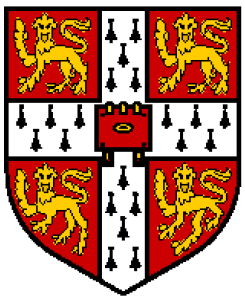
Phenotypic variation in erythrocyte invasion by *Plasmodium falciparum* isolates from Peru



A dissertation submitted for the degree of
Master of Philosophy

Richard Hesketh

Wellcome Trust Sanger Institute

Christ's College

University of Cambridge

Contents

Declaration

This thesis is a summary of research conducted within the Wellcome Trust Sanger Institute between September 2009 and August 2010. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. None of the research described, in its entirety or in part, has been submitted for any other qualification at any other University or similar institution.

Acknowledgements

I would like to thank my supervisor, Julian Rayner, for all of his support this year, both academically through supporting me to carry out the work described in this thesis and for accommodating the fulfilment of my sporting ambitions. I would also like to thank Michel Theron whose regular assistance and advice was invaluable throughout the year. In addition, all the members of T115, especially Leyla Bustamante and Matthew Jones, have provided priceless technical advice and companionship, especially with their help in fluorescence microscopy and antibody experiments. Other people who lent technical assistance included William Cheng and Bee Ling Ng (flow cytometry), David Goulding (electron microscopy) and Susana Campino and the other members of T112 who were responsible for sequencing the Peruvian samples. Lia Chappell will also play a vital role in the future of the project.

Dominic Kwiatkowski was instrumental in setting up my year in Cambridge and I would like to thank him and Christina Hedberg-Delouka, Annabel Smith (WTSI) and Marcus Coffey (Cardiff University) who worked very hard on my behalf to make this year possible.

My thesis committee, comprising Dominic Kwiatkowski, Oliver Billker, Teresa Tiffert and Virgilio Lew made excellent contributions towards keeping my project on track. I would like to thank them, particularly Teresa and Virgilio, for their time and enthusiasm.

I would also like to express my gratitude to Christ's College and the Wellcome Trust Sanger Institute for supporting my studies in Cambridge.

Outside the workplace, I would like to thank my family for everything they have done for me this year. As always, my mother and father have provided fantastic moral and financial support for which I will be eternally grateful. Becky, thank you, I could always count on your belief and encouragement even when you were halfway round the world. Last but not least, thanks to Rob for keeping morale high!

Abbreviations

A – adenine
bp – base pairs
BSA – bovine serum albumin
C – cytosine
cDNA – complementary DNA
CIDR – cysteine rich interdomain region
CSP – circumsporozoite protein
DBL - Duffy binding like
DDAO-SE – 7-hydroxy-9*H*-(1,3-dichloro-9,9-dimethylacridin-2-one) succinimidyl ester
DDT – dichlorodiphenyltrichloroethane
DMSO –dimethyl sulfoxide
DNA – deoxyribonucleic acid
DNase - deoxyribonuclease
EBA – erythrocyte binding antigen
EBL – erythrocyte binding ligand
EM – electron microscopy
EMP-1 –erythrocyte membrane protein-1
FITC – fluorescein isothiocyanate
FSC – forward scatter
G – guanine
GA – glutaraldehyde
GAG – glycosaminoglycan
GYP – glycophorin
Ht – haematocrit
MSP – merozoite surface protein
NTS – N-terminal segment
PBS – phosphate buffered saline
PCR – polymerase chain reaction
Pf –Plasmodium falciparum
PFA - paraformaldehyde
PMR – parasite multiplication rate
pRBC – parasitised erythrocytes
RBC – erythrocytes
Rh –reticulocyte binding protein homolog
RNA – ribonucleic acid
RNase – ribonuclease
SNP – single nucleotide polymorphism
sRBC – DDAO-SE stained erythrocytes
SSC – side scatter
T – thymine
TRAP – thrombospondin-related adhesive protein
TSP – thrombospondin
UNAP – Universidad Nacional Amazonia de Peruana
VSA – variant surface antigen

Abstract

Plasmodium falciparum invasion of erythrocytes marks the onset of the intra-erythrocytic stage of the parasite life cycle that is responsible for the generation of the symptoms and pathology associated with malaria in humans. The invasion process is complex and incompletely understood. However, specific ligand interactions between the parasite and host erythrocyte must take place to initiate the invasion process. Preventing these interactions inhibits erythrocyte invasion and therefore they have been closely studied as a potential vaccine target.

Plasmodium falciparum erythrocyte invasion is a variable phenotype and parasite isolates can use a number of alternative receptor-ligand interactions to catalyse invasion, resulting in a number of different invasion pathways. The redundancy in pathways allows *P. falciparum* parasites to invade erythrocytes of any age and reduces the possibility of host variation and immune responses having a detrimental effect upon parasite development. Invasion pathways are routinely studied by enzyme treatment of erythrocytes to remove specific receptors and observing the ability of the parasite to invade. Previously assays have been carried out using microscopy to detect invasion events. A two-colour flow cytometry-based assay has been developed at the Sanger Institute to increase the potential throughput of the assay, while also increasing the reproducibility and sensitivity.

Sequence polymorphism and differential expression of parasite ligands can both result in binding to different erythrocyte receptors, but clear genotype-phenotype associations for invasion pathways have not been established. Previous studies have focussed on selected single nucleotide polymorphisms in only a few specific genes. With the development of next generation sequencing technologies, whole genome and transcriptome sequencing of parasite isolates is now possible, allowing the use of genome-wide studies to identify the genotypes that are associated with specific pathways.

This study investigates invasion pathway variation using *P. falciparum* field isolates from Peru. It combines phenotyping, using a newly developed flow cytometry-based assay, with whole genome sequencing and transcriptome analysis to produce an in-depth study of the underlying mechanisms of variation.

Variation in the invasion pathways utilised between Peruvian field isolates was comparable to that in lab strains and field isolates from other countries. Whole genome sequencing and the powerful genotyping analysis tools available were used to identify non-synonymous SNPs in 62 invasion-related genes. Genotyping also revealed the presence of contamination in a number of Peruvian isolates. The project demonstrated that high throughput phenotyping and genotyping can be combined to produce detailed correlations with invasion, and highlights rate limiting steps for such future studies.

Chapter One

Introduction

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

History of Malaria

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

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

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

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

Epidemiology

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

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

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

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

The Control of Malaria Today

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

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

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

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

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

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

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

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

Pathology of Malaria

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

Pathogenesis of Severe Malaria

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

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

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

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

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

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

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

The Life Cycle of *Plasmodium*

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

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

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

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

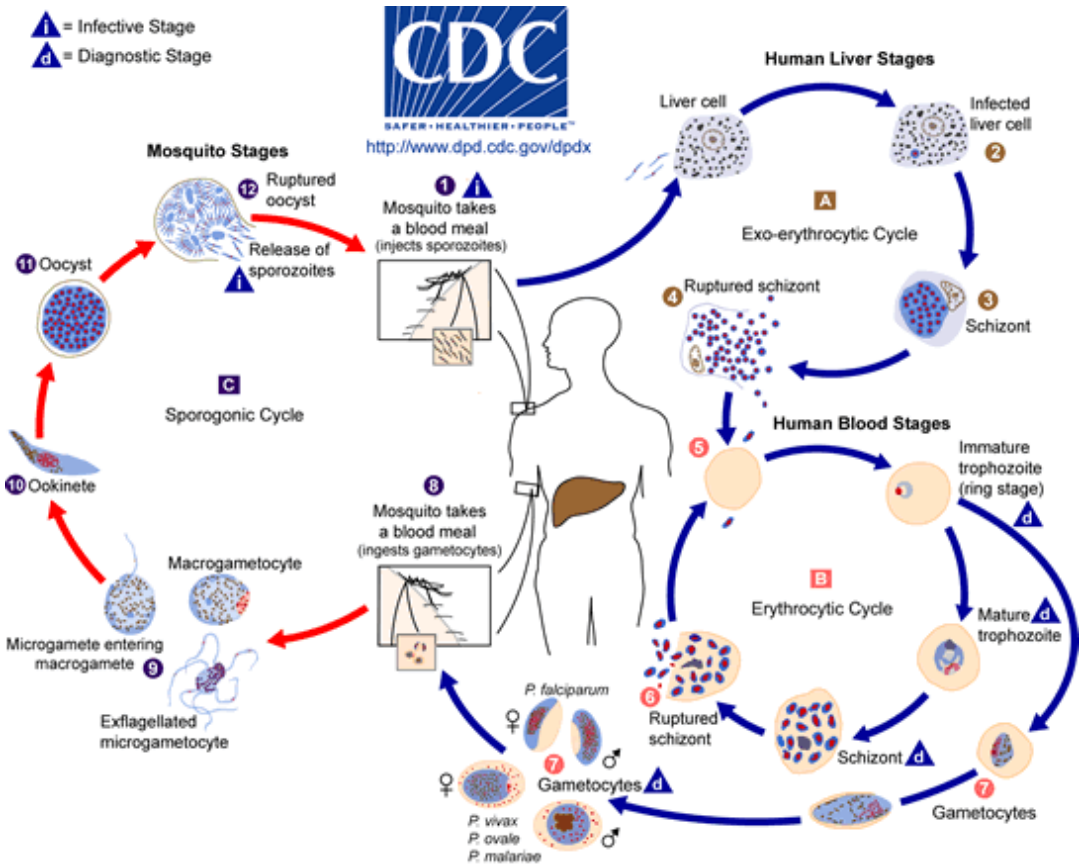


Fig. 1.1. The Life Cycle of *Plasmodium.* Reproduced from the Center for Disease Control & Prevention, Division of Parasitic Disease (DPDx) 2008.

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

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

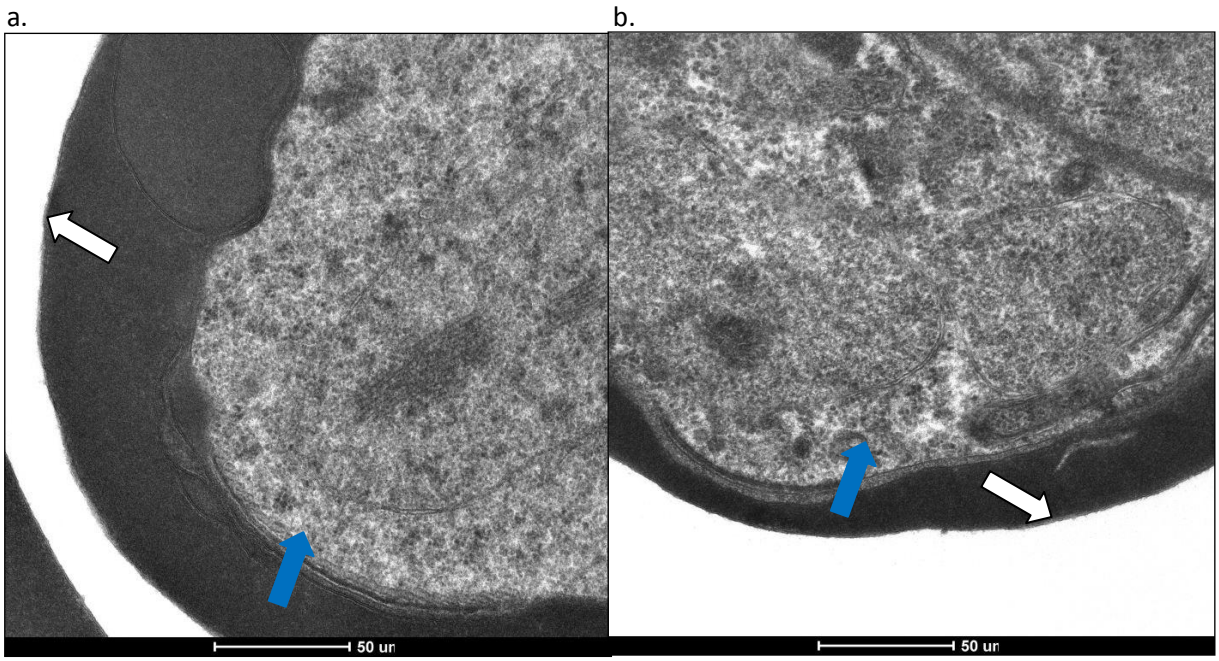


Fig. 1.2. Transmission electron micrographs of two Peruvian *P. falciparum* isolates in the erythrocytic stage. White arrows indicate the erythrocyte membrane and Blue arrows indicate the parasite. (a) Sample 3106. (b) Sample 3769.

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

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

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

Molecular Basis of Erythrocyte Invasion

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

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

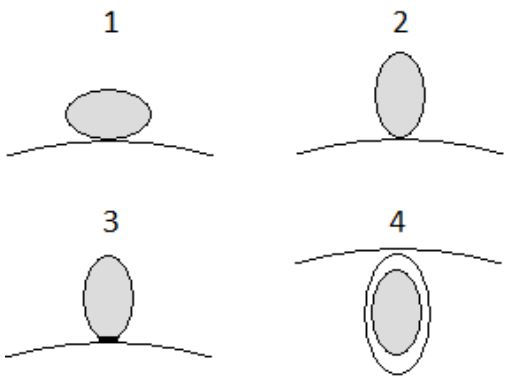


Fig. 1.3. The Stages of Erythrocyte Invasion. 1. Primary contact; 2. Reorientation; 3. Secondary ligand interactions; 4. Invasion.

1) Primary Contact

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

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

2) Reorientation

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

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

3) Secondary Ligand Interactions

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

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

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

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

4) Host Cell Invasion

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

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

Variation in Invasion Pathways

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

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

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

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

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

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

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

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

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

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

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

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

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

Rationale for This Study

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

1) The Sequencing Revolution

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Study Outline

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

- The two-colour flow cytometry phenotyping platform was used to generate a profile of invasion for each isolate, based on its ability to invade enzyme-treated erythrocytes, and to compare the parasite multiplication rate (PMR) of each isolate.

- DNA from each isolate was extracted and sequenced using the Illumina platform. Whole genome analysis was carried out using bioinformatics tools developed at the Wellcome Trust Sanger Institute specifically for the analysis of *P. falciparum* genomes.

- Parasite schizonts were isolated and RNA was extracted. This will be reverse transcribed to a cDNA library which will be sequenced using the Illumina platform allowing analysis of the whole transcriptome to be accomplished.

- Once investigations of each isolate were complete, multiple aliquots were frozen for future investigation.

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

Chapter Two

Materials and Methods

The experimental work was conducted between October 2009 and August 2010 and all the protocols are described in this chapter. All work with live *P. falciparum* was carried out *in vitro* in containment level 3 laboratories at The Wellcome Trust Sanger Institute. Except where noted, I performed all experiments described herein.

Cell Culture Reagents

- Incomplete medium
    - 10.43 g RPMI 1640 (Invitrogen)
    - 7.15 g HEPES (Sigma Aldrich)
    - 1 ml 1 mg/mL hypoxanthine (Sigma Aldrich)
    - 0.5 ml 1 mg/mL gentamicin (Sigma Aldrich)
    - 2 g Glucose (Sigma Aldrich)
    - 1 L Milli-Q water (Millipore)
- Complete medium containing 10% sera
    - 450 mL incomplete medium supplemented with:
    - 50ml human sera (NHS Blood Service)
    - 16 mL 7.5% sodium bicarbonate (Sigma Aldrich)
- Glycerolyte 57
    - 57 g glycerol (Sigma Aldrich)
    - 1.6 g sodium lactate (Sigma Aldrich)
    - 30 mg potassium chloride (Sigma Aldrich)
    - 51.7 mg monobasic sodium phosphate (monohydrate) (Sigma Aldrich)
    - 124.2mg dibasic sodium phosphate (anhydrous) (Sigma Aldrich)
    - 100 mL Milli-Q water

Solutions above were filtered through a 0.22 µm membrane to sterilise before use using Stericup® filter units (Millipore). Other solutions used were supplied as a sterile solution from the manufacturer.

In vitro **culture of** ***Plasmodium falciparum*** **parasites**

Samples of *Plasmodium falciparum* infected erythrocytes from patients in Peru were stored at −80°C and thawed in a water bath at +37°C. The volume of the sample was measured and 10 µL aliquots of 12% sodium chloride solution were added to a total of 0.2 times the sample volume. After 5 minutes incubation at room temperature 8 mL of 1.6% sodium chloride solution was added 100 µL at a time. The sample was centrifuged at 800 ***g*** for 5 minutes with no braking, the supernatant solution was removed and the pellet was resuspended in 8 mL of 0.2% D–glucose and 0.9% sodium chloride added in 100 µL drops. The sample was centrifuged as above and the pellet resuspended in complete medium containing human AB sera, with 400 µL of 50% haematocrit O^+ erythrocytes (NHS Blood & Transplant Service) in RPMI 1640 (Invitrogen). Prior to use, all erythrocytes were filtered using a Lymphoprep™ (Axis-Shield) density gradient to remove leucocytes, and washed with RPMI 1640. Samples were cultured in complete medium containing human AB sera for the first 48 hours before this was replaced with pooled sera. As AB sera lacks anti-A and anti-B antibodies, this 48 hour period prevents complement-mediated lysis before O+ erythrocytes have been successfully invaded by merozoites, as the blood group of the patient was unknown (Bakács et al. 1993).

Cultures were maintained at approximately 5% haematocrit, a parasitaemia of less than 10% and in an atmosphere of malaria gas (1% O_2, 3% CO_2 and 96% N_2). Complete medium was changed at least every 48 hours. Routinely, complete medium contained 10% human sera but this was increased to 15% if parasites failed to multiply.

Synchronisation of Parasites

Prior to most experiments, cultures were synchronised at the ring stage by adding a 5% sorbitol (Sigma Aldrich) solution (Lambros & Vanderberg 1979). Cultures were transferred to a Falcon tube and centrifuged at 800 ***g*** for 5 minutes with no braking and the supernatant solution was removed. The volume of the remaining erythrocyte pellet was estimated, and the pellet was resuspended in 5 times the pellet volume of 5% sorbitol solution. The suspension was mixed thoroughly by repeated pipetting and left at room temperature for 5 minutes, before centrifuging as above, washing in RPMI 1640 and finally

resuspending in complete medium and returning to culture conditions. Parasites were routinely sorbitol-treated twice at least 6 hours apart to ensure tight synchronisation.

Phenotyping Invasion Assay

Erythrocyte labelling

50% haematocrit O+ erythrocytes were diluted to 2% haematocrit in RPMI 1640 and centrifuged at 450 ***g*** for 3 minutes. The pellet was resuspended to 2% haematocrit with 10 µM 7-hydroxy-9*H*-(1,3-dichloro-9,9-dimethylacridin-2-one) succinimidyl ester (DDAO-SE) (Invitrogen) in RPMI 1640 and incubated for 2 hours at 37°C on a MACSmix™ tube rotator (Miltenyi Biotec). Following the incubation, erythrocytes were pelleted at 450 ***g*** for 3 minutes, washed with pre-warmed (37°C) complete medium, then resuspended to 2% haematocrit in complete medium and incubated for 30 minutes at 37°C on a MACSmix™ tube rotator. Erythrocytes were then pelleted at 450 ***g*** for three minutes, washed twice in complete medium, and resuspended at 2% haematocrit in incomplete medium (medium lacking human sera). The suspension was either used immediately or stored at 4°C for a period of no more than 24 hours.

Enzymatic treatment of DDAO-SE labelled O+ human erythrocytes

O+ human erythrocytes labelled with DDAO-SE (as described above) were centrifuged at 450 ***g*** for three minutes and the pellet was resuspended at 2% haematocrit with incomplete medium and aliquoted into individual 1.5 mL microfuge tubes. Neuraminidase from *Vibrio cholerae* (Sigma-Aldrich) was diluted in incomplete medium and added to the appropriate tubes to obtain a final concentration of 20 mU/mL. All tubes were then incubated at 37°C for 1 hour on a MACSmix™ tube rotator.

Following incubation, the tubes were centrifuged and the pellet was washed with incomplete medium. The pellets were resuspended to 2% haematocrit with incomplete medium. Trypsin (Sigma-Aldrich) or chymotrypsin (Sigma-Aldrich) was added to the appropriate tubes. The final concentration of trypsin was 50 µg/mL (low trypsin treatment) or 1 mg/mL (high trypsin treatment). The final concentration of chymotrypsin was 1 mg/mL. All tubes were incubated at 37°C for 1 hour while on a MACSmix™ tube rotator. Following

incubation, the tubes were centrifuged and the pellet was washed twice with incomplete medium before the pellet was resuspended at 2% haematocrit in complete medium.

Invasion Assays

For an invasion assay an aliquot of the culture was diluted with complete medium to 2% haematocrit and a parasitaemia of 1.5% before being transferred to a round-bottom 96-well plate in aliquots of 50 µL per well. Parasitaemia was determined prior to dilution by count of a thin smear stained with Field's stain. 50 µL of DDAO-SE labelled erythrocytes that had undergone differential enzyme treatment were then added to the appropriate wells resulting in a total culture volume of 100 µL in each well with a 1:1 ratio of parasitised erythrocytes (pRBC) to DDAO-SE labelled erythrocytes (sRBC). All invasion assays were carried out in triplicate, although if a post-invasion trypsin treatment was performed (described below) six replicate wells were started. Each well suspension was mixed thoroughly by repeated pipetting to try to ensure an equal probability of parasite reinvasion occurring into stained or unstained erythrocytes. Wells not in use were filled with PBS to reduce evaporation from the plate. The plate was placed in a gas chamber and malaria gas was run through the chamber for three minutes at high flow rate prior to sealing the chamber and incubation at 37°C for 48 hours.

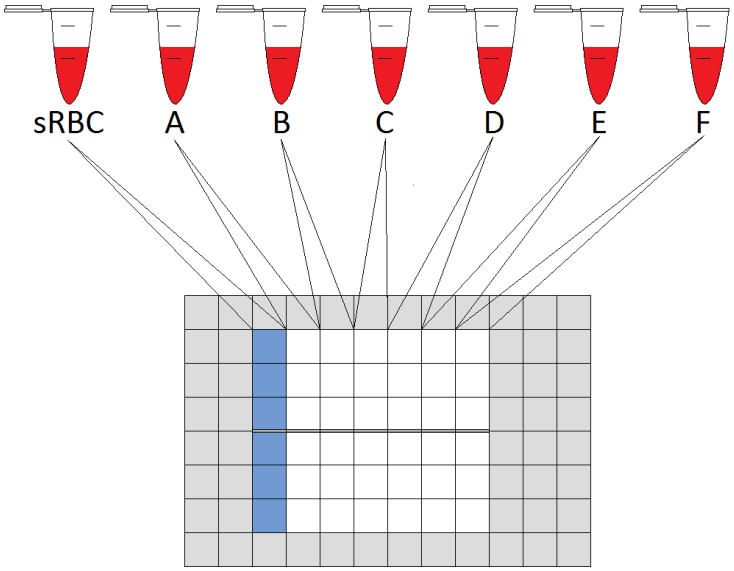


Fig. 2.1. Enzyme treatments in tubes and the 96-well plate set up.
Tubes: sRBC: stained untreated RBC; A: untreated (identical to Tube sRBC); B: neuraminidase treated; C: low trypsin treated; D: high trypsin treated; E: neuraminidase and high trypsin treated; F: chymotrypsin treated.
96-well plate: grey wells contain PBS; blue wells contain no parasites, only sRBC and unstained RBC (a control for flow analysis); white wells contain a 1:1 mix of sRBC and pRBC. Only the top three rows are trypsin treated post-invasion.

Post-Invasion Trypsin Treatment, Fixation and Parasite Labelling

After 48 hours, the plate was removed from incubation and centrifuged at 450 ***g*** for 3 minutes. 50 µL of supernatant solution was removed from each well and the pellets washed in 200 µL of PBS (Sigma Aldrich). A post invasion trypsin treatment was then performed for most assays, necessary because of a specific phenotype observed in some strains, as discussed in Chapter 3 "Results". For the trypsin treatment 100 µL of 1 mg/mL trypsin was added to the appropriate wells, and the wells were mixed thoroughly by repeated pipetting. The plate was incubated for 1 hour at 37°C before being centrifuged at 450 ***g*** for 3 minutes and each well washed once with 200 µL PBS. Fixation was performed by adding 200 µL of a 2% paraformaldehyde (PFA), 0.2% glutaraldehyde (GA) solution in PBS to each well and the plate incubated at 4°C for 1 hour. Each well was then washed once with 200 µL PBS.

Parasite labelling was achieved by using one of two DNA dyes. The first method required the cells to be permeabilised with 200 µL 0.3% Triton®X-100 (Sigma Aldrich, Dorset, UK) in PBS per well. The wells were mixed thoroughly by repeated pipetting and left at room temperature for 10 minutes. The plate was centrifuged at 450 ***g*** for 3 minutes and each well was then washed in 200 µL PBS. The pellet was resuspended with 200 µL 0.5 mg/mL ribonuclease A (MP Biomedicals) in PBS. Each well was mixed thoroughly and the plate was incubated at 37°C for 1 hour. Following incubation the plate was centrifuged at 450 ***g*** for 3 minutes and each well was then washed in 200 µL PBS. 200 µL of 1:5000 SYBR Green I (Invitrogen) in PBS was then added to each well and mixed thoroughly by repeated pipetting before the plate was incubated at 37°C for 1 hour. After incubation, the plate was centrifuged at 450 ***g*** for 3 minutes and each well was washed with 200 µL PBS. The PBS wash was repeated twice before each pellet was resuspended with 200 µL PBS and the plate was stored at 4°C in the dark for up to 72 hours.

The second method of parasite labelling was preferred as it did not require permeabilisation of cells or the RNase treatment. Following fixation and the subsequent wash with PBS, each pellet was resuspended in 200 µL of 2 µM Hoechst 33342 (Invitrogen) in RPMI 1640 and the plate was incubated at 37°C for 1 hour. After incubation, the plate was centrifuged at 450 ***g*** for 3 minutes before each well was washed three times with 200

µL PBS and the cells resuspended in 200 µL PBS and stored at 4°C in the dark for up to 72 hours.

