Plasmodium falciparum Population Genetics in Northern Ghana

Lucas N Amenga-Etego

Wolfson College,

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas Term 2012

Dedicated to the people of the Kassena-Nankana Districts
and the young people in our communities who strive to
unlock their potential and make a difference
and to the boys
Alenwine and Awingura
keep reaching for that rainbow!

Acknowledgements

I would like to thank my supervisors Dr Kirk Rockett and Professor Dominic Kwiatkowski (DK) for your tutelage, unwaivering support without which this work would have been impossible to accomplish.

Thank you to our Oxford lab team, Anna Richardson, Rachel Craik, Christina Hurbbard and Kate Rowlands for your team spirit. Particularly Rachel and Christina for your help in planning some of my bench work. My sincere gratitude goes to the administrative team, Victoria Cornelius (Vikki), Kimberly Johnson, Andria Diss for their love and support. I would like to mention Vikki in particular for your overwhelming support for me throughout this journey. God bless you, Vikki.

Thank you to the DK group members at the Sanger Centre. To the lab folks, Sarah Auburn, Susana Campino, Daniel Alcocks, and Eleanor Drury, thank you all for the support and friendship. I must mention Bronwyn MacInnis for her pivotal role in ensuring that my samples were prioritized for sequencing.

I am thankful to members of our Statistics/analysis team, Jack O'Brien, Chris Spencer, Si Quang Le and Gavin Band for building my confidence in the Statistics minefield, especially Jack for Statistics teaching and for taking direct interest in my work. My heartfelt gratitude goes to Olivo Miotto, though based in Thailand, you nonetheless found the time to support and encourage me. Not to forget Taane Clark, Kerin Small and YY Teo for teaching me the important basics.

To the IT/informatics teams both in Oxford and the Sanger Centre, many thanks for the wonderful systems that your collective effort put at my disposal. In particular, thank you to Robert Hutton, Valentin Ruano-Rubio and Mangus Manske for your support and generosity of friendship.

I thank the malaria team at the Navrongo Health Research Centre, Ghana for supporting my field work. My gratitude goes to the lab team, especially, Immaculate Anati and Christopher Baluah for their support.

Last but not least, my special thanks go to Paulina Tindana, my wife for the support and sacrifices you had to make because of my DPhil studies. Paulina is a rock!

Attributions

The work undertaken in this thesis is my own original work. The contribution of others is acknowledged below and in the main text.

Chapter 3

Assistance with training of two technicians to help with my field work- Dr Sarah Auburn, member of the DK group, Wellcome Trust Centre for Human Genetics (WTCHG), University of Oxford.

Assistance with QRT-PCR- Mr Daniel Alcocks, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK

Chapter 4 and 6

Assistance with statistical analysis- Dr John O'Brien (Jack), DK group, WTCHG, Oxford.

Chapter 6

Genomic DNA for parasite clones: Robert Pinches, Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DS.

Abstract

***Plasmodium falciparum* population genetics in Northern Ghana**
Lucas .N. Amenga-Etego
Degree of Doctor of Philosophy
Wolfson College
Trinity term, 2012

The main thrust of this thesis was to characterize *P. falciparum* genetic diversity in northern Ghana. To do this, I used simple techniques to purify *P. falciparum* DNA from clinical samples across a rural setting for whole-genome sequencing. The goal was to provide a framework for exploring host-parasite genetic interactions. Utilizing Illumina deep sequencing data for 277 isolates I analyzed *P. falciparum* genetic diversity and described within-host diversity across this area. I observed random mating (ie no population structure) in the local parasite population, and a high genetic diversity indicative of high out-crossing. Moreover, when I aggregated my data with similar published data from Burkina Faso and Mali (sites $\approx$ 500km apart), no population structure was evident. In contrast, sites sampled in Cambodia and Thailand ($\approx$ 800km apart) were found to have greater population structure and high potential for inbreeding. This may be driven by differences in transmission intensity between the sites sampled in West Africa and Southeast Asia.

To demonstrate the utility of deep sequencing data, I focused on the genomic regions of *pfdhfr*, *pfdhps* and *pfcrt*, known to be under antimalarial drug selection. I surveyed the full diversity of point mutations already characterized in these genes and discovered previously unknown variants. However, in order to provide a means to follow up on new variants or interesting candidate regions in large clinical samples with limited parasite DNA, I assessed the Sequenom iPLEX platform for high-throughput genotyping of *P. falciparum* polymorphisms. This necessitated developing a method appropriate for assigning genotypes in haploid genome mixtures common in natural infections. Finally, I used this method to type host and parasite markers in a case-control sample set from this region for exploring host-parasite genetic interactions. I found that children who have the sickle-cell trait and carry parasites that have *pfdhfr* resistant alleles lose their protection against severe malaria as compared to children who have normal haemoglobin and are infected with parasites with these resistant alleles.

Contents

List of Figures

Chapter 1

Introduction

This introductory chapter describes the general epidemiology of malaria and sets out the background for this thesis by discussing the key malaria control intervention efforts to date, and the challenges. In addition, the chapter reviews the relevant literature, highlighting the recent revolutionary advances in high throughput genetics, and the opportunity it offers for a deeper understanding of parasite biology and the potential to enhance the global fight against malaria. Lastly, I outline the fundamental questions that stimulate the work presented in the rest of this thesis.

1.1 Malaria epidemiology

Plasmodium falciparum is the most important eukaryotic pathogen[81,103]. Despite huge investment in malaria control, falciparum malaria continuous to impose a

huge morbidity and mortality burden on endemic population[242]. Malaria results from infection with *Plasmodium* parasites, members of the *apicomplexans* family that includes parasites responsible for other important human diseases such as toxoplasmosis, babesiosis and cryptosporidiosis[218]. Human malaria parasites include *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale* and *Plasmodium malariae*. *Plasmodium knowlesi*, a species that naturally infects monkeys, is known to cause incidental infections in humans[206]. In addition, there is strong evidence suggesting the occurrence of two nonrecombining sympatric forms of *Plasmodium ovale*, known as *Plasmodium ovale curtisi and Plasmodium ovale wallikeri*[160,219].

Of these parasites, *P. falciparum and P. vivax* cause the most disease worldwide, and *P. falciparum* is responsible for most mortality[80,81]. In malaria endemic populations, the majority of individuals are infected at any given time, but a relatively small proportion show clinical symptoms that warrant treatment with antimalarial drugs[116,127]. A further even smaller proportion ($\approx 1\%$) present with severe life-threatening disease, mostly children under the age of five[212]. Despite the low prevalence of severe life-threatening malaria, there were an estimated 655,000 malaria deaths in 2010, of which 91% were in Sub-Saharan Africa (SSA) alone[242]. Severe malaria mortality primarily occurs among children under 5 years[241,242,250]. Though, *P. falciparum* severe disease phenotypes vary in different endemic populations, two commonly studied complications include severe malarial anaemia (SMA), defined as a haemoglobin concentration $\leq$ 5g/dL and cerebral malaria (CM), which may be defined as an extreme deep unrousable coma. The SMA phenotype tends to occur in areas of intense malaria transmission whilst CM is

more prevalent in areas of moderate to low malaria transmission[126,157].

The greatest burden of SMA is carried by young children and pregnant women[134]. The prevalence of SMA in SSA vary between 31% and 90% in children[134,184], and between 60% and 80% in pregnant women[134]. Malaria in pregnancy contributes substantially to the number of maternal deaths and infant deaths resulting from low birth weight[42]. In SSA, strong evidence support an age dependent pattern of severe malaria disease, with SMA being more common in infants and young children under three years and CM in older children[77,134,214].

1.1.1 *Plasmodium falciparum* Life Cycle

Human malaria parasites have a complex, multistage life cycle involving female members of the mosquito species complex *Anopheles* and vertebrate hosts[218] (Figure 1). Human malaria is transmitted by the female *Anopheles* mosquito. When a female mosquito takes a blood meal from an infected individual, gametocytes, the sexual stages of the parasite could be ingested in the process[26]. The gametocytes give rise to male and female gametes that fertilize in the mosquito mid-gut to produce diploid zygotes, which develop into motile ookinetes that lodge beneath the mid-gut wall as oocysts[192]. Exponential growth and division of each oocyst produces thousands of the infective haploid forms known as sporozoites[26].

The sporozoites break out of the mid-gut wall after about 8-15 days of development, and invade the salivary gland lobes of the mosquito where they remain within salivary ducts[192]. During subsequent blood feeding, the malaria sporozoites are injected along with saliva into the human host, and are carried via

blood circulation into the liver where they invade liver cells (hepatocytes)[192]. The parasites feed, grow and divide asexually within hepatocytes and mature over 5-16 days into schizonts, containing tens of thousands of haploid forms called merozoites[26]. On maturity, the hepatocytes rupture and release these liver-stage parasites- the merozoites, into the blood stream. Parasite multiplication then ensues in the red blood cells (RBCs) from repetitive rounds of invasion, growth and division in a two-day period[137]. Typically, *P. falciparum* develops over 48 hours producing several billions of parasitized RBCs in the bloodstream[218].

Clinical disease is exclusively due to the cyclical multiplication of these asexual parasites in RBCs[137]. The febrile paroxysms of malaria occur during this period of merozoite release from infected RBCs[81].

A small proportion of the merozoites leave the rounds of asexual multiplication and switch to develop over a period of about 2 weeks into mature male and female gametocytes[192]. How such an important switch is biologically achieved is unclear, but, what is certain is that these sexual forms are in readiness to complete the parasite life cycle when these gametocytes are taken up during a blood meal by another mosquito, thereby sustaining the cycle of malaria transmission[26].

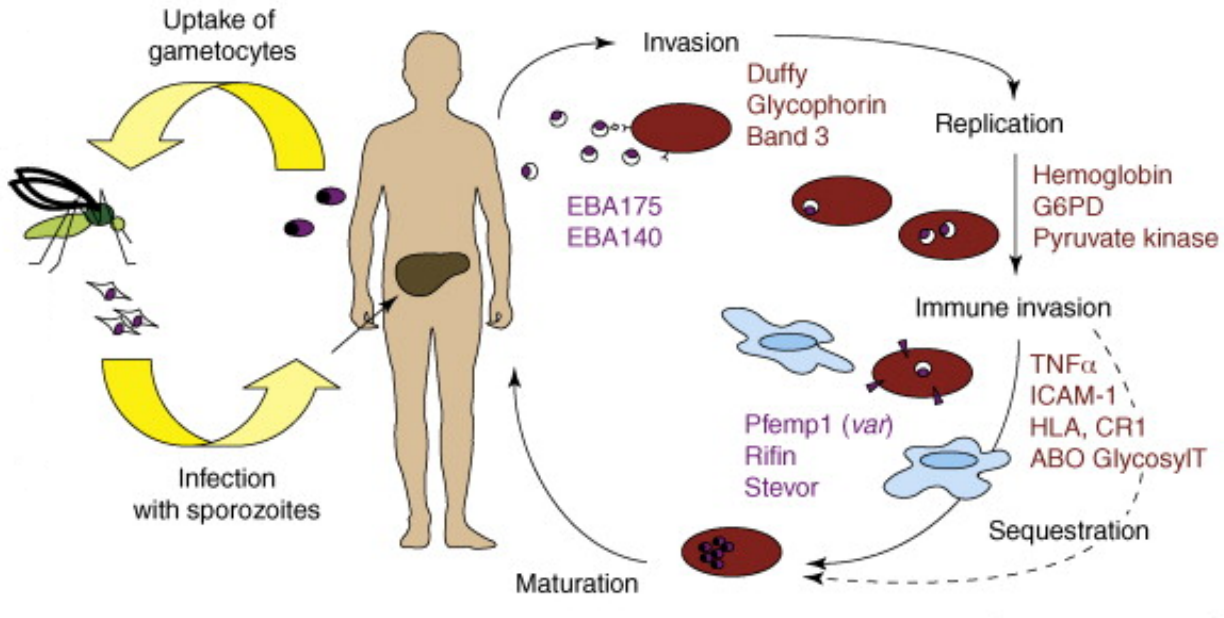


Figure 1.1: **Plasmodium falciparum Life Cycle**. Adopted from[26]. Also shown are various parasite and human host factors either implicated in invasion (i.e EBA-175 or 140, Duffy, and glycophorin band 3) or associated with parasite multiplication (haemoglobin, G6PD, pyruvate kinase), sequestration (ICAM-1, CR1) or severe malaria disease (TNFα, HLA type).

1.1.2 Complexity of natural infections

It can be deduced from the *Plasmodium* life cycle presented above that the mosquito's ability to successfully transmit the parasite may be as a result of several biting (or probing) attempts by the same or different mosquitoes. Thus, due to the temporary nature of this, it is possible for an individual to receive more than one infective bite, each with a genetically distinct parasite genome. Such a phenomenon is known as super-infection and is believed to be the main

factor generating multi-genome infections in areas of high malaria transmission[15]. It has also been shown that a single infective inoculation may also carry a diverse population of parasites[49], presumably as a result of feeding on an individual with a complex infection. Therefore, in areas of high malaria transmission, a child with typically 1% malaria infected red blood cells, an equivalent of 50,000 trophozoites µl (or ≈ 1250 trophozoites/200 WBCs)[228], may have tens of billions of parasites, which could be of diverse genetic type[227].

The number of parasite types (genomes) co-infecting a single host is traditionally known as multiplicity of infection (MOI)[169]. Conventionally, MOI is estimated by genotyping a relatively small number of polymorphic markers (eg MSP-1, MSP2 and GLURP)[159,169]. The MOI can be a useful indicator of the level of transmission and/or the immune status of the host[169,227]. Therefore, estimates of MOI tend to differ from one epidemiological setting to another, usually between 1 and 6 *P. falciparum* clones at any one time[19,159]. Increase in entomological inoculation rates (EIRs), that is, the number of infective bites that an individual is estimated to get in a year, is used to gauge levels of transmission and is generally postulated to correlate with the genetic complexity of infection[13]. Increased EIR estimates (ie intense transmission) is reported to be associated with a progressive rise in the average number of malaria parasite clones per host[19,39].

In areas of very high malaria transmission, MOI is age-dependent, with highest values reached in young children[19,99,210]. However studies from areas with low to moderate endemicity either reported little or no influence of age on infection complexity[17,145,254]. Multiplicity of infection also affects both the prevalence of parasite genetic markers such as those involved in antimalarial drug resistance[202],

and risk of clinical malaria[159].

Hence, the elusiveness of *P. falciparum* to control interventions and the devastating human and economic cost it exacts in endemic populations has many dimensions[213]. Its success can be attributed to its inherent evolutionary adeptness, the complexity of its life cycle shared between an invertebrate and a vertebrate host[26]. In addition, among ecological factors that promote its natural success, *P. falciparum* has over 5,000 genes with specialized proteins that enable it to survive in both intra- and extracellular environments, invade and grow in multiple cell types and evade host immune responses[81].

Subsection 1.1.1, alluded to the fact that malaria is a multi-genome disease involving the dynamic interaction of the human, parasite and mosquito genomes. Despite the important insight gained about each of these genomes over the years, several tiles remain missing in our understanding of how these genomes interact to ensure their evolutionary success. In the past, these genomes have been mostly studied independently, obviously because of the vast differences in the skill set required. However, the recent advances in high-throughput genetics, coupled with the success in building international research collaborations (or networks) across different specialities offer an opportunity to integrate research on all three genomes[151]. It is early days yet, but, the main thrust of my thesis is to explore the human genome and parasite genome and their interactions in this context. Conventionally, malaria control has focused on either knocking down populations of mosquitoes with chemicals or killing parasites *in vivo* in clinical cases with drugs, whilst exploring ways to exploit the genome interactions for vaccine development.

1.2 Malaria Control Interventions

Historically, global malaria control and elimination strategies have centered on two broad approaches: Interrupting malaria transmission via the control of vector mosquitoes using insecticides, and malaria case management with antimalarial drugs.

1.2.1 Interrupting malaria Transmission

Early attempts on a global scale by World Health Organization (WHO) to eradicate malaria in the late 1950s and early 1960s was through the use of dichlorodiphenyl-trichloroethane (DDT) in mass spraying exercises to kill mosquitoes en masse or radical cure through mass drug administration[220]. This initiative had some notable success though, it was largely unsuccessful in subtropical and tropical endemic regions as a result of a number of unforeseen obstacles that included weak national control programmes, emergence of resistance to DDT and sociocultural factors[81,132,220]. This notwithstanding, second generation insecticides such as pyrethroids (eg permethrin) have remained central in the global strategy to control malaria transmission. Various pyrethroid formulations including Insecticide treated-Nets (ITNs) are a key component of national programme strategies aimed at interrupting malaria transmission[81].

The effectiveness of ITNs was first demonstrated in four standardized large scale mortality trials in Kenya[201], Ghana[24], Burkina Faso[59] and The Gambia[6], all malaria endemic countries with different transmission intensities. These stud-

ies showed consistent evidence that ITNs could reduce child mortality between 17% and 33%[6,24,59,201]. Owing to this evidence, ITNs have taken centre stage in national malaria control programmes worldwide. They are advocated for and are now widely available through social marketing or free distribution to vulnerable groups such as pregnant women and children under five years by non-governmental organizations, notably UNICEF.

More recently, long-lasting ITNs, and indoor residual spraying coupled with artemisinin combination treatments (ACTs) have become pivotal in a renewed hope of malaria elimination in some SSA countries. A case in focus is the recent success of IRS in interrupting transmission in the Kwazulu-Natal province in South Africa[22].

1.2.2 Control of clinical malaria

The control of malaria disease has mainly been through the use of antimalarial drugs to manage clinical cases or prophylaxis to prevent disease. At the moment, no licensed antimalarial vaccine is available.

1.2.2.1 Antimalarial drugs

At the level of individual case management with antimalarial drugs, historically, chloroquine (CQ) was characterized by rapid efficacy and was widely used as an affordable first-line treatment for malaria in SSA[78]. The antifolate drug pyrimethamine and later a sulfadoxine-pyrimethamine (SP) combination, mar-

keted as a single dose therapy with almost no side effects, became a popular alternative to CQ for treatment of mild malaria in the early 1980s[81,194]. SP is also the current WHO recommended drug for Intermittent preventive treatment of malaria in pregnancy (IPTp)[216].

The global spread of resistance in *P. falciparum* against these previously effective and cheap first-line drugs, culminated in their withdrawal as first-line treatments for malaria in endemic countries[81]. Drug pressure was identified as a key factor in the emergence[40] and global spread of resistance[32,63,237]. Of note, however was an observed increased prevalence of CQ resistant alleles more rapidly in East Africa compared to West Africa[40]. This observation was supported by studies reporting CQ resistance levels of 70% in Kisumu, Kenya[25] and up to 44% clinical failure in at least one sentinel site in Uganda[146]. Meanwhile, around the same period, studies from West Africa reported comparatively lower levels of resistance to CQ, with CQ failure rates of under 20% from Burkina Faso[223], 23.2% in Ghana[55,102] and 14% in Mali[44]. Variation in drug pressure alone between these setting could not account for the observed differences. Hastings and D'Alessandro (2000)[86], suggested a role for transmission intensity, population immunity or people movement in attenuating the spread of CQ resistance in West Africa.

By the year 2000, CQ resistance had reached peak levels and the burden of malaria was spiraling, compelling the WHO to recommend the replacement of CQ and SP with artemisinin-combination therapy (ACT), where artemisinin or its derivatives were to be combined with appropriate partner compounds for the treatment of mild malaria in endemic countries[243,249]. The artemisinins are currently the world's most rapid and effective antimalarials, therefore, the ACTs

concept was aimed at protecting the artemisinins and their derivatives by making it more difficult for resistance to arise[40,47,248]. In this regard, recent compelling evidence of evolution of artemisinin resistance (slow parasite clearance) in Cambodia is a looming disaster[47,154]. Heritability studies support a genetic basis for the resistance phenotype conferred by this ongoing selection event[9]. Indeed, the introduction of this existing phenotype to Western Thailand was recently confirmed[33]. Other resistance foci have also been reported and are being actively evaluated[178]. There is eminent need to fully characterize the molecular basis for artemisinin resistance, and contain its spread in order to avert a momentous global disaster for malaria control[33].

1.3 Human genetics and malaria

Malaria has been responsible for the selection and maintenance of a number of human alleles in malaria endemic populations, with observed inter-ethnic differences in malaria susceptibility[38,92,108]. Over the last decade, the list of susceptibility alleles has rapidly expanded but reproducibility of results in replication studies remain a major draw back, particularly in African populations. The most evaluated candidates include the haemoglobinopathies, notably the sickle-cell trait and the thalassemias[57,244], the ABO blood groups[199] and other red blood cell mutations as well as polymorphisms in major histocompatibility complex class I genes[88], tumour necrosis factor (TNF-α)[133], and intercellular adhesion molecule 1 (ICAM-1)[60]. On the other hand, consensus evidence from multiple immunogenetic studies show the important role of cytokines in the pathogenesis and

clinical manifestation of malaria (reviewed in[37]). In this regard, higher baseline interleukin (IL)-10 plasma levels were found to correlate with higher parasite densities in Tanzania[89], and severe malaria phenotypes in Mali[114]. Khor *et al.,*(2010) showed that variants of the cytokine-inducible SRC homology 2 (SH 2) domain protein (CISH) gene are associated with susceptibility to diverse infectious diseases including malaria[98]. The molecular mechanisms that underpin these observations and genetic control, however, remain poorly understood. Overall, the haemoglobinopathies will be the main focus of this section as these are well evaluated and the sickle trait in particular is the most reproducible candidate polymorphism across many endemic populations.

Defects in haemoglobin synthesis that result in haemoglobinopathies; inherited disorders of haemoglobin structure and production were the first monogenic diseases in man to be linked to malaria protection[235,244]. The clinical effects of the haemoglobinopathies are variable. In their heterozygous state (carrier forms) all are benign conditions[244]. In their homozygous state, however, some, such as haemoglobin SS (HbSS) and β-thalassemia cause severe health problems and a reduced life-expectancy[87,161], others, exemplified by haemoglobin CC (HbCC) and haemoglobin EE (HbEE) are associated with mild clinical effects, notably haemolysis and splenomegaly, whilst homozygous α^+ -thalassemia, is clinically asymptomatic[57,244].

Although there are hundreds of structural variants of haemoglobins, some are widely known than others, for example HbS, HbC and HbE, with carrier frequencies reaching 15-20% in some parts of sub-Saharan Africa[235]. Normal adult hemoglobin is composed of two pairs of distinct globin chains ($\alpha_2\beta_2$; called

hemoglobin A or HbA). All of HbS, HbC and HbE affect the β-globin chain. In HbS, codon 6 of the β-globin chain has a single base change such that glutamine is substituted by valine ($\beta^{6Glu-Val}$), and in the HbC variant, an alternate mutation occurs that replaces glutamine with lysine ($\beta^{6Glu-Lys}$) also affecting codon 6, whilst the HbE mutation results in a change of glutamine to lysine ($\beta^{26Glu-Lys}$) at codon 26 of the β-globin chain[246].

The majority of case-control studies consistently support a protective advantage of HbS, over any other polymorphism[87,244], with over 90% protection of heterozygotes (HbAS) against severe life-threatening malaria and 60% protection against clinical malaria[246], and 60% protection against mortality in infants[4]. However, these individuals are not protected from asymptomatic parasitaemia[244,246] and the mechanism of protection remains a lively debate among proponents in the literature. The various positions include, the premature removal of infected red blood cells through some kind of innate immune mechanism[16], the involvement of a non-specific immune component as evidenced by an age-dependent increase in protection[245], and an accelerated acquisition of antibody responses to *P. falciparum* erythrocyte membrane antigen 1 (PfEMP1) by heterozgous individuals[31].

Population studies in Mali[3], Burkina Faso[141] and Ghana[140], all in West Africa, where the population frequency of HbC is high, found strong evidence of a malaria association[3,140,141]. In Burkina Faso a 29% reduction in clinical malaria for heterozygous (HbAC) individuals and over 90% reduction in homozygotes (HbCC) was evidenced[141]. Among the Dogon of Bandiagara, Mali, about 80% risk reduction for developing severe malaria was demonstrated for subjects carrying the HbC allele. The data suggest that the protection conferred by HbC may be

greater than that of HbS among the Dogon people of Mali[3].

Curiously, recent studies in Burkina Faso showed evidence implicating HbSS or HbCC with two-fold *in vivo* and four-fold *ex vivo* increased transmission of *P. falciparum* from the human host to the mosquito vector[76]. These findings have potential implication (e.g such as targeted strategies) for malaria control but have not been validated in other malaria endemic population.

1.4 *P. falciparum* population genetics

The majority of studies that have investigated the population structure of *Plasmodium* have done so from the perspective of capturing its global diversity and demographic history. Most of the early studies used microsatellite (MS) markers for their uniform distribution across the genome and their relative stability to gain insight[7,8,36]. The need for throughput and increased resolution led to the complementation of MS analyses with limited Sanger capillary[232] and short-gun sequencing[93], of a few tens to hundreds of culture adapted isolates[143,232]. These early studies also produced genetic diversity maps that were rapidly translated into SNP tiling arrays that brought a further boost to the genome-wide resolution of these analyses[43].

Overall, these studies provided evidence supporting geographical differentiation of *Plasmodium* at continental levels, with parasites sampled from Asia (Thailand, Cambodia and Papua New Guinea), Sub-Saharan Africa, and the Americas, showing extreme fixation in inter-population F_{st} estimates[7,8,36,93,232].

The global patterns of *Plasmodium* genetic diversity appeared to correlate with levels of transmission intensity, with highest diversity observed in African populations, intermediate in southeast Asia and lowest in South America[7,232]. Globally, *P. falciparum* allelic diversity, multiplicity of infection (MOI), out-crossing and gene flow are observed to be generally high in Africa, low in South America and intermediate in southeast Asia[7,36,121]. This is consistent with a geographical gradient of endemicity and transmission intensity[7,11].

Though the evidence was consistent across many studies, the small sample sizes made extrapolation to natural *Plasmodium* populations difficult. An additional source of skepticism was the fact that most of these early studies used culture-adapted parasites, that could have undergone *in situ* mutations during prolonged culture. There was eminent need to sample natural populations of *Plasmodium* and investigate this question at a high resolution and sample throughput.

