocytes is extremely limited. Given the lack of continuous *in vitro* culture, it is commonly inferred that much of the *P. vivax* invasion process mirrors that of *P. falciparum,* as many one-to-one orthologs of *P. falciparum* invasion proteins exist. In a few cases, such homologues have been directly shown to have a role in invasion in *P. vivax* or localize to the merozoite surface or rhoptries, including AMA1, MSP1 (Galinski et al., 1992, Galinski and Barnwell, 1996, Cheng and Saul, 1994, Barnwell and Galinski, 1991),

rhoptry neck protein 1 (RON1), and 6-cys proteins P12 and P38 (Li et al., 2012, Moreno-Perez et al., 2011, Mongui et al., 2008). In the majority of cases though, their function is simply inferred. It is also worth noting that many of the genes mediating receptor-ligand interactions in *P. falciparum* have no direct *P. vivax* homologs, and in most cases analogous receptor interactions have not yet been found in *P. vivax*. The fact that there are likely to be significant differences between the processes is most obvious in the fact that *P. vivax* displays a preference for invading reticulocytes. The commitment to invading reticulocytes is mediated at a minimum by *P. vivax* reticulocyte-binding proteins 1 and 2 (RBP-1 and RBP-2) (Galinski et al., 1992), although completion of the *P. vivax* genome made it clear that there are several other RBP homologues that may be involved (Carlton et al., 2008). The reticulocyte-specific receptor for these proteins remains unknown.

Only 1 receptor-ligand interaction has been extensively studied in *P. vivax*: that between *P. vivax* Duffy Binding Protein (PvDBP) and the human erythrocyte surface protein Duffy Antigen Receptor for Chemokines (DARC, also referred to as the Duffy antigen or Fy). DBP was first discovered to be the ligand for DARC-dependent invasion through the *P. knowlesi* homolog, PkDBP (Haynes et al., 1988), and subsequently confirmed for *P. vivax* by Wertheimer and Barnwell showing that pre-incubation with purified DARC glycoprotein blocked PvDBP protein binding to erythrocytes (Wertheimer and Barnwell, 1989). Using PkDBP as a probe, PvDBP was cloned in Louis Miller's laboratory in 1991 (Fang et al., 1991), and the binding to DARC was found to depend on the cysteine-rich region II (DBP-RII or DBP-II) (Chitnis and Miller, 1994). DBP interaction with DARC occurs after the initial merozoite attachment and reorientation, and may precipitate the formation of a tight junction between the merozoite and reticulocyte membranes which then moves along the surface of the merozoite (Galinski and Barnwell, 1996). DBP interaction with DARC is a dynamic process where two DBP-II domains initially form a heterotrimer with a single DARC receptor, then recruit a second DARC receptor, forming a heterotetramer (Batchelor et al., 2011, Paing and Tolia, 2014, Malpede and Tolia, 2014, Batchelor et al., 2014).

Duffy negativity, which is highly prevalent in sub-Saharan Africa, was thought to provide complete protection against *P. vivax* infection (Miller et al., 1976). However, the universality of this protection is becoming increasingly disputed, most notably by recent work showing that *P. vivax* infects Duffy-negative individuals in Madagascar (Menard et

al., 2010). Whether this is due to DBP-dependent invasion using a non-Duffy pathway, or represents a truly alternative DBP-independent invasion pathway is not currently known. Genome sequencing of additional *P. vivax* strains has led to the discovery of a DBP paralog (PvEBP) that was missing in the *P. vivax* Sal 1 isolate, which has the hallmarks of an erythrocyte binding protein, including a DBL and cysteine-rich C-terminal domain (Hester et al., 2013). However, the function of the gene remains unknown. The complete sequencing of 1 Malagasy *P. vivax* strain led to the discovery that a duplication of *DBP* was present in over half of Malagasy *P. vivax* infections while the duplication appears to be largely absent elsewhere (Menard et al., 2013). Madagascar has the highest rate of Duffy-negative infections reported thus far, which account for nearly 9% of *P. vivax* infections detected by PCR (Menard et al., 2010), leading to the hypothesis that the *P. vivax* infection of Duffy-negative in individuals is connected to *DBP* duplication. But it is currently unknown whether *DBP* duplication is enriched in *P. vivax* infections in Duffy-negative individuals.

1.4 Natural immunity to *Plasmodium* infections

Increasing our knowledge about the biology of *P. vivax* invasion has the potential for identifying new vaccine candidates. This is potentially augmented by understanding the targets of naturally-acquired immunity (NAI) that develops during infections. Humans experiencing repeated infections with *Plasmodium* frequently develop immunity, if not from infection, then from symptoms and the potential complications of severe disease. Understanding the development of antibody-based immunity and the specific antigens targeted, which effectively control infections, may inform and further prioritize candidates for vaccine development.