Data Acquisition by Flow Cytometry and Data Analysis

25 µL aliquots were transferred from each stained well into a new flat-bottomed 96-well plate containing 250 µL RPMI 1640 per well. The samples were examined with a 355 nm 20 mW UV laser, a 488nm 20 mW blue laser and a 633 nm 17 mW red laser on a BD LSRII flow cytometer (BD Biosciences). Hoechst 33342 was excited by a UV laser and detected by a 450/50 filter. DDAO-SE was excited by a red laser and detected by a 660/20 filter. BD FACS Diva (BD Biosciences) software was used to detect 100,000 events per well. The data collected were exported to FlowJo (Tree Star) for further analysis. Statistical analysis and phenotype profiles of parasitaemia were generated using GraphPad Prism (GraphPad Software). For the purpose of phenotyping, all experiments were carried out with six replicates (half underwent post-invasion trypsin treatment) and were performed twice in two separate life cycles, with data presented as the mean ± standard error of the mean of both experiments combined unless otherwise stated.

Parasite DNA Extraction and Sequencing

Preparation of Cultures for DNA Extraction

Cultures were synchronised and the volume was expanded to at least 50 mL at 5% haematocrit, with a minimum parasitaemia of 2%. When parasites were in the late trophozoite or schizont stage the culture was transferred to a Falcon tube and centrifuged at 800 ***g*** for 5 minutes with no braking. The pellet was washed in 20 mL RPMI 1640, centrifuged again, and then as much supernatant solution was removed as possible. The tube containing the pellet was then placed in a -20°C freezer prior to DNA extraction.

DNA Extraction

Parasite DNA was extracted and purified from erythrocyte pellets using the QIAamp Blood Maxi Kit (Qiagen®). Sequencing was then undertaken in three steps: library preparation, bridge amplification and sequencing by synthesis.

DNA extraction from isolates was carried out by Professor Dominic Kwiatkowski's team.

Library Preparation

Parasitic DNA was randomly fragmented by sonication and library preparation was carried out using an Illumina Library Prep Kit (Illumina). DNA fragments are purified and the overhangs resulting from sonication are repaired to blunt ends using T4 DNA polymerase and *E. coli* DNA polymerase I Klenow fragment. The 3' to 5' exonuclease activity of these enzymes removes 3' overhangs and the polymerase activity fills in the 5' overhangs. The 3' ends are adenylated using the polymerase activity of the Klenow fragment. The fragments can then be ligated to adapters that have a single thymine base overhang at their 3' ends. Solid Phase Reversible Immobilisation (SPRI) paramagnetic beads (Beckman Coulter) are then used to purify and size select fragments of 200 – 300 bp.

Bridge Amplification

Oligos complementary to the adapters are irreversibly bound to the surface of the flow cell. When the library fragments are added, the flow cell is heated to denature the DNA molecules into single strands which hybridise to the complementary oligos. Polynucleotide synthesis uses the oligo as a primer and the ssDNA as a template. The result is a double stranded DNA molecule with each strand anchored to the flow cell at one end by the attached oligo. Denaturation leaves two ssDNA templates for the next round of clonal amplification. In this fashion, several million unique ssDNA clusters are generated in each of the eight flow cell channels. The ends of each strand are then blocked and a sequencing primer is hybridised to each strand.

Sequencing by Synthesis

Clusters are sequenced simultaneously, base by base using an Illumina IIx Genome Analyser. Four fluorescently labelled, reversibly terminated bases are added to the flow cell and compete to hybridise to the template, with only the base complementary to the template able to hybridise. The cluster is then excited by a laser, with the colour emitted being different for each of the four bases allowing the sequence to be determined. The

fluorescent label and blocking group that prevents the addition of more than one base at a time are then removed and the cycle is repeated. From the 200 – 300 bp fragments generated in library preparation of Peruvian field isolates, 76 bp were sequenced from each end.

Genotyping Analysis

Whole genome sequencing analysis was carried out using tools developed by Professor Kwiatkowski's team at the Wellcome Trust Sanger Institute.

SNP-o-matic is a tool that aligns paired Illumina reads to the 3D7 reference sequence. Putative variable positions are specified at the outset. Reads are aligned only if they are perfect matches to the reference sequence, unless the variation occurs in one of the specified variable positions. The requirement for perfect matches means that false SNP calls are not made from misaligned reads (Manske & Kwiatkowski 2009).

MapSeq (http://www.sanger.ac.uk/MapSeq) is a browser-based tool for analysing *Plasmodium falciparum* genomes. MapSeq generates a library of all single nucleotide polymorphisms (SNPs) for a given strain, and lists them in order of location in the genome. It allows SNPs to be browsed and analysed across the whole genome, or by specific genome region or gene. Many parameters are user-defined, such as the type of SNP to be analysed (in coding or non-coding regions, synonymous or non-synonymous), or the stringency of evidence required to call SNPs (minimum number of reads required, minimum percentage of samples that a SNP has to be seen in). For analysis of Peruvian strains the default settings were used: a minimum of 10 reads were required for a SNP call, and a minimum of 75% of samples had to be called at that position.

MapSeq also allows the user to link to LookSeq, a browser-based tool to view read alignments from a whole chromosome down to a single base (Manske & Kwiatkowski 2009). Multiple samples can also be viewed and compared simultaneously. SNPs in erythrocyte invasion ligands were all verified by direct inspection of the individual reads via LookSeq.

Mutabo! is a browser-based program that generates a list of all amino acid changes for non-synonymous SNPs. It does not allow the minimum number of reads required for a

call to be pre-determined so all amino acid changes generated in Mutabo! had to be checked with MapSeq to verify sequence quality.

Parasite RNA Extraction and Sequencing

Schizont Enrichment in Preparation for RNA Extraction

Each culture was expanded to 50 mL at 5% haematocrit and synchronised at the ring stage using 5% sorbitol treatments, as detailed previously. A minimum parasitaemia of 2% was required. The development of the parasites was followed by regular blood smear microscopy and Field's staining, as above. When the majority of parasites were found to be schizonts, these were purified using a 60% Percoll™ gradient (GE Healthcare) (Saul et al. 1982). For Percoll™ purification, the erythrocytes were pelleted at 800 ***g*** for 5 minutes with no braking, to produce a 2.5 mL erythrocyte pellet. All subsequent centrifugations were carried out at 4°C and all solutions were kept on ice. The pellet was resuspended with 20 mL complete medium containing 10% heat-inactivated fetal bovine serum (FBS) and centrifuged at 800 ***g*** for 5 minutes with no braking. The supernatant solution was removed and the pellet washed again in FBS-containing complete medium. The 2.5 mL erythrocyte pellet was then resuspended in 22.5 mL of FBS-complete medium and loaded on top of 31.25 mL of a 60% Percoll solution before centrifugation for 20 minutes at 1500 ***g*** with no braking. The top layer containing the enriched schizonts was then collected into a new tube containing 20 mL PBS, the schizonts were pelleted as above and washed with PBS. The purity of the final pellet was evaluated by light microscopy with Field's staining, before the pellet was resuspended in 1 mL RNALater (Ambion®) and stored at -20°C.

RNA Extraction from Schizonts

Purified schizonts stored in RNALater were thawed and pelleted by centrifugation at 8000 ***g*** for 1 minute. The supernatant solution was removed and RNA was extracted from the schizonts using a Qiagen QIAamp® RNA Blood Mini Kit (Qiagen®). The protocol used was "Purification of Total Cellular RNA from Human Whole Blood", and included the "Optional On-Column DNase Digestion with the RNase-Free DNase Set". The final elution with RNase-free water was performed with a final volume of 60 µL. The quantity and purity

of RNA was then assessed by placing 3 – 4 µL on an Implen NanoPhotometer™ (Implen). The samples were stored at -80°C.

Freezing of Parasites

Aliquots were frozen from cultures with a parasitaemia of at least 2% rings. The culture was centrifuged at 800 ***g*** for 5 minutes with no braking and the pellet was washed once with 10 mL RPMI 1640. After centrifugation, the pellet volume was estimated and half the pellet volume of glycerolyte 57 solution was added one 25 µL drop at a time. The suspension was incubated at room temperature for 5 minutes before a further one and a half times the pellet volume of glycerolyte 57 solution was added one 50 µL drop at a time. The suspension was then aliquoted to tubes with approximately 200 µL of pellet per tube. These tubes were placed in a -80°C freezer for at least 24 hours before being transferred to storage in liquid nitrogen vapour at -130°C.

Fluorescence Microscopy DDAO-SE or Hoechst 33342 Stained Samples

Thin blood smears were made on a slide and air dried for a minimum of 4 hours. The samples were fixed by exposure to a 1:1 methanol & acetone solution. After drying, a drop of Fluoro-Gel with Anti-Fading Mounting Medium (Electron Microscopy Sciences) was placed on the slide and a coverslip was put on top. The slide was dried at 4°C for 24 hours. Slides were observed using either a Leica DM2500 fluorescence microscope and a Leica LAS AF camera (Leica Microsystems) or a Zeiss LSM510 laser scanning confocal microscope (Carl Zeiss) and a Leica DFC420C camera (Leica Microsystems). Leyla Bustamante lent invaluable assistance in performing confocal microscopy.

Additional Techniques Used

When two Peruvian field isolates (3135 & 3769) were phenotyped by invasion assay, extra populations were observed by flow cytometry. The following techniques were used to investigate the cause of these unforeseen populations.

Erythrocyte Rosetting Assay

An aliquot from Peru strain 3769 was diluted to 2% haematocrit with complete media. 1 mL was centrifuged before the pellet was resuspended with 2 µM Hoechst 33342 in complete media, maintaining a 2% haematocrit. The suspension was incubated at 37°C for 1 hour before being centrifuged and washed with complete media three times. The pellet was then resuspended and diluted to 0.5% haematocrit with complete media and 4 µL was placed on a slide with a coverslip sealed on top with nail varnish. Using a Leica DM2500 fluorescence microscope, 200 parasitised red blood cells were counted. Rosettes were defined as one parasitised erythrocyte binding two or more uninfected erythrocytes (Doumbo et al. 2009).

Electron Microscopy of Parasites

A 10 mL culture at 5% haematocrit was cultured to a high parasitaemia before being fixed in 2% PFA / 0.2% GA in PBS. David Goulding was responsible for preparing the samples further and imaging.

Merozoite Detection Using anti-MSP-1 Antibody

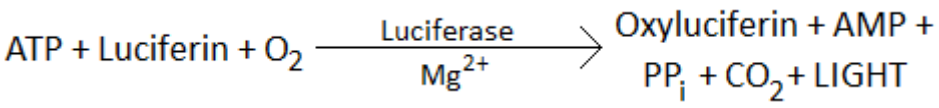
Aliquots from two Peruvian strains, 3769 and 3541, were both adjusted separately to 1% haematocrit, synchronised cultures of 6% rings and 2% schizonts. 100 µL aliquots were transferred to a round-bottom 96 well plate. The plate was centrifuged at 450 ***g*** for 3 minutes, and the supernatant solution was removed. Wells containing schizonts were resuspended in 100 µL 2% PFA / 0.2% GA in PBS, while wells containing rings were resuspended in PBS. The plate was incubated at 4°C for 1 hour before being centrifuged as above, and all wells were washed with PBS. Wells containing schizonts were then permeabilised by resuspending the pellet with 100 µL 0.3% Triton®X-100 in PBS and incubating the plate for 10 minutes at room temperature. Wells containing rings were

resuspended in 100 µL PBS. Following the permeabilisation step, the plate was centrifuged as above, and all pellets were washed three times with 100 µL PBS. 100 µL 1% bovine serum albumin (Sigma Aldrich) in PBS was added to each pellet and the plate was incubated for 30 minutes at room temperature. The plate was centrifuged (as above) and the supernatant solution removed. Mouse monoclonal anti-MSP-1 19 kDa IgG antibody (abcam) was diluted 1:500 in 1% bovine serum albumin (BSA) in PBS and 100 µL was added to each of the appropriate wells. Control samples only being exposed to the secondary antibody were resuspended with 100 µL 1% BSA in PBS. The plate was incubated at room temperature for 1 hour. Subsequently, the plate was centrifuged as above and all wells were washed with 100 µL 1% BSA in PBS three times. Goat polyclonal IgG secondary antibody conjugated with FITC (abcam) was then diluted 1:2000 with 1% BSA in PBS and 100 µL was added to all wells. The plate was incubated at room temperature for one hour before all wells were washed three times with 100 µL 1% BSA in PBS. Wells containing rings were fixed with 2% PFA / 0.2% GA in PBS using the same protocol as above. All pellets were then washed three times with 100 µL PBS, before being resuspended in 100 µL of PBS. Wells were acquired using a BD Calibur flow cytometer (BD Biosciences) and 100,000 events were counted.

Peruvian isolate 3541 at a parasitaemia of 6% rings was also probed with an antibody to CD147. The method was the same as used above for wells containing rings except for the steps involving antibodies. Prior to adding the primary antibody, an aliquot of 6 µL of the 1% haematocrit suspension was added to a new well containing 94 µL 1% BSA in PBS (i.e. the haematocrit of the new well was 0.06%). 0.5 µL of primary monoclonal mouse anti-CD147 IgG antibody (Exbio Antibodies) was added to the appropriate wells and the plate was incubated at room temperature for 1 hour. After centrifugation and three washes with 1% BSA in PBS, the wells were resuspended in 1% BSA in PBS. 0.5 µL of goat polyclonal IgG secondary antibody (abcam) conjugated with FITC was added to the appropriate wells and the plate was incubated at room temperature for 1 hour. After washing with PBS, as above, 10,000 events were acquired using a BD Calibur flow cytometer.

Mycoplasma Detection Assay

Three cultures, Peru 788, 3769 and laboratory strain 7C126 were cultured from 0.6% rings for 72 hours without replacing the media. Each culture was then tested using a MycoAlert® Mycoplasma Detection Kit (Lonza). 100 µL of the culture supernatant solution was transferred to a luminometer cuvette. 100 µL of MycoAlert® Reagent was added and each sample was incubated for 5 minutes at room temperature. The purpose of the MycoAlert® Reagent was to lyse any mycoplasma present. A reading (A) of luminescence was taken using an Orion II Microplate Luminometer (Berthold Detection Systems). 100 µL of MycoAlert® substrate was added and each sample was incubated for 10 minutes at room temperature. A second reading (B) of luminescence was taken. Enzymes released by the lysing of the mycoplasma react with MycoAlert® substrate, catalysing the phosphorylation of ADP to ATP. If mycoplasma are present the rise in ATP is detected by bioluminescence according to the following reaction:



A ratio of luminescence readings B:A >1 is indicative of mycoplasma contamination.

Chapter 4

Discussion

The work described in this thesis was dependent upon the successful establishment of cultures of *Plasmodium falciparum* parasites from blood samples collected by collaborators in Peru and supplied as frozen isolates of samples that had been cultured initially in Peru. Culture adaptation of field isolates has typically had success rates around 60% (Gomez-Escobar 2010; Baum et al 2003), so the 24% growth success rate seen in the Peruvian field isolates is particularly low. This is compounded by the fact that these isolates were previously culture-adapted in Peru. There is no single reason for the low success rate of growth, however some samples appeared to have thawed and re-frozen either in Peru (where power outages are frequent) or in transit from Peru, which would severely impact their viability. The volumes of other samples suggested that they had not been frozen in a suitable volume of glycerolyte 57 which is also likely to impact viability. Although these technical errors exacerbated the problem, even once successfully grown at the Sanger Institute, Peruvian field isolates demonstrated an intrinsic susceptibility to the freezing process. All non-contaminated isolates grown and frozen at the Sanger Institute failed to be re-grown during multiple attempts (278 – 2 attempts, 6390 – 3 attempts; 9050 – 4 attempts). Only in the case of 6390 was a culture eventually successful after a fourth tube was thawed. It was observed that upon re-culturing these isolates, parasites were initially present but asexual parasites were lost to gametocytogenesis within the first few weeks and the culture subsequently died. Re-growth of the other successfully grown Peruvian isolates (5802, 5809 and 5814) has yet to be attempted. This susceptibility contrasted with the behaviour of the isolates that were confirmed as W2 contaminants (788, 3106, 3135, 3541 and 3769) which showed no such problems with the freezing process.

Among the non-contaminated Peruvian isolates that were successfully invasion phenotyped (278, 5802, 5809, 5814 and 6390) there was substantial variation in the dependence upon sialic acid-containing and chymotrypsin-sensitive receptors, although all isolates were highly dependent upon trypsin-sensitive receptors. None of the isolates were dependent upon sialic acid residues to the same extent as Dd2 or the W2 contaminants. The increased ability of these isolates (particularly 278 and 5809) to invade sialic acid depleted erythrocytes relative to Dd2 suggests that they are not reliant solely upon ligands of the DBL family for invasion which recognise the major sialoglycoproteins, the glycophorins. In the

case of isolate 5809, high resistance to neuraminidase and high sensitivity to trypsin- and chymotrypsin- mediated depletion of erythrocyte ligands implies the role of the PfRh4 pathway of invasion. The moderate sensitivity of the other isolates to neuraminidase and trypsin treated erythrocytes makes it harder to propose a reliance on a single receptor or ligand although for 6390, which invades chymotrypsin treated erythrocytes relatively well compared to those treated with neuraminidase or trypsin, arguments could be made that it preferentially invades cells expressing either glycophorin A or C, which are both chymotrypsin resistant. It should also be emphasised that the number of erythrocyte receptors that have been shown to bind merozoite ligands is not fully defined and therefore alternative interactions with similar enzyme sensitivities could be taking place that have yet to be identified.

Previous studies have defined sensitivity to an enzyme treatment as an invasion efficiency of less than 50%. In Brazil four different invasion phenotypes were observed in fourteen isolates. In the Peruvian samples three different invasion phenotypes were observed in the five non-contaminated isolates. The neuraminidase resistant, trypsin sensitive and chymotrypsin sensitive phenotype shared by 278 and 5809 was not seen in Brazil, however this profile was the predominant invasion phenotype seen in Kenyan isolates (Deans et al. 2007). 5802, 5814 and 6390 displayed invasion profiles similar to those seen in Brazil.

Although the sample numbers are low, the high level of diversity in invasion phenotype from Peruvian isolates already observed is an indication that natural variation of invasion pathways is intrinsically present, even without the high genetic diversity associated with hyper-endemic transmission regions such as in sub-Saharan Africa.

The negative correlation between invasion efficiency into neuraminidase- and chymotrypsin-treated erythrocytes was contrary to studies of samples obtained from Tanzania and Senegal. In Tanzanian samples, significant positive correlation was found between invasion into trypsin- and chymotrypsin-treated erythrocytes. However, once the phenotypes of W2 contaminated samples were removed, the Peruvian sample numbers were no longer sufficient to produce significant correlation (Fig. 4.1). The positive

correlation between PMR and invasion into chymotrypsin-treated erythrocytes was also found to be insignificant when the contaminants were excluded (Fig. 4.2).

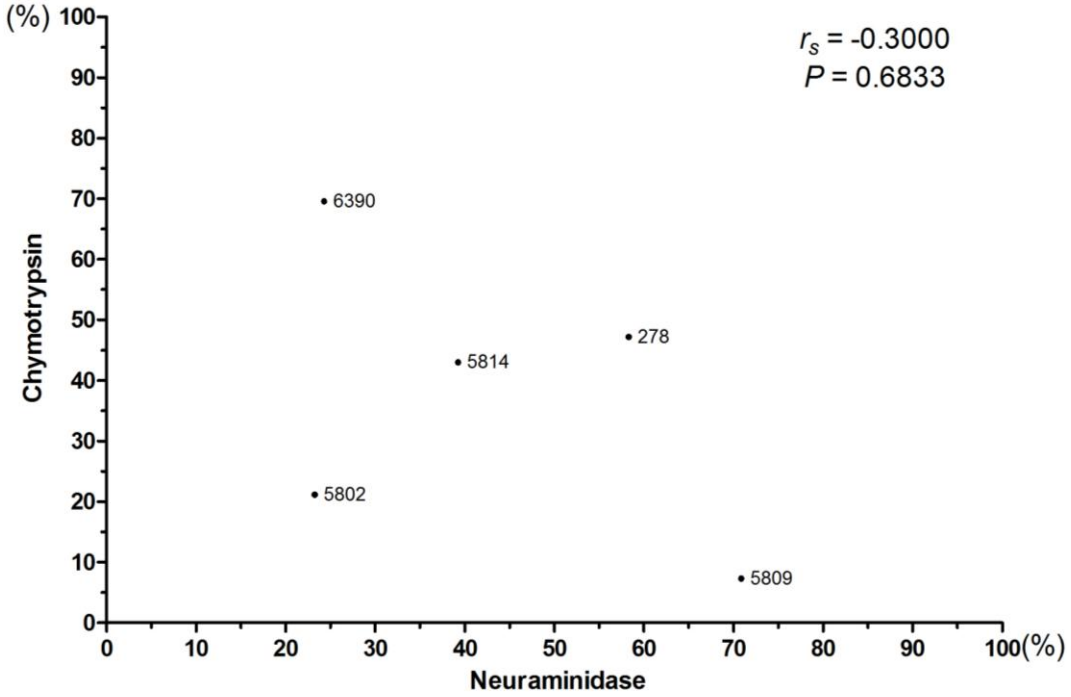


Fig. 4.1. Invasion efficiencies (%) into neuraminidase treated (x-axis) and chymotrypsin treated (y-axis) erythrocytes. Spearman's rank correlation coefficient (r_s) and a two-tailed *P*-value are given for each plot. Correlation with a *P*-value of <0.05 was considered significant.

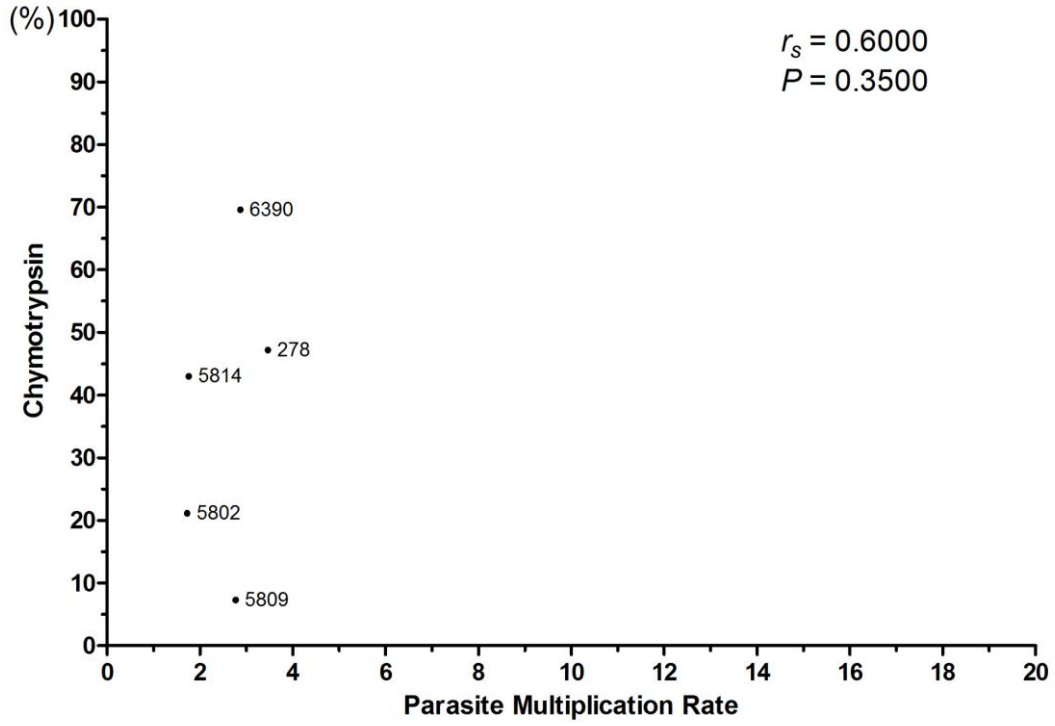


Fig. 4.2. PMR (x-axis) and invasion efficiency (%) into chymotrypsin treated (y-axis) erythrocytes. Spearman's rank correlation coefficient (r_s) and a two-tailed *P*-value are given for each plot. Correlation with a *P*-value of <0.05 was considered significant.

The aim of this study was to generate phenotype – genotype correlations to study erythrocyte invasion. With only one of the phenotyped samples having been genotyped at the time of writing, this has not been possible. However the remaining samples should be sequenced within a few months and although still limited by the small number of samples, preliminary correlations between sequence polymorphism and invasion phenotype may be possible.

What the study has clearly demonstrated is the potential to apply whole genome sequencing and high throughput phenotyping to study erythrocyte invasion. The identification of 93 high confidence, non-synonymous SNPs and 80 corresponding amino acid changes across 62 genes associated with invasion in a single isolate, 6390, highlights the power of the technology. Previously, genotyping of samples has been limited to a few polymorphic loci and used primarily for the purpose of detecting multiple infections (Okoyeh et al. 1999; Lobo et al. 2004; Deans et al. 2007). This new in-depth approach could give new insight into whether genetic variation dictates the invasion pathway utilised.

The tools available in MapSeq were used to group samples by PCA analysis according to their genetic similarities. The two non-contaminated isolates from Peru grouped with HB3, the lab strain of Honduran origin. It will be interesting to see whether future Peruvian isolates also group in a geographical manner. The addition in the near future of around 400 other samples from around the world to the MapSeq database, will provide valuable comparators for further analysis of the Peruvian samples. The origins of *P. falciparum* in South America are unclear, although it is widely believed to have been transferred with European explorers, possibly via African slaves. Whole genome sequencing analysis of multiple Peruvian and South American isolates will make the origin of *P. falciparum* malaria in South America easier to decipher.