To this end, the malariaGEN successfully used Illumina next generation sequencing technology[23,124], to sequence natural populations of *P. falciparum* from West Africa (Burkina Faso, Mali), East Africa (Kenya), and Asia (Cambodia, Thailand and Papua New Guinea). These were parasites extracted directly from patient blood without a culture step or in a few cases a minimal overnight culture[121]. The group developed informatics tools for mapping Illumina short reads against the 3D7 reference genome to call SNP variants[119,120].

Data from this initiative revealed further interesting parasite biology and consolidated not only the evidence of extreme continental differentiation previously observed, but also showed intra-regional variation in population structure and

Plasmodium genetic diversity in some endemic regions but not others[121].

1.5 Parasite genetic determinants of malaria

1.5.1 Antigenic variation and virulence

The erythrocytic stage of the parasite life-cycle is characterized by variant surface antigens (VSAs) that are variably expressed on the surface of infected erythrocytes (IEs)[109,153]. The VSAs are targets of antibody-mediated immune responses, resulting in clearance of the recognized parasites[29,30,126]. The most characterized VSA was first identified by Leech *et al.*,(1984) to be a family of polymorphic proteins termed *P. falciparum* Erythrocyte Membrane Protein-1 (PfEMP-1) encoded by *var* genes[111]. Any one parasite contains about 60 members of the *var* gene family. At any one time only one is expressed on the parasite surface (or surface of intact IE). It is the shuffling of expression of these genes that enable a subpopulation of parasites to evade the mounting immune response to previous variants[27].

Consolidated evidence demonstrate that one member of PfEMP-1 mediates adhesion of IE to variety of host ligands. It also contributes significantly to malaria pathology particularly, adhesion to microvascular endothelial cells and sequestration of IE within vital organs such as the brain may lead to cerebral malaria[109,153].

Besides mediating sequestration, another member of PfEMP-1 is known to mediate rosette formation[196]. Rosetting is a severe malaria phenotype resulting from *P*

falciparum- IEs binding to uninfected erythrocytes to form rosette-like clumps of cells associated with malaria severity in African children[195]. The parasite ligand for rosetting has been shown to be certain members of PfEMP-1[196]. Independent studies in many parts of Sub-Saharan Africa have consistently associated rosetting with severe malaria[152,172,193,195,198].

Other forms of the variant surface antigen PfEMP1, have been demonstrated to bind to chondroitin-4-sulfate A (CSA)[51]. CSA is present in placental intervillous spaces and is associated with pregnancy-associated malaria[191]. Two related forms of varCSA are known to be expressed by placental isolates[52], that could mediate the sequestration of IEs in the placenta, and result in life-threatening maternal malaria, especially first-time mothers in endemic countries[67,68]. At the moment, the mechanism behind the natural switching of *var* gene expression between members is not fully understood, but, the body of evidence support host immunity as a strong component of the factors that trigger *var* gene reshuffling[27].

1.5.2 Antimalarial drug resistance

Genetic diversity among contemporary *P. falciparum* parasite populations is another important virulence factor, particularly diversity at genes responsible for antimalarial drug resistance[43,225]. Genetically, resistance may develop when parasites undergo single nucleotide mutations or gene amplification events that confer a fitness advantage[2]. The best known of these genetic polymorphisms is encoded by the *P. falciparum* chloroquine-resistance transporter-gene (*pfcrt*) on chromosome 7[236]. The SNPs in this gene are known to confer resistance to the popular

quinoline-based drugs, such as chloroquine (CQ), mefloquine (MFQ), and quinine (QN)[226]. Notably, an amino acid substitution at codon 76 (K76T) is known to confer chloroquine resistance (CQR)[44,223,226]. Eight SNPs have so far been identified in this gene (M74I, N75E, K76T, A220S, Q271E, N326S, I356T, and R371I), the haplotype combinations of which distinguishes resistant and sensitive parasites[61].

Another gene that is important in resistance to the quinoline class of antimalarials is the *P. falciparum* multidrug resistance gene (*pfmdr-1*)[64,247]. The SNPs that have so far been characterized in this gene include the N86Y, Y184F, S1034C, N1042D, and D1246Y[66]. Epidemiological evaluations of these mutations, however, generated conflicting results; with the N86Y basal mutation associating with high-levels of CQR in some African parasites[20,44] whilst ironically correlating with increased sensitivity to MFQ in Thailand[185,186]. The other mutations appear to confer either susceptibility to MFQ or resistance to QN in different settings[62,147,179].

Amplification of the *pfmdr1* gene has also been associated with lumefantrine and MFQ resistance in southeast Asia[144,187]. Pickard *et al.,* (2003)[179] observed that increased gene copy number (greater or equal 3) was associated with decreased susceptibility to MFQ and artemisinin compounds. Further evidence of *pfmdr1* gene amplification and MFQ resistance, and reduced sensitivity to artesunate was provided by Price *et al.,* (2004) in a large study of parasite isolates from endemic communities in the Western Thailand borders[186].

The next most evaluated genes are the *dhfr* (dihydrofolate reductase) and *dhps*

(dihydropteroate synthase) that mediate *P. falciparum* resistance to the antifolate drugs pyrimethamine (P), cyloguanil (CY) and the combination drug sulfadoxine-pyrimethamine (SP). This category of antimalarials have well defined targets in the folate biosynthesis pathway. The DHFR and DHPS enzymes together with other enzymes catalyze the de novo synthesis and salvage of folate cofactors[2,234].

In *Plasmodium* species, DHFR exist as a bifunctional enzyme that includes thymidylate synthase (TS). This enzyme is sensitive to inhibition by P, CY and other derivatives of these compounds that bind strongly to the active site of the malarial enzyme[82]. In *P. falciparum* DHFR SNPs resulting in decreased affinities of drug binding to the enzyme were demonstrated to be the main cause of resistance[65]. Mutations have been found at various positions in the enzymes, including those at codons specifying amino acid residues 16, 50, 51, 59, 108 and 164[177,182].

The evolution of these mutations are postulated to be stepwise, beginning with the serine to asparagine mutation at codon 108 (S108N), which is known to result in moderate resistance (decreased inhibitor binding affinity)[207,208]. Subsequent mutations at other positions, namely, an asparagine to isoleucine at codon 51 (N51I), a cysteine to arginine at codon 50 or 59 (C50R;C59R), and an isoleucine to leucine at codon 164 (I164L) leads to progressively higher levels of resistance[2,82].

Resistance to combinations of DHFR and DHPS inhibitors such SP is observed to occur when the parasite has multiple mutations in the *dhfr* gene, exacerbated by additional mutations in *dhps*[182,224]. In close similarity with *dhfr*, mutations in *dhps* may occur sequentially beginning with an alanine to glycine change

at codon 437 (A437G), and followed by a variety of other mutations including mutations at codons 436 (S436A;S436F), 540 (K540E), 581 (A581G), and 613 (A613T;A613S)[2,224]. However, it is believed that the alanine to glycine (A437G) and lysine to glutamic acid (K540E) mutations are the primary and most critical mutations for SP resistance[2].

As noted earlier, the deployment of ACTs was aimed at prolonging the effective life of the artemisinins. However, evidence of the emergence of an artemisinin resistance phenotype has been identified and confirmed in Cambodia[47]. In addition, a recent study confirmed the spread of this phenotype to Western Thailand[33]. Though this trait has been demonstrated to be heritable[9], its genetic basis is not known, and the factors driving its spread are poorly understood.

It is anticipated that the recent achievement of next-generation sequencing of parasites from clinical samples, and the opportunity it provides for integrating deep sequencing data with epidemiological data from different settings will yield a deeper understanding of parasite genetic diversity and boost the search for the mutation responsible for the artemisinin resistance phenotype.

1.6 Advances in high-throughput genetics

Recent advances in high throughput genetic technologies have afforded whole genome sequencing of hundreds of samples with revolutionary impact on population genetics[122]. These next-generation sequencing (NGS) methods, have facilitated variant discovery, through whole genome sequencing and targeted re-

sequencing of biologically interesting genomic regions[136]. It is an exciting era for comparative genomics, and genomic epidemiology of leading disease such as malaria, with much anticipated expansion in the basic biology of medically important pathogens[123].

1.6.1 Next-generation DNA sequencing

Automated Sanger sequencing chemistry was the primary source of sequence data until 2004 when massively parallel next-generation DNA sequencing technologies became commercially available[122]. The three platforms that gained prominence include, the Illumina/Solexa Genome Analyzer, the Roche/454 FLX Pyrosequencer, and the Applied Biosystems SOLiD®Sequencer[10,122]. Two others, the Helicos Heliscope®and Pacific Biosciences SMRT technologies were later introduced[122]. Therefore, there is a variety of NGS features, and multiple platforms.

The backbone of these second generation technologies is a reliance on a combination of strategies that include template preparation, sequencing and imaging, and data analysis[136].The NGS techniques offer enormous volume of sequence data relatively cheaply, in some cases in excess of one billion short reads per instrument run[10,136]. The unique combination of specific protocols distinquish one technology from another and determines the type of data generated from each platform[136].

1.6.1.1 Template preparation

In general template preparation involves randomly breaking up genomic DNA into smaller sizes, from which either fragment templates[23], or mate-pair templates[136] are created. A common feature among NGS platforms is that the template is attached to a solid surface, which allows several thousands to billions of sequencing reactions to be done on the immobilized spatially separated sites[136]. Current NGS methods either use clonally amplified templates or single DNA molecule templates for NGS reactions.

Clonally amplified templates: Clonal amplification of templates is achieved by one of two methods; emulsion PCR[48] or solid-phase amplification[58]. A library of fragments or mate-pair targets is created, and adaptors containing universal priming sites are ligated to the target ends, allowing genomes of different complexity to be amplified with common PCR primers[136,204]. After ligation, the DNA is separated into single strands and captured onto beads under conditions favourable for one DNA molecule per bead (Figure 1.2 a)[136]. After the successful amplification and enrichment of emulsion PCR beads, millions can be immobilized in a variety of ways including deposition into individual Pico TitrePlate wells in which NGS reaction can be performed as pertains in the Roche/454 platform[110]. In the case of Illumina/Solexa platform, solid-phase amplification (Figure 1.2 b) is used to produce 100-200 million spatially separated template clusters, providing free ends to which a sequencing primer can be hybridized to initiate the NGS reaction[23,136].

Single-molecule template: One caveat of clonal amplification is the large

amount (3-20 µg) of genomic DNA material required[136]. However, single molecule templates are PCR free and require $<$ 1 µg of genome DNA as starting material[136].

Prior to initiating the NGS reaction, single molecule templates are usually immobilized on solid surface using one of three different approaches[204]. In the first, spatially distributed individual primer molecules are covalently attached to the solid surface[84]. The template which is prepared by randomly fragmenting genomic starting material into sizes of $\approx 200 - 300$bp, and adding adaptors to the ends, is then hybridized to the immobilized primer (Figure 1.2 c)[23,136]. In the second approach, spatially distributed single molecule templates are covalently attached to the solid surface by priming and extending single-stranded, single-molecule templates from immobilized primers (Figure 1.2 c)[136]. A common primer is then hybridized to the template (Figure 1.2 d)[136]. In either approach DNA polymorase can then bind to the immobilized primed template to initiate NGS reaction[84]. Both of these approaches are used by Helicos BioSciences. In a third approach, spatially distributed single polymerase molecules are attached to the solid support to which a primed template molecule is bound (Figure 1.2 e)[56].

1.6.1.2 Sequencing and imaging

Sequencing and imaging clonally amplified and single-molecule templates are fundamentally different, each with its unique challenges[136]. Sequencing by synthesis is by far the most popular and uses cyclic reversible termination (CRT) that involves nucleotide incorporation, fluoresence imaging and cleavage[135,136]. Firstly,

DNA polymerase bound to the primed template adds just one fluorescently modified nucleotide, which represent the complement of the template base[122]. The termination of synthesis after this step is the hallmark of this approach[136]. The remaining unincorporated nucleotides are washed away. Imaging is then performed to determine the identity of the incorporated nucleotide[23]. This is followed by a cleavage step, which removes the terminating group and the fluorescent dye (Figure 1.3 a). Additional washing is done before initiating the nucleotide next incorporation step[122,136]. The Illumina/Solexa NGS platform utilizes this approach with a four-colour sequencing by synthesis cycle (Figure 1.3 a), and Helicos BioSciences uses a one-colour (Figure 1.3 c) sequencing by synthesis cycle[136].

Currently the Illumina/Solexa variety of platforms dominate the NGS market[136]. It uses the clonally amplified template method illustrated in Figure 1.2 b, coupled with the four-colour CRT imaging method depicted in Figure 1.3 a. The four colours are detected and imaged using lasers techniques, the output of which is shown in Figure 1.3 b. The Illumina platform has an eight channel flow cell which allows independent samples to be run simultaneously.

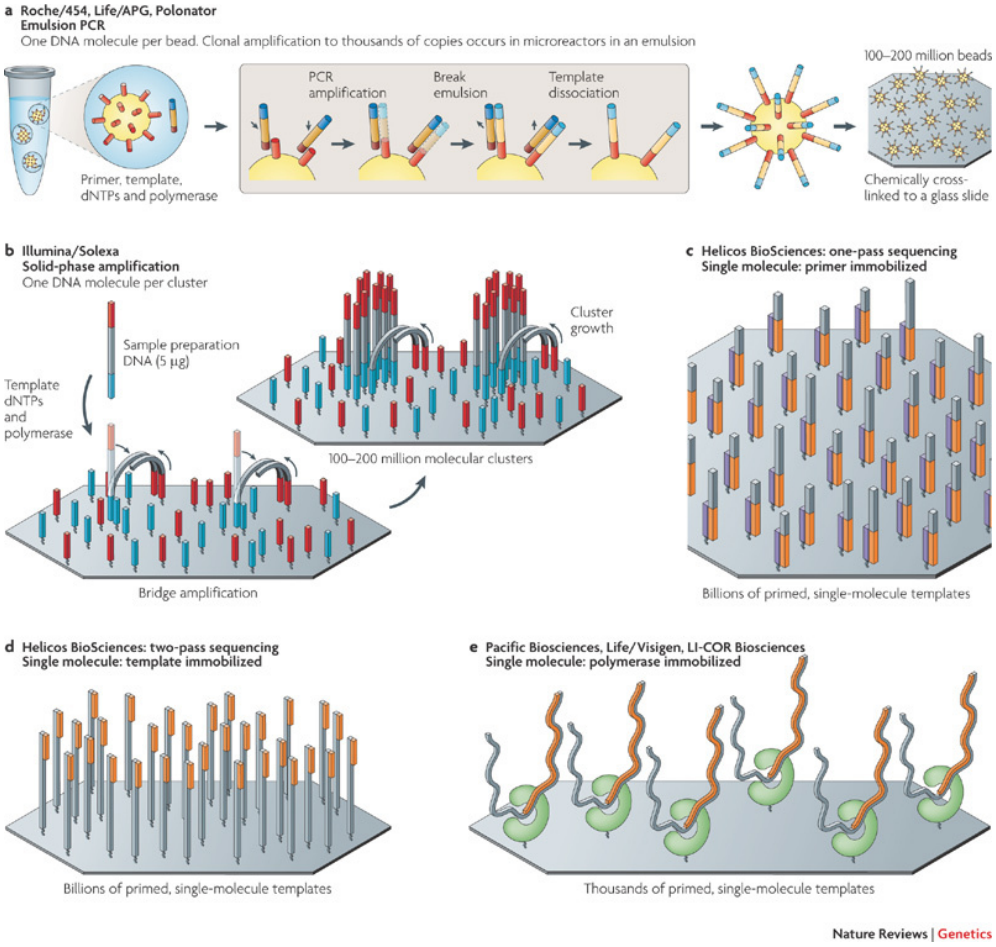


Figure 1.2: **NGS workflow: library construction ligates platform specific adapters to DNA fragments for immobilized amplification via a variety of approaches. Adapted from**[136]

In emulsion PCR (**a**), a reaction mixture consisting of an oil-aqueous emulsion is created to encapsulate bead-DNA complexes into single aqueous droplets. PCR amplification is done within these droplets so that beads containing several thousand copies of the same template sequence are produced. Solid-phase amplification (**b**) involves two steps: initial priming and extending of the single-stranded, single molecule template, and bridge amplification of the immobilized template with immediately adjacent primers to form clusters. (**c**) immobilization of template; and (**d**) immobilization of a polymerase. (**e**) 2′-deoxyribonucleoside triphosphate (dNTPs) incorporation.

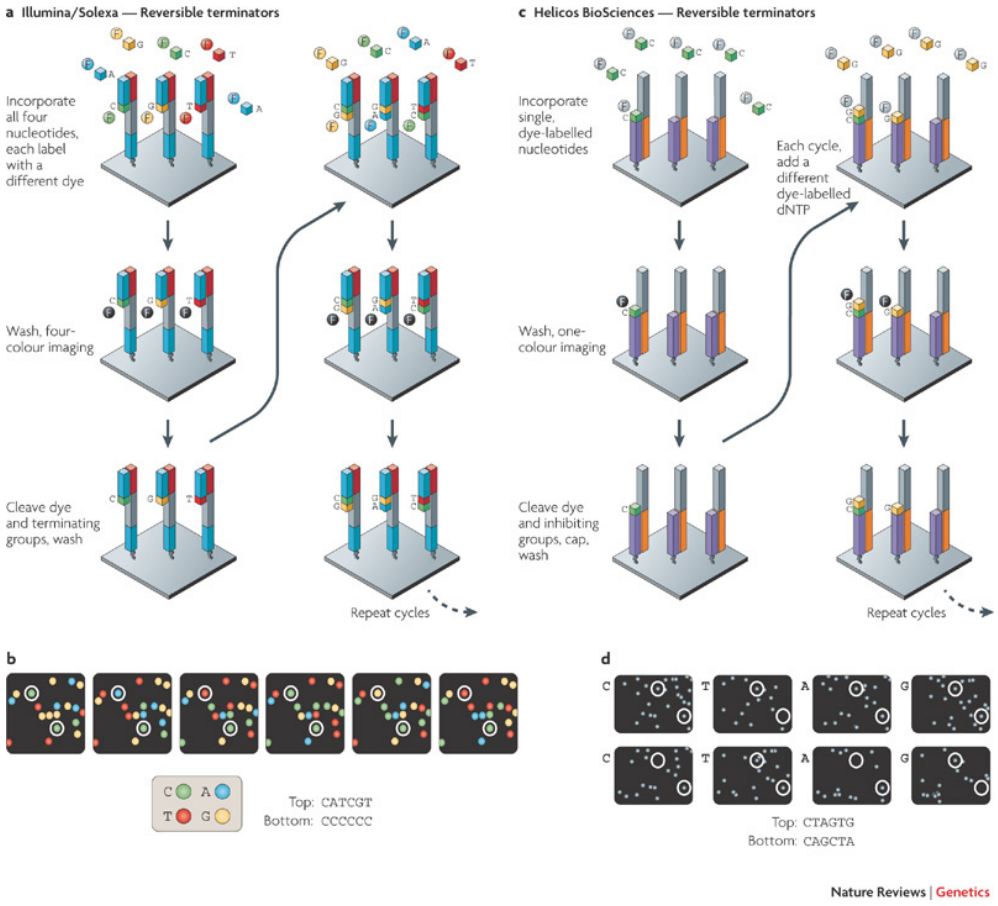


Figure 1.3: **NGS Imaging: Four- and one-colour cyclic reversible termination methods. Adapted from**[136]

(**a**); the four-colour cyclic sequencing by synthesis method utilizes Illumina/Solexa reversible terminator chemistry, using solid-phase amplified clusters. (**b**); The four-colour images highlight the sequencing data for two clonally amplified templates. (**c**); For Helicos Biosciences, terminators are labeled with the same dye and dispensed individually in a predetermined order. (**c**); The one colour-images highlight the sequencing data from two single-molecule templates.

1.6.2 Sequenom massARRAY iPLEX platform

Briefly, the Sequenom massARRAY platform (Sequenom®, San Diego, USA) is one other high-throughput technique, which has evolved from an expert-user-technology, to a routine genotyping method[94] . The Sequenom massARRAY iPLEX platform[70] is a validation and not a discovery technique that was originally designed for analyzing human SNPs[94]. It involves a step-wise achievement of throughput that may begin with genome-wide amplification of the target DNA. A highly multiplexed (up to 40 reactions per well) locus-specific PCR is then used to amplify the region of interest (Figure 1.4(**a**)), followed by single-base extension using mass-modified dideoxynucleotide terminators of an oligonucleotide primer[70].

The primer anneals immediately upstream of the polymorphic site and undergoes a single base extension (Figure 1.4 (**c**))[70]. Using Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry, the distinct mass of the extended primer identifies the SNP allele[95]. Genotypes are assayed real-time using the spectroTyper software[94]. The MASSarray iPLEX platform enables simultaneous processing of 384 samples, at a current cost of about 0.035 USD per SNP[70]. Thus, typically several thousands of samples for a few hundreds of SNPs can be assayed in a couple of weeks.

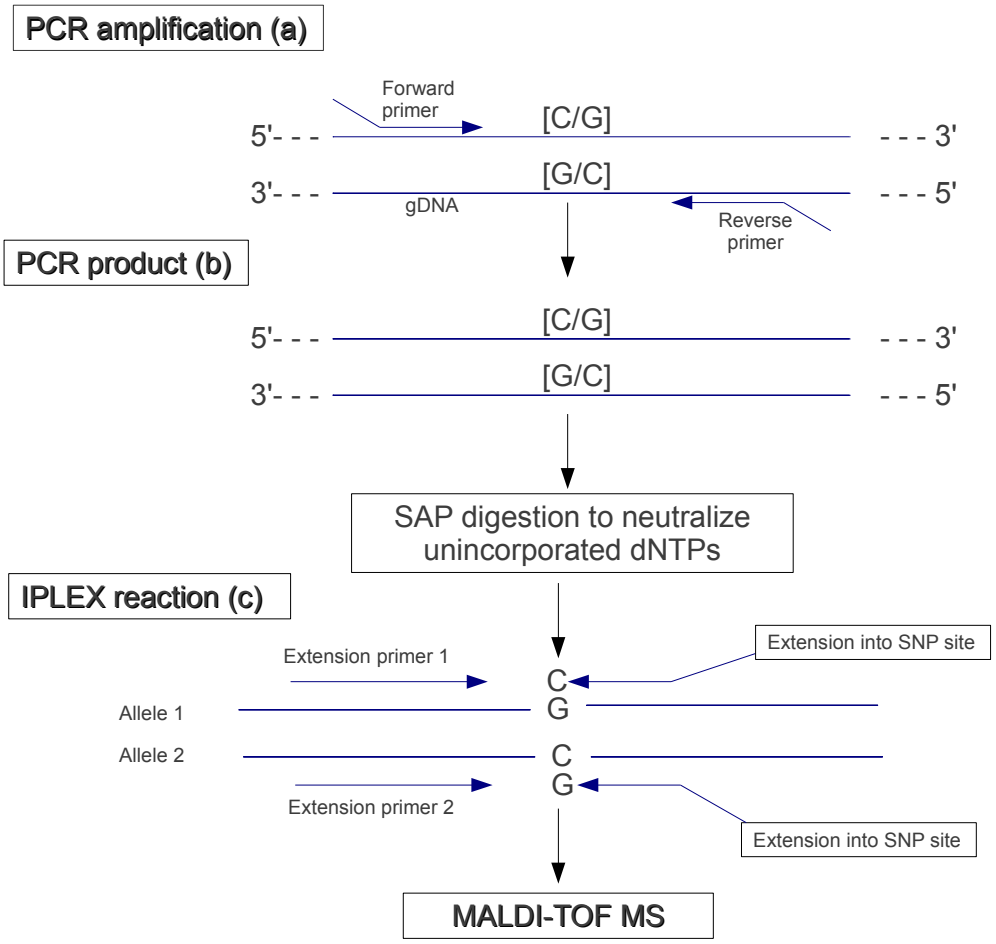


Figure 1.4: A schematic representation of the Sequenom iPLEX genotyping.

1.7 Motivation for work

Recent work on *P. falciparum* population genetics has focused mostly on global and regional level analyses and has demonstrated immense variability and the importance of historical events such as long-distance migration in the global spread of the parasite[9,121,232]. Recently high resolution Illumina sequencing data from clinical isolates consolidated the evidence on continental differentiation in *P. falciparum* populations[121]. What is fundamentally lacking is any deep understanding of the role of contemporary factors or environmental heterogeneity in structuring *P. falciparum* genetic diversity, and shaping transmission patterns at local levels in endemic communities.

Understanding how this parasite diversity is structured across populations is critical to unearthing the dynamics of parasite gene flow and is particularly informative in tracking phenotypes of interest such as novel antigenic variants, vaccine escape mutants or drug resistance emergence[121,144]. Antimalarial drug resistance remains the single most important threat to malaria control and elimination[33,47?] . The capacity to survey quickly, archived and/or new filter paper blood blots to confirm the prevalence of existing or new mutants in endemic populations is critical, but, currently lacking. High-throughput genotyping technologies such as the Sequenom iPLEX platform hold promise to this end. In addition, a complementary use of Sequenom iPLEX technology with the massively parallel NGS technologies will assure effective surveillance where new variants are identified and tracked in endemic populations.

The growing evidence of diverse genetic adaptations to malaria that have arisen

in different human populations[38,107], attest to the life history trade-offs in the host-parasite arms race[75]. Hence, it is possible that certain strains of parasites may be better adapted to some host polymorphisms[176]. In the light of this, there is enormous interest in host-parasite genetic interactions with the hope of deriving signatures that may be vital in malaria control and elimination[183,200,229]. The major roadblock, however, is the lack of a sampling framework that will permit population-specific testing of hypotheses involving parasite genetics, human genetics and their interactions. Previously, my work as a malariaGEN data Fellow, helped to set up a sampling framework for collecting clinical samples to study human genetics of severe malaria in the KNDs of northern Ghana[222]. Owing to my experience, when malariaGEN was embarking on Ilumina NGS of *P. falciparum* directly from clinical samples, it was opportune for me to focus my thesis on sampling and purifying *P. falciparum* DNA across our rural population for deep sequencing. Along with this project came new challenges including but not limited to engaging rural communities for pathogen genomics, obtaining consent for integrating the previous collections with this new one, and practical challenges of separating human white blood cells from infected red blood cells to reduce human DNA contamination. My thesis therefore, sought to purify *P. falciparum* DNA from clinical samples across rural communities in northern Ghana for whole-genome sequencing, and to provide a platform to integrate our previous case-control data with the parasite data for exploring parasite genetics, host genetics and their interactions in this well-characterized population.