As reviewed by Doolan et al., one of the first studies of induced immunity against *Plasmodium* came with experimental “malariotherapy” as a treatment for neurosyphilis in 1917 by von Wagner-Jauregg. Patients diagnosed with the disease were intentionally infected with *Plasmodium* parasites of different species leading to the cure of 1 in 3 patients through the fevers associated with the resulting malaria episode. NAI in the adult subjects was observed to occur sometimes after 1 exposure, though often requiring multiple infections. Protection to *P. falciparum* appeared to be acquired more slowly than to *P. vivax*. Immunity also appeared to be species-specific, and protective immunity to at

least *P. vivax* did not persist for long periods after treatment [reviewed in (Doolan et al., 2009)]. Based on the work from Brown and Brown in 1965 on antigenic variation in *P. knowlesi*, a theory of immunity development emerged (Brown and Brown, 1965). Repeated exposures against a wide repertoire of antigenic variants were required to develop protective memory against a highly antigenically variable parasite population.

This cumulative-exposure model (strain-specific immunity acquisition) has been questioned by Doolan et al., however. They reviewed data that age is a more significant factor for acquiring immunity, which is inversely protective depending on whether exposure is acute or chronic. In acute scenarios where non-immune adults and children are exposed to *P. falciparum* for the first time, adults are more susceptible to severe disease than children. In contrast, in settings with chronic exposure, adults will develop immunity and protection from severe disease faster than children. Age-dependent (which may differ from cumulative exposure-dependent) acquisition of immunity appears to be very important in achieving "antiparasite immunity" in *P. falciparum* settings. The fact that adults can develop robust protective and strain-transcending immunity to clinical disease after relatively few episodes (3-4) has potential positive implications for vaccine design, as antigen variation may not represent a barrier to cross-protection against a diverse set of parasite clones (Doolan et al., 2009).

1.4.1 Immunity development to *P. falciparum*

As with most studies in malaria, the research into the development of NAI is much more advanced for *P. falciparum* than for *P. vivax*. Protective immunity to *P. falciparum* develops slowly with repeated exposure and involves the development of a protective IgG response [reviewed in (Langhorne et al., 2008, Doolan et al., 2009)]. In general, protection is defined as the lower risk of clinical disease through the absence of fever (axillary temperature >37.5°C) even if parasites are present in the blood. In holoendemic areas across sub-Saharan Africa, populations are often continuously exposed and first develop "antidisease immunity" relatively quickly in young childhood, with significant protection from severe disease complications and death even with higher parasite densities. Subsequently, a more slowly acquired "antiparasite immunity" develops, in which parasite densities are significantly reduced with additional protection from disease. Clearing parasites while limiting host pathology involves a series of pro- and anti-inflammatory or regulatory cytokine responses. An early pro-inflammatory response with

increased tumor necrosis factor (TNF)-α and interferon (IFN)-γ leads to fast clearance of parasites (Kremsner et al., 1995, Walther et al., 2006, D'Ombrain et al., 2008) while an increase in regulatory cytokines such as interleukin (IL)-10 and transforming growth factor (TGF)-β is then important for limiting severe disease (Day et al., 1999, Othoro et al., 1999, Perkins et al., 2000).

NAI is not sterilizing, however, and asymptomatic adults with low parasite densities are the norm in high-transmission areas. The most high-risk periods for complications from malaria are from about age 3 months to 5 years as children experience exposure to repeated malaria episodes. Women experiencing first and second pregnancies are also particularly vulnerable due to immunosuppression and the potential exposure to a new *var* gene variant (VAR2CSA) which mediates cytoadherence to chondroitin sulphate A in the placenta (Beeson et al., 2002, Fried et al., 2006).

While several aspects of the human immune response are under active study, the development of IgG antibodies remains the only immune response definitively shown to protect in human studies (Cohen et al., 1961). Protective IgG responses are relevant for vaccine design and there are numerous studies evaluating whether IgG to specific *P. falciparum* proteins correlate with protection. The focus of many studies has been the cytoadherence-mediating PfEMP1, and several studies suggest that the majority of a protective IgG response is directed against this family of proteins and leads to variant-specific protective immunity (Horrocks et al., 2004, Duffy et al., 2001, Kraemer and Smith, 2006). But the development of PfEMP1 as a vaccine candidate is likely limited due to the significant variation and frequent recombination present in the *var* gene repertoire (Claessens et al., 2014). Merozoites also appear to be common targets of NAI, as passive transfer of immune serum following schizogony significantly reduced parasitemia (with apparently little effect on developing or mature trophozoites) (McGregor, 1964). More recently, the search for correlations with protection has been expanded into the evaluation of IgG directed against a panel of *P. falciparum* merozoite proteins in a Kenyan cohort study (Osier et al., 2014). The results concluded that the breadth of antibody response rather than the response to any single antigen was a significant predictor for protection from clinical disease. These findings suggest that a multivalent vaccine with multiple antigen targets may be a winning vaccine strategy rather than a strategy that relies on a single, partially protective antigen.