The other benefit of genome comparison and PCA analysis was that it immediately identified five samples as being virtually identical to Dd2 / W2. Phenotypically, the contaminated isolates are similar, but by no means identical to the laboratory strains, with slight variation existing in both trypsin and chymotrypsin sensitivity. There was also significant variation present within the PMRs of contaminated phenotypes. PMR can be affected by many variables including the quality of erythrocytes, quantity and quality of

medium and variations in temperature. PMR also relied upon the determination of starting parasitaemia by counting parasites using slide microscopy. Slide microscopy based counts involve only 1-2,000 erythrocytes, as opposed to the 100,000 or more that can be counted by flow cytometry. Therefore there will inevitably be some variability in actual starting parasitaemia, which could account for the variation in PMR between W2 contaminated strains. However, the large variation seen between the PMR of contaminated samples combined with the small variations in invasion profile may have meant that if genotyping had not been performed, these samples may have been treated as true field isolates, leading to an inaccurate picture of invasion variation at the study site.

While it was hoped that both genomic and transcriptomic sequence analysis of these isolates would be carried out, the method for sequencing *Plasmodium* RNA is still being adapted from the human RNA sequencing protocol and currently requires at least 10 µg of RNA for cDNA library preparation. Some Peruvian samples had as little as 1.87 µg of RNA extracted so RNA sequencing will await further technical development.

The discovery of extra populations in Peruvian isolates 3106 and 3135 presented a problem to the use of flow cytometry to phenotype isolates for invasion. When the four separate populations that were normally present were merged by the presence of two extra populations with intermediate DNA staining, the manual placement of gates to count events was not accurate. This population was found to be invariable, occuring in every assay performed upon the two isolates, but it was not found in any other isolates. Surprisingly, this was not a feature of a Peruvian isolate, but in two of the W2 contaminated isolates. A solution to the problem was quickly found; a post-invasion treatment of all wells with a protease removed the extra populations but did not substantially affect the assessment of true parasitaemia or uninfected cell populations. Trypsin and chymotrypsin were equally effective at removing the intermediate populations but trypsin was used for post-invasion treatment.

Prior to the discovery of extra populations, invasion assays had been carried out using SYBR Green I to stain the parasitic DNA. SYBR Green I has the advantage over Hoechst 33342 that it does not require a UV laser, an expensive and rare addition to most flow cytometers. With the future potential for phenotyping to be performed in Peru the assay

was designed to keep costs to a minimum. However the SYBR Green I staining protocol requires permeabilisation and RNase treatment steps. When an hour-long trypsin treatment was added the post-invasion protocol for SYBR Green I staining was taking around six hours. Hoechst 33342 staining does not require permeabilisation and RNase treatment and therefore only takes 4 hours. To standardise the procedure for obtaining invasion profiles, invasion assays with post-invasion trypsin treatment were run in parallel to invasion assays without trypsin treatment, using the plate set up in Methods Fig. 2.1, and Hoechst 33342 staining was used. This method would have been used on all isolates but isolate 278 could not be re-grown so the original data had to be used. Theron et al. (manuscript submitted) compare parasitaemia counts between SYBR Green I and Hoechst 33342 staining and the methods produce almost identical results, so there is little concern that the use of a different dye for isolate 278 phenotyping will result in significant variation.

The source of the intermediate population was not established. The rosetting hypothesis was rejected and there was no evidence to support the hypothesis that merozoites were adhering to the surface. However the sensitivity of the population to protease enzyme treatment suggests a mechanism of extracellular adherence of some description. The high specificity of DNA stains, particularly Hoechst 33342 which binds DNA in the minor groove (Filatov et al. 1994), implies that this population contains DNA but in a smaller quantity than an erythrocyte containing a single parasite ring. The population was indiscernible using any method other than flow cytometry and no changes to the parasite or parasitised erythrocyte could be seen by electron microscopy. The presence of mycoplasma contamination in the culture would explain the smaller quantity of DNA being stained. The two Peruvian samples tested (both W2 contaminants) were found to contain mycoplasma but by flow cytometry only one of these had a significant extra population. As the mycoplasma assay is not quantitative, the amount of mycoplasma present in 3769 could be much higher than that of 788 and therefore is seen by flow cytometry. In the future, mycoplasma contamination testing will be routinely carried out across all isolates.

Improvements and Future Work

This study has shown that a high throughput invasion phenotyping assay can be used in conjunction with next generation sequencing to study genotype association with invasion phenotyping. Although contamination issues have reduced the power of the data generated, as a pilot for future studies it was clearly successful. The next stage of the project is to produce genotype – phenotype correlations as well as transcriptome analysis on a larger number of isolates. While the phenotyping assay itself is high throughput, especially compared to previous slide microscopy based techniques, the culture of Peruvian isolates has been limiting due to the poor culture success rates and very low growth rates once in culture. If the assay is going to become truly high throughput, then phenotyping would have to be done with samples straight from the arm. This has the advantages that no culture is required and eliminates the possibility that culture-adapted isolates may change their invasion pathway usage during the adaptation process. With the simple modifications made to be able to phenotype isolates that exhibited extra populations, the assay has already shown its adaptability. Only very minor modifications would have to be made to have the assay working in the field. Recently a BD Calibur flow cytometer was purchased by the group in Peru. The BD Calibur does not possess a UV laser, but it can be used to detect both DDAO-SE and SYBR Green I fluorescence. Studies involving samples transported straight to the lab from the field clinic, to allow phenotyping of the first round of erythrocyte invasion *in vitro,* are therefore possible. This approach, combined with the continual development of next generation sequencing technologies to be able to handle large numbers of samples, clearly paves the way for larger scale association studies.

Chapter 3

Results

Invasion Phenotyping

A total of 46 Peruvian *P. falciparum* samples were thawed and cultured *in* vitro. All samples had been isolated from patients in the Zungoracocha community near Iquitos, Peru, and had been cultured for between 1 week and 3 months at the Universidad Nacional Amazonia de Peruana. All samples were cultured for a minimum of six weeks after which, if no growth occurred, the sample was recorded as negative for growth.

Eleven isolates grew successfully (24%). Of these, ten isolates underwent phenotyping, genotyping and RNA extraction (Table 3.1). 9050 was genotyped and then frozen again, but as it could not be re-cultured successfully no phenotyping or RNA extraction was performed.

Sample	Phenotyped?	Sequenced?	RNA Extracted?	Gametocytogenesis?
278	Yes	Yes	Yes	Yes
788	Yes	Yes	Yes	No
3106	Yes	Yes	Yes	No
3135	Yes	Yes	Yes	No
3541	Yes	Yes	Yes	No
3769	Yes	Yes	Yes	No
5802	Yes	Yes	Yes	Yes
5809	Yes	Yes	Yes	Yes
5814	Yes	Yes	Yes	Yes
6390	Yes	Yes	Yes	Yes
9050	No	Yes	No	Yes

Table 3.1. The 11 isolates that were cultured successfully.

From the 35 samples that could not be cultured, three strains (including 9050) were successfully thawed and asexual parasites were observed in the first few days, but within two weeks the only parasites present were in the sexual gametocyte stage. By four weeks, all gametocytes had died. No parasites were observed in the other 32 samples at any point (Appendix Table A1).

In the eleven isolates that grew, parasites were observed no later than the second day of culture. Multiplication rates were very variable, with samples taking up to four weeks to reach a parasitaemia that was sufficiently high and synchronous to perform phenotyping assays. The time taken for cultures to reach sufficient parasitaemia was further elongated by the requirement of large volumes of culture for DNA and RNA extraction and for freezing

aliquots of each isolate. Samples with low multiplication rates also had a tendency to produce gametocytes in culture.

From the eleven samples that were cultured successfully, invasion assays were carried out on ten samples to assess the parasites' ability to invade receptor-depleted erythrocytes (9050 could not be re-cultured and was therefore not phenotyped, as noted above). Phenotyping was carried out using a two-colour flow cytometry based assay (Theron et al. 2010; see Chapter 2 "Materials and Methods"). This approach uses erythrocytes that have been labelled with an intracellular fluorescent dye (DDAO-SE) and then treated with one of several different enzymes that remove a subset of receptors utilised in parasite invasion from the surface of labelled erythrocytes. The labelled erythrocytes are mixed in equal volume with unlabelled *P. falciparum* infected erythrocytes and cultured for 48 hours before parasite density is measured using fluorescent DNA dyes, either SYBR Green I or Hoechst 33342. Dot plots were produced by a BD LSRII flow cytometer and from these parasitaemias in labelled and unlabelled erythrocytes were calculated (Fig. 3.1). Fluorescence labelling was confirmed by confocal microscopy (Fig. 3.2).

Nine samples were phenotyped using Hoechst 33342 to stain the parasitic DNA and one using SYBR Green I. Two isolates (3135 and 3769) required post-invasion trypsin treatment in order to be phenotyped because of the presence of a secondary population on dot-plots (See Chapter 3 "Results: Extra Populations"). To maintain consistency, all nine isolates were phenotyped using the post-invasion trypsin step, although seven of these samples were also phenotyped without post-invasion trypsin treatment (Appendix Figs. A1 & A2). One sample (278) could not be re-cultured after being frozen and was therefore phenotyped using SYBR Green I as a DNA dye and did not undergo post-invasion trypsin treatment. All isolates were phenotyped on two independent occasions and each assay condition was performed in triplicate in each experiment and the mean parasitaemias for each enzyme treatment were determined (Fig. 3.3).

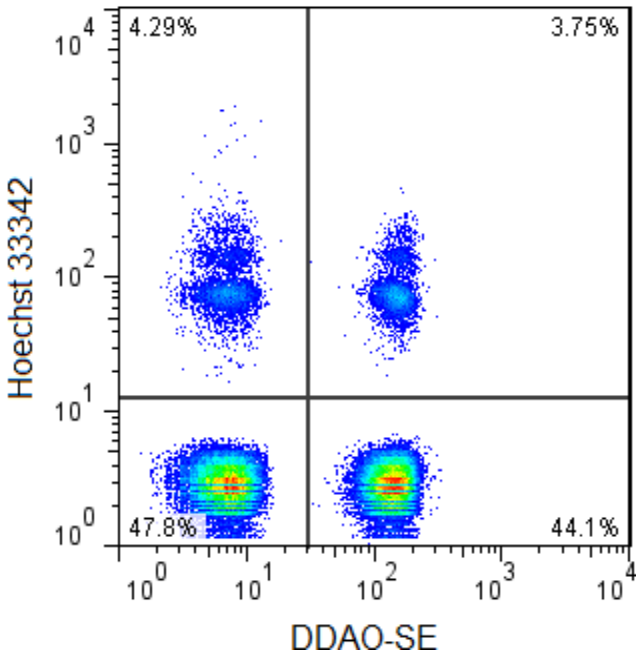


Fig. 3.1. Dot plot of an invasion assay. The plot indicates invasion of strain 3D7 into untreated erythrocytes. The x-axis is DDAO-SE cell labelling and the y-axis is Hoechst 33342 parasite DNA labelling. The four populations (clockwise from bottom right) are: target uninfected cells, donor uninfected cells, donor infected cells and target infected cells. Target infected cells (top right), is the population that varies with enzyme treatment and determines the phenotype.

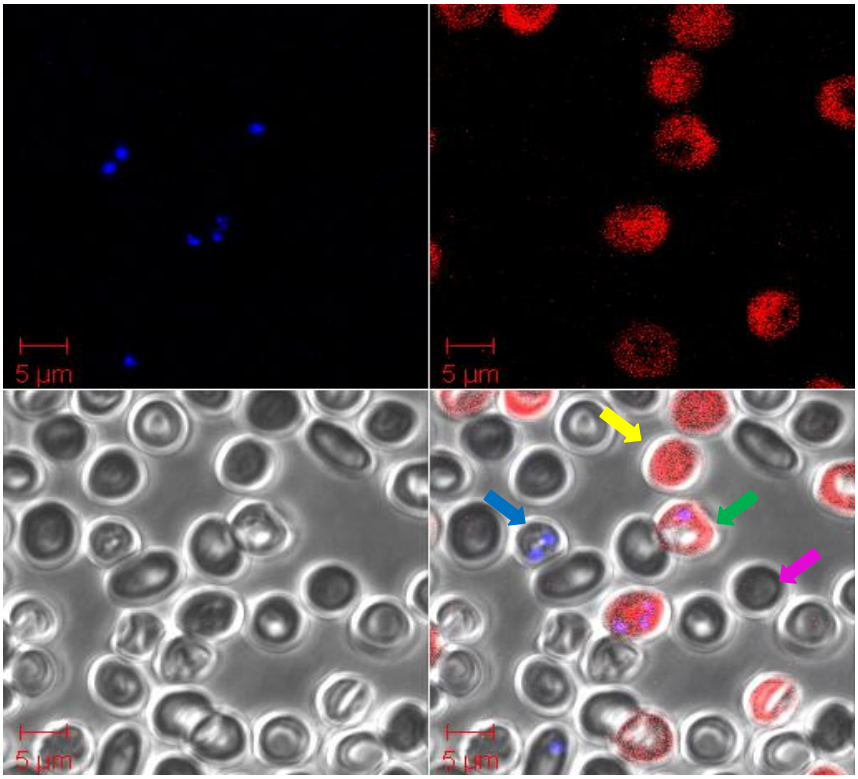


Fig. 3.2. Confocal fluorescence microscopy of erythrocytes from the same well as the dot plot above. The four panels are of the same field with different fluorescence excitation. Clockwise (from bottom right): All fields merged; brightfield; violet laser – Hoechst 33342 parasite labelling; red laser – DDAO-SE cell labelling. Blue arrow: infected, donor erythrocyte. Yellow arrow: uninfected, target erythrocyte. Green arrow: infected, target erythrocyte. Pink arrow: uninfected, donor erythrocyte.

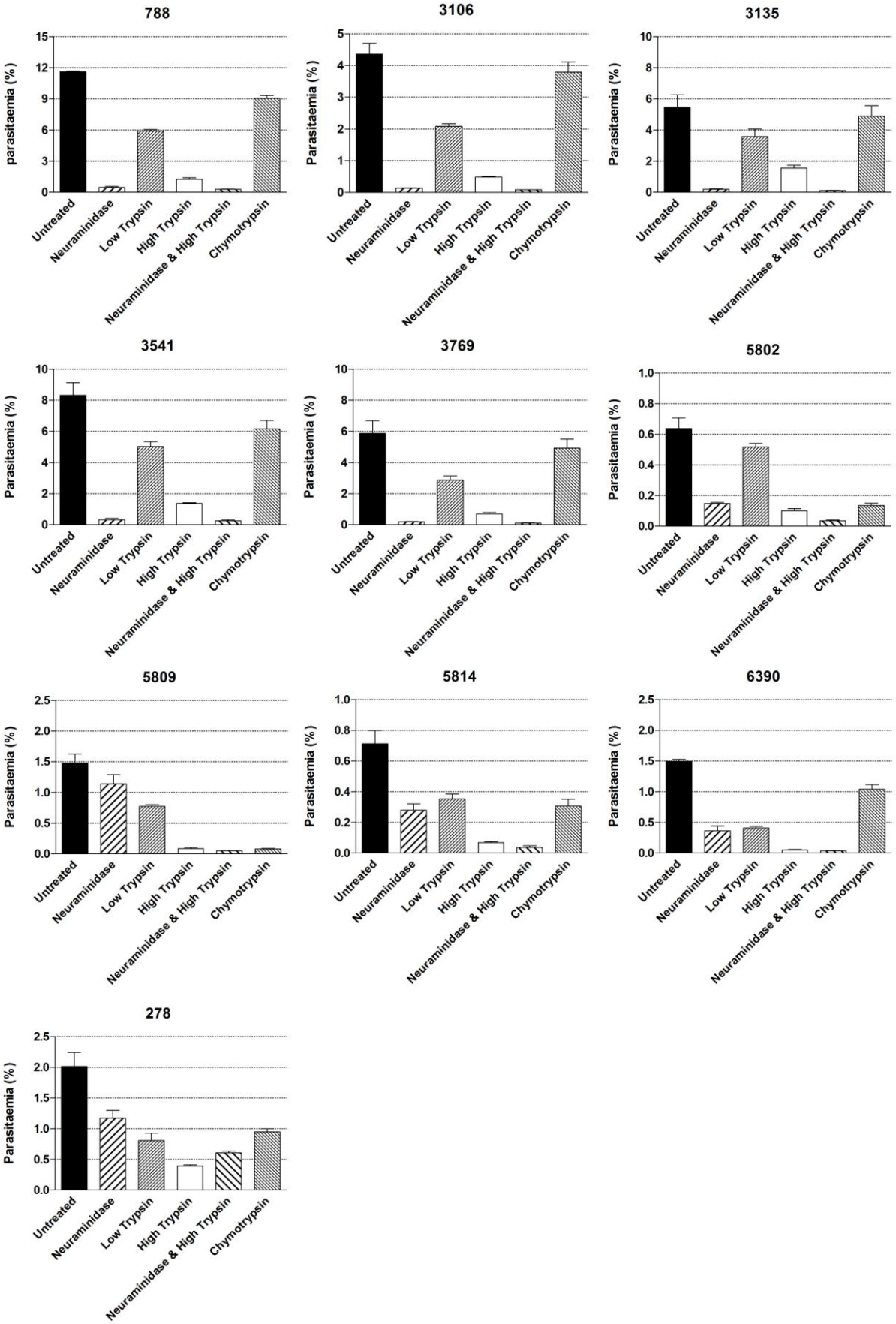


Fig. 3.3. Parasitaemia bar graphs of each phenotyping invasion assay. The bars represent the mean parasitaemia calculated from two experiments of three replicates. Error bars are SEMs.

From the parasitaemias measured in labelled erythrocytes (Fig. 3.3), mean invasion efficiencies for all 10 Peruvian isolates were calculated by expressing the parasitaemia in labelled erythrocytes that underwent each enzyme treatment as a percentage of the parasitaemia in labelled untreated erythrocytes (Fig. 3.4). For example, if an enzyme treatment causes a 20% decrease in parasitaemia with respect to untreated erythrocytes, the invasion efficiency would be 80%.

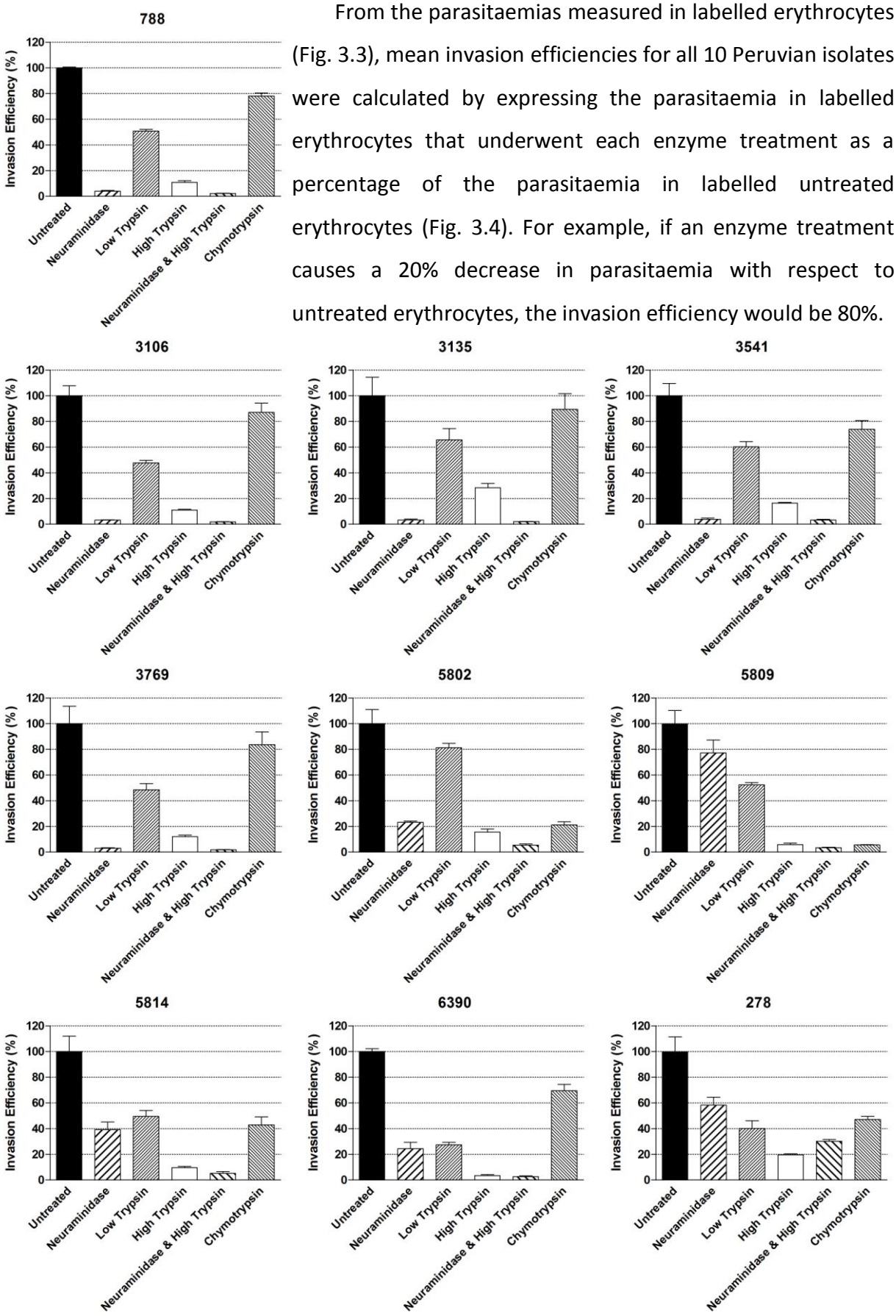


Fig. 3.4. Invasion phenotypes of Peruvian field isolates. Error bars are SEMs.

Clustering of Invasion Phenotypes

Comparison of the invasion profiles of all ten isolates revealed two clusters of phenotypes. This can be seen when invasion profiles are ordered according to their neuraminidase sensitivity (Fig. 3.5). Five isolates exhibited very high sensitivity to neuraminidase and high trypsin but were chymotrypsin-resistant. The isolates of this cluster were 788, 3106, 3135, 3541 and 3769 and are referred to hereafter as Type I phenotypes. The remaining isolates, 278, 5802, 5809, 5814 and 6390, were not grouped by specific criteria but lacked the characteristics of the Type I cluster and were therefore classified as having a Type II phenotype. The differences between the two clusters are summarised in Table 3.2.

	Phenotype Cluster % mean (SEM)	
Enzyme Treatment	**Type I**	**Type II**
Neuraminidase	3.5 (0.2)	44.5 (10.4)
Low Trypsin	54.7 (3.6)	50.3 (8.9)
High Trypsin	15.7 (3.3)	10.8 (3.0)
Neuraminidase & High Trypsin	2.3 (0.3)	9.4 (5.2)
Chymotrypsin	82.5 (2.9)	32.0 (10.5)

Table 3.2. Mean invasion efficiencies for isolates of each phenotype cluster. Standard error of the mean (SEM) is in brackets. All figures are given as percentages (%).

Greater heterogeneity existed in Type II phenotypes than Type I, with no two isolates having identical phenotypes. Within Type II parasites, similarities could be drawn between the invasion profiles of 5802, 5814 and 278, with neuraminidase and chymotrypsin sensitivity appearing to be linked. However, 6390 had high sensitivity to neuraminidase but was relatively chymotrypsin-resistant and of the Type II phenotypes it bore the most resemblance to a Type I profile. Conversely, 5809 had high resistance to neuraminidase treatment but was highly sensitive to chymotrypsin. 278 had unusually high resistance to the combined treatment of neuraminidase and high trypsin, with mean invasion efficiency being 30.2% (SEM 1.4%). The greater variation found amongst Type II phenotypes is evident from the greater SEM values (except for high trypsin treatment: Table 3.2). In general, this cluster showed greater resistance to neuraminidase than Type I isolates with 5802 having the lowest invasion efficiency into neuraminidase-treated erythrocytes (23.2%, SEM 1.0%). As with Type I, high sensitivity to high trypsin treatment was a feature of all Type II phenotypes but, with the exception of 6390, Type II isolates had mean invasion efficiencies under 50% into chymotrypsin-treated erythrocytes. The variation in invasion efficiency between individual isolates is illustrated in Fig. 3.5.