1.7.1 General aims

1. To establish a sampling framework for purifying *P. falciparum* DNA at community health centres across the KNDs of northern Ghana for deep sequencing.

2. Analysis and use of deep sequencing data to characterize *P. falciparum* genetic diversity in the KNDs, Ghana.

3. Use of deep sequencing data to investigate the prevalence of SNP variation in genes known to be under antimalarial drug selection.

4. To assess the utility of the Sequenom iPLEX high-throughput genotyping platform to measure (type) *P. falciparum* polymorphims in field samples.

5. To explore the association of parasite genetic factors and host phenotypes and host-parasite genetic interactions in the KNDs.

Each aim forms the focus of each of my results Chapters. Within each chapter, I provide a brief introduction to the topic as well as a give more details of the specific objectives.

Chapter 2

Materials and Methods

2.1 The study site

The study was conducted in the Kassena-Nankana east and west districts (KNDs) (Figure 2.1, top panel), two adjoining districts of the Upper East region of Northern Ghana. The districts together cover an area of about 1675km^2 and lie between latitude 10.30′ and 11.10′ North and longitude 1.1′ West close to the Ghana-Burkina Faso border. The area is characterized by a short rainy season spanning July through October and a long dry season, giving an average rainfall of ≈1300mm and a Guinea Savannah vegetation. The KNDs population is estimated at about 153,000 inhabitants with population density of 91.5km^2 that is under continuous demographic surveillance by the Navrongo health and demographic surveillance system (HDSS)[158]. The major ethnic groups in the area are the Kassem and the Nankani with a minority Buli[157]. One of the main features of the area is a large reservoir in the district, which provides water for irriga-

tion throughout the year. There is one hospital, the Navrongo War Memorial Hospital (NWMH) that serves as a referral hospital for the population. There are also four Health centres strategically located across the districts to support the NWMH. Malaria epidemiology has been extensively characterized in this area of Northern Ghana[21,100,101,157,168,170]. Malaria transmission is intense, with seasonal fluctuation giving an overall entomological inoculation rate of 418 infective bites/person-year in 2003[12] (relevant to the 2002-2004 case-control studies discussed in Chapter 7). More recently, however, malaria transmission in sub-Saharan Africa has been on the general decline[164]. Recent unpublished data from the KNDs show an annual EIR of 140 bites per person-years, with seasonal variation of up to 170 in the rainy season and 77 in the dry season (Asoala et al., 2013 manuscript in preparation).

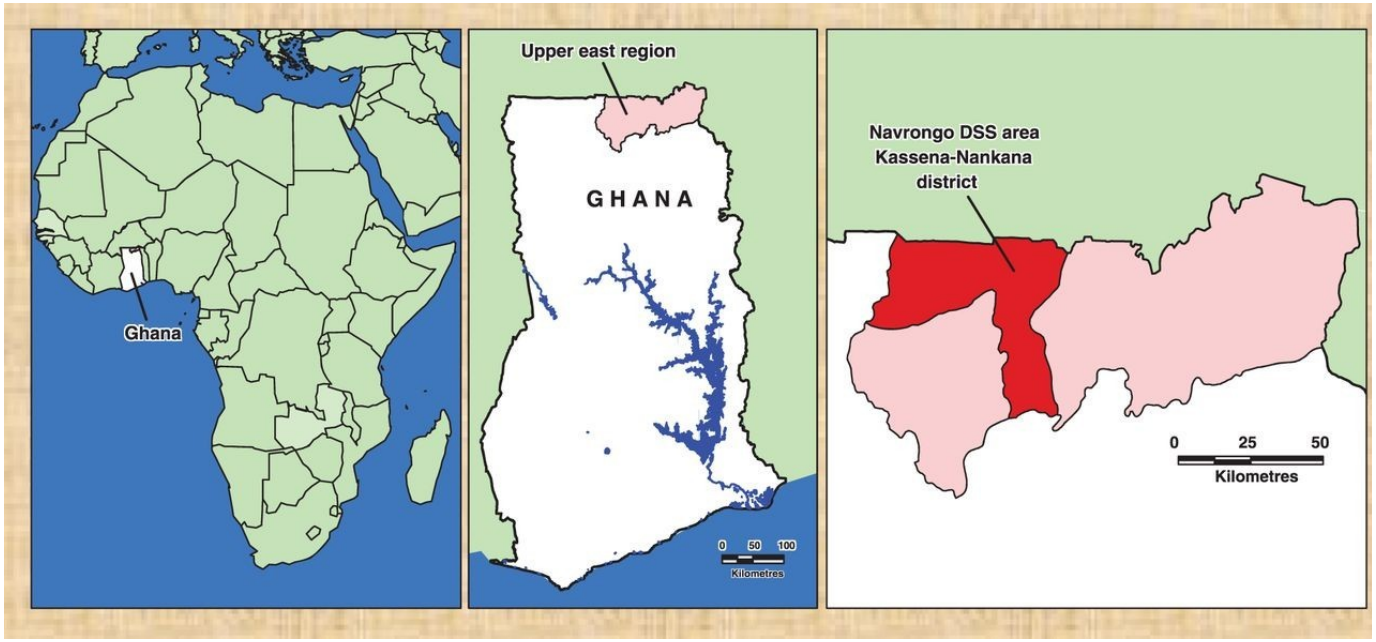


Figure 2.1: **Location of Ghana in Africa, the upper east region in Ghana and the Kassena-Nankana districts (KNDs) of the upper east region. Adapted from[158].**

2.1.1 Study design and population

I used a prospective longitudinal study design. For wholly practical reasons, I set out to implement on a broader scale field methods and protocols for removing leukocytes from whole blood and enriching parasite infected red blood cells for extracting DNA suitable for parasite sequencing. Since at the time, these methods were largely untested at a community-scale, the practicalities of sampling individuals within endemic communities was an important component of the experimental design. Therefore, in order to maximize success, I used a design that targeted malaria positive individuals presenting at community clinics. Therefore, potential participants were individuals who had uncomplicated malaria and children 0-5 years who were admitted to the NWMH paediatric ward and diagnosed with severe malaria.

2.1.2 Participant Recruitment

I collected *P. falciparum* parasites from all-age patients attending clinic in five locations in the KNDs from August 2009 to November 2011. These included, Chiana, Paga, The Kassena-Nankana East (KNE), and Kologo clinics and the out-patient department of NWMH. I screened patients reporting to clinic in these locations who had fever with Paracheck-Pf (Orchid Biomedical Systems, India), a rapid malaria diagnoses test (RDT) kit, which is based on *P. falciparum* specific Histidine-Rich Protein 2 antigen detection in peripheral blood[188]. The RDT positive individuals were then selected for further evaluation and possible inclusion into the study.

I used standard microscopy to assess parasite densities. A lancet stick of the finger was used to prepare thin and thick blood films. Thin films were fixed in methanol, and both thin and thick films were stained with 10% Giemsa for malaria parasites detection and quantification. Parasite density was scored as number of parasites per 200 white blood cells (WBCs) on a thick film. 200 high power fields of the thick films were examined at 1000x magnification to assign a negative result. Out-patients with parasitaemias ≥ 50 parasites per 200 WBCs (about 2000 parasites μl^{-1} of blood) were preferentially enroled.

Additionally, young children who were admitted to The NWMH with a diagnosis of severe malaria were included in the study. Criteria for diagnosis of severe malaria were those included in the standard WHO definition: (1) cerebral malaria (CM) defined as unrousable coma or Blantyre Coma Score of 3 or less; coma persists for more than 30 minutes after fits have ceased and normal CSF findings, (2) repeated or prolonged generalized convulsions, generalized convulsions lasting more than 30 minutes, or more than 2 convulsions in 24 hours despite cooling, (3) severe malarial anemia (SMA) defined as a hemoglobin concentration < 5 g/dl,
(4) respiratory distress- presence of alar flaring, intercostal or subcostal chest recession, use of accessory muscles of respiration, or abnormally deep respiration, (5) hypoglycemia; blood glucose < 2.2 mM/L and
(6) circulatory collapse- systolic blood pressure < 50 mm Hg (7) renal failure. In order to learn the impact of seasonal variation on parasite population dynamics, sampling was done across the high and low transmission seasons.

2.1.3 Ethical considerations

This study was reviewed and approved by the Navrongo Ethical Review Board (IRB) and Oxford Tropical Research Ethics Committee (OXTREC). In all patient categories blood samples were taken only after obtaining written informed consent. Individual consent was obtained for patients above 18 years (the legal age of adulthood in Ghana), both parental or legal guardian's consent and assent for children 12-17 years and parental or legal guardian's consent for minors (0-11 years). All consent forms were translated into the two major languages in the KNDs and administered alongside the English versions.

2.2 Materials and Methods

All materials that were used in the studies discussed in this thesis are list in appendix A.

2.2.1 Processing of whole blood

Plasmodium falciparum DNA was extracted directly from patient blood. Whole blood (2-3ml) was obtained from malaria patients via venipuncture in EDTA-containing Vacutainer tubes (Becton Dickinson). The blood was then processed as soon as possible with either CF11 filters or Plasmodipur® filters to remove white blood cells (mononuclear cells) prior to parasite DNA extraction.

The commercially available Plasmodipur® filters (Euro-diagnostica) were pur-

chased whilst the CF11 cellulose powder (Whatman®) was purchased and used to construct the CF11-based filter on site[215]. Plasmodipur® filters are designed to trap human nucleated cells and platelets while allowing parasite infected red blood cells (IRBCs) to be collected. They are relatively more expensive, and most studies tend to use more than one filter per sample.

CF11 powder on the one hand, is a fibrous, medium cellulose powder for general column chromatography that is readily available. It is relatively cheap and easy to use in assembling columns as filters for processing infected whole blood. Once assembled CF11 filters are portable and remarkably easy to use in filtering malaria infected blood under field conditions.

2.2.1.1 CF11 column construction

A 10ml centered outlet syringe (with the plunger removed) was tipped with either a 1.0cm^2 piece of Grade 105 lens cleaning tissue (Whatman) or a thin layer of cotton wool. The tissue or disc of cotton wool was carefully placed in the syringe so as to cover the outlet lumen. CF11 cellulose powder was then loosely added to the syringe up to the 10 ml mark and then packed down to about 5.5ml of packed cellulose using the plunger. When ready to use, 5 µl of isotonic phosphate buffered saline (PBS) solution (pH 7.3) was first run down the column to wet it.

2.2.2 Lymphoprep density-gradient centrifugation

One blood volume of Lymphoprep (Axis-shield) was added to a 15ml centrifuge tube. Whole blood samples were diluted with an equal volume of isotonic phos-

phate buffered saline (PBS) solution [pH 7.3] (Sigma) and carefully layered over the Lymphoprep, taking absolute care to layer and not mix. A maximum of 5ml diluted blood over 2.5ml Lymphoprep was used per tube and few larger blood volumes split accordingly. The samples were then centrifuged at 800g for 20 min at 20-22°C. The mononuclear cells, which form an obvious band at the sample/PBS medium interface were aspirated and discarded. The RBC pellet was re-suspended in 10ml PBS to wash and centrifuged at 800g for 5 min. The supernatant was aspirated and discarded and the RBC pellet was used for further processing by Plasmodipur filters.

2.2.3 Plasmodipur filtration

Plasmodipur filters (Euro-diagnostica) were pre-wet with 5ml PBS. The RBC pellets obtained from Lymphoprep density-gradient centrifugation were diluted in 3 volumes of PBS and drawn into a 5ml syringe before mounting onto a Plasmodipur filter. The syringe and Plasmodipur unit was then propped over a 50ml falcon tube. Gentle pressure was applied to the syringe and the diluted blood was gently and carefully passed through the Plasmodipur filter and collected into the falcon tube. An additional 10-20ml of PBS was used to wash through any RBCs remaining in the filter. The filtered blood was centrifuged at 800g for 10 min at 20-22°C. The supernatant was discarded and RBC pellet stored at -20°C until DNA extraction.

2.2.4 CF11 column filtration

Whole blood samples were diluted 1 volume of PBS, and the post-Lymphoprep RBC pellets were resuspended in 3 volumes of PBS and carefully added to CF11 columns that were pre-wetted with 5ml of PBS, and propped over 50ml falcon tube. The blood was then allowed to filter gradually by gravity. The filtered blood was centrifuged at 800g for 10 minutes at 20-22°C and the plasma supernatant was discarded and RBC pellet stored at -20°C until DNA extraction..

2.2.4.1 DNA extraction and Quantification

DNA extraction was done using the QIAamp DNA Blood Midi/Maxi kit (Qiagen) as per manufacturer's protocol. Qiagen protease (proteinase K) [120µl] (for $\leq$ 1.0ml of whole blood) or 220µl (for 1-2ml of whole blood) was added to a 15ml centrifuge tube. The RBC pellets were equilibrated to room temperature (15-25°C) and diluted with PBS to 1 or 2ml and added to the Qiagen protease. Buffer AL (1.2ml) or 2.4ml was further added and mixed thoroughly by inverting at least 15 times, followed by additional vigorous shaking for 1 min to ensure adequate lysis. The tube was incubated at 70°C for 10 min. One millilitre or 2ml ethanol (96-100%) was added to the sample and mixed thoroughly by inverting for 10 min, followed by additional vigorous shaking to ensure efficient binding with the DNA. All or one-half of the solution was carefully transferred onto a QIAamp Midi column placed into a 15ml centrifuge tube, taking care not moisten the rim. The solution was centrifuge at 3000rpm for 3 min. The filtrate was discarded and the column placed back into the 15ml centrifuge tube. Buffer AW1 (2ml)

was carefully added to the column, taking care not to moisten the rim. The solution was centrifuged at 4000 rpm for 2 min. Buffer AW2 (2ml) was carefully added to the column, again taking care not to moisten the rim. The solution was centrifuged at 4000 rpm for 15 min. The column was incubated at 70°C for 10 min to evaporate residual ethanol. The column was placed in a new 15ml tube whilst the collection tube was discarded together with the filtrate. Buffer AE (200µl) or 300µl was pipetted directly onto the QIAamp Midi column, incubated at room temperature for 5 min, and centrifuged at 4000 rpm for 4 min. For maximum DNA yield, a fresh 200µl or 300µl of buffer AE was pipetted onto the QIAamp Midi column membrane; further incubated at room temperature for 5 min, and centrifuged at 4000 rpm for 4 min.

Note: QIAamp Maxi columns were used for whole blood volumes greater than 2.0 ml.

DNA was quantified using the Quant-iT^{TM} PicoGreen dsDNA Assay kit (Molecular probes-invitrogen) and stored at −20°C.

2.2.4.2 Quantitative Real-time PCR

Quantitative Real-time PCR (QRT-PCR) was used to quantify the amount of human DNA remaining after WBC-depletion and the parasite DNA yielded. QRT-PCR was undertaken on 2-5ng of each DNA sample using the Applied Biosystems StepOne PCR System with human primers and *Plasmodium-specific* primer sets as shown in Table 2.1. Human and *P. falciparum* DNA negative controls were added in the respective reactions to confirm primer species-specificity. Each primer set was unique within the respective genomes. Pure human and pure *P.*

falciparum standards were prepared to range 0.1-100ng/µl concentrations. Test samples were then diluted as required to fit this range, and all samples and standards were tested in duplicate.

One µl DNA was added to a 24µl PCR solution containing 8.5µl water, 12.5µl PCR Sybergreen Master Mix and 1.5µl of each primer (Stock at 2µM). The mixture was amplified for 5 cycles of 94°C for 45 seconds, 56°C for 45 seconds, 72°C for 45 seconds, followed by 30 cycles of 94°C for 45 seconds, 65°C for 45 seconds (*data collection point*), and 72°C for 45 seconds. The data obtained was analysed using the Applied Biosystems StepOne version 2.0 software.

Target	Gene	Primer pairs (5′-3′)
Human	TLR-9 (F)	ACGTTGGATGCAAAGGGCTGGCTGTTGTAG
	TLR-9 (R)	ACGTTGGATGTCTACCACGAGCACTCATTC
Parasite	EBA-175 (F)	ACGTTGGATGCACCAGTGAAGAAACTACAG
	EBA-175 (R)	ACGTTGGATGCTTCATATTCCTTAGTAAGCG

Table 2.1: Primers for QRT-PCR. The first 10 bases for each primer are the same because these are sequenom first round primers known to work together very well that I used for this exercise.

2.2.5 Sequenom iPLEX Genotyping

The MassARRAY iPLEX platform (Sequenom, San Diego, USA) was used for assessing the *P. falciparum* SNPs discussed in chapters 4 and 5 (section 6.7), and the human SNPs discussed in chapters 7. This platform, however, was originally designed for genotyping human polymorphisms. MassARRAY SNP genotyping combines the benefits of accurate primer extension chemistry with the detection of extension products by Matrix Assisted Laser Desorption/Ionization-Time of Flight(MALDI-TOF) Mass Spectrometry (MS). Multiplex design for the iPLEX methodology was achieved using the MassARRAY Assay Design Software (version 3.1), which has a design capacity of up to 40 SNPs per reaction.

2.2.5.1 Primer Extension Pre-amplification

Prior to genotyping whole genome amplification of all DNA samples was done using the Primer Extension Pre-amplification (PEP) method[253]. The PEP method utilizes 15 base random sequence primers (N15) to amplify the entire genome. Typically, PEP products are usually of the order of 600 - 800bp in length.

Five nanograms of gDNA was amplified in 96-well PCR plates in a final reaction volume of either 25 µl or 50 µl. The PEP reaction mix of 50µl contained 0.04nmol/µl of each oligonucleotide primer (Genetix, Cambridge, United Kingdom), 1x buffer (Bioline, Cambridge, United Kingdom), 2.5 mM $MgCl_2$ (Bioline), 0.2 mM each of dNTPs (Bioline), and 2.5 units of Taq polymerase (Bioline). Extracted DNA (1ng/µl) was used as template for the amplification reaction. The reaction mix was thoroughly mixed by vortexing, followed by centrifugation at

1500 rpm for 1 minute (min). The temperature profile for the reactions were 94°C for 3 min, followed by 50 cycles of 94°C for 1 min, 37°C for 2 min, and then slowly increasing the cycling temperature by 0.1°C each cycle to 55°C, and held for 4 min and then a final extension step of 72°C for 5 min.

2.2.5.2 Primer extension and detection by MALDI-TOF MS

Sequenom's iPLEX MassEXTEND is a SNP genotyping procedure that is based on two levels of specificity. First, a locus-specific 'first round ' iPLEX PCR reaction is done to amplify the region of DNA flanking the SNP of interest to produce unique targets (typically 100 base pairs in length) in sufficient concentrations to enable SNP capture.
Second, a locus-specific primer extension reaction in which an extension primer hybridizes immediately upstream (3′) of the polymorphic site and is extended based on one of four complementary mass-modified terminator molecules that is specific to the sequence of the variant site. For example: genotyping a G/T SNP with the termination mix dd**T**TP/dd**A**TP/dd**C**TP/dd**G**TP, a **G** change will result in a single base extension C, and a **T** change will result in a single base extension A. In like manner it is feasible to genotype tri-allelic variations in SNP. The mass differences of the extended primer can be determined by MALDI-TOF mass spectrometry[70]. Thus, allowing for the differentiation of alleles at any given SNP site in DNA, and homozygous and heterozygous samples in diploids.

The sequences and other essential design data for the first round PCR primers and the extension primers are listed in Table 2.2. All PCR primers were initially tested to ensure that a clean single band product of the required length is achieved. The

iPLEX PCR reaction mix per samples was 0.45µl primers, 0.49µl of 25mM $MgCl_2$, 0.15µl of 25mM dNTPs, 0.94µl of 10x PCR buffer (Bioline), 0.3µl of 5units/µl Taq DNA polymerase (Bioline), 2.18µl of milliQ water, and 3µl of 1ng/µl DNA. The final reaction volume was 7.5µl. The cycling conditions were 94°C for 15 minutes, 94°C for 20 seconds, 56°C for 30 seconds, 72°C for 1 minute, then 44 cycles of 94°C for 20 seconds, 72°C for 3 minutes, 15°C for 15 minutes. All PCR reactions were carried out in 384 well plates and all thermal cycling was on MJ Tetrad (Bio-Rad) and the heated sealed lid option was used to avoid evaporation. All PCR products were resolved on a 3% agarose gel.

The iPLEX PCR reaction products were further processed to remove all excess dNTPs using iPLEX shrimp alkaline phosphatase (SAP). A multimek robot at the Core Genomics Group, Wellcome Trust Centre for Human Genetics, Oxford, was used to transfer 2µl of SAP mix into each well of the 384 well plate (containing 5µl of PCR product) such that each sample contained 1.53µl of SAP 10x buffer (Bioline), 0.17 milliQ water, 0.30µl of 1.7units/µl of SAP. The SAP mix was incubated at 37°C for 40 minutes (*Digestion*), followed by SAP inactivation at 87°C for 5 minutes, and 15°C for 15 minutes.

AT this stage, the PCR products are now ready for primer extension and termination reaction that incorporates one of four dd**N**TPs. Single primer extension over the SNP was carried out in a final concentration of between 0.625µM and 1.5µM of each extension primer, depending on the mass of the probe, iPLEX termination mix (Sequenom) and 1.35 units of iPLEX enzyme (Sequenom) and cycled using a two-step 200 short cycles program; 94°C for 30 seconds, 94°C for 5 seconds, 52°C for 5 seconds, 80°C for 5 seconds, followed by 4 cycles of 52°C

for 5 seconds, 39 cycles of 94°C for 5 seconds, and then 72°C for 3 minutes, and 15°C for 15 minutes.

The reaction was then desalted (removes Mg^{2+}, ddNTPs and dNTPs) by addition of 6mg cation exchange resin (SpectroCLEAN resin, Sequenom) followed by mixing and centrifugation to settle the contents of the tube. The extension product was then spotted onto a 384 well spectroCHIP before being subjected to the MALDI-TOF mass spectrometry for genotype calling in real time. The Sequenom SpectroTYPER software was used for visualization of the genotype calls and making an initial assessment of the homozygous and heterozygous resolutions for the human SNPs . The genotype data was then exported as a Microsoft Excel file for further processing and joining with meta data for analyses.

Assay	**Description**	**Sequence (5′-3′)**
DHFRA16V	First-round	ACGTTGGATGAACAAGTCTGCGACGTTTTC
		ACGTTGGATGGATTCATTCACATATGTTG
	Extension	TCGATATTTATGCCATATGTGC
DHFRC59R	First-round	ACGTTGGATGAACAAGTCTGCGACGTTTTC
		ACGTTGGATGGATTCATTCACATATGTTG
	Extension	GATTCATTCACATATGTTGTAACTGCACA
DHFRI164L	First-round	ACGTTGGATGTCTTGATAAACAACGGAAC
		ACGTTGGATGGATCTAATAGTTTTACTTGGG
	Extension	TAAACAACGGAACCTCCTAA
DHFRN51I	First-round	ACGTTGGATGAACAAGTCTGCGACGTTTTC
		ACGTTGGATGGATTCATTCACATATGTTG
	Extension	GAGTATTACCATGGAAATGTAA
DHFRS108NT	First-round	ACGTTGGATGAGGTTCTAGACAATATAAC
		ACGTTGGATGCAAAATGTTGTAGTTATGGG
	Extension	TGGAATGCTTTCCCAGC
DHPSA581G	First-round	ACGTTGGATGGTGGATACTCATCATATAC
		ACGTTGGATGGGATACTATTTGATATTGG
	Extension	TTAATAGATTGATCATGTTTCTTCC
rs45243608	First-round	ACGTTGGATGTGAAGGTGCTAGTGTTATAG
		ACGTTGGATGGGTACTACTAAATCTCTTTC
	Extension	GATATAGGTGGAGAATCCTC
rs45319633	First-round	ACGTTGGATGGTAGCTAGTATAATATTTATC

Continued on next page

Table 2.2 – continued from previous page

Assay	Description	Sequence (5′-3′)
		ACGTTGGATGGAAAATTATTTATTAGAATG
	Extension	TAAATTATTTTTCATTATTTCGTCATTC
rs45239511	First-round	ACGTTGGATGCTAAATCACTTGTATCTAC
		ACGTTGGATGCATATTGGATGAGGAAGAAG
	Extension	ggAATCACTTGTATCTACTGTATCA
PfcrtK76T	First-round	ACGTTGGATGAACATAATCATACAAATAAAGT
		ACGTTGGATGTCGACCTTAACAGATGGCTCAC
	Extension	GTTTAAAGTTCTTTTAGCAAAAATTG
rs45241656	First-round	ACGTTGGATGAAACCTTCGCATTGTTTTCC
		ACGTTGGATGTCCTCTTGTATGTATCAACG
	Extension	CTTCGCATTGTTTTCCTTCTTTA
rs45241660	First-round	ACGTTGGATGCTTTTTAATTTTATAGGGTG
		ACGTTGGATGTCACACTTACCAAAGTTACG
	Extension	AATTTTATAGGGTGATGTTGTAAG
rs45244585	First-round	ACGTTGGATGGAAGGATCCAAACCAATAGG
		ACGTTGGATGTGTAAATGCAGCTTTATGG
	Extension	AATAATTGAGCGCTTTGACA
rs45254309	First-round	ACGTTGGATGCTCCTGAATCTTCTTGTTGG
		ACGTTGGATGCAACGAACATAGGGAACATC
	Extension	TCTTCTTGTTGGTATGTATGTTC
PfAMA1-R512K	First-round	ACGTTGGATGGGAATATCTGCATATTCATC

Continued on next page

Table 2.2 – continued from previous page

Assay	Description	Sequence (5′-3′)
		ACGTTGGATGTTTCTTTGTATGTAAATGTG
	Extension	ATGTTACTTCTGCCCTTC
rs45296956	First-round	ACGTTGGATGCAGCAAACAATGATGGTGAA
		ACGTTGGATGTATTGTACTTTACTTTTTGG
		ACGTTGGATGGAAACACGTCCACATGGTAG
	Extension	AAAAAGGATATTTTGTGATTTAAAAAA
rs45318952	First-round	ACGTTGGATGAGCATCAGTTTCAGGTCTTG
		ACGTTGGATGAGGACCAAAAGGAAATGAA
	Extension	TTACTCAAACTATCGTCATCACGTTC
PfEBA175G1100D	First-round	ACGTTGGATGCTTCATATTCCTTAGTAAGCG
		ACGTTGGATGCACCAGTGAAGAAACTACAG
	Extension	TTAGTAAGCGTTCTCTCAC
PfEBA175P390S	First-round	ACGTTGGATGCTTTAGCATCATTCTTGTT
		ACGTTGGATGGGCATACGTTATCGAAAGAA
	Extension	TTTCCGCATTTTCCTTTGA
rs45323313	First-round	ACGTTGGATGGTCAAACACAAATACATGTTC
		ACGTTGGATGTGTCTTCAGAACATGTTTC
	Extension	TGTCAATATATGGAAAGAAATAAGA
		TGTCAATATATGGAAAGAAATAAGT
rs45323318	First-round	ACGTTGGATGCAATATCCTTACACTCAGAAC
		ACGTTGGATGCGATATACTATATTTAGGAG

Continued on next page

Table 2.2 – continued from previous page

Assay	Description	Sequence (5′-3′)
	Extension	AAATCCAAAGCATTATCAGGG
rs45326723	First-round	ACGTTGGATGTATTCCTCTTAGCACCACTC
		ACGTTGGATGCTAGTGATAGTAAGAAAATG
	Extension	gTAGCACCACTCCTACTAC
rs45328613	First-round	ACGTTGGATGCCTGGCTCTGCTAATAAACC
		ACGTTGGATGCGACATTAAACACACTGGAAC
	Extension	acTAAACCTAAAGACGAATTAGATTATGC
rs45333817	First-round	ACGTTGGATGCCATTTTTTCCTATTTTCCGC
		ACGTTGGATGGACATGGAAGGACTTGGTAG
	Extension	TTTCCTATTTTCCGCAGATTTCTC
rs45333829	First-round	ACGTTGGATGCGGGTCGCTTGAAAGTGATA
		ACGTTGGATGGGATACTTTAATGTATTAGCG
	Extension	AGATGATATAAGTAATATAGAAGGAGA
PfMSP1T1667K	First-round	ACGTTGGATGTAATGGTGGATGTGATGCAG
		ACGTTGGATGTCAGGTTTAGTACATTCAC
	Extension	ATGTGATGCAGATGCCAC
PfTRAPH488Y	First-round	ACGTTGGATGTGGCTTTTCATGTTCTTCCC
		ACGTTGGATGGAAACACGTCCACATGGTAG
	Extension	ATGTTCTTCCCTTTCAGGATA
PfTRAPL277I	First-round	ACGTTGGATGGAAATCTTACACGAAGGATG
		ACGTTGGATGTGGAACATCTAATGGTTCCC

Continued on next page

Table 2.2 – continued from previous page

Assay	Description	Sequence (5′-3′)
	Extension	CGAAGGATGTACAAGTGAAA
PfTRAPV437I	First-round	ACGTTGGATGGGGTCACTTTGTTTCCTTTC
		ACGTTGGATGCCAAATGATAAAAGTGATAG
	Extension	TTGTTTCCTTTCATTATCCAAAAC

Table 2.2: **Primers for *P. falciparum* Sequenom Genotyping**

2.2.6 Data Analysis

The open source R statistical software[189] was used for all the statistical analyses undertaken in this thesis. I implemented a variety of statistical test in R, including Chi-square and Fisher's exact tests to assess deviations from the null-hypotheses. In all cases, a two-tail 95% confidence interval was used to ascertain the level of statistical significance.