1.4.2 Immunity development to *P. vivax*

Research into the development of NAI to *P. vivax* is far less advanced, and was recently reviewed in detail (Longley et al., 2016). In contrast to *P. falciparum* rates in Africa, the majority of *P. vivax*-endemic areas experience lower transmission rates. Consequently, populations in low-transmission areas are not repeatedly exposed, and people of all ages experience clinical disease. The exception to this is Papua New Guinea (PNG) where transmission of both *P. falciparum* and *P. vivax* can reach rates as high as hyperendemic regions in Africa. In higher-transmission co-endemic settings such as these, the incidence of *P. vivax* infection peaks at an earlier age than *P. falciparum* (Michon et al., 2007, Maitland et al., 1996, Phimpraphi et al., 2008, Mueller et al., 2009b, Lin et al., 2010). This corroborates reports from experimental infections showing that immunity to *P. vivax* develops more quickly than to *P. falciparum* (Jeffery, 1966, Ciuca, 1934). *P. vivax* infections are frequently polyclonal, and recent work in PNG by the Mueller laboratory suggests this speed of immunity acquisition may be partly due to the greater force of exposure to genetically distinct parasites that occurs in *P. vivax* infections, referred to as the molecular force of blood-stage exposure ($_{Mol}$FOB) (Koepfli et al., 2013).

1.4.2.1 Naturally-acquired cellular immunity to *P. vivax*

Naturally-acquired cellular immunity to *P. vivax* infection has been the subject of limited study and frequently with contradictory findings. It has been known for decades that T cells are essential for eliminating *Plasmodium* parasites during liver stages (Hoffman et al., 1989, Weiss et al., 1990, Renia et al., 1993) but whether any cellular responses are directed at dormant *P. vivax* hypnozoites is unknown. The induction of cytokines to *P. vivax* infection has been evaluated most commonly through measuring cytokines in plasma [reviewed in (Longley et al., 2016)], which eliminates the possibility of determining the cellular source of the cytokines. As with *P. falciparum* infection, *P. vivax* infection induces a pro-inflammatory TNF response (Karunaweera et al., 1992), which has been shown to limit infection for *P. falciparum* (Kremsner et al., 1995) but with the dual potential for inducing immunopathology or severe disease (Kwiatkowski et al., 1993, Grau et al., 1989). Multiple subsequent studies have shown the induction of strong pro-inflammatory responses to *P. vivax* infections, such as TNF, IFγ, IL-12, IL-16, IL-1β and IL-8 [reviewed in (Longley et al., 2016)]. It is largely accepted that *P. vivax* pro-inflammatory cytokine responses are higher at lower parasitemias than during *P. falciparum* infections, thus producing fevers with far fewer parasites (Price et al., 2007).

But the available data is in conflict with several studies finding comparatively higher levels of pro-inflammatory cytokines per parasite during *P. vivax* infections compared to *P. falciparum* infections (Hemmer et al., 2006, Karunaweera et al., 1992), while several other studies find no difference during infection with either species either by measuring cytokines per parasite or comparing cytokine levels for similar parasitemias (Fernandes et al., 2008, Goncalves et al., 2012, Rodrigues-da-Silva et al., 2014). These differing findings may be the result of a variety of factors including subject age, transmission level, location, stage of infection and methods used for measuring cytokines, suggesting the need for standardized assays to resolve the conflicting data (Longley et al., 2016).

The induction of anti-inflammatory or regulatory cytokine IL-10 during *P. vivax* infection has been reported in numerous studies [reviewed in (Longley et al., 2016)], with most studies reporting that levels of IL-10 were higher during *P. vivax* infection than *P. falciparum* infection (Praba-Egge et al., 2003, Fernandes et al., 2008, Goncalves et al., 2012, Yeo et al., 2010). IL-10 has been shown to have a protective effect against experimental cerebral malaria in mice with *P. berghei* infections (Kossodo et al., 1997). The connection between IL-10 levels and severe disease in *P. vivax* is in conflict with one study reporting lower IL-10 levels in severe compared to asymptomatic *P. vivax* cases (Andrade et al., 2010) and several other studies reporting the opposite finding (Goncalves et al., 2012, Mendonca et al., 2013). Additional research is needed to clarify the the role of immunoregulatory signals in severe, symptomatic and asymptomatic *P. vivax* infections.

1.4.2.2 Naturally-acquired humoral immunity

Antibody responses to individual *P. vivax* antigens have been evaluated in numerous immunoreactivity studies, but studies evaluating protective associations are few. A recent review of *P. vivax* immunoepidemiological studies found only 3 antigens – MSP1, MSP3.10 (MSP3α), and MSP9 – were consistently associated with protection (Cutts et al., 2014). DBP has also been a central target in clinical protection studies given its near universal requirement for invasion of reticulocytes, with some exceptions, which have been discussed elsewhere in this introduction, for example (Menard et al., 2010). DBP elicits antibody responses in naturally-infected individuals as first reported by Fraser et al. in 1997, and it was later determined that antibodies directed against the RII domain can block binding to erythrocytes (Fraser et al., 1997, Michon et al., 2000). Antibody levels to

DBP-RII have also been shown to increase with age (Michon et al., 1998, Xainli et al., 2003), which supports the potential development of protective immunity to DBP-RII in natural infections. The general sparseness of data evaluating correlations with protection underline the great need for expanding the list of proteins studied.