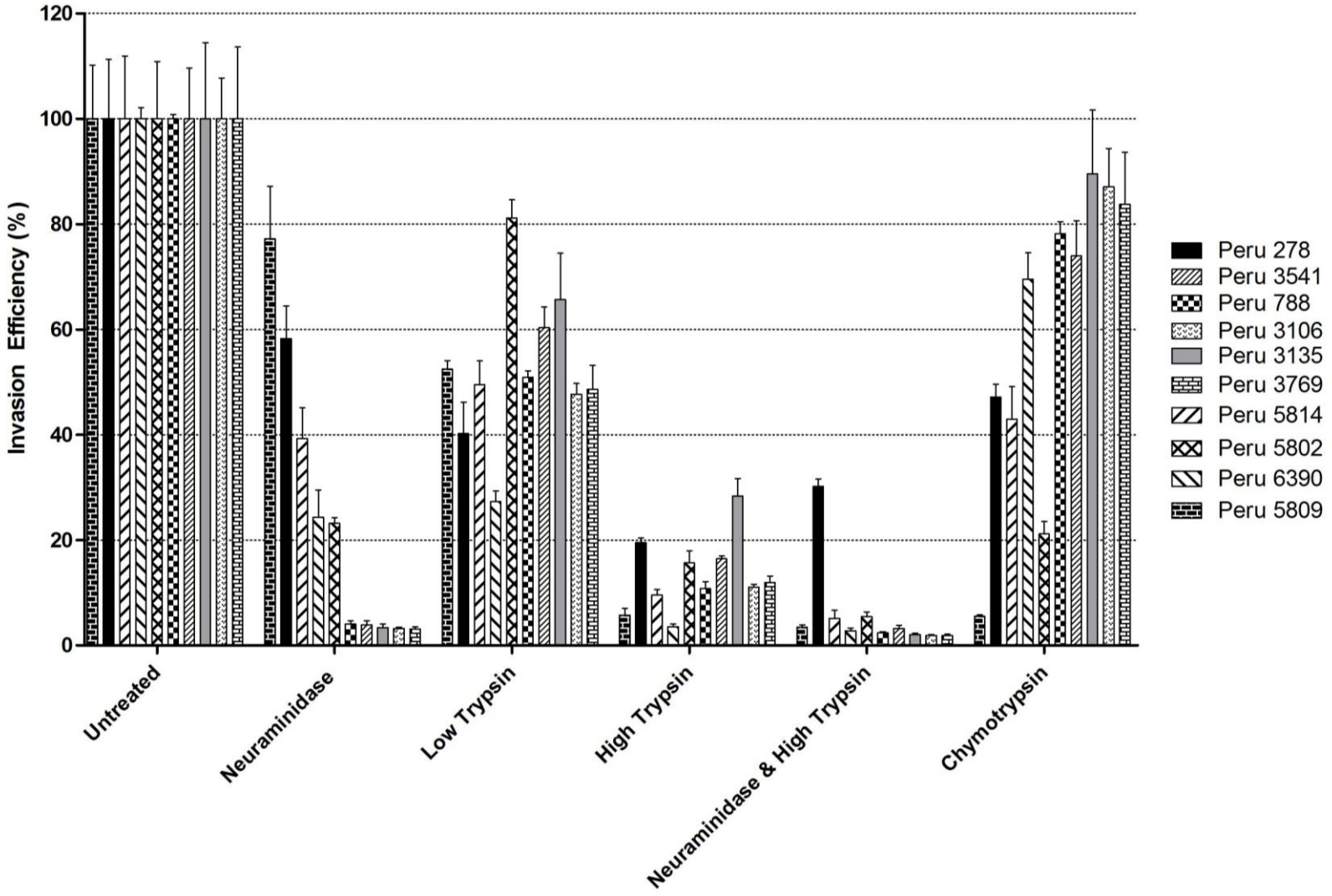


Fig. 3.5. Invasion profiles of all Peruvian isolates. Strains are ordered by decreasing neuraminidase sensitivity, so that the first five isolates for each enzyme treatment correspond to the Type II phenotypes.

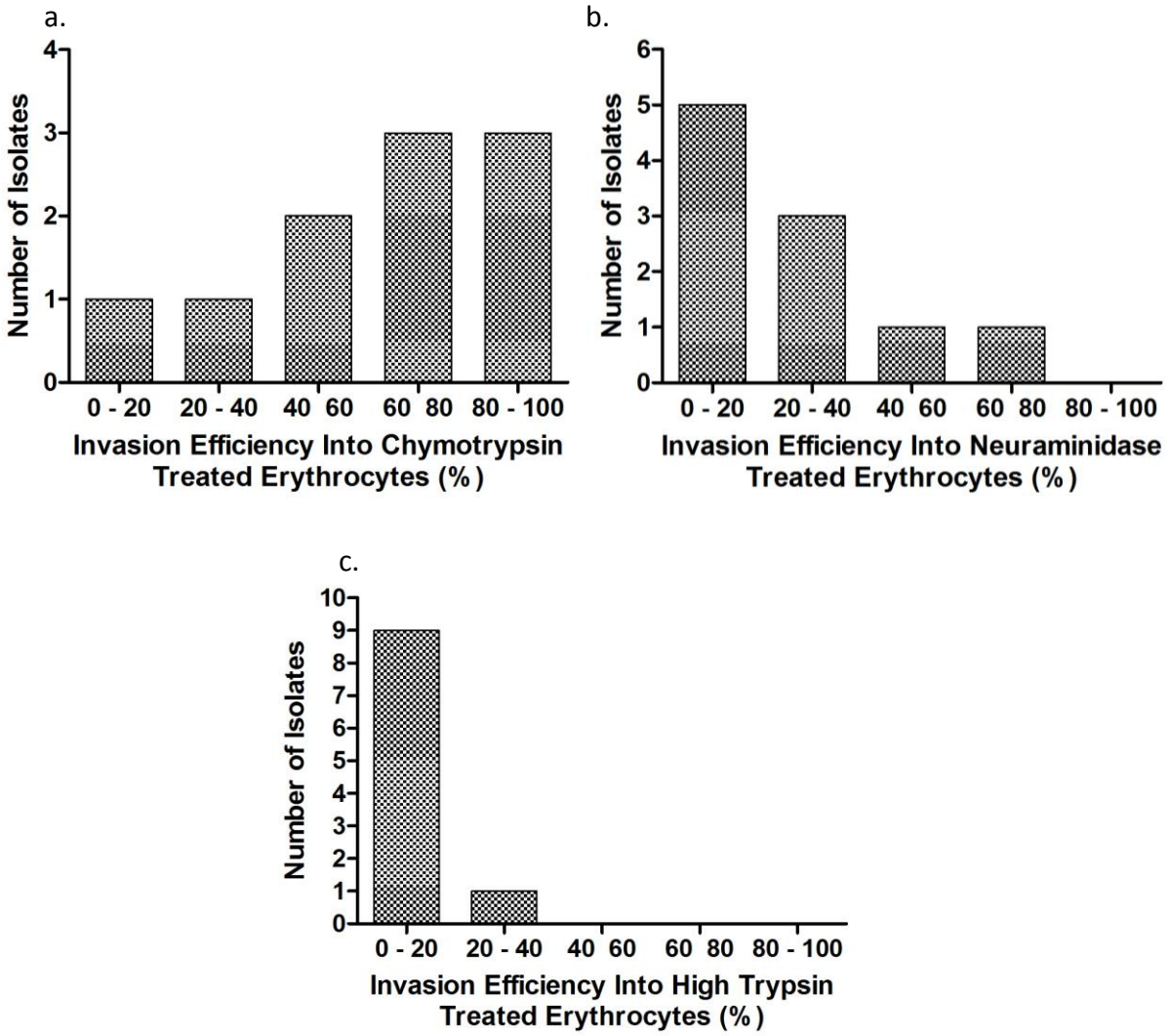


Fig. 3.6. Invasion efficiency into enzyme treated cells. (a) chymotrypsin, (b) neuraminidase, (c) high trypsin.

Across all ten Peruvian isolates invasion efficiency into chymotrypsin-treated cells showed the greatest diversity, ranging from 7.3% (5809) – 89.5% (3135) and also the most even distribution of values (Fig. 3.6a). Inhibition of invasion by neuraminidase treatment resulted in a range of invasion efficiencies from 3.1% (3769) to 70.9% (5809) (Fig. 3.6b). However, despite the large range, the majority of isolates had invasion efficiencies <25% into neuraminidase-treated cells and all isolates with Type I invasion profiles fell into this category. The least variation amongst isolates was seen when invading high trypsin-treated erythrocytes, all but one isolate (3135 – 28.4%) having an invasion efficiency under 20% (Fig. 3.6c).

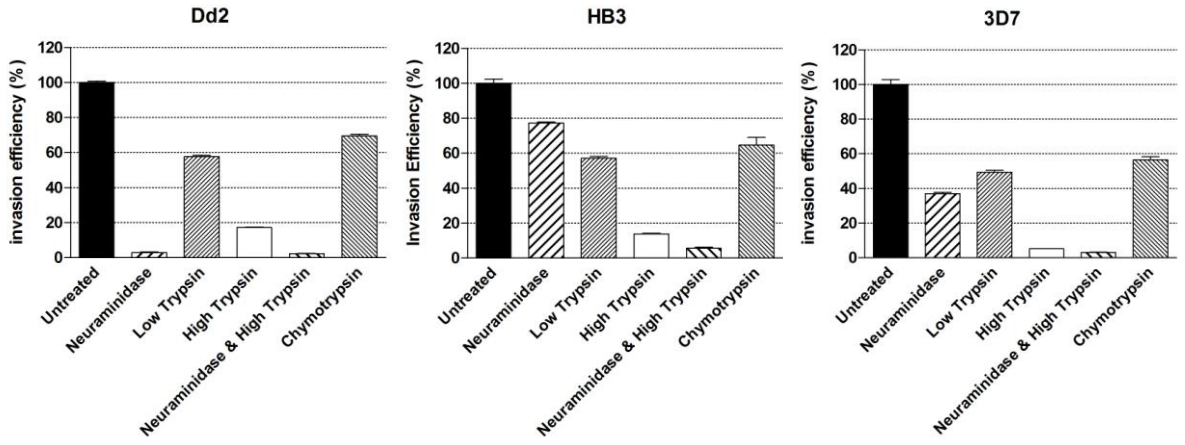


Fig. 3.7. Invasion profiles of lab strains Dd2, HB3 and 3D7. Error bars are SEMs.

Comparing the invasion profiles of Peruvian field isolates to lab strains (Fig. 3.7) illustrates that the Type I phenotypes are very similar to that of Dd2, a strain that originates from South-East Asia. The invasion profiles of Type II isolates are clearly different as they are not completely dependent on neuraminidase-sensitive pathways. 5814 and 5802 are comparable in profile to 3D7 (a lab strain that originates from Africa), while 278 appears more similar to HB3 (a lab strain that originates from Honduras). The invasion efficiency of 6390 into neuraminidase-treated cells is low, falling between the sensitivities of 3D7 and Dd2, while its ability to invade chymotrypsin-sensitive cells is high, similar to Dd2 and HB3. 5809 exhibits neuraminidase sensitivity comparable to HB3, but is unique in its relative inability to invade chymotrypsin-treated cells. This range of phenotypes within Type II isolates emphasises the fact that erythrocyte invasion is a very variable phenotype, even in a region of low malaria transmission and low *P. falciparum* genetic diversity such as the Peruvian Amazon.

Correlations within Invasion Phenotypes and Between Invasion and Multiplication Rates

In addition to invasion profiling, parasite multiplication rates (PMRs) were calculated for each isolate during phenotyping by dividing the final parasitaemia in untreated erythrocytes by the starting parasitaemia of 0.75% (as detailed in Chapter 2: “Materials & Methods”). The final parasitaemia was determined from Hoechst 33342 or SYBR Green I positive events in both donor and target erythrocytes (i.e. the top two populations of the

dot plot in Fig. 3.1), and assumes that all parasites detected in the donor cells are new invasions (Fig. 3.8).

The PMR for Peruvian strains fall into the same clusters produced by invasion profiling. For isolates with Type II invasion profiles, PMR values were uniformly low with a mean of 2.6 (SEM 0.4). Isolates with a Type I invasion phenotype had higher but more variable PMRs with a mean of 10.2 (SEM 1.7). Peruvian isolate 788 had a particularly high PMR of 16.1, and both 788 and 3541 exceeded the PMR of Dd2, the highest of the three lab strains. HB3 had the lowest PMR of a lab strain at 6.8, and the lowest PMR found in a field isolate with a Type I invasion profile was 6.6 in isolate 3106.

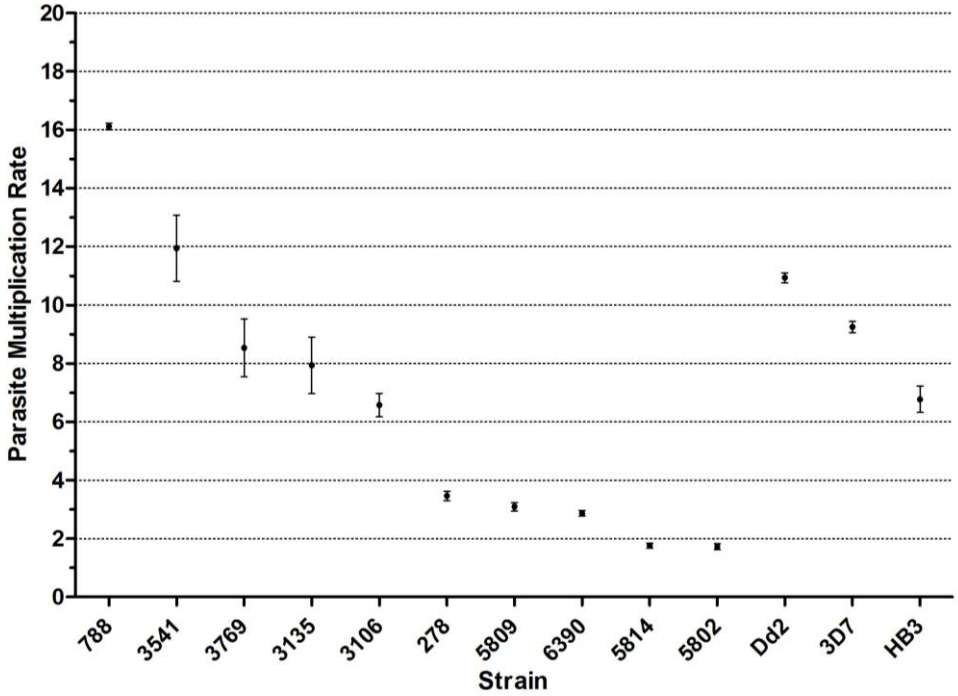
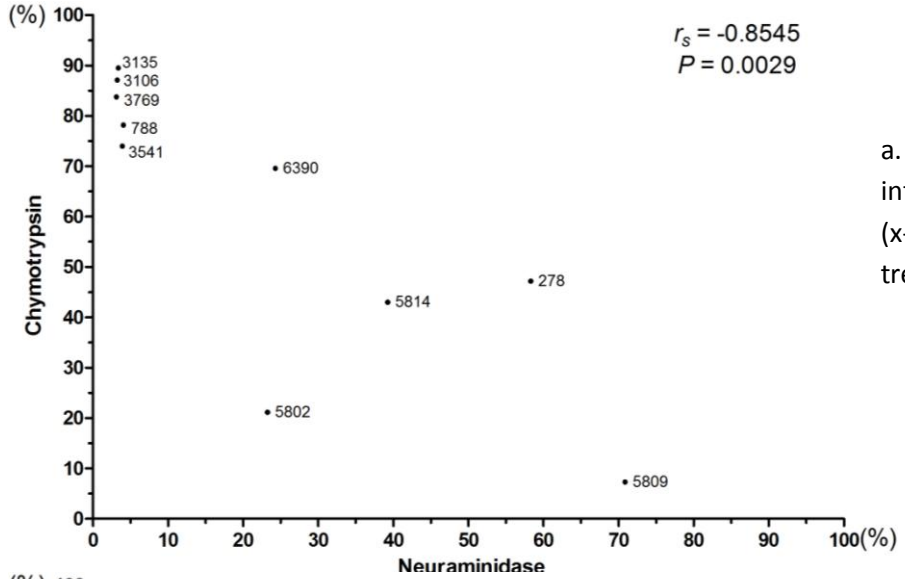


Fig. 3.8. Parasite Multiplication Rates (PMRs) found in Peruvian field strains and three lab strains, Dd2, 3D7 and HB3. PMRs fall into the same clusters as seen in invasion profiles with isolates that have Type I invasion profiles also having PMRs of above 6, while Type II invasion profiles correlate to a PMR below 4.

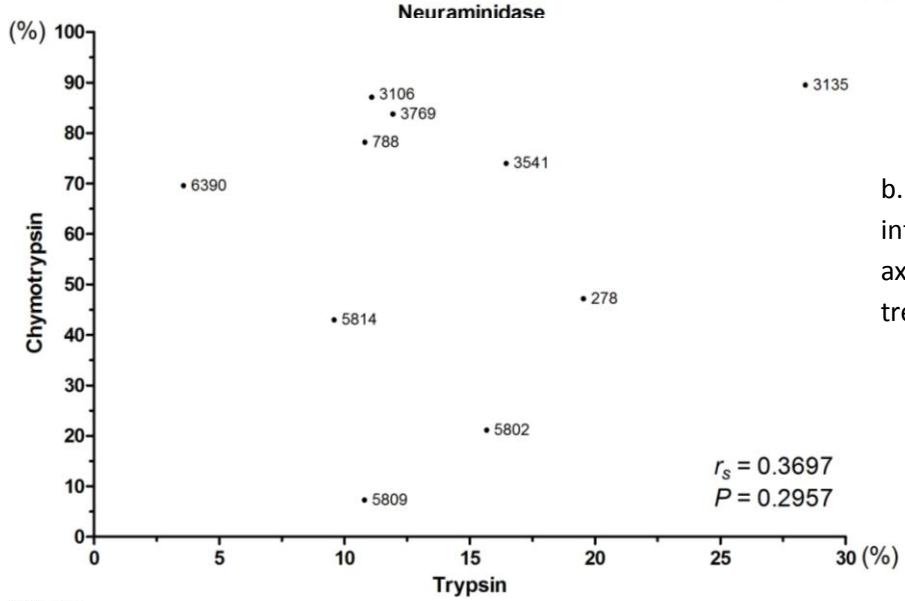
Invasion efficiencies of the ten isolates were compared for correlation between different enzyme treatments (Fig. 3.9). No correlation was found between invasion into neuraminidase and high trypsin-treated cells (Rs = -0.35, *P* = 0.33; Fig. 3.9c), or invasion into high trypsin-treated and chymotrypsin-treated cells (Rs = 0.37, *P* = 0.30; Fig. 3.9b). However, invasion efficiency into neuraminidase- and chymotrypsin-treated erythrocytes had a strong negative correlation (Rs = -0.85, *P* = 0.0029; Fig. 3.9a).

Given the fact that Type I and II isolates had substantially different mean PMRs and the two clusters differed in their sensitivity to neuraminidase treatment, the PMR for each isolate was also compared to invasion efficiencies into enzyme-treated erythrocytes (Fig. 3.10). Positive correlation was found between PMR and invasion efficiency into chymotrypsin-treated erythrocytes (r_s = 0.73, P = 0.02; Fig. 3.10a). A positive correlation was also found between invasion into neuraminidase-treated erythrocytes and PMR, but it was not statistically significant (r_s = 0.61, P = 0.07; Fig. 3.10b). No correlation was found between invasion into trypsin-treated erythrocytes and PMR (r_s = 0.32, P = 0.37; Fig. 3.10c).

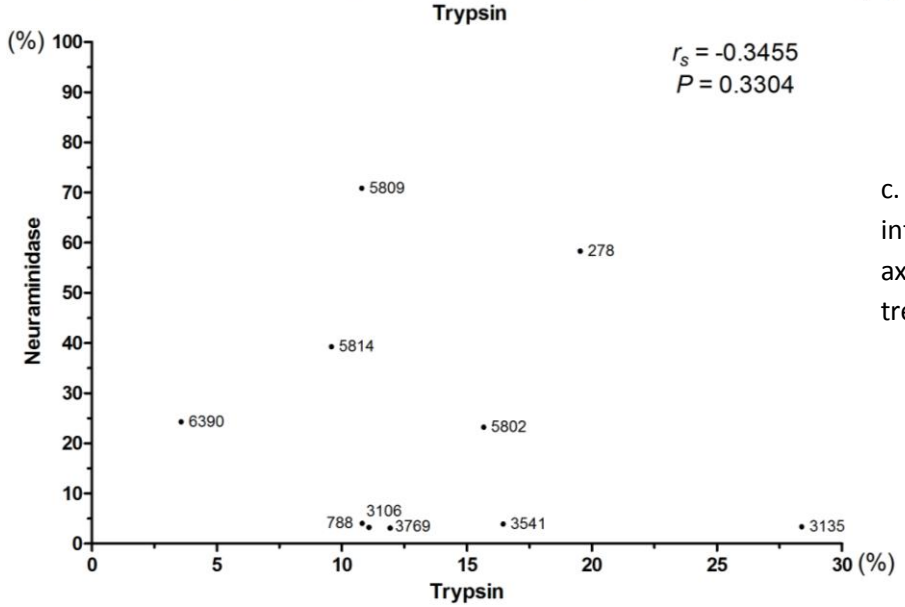
The phenotyping data therefore indicate that the Peruvian isolates fall into two distinct clusters. Type I isolates have variable but generally high PMR and invasion of these isolates is strongly and uniformly inhibited by neuraminidase treatment but is resistant to chymotrypsin treatment. Type II isolates have uniformly low PMR but more variable invasion phenotypes, which are generally characterised by some level of resistance to trypsin treatment. In order to establish whether a genetic basis for these differences could be identified, genomic DNA from all isolates was submitted for whole-genome re-sequencing using an Illumina IIx Genome Analyzer.



a. Invasion efficiencies (%) into neuraminidase treated (x-axis) and chymotrypsin treated (y-axis) erythrocytes.

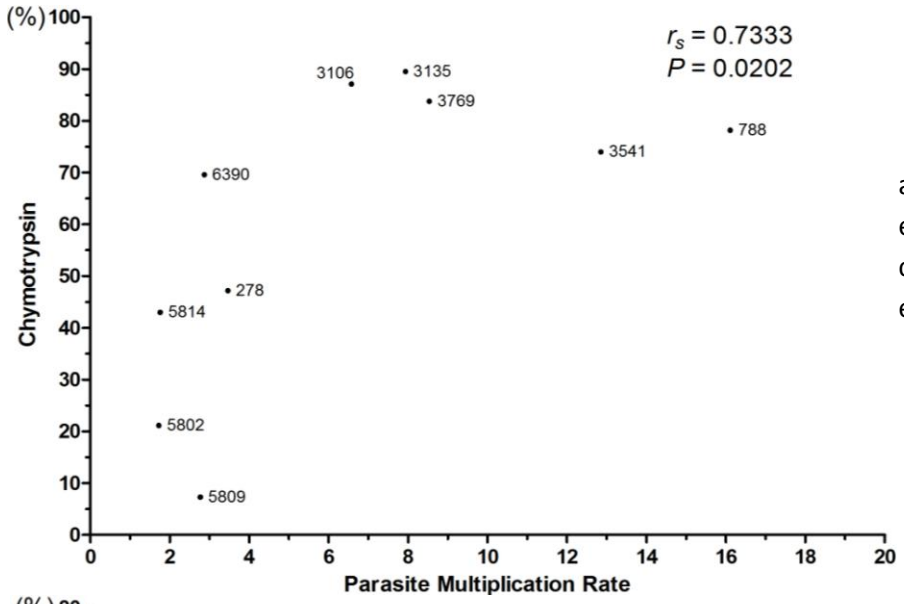


b. Invasion efficiencies (%) into high trypsin treated (x-axis) and chymotrypsin treated (y-axis) erythrocytes.

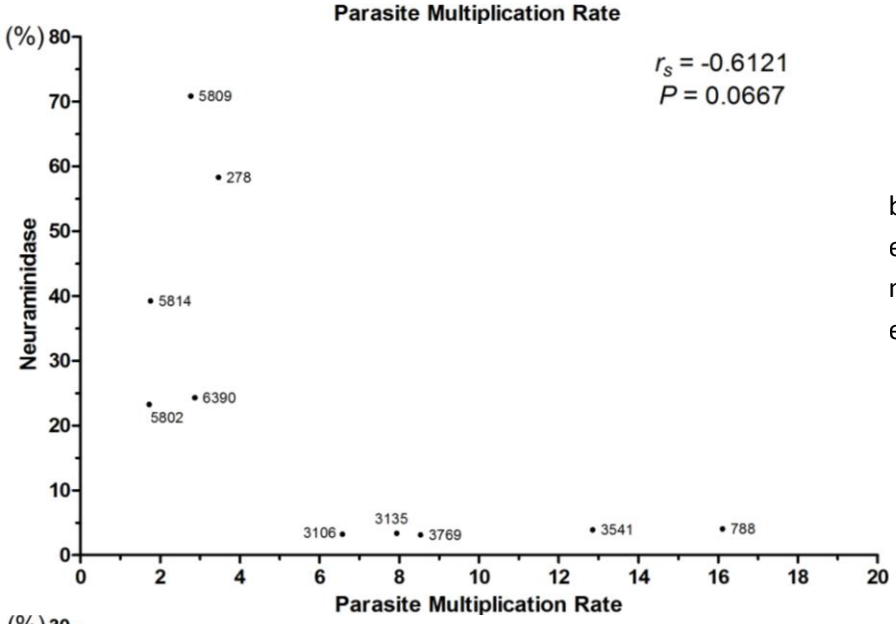


c. Invasion efficiencies (%) into high trypsin treated (x-axis) and neuraminidase treated (y-axis) erythrocytes.

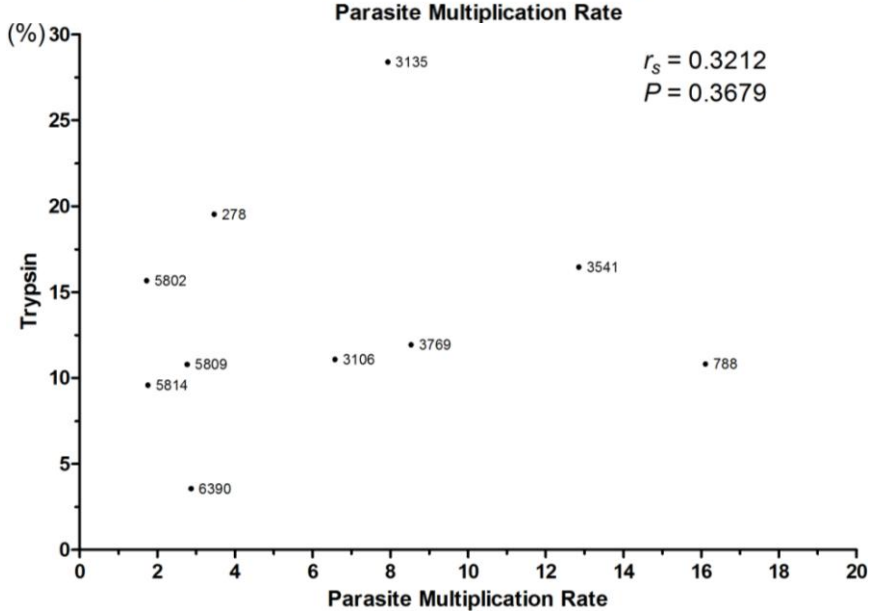
Fig. 3.9. Invasion efficiencies of Peruvian field isolates into enzyme treated erythrocytes. Spearman's rank correlation coefficient (r_s) and a two-tailed *P*-value are given for each plot. Correlation with a *P*-value of <0.05 was considered significant.



a. PMR (x-axis) and invasion efficiency (%) into chymotrypsin treated erythrocytes (y-axis).



b. PMR (x-axis) and invasion efficiency (%) into neuraminidase treated erythrocytes (y-axis).



c. PMR (x-axis) and invasion efficiency (%) into high trypsin treated erythrocytes (y-axis).