Chapter 3

Establishing a framework for purifying parasite DNA

3.1 Introduction

Recent advances in high-throughput genotyping technologies such as NGS present a unique opportunity to examine questions in host and parasite genetics[122]. Particularly, the ability to use these methods to sequence parasite DNA extracted directly from patient blood will pave the way for a better understanding of 'natural' parasite genetic diversity and parasite factors responsible for disease outcomes[14,121]. The potential exists to correlate such deep-sequencing data with human host genetic diversity data to gain understanding of host-parasite genetic interactions. What is required, however, is an appropriate community-based sampling framework that could set the stage for addressing relevant epidemiological questions. Following on from my experience with setting up a framework for sampling clinical cases to study human genetic association with severe malaria[222], I

set-up a similar framework for purifying *P. falciparum* DNA from clinical samples for deep sequencing. This chapter details the activities I undertook to achieve this in endemic communities in Kassena-Nankana district of northern Ghana. Notwithstanding the common challenges of ethics and the practicality of taking blood from a rural setting, the need to reduce human DNA contamination by using simple methods in separating human white blood cells (WBCs) from parasite infected red blood cells (iRBCs) was critical.

3.2 Objectives

1. To purify *P. falciparum* DNA from clinical samples across rural communities in the KNDs for deep sequencing,

2. To assess key malariological parameters in this rural population with high malaria endemicity

3. To compare the efficacy of WBC/iRBC separation protocols employing either Plasmodipur® or CF11-based filters.

3.3 Materials and Methods

3.3.1 Patient recruitment

I set up five recruitment centres in the KNDs; four peripheral outpatient clinics and the Navrongo War Memorial Hospital (NWMH). Patients reporting to these clinics with fever, were screened with *P. falciparum* specific rapid-diagnostic test

(RDT) kit to confirm malaria infection. The RDT positive patients were then evaluated by standard microscopy for their inclusion into the study. Patient recruitment is described in detail in chapter 2 (section 2.1.2).

3.3.2 Processing of whole blood

Whole blood taken from patients was initially separated by lymphoprep density gradient centrifugation (see chapter 2 section 2.2.2). This initial step drastically reduced the number of peripheral blood mononuclear cells. Subsequently the blood was then either further processed with Plasmodipur® filters (Eurodiagnostica) or CF11-based filters constructed on site. Details of CF11 filter construction are described in chapter 2 (section 2.2.1.1). Processing of whole blood by these methods was described in detail in chapter 2, (see section 2.2.3 for Plasmodipur® filteration and section 2.2.4 for CF11-based filteration).

3.3.3 DNA extraction and quantification

DNA extraction was undertaken using the QIAamp DNA Blood Midi/Maxi kit (Qiagen) according to the manufacturer's protocol. I used the QIAamp Maxi columns to process red blood cell pellets of greater than 2.0 ml, while low blood volumes were processed with the Midi columns. Details of the DNA extraction protocol was described in chapter 2 (section 2.2.4.1). DNA was quantified using the Quant-iT^{TM} PicoGreen dsDNA Assay kit (Molecular probes-invitrogen), and the amount of human DNA contamination was assessed using real-time PCR (Chapter 2 section 2.2.4.2).

3.4 Results

A total of 551 participants met the study inclusion criteria (briefly, a positive malaria by RDT and $\geq$ 50 trophozoites/200WBCs) and were enrolled into the study. The proportion of samples collected by clinic and relevant malariometric indices are shown in Figure 3.1.

3.4.1 Baseline characteristics of participants

3.4.1.1 Demographic Characteristics

The demographic characteristics of participants was reflective of the general demography of the KNDs. Of the 493 participants who had mild malaria (MM), the majority Kasem ethno-linguistic group formed 66.4%, followed by the Nankam who represented 42.6%, and 1.4% were from the minority Buli ethno-linguistic group. The demographic and ethno-linguistic distribution of severe malaria (SM) patients followed a similar pattern, with 67.2% being Kasem and 29.3% Nankam. In terms of gender, 54.4% of SM cases and 47.3% of the MM cases were males. Further details may be reviewed in Table 3.1.

3.4.1.2 Clinical Characteristics

Of the 551 patients recruited, 493 (89.5%) were mild malaria (MM) patients and 58 (10.5%) were children 0-5 years admitted to the NWMH with severe malaria (SM). The mean haemoglobin concentration for MM cases was 10.0 $\mathrm{g\,dl^{-1}}$ (range: 8.7-18.3) and that for SM cases was 8.7$\mathrm{g\,dl^{-1}}$ (range: 0.9-11.7). The mean age was

8.0 and 3.2 years for MM and SM patients respectively. By and large, gender was equally split in both disease categories, but, females were proportionately higher in MM and the reverse was true for SM. A 25.4% of severe malaria cases had severe malarial anaemia (i.e haemoglobin concentrations below 5g/dL) and were given blood transfusion. The peak prevalence of parasitaemia followed a unimodal distribution that coincided with the months of peak malaria transmission (rainy season). Though, I observed a wide variation in parasite densities through each month, the peak of mean parasite densities occurred in the months of August through October (Figure 3.2).

In terms of location-specific parasite densities, the lowest mean parasite density of 35100 (range:1360-166600) was observed among patients attending the Paga clinic in the north close to the border with Burkina Faso (Figure 3.1). This level of parasitaemia was closely mirrored by the parasite density in communities in the rocky highlands situated west of the KNDs (mean: 36810; [range: 1600-169500]), where patients were sampled from the Chiana clinic. Areas with high parasite densities were the communities in the south sampled from the Kologo clinic (mean: 61860;range:[3260-290600]), and the NWMH in the central district (mean: 59520; [range:4360-204600]). Patients sampled from communities situated in lowland Savannah areas in the east, and attending the Kassena-Nankana east clinic had intermediate parasite densities (mean: 50090; [range:1600-22630]).

Since WBCs counts were not measured in this study, the conventional approach that assumes a WBC count of 8000 per microlitre of blood was used to convert parasite counts per 200 WBCs to counts per microlitre of blood. Overall, the mean parasite density was 49191.7 parasites μl^{-1} of blood (range:1360-290560)

for MM cases and 111315.9 parasites μl^{-1} of blood (range:17120-337880) for SM cases. In general, hyperparasitaemia, defined as, parasitaemia $\geq$ 100,000 parasites μl^{-1} of blood or 2% parasitaemia (shown as a horizontal dark red dotted line in Figure 3.2), is a common ($\approx$ 10% of cases) manifestation in the KNDs.

Further details of the clinical features of participants is presented in Table 3.1.

Characteristic	**Mild malaria** (N = 493)	**Severe malaria** (N = 58)
Ethnic group:		
Kasem	270 (54.8)	39 (67.2)
Nankam	210 (42.6)	17 (29.3)
Buli	7 (1.4)	1.0(1.7)
Fulani	5 (1.0)	-
Other	1 (0.2)	1.0 (1.7)
Age (yrs)	8.0 [0-80]	3.2 [0-5]
Gender:		
Male	233 (47.3)	31 (53.4)
Female	260 (52.7)	27 (46.6)
Clinical:		
Fever (°C)	37.7 [36.9-40.3]	38.9 [38.3-40.5]
Fever duration (days)	2.6 [1.0-8.0]	2.5 [1.0-11.7]
Haemoglobin ($g\,dl^{-1}$)	10.0[8.7-18.3]	8.7 [0.9-11.7]
Parasite density (μl^{-1})	49191 [1360-290560]	111315 [17120-337880]
GM parasite density	30012.4	84543.3

GM-Geometric mean, n (%), Mean[range]

Table 3.1: Baseline characteristics of participants

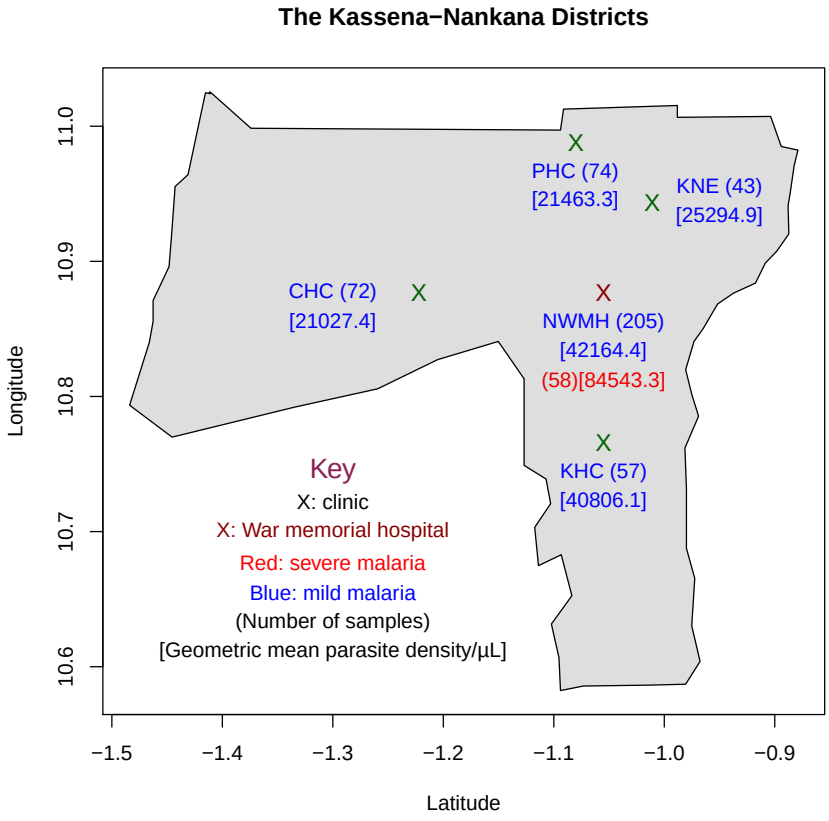


Figure 3.1: **Map of KNDs showing the location specific distribution of samples collected.**

MM cases (blue) were recruited from outpatient clinics; **CHC**-Chiana Health Centre serves the Western rocky highland communities, **PHC**-Paga Health Centre serves communities north of the district along the Ghana Burkina Faso border, **KNE**-Kassena-Nankana East Health Centre serves communities in the far East of the district, **KHC**-Kologo Health Centre serves communities in the South, **NWMH**-The Navrongo War Memmorial Hospital outpatient department serves the Navrongo town and surrounding communities and all SM cases (red) are referred to the NWMH paediatric ward

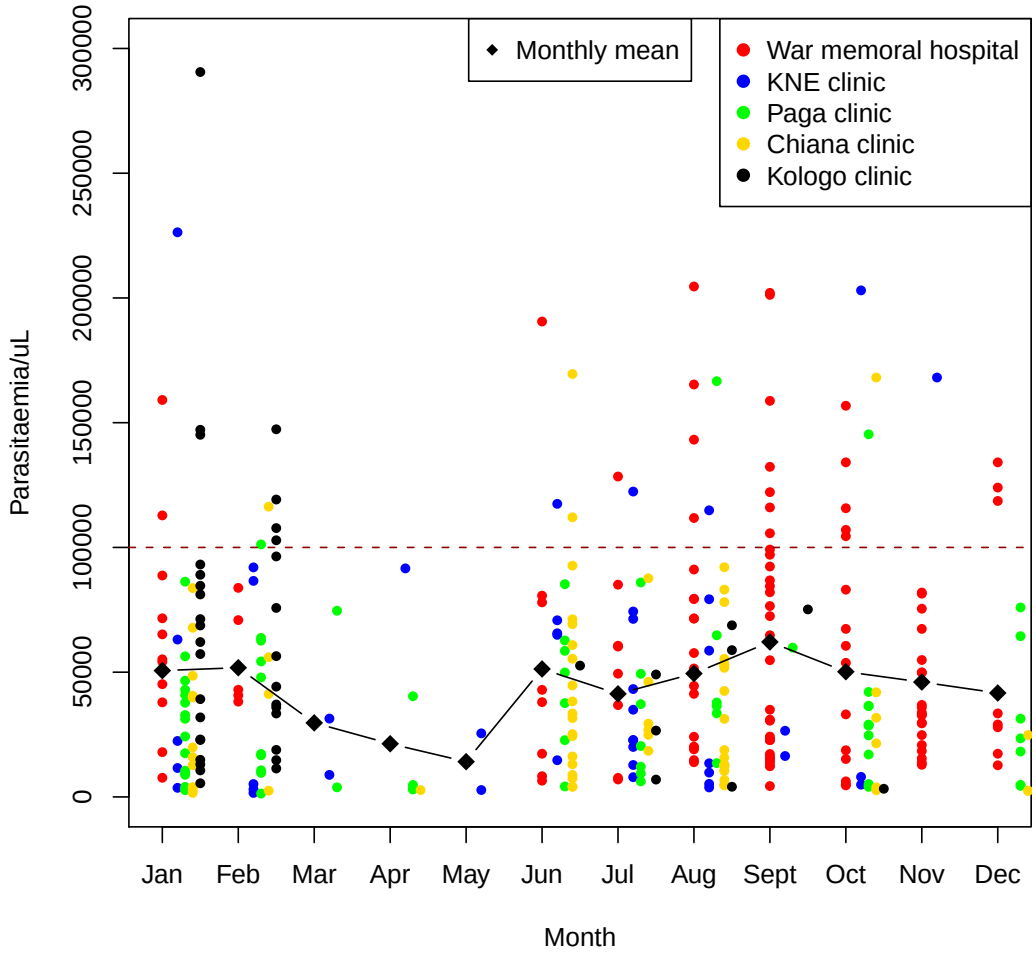


Figure 3.2: **A scatter plot of the monthly distribution of parasite densities. Data presented spans August 2009 to November 2011.** Points are colour coded according to the clinic from which patients were sampled. The dark red dotted line indicates the 2% parasitaemia mark, whilst the solid dark line shows the pattern of monthly mean parasite densities

3.5 Total DNA recovered

In general there was wide variation in the total DNA recovered. Overall the mean total DNA (human plus parasite DNA) yield was 5.85µg ml^{-1} of blood taken (range:0.00-51.94). As shown in table 3.2, the quantity and quality of DNA yield varied considerably between the five sampling sites. The average blood draw was 3.3ml (range:1.0-5.5), and 75% of samples had <50% human DNA contamination. I observed that overall, there was a positive correlation (r=45%) between blood draw and total DNA yield (Figure 3.3), however, the linearity of the relationship could potentially be affected by other parameters, notably parasite density, batch-effect in the processing of samples, and variation in the separation method used. There was significant difference (p=0.0001) between the variances of total DNA recovered at blood draws (<4ml) and higher blood draws (>3ml). Therefore, the Wilcoxon ranked sum test (a non-parametric test) was to compare total DNA in the two categories, which was observed to be statistically significant (Wilcoxon test statistic, $W = 6594.5$, $p = 0.0004$). Human DNA contamination was similarly distributed across all sites. I did not find evidence of a systematic site-specific influence on the total DNA extracted. Overall, the majority of samples (62.3%) had a human DNA contamination under 10% (Table 3.2).

3.6 Analysis of leukocytes depletion methods

Plasmodipur® filters and CF11-based filters were used interchangeably, based on availability, to deplete whole blood of leukocytes prior to parasite DNA extrac-

tion. Overall, 171 samples were processed with Plasmodipur® filters and 115 samples with CF11-based filters in the same malaria season, and were used for this analysis. It is therefore, important to note that the comparisons discussed were global and not a head-to-head comparison of the two methods.

3.6.1 DNA yield by Plasmodipur® and CF11 filters

There may be several parameters that influence WBC separation by Plasmodipur® and CF11-based filters, including volume of blood draw, ratio of parasitized RBCs and haemoglobin levels. Notwithstanding these, I found the following:

First, the median blood draw for samples processed with Plasmodipur® filters was 4.0 ml (range: 1.5-5.5) and that for the CF11-based filters was 2.5 ml (range: 1.0-4.5). Therefore, correcting for volume, the median DNA yielded for samples processed by Plasmodipur® filters was $0.6\mu g\,ml^{-1}$ of blood drawn (mean: 1.5, range: 0.0-13.1), and that for samples processed by CF11-based filters was $12.3\mu g\,ml^{-1}$ of blood drawn (mean: 16.6, range: 0.1-62.6).

Second, as shown in Figure 3.4, in general for both methods, there is a linear relationship between total DNA recovered with blood volume. However, as expected parasite density (size of circles in Figure 3.4) was an important variable in the relationship between total DNA recovered and volume of blood draw. This influence may be dependent on other factors, such as haemoglobin status and may not be linear.

I used a generalized linear model to examine these relationships, taking blood draw and parasitaemia as the two main effects on total DNA recovered. Par-

asitaemia was found to be highly significant for both methods (p = 0.0007 for CF11-based filters and p = 0.006 for Plasmodipur filters). Blood volume was significant for both filters as well (p = 0.02 for CF11-based filters and p = 0.03 for Plasmodipur filters).

Finally, It was worth establishing whether there was any evidence of non-linearity in the response of total DNA yield to blood volume with parasite density as co-variate. By fitting a quadratic term for parasitaemia, it became evident that the relationship was significantly non-linear (p = 0.006 for CF11-based filters and p = 0.03 for Plasmodipur filters). I fitted generalized additive model (GAM) to the response variable total DNA yield with smooth terms for volume of blood draw and parasite density (trophozoites/200WBCs) to further investigate these non-linear relationships using all the data. The GAMs are non-parametric smoothers, excellent at showing the humped relationship between variables when there are covariates that are continuous (ie parasite density), and could highlight the possibility for a threshold parasitaemia at which the rather complicated relationship between total DNA and blood volume becomes non-linear.

The tight 95% confidence intervals (95% CI) show that volume of blood draw is a good predictor of total DNA yield (Figure 3.5 top panel). Also apparent is the fairly constant effect between 2-4ml of blood draws, which suggests that within this range the levels of parasitaemia may be the overriding factor in the amount of DNA recovered. Interestingly, below 2ml and above 4ml of blood draw there is an increase in total DNA yield. Increased total DNA yield at higher blood draw is to be expected, however, the apparent increase at lower blood draws could be due to many factors, including extreme care in processing merely because

of the small volume obtained, more effective action by the lysis buffer during DNA extraction, higher parasite densities, and the like. On the other hand, the relationship between parasitaemia and total DNA recovered is completely linear (Figure 3.5 bottom panel), but, as parasitaemias increases this linearity becomes less and less predictive of total DNA yield as depicted by the widening 95% CI. Also apparent is a possible cut-off for minimum parasitaemia at about 2000 trophozoites μl^{-1} of blood (50 trophozoites/200 WBCs), as depicted by the extremely tight 95% CI.

3.6.1.1 Human DNA contamination

To further examine the influence of parasitaemia on the quality of the total DNA recovered, ten equally sized bins of parasite density were created and the mean percentage of human DNA contamination was calculated for each bin. As shown in Figure 3.6, mean human DNA contamination for both methods was generally higher up to parasite densities of about 3000 μl^{-1} of blood (75 trophozoites/200 WBCs) for CF11-based filters and Plasmodipur filters respectively. In general beyond a threshold of 75 trophozoites/200 WBCs, the mean human DNA contamination dropped, but, more rapidly with CF11-based filters than Plasmodipur filters. This further suggests the relatively better efficacy of the CF11-based filters to rapidly deplete WBCs as compared to Plasmodipur filters

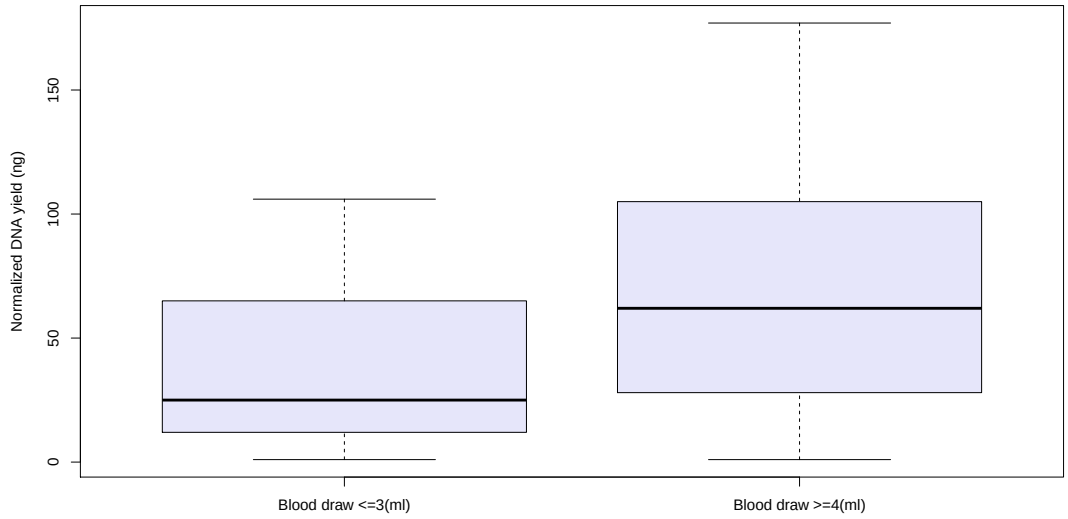


Figure 3.3: **Influence of blood draw on total DNA recovered**. Boxplots of DNA yield (normalized for blood draw) per two categories of blood draw. Boxes show inter-quartile range (black line in boxes are medians); whiskers extend to data extremes

	Sampling locations					
Parameter	CHC	PHC	KNE	NWMH	KHC	Overall
Number	72	74	43	304	57	551
Mean blood draw (ml)	±3.5	±3.0	±2.5	±3.0	±3.2	±3.3
Mean parasite density μl^{-1}	36810	35100	50090	59520	61860	54400
Mean DNA ($\mu g\, ml^{-1}$ blood drawn)	4.61	5.24	3.92	5.89	4.92	5.85
Range	0.0-26.6	0.0-32.3	2.2-44.3	2.3-51.9	0.0-26.6	0.0-51.9
$huDNA_{50}^{c}$	64 (88.9)	34 (45.9)	19(44.2)	223(73.4)	19 (33.3)	412 (74.8)

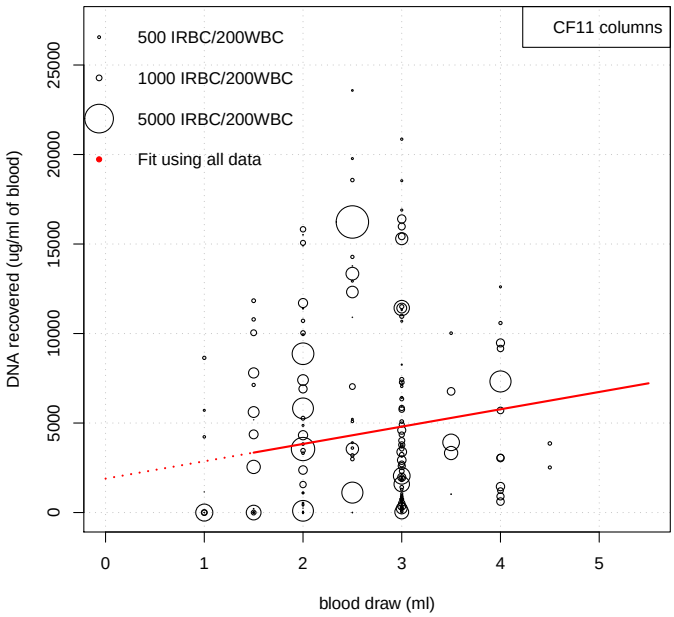
[c] number (%) of samples having <50% human DNA contamination.