1.5 Approaches to a *Plasmodium* vaccine development

Control measures combatting *Plasmodium* have proven effective at reducing malaria deaths over the last decade. But these gains are threatened by the development of artemisinin resistance in *P. falciparum* in Southeast Asia and insecticide resistance in *Anopheles* vectors worldwide. A vaccine targeting *Plasmodium* parasite development is a universally recognized essential component of any global malaria eradication campaign. *P. vivax* vaccine development lags significantly behind *P. falciparum* due to both financial resources and the lack of critical information surrounding *P. vivax* biology.

1.5.1 Vaccine targets at different stages of the life cycle

Vaccine development efforts have focused on several parasite life stages with different goals for blocking parasite development. The majority of efforts have historically been directed at the asexual blood stages for several reasons. This section of the life cycle causes the symptoms of malaria, and blocking merozoite development and invasion would halt the disease progression. It is also the stage most tractable to study, as *P. falciparum in vitro* cultures provide a ready source of parasites. Merozoites are exposed to the immune system briefly after schizont egress and erythrocyte invasion, and IgG responses directed against merozoite proteins have shown correlations with protection. Early experiments suggested that the asexual-blood stages are prime targets of NAI, as serum transfers from immune donor patients led to rapid declines in parasitemia for recipient patients (Boyd, 1939). A vaccine mimicking this natural response may prove to be a useful tool (Greenwood and Targett, 2011).

Targeting the asexual blood stage would not, however, prevent initial sporozoite infection of hepatocytes, and additional vaccine development efforts specifically focus on the pre-erythrocytic-stages. Sporozoites are exposed to the human immune system from bite through to hepatocyte invasion, a process estimated to take at least 15 minutes, and may represent an attractive target for preventing initial infections from taking hold. Evidence

as far back as 1946 from a neurosyphilis patient inoculated with infectious bites and with salivary-gland emulsion demonstrated that pre-erythrocytic forms could be the target of induced immunity (Covell and Nicol, 1951). The two most advanced vaccine candidates for *P. falciparum* target this stage and are discussed in more detail below.

A third strategy in vaccine design is termed "transmission-blocking" and specifically targets the sexual stages. The primary goal in this approach is to prevent parasite fertilization in the mosquito by either targeting surface antigens on gametocyte-infected cells (thus targeting them for degradation) or antigens directly on the surface of sexual forms (gametocytes, gametes, ookinetes). In the latter scenario, mosquitoes would ingest not only gametocytes but also human antibodies blocking critical processes for maturing in mosquitoes. While such a vaccine would have no effect on the progression of illness in the current human host, it would effectively limit the spread of the disease to future hosts. Such an approach may be important as gametocytes in *P. falciparum* can often persist for several weeks and would not be targeted by a blood-stage vaccine administered after symptoms appear. Gametocytes appear almost simultaneously with *P. vivax* asexual forms, potentially indicating their development in liver stages and therefore the possibility of transmission even with protection from a blood-stage vaccine.

Overall, each strategy has some benefits and potential limitations, and it is likely that a combination of candidate antigens targeting multiple stages would provide the greatest chance of success in halting disease and transmission.

1.5.2 *P. falciparum* vaccine development

The majority of *Plasmodium* vaccine efforts have focused on *P. falciparum*, and were historically aimed at the asexual blood stages. The vast majority of initial work focused on the merozoite surface proteins MSP1 and AMA1, but phase II trials have failed to demonstrate significant protection (Geels et al., 2011, Schwartz et al., 2012). These proteins are highly abundant on the merozoite surface and experience significant immune pressure, resulting in a high degree of polymorphism. Antibodies directed against them have not proven to provide cross-protection with other haplotypes, and their usefulness as vaccine candidate may be limited (Schwartz et al., 2012, Hill, 2011, Geels et al., 2011). There are currently 7 merozoite proteins approved for clinical testing: MSP1, MSP2, MSP3, AMA1, EBA175, GLURP, and SERA5 (Schwartz et al., 2012). More recently, antibodies generated against several recombinant merozoite proteins, including *P.*

falciparum RH5, CyRPA, and RIPR have been shown to block invasion, and are all potential vaccine candidates under development (Chen et al., 2011, Reddy et al., 2015). As transmission-blocking vaccine candidates, *P. falciparum* Pfs25, a GPI-anchored surface protein of mosquito-stage zygotes and ookinetes, is the most active candidate (Table 1.1). Rabbits immunized with recombinant forms of Pfs25 and *P. vivax* Pvs25 were shown to block transmission to mosquitoes in membrane-feeding assays (Miura et al., 2007).