Fig. 3.10. Invasion efficiencies into enzyme treated erythrocytes and PMR of Peruvian field isolates. Spearman's rank correlation coefficient (r_s) and a two-tailed *P*-value are given for each plot. Correlation with a *P*-value of <0.05 was considered significant.

Genotyping

Analysis of the samples was carried out using MapSeq – a web browser-based tool for comparison of samples (Fig. 3.11), LookSeq – a web browser-based interface for visualising Illumina reads (Fig. 3.12), and Mutabo! – a tool for correlating non-synonymous SNPs to corresponding amino acid changes. All three tools were developed by Dominic Kwiatkowski's team at WTSI.

Fig. 3.11. The MapSeq interface. Genotyping view allows the user to select samples and genome region and identifies the SNPs present with respect to the 3D7 reference sequence. The "compare groups" function creates a list of mutations that are present across a cohort of samples. Analysing populations uses principal component analysis (PCA) to compare different samples.

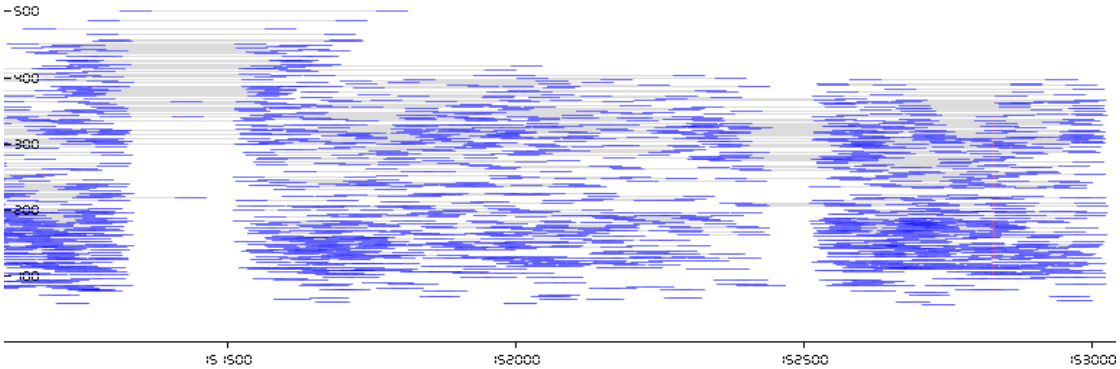


Fig. 3.12. The LookSeq genome browser: a short segment of chromosome 1 of Peruvian isolate 3135. Base number is plotted on the x-axis, and paired read length is on the y-axis. Reads are in blue, the darker blue indicates greater coverage of that area. The grey strips indicate the sequence between paired reads. Red areas signify SNPs called with respect to the 3D7 reference sequence (far right). On the left is an insertion / deletion characterised by the gap in reads and the increase in paired read length either side.

All eleven of the cultured Peruvian field isolates were sequenced using the Illumina platform. Of the eleven isolates, only seven had undergone genome assembly and been analysed at the time of thesis submission, the remaining four are still in the sequencing pipeline. The seven isolates to be analysed include all those with Type I invasion profiles (788, 3106, 3135, 3541 and 3769), one Type II isolate (6390), and one isolate that was not phenotyped because it could not be re-cultured after freezing (9050). 278, 5802, 5809 and 5814 are still in the sequencing pipeline.

Principal component analysis (PCA) was used to compare SNPs in Peruvian isolates to previously sequenced samples from other isolates (Fig. 3.13). PCA is useful for identifying patterns in data and displaying the results to highlight their similarities and differences, and is particularly useful when there is a large number of variables. By transforming data values to new sets of variables, the principal components, the dimensionality of the data set is reduced but the majority of the variability is retained (Jolliffe 2002). The majority of variability is summarised in the principal component (x-axis), while remaining variability is assessed in the second principal component (y-axis). In the context of *Plasmodium* genotyping, samples can be analysed using two principal components. Isolates will group together if they are genetically similar and will be separated if they are genetically diverse. This allows a large number of samples to be compared for similarities quickly and simply.

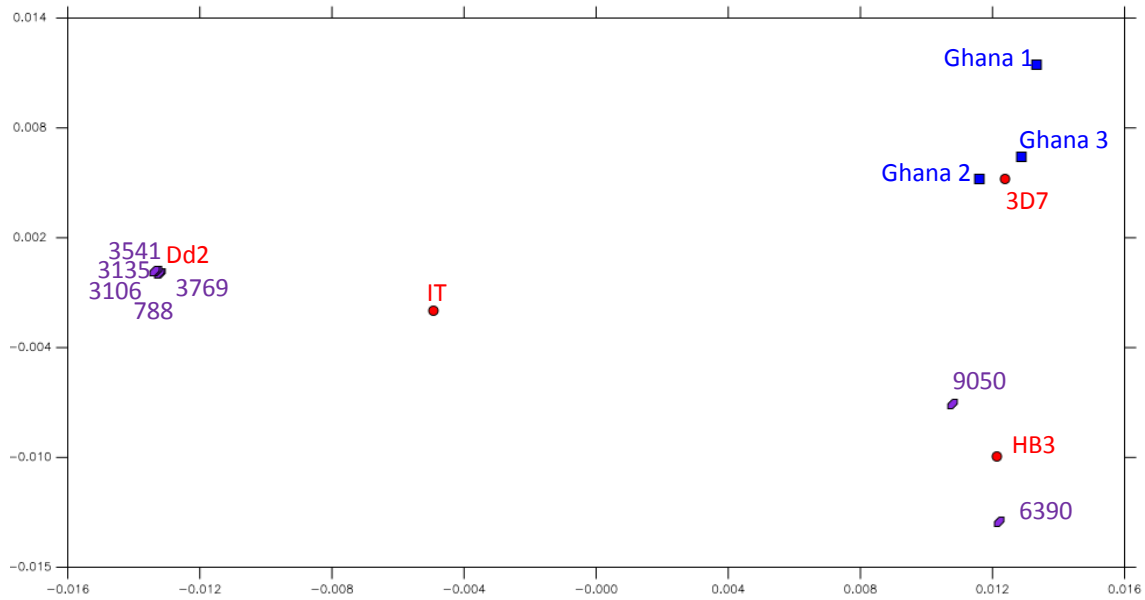


Fig. 3.13. PCA plot of all sequenced *Plasmodium* samples. Peru strains are purple hexagons, red circles are lab strains and blue squares are field isolates from Ghana. The plot is generated from PCA analysis of all SNPs found in the genome, with the minimum of ten reads for a SNP call to be made.

The PCA plot (Fig. 3.13) shows that two of the Peruvian isolates are genetically similar to the lab strain HB3 originally from Honduras. The remaining five Peruvian isolates that all had Type I invasion profiles are genetically very similar to Dd2 (Fig. 3.13 / Fig. 3.14). The field isolates from Ghana group with 3D7, a lab strain thought to originate from Africa. IT, another lab strain originating from Brazil, does not appear to be genetically similar to the other isolates studied here.

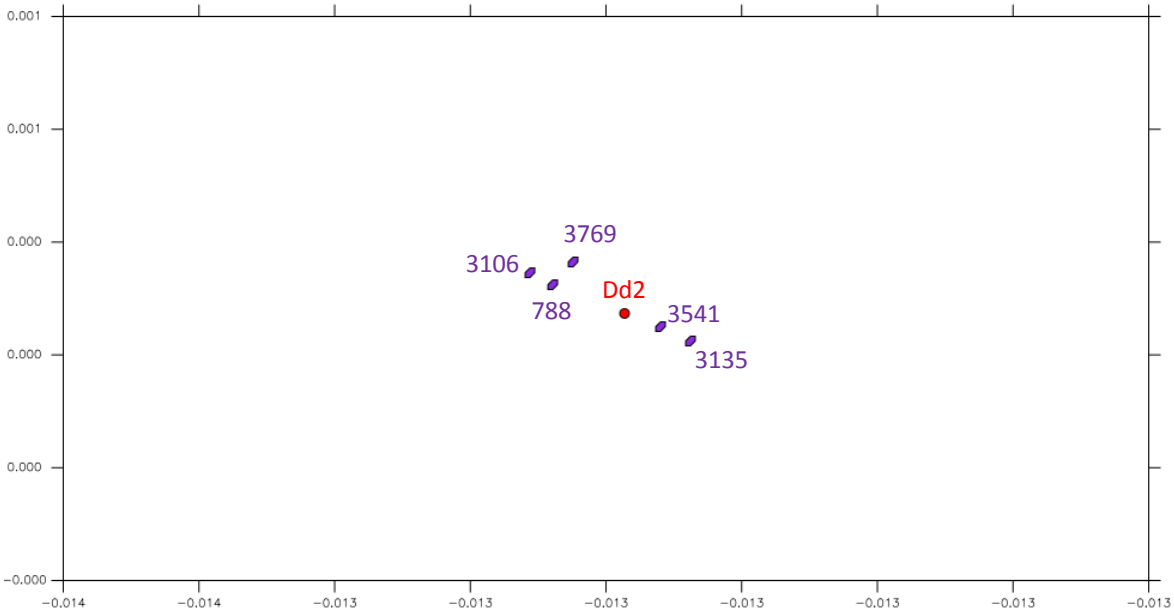


Fig. 3.14. Peruvian isolates with type one invasion profiles are closely related to Dd2. This is a magnified view of the group on the extreme left of Fig. 3.13. As can be seen from the scales, there is virtually no variation between these isolates. Peruvian isolates are represented as purple hexagons, while Dd2 is a red circle.

The genomic similarities between Peruvian isolates with Type I phenotypes and Dd2 were consistent with the similarities seen in phenotype. However, two features of the PCA plot were unusual. Firstly, the genetic distance between the Type I isolates and Dd2 was significantly lower than the distance between any two of the 400 *P. falciparum* isolates that have now been sequenced at WTSI (Kwiatkowski et al. unpublished data). Secondly, at the scale of PCA shown in Fig. 3.13, *P. falciparum* isolates always group by geographic origin. The fact that Type I Peruvian isolates cluster with a South Asian strain, while two other Peruvian isolates cluster with a strain from Honduras is unusual. The simplest explanation of the data is that the Type I Peruvian isolates were contaminated with a Dd2 like strain.

To test this hypothesis, all isolates that had Type I invasion profiles were tested for *msp*-1 and *msp*-2 genotypes. MSP-1 and MSP-2 are encoded by highly variable genes (*msp1* and *msp2* respectively) that are routinely used for genotyping studies (Baum et al. 2003; Lobo et al. 2004; Deans et al. 2007). The *msp1* block 2 and *msp2* block 3 loci containing highly polymorphic repeat sequences were amplified by PCR and the sizes of DNA fragments produced were analysed by gel electrophoresis (Table 3.3).

	Fragment size (bp)	
Isolate	***msp1***	***msp2***
3D7	472	748
Dd2	430	515
W2	436	513
3135	431	513
3541	435	510
788	430	518
3769	436	514
3106	434	514
9050	455	854
6390	469	790
278	467	776
5802	428	680
5809	454	848

Table 3.3. Fragment sizes of polymorphic loci from *msp1* and *msp*2. Highlighted in red are fragment sizes from Peruvian isolates that had loci that were of very similar size to lab isolates Dd2 or W2.

All Type I isolates that were found to be genetically similar to Dd2 by PCA had polymorphic fragment sizes from *msp1* block 2 and *msp2* block 3 that were homologous to Dd2 (highlighted in red in Table 3.3). In contrast, Type II isolates had clearly different *msp1* and *msp2* genotypes; although the fragment of *msp1* amplified from Peruvian isolate 5802 was similar to Dd2 the *msp2* fragment was very different.

From the identical phenotypes, PCA and polymorphic fragment sizes it was clear that Peruvian isolates 3135, 3541, 788, 3769 and 3106 were contaminated. While Dd2 had never been grown in Peru, its parental lab strain W2 had been in culture in Iquitos on multiple occasions as a positive control for the culture procedure. Checking the culture records at UNAP established that W2 was in culture at the times when the Type I isolates had been

frozen. Conversely, when the non-contaminated isolates (278, 5802, 5809, 5814 and 6390) had been in culture, W2 was not being grown.

Non-Synonymous SNPs and Amino Acid Changes in Peruvian Isolate 6390

Of the non-contaminated samples only 6390 had been genotyped at the time of thesis submission; sequences for the remaining isolates are expected within the next month. For this isolate, non-synonymous SNPs were identified across 62 genes identified as having a role in invasion, and the consequent amino acid changes were determined. The data are summarised below and are presented in full in Appendix Table A2.

98 non-synonymous SNPs were identified in these 62 genes. The mean number of reads for non-synonymous SNP calls was 86, but five SNPs were called from fewer than ten reads resulting in a lower confidence in the calls. The 93 high confidence non-synonymous SNPs code for 80 amino acid changes, the disparity between numbers of SNPs and amino acid changes being due to multiple SNPs being present in a single codon.

Of the genes known to act as invasion ligands, no SNPs were found in PfRh2a, PfRh2b, PfRh4, EBL-1 or EBA-181. However, a single SNP was found in each of EBA-140, EBA-175 and PfRh5. The greatest level of polymorphism that was observed in an invasion ligand was in PfRh1, with eight SNPs. MSP3.8, a protein of unknown function that is a member of the MSP3 gene family on chromosome 10 (Singh et al. 2009), had the largest number of SNPs (45), suggesting that it is highly polymorphic between isolates.

With more sequences available, correlation between ligand sequence and invasion phenotype will be possible, as discussed below.

RNA Extraction

RNA was extracted from all isolates that were phenotyped. The minimum quantity extracted from an isolate was 1.87 µg from 3106 (Appendix Table A1). As 10 µg is the current lower limit for RNASeq, transcriptomic analysis of these isolates awaits further technology improvement.

Extra Populations

Invasion phenotyping of two Peruvian isolates yielded an additional cell population not visible in any of the other samples. The invasion assays using the two-dye method routinely yielded four clearly separable populations observed by flow cytometry without using a post-invasion trypsin treatment (Figs. 3.1 & 3.15 a). However, in the case of Peruvian strains 3135 and 3769 extra populations with an intermediate level of DNA staining were observed (Fig. 3.15 b & c), which made parasite quantification difficult.

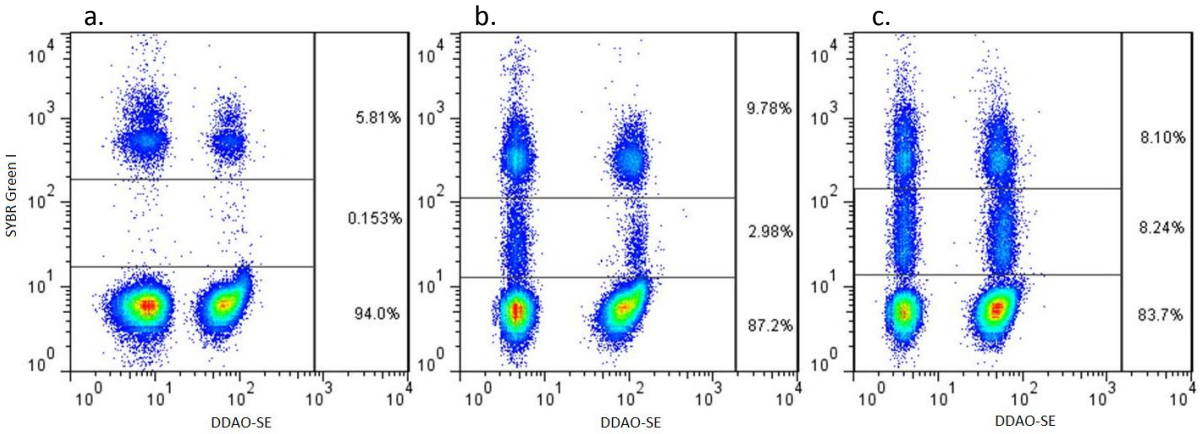


Fig. 3.15. Dot plots from Peruvian phenotyping assays of the invasion of erythrocytes that have not been enzyme treated. Staining: DDAO-SE (x-axis), SYBR Green I (y-axis). (a) 788; (b) 3135; (c) 3769. 788 produced four well separated populations as expected, also seen in Fig. 3.1. 3135 and 3769 have an extra, intermediate population between events that are SYBR Green I positive (parasitised) and SYBR Green I negative (uninfected).

For all strains previously studied, gates could be manually placed to count the events in each quadrant. With the four populations of 3135 and 3769 having a “smear” between them, accurate manual gating was impossible due to the diffuse boundary between events caused by uninfected erythrocytes and parasitised events. The extra populations were also highly consistent and always present in invasion assays using these two isolates, regardless of whether SYBR Green I or Hoechst 33342 was used to stain DNA.

This section deals with the two questions that were asked of the extra populations observed by flow cytometry: How can strains showing this phenomenon be accurately phenotyped, and what are the causes of this population?

Quantification of the Intermediate Population

Quantification of the intermediate population was compared across all Peruvian isolates by manually placing gates around each intermediate population (as shown in Fig. 3.15) for positive control invasion wells, where invasion is carried out using labelled untreated erythrocytes. The results are displayed in Fig. 3.16; points are the mean of three replicates, and error bars represent SEM.

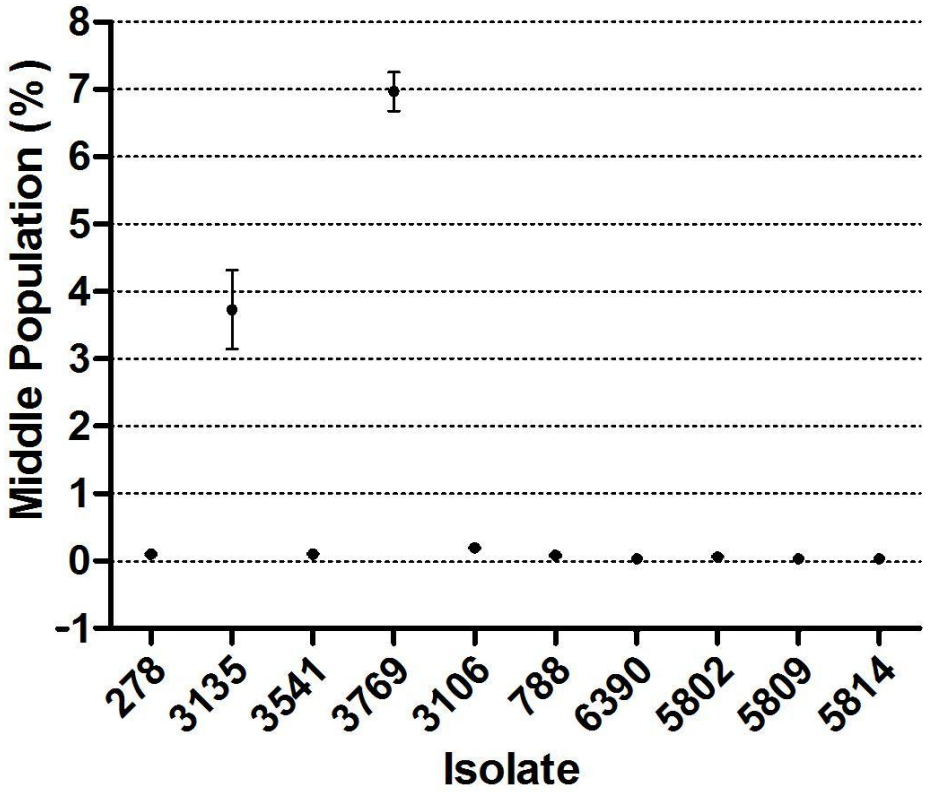


Fig. 3.16. Quantification of the intermediate population in Peruvian field isolates. By using manual gating the intermediate population is represented as a percentage of all events observed by flow cytometry.

Fig. 3.16 shows that only in two Peruvian isolates, 3135 & 3769, is the intermediate population of significant size. In 3135 the mean size of the intermediate population is 3.73% (SEM 0.59) of the sample; in 3769 it is 6.97% (SEM 0.29). In the other samples the intermediate population is negligible; 3106 is the largest of these, representing 0.20% (SEM 0.009) of the sample. In all samples other than 3135 and 3769 when the intermediate population is small the exact placement of the gate has an insignificant effect on the overall parasitaemia recorded (Fig. 3.15 a). However, for isolates 3135 and 3769 a small difference in gate placement could have a large effect on the recorded parasitaemia (Fig. 3.15 b & c).

The Intermediate Population: Parasite-based Events?

To determine whether the intermediate population is comprised of parasitised erythrocytes, an invasion assay was set up with two wells using unstained, untreated erythrocytes mixed with Peruvian isolate 3135. The method was the same as detailed in the "Phenotyping Invasion Assay" section of the "Methods", except the target erythrocytes were not stained and only invasion into erythrocytes that had not been enzyme-treated was studied.

After the 48 hour incubation period, one well was used to make a blood smear on a slide. This slide then underwent methanol fixation and Field's staining, using the standard protocol. The parasitaemia was determined by counting 1,000 erythrocytes and calculating the proportion of infected erythrocytes present by light microscopy.

The other well underwent fixation with 2% PFA / 0.2% GA and Hoechst 33342 staining using the standard protocol. An aliquot of the suspension was placed on a glass slide, air dried and fixed with methanol. The parasitaemia was determined by counting 1,000 erythrocytes and determining the proportion of Hoechst 33342 positive erythrocytes present by fluorescence microscopy. The parasitaemia was also determined by counting 100,000 events from the well using a BD LSRII flow cytometer. The gate placement for determination of parasitaemia by flow cytometry is shown in Fig. 3.17.

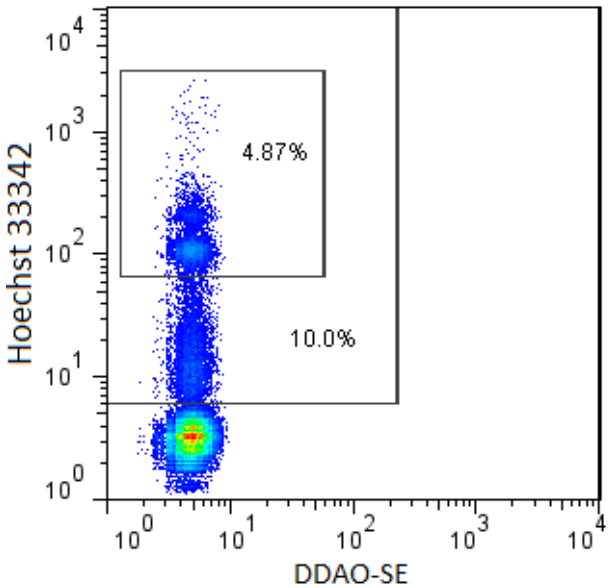


Fig. 3.17. Two different gate placements to determine parasitaemia in Peruvian isolate 3135. One gate was placed around both the intermediate population and the top population (parasitaemia 10.0%). The other gate was placed around only the top population (parasitaemia 4.87%).

Parasite staining	Viewing technique	Parasitaemia (%)
Field's staining	Light microscopy	4.48
Hoechst 33342	Fluorescence microscopy	5.43
Hoechst 33342	Flow cytometry (top population only)	4.87
Hoechst 33342	Flow cytometry (intermediate and top populations)	10.00

Table 3.4. The parasitaemia of Peruvian isolate 3135 determined using different methods.

The parasitaemia determined from the different methods is detailed in Table 3.4. Comparison of parasite counts using the three methods suggest that only the top gate contains events caused by intra-erythrocytic parasites. The parasitaemias determined by Field's staining, fluorescence microscopy and flow cytometry fell within 0.95% of each other when only the top population was defined as parasitic. In contrast, when the intermediate population was also included in the parasitaemia count by flow cytometry, the parasitaemia was 10.00%, significantly different to the microscopy-based counts. Further investigations of the intermediate population are discussed below.

Phenotyping Using Post-Invasion Enzymatic Treatments

As previously discussed, when parasites from isolates 3135 and 3769 invaded erythrocytes that had undergone no enzyme treatment, an extra population was present in both target and donor erythrocytes (Fig. 3.15 b & c). However, the intermediate population was not consistent across the phenotyping invasion assay (Fig. 3.18).