Table 3.2: **Summary of location specific parameters and total DNA yield**

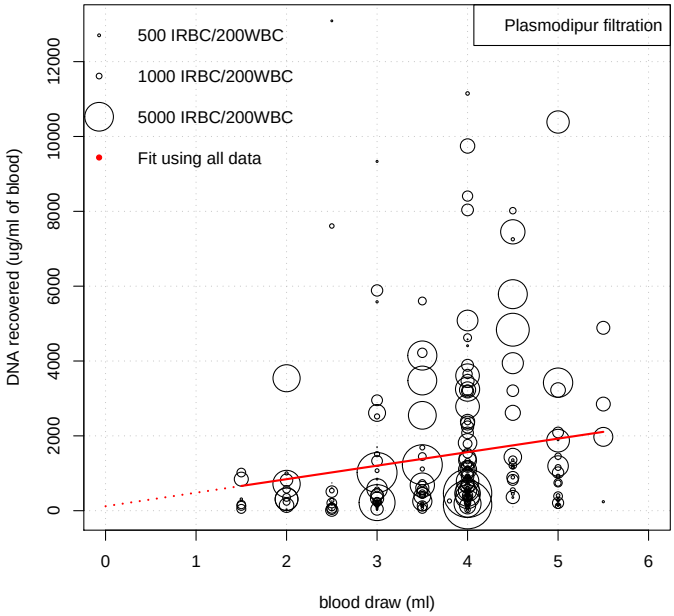
Method	**Plasmodipur filters**	**CF11-based filters**	**P-value**[1]
Total DNA yield (ng)	15.0 (0.0-173)[2]	12.7 (0.1-164)	0.035
Parasite DNA yield (%)	65.9 (0.0-100)	94.7 (45.9-100)	<2.2e-16

[2] Mean (range);[1] Wilcox ranked sum test with continuity correction

Table 3.3: **Total DNA and percent parasite recovered**



(a) CF11-based filters



(b) Plasmodipur® filters

Figure 3.4: Relationship between blood draw and total DNA yield.The influence of parasitaemia is also shown (size of circle). **Note that the Figures have different axis.**

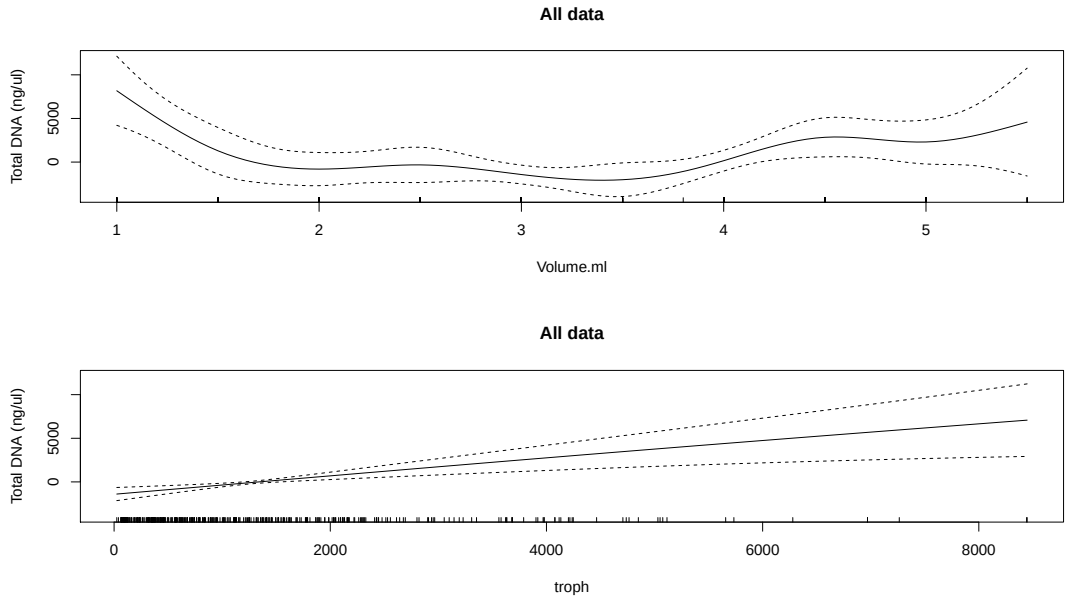
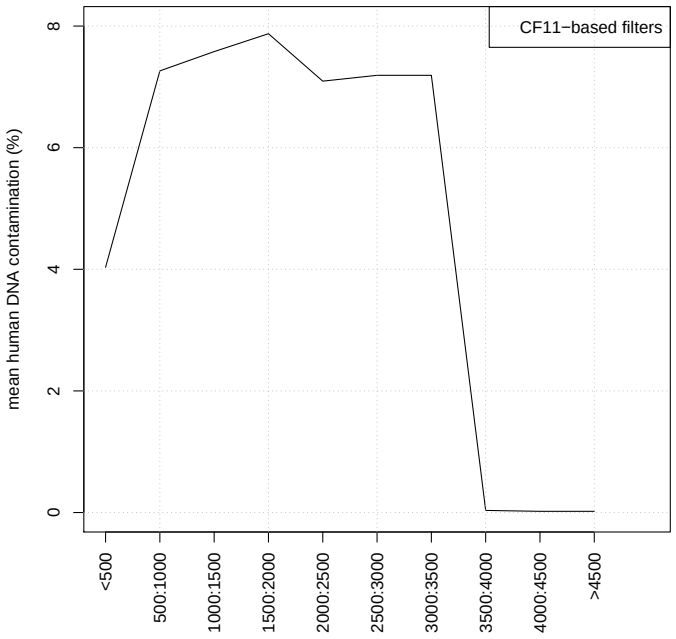
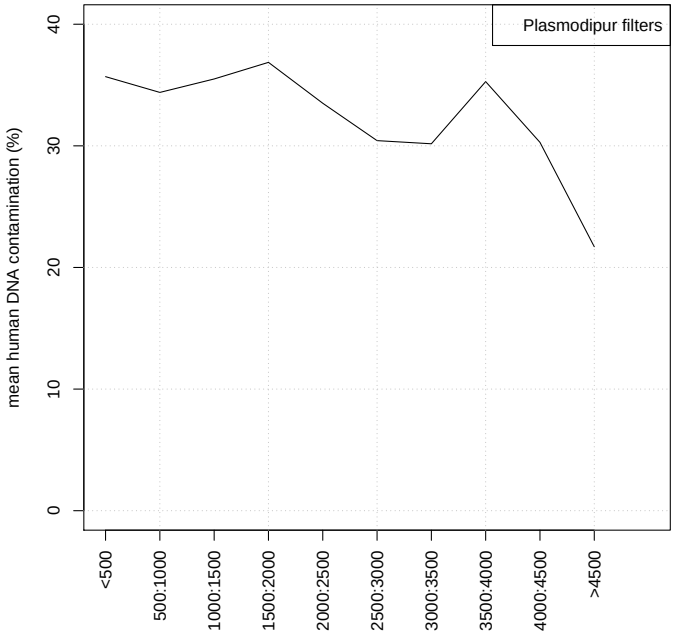


Figure 3.5: **Estimated functions for volume of blood** ml^{-1} **(top panel) and parasite density** $\mu\mathrm{l}^{-1}$ **of blood (bottom panel).** The region between the two broken lines represent the 95% confidence intervals of estimated function (solid line).



(a) CF11-based filters



(b) Plasmodipur® filters

Figure 3.6: **Relationship between parasite density and mean percentage human DNA contamination**. The x-axis shows 10 bins for parasite density μl^{-1} of blood.

3.6.2 Depletion of Leukocytes and retention of IRBCs

The relative effectiveness of each method in depleting malaria-infected blood of WBCs and concentrating IRBCs should lead to differences in the percentage of human DNA contamination. This should be dependent on the initial WBC count as high counts may lead to early saturation of the filters. However, since the goal was to purify parasite DNA at front-line community health centres, WBC counts were not determined for logistic reasons. Also, because the end-point for the WBC separation protocols was the proportion of human DNA depleted (expressed as a percentage of total DNA recovered), removing the leukocytes sooner than later to avoid lysis and release of human DNA was critical. These comparisons are therefore, in the light of the relative strengths of these methods in removing WBCs and enriching infected RBCs for DNA extraction. Thus, I next assessed the frequency distribution of human DNA contamination by each method. As evidenced in Figure 3.7 both filters were adequate in trapping the WBCs as the majority of the samples (62%) had low amount of human DNA contamination (<10%). The mean human DNA contamination for samples processed with Plasmodipur® filters was 35.7% (range:0.0-100) and the median was 17.5%. On the other hand, samples processed with CF11-based filters had a mean contamination of 7.7% (range:0.0-54.1) and the median was 0.46%.

To shed more light on how human DNA affected the variation in total DNA by each method, I further assessed the distribution of total DNA among three categories (<30%, 30 to 60% and >60%) of human DNA contamination. As shown in Figure 3.8, a similar distribution of variance in Total DNA yield was obtained at low (<30%) and intermediate (30-60%) human DNA contamination for both

methods. The CF11-based filter distribution did not change with increasing percentage of human DNA, suggesting a more consistent process and somewhat early saturation of these filters. The Plasmodipur® filters at similar levels of human DNA contamination yielded much less total DNA, but, tended to have many outliers. However, beyond 60% contamination the total yield for Plasmodipur® filters was found to improve, but largely human DNA. A detailed examination of these highly contaminated Plasmodipur separated samples revealed that 98% of them were taken from peripheral clinics and transported to the laboratory for depletion. This was necessitated because the local clinic had run out of Plasmodipur filters. The possibility of cell lysis (haemolysis) prior to WBC separation therefore could not be ruled out, emphasizing the need for immediate separation of WBCs rather than later.

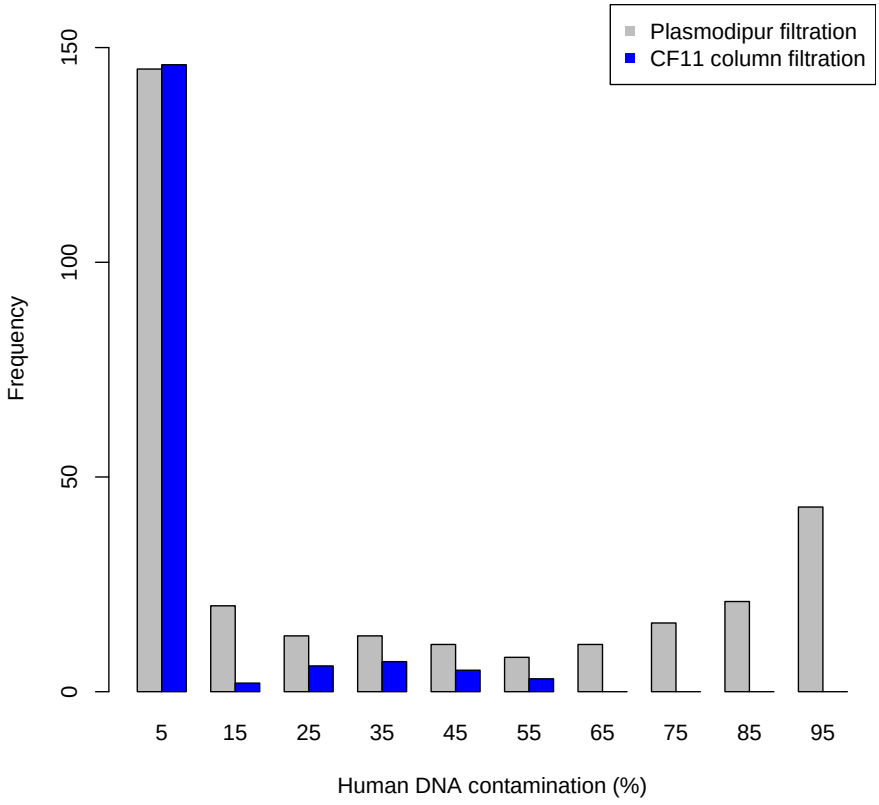


Figure 3.7: Frequency of human DNA contamination by method. Percentage of human DNA contamination is comparatively lower with CF11-filters (majority of samples had less 50% contamination) showing the efficacy of CF11-based filters in depleting leukocytes (nucleated cells that contain human DNA)

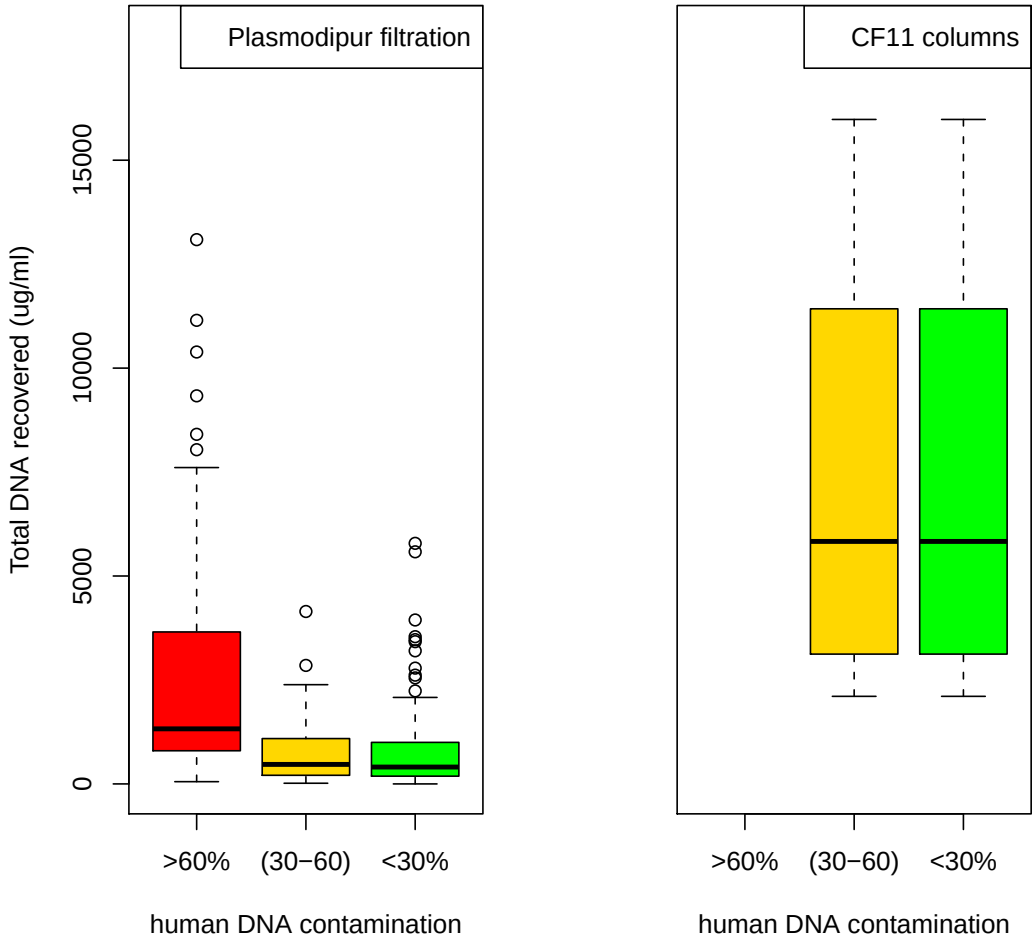


Figure 3.8: **Effect of human DNA contamination on the total DNA yield**. CF11-based filters gave a much higher median DNA yield and a wider spread. Plasmodipur® filters deplete leukocytes poorly and retains little parasite DNA beyond 60% contamination. Outliers are common with Plasmodipur® filtration

3.7 Discussion

The goal of the work presented in this chapter was to purify *P. falciparum* DNA from clinical samples for deep sequencing and collect clinical and demographic data for assessing relevant malariological parameters. To this end, I employed simple field techniques to separate WBCs (containing human DNA) from whole blood taken directly from patients whilst purifying parasite IRBCs for DNA extraction. In this regard, the sampling locations were purposely selected to capture the socio-demographic and ecological variation in this areas of perennial but, seasonal malaria transmission. This discussion will therefore, entail the demographic characteristics and malariological patterns observed and the relative successes of the methods used.

To start with, the majority (55%) of samples were collected from the central district, which has a comparatively wider catchment area and least (8%) from the Kassena-Nankana east clinic. Samples were similarly distributed in the other three locations. Of the patients sampled 55% were of the majority Kasem ethnolinguistic group, and 43% of the Nankam group. The minority Buli ethnolinguistic group represented about 2% of cases recruited. Interestingly, 1% of Fulani, a nomadic ethnic group in scattered all over West Africa.

Overall, gender was equally split within each disease category, with slightly more females (53%) in the mild malaria group and slightly more males (53%) in the severe malaria group. The mean age of 3 years for children enroled with severe malaria reflects the general high burden of life-threatening malaria among infants in areas of high malaria transmission[163]. A 25.4% of severe malaria cases

had haemoglobin concentrations below 5g/dL requiring blood transfusion, further evidence that severe malarial anaemia is an important manifestation of severe malaria in the KNDs[157].

In general, the monthly distribution of parasitaemia (Figure 3.2) was consistent with the pattern of malaria transmission reported in the KNDs[12]. Peak parasitaemia levels coincided with the peak of the rainy season from August through October, when malaria transmission is highest. These data are consistent with previous reports from clinical studies that found peak hospital admissions among children to occur during the months of August through October, mostly with severe malaria[157]. The location-specific prevalence of parasitaemia was high but variable with mean parasite densities ranging from 35100 parasite μl^{-1} of blood in the north to 61860 mean parasites μl^{-1} of blood in the south. Notwithstanding the sensitivity of the mean to outliers and variations in sample sizes at the different locations, these data suggests that malaria transmission across the KNDs is intense. However, malaria control interventions such as insecticide treated bednets aimed at interrupting transmission and the prevalence of the main vector mosquitoes (eg *Anopheles gambiae and An. funestus*) may not be uniform across the district. Indeed, a study by Appawu *et al.,* 2004, reported micro-ecological differences in malaria transmission in the KNDs[12].

Also evidenced was prevalence of uncomplicated hyperparasitaemia ($>$ 2% parasitaemia), which seems to be a common outcome of malaria infection in the KNDs during the peak transmission period. Hyperparasitaemia is associated generally with higher treatment failure rates, and of particular concern is the fact that antimalarial drug resistance is likely to arise in children with heavy parasite bur-

dens but with little or no immunity[239]. The current recommended treatment for mild malaria in Ghana is a combination of artesunate (a short half-life drug) and amodiaquine (a long half-life drug). It is anticipated that artesunate will achieve substantial and rapid parasite killing that will reduce the parasite biomass considerably, to allow the long acting amodiaquine to kill off the remaining parasites[209,238]. However, in view of the prevalence of hyperparasitaemia observed here, if therapeutic levels of amodiaquine are not achieved, inasmuch as patient compliance to the treatment regimen remains a problem, such a situation may encourage the evolution of resistance to amodiaquine in the long term.

In terms of DNA recovery, simple field-based techniques were used to purify *P. falciparum* DNA from whole blood across communities in this rural setting. Of the 551 *P. falciparum* clinical isolates obtained, the overall mean total DNA recovered was 5.85 $\mu g\,ml^{-1}$ (range: 0.0-59.1). Such a wide variation in total DNA recovered may reflect the enormous variation in WBC counts observed in clinical samples, particularly among sick children. Of the isolates obtained, 75% (412/551) of samples had high quality parasite DNA with $<$50% human DNA contamination (Table 3.2). In general, all the samples obtained could potentially be used for molecular studies involving *P. falciparum* DNA, and majority of these samples met the current criteria of at least 5 µg of DNA and $<$60% human DNA contamination that is required for Illumina next-generation sequencing.

Regarding the relative performance of Plasmodipur® filters versus CF11-based filters in depleting the patient blood of WBCs in order to reduce human DNA contamination, the volume of blood draw and parasitaemia were found to be critical in the total DNA yielded regardless of the method used. I observed a

linear relationship between total DNA and volume of blood draw (Figure 3.4) for each method. As one would expect, parasite density was found to have a linear relationship with Total DNA recovered by both methods. Indeed, the non-parametric generalized additive models (GAMs) with smooth terms were useful in providing clarity on the influence of parasitaemia, the 95% confidence intervals showed a minimum threshold parasitaemia at about 2000 trophozoites μl^{-1} of blood (50 trophozoites/200 WBCs) [Figure 3.5]. The GAM procedure described also showed that smaller blood draws (in the range of 1 to 2 ml) could recover a useful amount of DNA, which supports the current effort by MalariaGEN to use miniature CF11-based columns to process small blood draws from a lancet stick of the finger instead of veni-puncture. The mean percent parasite DNA obtained was significantly higher with CF11-based filters than the Plasmodipur® filters ($p = 0.001$)[Table 3.3]. These findings are evidence of the fact that with the current protocols, higher parasite density alone does not guarantee a higher parasite DNA yield. Other factors such as haemoglobin (or haematocrit) levels and stage of the infection may and will influence DNA quality.

As observed in other studies[14,215], Plasmodipur® filters performed poorly relative to the CF11-based filters in trapping WBCs for removal and concentrating parasite infected RBCs for parasite DNA extraction. Though the used of two Plasmodipur® filters per sample may have some added benefit, the high cost ($\approx$ 12 USD) of these filters make such an option not sustainable, especially when sample sizes increase.

The level of parasitaemia at which human DNA contamination begun to drop was about 3000 μl^{-1} of blood for both methods. The data shows that a mini-

mum blood draw of 1.0ml (average 3.0ml) and minimum parasitaemia of 2000 trophozoites μl^{-1} of blood is adequate for recovering good quality parasite DNA suitable for sequencing.

In summary, I sampled *P. falciparum* isolates from rural community clinics in the KNDs of northern Ghana. As evidenced from the data presented, sampling was reflective of ethno-linguistic distribution of the KNDs population, and malaria related parameters discussed also closely mirrored those reported by previous studies[100,101,157]. Therefore, these epidemiological evidences show that the parasite samples obtained from this endeavour are representative of the general epidemiological background in this setting. In addition, I tested out WBCs/iRBCs separation protocols and methods, which show that CF11-based filters are effective in separating WBCs from whole blood taken from malaria patients and concentrating iRBCs for extracting parasite DNA suitable for sequencing. The cellulose powder and syringes used to make CF11-based filters are cheap and the columns are very easy to prepare and set-up, therefore, they are suitable for collecting parasites from rural settings. In addition, the length of the columns are not fixed and in principle could be prepared at different lengths for processing different volumes of blood. Indeed, miniature CF11-based filters are currently being tried by malariaGEN for processing smaller blood volumes in the order of what is typically a filter paper blood blot. The data also shows that if money and blood volume are not an issue, Plasmodipur® filters are an acceptable alternative to CF11-based filters for separating WBCs in the field. These filters are pre-designed and convenient to use and as evidenced in this work, samples processed by Plasmodipur® filters are more likely to have human DNA present

in them, offering a better opportunity than CF11-based filters for both human and parasite genotyping.

3.8 Summary

- This chapter describes a sampling framework that was established to purify *P. falciparum* DNA from clinical samples in the KNDs, Ghana.
- Parasite densities in these clinical samples from the KNDs was observed to be high, with overall mean parasite density of 54400.0 parasites μl^{-1} of blood.
- Mean parasite densities tended to be high in the south (61860 parasites μl^{-1}) and central district (59520 parasites μl^{-1}) but lower in the west (36810 parasites μl^{-1}) and north (35100 parasites μl^{-1}).
- Communities in the east of the districts had intermediate parasite prevalence with a mean density of 50090 parasites μl^{-1}.
- This study also shows that uncomplicated hyperparasitaemia, a phenomenon known to have a high risk of progress to severe disease, is relatively common during peak transmission.
- Overall, of the isolates obtained, 75% (412/551) had high quality parasite DNA with $<$50% human DNA contamination.
- A minimum blood draw of 1.0 ml (average of 3.0 ml), and parasitaemia of 2000 trophozoites μl^{-1} of blood (50 trophozoites/200 WBCs) is adequate for obtaining quality DNA suitable for sequencing.

- A comparative analysis showed that CF11-based filters were more effective in depleting whole blood of leukocytes and drastically reducing the amount of human DNA contamination in the total DNA recovered.
- Plasmodipur® filters were found to be inferior to CF11-based filters in depleting leukocytes and concentrating parasite infected red blood cells for extraction.
- The additional downside is that Plasmodipur® filters are very expensive ($\approx$ 50-55 USD per filter). Therefore, CF11-based filters will make for a cheaper ($\approx$ 1 USD per filter), simple and suitable method for use in the field over Plasmodipur® filters in sampling of *Plasmodium* DNA from clinical samples.

3.9 Conclusion

The sampling framework detailed in this chapter has provided an update of the epidemiological background of the KNDs, the reference population for this thesis, and highlighted key malaria and demographic parameters that are representative of the KNDs population.

- CF11-based filters are more effective than Plasmodipur® filters for purifying parasite DNA in rural settings that is suitable for sequencing.
- The data suggests an average blood draw of 3.0ml, and minimum parasitaemia of 50 trophozoites/200 WBCs is adequate for obtaining good quality parasite DNA for next-generation sequencing.

Chapter 4

P. falciparum population genetics

4.1 Introduction

In chapter 3, I described the implementation of a community-based sampling framework for purifying *P. falciparum* DNA from clinical samples for characterizing parasite diversity in northern Ghana. Genetic characterization of *P. falciparum* populations will unravel the natural history and population diversity of this important pathogen[35]. To this end, previous studies based on limited microsatellite (MS) data[7], and later capillary sequencing of a few hundreds of cultured isolates provided some insight[232]. More recent evidence was revealed by over 86,000 SNPs derived from Illumina next-generation sequencing of field isolates sampled from West Africa, East Africa, southeast Asia, and Oceania[121]. Consensus evidence from all these studies support population structure at continental and regional levels, and a high genetic diversity[7,36,121,232]. Much less is known about the genetic diversity and gene flow at local scales in malaria endemic

communities. Recent evidence emerging from studies of the ongoing artemisinin selection events in western Cambodia and western Thailand corroborate the need for an in-depth understanding of *P. falciparum* genetic diversity and gene flow at local levels in endemic populations[33]. In this chapter, I used a subset of the *P. falciparum* field isolates described in chapter 3, to characterize the parasite population in the Kassena-Nankana Districts (KNDs) of Northern Ghana.