There are two pre-erythrocytic *P. falciparum* vaccine candidates in advanced testing, RTS,S and irradiated sporozoites, each of which has drawbacks. RTS,S targets the circumsporozoite protein (CSP) present on the sporozoite surface and has demonstrated only moderate efficacy, with a 36% reduction in malaria episodes but no effect on mortality (RTS, 2015). This represents a useful first step in reducing the number of malaria episodes in highly-endemic regions in Africa, but is likely to have little overall effect in reducing transmission. Improvements in efficacy will be absolutely crucial to making significant changes to the global malaria burden. The second most advanced vaccine involves the intravenous injection of irradiated sporozoites. It has thus far proven to be much more efficacious, with 6/6 participants receiving 5 doses protected by subsequent challenge with *P. falciparum* (Seder et al., 2013). However, requirement for manual dissection of mosquitoes, liquid-nitrogen storage, and multiple intravenous inoculations render the worldwide rollout of this vaccine unlikely (Garcia et al., 2013). While these are incredibly useful first steps, it is likely that additional candidates will be needed to produce a cost-effective, widely distributable, and efficacious vaccine.

1.5.3 *P. vivax* blood-stage vaccine candidate: Duffy binding protein

Very few *P. vivax* candidates are in preclinical or clinical vaccine trials, including only DBP (Phase 1a) and CSP (preclinical) (Table 1.1). The significant role that DBP plays in *P. vivax* invasion of reticulocytes makes it the prime blood-stage vaccine candidate. Previous studies have shown that antibodies against the recombinant DARC-binding DBP region II (DBPII) inhibit merozoite binding and invasion of erythrocytes (Fraser et al., 1997, Michon et al., 2000, Singh et al., 2001, Grimberg et al., 2007). However, the DBPII under consideration is highly polymorphic and under immune pressure (VanBuskirk et al., 2004a, Cole-Tobian et al., 2002), and antibody responses to this domain have been shown to be strain-specific (Ceravolo et al., 2009). Epitopes critical to binding between

DBP and DARC are highly conserved (Batchelor et al., 2011, Singh et al., 2006), suggesting that DBPII remains a promising vaccine target. This is further supported by a longitudinal cohort treatment re-infection study in Papua New Guinea showing that naturally-acquired high titer antibodies directed at DBPII blocked DBPII-DARC binding and were associated with protection from *P. vivax* infection (King et al., 2008). While the diversity of the DBPII domain makes the development of a strain-transcending vaccine a complex task, strategies to overcome this challenge are being actively pursued, most notably through the use of a synthetic DBPII-based vaccine candidate, termed DEKnull, which lacks an immunodominant variant epitope (Ntumngia et al., 2014, Chen et al., 2015).

1.5.4 Other blood-stage candidates

A limited number of additional proteins have been investigated for their potential as vaccine candidates [reviewed in (Longley et al., 2016)]. These include the *P. vivax* circumsporozoite protein PvCSP (currently in preclinical testing), PvAMA1, PvRBP1, PvRBP2, and $PvMSP_{1\text{-}19}$ (Kocken et al., 1999, Galinski et al., 1992, Galinski et al., 2000, Galinski and Barnwell, 1996, Rogers et al., 1999). In the case of a recombinant fragment of RBP1, antibodies directed against the fragment were high but were not protective against infection in an *Aotus* monkey *in vivo* model (Caraballo et al., 2007). More recently, antibody responses to MSP3, MSP9, several TRAgs, Virs, and gametocyte antigen 1 (GAM1) were evaluated [reviewed in (Longley et al., 2016)]. A meta-analysis of the available studies by Cutts *et al*. found correlations with protection consistently with MSP3.10 (MSP3α), MSP9, and MSP1 (Cutts et al., 2014). However, limited progress beyond initial studies has been made for any candidates thus far.

Table 1.1: Preclinical and clinical *Plasmodium* vaccine candidates*

Preclinical	Clinical				
	Translational projects			Vaccine candidates	
	Phase 1a	Phase 2a	Phase 1b	Phase 2b	Phase 3
CSP	ChAd63/MVA ME-TRAP + Matrix M	RTS,S-AS01 fractional dose	R21/Matrix-M1	ChAd63/MVA ME-TRAP	RTS,S-AS01E
P. vivax CSP	PfCelTOS FMP012	PfSPZ-CVac (PfSPZ Challenge + chlor. or + chlor./pyrim.	AMA1-DiCo	*Pf*SPZ	
VAR2CSA	R21/AS01B	PfSPZ-CVac (PfSPZ Challenge + chloro.)	P27A		
EBA175 /Rh5	PfPEBS	FMP2.1/AS01B	SE36		
RH5.1	ChAd63 RH5 +/- MVA RH5	M3V-D/Ad- PfCA (CSP/AMA1 or CSP/AMA1/SSP2/TRAP)	PRIMVAC (VAR2CSA)		
Pfs 48/45	ChAd63/MVA PvDBP			*P. falciparum*	*P. vivax*
	*Pfs*25-VLP			Pre-erythrocytic	Pre-erythrocytic
Pfs25	ChAd63 Pfs25-IMX313/MVA Pfs25-IMX313			Blood-stage	Blood-stage
				Transmission-blocking	
Combo: PE: R21, ME-TRAP E:RH5 TB: Pfs230 and Pfs25	Pfs230D1M-EPA/Alhydrogel and/or Pfs25-EPA/Alhydrogel			Combination	