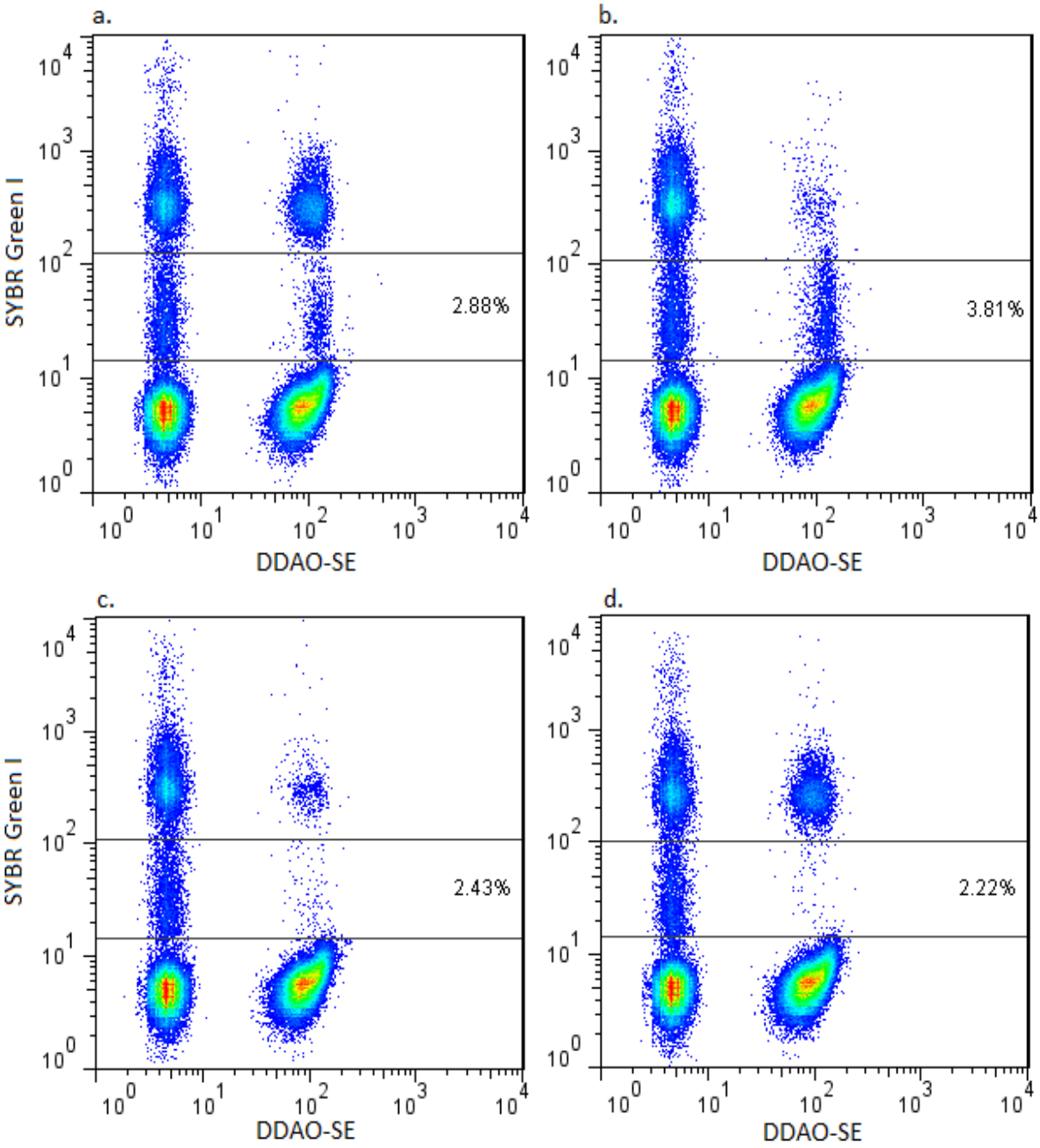


Fig. 3.18. Four flow cytometry dot plots of Peruvian isolate 3135. The target erythrocytes underwent different enzyme treatments: (a) no enzyme treatment; (b) neuraminidase treatment; (c) high trypsin treatment; (d) chymotrypsin treatment.

When the DDAO-SE-labelled, target erythrocytes were treated with trypsin or chymotrypsin the population was greatly reduced in labelled erythrocytes, but not in unlabelled erythrocytes (Fig. 3.18 c & d). However, neuraminidase treatment of target erythrocytes had no effect on the intermediate population (Fig. 3.18 b). This observation suggests that trypsin or chymotrypsin treatment of erythrocytes could eliminate the intermediate population.

To test this hypothesis an invasion assay was set up using unlabelled, untreated cells in the standard manner but, after 48 hours of *in vitro* culture, the cell pellet underwent one of four different treatments:

- Treatment A: Control. Cells resuspended in 100 µL PBS.
- Treatment B: 100 µL 20mU/mL neuraminidase was added to each well and the suspension mixed thoroughly.
- Treatment C: 100 µL 1 mg/mL trypsin was added to each well and the suspension mixed thoroughly.
- Treatment D: 100 µL 1 mg/mL chymotrypsin was added to each well and the suspension mixed thoroughly.

After individual treatments the plate was incubated for one hour at 37°C before all wells were washed with PBS and stained with Hoechst 33342 using the previously described method. Wells were acquired using a BD LSRII flow cytometer and 100,000 events were counted.

The effects of the post-invasion enzymatic treatments are shown in Fig. 3.19. Neuraminidase treatment (Fig. 3.19 b) had no impact on the intermediate population. However both trypsin and chymotrypsin removed the intermediate population while having no impact on the top population (Hoechst 33342 positive), believed to be the true parasitaemia.

Although chymotrypsin appeared just as effective as trypsin at removing the extra populations, trypsin was used for further experiments.

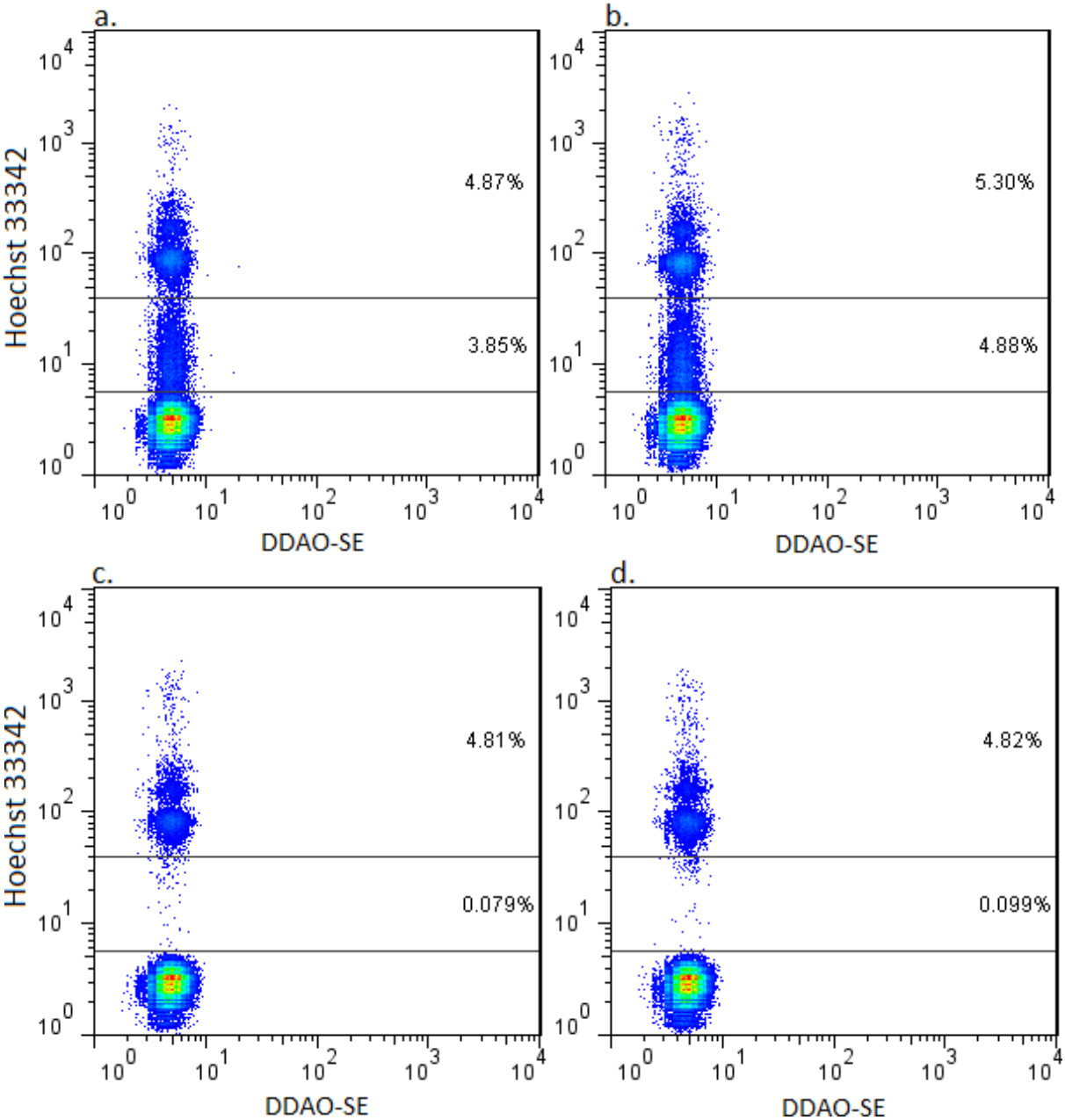


Fig. 3.19. The effects of post-invasion enzymatic treatments on the intermediate population.
(a) No enzyme treatment; (b) neuraminidase treatment; (c) trypsin treatment; (d) chymotrypsin treatment.

Validation of Post-Invasion Trypsin Treatment

Confirmation was required that the post-invasion trypsin treatment did not have a significant impact upon parasitaemia and could therefore be used in phenotyping invasion assays. 3D7 and Dd2, two lab strains that did not have any unexpected extra populations when phenotyped, were used for this purpose.

Invasion assays of 3D7 and Dd2 were set up and carried out exactly as detailed in Chapter 2 “Materials and Methods: Phenotyping Invasion Assay”, with one set of triplicate wells being trypsin-treated after the 48 hour incubation period, and the other set of triplicate wells not undergoing post-invasion treatment. After acquiring the data, the parasitaemia of the two sets of triplicate wells were compared. The experiment was repeated and the combined results of both experiments are shown in Figs. 3.20 & 3.21.

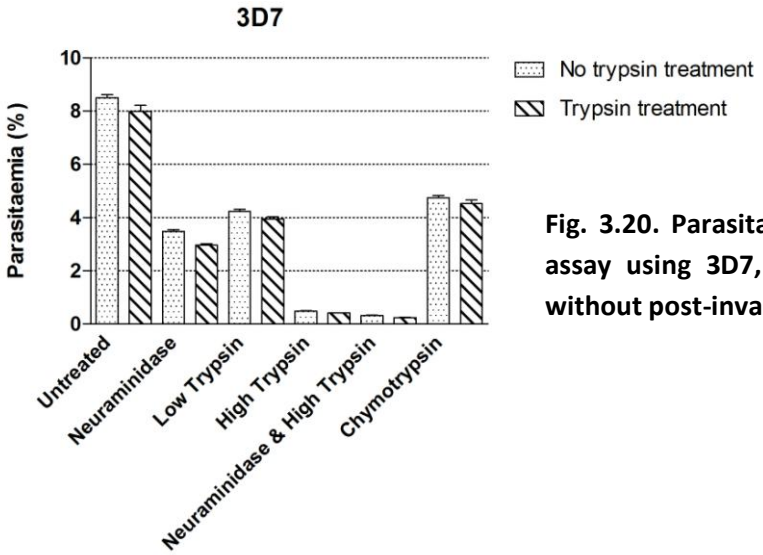


Fig. 3.20. Parasitaemia of a phenotyping invasion assay using 3D7, comparing replicates with and without post-invasion trypsin treatment.

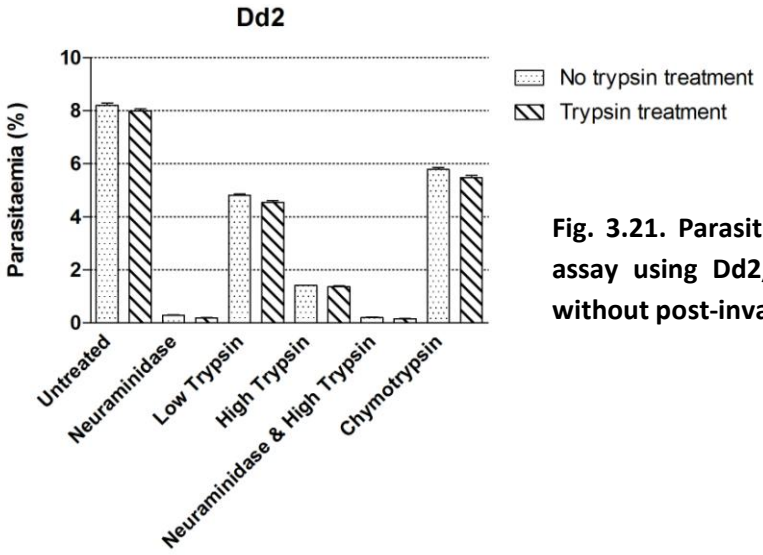


Fig. 3.21. Parasitaemia of a phenotyping invasion assay using Dd2, comparing replicates with and without post-invasion trypsin treatment.

Parasitaemia values for replicates undergoing post-invasion trypsin treatment were very similar to replicates that were untreated post-invasion. There was a slight decrease in mean parasitaemia in all wells where post-invasion trypsin treatment was carried out, ranging from 0.52 ± 0.10 (3D7 – neuraminidase) to 0.04 ± 0.02 (Dd2 – high trypsin and neuraminidase), perhaps due to lysis of protease-treated infected cells.

The confirmation that post-invasion trypsin treatment did not significantly affect the true parasitaemia meant that the two isolates that exhibited the extra population could now be accurately phenotyped. Resolution of this issue meant that all subsequent phenotyping assays were carried out as described in Chapter 2 "Materials and Methods: Phenotyping Invasion Assay", including post-invasion trypsin treatment and Hoechst 33342 staining. Isolates that had been phenotyped prior to this discovery were re-phenotyped using the new method. As discussed previously, phenotypes without post-invasion trypsin treatment do not differ significantly from those that include post-invasion trypsin treatment (Appendix Figs. A1 & A2). The only isolate that was not re-phenotyped using this method was 278, which could not be re-cultured after being frozen. Consequently, the values for 278 calculated in Figs. 3.3 & 3.4 are from an invasion assay using SYBR Green I staining, without post-invasion trypsin treatment.

Determining the Source of the Intermediate Population

Although isolates that displayed an intermediate population, only visualised by flow cytometry, could now be phenotyped for invasion accurately by using a post-invasion trypsin treatment to remove the population, the source of this population was still undetermined. The following experiments were undertaken to verify a number of postulates regarding the cause of the intermediate population.

Erythrocyte Rosetting

Clumping of erythrocytes, or rosetting, is a possible reason for a decrease in fluorescent intensity of DNA stains. It was hypothesised that if one parasitised erythrocyte was surrounded by two or more uninfected erythrocytes, the fluorescence intensity from the stained parasite DNA may essentially be masked by the presence of the surrounding, uninfected cells. The clumping of erythrocytes would be consistent with the post-invasion

protease treatment removing the intermediate population, if the ligands responsible for clumping were being digested by the enzyme.

As well as measuring fluorescence intensity, flow cytometry also measures forward scatter and side scatter from each cell. Forward scatter represents the size of the particle, while side scatter corresponds to the intracellular complexity of the cell. If the intermediate population arose from cell clumps, both forward scatter and side scatter would be expected to differ from that of an individual cell. In Fig. 3.22 forward and side scatter were assessed for Peruvian isolate 3135. Fig. 3.22 b shows the forward and side scatter values for just the intermediate population, a subset of the whole sample of 3135 shown in Fig. 3.22 a. The intermediate population does not appear to have a different size or intracellular complexity compared to the rest of the sample.

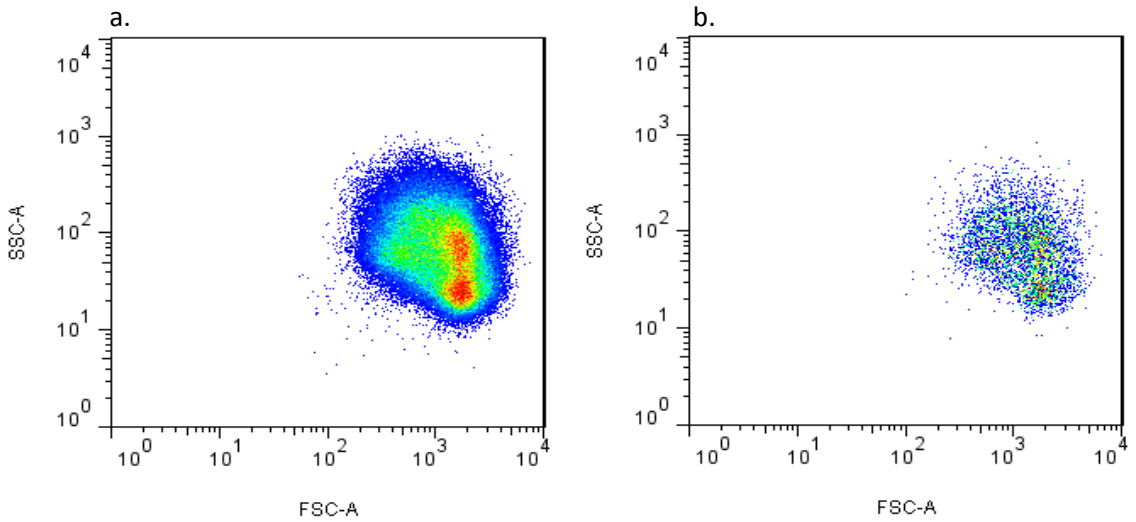


Fig. 3.22. Flow cytometry dot plots of forward scatter (x-axis) and side scatter (y-axis) from Peruvian isolate 3135, to compare the relative size and intracellular complexity of the intermediate population compared to the remainder of the sample. (a) The whole sample of 3135. (b) The intermediate population as a subset of the whole population in (a).

In addition to the flow cytometric analysis of clumping in Fig. 3.22, an erythrocyte rosetting assay was performed using strain 3135 (Chapter 2 "Materials and Methods: Erythrocyte Rosetting Assay"). Using the definition of rosettes as one parasitised erythrocyte in close contact with two or more uninfected erythrocytes, 7/200 parasitised erythrocytes were found in rosettes; a rosetting frequency of 3.5%. In Fig. 3.19 a, the intermediate population is quantified as 3.85% and the parasitised population is 4.87% of the total sample. If every event in the intermediate population involved a rosetted,

parasitised erythrocyte, a rosetting frequency of 44% would be expected. This, combined with the scatter data, suggest that rosettes are not an explanation for the intermediate population.

Merozoite Adhesion

The second possibility was that the intermediate population arose from merozoites adhering to but failing to invade erythrocytes. While a merozoite contains a single copy of the genome and would be expected to bind fluorescent dyes to the same extent as rings or early trophozoites, it cannot survive outside the erythrocyte, and subsequent degradation of the merozoite should lead to a decrease in fluorescence signal.

Variation of the Intermediate Population over the Life Cycle

If merozoite adherence and degradation caused the intermediate population it would therefore be expected that the intermediate population would be at its largest immediately after re-invasion, when the most merozoites are still adhered. This population should then decrease over time. To test this hypothesis, the presence of the intermediate population was measured over time over two separate life cycles of Peruvian isolate 3769. Time course experiments one and two were performed in the same life cycle, while the third experiment was performed in the subsequent life cycle. Time points in the parasite intra-erythrocytic life cycle were estimated by Field's staining and thin smear light microscopy.

Time Course 1 & 2 (h)	Parasite stage
9	Rings
25	Late rings
35	Trophozoites
1	Early rings / late schizonts

Time Course 3 (h)	Parasite stage
43	Late trophozoites / early schizonts
47	Schizonts
3	Early rings / late schizonts
19	Rings
25	Late rings
43	Late trophozoites / early schizonts

Table 3.5. The elapsed time (hours) of the parasite intra-erythrocytic life cycle when measurements of the intermediate population were taken after initiating an invasion assay. It should be noted that the time of the life cycle is an estimate accurate ± six hours. Where more than one parasite life cycle stage was seen in the culture, the predominating stage is listed first.

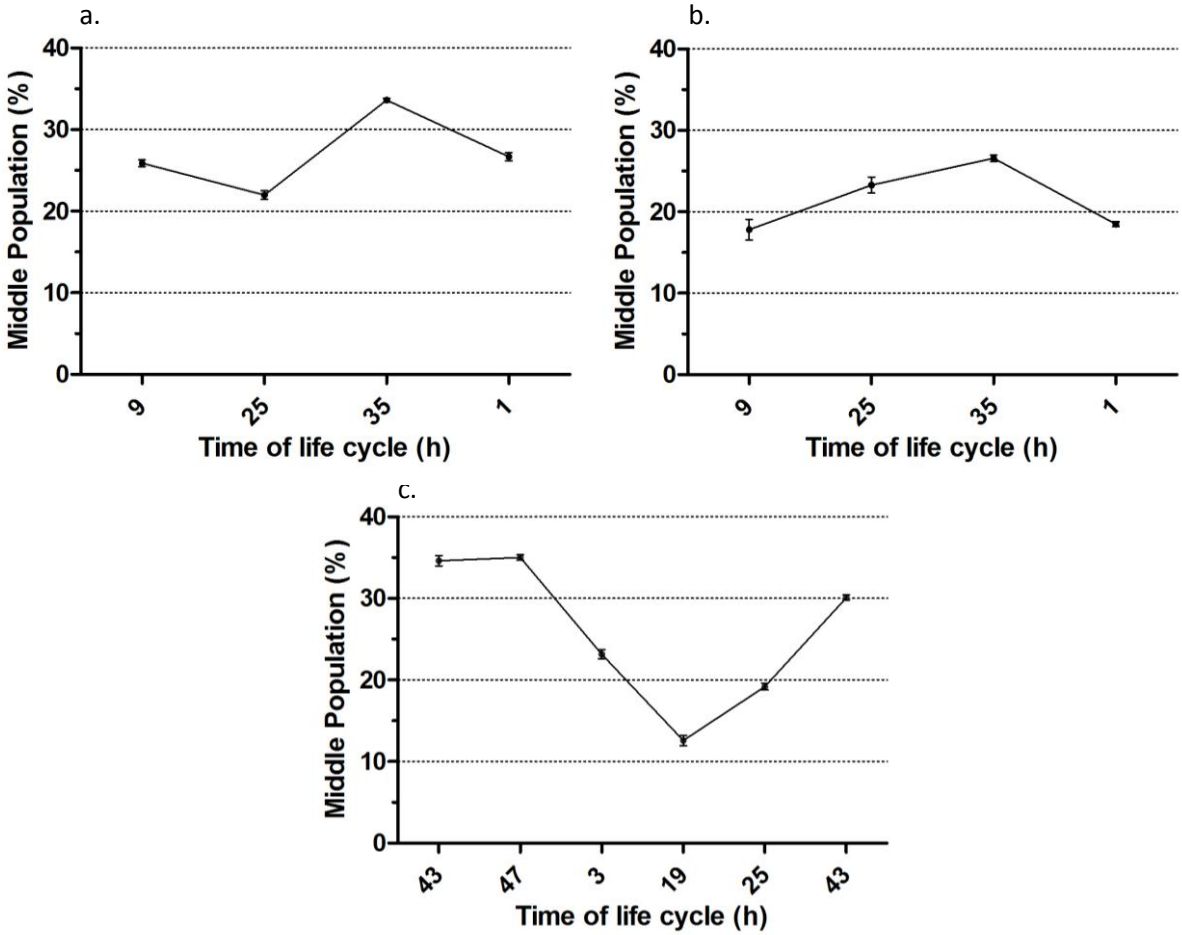


Fig. 3.23. Variation of the intermediate population with life cycle. (a) Time course 1; (b) Time course 2; (c) Time course 3. The intermediate population (y-axis) is expressed as a percentage of the Hoechst 33342 events (the top two populations in Fig. 3.17).

From Fig. 3.23, rather than the hypothesised increase in expression of the intermediate population in the early stages of the parasite life cycle, the population decreased in the ring stage and was largest in the late trophozoite and schizont stages. From the third time course (Fig. 3.23 c), this population accounted for 12.6% (SEM 0.6) of the Hoechst 33342 positive events after 19 ± 6 hours (Hoechst 33342 positive events are the top two populations in Fig. 3.17. This excludes uninfected erythrocytes, which were defined as Hoechst 33342 negative and comprises the bottom population in Fig. 3.17). In the third experiment the population was maximal after 47 hours of the parasite life cycle, representing 35.0% (SEM 0.3) of the Hoechst 33342 positive events. In the first two experiments (Fig. 3.23 a & b), although fewer measurements were taken, the same pattern was apparent with the proportion of Hoechst 33342 positive events present in the intermediate population increasing in later stages of the life cycle.

Anti-MSP-1 19 kDa Antibody Test

To detect specifically the presence of adhered merozoites on the surface of erythrocytes, antibodies to the 19 kDa fragment of MSP-1 were used (Figs. 3.24 to 3.27). MSP-1 is a merozoite surface protein, so would be detectable on extracellular merozoites but not in parasites that have invaded erythrocytes, unless the cells are permeabilised. Peruvian isolate 3769 was used to study the intermediate population, while Peruvian isolate 3541 (which does not express the intermediate population) was used as a control. Anti-MSP-1 19 kDa binding was measured in schizonts and rings using a FITC-conjugated secondary antibody. Anti-MSP-1 19 kDa binding was compared between rings and schizonts, with and without permeabilisation, to differentiate extracellular and intracellular merozoites.

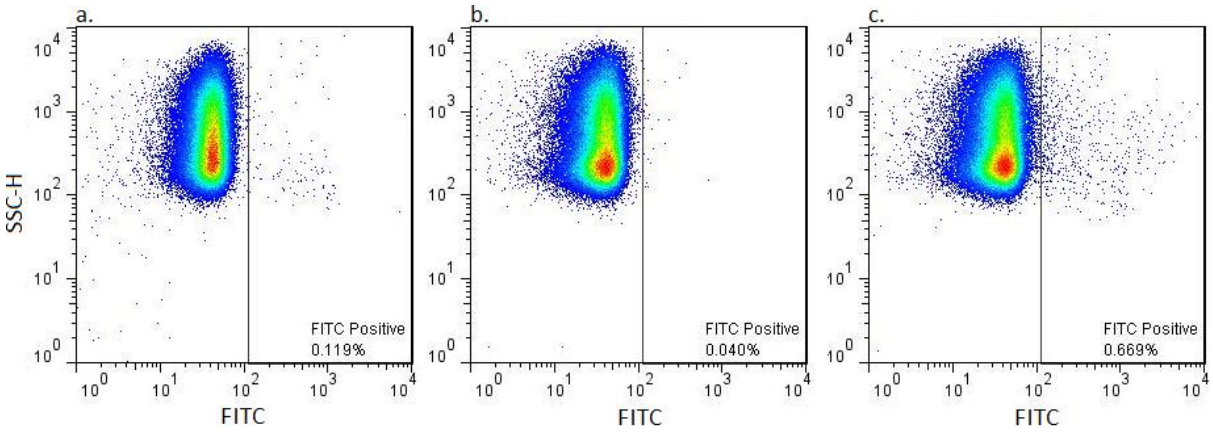


Fig. 3.24. Antibody specificity controls. (a) Uninfected erythrocytes: 1° anti-MSP-1 19 kDa and 2° antibody (negative control). (b) Peruvian isolate 3541 containing 2% schizonts: 2° antibody only (negative control). (c) Peruvian isolate 3541 containing 2% schizonts: 1° anti-MSP-1 19 kDa and 2° antibody (positive control).