4.2 Objectives

To use Illumina deep sequencing genotype data to:

1. characterize the structure of parasite population in the KNDs,
2. characterize within-host diversity using the genome-wide F_{WS} metric.
3. explore the epidemiologic profile of the F_{WS} metric in this area of intense malaria transmission.

4.3 Materials and Methods

4.3.1 Study site

The study was conducted in the Kassena-Nankana Districts (KNDs). As detailed in chapter 2, the KNDs are administrative areas in northern Ghana that cover a unified geographic area, of 1.675km^2 with a population density of 91.5 persons per km^2[158]. The estimated 153,000 residence of the KNDs are under continuous

demographic surveillance by the Navrongo health and demographic surveillance system[158]. This area holds the Tono irrigation dam, a large irrigation scheme that supports both subsistence and commercial farming activities, and is the main source of continous breeding of vector mosquitoes. The predominant vector is *Anopheles gambiae* (86%), followed by *Anopheles funestus* (14%)[12]. The average annual entomological inoculation rate (EIR) was estimated at 418 infective bites/person-years, with micro-ecological variations in EIRs over the entire area[12]. However, there is evidence of a changing trend in the burden of malaria across sub-Saharan Africa[164]. Indeed, recent unpublished data from the KNDs estimated an annual EIR of 140 infective bites/person-years (Asoala et al 2013, manuscript in preparation). Overall, malaria epidemiology is sufficiently characterized in this region[21,100,101,157,168,170].

4.3.2 Patient recruitment

I used a prospective longitudinal study design to sample field isolates among all-age groups reporting sick at the Navrongo War Memorial Hospital (NWMH), and four other health centres. I sampled parasites to reflect the ecological variation and spatial distribution of local villages in the KNDs (Figure 4.1), and the reported micro-ecological variation in transmission intensity[12]. I first screened patients for malaria with Paracheck-Pf® (Orchid Biomedical Systems, India), a rapid diagnoses test (RDT) kit[188], which is *P. falciparum* specific, reliable, and easy to use. Patients with parasitaemias $\geq$ 50 parasites per 200 WBCs ($\approx$ 2000 parasites μl^{-1} of blood) were preferentially recruited. Further details are described in Chapter 2 (section 2.1.2).

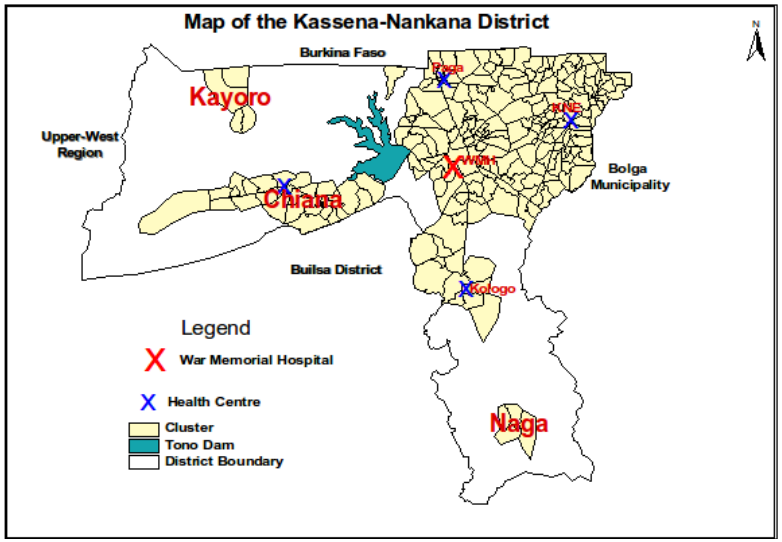


Figure 4.1: Sampling frame showing location of clinics (denoted X) that formed the sampling units. **NWMH**-Navrongo War memorial hospital, **KHC**-Kologo health Centre, **KNE**-Kassena-Nankana East health centre, **PHC**-Paga health centre and **CHC**-Chiana health centre.

4.3.3 Field laboratory activities

Patients who met my inclusion criteria were bled by venupuncture to obtain 2-5 ml of whole blood for parasite DNA extraction. Plasmodipur® filters and CF11-based filters were used to deplete whole blood of human leukocytes (to minimize human DNA contamination) and to concentrate parasite-infected red blood cells for parasite DNA extraction. I extracted parasite DNA using the QIAamp DNA Blood Midi/Maxi kit (Qiagen) according to manufacturer's instructions as detailed in chapter 2 (see section 2.2.1). Total DNA was quantified using the Quant-iT^{TM} PicoGreen® dsDNA Assay kit (Molecular probes-invitrogen) and stored at -20°C until ready for shipment to the Kwiatkowski laboratory for further processing.

4.3.4 Quantifying human DNA contamination

I used Quantitative Real-Time PCR (QRT-PCR) to assess the amount of human DNA contamination in the total DNA extracted. I used 2-5ng aliquots of each sample for QRT-PCR, using the Applied Biosystems StepOne PCR System with sets of human primers and primers specific to *Plasmodium spp.* Details of primer sequences and reaction profile is described in chaper 2 (section 2.2.4.2)

4.3.5 Illumina DNA sequencing

Illumina DNA sequencing was undertaken within the framework of a large multicentre *P. falciparum* genome variation project to facilitate variant discovery

and population genetic characterization[121]. I submitted samples in batches to the sequencing facility at the Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire. Samples were either sequenced using the Illumina Genome Analyzer II or the Illumina HiSeq2000 machines. Sequencing preference was given to parasite DNA samples with less than 60% human DNA contamination. The sequencing team prepared standard sequencing libraries according to manufacturer's protocol, except that genomic DNA was fragmented to $\approx$ 300 bps using Covaris device (Covaris Inc., Woburn, MA) rather than nebulisation[23]. Within the course of this project, sequencing read lengths increased from 37 to 105 bp paired-end reads.

4.3.6 Detection of polymorphisms

Genome-wide SNP discovery was undertaken as part of an established pipeline for the MalariaGEN community project on *P. falciparum* genome variation (PGV project) [Figure 4.2]. The *P. falciparum* genome has a high AT content (81%), particularly in non-coding regions (87%), which makes it comparatively more difficult to sequence[71]. Such low complexity sequence makes it especially challenging for accurate alignment of short sequence reads to the reference genome. Further compounding issues were the occurrence of genome regions with multiple polymorphisms, and the fact that multiple parasite genomes could be found within a single sample[15,121]. Numerous steps were, therefore, integrated into the PGV variant discovery pipeline in order to mitigate these challenges and obtain high confidence SNPs. As illustrated in Figure 4.2, it is a complex workflow that uses a couple of sequence mapping tools to align short sequence reads against the *P.*

falciparum 3D7 reference sequence (version 2.1.4)[112] to detect variable positions. A series of multi-layered quality filters were then used to assess the validity of each position across all samples (from different studies) in the pipeline at a given time. Thus, genotype calls are released by MalariaGEN to contributing investigators in a version controlled fashion. Details of PGV pipeline is described in[121].

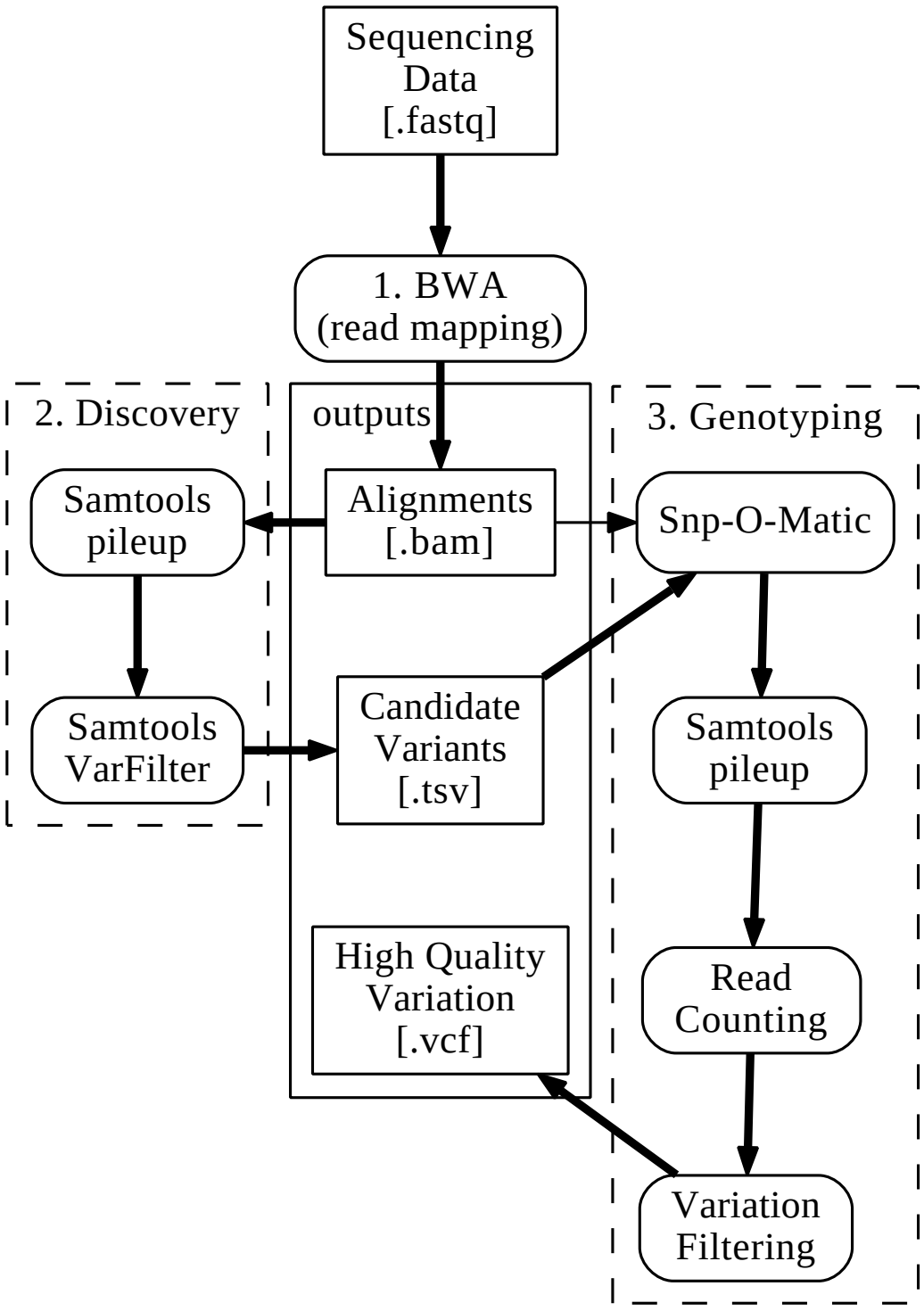


Figure 4.2: **Schematic representation of the PGV variant calling pipeline.** Squared boxes indicate data elements (reads, alignments etc), Rounded boxes indicate processes (mapping, genotyping etc). Arrows indicate execution dependencies,but, **thicker** arrows compose the critical execution path.

4.4 Statistical analysis and results

Notation	Definition
N	Total number of samples
M	Total of SNPs
$\hat{p}$	Non-reference allele frequency
$\hat{h}_i$	SNP heterozygosity
H_W^O	Observed within-sample heterozygosity
H_S^E	Expected population level heterozygosity

Table 4.1: Notations used to describe the parasite population

Ghana	Samples	MalariaGEN	Samples
Central	91	Burkina Faso	47
East	27	Mali	51
North	38	Kenya	26
South	50	Papua New Guinea	21
West	43	Cambodia	35
		Thailand	46
Total	249	Total	226
Total sites genotyped 400,000			
Number variable in Ghana 363,330			

Table 4.2: **Geographical distribution and counts of samples.** The local distribution of my samples is shown as well as malariaGEN open access data

Overall, I submitted 550 parasite DNA samples with between 80-500ng total DNA and a mean human DNA contamination of 15% (range: 0-96%) for Illumina next-generation sequencing. Of these 550 samples, genotype data for 277 parasite isolates were obtained as part of MalariaGEN release 2.0, and were used for this analysis. Genotypes were called at 400,000 high confidence SNP positions, and 91% (363,330/400,000) of these were variable in the KNDs population (Table

4.2). Of the 277 samples sequenced, 28 samples (8.8%)) were removed from this analysis because of poor mapping quality. Table 4.2 above summarizes the geographic distribution of samples used for this analysis.

4.4.1 Determination of allele frequencies

To determine allele frequencies at variable positions I first estimated the non-reference allele frequency (NRAF) as the proportion of non-reference sequence read counts at the j^{th} position within each sample (Equation 4.2). Let n_{ij} and r_{ij} denote, respectively, the number of non-reference and reference read counts in sample i at position j. If p_{ij} denotes the true NRAF, then for a given SNP j, the NRAF $\hat{p}_{ij}$ is given by:

$$\hat{p}_{ij} = \frac{n_{ij}}{n_{ij} + r_{ij}} \ . \tag{4.1}$$

I then compute the minor allele frequency (MAF) as:

$$MAF = \begin{cases} \hat{p}_{ij} & \text{if } \hat{p}_{ij} \leq 0.5 \\ 1 - \hat{p}_{ij} & \text{otherwise} \end{cases} \ .$$

I then calculated the population-level NRAF at a given SNP j by averaging across all samples:

$$\hat{p}_j = \frac{1}{N} \frac{\sum_{i=1}^{N} n_{ij}}{\sum_{i=1}^{N} n_{ij} + r_{ij}} \ . \tag{4.2}$$

As evidenced in Figure 4.3 the *P. falciparum* population in the KNDs is characterized by low frequency (<0.5) variants. In particular, rare variant (<5%) predominate the allele frequency spectrum in this highly malarious population.

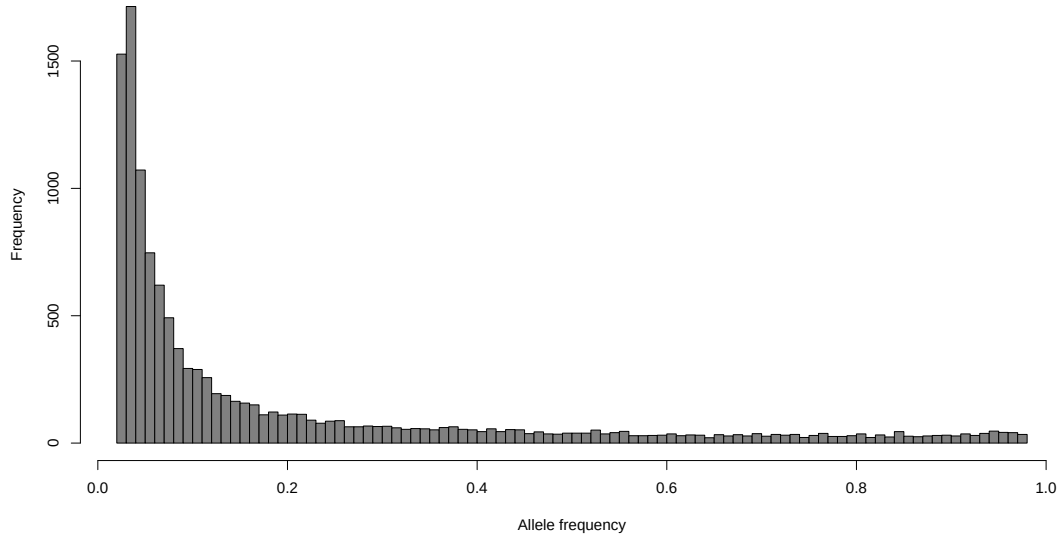


Figure 4.3: Allele frequency distributions in the KNDs

4.4.2 Calculation of heterozygosity

note: For the rest of this section, I will index samples by i*, and SNPs by* j*.*

The 400K high-quality SNPs obtained from our genotyping pipeline enabled me to carry out a genome-wide assessment of within-host diversity. For a given malaria infection, there is a possibility of occurrence of mixed infections (multiple genotypes). However one can not directly observe these using next-generation sequencing of clinical isolates. What can be observed is the occurrence of multiple alleles at given loci, which we refer to as heterozygosity. To garner an understanding of this process, I estimated the heterozygosity at each biallelic SNP j where there was sufficient read coverage for genotyping ($\geq$ 5 reads). Assuming that each allele has an equal chance of infecting a given host, mating is random, there

are no mutations or natural selection occurring, and no significant migration (ie the Hardy-Weinberg equilibrium holds), then the expected heterozygosity within a sample should be simply the fraction of variants expected to be[85]:

$$\hat{H}_{ij}^{E} = 2\hat{p}_j(1-\hat{p}_j) \ . \tag{4.3}$$

However it is possible to estimate within-sample observations. For a given SNP j, within sample i, I estimate observed heterozygosity by,

$$\hat{H}_{ij}^{O} = 2\hat{p}_{ij}(1-\hat{p}_{ij}) \ . \tag{4.4}$$

By averaging the observed heterozygosity at each SNP ($\hat{H}_{ij}^{O}$) across all SNPs for each sample, I derive the observed within-sample heterozygosity ($\hat{H}_{W}^{O}$), that is,

$$\hat{H}_{W}^{O} = \frac{\sum_{i=1}^{M} \hat{H}_{ij}^{O}}{M} \ . \tag{4.5}$$

Similarly the expected population-level heterozygosity was estimated as:

$$\hat{H}_{S}^{E} = \frac{\sum_{i=1}^{M} \hat{H}_{ij}^{E}}{M} \ . \tag{4.6}$$

Natural *P. falciparum* infections usually contain several thousands of parasites that could be of different genetic type. Therefore, I estimated the proportion of SNPs that carried different alleles (ie heterozygous) in each sample, and found that 66% of patients were heterozygous (Figure 4.4). As shown in Figure 4.4, a minority of infections were essentially monoclonal with no variable loci genome-

wide.

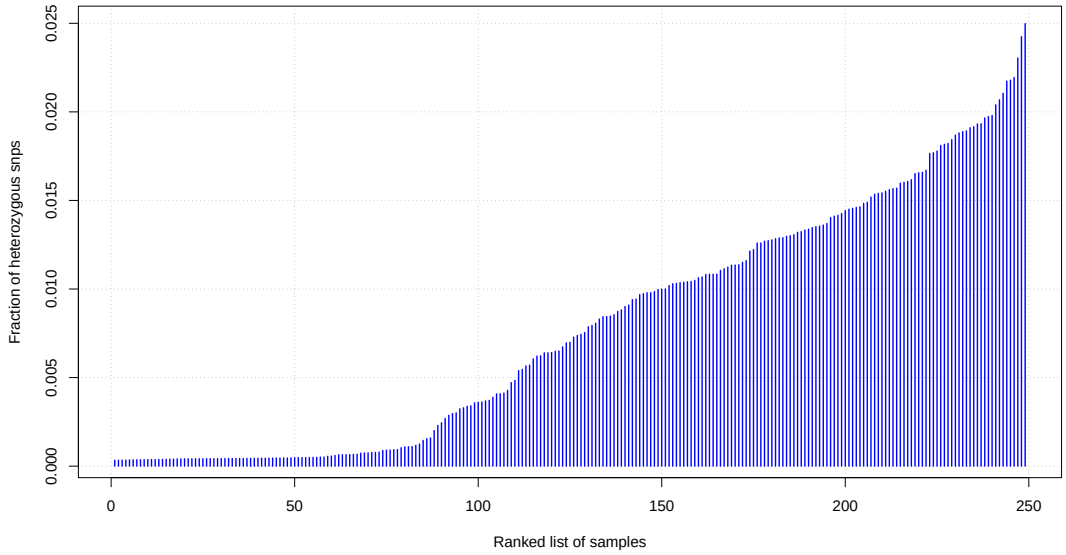


Figure 4.4: Fraction of heterozygous SNPs among patients

4.4.3 Calculation of F_{WS}

At this stage I would like to transform these SNP-level calculations into a metric for the degree of mixed infections within a sample and risk of out-crossing (or inbreeding). In this regard, I utilized the F_{WS} metric, which was introduced by Manske *et al.,* (2012). Details of calculations used to derive this metric may be reviewed in[121]. The F_{WS} metric is a linear approximation of the heterozygosity in the individual sample against that expected from the overall local population of parasites. That is, the F_{WS} metric uncovers the fraction of heterozygosity observed that is explained by the expected levels of heterozygosity across all samples. If we represent the observed within-sample heterozygosity as H_W and the expected within-population heterozygosity as H_S, then it implies that; $H_W = \hat{H}_W^O$ and $H_S = \hat{H}_S^E$.

From[121], F_{WS} can be writen as

$$F_{WS} = 1 - \frac{H_W}{H_S} \ . \tag{4.7}$$

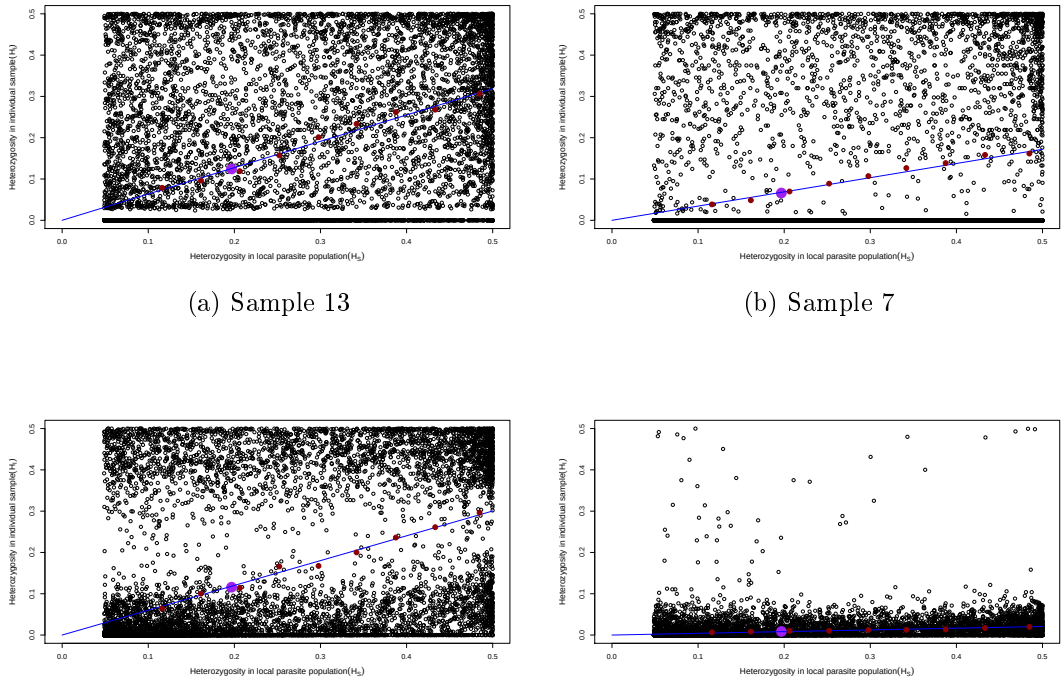
Hence the estimate of F_{WS} from this data is given by:

$$\hat{F}_{WS} = 1 - \frac{\hat{H}_W^O}{\hat{H}_S^E} \ . \tag{4.8}$$

As noted earlier, at the population level, the allele frequencies at the SNP were estimated as the mean of the allele frequencies, taken across all samples. The frequency of the least common allele at a given SNP across all samples was regarded as the population level MAF. As shown in figure 4.5 panel 1-4, there was

a considerable scatter in the ratio $\frac{H_W}{H_S}$ observed for different SNPs within a given sample, but, I observed the close linear correlation between H_W and H_S (purple dot) reported in[121]. However, as evidenced in the four different samples depicted in Figure 4.5, although each sample shows this linear relationship, the slope ($\frac{H_W}{H_S}$) of the line varies between samples.

Since heterozygosity depends on allele frequencies each of the SNPs was assigned to one of ten equally-sized MAF bins ([0.0-0.05],[0.05-0.1]...[0.45-0.5]). Estimates of H_W and H_S were computed for each MAF interval as the means of H_W and H_S across all SNPs in that interval. For each sample, H_W estimates were plotted against the corresponding H_S estimates for all MAF interval (Figure 4.6). The actual variation of the estimates of H_W and H_S for each minor allele frequency (MAF) bin for each sample is plotted as red dots in Figure 4.5. The F_{WS} metric for each sample was derived from the slope of the resulting straight-line plots for each sample as a computation of the ratio $\frac{\Delta H_W}{\Delta H_S}$ (Figure 4.6).



(a) Sample 13

(b) Sample 7

(c) Sample 28

(d) Sample 21

Figure 4.5: **Variation between samples in the correlation of within-host heterozygosity within-population heterozygosity (H_W, H_S).** The four panels each represent a different sample. Notice the changing gradient of the regression line, beginning in subfigure 4.5(a) with a much steeper gradient, and deep mixing of genomes. Subfigure 4.5(b) depicts an infection with moderate levels of genome mixing whilst in Subfigure 4.5(c) the allele frequency banding suggests the mixing of essentially two genomes (MOI of 2) in this infection. The bottom right (Subfigure 4.5(d)) shows an infection that was of very low genetic diversity (but not monoclonal).