*Compiled from WHO Rainbow Table of malaria vaccine projects, updated March 2016 (http://www.who.int/immunization/research/development/Rainbow_tables/en/)

1.5.5 Future *P. vivax* vaccine development

DBP remains the most promising *P. vivax* vaccine candidate. However, strong reliance on this single target may prove problematic for successful vaccine design for several reasons. First, as discussed, DBP is highly polymorphic and strain-transcending protection may be difficult to achieve. Second, corollaries in *P. falciparum* vaccine studies targeting single blood-stage antigens, such as AMA1 and MSP1, have failed to generate protective immune responses (Sagara et al., 2009, Ogutu et al., 2009). It is possible that a cumulative immune response to a multivalent vaccine will be needed to produce protective immunity.

Identifying new protein candidates for vaccine trials will be important for ultimate success. The whole genome sequencing and transcriptome studies of both *P. falciparum* and *P. vivax* supplied a long list of unexplored blood-stage candidates. But functions remain unclear for the vast majority of proteins. Considering a panel of candidates in high-throughput functional and immunoepidemiological screens could help to address this by prioritizing candidates for further investment.

1.6 Characterization of *Plasmodium* vaccine antigens

1.6.1 Producing *Plasmodium* vaccine antigens

The current state of vaccine development for *P. vivax* underlines the need for additional proteins to evaluate as vaccine candidates. Given the intractability of *P. vivax* to *in vitro* culture, it is largely impossible to obtain quantities of protein necessary to study and functionally test any desired candidates. New candidate antigens therefore must be expressed and studied as recombinant proteins.

Several systems have been used previously for producing recombinant *Plasmodium* proteins, which are reviewed in (Birkholtz et al., 2008). Most recombinant protein expression aimed at vaccine development has focused on *P. falciparum*, for which protein expression is particularly difficult due to the high AT-content, biased codon-usage, and the presence of repetitive amino acid sequences (Tsuboi et al., 2008). Membrane-bound proteins are often the focus of vaccine research, given their potential roles in invasion. Such proteins are difficult to express in properly-folded and soluble form, and thus most research has focused on producing soluble extracellular ectodomains. The most widely

used and inexpensive system remains *E. coli*, despite frequent issues with solubility that necessitate refolding procedures. The process for obtaining soluble, properly-folded protein fragments in *E. coli* often requires protein-specific protocols, and is therefore limited for high-throughput applications (Birkholtz et al., 2008).

Yeast expression systems (*Saccharomyces cerevisiae*, *Pichia pastoris*) are also frequently used to express vaccine candidates including *P. falciparum* EBA175, AMA1, MSP1, MSP3, and Pfs25 [reviewed in (Birkholtz et al., 2008)]. These systems have very high yield (10-75 mg/l) and overcome many of the solubility issues of the *E. coli* systems by linking expression constructs to an N-terminal yeast pheromone that directs protein secretion into the culture supernatant. As a eukaryotic system producing eukaryotic proteins, folding and post-translational modifications are likely to be much improved over the *E. coli* system. However, this is a drawback in the case of N- and O- linked glycosylation, which do not significantly occur in *Plasmodium*. Disulfide bonding has been shown to be heterogeneous, which may impact functional studies (Stowers et al., 2001). The *S. cerevisiae* system also requires the use of codon-optimization (at least in the case of *P. falciparum* proteins), as certain A-T containing codons are recognized as stop codons, a particular problem in the A-T-rich *P. falciparum* genome.

The baculovirus-infected insect cell system has also been applied for *Plasmodium* vaccine candidate expression, including fragments of *P. falciparum* MSP1, CSP, PfEMP1, EBA175, and SERA [reviewed in (Birkholtz et al., 2008)]. It may have advantages in proper protein folding over yeast expression systems as all of the listed proteins elicited antibody responses, while a yeast-expressed *P. falciparum* MSP1 contained conformational changes that rendered it immunologically inactive (Chang et al., 1992). Several other *P. falciparum* proteins (EBA175, SERA) also proved to be functionally active or were processed into fragments mirroring *in vivo* fragment sizes (Li et al., 2002, Daugherty et al., 1997).

Cell-free expression systems have been on the rise more recently and have several additional advantages over previous expression systems. The most common system involved using ribosomes, translation factors, and post-translational components taken from wheat embryos (Farrokhi et al., 2009). The system is well established for producing *Plasmodium* proteins including the *P. falciparum* vaccine candidates Pfs25, CSP, and AMA-1 (Tsuboi et al., 2008, Crompton et al., 2010, Trieu et al., 2011). The main

advantage of the system is in producing properly-folded proteins in plate format, thus facilitating the use of downstream functional or immunoepidemiological screens. For example, the wheat germ cell-free system was recently used to express a panel of 89 (Chen et al., 2010) or 143 (Lu et al., 2014) *P. vivax* recombinant protein ectodomain fragments; subsequent screening using malaria-exposed patient plasma from Korea established that at least a subset (19% and 28%) of these proteins was highly immunoreactive. The primary drawbacks of the system are that reagents and running costs are high and obtaining extracts is difficult, which limits the usage of this technology in smaller individual laboratories (Farrokhi et al., 2009).