Permeabilised Peruvian isolate 3541 at a parasitaemia of 2% schizonts exhibited higher levels of anti-MSP-1 19 kDa antibody binding (FITC positive, 0.67%; Fig. 3.24 c) than the negative controls of uninfected erythrocytes (FITC positive, 0.12%; Fig. 3.24 a) or 2% schizonts from Peruvian isolate 3541 with only secondary antibody (FITC positive, 0.04%; Fig. 3.24 b). This control showed that the 1° antibody binds specifically to MSP-1 and not to uninfected erythrocytes, and that there was no non-specific binding of the 2° antibody.

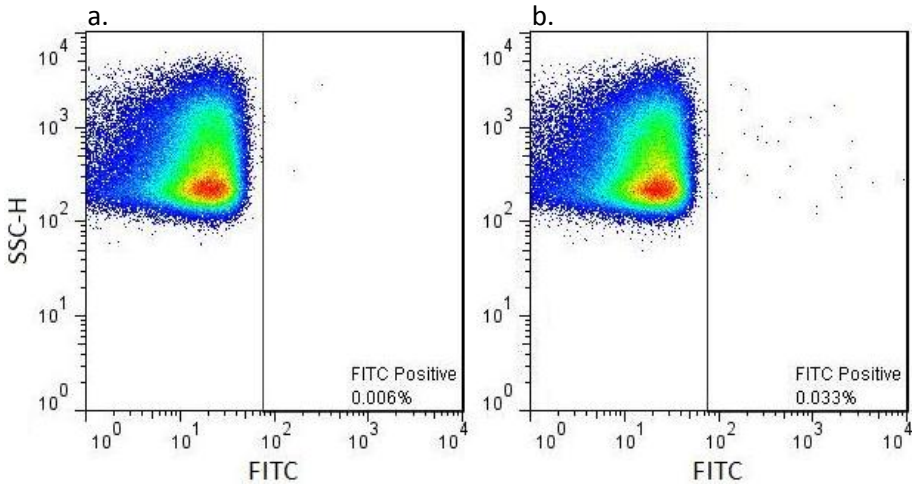


Fig. 3.25. Peruvian isolate 3541 at a parasitaemia of 6% rings. (a) 2° antibody only (negative control). (b) 1° anti-MSP-1 19 kDa antibody and 2° antibody (negative control).

Peruvian isolate 3541 at a parasitaemia of 6% rings was used as a negative control to show that there was no extracellular MSP-1 present in isolates lacking the intermediate population (Fig. 3.25). No MSP-1 and associated merozoites were detected with 1° anti-MSP-1 19 kDa antibody and FITC conjugated 2° antibody (Fig. 3.25 b), compared to 2° antibody used alone (Fig. 3.25 a).

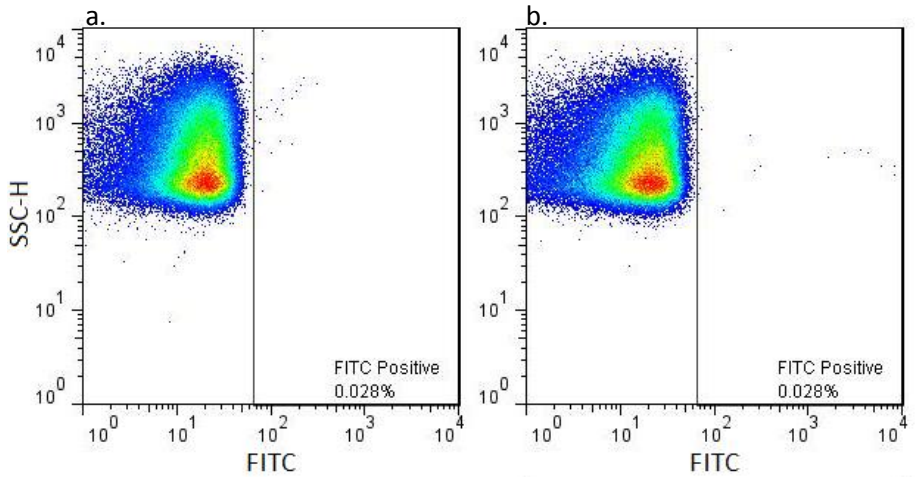


Fig. 3.26. Peruvian isolate 3769 at a parasitaemia of 6% rings. (a) 2° antibody only (negative control). (b) 1° anti-MSP-1 19 kDa antibody and 2° antibody.

Peruvian isolate 3769 contained the intermediate population hypothesised to be caused by merozoite adherence. However, no extracellular MSP-1 specific labelling was detected on this isolate (Fig. 3.26b).

References

Alonso PL, Sacarlal J, Aponte JJ, Leach A, Macete E, Aide P, Sigauque B, Milman J, Mandomando I, Bassat Q, Guinovart C, Espasa M, Corochan S, Lievens M, Navia MM, Dubois MC, Menendez C, Dubovsky F, Cohen J, Thompson R, Ballou WR. 2005. Duration of protection with RTS,S/AS02A malaria vaccine in prevention of *Plasmodium falciparum* disease in Mozambican children: single blind extended follow-up of a randomised controlled trial. *The Lancet.* Vol. 6736(5); pp. 669-676.

Arnot DE, Gull K. 1998. The *Plasmodium* cell-cycle: facts and questions. *Annals of Tropical Medicine & Parasitology.* Vol.92, No. 4, pp. 361-365.

Bakacs T, Vegh Z, Merry AH, Sim RB, Varga L, Kertesz Z, Tusnady G, Klein E. 1993. Interactions in the complement-mediated lysis of blood group AB erythrocytes sensitized simultaneously with anti-A and anti-B monoclonal antibodies. *Immunol Lett.* Vol. 35(3); pp. 219-28.

Baker DA. 2010. Malaria gametocytogenesis. *Molecular and Biochemical Parasitology.* Vol. 172;2, pp. 57-65.

Bannister LH, Hopkins JM, Dluzewski AR, Margos G, Williams IT, Blackman MJ, Kocken CH, Thomas AW, Mitchel GH. 2003. *Plasmodium falciparum* apical membrane antigen-1 (*Pf*AMA-1) is translocated within micronemes along subpellicular microtubules during merozoite development. *Journal of Cell Science.* Vol. 116; pp. 3825-3834.

Bate CA, Kwiatkowski D. 1994. Inhibitory immunoglobulin M antibodies to tumour necrosis factor-inducing toxins in patients with malaria. *Infection and Immunity.* 62(8):pp. 3086-3091.

Baum J, Papenfuss AT, Baum B, Speed TP, Cowman AF. 2006. Regulation of apicomplexan actin-based motility. *Nature Reviews Microbiology.* Vol 4; pp. 621-628.

Baum J, Pinder M, Conway DJ. 2002. Erythrocyte invasion phenotypes of *Plasmodium falciparum* in The Gambia. *Infection and Immunity.* Vol. 71(4): pp. 1856-1863.

Bei AK, Membi CD, Rayner JC, Mubi M, Ngasala B, Sultan AA, Premji Z, Duraisingh MT. 2007. Variant merozoite protein expression is associated with erythrocyte invasion phenotypes in *Plasmodium falciparum* isolates from Tanzania. *Molecular & Biochemical Parasitology*. Vol. 153; pp. 66-71.

Beier JC. 1998. Malaria Parasite Development in Mosquitoes. *Annual Review of Entomology.* 43: pp. 519-543.

Billker O, Lindo V, Panico M, Etienne AE, Paxton T, Dell A, Rogers M, Sinden RE, Morris HR. 1998. Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito. *Nature.* 392; pp.289-292.

Binks RH, Conway DJ. 1999. The major allelic dimorphisms in four *Plasmodium falciparum* merozoite proteins are not associated with alternative pathways of erythrocyte invasion. *Molecular and Biochemical Parasitology*. Vol. 103; pp. 123-127.

Bojang KA, Milligan PJM, Pinder M, Vigneron L, Alloueche A, Kester KE, Ballou WR, Conway J, Reece WHH, Gothard P, Yamuah L, Delchambre M, Voss G, Greenwood BM, Hill A, McAddam KPWJ, Tornieporth N, Cohen JD, Doherty T. 2001. Efficacy of RTS,S/AS02 malaria vaccine against *P. falciparum* infection in semi-immune adult men in The Gambia: a randomised trial. *The Lancet.* Vol. 358(9297); pp. 1927-1934.

Boyle MJ, Richards JS, Gilson PR, Chai W, Beeson JG. 2010. Interactions with heparin-like molecules during erythrocyte invasion by *Plasmodium falciparum* merozoites. *Blood.* Vol. 115(22): pp. 4559-4568.

Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, BeRisi JL. 2003. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biology.* Vol. 1; 1 pp. 85-100.

Branch OH, Casapia WM, Gamboa DV, Hernandez JN, Alava FF, Roncal N, Alvarez E, Perez EJ, Gotuzzo E. 2005. Clustered local transmission and asymptomatic *Plasmodium falciparum* and *Plasmodium vivax* malaria infections in a recently emerged, hypoendemic Peruvian Amazon community. *Malaria Journal*. Vol. 4(27).

Carme B, Bouquety JC, Plassart H. 1993. Mortality and sequelae due to cerebral malaria in African Children in Brazzaville, Congo. *Am. J. Med. Hyg.* Vol. 48(2); pp. 216-221.

Carter R, Mendis K. 2002. Evolutionary and Historical Aspects of the Burden of Malaria. *Clin. Microbiol. Rev.* 15(4): pp.564-594.

CDC DPDx. 2008. Life cycle of *Plasmodium* spp. Accessed online at [<http://www.dpd.cdc.gov/dpdx/HTML/Malaria.htm> on 16/07/2010].

Cele-Conde CJ, Ayala FJ. 2003. Genera of the human lineage. *PNAS.* 100(13): pp. 7684-7689. [Online] <http://www.pnas.org/content/100/13/7684.full.pdf+html> [Accessed on: 09/07/2010].

Chen Q, Barragan A, Fernandez V, Sundstrom A, Schlichtherle M, Sahlen A, Carlson J, Datta S, Wahlgren M. 1998. Identification of *Plasmodium falciparum* erythrocyte membrane protein (*Pf*EMP-1) as the resetting ligand of the malaria parasite *Plasmodium falciparum. J. Exp. Med.* 187(1): pp. 15-23.

Chen Q, Schlichtherle M, Wahlgren M. 2000. Molecular Aspects of Severe Malaria. *Clinical Microbiology Rev.* Vol.13(3): pp. 439-450.

Chen TR. 1977. *In situ* detection of mycoplasma contamination in cell cultures by the fluorescent dye Hoechst 33258. *Exp. Cell Res.* Vol. 104; pp. 255-262.

Chotivanich K, Udomsangpetch R, Simpson JA, Newton P, Pukrittayakamee S, Looareesuwan S, White NJ. 2000. Parasite multiplication potential and the severity of Falciparum malaria. *Journal of Infectious Diseases.* Vol. 181; pp. 1206-1209.

Corby-Harris V, Drexler A, Watkins de Jong L, Antonova Y, Pakpour N, Ziegler R, Ramberg F, Lewis EE, Brown JM, Luckhart S, Riehle MA. 2010. Activation of Akt signalling reduces the prevalence and

intensity of malaria parasite infection and lifespan in Anopheles stephansi mosquitoes. *PLoS Pathogens.* Vol. 6(7); pp. e1001003.

Cowman AF, Crabb BS, 2006. Invasion of Red Blood Cells by Malaria Parasites. *Cell.* 124, pp. 755-766.

David PH, Handunnetti SM, Leech JH, Gamage P, Mendis KN. 1988. Rosetting: A new cytoadherence property of malaria infected erythrocytes. *Am. J. Trop. Med. Hyg.* 38(2); pp. 289-297.

Deans AM, Lyke KE, Thera MA, Plowe CV, Kone A, Doumbo OK, Kai O, Marsh K, Mackinnon MJ, Raza A, Rowe JA. 2006. Low multiplication rates of African *Plasmodium falciparum* isolates and lack of association of multiplication rate and red blood cell selectivity with malaria virulence. *Am. J. Med. Hyg.* Vol. 74(4); pp. 554-563.

Deans AM, Nery S, Conway DJ, Kai O, Marsh K, Rowe JA. 2007. Invasion profiles and malaria severity in Kenyan *P. falciparum* clinical isolates. *Infection and Immunity.* Vol. 75(6); pp. 3014-3020.

Doumbo OK, Thera MA, Kone AK, Raza A, Tempest LJ, Lyke KE, Plowe CV, Rowe JA. 2009. High levels of *Plasmodium falciparum* rosetting in all clinical forms of severe malaria in African children. *Am. J. Trop. Med. Hyg.* Vol. 81(6); pp. 987-993.

Duraisingh MT, Triglia T, Ralph SA, Rayner JC, Barnwell JW, McFadden GI, Cowman AF. 2003. Phenotypic variation of *Plasmodium falciparum* merozoite proteins directs receptor targeting for invasion of human erythrocytes. *EMBO Journal*. Vol 22; pp. 1047-1057.

Egan JE, Hoffman SL, Haynes D, Sadoff JC, Schneider I, Grau GE, Hollingdale MR, Ripley Ballou W, Gordon DM. 1993. Humoral immune responses in volunteers immunized with irradiated *Plasmodium falciparum* sporozoites. *Am. J. Trop. Med. Hyg.* Vol. 49(2); pp.166-173.

Egan T. 2007. Haemozoin formation. *Molecular and Biochemical Parasitology.* Vol. 157;2. Pp. 127-136.

Filatov MV, Varfolomeeva EY. 1995. Active dissociation of Hoechst 33342 from DNA in living mammalian cells. *Mutation research / Fundamental and Molecular Mechanisms of Mutagenesis.*Vol. 327(1-2); pp. 209-215.

Fujioka H, Aikawa M. 2002. Structure and life cycle. *Chem Immunol. Basel, Karger.* Vol. 80, pp.1-26.

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Paterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DMA, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum. Nature.* Vol. 419; pp. 498-511.

Gilson PR, Nebl T, Vukcevic D, Moritz RL, Sargeant T, Speed TP, Schofield L, Crabb BS. 2006. Identification and stoichiometry of GPI-anchored membrane proteins of the human malaria parasite *Plasmodium falciparum. MCP.* Vol. 5; pp. 1286-1299.

Girard MP, Reed ZH, Friede M, Kieny MP. 2007. A review of human vaccine research and development: Malaria. *Vaccine.* Vol. 25; pp. 1567-1580.

Gomez-Escobar N, Amambua-Ngwa A, Walther M, Okebe J, Ebonyi A, Conway DJ. 2010. Erythrocyte invasion and merozoite ligand gene expression in severe and mild *Plasmodium falciparum* malaria. *Journal of Infectious Disease.* Vol. 201(3); pp. 444-452.

Green JL, Martin SR, Fielden J, Ksagoni A, Grainger M, Yim Lim BYS, Molloy JE, Holder AA. 2006. The MTIP-Myosin A Complex in Blood Stage Malaria Parasites. *Journal of Molecular Biology.* Vol. 355(3); pp. 933-941.

Guar D, Singh S, Singh S, Jiang L, Diouf A, Miller LH. 2007. Recombinant *Plasmodium falciparum* reticulocyte homology protein 4 binds to erythrocytes and blocks invasion. *PNAS.* Vol. 104(45); pp. 17789-17794.

Guarda JA, Asayag CR, Witzig R. 1999. Malaria reemergence in the Peruvian Amazon Region. *Emerging Infectious Diseases.* Vol. 5(2). [Accessed online <http://www.cdc.gov/ncidod/eid/vol5no2/aramburu.htm> on 19/06/2010].

Guinet F, Dvorak JA, Fujioka H, Keister DB, Muratova O, Kaslow DC, Aikawa M, Vaidya AB, Wellems TE. 1996. A Developmental Defect in *Plasmodium falciparum* Male Gametogenesis. *Journal of Cell Biology*. Vol. 135;1. pp. 269-278.

Hayton K, Gaur D, Lui A, Takahashi J, Henschen B, Singh S, Lambert L, Furuya T, Bouttenot R, Doll M, Nawaz F, Mu J, Jiang L, Miller LH, Wellems TE. 2008. Erythrocyte binding protein *Pf*Rh5 polymorphisms determine species-specific pathways of *Plasmodium falciparum* invasion. *Cell Host & Microbe.* Vol. 4; pp. 40-51.

Holder AA. 1994. Proteins on the surface of the malaria parasite and invasion. *Parasitology.* Vol. 108; pp. S5-18.

Imwong M, Snounou G, Pukrittayakamee S, Tanomsing N, Kim JR, Nandy A, Guthmann JP, Nosten F, Carlton J, Looareesuwan S, Nair S, Sudimack D, Day NPJ, Anderson TJC, White NJ. 2007. Relapses of *Plasmodium vivax* infection usually result from activation of heterologous hypnozoites. *Journal of Infectious Diseases.* 1995: pp.927-933.

Jennings CV, Ahouidi AD, Zilversmit M, Bei AK, Rayner JC, Sarr O, Ndir O, Wirth DF, Mboup S, Duraisingh MT. 2007. Molecular analysis of erythrocyte invasion in *Plasmodium falciparum* isolates from Senegal. *Infection and Immunity*. Vol. 75(7); pp. 3531-3538.

Jolliffe I. 2002. Principal Component Analysis. *Springer. 2nd Edition.* pp. 1-6. ISBN: 978-0387954424.

Jones ML, Kitson EL, Rayner JC. 2006. *Plasmodium falciparum* erythrocyte invasion: A conserved myosin associated complex. *Molecular and Biochemical Parasitology.* Vol. 147; pp. 74-84.

Kadekoppala M, Holder AA. 2010. Merozoite surface proteins of the malaria parasite: the MSP-1 complex and the MSP-7 family. *International Journal of Parasitology.* Vol. 40(10); pp. 1155-1166.

Kain KC, Orlandi PA, Haynes JD, Sim KL, Lanar DE. 1993. Evidence for two-stage binding by the 175-kD erythrocyte binding antigen of *Plasmodium falciparum. Journal of Experimental Medicine.* Vol. 178; pp. 1497-1505.

Kappe SHI, Vaughan AM, Boddey JA, Cowman AF. 2010. That was then but this is now: malaria research in the time of an eradication agenda. *Science.* Vol 328(862); pp. 862-866.

Krief S, Escalante AA, Pacheco MA, Mugisha L, Andre C, Halbwax M, Fischer A, Krief JM, Kasenene JM, Crandfield M, Cornejo OE, Chavatte JM, Lin C, Letourneur F, Gruner AC, McCutchan TF, Renia L, Snounou G. 2010. On the diversity of malaria parasites in African Apes and the origin of *Plasmodium falciparum* from Bonobos. *PLoS Pathogens.* Vol. 6(2); pp. E1000765.

Kwiatkowski D. 1989. Febrile temperatures can synchronise the growth of *Plasmodium falciparum* in vitro. *Journal of Experimental Medicine*. 169(1) pp. 357-361.

Kyes S, Horrocks P, Newbold C. 2001. Antigenic Variation at the Infected Red Cell Surface in Malaria. *Annual Rev. of Microbiology.* 55: pp. 673-707.

Lambros C, Vanderberg JP. 1979. Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *Journal of Parasitology.* Vol. 65(3); pp. 418-420.

Lobo CA, de Frazao K, Rodriguez M, Reid M, Zalis M, Lustigman S. 2004. Invasion profiles of Brazilian Field Isolates of *Plasmodium falciparum:* Phenotypic and genotypic analyses. *Infection and Immunity.* Vol. 72(10); pp. 5886-5891.

Malaria Site. 2010. *MalariaSite.* [Accessed online <http://www.malariasite.com/malaria/history_control.htm>] [Accessed on: 21/06/2010].

Manske HM, Kwiatkowski DP. 2009. LookSeq: a browser-based viewer for deep sequencing data. *Genome Research.* Vol. 19(11); pp. 2125-2135.

Manske HM, Kwiatkowski DP. 2009. SNP-o-matic. *Bioinformatics.* Vol. 25(18); pp. 2434-2435.

Mardis ER. 2008. The Impact of next-generation sequencing technology on genetics. *Trends in Genetics.* Vol. 24(3); pp. 133-141.

Martin RE, Marchetti RV, Cowan AI, Howitt SM, Bröer S, Kirk K. 2009. Chloroquine transport via the malaria parasite's chloroquine resistance transporter. *Science.* Vol. 325(5948); pp. 1690-1682.

Mayer DCG, Mu JB, Feng X, Su X, Miller LH. 2002. Polymorphism in a *Plasmodium falciparum* erythrocyte binding ligand changes its receptor specificity. *JEM.* Vol. 196(11); pp. 1523-1528.

Mayer DCG, Mu JB, Kaneko O, Duan J, Su X, Miller LH. 2004. Polymorphism in the P. falciparum erythrocyte binding ligand JESEBL/EBA-181 alters its receptor specificity. *PNAS.* Vol. 101(8); pp. 2518-2523.

Miller LH, Baruch DI, Marsh K, Doumbo OK. 2002. The pathogenic basis of malaria. *Nature.* 415; pp.673-679.

Miller LH, Good MF, Milon G. 1994. Malaria Pathogenesis. *Science*. Vol. 264; 5167. pp. 1878-1883.

Miller LH, Mason SJ, Clyde DF, McGinniss MH. 1976. The resistance factor to Plasmodium vivax in blacks. The Duffy blood group genotype FyFy. *New England Journal of Medicine.* Vol. 295(6); pp. 302-304.

Mital J, Meissner M, Soldati D, Ward GE. 2005. Conditional expression of *Toxoplasma gondii* apical membrane antigen-1 (TgAMA1) demonstrates that TgAMA1 plays a critical role in host cell invasion. *Mol. Biol. Cell*. 16; pp. 4341-4349.

Molyneux M, Fox R. 1993. Diagnosis and treatment of malaria in Britain. *British Medical Journal.* Vol. 306. Pp. 1175-1180.

Moorthy VS, Good MF, Hill AVS. 2004. Malaria vaccine developments. *The Lancet.* Vol 363; pp. 150-156.

Moskes C, Burghaus PA, Wernli B, Sauder U, Durrenberger M, Kappes B. 2004. Export of *Plasmodium falciparum* calcium-dependent protein kinase 1 to the parasitophorous vacuole is dependent on the N-terminal membrane anchor motifs. *Mol. Microbiology.* Vol. 54(3); pp. 676-691.

Moustafa I, Connaris H, Taylor M, Zaitsev V, Wilson JC, Kiefel MJ, von Itzstein M, Taylor G. 2004. Sialic acid recognition by *Vibrio cholerae* neuraminidase. *The Journal of Biological Chemistry.* Vol. 279; pp. 40819-40826.

Mu J, Awadalla P, Duan J, McGee KM, Joy DA, McVean GAT, Su X. 2005. Recombination hotspots and population structure in *Plasmodium falciparum. PLoS Biology.* Vol. 3(10); pp. 1735-1741.

Narum DL, Haynes JD, Fuhrmann S, Moch K, Liang H, Hoffman SL, Sim BKL. 2000. Antibodies against the *Plasmodium falciparum* receptor binding domain of EBA-175 block invasion pathways that do not involve sialic acids. *Infection and Immunity.* Vol. 68(4); pp. 1964-1966.

Nery S, Deans AM, Mosobo M, Marsh K, Rowe JA, Conway DJ. 2006. Expression of Plasmodium falciparum genes involved in erythrocyte invasion varies among isolates cultured directly from patients. *Mol. Biochem. Parasitology.* Vol. 149(2); pp. 208-215.

Okoyeh JN, Pillai CR, Chitnis CE. 1999. *Plasmodium falciparum* field isolates commonly use erythrocyte invasion pathways that are independent of sialic acid residues of glycophorin A. *Infection and Immunity.* Vol. 67(11); pp. 5784-5791.

Parekh FK, Hernandez JN, Krogstad DJ, Casapia WM, Branch OH. 2007. Prevalence and risk of *Plasmodium falciparum* and *P. vivax* malaria among pregnant women living in the hypoendemic communities of the Peruvian Amazon. *Am. J. Trop. Med. Hyg.* Vol. 77(3); pp. 451-457.

Prugnolle F, Durand P, Neel C, Ollomo B, Ayala FJ, Arnathau C, Etienne L, Mpoudi-Ngole E, Nkoghe D, Leroy E, Delaporte E, Peeters M, Renaud F. 2010. African great apes are natural hosts of multiple related malaria species, including *Plasmodium falciparum. PNAS.* Vol. 107(4); pp. 1458-1463.

Rayner JC, Vargas-Serrato E, Huber CS, Galinski MR, Barwell JW. 2001. A *P. falciparum* homologue of *Plasmodium vivax* reticulocyte binding protein (PvRBP1) defines a trypsin resistant erythrocyte invasion pathway. *Journal Exp. Med.* Vo. 194; pp. 1571-1581.