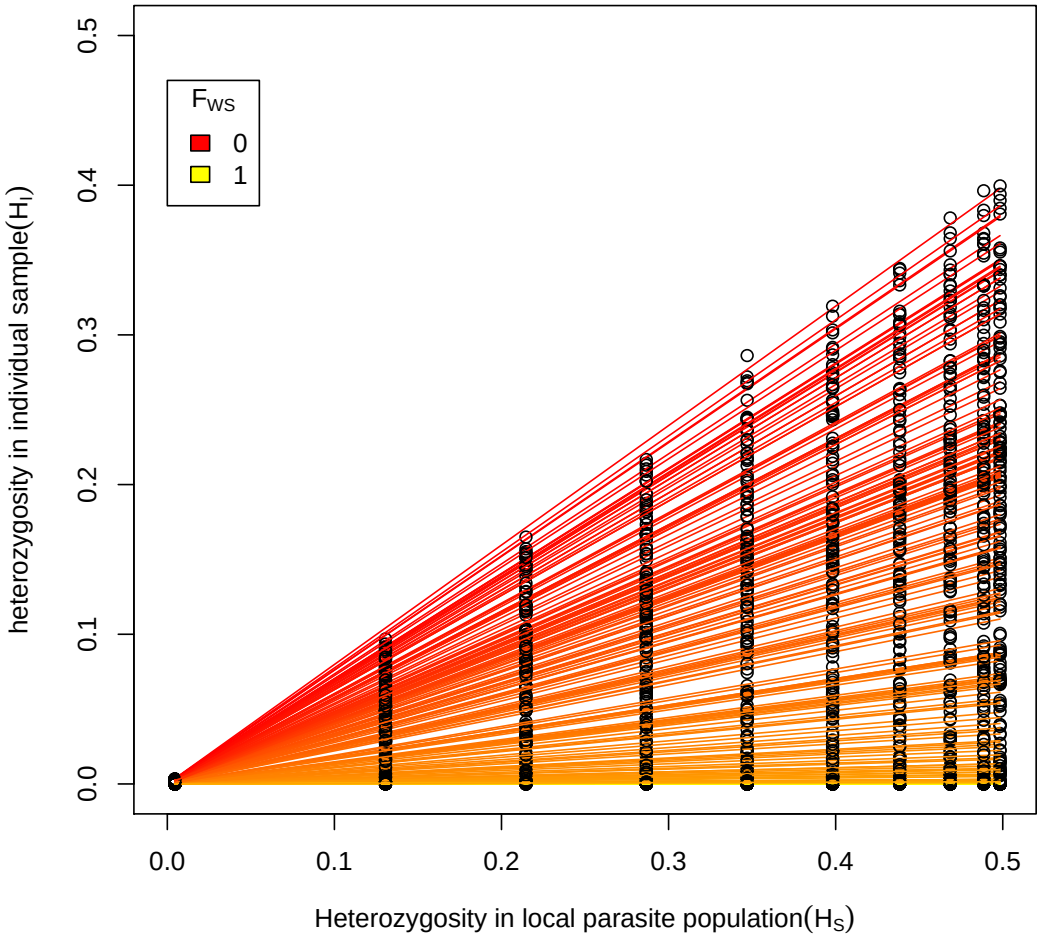


Figure 4.6: **Correlation of within-individual heterozygosity with population-level heterozygosity.** Each line represents a different sample. Thus, each different β value for the gradient (the ratio $\frac{\Delta H_W}{\Delta H_S}$) is technically equivalent to F_{WS}, which ranges from 0 to 1. The colour coding illustrates the variability of samples within this range (red meaning highly diverse and yellow clonal). An F_{WS} of 0 means that the diversity in that sample is equivalent to the population-level diversity giving a high risk of out-crossing. Apparent is the high risk of out-crossing in this parasite population (spectrum of red colour coded samples).

4.4.4 Parasite population structure

The genetic relatedness between samples was investigated first, by principal component analysis (PCA)[1]. The PCA is a multivariate technique that analyzes a data table in which observations are described by several inter-correlated quantitative variables. The aim of PCA is to reduce data dimensionality into a smaller number of uncorrelated variables (components) that enables the extraction of those components that account for most variability in the data. All possible correlated variables (ie all variable positions between samples) are transformed into a smaller number of predominantly uncorrelated variables- the principal components. The first PCA accounts for as much of the variability in the data as possible, with each successive component accounting for a proportion of the remaining variability. The pattern of similarity of the observations and of the variables can then be displayed as points in maps.

The variable SNPs (363,330) were used to derived a pairwise distance matrix for 249 samples based on pairwise -complete genotypes. Where one or both samples could not be genotyped, I discarded such SNPs. I calculated the pairwise distance as the proportion of SNPs at which the two samples were genotyped with different alleles. This distance was computed for all sample pairs, creating an $N \times N$ symmetric matrix.

Second, the same pairwise distance matrix was used to assess genetic relatedness using the *neighbour-joining tree* method in the R *ape package*[173]. The *neighbour-joining tree* method is a cluster analysis technique, which does not require that all lineages have diverged by the same amounts. This method is particularly suit-

able for datasets comprising lineages with varying rates of evolution (eg different branch lengths).

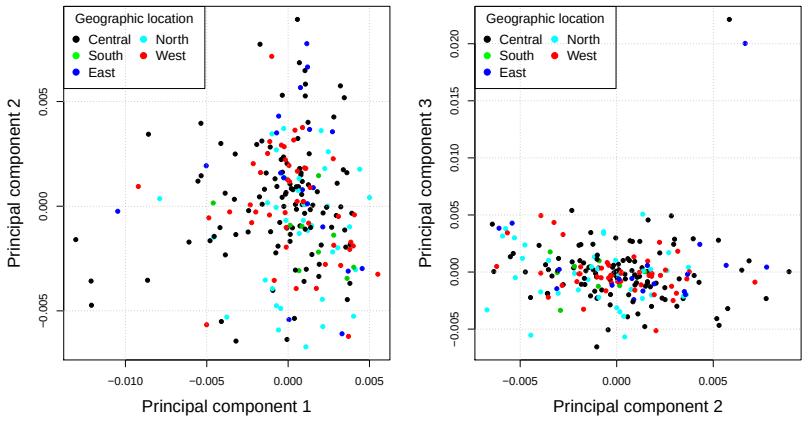
The PCA (subfigure 4.7(a)) did not reveal any substantial population structure among parasites sampled from the different micro-ecological regions of the KNDs. Additionally, as shown in subfigure 4.7(b) a *neighbour-joining tree* analysis did not show any distinctive clustering of parasites; both approaches suggest the absence of population structure (i.e random mating) in the KNDs parasite population. There is however a high level of diversity (branch length) that is uniform across the districts and indicative of a monophyletic parasite population likely to have originated from a common ancestor.

In order to assess the regional parasite population structure, I re-analyzed my data together with MalariaGEN published data from Burkina Faso and Mali. These are MalariaGEN open access data associated with the community project on *P. falciparum* population genomics[121]. Details of the individual studies may be reviewed at (`www.malariagen.net`). In brief, the Burkina Faso field isolates were collected from all age patients attending three clinics located in Colsamma, Ouezzin-ville and Sakaby, all in the urban area of Bobo-Dioulasso. This study was conducted by John Bosco Ouedraogo and colleagues to investigate parasite diversity and local population structure. The samples from Mali on the one hand, are from similar clinical studies conducted by Abdoulaye Djimde and colleagues. Parasite isolates were collected from clinics situated in Kolle and Faladje, two rural communities. As shown in Figure 4.8, no substantial population structure was still evident from this regional analysis. Parasites from all three countries in West Africa remained strongly intermixed. However, there were a few obvious

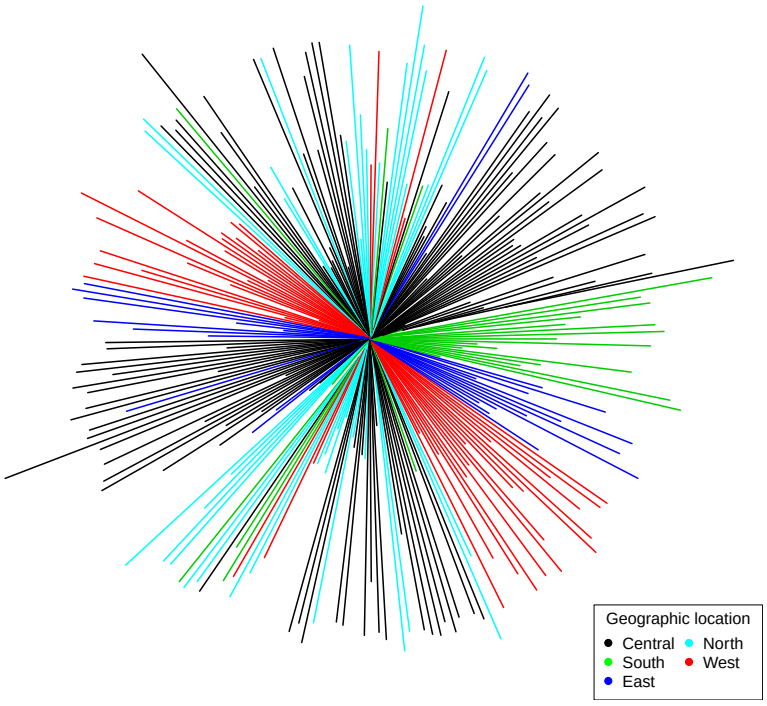
outlier parasites that appeared to be uniquely different from the bulk of the West African populations.

Finally, to assess global patterns previously reported, I extended my analysis to include further MalariaGEN published data from Kenya, Cambodia, Thailand and Papua New Guinea. The samples from Kenya were collected as part of clinical studies conducted in the Kilifi district. These isolates were adapted to laboratory culture for drug sensitivity testing by Allex Nzila and colleagues in the Kenyan Medical Research Institute, Kilifi. On the other hand, the Cambodian sample set was contributed by Xin Zhuan Su and Rick Fairhurst to the MalariaGEN from clinical studies conducted at a field site in Pursat province, Cambodia. The isolates from Thailand were collected from clinical studies as part of ongoing studies by Tim Anderson and Francois Nosten at a field site situated on the Thai-Burmese border.

I observed a marked contrast between the parasite population structure in south-east Asia and West Africa (Figure 4.9). As previously observed[121], a greater population structure was observed between sites in Cambodia and Thailand, whilst the West African populations (Ghana, Burkina Faso and Mali) still remained intermixed (Figure 4.9).

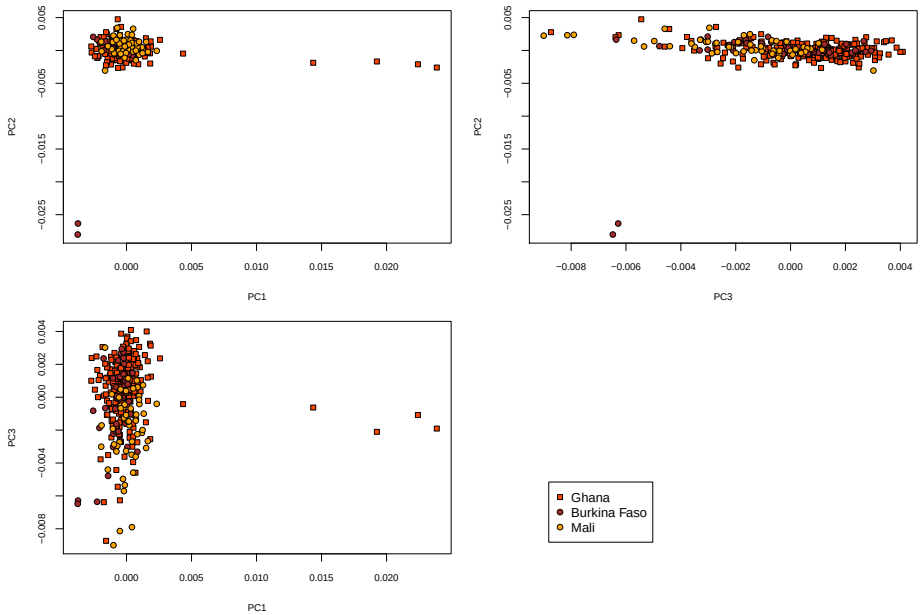


(a) First two principal components

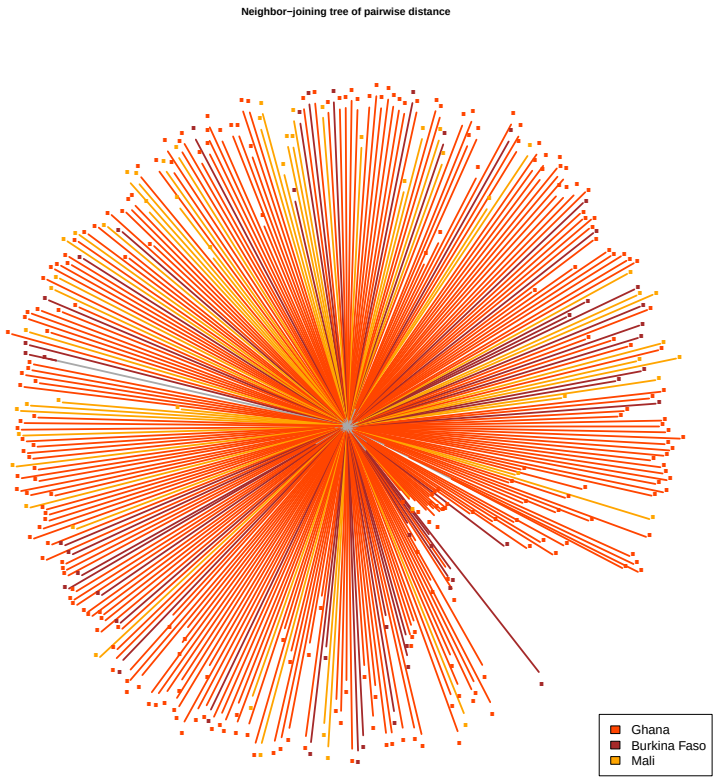


(b) Unrooted neighbour-joining tree.

Figure 4.7: **Genetic relatedness of parasites in Ghana**. In the top panel, subfigure 4.7(a) shows the first two PCA plots and in the bottom panel subfigure 4.7(b) shows a neighbour-joining tree analysis for population structure in Ghana. PCA points and Leaf braches in neighbour-joing tree are colour coded according to geographical area.

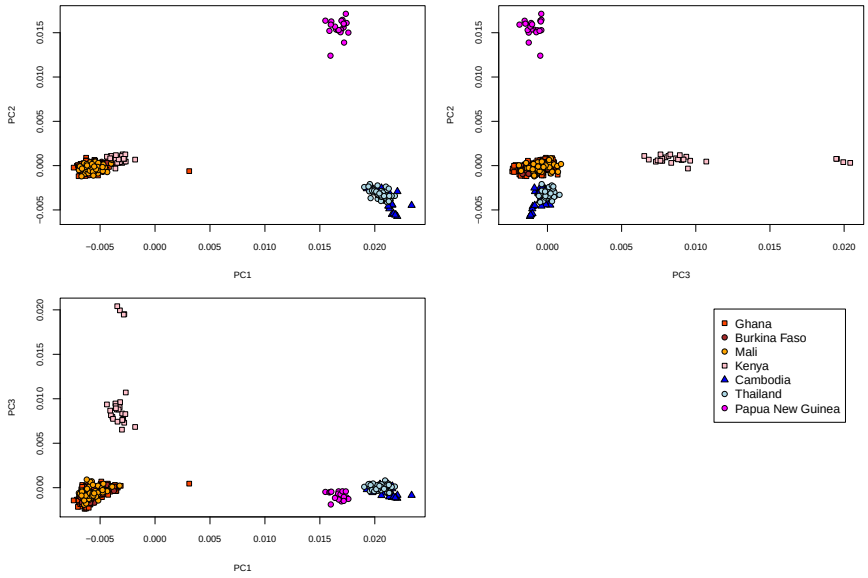


(a) First three principal components

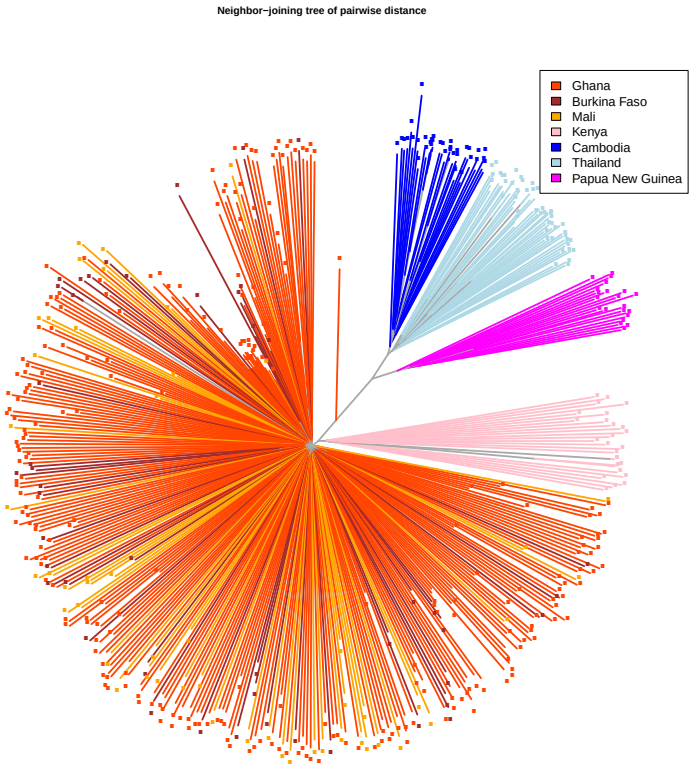


(b) Unrooted neighbour-joining tree.

Figure 4.8: **West African genetic relatedness of parasites**. In subfigure 4.8(a) PCA shows no distinct clustering of principal component points by country. Likewise neighbour-joining tree analysis shows an intermixed population. PCA points and Leaf braches in neighbour-joing tree are colour coded according to country.



(a) Plots of first three principal components



(b) Rooted neighbour-joining tree.

Figure 4.9: **Regional and global genetic relatedness of parasites**. PCA and neighbour-joining tree analysis revealed greater distinction between sites in southeast Asia but West Africa remains intermixed. PCA points and Leaf braches in neighbour-joing tree are colour coded according to country.

4.4.5 Epidemiological correlates of the $\hat{F}_{WS}$ metric

This study was set up to sample *P. falciparum* from communities located in different parts of the KNDs to reflect the socio-demographic and ecological variation in this region as important determinants of parasite genetic diversity. This section explores the $\hat{F}_{WS}$ metric as a measure of within-host diversity and the risk of out-crossing (or inbreeding) in relation to time and space, patterns of transmission intensity, clinical malaria and age at infection.

4.4.5.1 Time and space variation of $\hat{F}_{WS}$

As indicated earlier, $\hat{F}_{WS}$ estimates vary between 0 and 1, higher estimates imply low diversity (little or no population-level variability), which implies a high risk of inbreeding, and lower estimates are indicative of high within-host diversity (that is shared at the population-level) and thus high risk of out-crossing. As illustrated in Figure 4.6, the genetic complexity of malaria infections in the KNDs was enormous, and the risk of out-crossing high. But how is diversity distributed over time and location in this Savannah landscape with a short rainy season but a large irrigation dam and enough scatter of water bodies to maintain perennial transmission?

First, there was huge variability of $\hat{F}_{WS}$ scores within each month, with extreme (outliers) values (infections with very low diversity) occurring in January and June/July (the onset of the dry season and brink of the rainy season respectively) [Figure 4.10]. This period delimits the low transmission season, and suggests a shift towards low genetic diversity at low transmission, but, monoclonality is not

achieved because of the return of high transmission (leading to break down of the incipient low diversity). This pattern is supported by the observed occurrence of the two extremes of diversity (lower and high diversity) in the months of August through October (rainy season and peak of high transmission). These observations suggest a role for malaria transmission intensity in the structuring of within-host diversity and overall risk of out-crossing or breeding. The predominance of low $\hat{F}_{WS}$ estimates over time also underscore a potential for the rampant high diversity to mask useful information in the aggregated data, and suggests categorization of the data to yield useful interpretation.

In the light of this, I next assessed the geographical distribution (or ecological variation) of $\hat{F}_{WS}$ estimates using two approaches; (1) utilizing all the data, and (2) focusing on samples with low genetic diversity using a cut-off of $\hat{F}_{WS} > 0.5$. As illustrated in Figure 4.11 (left panel), there was large variation in $\hat{F}_{WS}$ estimates across the different micro-ecological locations in the KNDs, with the lowest and highest median scores occurring in the east and north respectively. The west however, showed the most extreme (whiskers) of $\hat{F}_{WS}$ scores. On the other hand, by focusing on samples with high $\hat{F}_{WS}$ scores (right panel), a pattern of occurrence of low genetic diversity (higher median scores) became apparent in the central, north and west that is suggestive of a relative potential for inbreeding. Conversely, infections with high genetic diversity were found in the East and South, that suggests an overall high risk of out-crossing in these subareas. Admittedly, there is a complex interplay of several variable epidemiological factors that could underscore the spatial distribution of malaria infection complexity. Unpacking the individual effects of these potential confounding variables to derive informa-

tion that could be of public health benefit remains an important challenge.

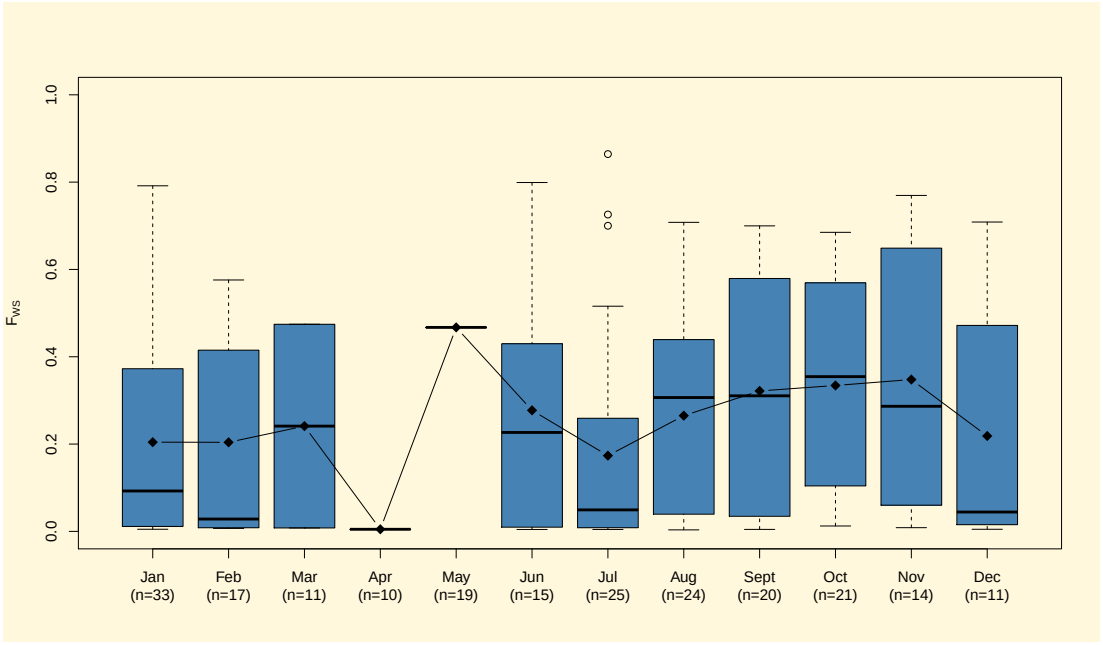


Figure 4.10: **Monthly distribution of $\hat{F}_{WS}$ scores.** *Monthly data presented are cummulative from August 2009 to November 2011.* Boxplot of $\hat{F}_{WS}$ by month sample was taken. Black line shows the distribution of monthly means (diamond shaped point). Boxes show the inter-quartile range and whiskers data extremes. April and May are the peak low transmission periods, the reason for fewer samples in these months.

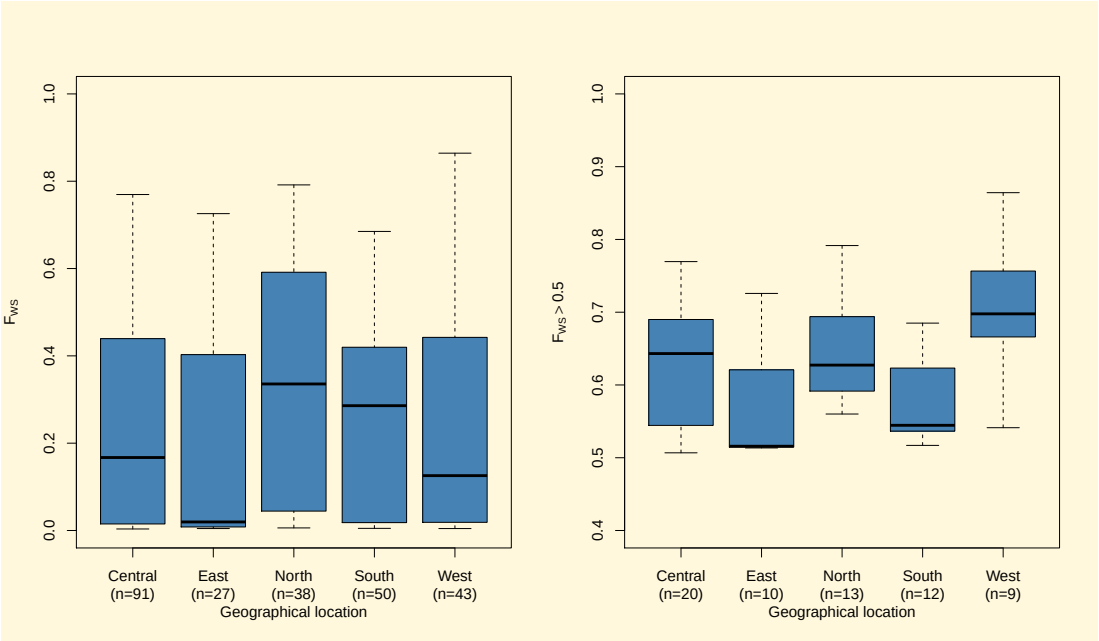


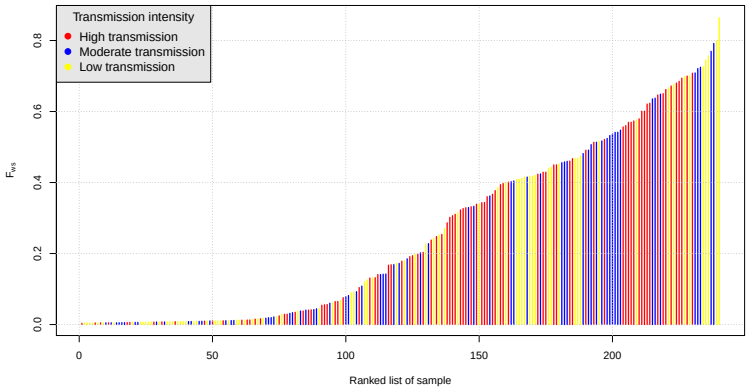
Figure 4.11: **Distribution of $\hat{F}_{WS}$ scores by geographical location.** The left plot shows the distribution of all the data and the right plot shows the distribution of $\hat{F}_{WS} > 0.5$.