Mammalian expression systems are well characterized and support proper folding of eukaryotic proteins, but were not used significantly to produce recombinant *Plasmodium* proteins due to the low yields achieved in adherent cultures (Birkholtz et al., 2008). This drawback has been largely overcome, however, with the development of liquid cultures leading to protein production on the milligram to gram range (Tom et al., 2008). The most common cell line used for large scale recombinant protein production is the human embryonic kidney 293 (HEK293E) line, which stably expresses the Epstein-Barr virus nuclear antigen 1 (Tom et al., 2008). Prior work in the Wright and Rayner laboratories established the HEK293E expression system for the successful expression of high-quality *P. falciparum* proteins that were functional and immunogenic (Crosnier et al., 2011, Taechalertpaisarn et al., 2012, Osier et al., 2014).

1.6.2 High-throughput screening of vaccine antigens

Fully characterizing the function of potential vaccine antigens and their immunological underpinnings is an important next step for expanding *P. vivax* vaccine development. Most vaccine research aims at targeting candidates important for host cell interactions, as blocking such interactions halts the invasion of parasites into host cells (i.e., sporozoites into hepatocytes or merozoites into erythrocytes). Interactions between parasites and host cells are often transient and low-affinity, and therefore difficult to detect (Wright, 2009, Bei and Duraisingh, 2012). Interaction and immunoepidemiological screens have traditionally utilized individual assays with individually purified proteins, often with limited ability to scale-up for protein libraries (Wright, 2009, Bei and Duraisingh, 2012). This type of approach often involves a large investment of resources and time focused on

a specific candidate, based on some prior knowledge or predicted homology-based function, which is often the case for *P. vivax*.

This individual approach is frequently used in mammalian expression system, cell-based binding assays to identify *Plasmodium* merozoite proteins and the sub-domains therein that mediate erythrocyte binding. This is reviewed in Birkholtz et al., and the list includes *P. falciparum* EBA175, EBA181, MAEBL, MSP1, AARP, and *P. vivax* DBP [e.g., (Sim et al., 1994, Tolia et al., 2005, VanBuskirk et al., 2004b, Han et al., 2004, Mayer et al., 2009, Mayer et al., 2004, Wickramarachchi et al., 2008)]. These have primarily utilized COS-7, a green monkey kidney cell-line, but have also used CHO-K1 and HeLa cell lines. This approach was used for determining the region of binding of *P. vivax* DBP to erythrocytes. DBP fragments were expressed as fragments embedded in the membrane of COS cells and incubated with erythrocytes. Binding events were visible as clusters of erythrocytes (rosettes) around COS cells. The method was highly effective for determining the required region II domain necessary for binding (VanBuskirk et al., 2004b). These approaches require significant resources and labor devoted to single candidates, however, which is a significant limitation in studying a larger protein library.

An alternative approach to this individual protein strategy is to test a panel of potential proteins in high-throughput interaction screens and immunoepidemiological studies. Evaluating proteins in parallel would prove a less biased approach, as proteins produced on the same platform could be compared in a more systematic way. Indeed, this was the goal for the panels of *P. vivax* protein fragments produced in the wheat germ cell-free system (described in the previous section) most recently used to evaluate general reactivity in Korean isolates (Lu et al., 2014, Chen et al., 2010). Large-scale yeast two-hybrid interaction screens were tested in the mid-2000s to identify potential *P. falciparum* interactions (LaCount et al., 2005), but these studies often used protein fragments with proteins potentially lacking proper conformations, which was a significant limitation.

The protein library approach was developed recently in the Wright and Rayner laboratories to screen a panel of *P. falciparum* proteins against a library of erythrocyte receptors utilizing the HEK293E expression system. The screen uncovered the *P. falciparum* Rh5 interaction with the erythrocyte receptor basigin, which was an essential interaction for invasion in all parasite isolates tested (Crosnier et al., 2011). The primary tool used in the work involved an avidity-based extracellular interaction screen

(AVEXIS) assay which reliably detects highly-transient interactions with half-lives ≤ 0.1 seconds (Figure 1.6)(Bushell et al., 2008); we adapted this assay to screen a *P. vivax* recombinant protein library (discussed in Chapter 4).