Rayner JC. 2009. The merozoite has landed: reticulocyte-binding-like ligands and the specificity of erythrocyte recognition. *Trends in Parasitology.* Vol. 25(3); pp. 104-106.

Rich SM, Leendertz FH, Xu G, LeBreton M, Djoko CF, Aminake MN, Takang EE, Diffo JLD, Pike BL, Rosenthal BM, Formenty P, Boesch C, Ayala FJ, Wolfe ND. 2009. The origin of malignant malaria. *PNAS.* Vol. 106(35); pp. 14902-14907.

Richard D, MacRalid CA, Tiglar DT, Chan JA, Foley M, Baum J, Ralph SA, Norton RS, Cowman AF. 2010. Interaction between *Plasmodium falciparum* apical membrane antigen-1 and the rhoptry neck protein complex defines a key step in the erythrocyte invasion process of malaria parasites. *The Journal of Biological Chemistry*. Vol. 285; pp. 14815-14822.

Richards JS, Beeson JG. 2009. The future for blood-stage vaccines against malaria. *Immunology and Cell Biology.* Vol. 87; pp. 377-390.

Rodriguez M, Lustigman S, Montero E, Oksov Y, Lobo CA. 2008. *Pf*Rh5: A novel reticulocyte-binding family homolog of *P. falciparum* that binds to the erythrocyte, and an investigation of its receptor. *PLoS ONE.* Vol. 3(10): pp. e3300.

Sam-Yellowe TY.1996. Rhoptry organelles of the apicomplexa: Their role in host cell invasion and intracellular survival. *Parasitology Today.* Vol. 12(8); pp. 308-316.

Sanders PR, Gilson PR, Cantin GT, Greenbaum DC, Nebl T, Carucci DJ, McConville MJ, Schofield L, Hodder AN, Yates III JR, Crabb BS. 2005. Distinct protein classes including novel merozoite surface antigens in raft-like membranes of *Plasmodium falciparum. Journal of Biological Chemistry.* Vol. 280(48): pp. 40169-40176.

Sanger F, Nicklen S, Coulson AR. 1977. DNA Sequencing with chain-terminating inhibitors. *PNAS.* Vol. 74(12) pp. 5463-5467.

Saul A, Myler P, Elliott T, Kidson C. 1982. Purification of mature schizonts of *Plasmodium falciparum* on colloidal silica gradients. *Bulletin of the World Health Organisation.* Vol. 60(5); pp. 755-759.

Schlagenhauf-Lawlor. 2008. Travellers' Malaria. *BD Decker.* Second Edition. pp. 10-23.

Shendure J, Hanlee J. 2008. Next generation DNA sequencing. *Nature biotechnology.* Vol. 26(10); pp. 1135-1145.

Sherman IW. 1998. Malaria: Parasite Biology, Pathogenesis and Protection. *American Society for Microbiology.* pp. 3-10. ISBN-13: 978-1555811310.

Silvie O, Rubinstein E, Franetich JF, Prenant M, Belnoue E, Rénia L, Hannoun L, Eling W, Levy S, Boucheix C, Mazier D. 2002. Hepatocyte CD81 is required for *Plasmodium falciparum* and *Plasmodium yoelii* sporozoites infectivity. *Nature Medicine.* 9 pp. 93-96.

Sim BK, Chitnis CE, Wasniowska K, Hadley TJ, Miller LH. 1994. Receptor and ligand domains for invasion of erythrocytes by *Plasmodium falciparum. Science.* Vol. 264(5167); pp. 1941-1944.

Singh S, Soe S, Weisman S, Barnwell JW, Perignon L, Druilhe P. 2009. A conserved multi-gene family induces cross-reactive antibodies effective in defense against *Plasmodium falciparum. PLoS One.* Vol. 4(4); pp. E5410.

Smith JD, Gamain B, Baruch DI, Kyes S. 2001. Decoding the language of *var* genes and *Plasmodium falciparum* sequestration. *Trends in Parasitology.* Vol.17;11. pp. 538-545.

Snow RW, Craig M, Deichmann U, Marsh K. 1999. Estimating mortality, morbidity and disability due to malaria among Africa's non-pregnant population. *World Health Organisation Bulletin.* Vol. 77(8); pp. 624-640.

Sowumni A. 1997. Clinical study of cerebral malaria in African children. *African Journal of Medical Science.* Vol. 26(1-2); pp. 9-11.

Stanley C. Oaks, Jr., Violaine S. Mitchell, Greg W. Pearson, and Charles C.J. Carpenter. 1991. Malaria: Obstacles and Opportunities. pp. 37-39. ISBN-13: 978-0-309-04527-8.

Stark KR, James AA. 1996. Salivary gland anticoagulants in culicine and anopheline mosquitoes. *J Med Entomol.* Jul;33(4):pp.645-50.

Taylor HM, Grainger M, Holder AA. 2002. Variation in the expression of a *Plasmodium falciparum* protein family implicated in erythrocyte invasion. *Infection and Immunity.* Vol. 70(10); pp. 5779-5789.

Tham WH, Wilson DW, Reiling L, Chen L, Beeson JG, Cowman AF. 2009. Antibodies to reticulocyte binding protein-like homologue 4 inhibit invasion of *P. falcparum* into human erythrocytes. *Infection and Immunity.* Vol. 77(6); pp. 2427-2435.

Tham WH, Wilson DW, Lopaticki S, Schmidt CQ, Tetteh-Quarcoo PB, Barlow PN, Richard D, Corbin JE, Beeson JG, Cowman AF. 2010. Complement receptor 1 is the host erythrocyte receptor for *Plasmodium falciparum Pf*Rh4 invasion ligand. *PNAS.* Vol. 107(40); pp. 17327-17332.

Theron M, Hesketh RL, Subramanian S, Rayner JC. 2010. An adaptable two-color flow cytometric assay to quantitate the invasion of erythrocytes by *Plasmodium falciparum.* 2010. *Cytometry A.* Vol. 77(11); pp. 1067-1074.

Trucco C, Fernandez-Reyes D, Howell S, Stafford WH, Scott-Finnigan TJ, Grainger M, Ogun SA, Taylor WR, Holder AA. 2001. The merozoite surface protein 6 gene codes for a 36 kDa, protein associated

with the *Plasmodium falciparum* merozoite surface protein-1 complex. *Molecular and Biochemical Parasitology.* Vol. 112(1); pp. 91-101.

VanBuskirk KM, Sevova E, Adams JH. 2004. Conserved residues in the *Plasmodium vivax* Duffy-binding protein ligand domain are critical for erythrocyte receptor recognition. *PNAS*. Vol. 101(44); pp. 15754-15759.

Van Diggelen OP, Niermeijer MF. 1986. Elimination of mycoplasmas from infected cell cultures by combined trypsin / antibiotics treatment. *J. Inher. Metab. Dis.* Vol. 9; pp. 398-399.

Ward GE, Miller LH, Dvorak JA. 1993. The origin of parasitophorous vacuole membrane lipids in malaria infected erythrocytes. *Journal of Cell Science.* Sept;106 (Pt 1) pp. 237- 248.

Woehlbier U, Epp C, Kauth CW, Lutz R, Long CA, Coulibaly B, Kouyate B, Arevalo-Herrera M, Herrera S, Bujard H. 2006. Analysis of antibodies directed against merozoite surface protein 1 of the human malaria parsite *P. falciparum. Infection and Immunity.* Vol. 74(2); pp. 1313-1322.

World Health Organisation. 2009. *World Malaria Report.* [Online] World Health Organisation. Available at <http://whqlibdoc.who.int/publications/2009/9789241563901_eng.pdf> [Accessed on: 10/07/2010].

Appendix

Table A1. Summary of all field isolates received from Peru.

Strain	Freezing date	Cultured?	Phenotyped	Sequenced	RNA yield (ng)	Gametocytogenesis?
278	02/05/2008	Grew	Yes	Yes	3472	Yes
788	28/09/2008	Grew	Yes	Yes	2052	No
3106	27/04/2008	Grew	Yes	Yes	1870	No
3135	04/04/2007	Grew	Yes	Yes	6612	No
3541	07/01/2006	Grew	Yes	Yes	3420	No
3769	20/12/2004	Grew	Yes	Yes	2348	No
5802	15/03/2004	Grew	Yes	Yes	2192	Yes
5809	10/03/2004	Grew	Yes	Yes	3421	Yes
5814	19/03/2004	Grew	Yes	Yes	2751	Yes
6390	24/02/2007	Grew	Yes	Yes	4018	Yes
9050	04/02/2007	Grew	No	Yes	N.e.	Yes
5188	23/02/2004	Lysed on thawing	No	No	N.e.	No
93	31/10/2007	Did not grow	No	No	N.e.	No
229	28/09/2008	Did not grow	No	No	N.e.	No
254	No date	Did not grow	No	No	N.e.	No
255	24/04/2006	Did not grow	No	No	N.e.	No
303	15/03/2004	Did not grow	No	No	N.e.	No
527	19/04/2002	Did not grow	No	No	N.e.	No
806	21/07/2008	Did not grow	No	No	N.e.	No
808	21/06/2008	Did not grow	No	No	N.e.	No
1546	19/08/2007	Did not grow	No	No	N.e.	No
1551	17/07/2007	Did not grow	No	No	N.e.	No
1975	19/07/2007	Did not grow	No	No	N.e.	No
2215	13/05/2006	Did not grow	No	No	N.e.	No
2790	26/03/2006	Did not grow	No	No	N.e.	No
3265	19/07/2007	Did not grow	No	No	N.e.	No
3307	12/06/2007	Did not grow	No	No	N.e.	No
3329	25/07/2007	Did not grow	No	No	N.e.	No
3626C	28/05/2004	Did not grow	No	No	N.e.	No
4333D	07/11/2006	Did not grow	No	No	N.e.	No
4382	15/03/2007	Did not grow	No	No	N.e.	No
4406	29/07/2007	Did not grow	No	No	N.e.	No
4406	10/08/2007	Did not grow	No	No	N.e.	No
4416	08/10/2007	Did not grow	No	No	N.e.	No
4418	28/10/2007	Did not grow	No	No	N.e.	No
5777	24/02/2004	Did not grow	No	No	N.e.	No
5803	15/03/2004	Died @ 3 weeks	No	No	N.e.	Yes
5849C	12/05/2004	Did not grow	No	No	N.e.	No
6164z	27/04/2006	Did not grow	No	No	N.e.	No
6184z	27/04/2006	Did not grow	No	No	N.e.	No
6215z	24/05/2006	Did not grow	No	No	N.e.	No
8033	25/07/2007	Did not grow	No	No	N.e.	No
8033	17/07/2007	Did not grow	No	No	N.e.	No
9051	07/02/2007	Did not grow	No	No	N.e.	No
9704z	25/03/2006	Did not grow	No	No	N.e.	No
5763	16/02/2004	Died @ 2 weeks	No	No	N.e.	Yes

Table A1. Summary of all field isolates received from Peru. N.e. = RNA not extracted.

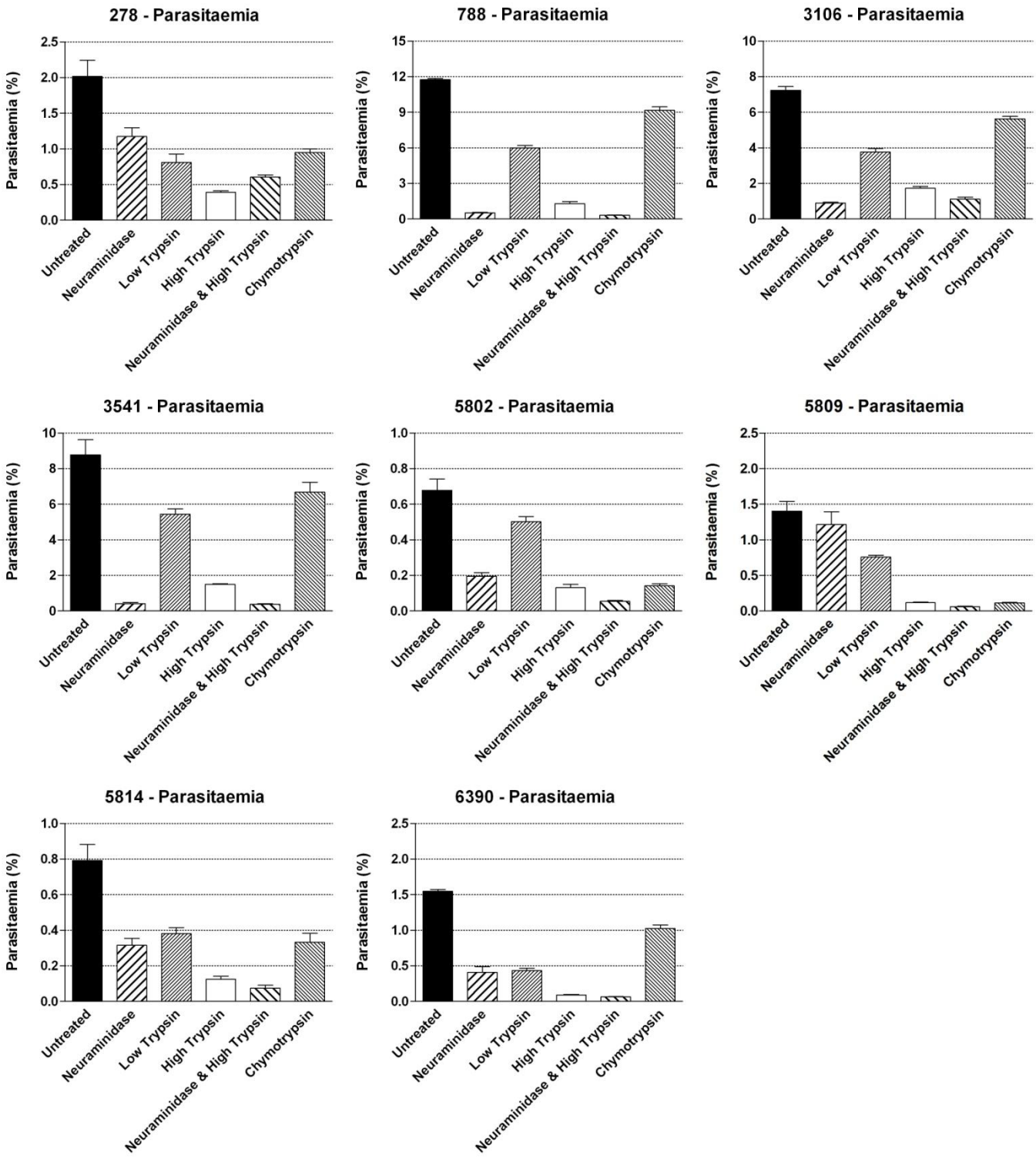


Fig. A1. Parasitaemia of each phenotyping invasion assay, without using post-invasion trypsin treatment. The bars represent the mean parasitaemia calculated from two experiments of three replicates. Error bars are SEMs. These profiles are quantitatively similar to those in Results Fig. 3.3, due to post-invasion trypsin treatment having very little impact on true parasitaemia.

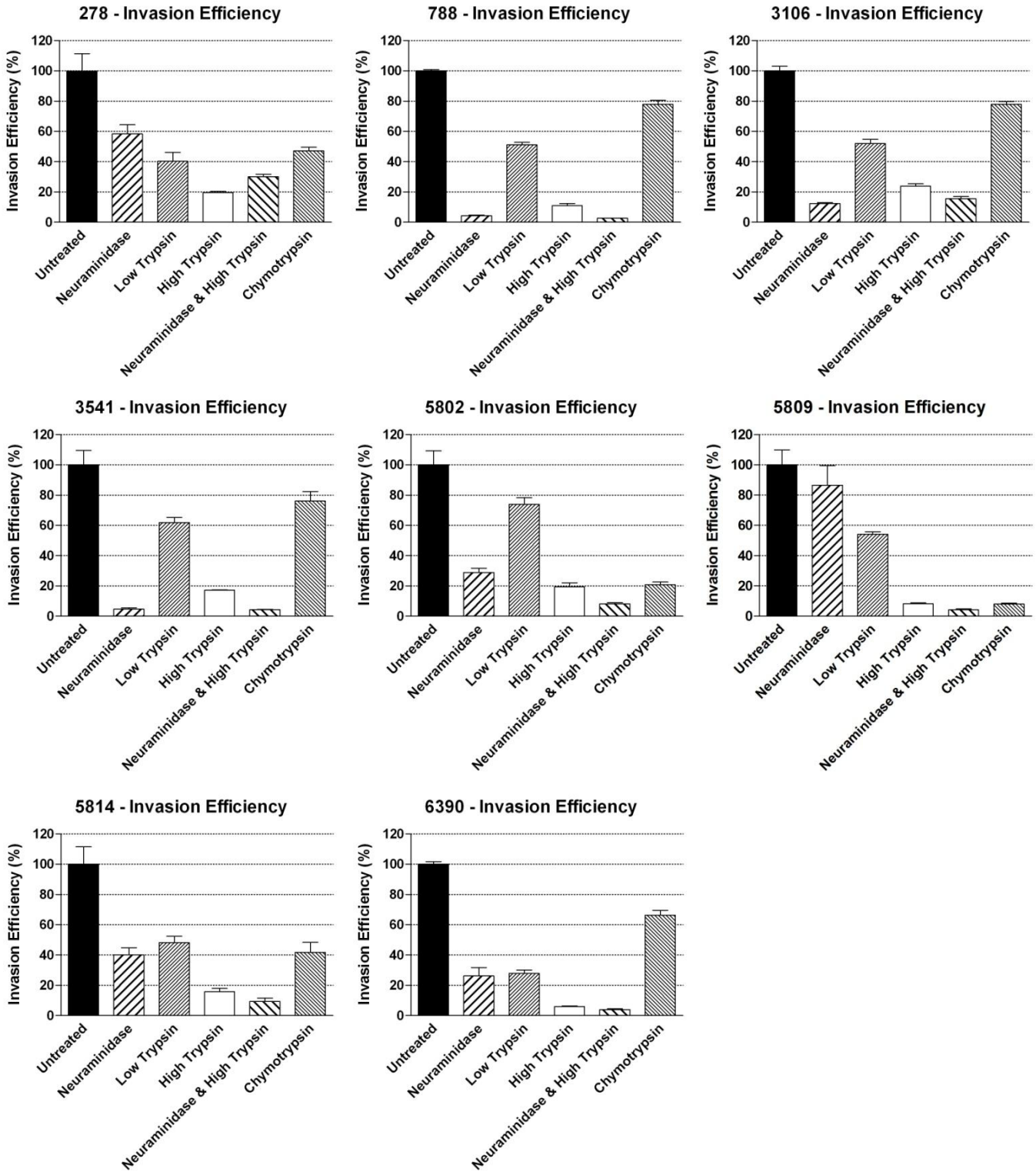


Fig. A2. Invasion phenotypes of Peruvian field isolates, without using post-invasion trypsin treatment. Error bars are SEMs. These profiles are quantitatively similar to those in Results Fig. 3.4, due to post-invasion trypsin treatment having very little impact on true parasitaemia.

Table A2. SNPs present in Peruvian isolate 6390.

Gene	Annotation	Chromosome	Start	Finish	Size (aa)	SNP Position	3D7 Reference base	SNP base	Reads	Reference AA	Mutated AA	AA Position
MSP-1	PFI1475w	MAL9	1201801	1206964	1720	1202181	G	T	190	S	I	127
						1202806	T	A	4	D	E	335
						1202850	A	G	1	N	S	350
						1202853	C	A	4	T	K	351
						1203239	A	C	63	K	Q	480
						1203377	A	G	49	N	D	526
						1205108	C	T	69	H	Y	1103
						1205322	T	C	92	V	A	1174
MSP-10	PFF0995c	MAL6	851374	852952	525	851416	A	G	23	F	L	513
Pf92	PF13_0338	MAL13	2564890	2567281	796	2565982	A	C	66	K	T	364
						2566523	G	A	94	E	K	545
						2567063	A	C	65	T	P	725
RAMA	MAL7P1.208	MAL7	394420	396777	785	395539	G	C	129	E	Q	374
						396574	A	G	73	K	E	719
						396583	G	A	75	D	N	722
EBA-140	MAL13P1.60	MAL13	89421	93455	1210	93122	G	A	71	L	F	112
EBA-175	MAL7P1.176	MAL7	1413430	1418305	1462	1414265	A	G	83	K	E	279
MTRAP	PF10_0281	MAL10	1181002	1182499	498	1181681	T	C	66	Y	H	227
MSP3	PF10_0345	MAL10	1404191	1405256	354	1404394	T	C	116	L	S	68
MSP7	PF13_0197	MAL13	1419283	1420339	351	1419437	T	A	88	K	N	301
						1419488	A	C	106	D	E	284
Pf41	PFD0240c	MAL4	273708	274845	378	274612	C	A	61	K	N	78
RAP2	PFE0080c	MAL5	84040	85237	398	85096	T	G	61	N	H	48
RhopH2	PFI1445w	MAL9	1175192	1180752	1378	1175691	A	G	60	I	V	167
						1175895	G	A	54	A	T	235
Rh1	PFD0110w	MAL4	144097	153112	2971	144766	G	A	38	E	K	191
						145010	T	A	22	I	K	272
						145012	A	G	26	K	E	273
						146968	C	T	35	L	F	925
						147183	A	T	36	L	F	996
						148354	G	A	41	E	K	1387
						149371	G	A	29	E	K	1726

Gene	Annotation	Chromosome	Start	Finish	Size (aa)	SNP Position	3D7 Reference base	SNP base	Reads	Reference AA	Mutated AA	AA Position
						149930	T	C	29	V	A	1912
Rh5	PFD1145c	MAL4	1086801	1088589	526	1087775	C	T	64	C	Y	203
PTRAMP	PFL0870w	MAL12	703887	704946	352	704481	T	A	57	N	K	198
TLP	PFF0800w	MAL6	684738	688854	1371	685237	G	A	48	V	I	167
						687145	A	T	80	S	C	803
PF10_0323	PF10_0323	MAL10	1335219	1336287	355	1335366	T	A	83	N	K	49
PFL2505c	PFL2505c	MAL12	2114613	2122473	2215	2115215	G	C	37	T	R	2016
						2115685	A	T	57	N	K	1859
						2116716	G	C	47	Q	E	1516
						2117920	T	A	50	K	N	1114
RON2	PF14_0495	MAL14	2134233	2140803	2189	2138624	A	C	114	E	D	1464
						2138627	A	G	109	N	S	1465
GAMA	PF08_008	MAL8	1238356	1240573	738	1239800	C	T	53	M	I	258
						1239889	T	C	69	T	A	229
						1240374	T	A	53	N	I	67
PTRAMP	MAL12P1.174	MAL12	703887	704946	281	704481	T	A	57	N	K	198
PFL0865W	MAL12P1.173	MAL12	700523	702204	309	701619	G	T	25	E	D	289
PF10_0350	PF10_0350	MAL10	1420529	1422668	712	1421655	A	G	26	T	A	376
MSP3.8	PF10_0355	MAL10	1432494	1434783	762	1432628	A	T	50	E	D	45
						1432629	A	T	46	E	D	45
						1432699	A	G	47	K	E	69
						1433171	A	A/C	28/24	K	T	226
						1433179	G	G/A	52/28	D	K	229
						1433181	T	T/A	51/31	D	K	229
						1433190	A	A/T	58/24	K	N	232
						1433204	A	A/C	80/26	Y	S	237
						1433209	T	T/G	84/22	S	G	239
						1433210	C	C/G	86/21	S	G	239
						1433211	A	A/T	95/14	S	G	239
						1433242	A	A/G	91/25	N	G	250
						1433243	A	A/G	89/26	N	G	250
						1433251	T	T/A	77/41	L	T	253
						1433252	T	T/C	79/46	L	T	253
						1433264	A	A/T	88/77	H	L	257

Gene	Annotation	Chromosome	Start	Finish	Size (aa)	SNP Position	3D7 Reference base	SNP base	Reads	Reference AA	Mutated AA	AA Position
						1433274	A	A/G	92/96	I	M	260
						1433276	G	G/A	90/98	R	K	261
						1433281	A	A/G	88/108	S	G	263
						1433305	A	A/G	82/104	I	V	271
						1433308	A	A/C	97/64	R	Q	272
						1433309	G	G/A	87/90	R	Q	272
						1433314	G	G/A	87/103	D	N	274
						1433325	G	G/A	82/98	M	I	277
						1433343	A	A/T	71/92	K	N	283
						1433344	G	G/A	70/86	E	K	284
						1433351	T	T/A	69/98	I	K	286
						1433365	A	A/G	54/89	K	E	291
						1433374	A	A/G	48/80	K	G	294
						1433375	A	A/G	46/70	K	G	294
						1433378	T	T/A	43/60	I	K	295
						1433379	T	T/A	54/41	I	K	295
						1433380	T	T/C	47/68	Y	Q	296
						1433382	T	T/A	55/57	Y	Q	296
						1433383	A	A/T	58/64	N	Y	297
						1433386	G	G/A	56/62	E	K	298
						1433392	A	A/G	59/69	N	V	300
						1433393	A	A/T	68/55	N	V	300
						1433395	G	G/A	66/82	D	N	301
						1433398	A	A/G	75/83	K	D	302
						1433400	A	A/T	74/83	K	D	302
						1433669	T	C	7	M	T	392
						1433754	T	G	9	N	K	420
						1433797	G	A	10	E	N	435
						1433799	A	T	18	E	N	435
MAL4P1.41b	PFD0207c	MAL4	240477	243386	639	240827	A	T	38	F	Y	565
RON4	PF11_0168	MAL11	602981	606917	1201	606433	A	T	58	I	N	162
RON5	MAL8P1.73	MAL8	806922	814591	1156	812208	G	T	24	F	L	288

Table A2. SNPs present in Peruvian isolate 6390. In red: SNPs called on fewer than 10 reads. In green: 2 SNPs correspond to one amino acid change. In blue: 3 SNPs correspond to one amino acid change.

List of Figures