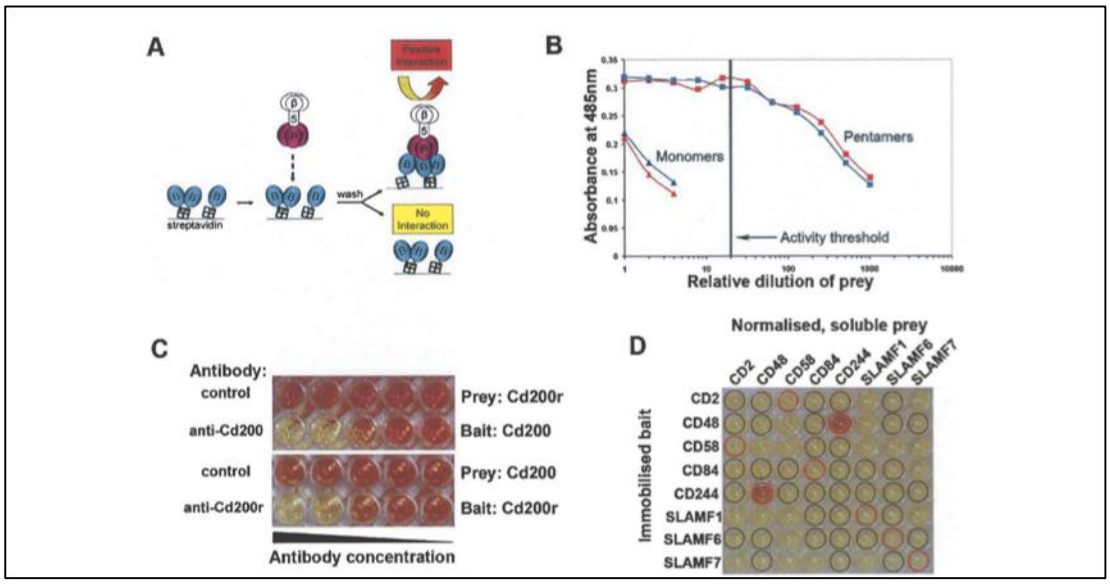

Figure 1.6: AVidity-based EXtracellular Interaction Screen (AVEXIS)
(A) Depiction of AVEXIS showing biotinylated bait proteins bound to streptavidin-coated plates. Baits are probed by pentamerized prey proteins tagged with ß-lactamase. Interactions are detected by nitrocefin turnover resulting in a color change from yellow to red. (B) Detection is improved using pentamerized preys over monomeric preys. (C) Interactions can be blocked in a dose-dependent manner using inhibitory antibodies. (D) The assay has a low false-positive rate with expected interactions highlighted in red and known non-interactors highlighted in black. Figure reprinted from (Bushell et al., 2008) under the Creative Commons Attribution (CC BY) license.

1.7 Specific Aims

The fact that only a relatively limited repertoire of potential *P. vivax* antigens are currently being considered for vaccine development is due in large part to a lack of understanding of the molecular details of *P. vivax* biology in general, and invasion in particular, and the fact that there have been few large studies of immunological responses generated during *P. vivax* infection, or assessment of whether particular responses correlate with protection. A comprehensive and systematic approach is needed to advance our understanding of *P. vivax* invasion and to identify additional vaccine candidates. My project aimed to enhance our basic understanding of the proteins important for *P. vivax* invasion of erythrocytes, and the development of immunity against those proteins.

My first aim was to study the transcription of *P. vivax* parasites just prior to invasion, which is discussed in Chapter 3. This required the testing and development of RNA extractions for *P. vivax* clinical samples from Cambodia and the bespoke production of strand-specific Illumina RNA-Seq libraries. These data would be used to detect genes with increased transcription in the late blood stages, and hence identify novel vaccine targets. Additionally, it would enable the confirmation or correction of existing gene model predictions in the *P. vivax* reference genome, the prediction of 5' and 3' untranslated regions, and potentially uncover novel gene transcripts and alternative splicing events.

My second aim was to produce a library of blood-stage *P. vivax* proteins to study in functional and immunoepidemiological studies, which is described in Chapter 4. The prioritization of candidates would be enhanced by RNA-Seq abundance data, and would focus on proteins with known or predicted localization to the merozoite surface or invasive secretory organelles, and on homology to *P. falciparum* vaccine candidates. The mammalian expression system, HEK293E cells, has previously been used successfully to express *P. falciparum* proteins and would provide the basis for the development of the library. High-throughput interaction screens developed in the Wright and Rayner laboratories would provide the foundation for investigating protein function.

Next I wanted to utilize the *P. vivax* merozoite protein library to conduct immunoepidemiological studies discussed in Chapter 5. Understanding general protein reactivity and correlation with either exposure or protection from clinical disease can help to prioritize candidates for further functional studies as vaccine candidates. Prior studies have often focused on one or several full-length recombinant proteins, and utilizing a panel of proteins will enable more systematic conclusions to be drawn. Ultimately patient collections from 3 *P. vivax*-endemic countries (Cambodia, Solomon Islands, and Papua New Guinea) were screened to further our understanding of naturally-acquired immunity to *P. vivax*.

This project aims to add significant new knowledge to our understanding of *P. vivax* invasion and the development of natural immune responses to infection. The *P. vivax* recombinant protein library will not only serve as a resource for the scientific community, but also aims to expand the list of potential blood-stage vaccine targets.