4.4.5.2 $\hat{F}_{WS}$ and transmission intensity

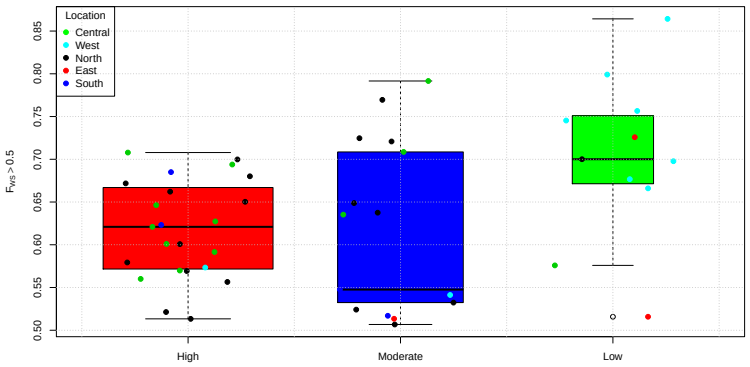
There are reported micro-ecological variations in EIR estimates in the KNDs with lower estimates in the rocky highland communities in the west[12]. The difficulty, however, lies in accurately delineating windows of transmission that would sufficiently reflect natural transmission patterns in the area. I relied on the monthly pattern of mean parasitaemias observed in this work (Chapter 3, *cf:* Figure 3.2) coupled with the malaria transmission literature[12,24,169] from the KNDs to determine monthly cut-offs and define windows of High, Moderate and Low malaria transmission seasons. I however, observed a sampling bias that led to significant differences between the number of samples obtained at these transmission seasons between geographical locations ($\chi^2 = 55.54$, p = 3.5e-09). Therefore, such a bias needed to be accounted for in the analysis of the effects of transmission seasons as defined.

In general, an analysis of variance (ANOVA) of the aggregate data did not reveal significant (F = 0.43, p = 0.51) differences between levels of infection diversity (or out-crossing potential) between high, moderate and low malaria transmission seasons in the KNDs (Figure 4.12, Subfigure 4.12(a)). The extent of genetic diversity in patient samples was more or less stochastic, as both low and high level diversity was observed across all transmission seasons. What could be interesting though, was the distribution of the randomly occurring low diversity infections in such a region with perennial intense malaria transmission. To this end, when I limited the analysis to $\hat{F}_{WS} > 0.5$, a pattern of distribution of estimates across these transmission windows, which is consistent with previous studies[7,11,121], begun to emerge (Figure 4.12; Subfigure 4.12(b)). An ANOVA showed that low

diversity infection and high risk of inbreeding was significantly more prevalent during low transmission compared with moderate to high transmission (F = 5.69, p = 0.006). After adjusting for the effect of geographical location, this difference remained statistically significant (F = 5.13, p = 0.01). The observed $\hat{F}_{WS}$ scores were however, highly heterogeneous, with considerable variability during moderate transmission. A period that is difficult to accurately define, owing to the fact that the transition between high and moderate transmission is rather gradual and dependent on when the last rainfall of the season occurred.



(a) Distribution of $\hat{F}_{WS}$ across transmission season



(b) Distribution of $\hat{F}_{WS} \geq 0.5$ by transmission season

Figure 4.12: **Within-host diversity and transmission season in the KNDs**. In the top panel (subfigure 4.12(a)), within-host diversity and risk of out-crossing as measured by $\hat{F}_{WS}$ is shown across three transmission windows (High, moderate and low). Subfigure 4.12(b) is a boxplot of $\hat{F}_{WS}$ estimates per transmission season. Boxes show inter-quartile range (black line in boxes are medians); whiskers extend to data extremes (outliers beyond whiskers).The points show the actual data (High= 23, Moderate = 17, Low = 11) per boxplot and colour coded by sampling location.

4.4.5.3 Relationship between $\hat{F}_{WS}$ and clinical malaria

Three categories of $\hat{F}_{WS}$ estimates were used to correlate different levels of genetic diversity with clinical malaria (Table 4.3). I classified estimates ≤ 0.25 as extreme genetic diversity, and estimates in the range $0.25 < F_{WS} < 0.5$ as high genetic diversity, whilst $\hat{F}_{WS}$ estimates > 0.5 were denoted as low genetic diversity at several thousands of loci.

Again, I observed heterogeneity in $\hat{F}_{WS}$ scores among severe and mild malaria patients (Figure 4.13). The Fisher exact test, however, did not show statistically significant association of $\hat{F}_{WS}$ estimates and clinical malaria (OR:1.59,CI:0.34-10.45, p = 0.73). Severe malaria infections had a mean $\hat{F}_{WS}$ scores of 0.31 and mild malarias 0.25 but the difference was not statistically significant by the Wilcox rank sum test (p = 0.99). Overall, 80% of mild malaria infections and 76% of severe malarias exhibited high genetic diversity (ie $\hat{F}_{WS}$ estimates $\leq$0.5)[Table 4.3].

4.4.5.4 Relationship between $\hat{F}_{WS}$ and Age

I observed a negative correlation between age at infection and the genetic diversity of the infection (Figure 4.14). This is supported by other studies, which suggest a decreasing multiplicity of infection with age in areas of intense malaria transmission[90,170]. By categorizing age to reflect the relative risk of severe malaria disease (or conversely semi-immunity), I observed a distribution of $\hat{F}_{WS}$ estimates, which showed high diversity in children under 5 years and low diversity in older patients (Figure 4.14).

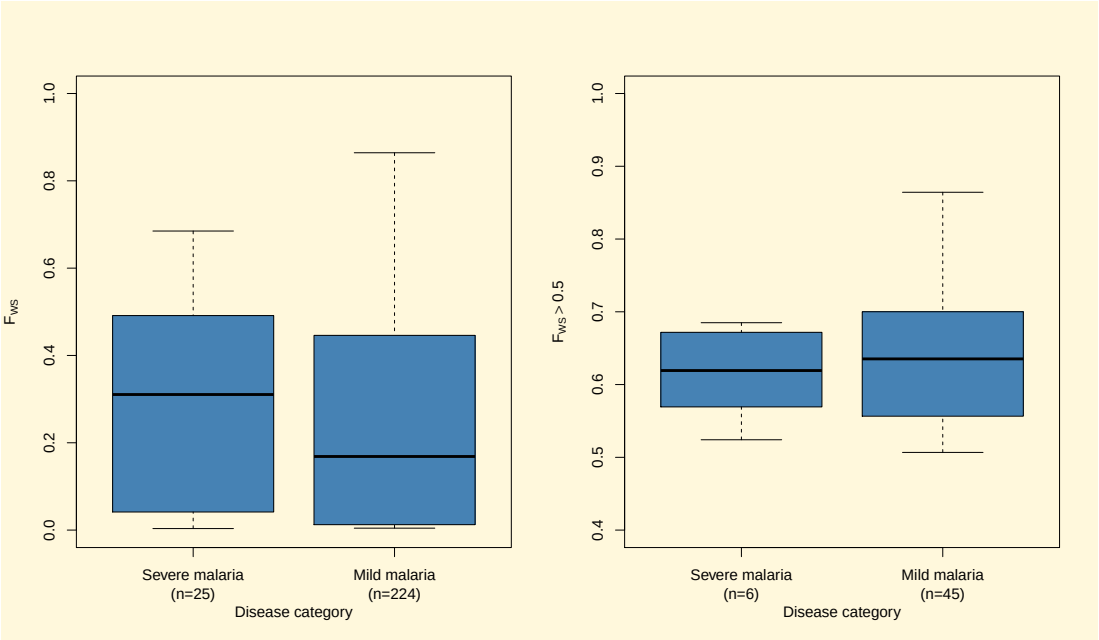


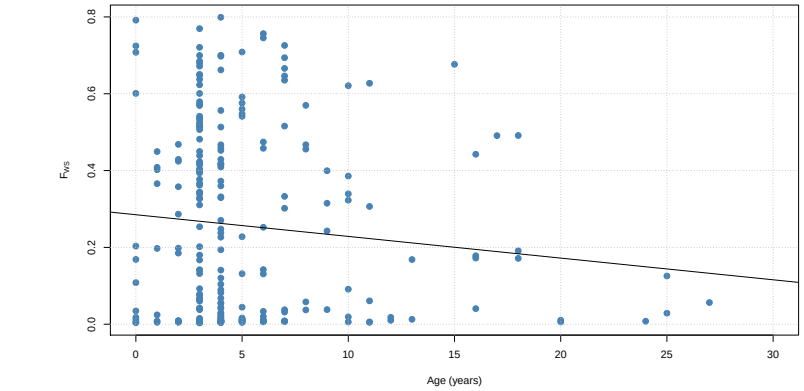
Figure 4.13: **Distribution of $\hat{F}_{WS}$ scores among disease phenotypes.** Boxplot (left) of $\hat{F}_{WS}$ estimates per disease phenotype. Boxplot (right) of $\hat{F}_{WS} > 0.5$ per disease phenotype. Both plots show considerable heterogeneity of $\hat{F}_{WS}$ estimates among disease phenotypes

.

	Mild malaria N=224				Severe malaria N=25		
Diversity	$\hat{F}_{WS}$ values	n	Mean	Range	n	Mean	Range
Low	> 0.5	45	0.64	0.51-0.86	6	0.61	0.52-0.69
High	0.25-0.5	52	0.39	0.25-0.49	8	0.40	0.25-0.49
Extreme	≤ 0.25	127	0.05	0.00-0.24	11	0.06	0.00-020

$\chi^2 = 0.75, P = 0.68$

Table 4.3: **Variation of $\hat{F}_{WS}$ with disease phenotypes.**



(a) Relationship between $\hat{F}_{WS}$ scores and age at infection

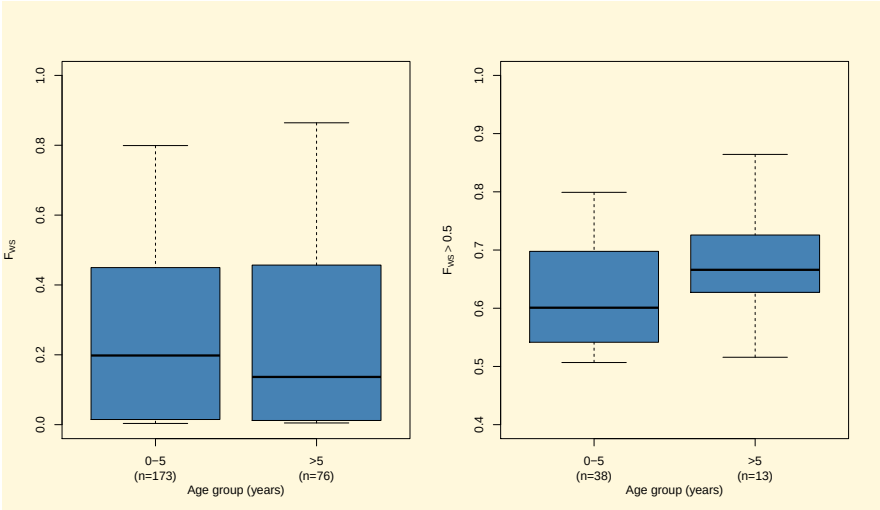


Figure 4.14: **Distribution of $\hat{F}_{WS}$ by age category.** Relationship between within-host diversity and age at infection

4.5 Discussion

The allelic architecture of *P. falciparum* populations in Northern Ghana is characterized by low frequency variants and is primarily dominated by rare variants (<5%) (Figure 4.3). The genome-wide over representation of low frequency variants in the KNDs parasite population is further support for a general low frequency allelic spectrum in areas of high malaria transmission, as observed in previous studies in other West African countries (Burkina-Faso and Mali) known to have perennial high malaria endemicity[121].

The lack of significant population structure either at local levels in Ghana or the West African (sites in Burkina Faso and Mali, are ≈ 500km apart and ≈ 800km between Ghana and Mali) sub-region by PCA and neighbour-joining tree analysis (Figure 4.8) is further evidence of a genetically intermixed and closely related parasite population across this whole region. Particularly among parasites in the respective local populations. This study confirms that these observations are not due to small sample size or culture issues but are a real biological effect. Also apparent was a group of outliers that show some level of diversification from the bulk of the population in Ghana (subfigure 4.8(a)). This small group of parasites may be diversifying from the local population as it were or experiencing some kind of population bottleneck, but, at this stage, I am unable to characterize these any further because of the rather small number of samples involved. However, this definitely requires farther investigation. The parasite population dynamics in West Africa is in sharp contrast to sites sampled in Cambodia and Thailand (≈ 800km apart) by malariaGEN, where a greater population structure and high

potential for inbreeding was observed[121]. This may be driven by transmission intensity, which is relatively high but variable across West Africa. Estimates range between 10-25 infective bites per person per month during the rainy season at the sites sampled in Mali[83], and about 17.0 bites per person per night in the KNDs, Ghana (Asoala et al.,2013, manuscript in preparation), and under 5 infective bites/person-years in the sites in Cambodia and Thailand[121]. These findings are consistent with previous evidence of panmixia in parasite populations under intense transmission[7,11,174]. Also relevant in this regard, may be the demography and settlement patterns in the West African countries sampled (Ghana, Burkina Faso and Mali), where a scater of large clusters of villages in a Sahel Savannah grassland provides conducive conditions for superinfection and random mating.

This study observed a high level of multi-genome infections, defined here as the genome-wide fraction of heterozygous SNPs (loci with both reference and non-reference genotypes) per infection [Figure 4.4]. This suggests a high genetic complexity of infections that will be extremely challenging to capture by the traditional method of assessing multiplicity of infection (MOI). Conventionally, the MOI of clinical samples is estimated using PCR to assess the number of repeat length variants observed at a few highly diverse genes, most frequently variable number tandem repeats in the genes encoding the merozoite surface protein 1 (MSP1) and 2 (MSP2) and the *P.falciparum* Glutamate rich protein (GLURP)[15,170,211,231].

Manske *et al.,* (2012)[121], introduced the $\hat{F}_{WS}$ metric for assessing within-host diversity and relating it to the population-level diversity using genome-wide SNP data. This metric captures not only the overall genetic diversity within an in-

dividual but also the level of similarity between the parasites at the population-level[15,121]. Thus, it provides a basis to gauge the risk of out-crossing (or in-breeding) in a given population. In this chapter, I explored this metric in a specific epidemiological setting in northern Ghana with perennial intense malaria transmission, and provided some relevant clarity about this metric that was not apparent in the original publication. I then used the metric to estimate the genetic diversity of infections at several thousands of loci in over 200 samples to provide population specific evaluation of this new statistic.

First, the $\hat{F}_{WS}$ estimates were indicative of an overall high within-host diversity and risk of out-crossing in this population (Figure 4.6). The data showed a high likelihood of out-crossing during sexual recombination in the mosquito mid-gut to provide new recombinant progeny. The variability of $\hat{F}_{WS}$ estimates over a time-scale in months (Figure 4.10) appeared to mimic the general pattern of the malaria case burden in this region, with extreme variability coinciding with the rainy months of August through October during which hospital admissions are reported to peak[157]. Recent transmission studies estimate 17.0 infective bites per person per night during the rainy season and 8.0 during the dry season (Asoala et al.,2013, Manuscript in preparation). Of note in the time course of $\hat{F}_{WS}$ variation, however was the spike of low genetic diversity (high $\hat{F}_{WS}$ estimates) observed in January, which marks the onset of low malaria transmission and in June/July which mark the onset of rains and beginning of high transmission. This period delimits the low transmission season and suggests a cyclical wave of genetic diversity, where during low transmission prevalence of infections with low genetic diversity begins to increase, but it is not long enough before the onset of rains and

high transmission. Thus, interrupting the incipient selection of parasites with low genetic diversity and breaking up whatever haplotypes were under selection via recombination (during high transmission). This seasonal cycle of genetic diversity suggests a propensity for mutations that confer a fitness advantage (eg drug resistance mutations) to rapidly spread, but, may struggle to reach fixation (or high frequencies). This model is supported by the evidence from *pfcrt*, when the 76T mutation associated with chloroquine resistance had risen to high frequencies in East Africa and Asia, but, remained under 50% frequency for a long time in West Africa[44,50].

Second, In terms of spatial distribution, there was high variability in the location specific $\hat{F}_{WS}$ estimates, with estimates from communities in the north and south showing a higher median score of $\hat{F}_{WS}$ (ie low genetic diversity) (Figure 4.11). The distribution of samples with low genetic diversity revealed interesting findings that were previously masked by the aggregated data; that communities in the central district (both urban and peri-urban to rural) that enjoys unfettered access to health interventions, similar to communities in the North may have a higher prevalence of low genetic diversity (high risk of inbreeding). Also in this group are the isolated communities in the rocky highland subareas west of the KNDs, where EIRs are reported to be comparatively lower[12]. Conversely, communities in the east and south tended to have high genetic diversity (low median $\hat{F}_{WS}$ scores), suggestive of an overall high risk of out-crossing in these suburbs.

Third, to define transmission intensities (high, moderate and low) for assessing the effects of varying transmission on $\hat{F}_{WS}$ scores, one challenge worth highlighting was the difficulty in accurately delineating a season for moderate transmission.

This season is heavily dependent on when the last couple of rains during high transmission occurred to allow for a gradual onset of a rather short moderate transmission season. Hence, the huge variability of $\hat{F}_{WS}$ scores observed at this season underscores this difficulty. Consistent with studies that have assessed within-host diversity in areas of moderate to low malaria transmission[7,11,174], I observed significant over representation of infections with low genetic diversity during the low transmission (adjusted P = 0.01)[Figure 4.12]. This is consistent with a recent study, which had global *P. falciparum* representation found low genetic diversity in countries sampled from Southeast Asia were transmission is generally low[121].

Fourth, in terms of the relationship between $\hat{F}_{WS}$ estimates and clinical malaria in the KNDs, overall, estimates of $\hat{F}_{WS}$ were heterogeneous with a wide range of variability within severe and mild malaria infections. There was no significant difference in the distribution of $\hat{F}_{WS}$ estimates between the two. Some previous studies have shown an inverse association between MOI (estimated by conventional methods) and clinical malaria attacks[5,69,145], whilst others have shown a positive correlation between MOI and clinical malaria[28,131,159]. In general, consistent with a negative correlation, a high proportion of samples exhibited low $\hat{F}_{WS}$ estimates in both disease categories, 80% of severe malaria and 76% of mild malaria (Table 4.3).

Finally, I observed a negative correlation of the distribution of $\hat{F}_{WS}$ estimates by age at infection (Figure 4.14). This finding is supported by other studies that assessed the relationship between age and malaria infection complexity in areas of high malaria endemicity[159,170].

Taken together, the mean $\hat{F}_{WS}$ score was 0.25 (range: 0.00-0.86) for the KNDs parasite population (Table 4.3). A high proportion, 88% (218/249) of infections had $\hat{F}_{WS}$ scores < 0.6, and 51% were of extremely high genetic diversity ($\hat{F}_{WS}$ scores ≤ 0.25), showing over representation of high genetic diversity in this population. Conversely, no samples exhibited extreme low genetic diversity ($\hat{F}_{WS}$ scores $\geq$0.95) indicative of monoclonal infections. The mean $\hat{F}_{WS}$ estimate in the northern Ghana is much lower than that observed in Mali (ie 0.81; $\hat{F}_{WS}$ 29% $\geq$0.95) and Burkina-Faso (ie 0.69; $\hat{F}_{WS}$ 30% $\geq$0.95)[15,121], reflecting a higher *P. falciparum* population diversity and risk of out-crossing in this setting compared to the areas sampled in Burkina Faso and Mali. The absence of population structure by PCA and phylogenetic analysis are key evidence of a large degree of random mating in this parasite population. Evidence that is consistent with a lack of population differentiation in parasite populations sampled in West Africa[121].

4.6 Summary

- The analysis undertaken in this chapter has shown that the allelic spectrum of *P. falciparum* population in the KNDs of Northern Ghana is dominated by low frequency variants.

- The evidence supports random mating and a close genetic relatedness in the parasite population both in Ghana and West Africa as a whole; notwithstanding, a few outlier populations in Ghana.

- A high degree of genome-wide within-host diversity and risk of out-crossing estimated by the $\hat{F}_{WS}$ metric was observed, but, with occasional risk of

inbreeding (particularly within communities in the west of KNDs).

- An overall mean $\hat{F}_{WS}$ score of 0.25 was observed, and a high proportion, 88% of infections had $\hat{F}_{WS}$ scores < 0.6, and 51% were of extremely high genetic diversity ($\hat{F}_{WS}$ scores ≤ 0.25), showing over representation of high genetic diversity in this population.

- There was high variability of $\hat{F}_{WS}$ estimates between different geographical location sampled in the KNDs.

- Samples with low genetic diversity were significantly more prevalent in the low transmission season

- No correlation of $\hat{F}_{WS}$ estimates with mild or severe malaria was observed, though low estimates (high genetic diversity) were more prevalent in both disease phenotypes, suggesting a correlation with clinical malaria.

- A negative correlation was observed between $\hat{F}_{WS}$ estimates and age at infection and higher prevalence of infections with low genetic diversity was found in older children and adults.

Chapter 5

Molecular Assessment of Antimalarial drug selection

5.1 Introduction

Sulfadoxine-pyrimethamine (SP) was introduced in Ghana in the 1980s first, as a rescue drug for chloroquine (CQ) treatment failures[194], but, quickly became a preferred first-line treatment for its single dose and absence of side effects. By 2004 as a consequence of a worsening burden of malaria due to wide-spread CQ and SP resistance[50,55,139,148], ACT was introduced in Ghana, with Artesunate-amodiaquine becoming the front-line treatment for uncomplicated malaria[156]. However, to date, SP remains the sole drug for intermittent preventive treatment of malaria in pregnancy (iPTP) in Ghana[79].

The genetic basis of resistance to sulfadoxine and pyrimethamine has been shown to be point mutations in the genes encoding their target enzymes-dihydropteroate

synthase (DHPS) and dihydrofolate reductase (DHFR) respectively[155,175,194]. The basal mutation known to confer CQ resistance (CQR) is the K76T mutation in the *P. falciparum* chloroquine resistance transporter gene (*pfcrt*)[44,61,139]. Resistance to SP is caused by the point mutations that accumulate at several sites in the *pfdhfr* and *pfdhps* genes[194,233]. In both genes, sequential acquisition of mutations incrementally increases the parasite's tolerance to the drug *in vitro*[194,224]. Recent unpublished data from a multi-centre phase IV study in the KNDs, revealed continued use of CQ via the private health sector, and common use of artemisinin derivatives such as artesunate and dihydroartemisinin as monotherapies (Atuguba *et al.,* personal communication). More importantly the confirmed reports of evolution of a heritable trait that confers resistance to artemisinin (prolonged clearance of parasitaemia) in Cambodia[47], and its rapid spread to western Thailand[33], advocates monitoring of drug resistance profiles in endemic communities where ACTs have been deployed.

The goal of this chapter is to use deep sequencing data to characterize known loci under antimalarial drug selection. In particular, to ascertain the prevalence of *pfcrt*, *pfdhfr* and *pfdhps* resistant alleles, and assess the pattern of variation of CQ resistance alleles (*pfcrt*76T) six years after withdrawal of CQ as first-line treatment for malaria in Ghana.

5.2 Objectives

- To assess the population prevalence *pfdhfr*, *pfdhps* and *pfcrt* mutations in the KNDs.

- To ascertain the prevalence of inferred resistance-associated haplotypes in this population.
- To assess the geographic variation in the prevalence of resistant alleles in these genes.

5.3 Materials and Methods

5.3.1 Samples

I used the full set of 277 *P. falciparum* field isolates that were sequenced by Illumina and discussed in chapter 4. These samples were collected from five clinics across the KNDs (see chapter 3). *Plasmodium falciparum* DNA was extracted directly from patient blood using the QIAamp DNA Blood Midi/Maxi kit (Qiagen) as per manufacturer's protocol, after removing WBCs. Further details may be reviewed in Chapter 2 section 2.2.1.

5.3.2 SNP Genotyping

This chapter focuses on SNPs present in *pfdhfr*, *pfdhps* and *pfcrt*, and specifically, five point mutations encoding five amino acid changes in the *pfdhfr* gene (A16I, N51I, C59R, S108N, and I164L), the *pfdhps* gene (E189Q, S436A, A437G, L540E, and A581G) and six point mutations encoding amino acid changes in the *pfcrt* gene (M74I, N75D, K76T,Q271E and I326T) [Table 5.1]. These SNPs were detected as part of an established pipeline for the malariaGEN community project

on *P. falciparum* genome variation, which is described in chapter 4 (see section 4.3.6).

5.3.3 Analysis

I used the open source statistical programme R to analyze the data[189]. I performed simple cross tabulations to assess the frequencies of resistant alleles, and used the Fisher's exact and chi-square test to assess statistical significance in all comparisons. All confidence levels were set at 95%.

Polymorphism	Gene	Chr position	REF allele	NREF allele	NRAF
DHFR-TSA16V	PFD0830w	755115	A	T	1.00
DHFR-TSN51I	PFD0830w	755220	A	T	0.47
DHFR-TSC59R	PFD0830w	755243	T	C	0.65
DHFR-TSS108N	PFD0830w	755391	C	A	0.35
DHFR-TSI164L	PFD0830w	755558	A	T	0.01
DHPSE189Q	PF08_0095	550061	G	C	0.76
DHPSS436A	PF08_0095	550802	T	G	0.62
DHPSG437A	PF08_0095	550806	G	C	0.15
DHPSK540E	PF08_0095	551114	A	G	0.01
DHPSA581G	PF08_0095	551238	C	G	0.01
PfcrtM74I	MAL7P1.27	458998	G	T	0.24
PfcrtN75D	MAL7P1.27	458999	A	G	0.23
PfcrtK76T	MAL7P1.27	459003	A	C	0.24
PfcrtQ271E	MAL7P1.27	460214	C	G	0.27
PfcrtI326T	MAL7P1.27	460978	T	C	0.02

Table 5.1: **Drug resistance loci that had sufficient read coverage for genotyping**. All are previously identified point mutations except Glu-Gln substitution in DHPS codon 189 which is novel in this population. Also shown are the reference (REF) and non-reference (NREF) alleles and non-reference allele frequencies (NRAF).

5.4 Results

Owing to poor mapping coverage sufficient for calling genotypes in some samples, 11, 15 and 37 samples out of the 277 sequenced were dropped from this analysis for *pfdhfr*, *pfdhps* and *pfcrt* respectively. As described in chapter 4, the presence of an allele required a minimum threshold of 5 read counts as evidence for genotyping.

The overall frequency of drug resistant alleles was 82% (219/266) for *pfdhfr*, 77% (202/262) for *pfdhps*, and 34% (81/240) for *pfcrt* (Tables 5.3 and 5.4). The proportion of infections that were fully sensitive (ie no known mutations at any loci) was 18% (47/266) for *pfdhfr*, 22% (57/262) for *pfdhps* and 66% (159/240) for *pfcrt*. I observed 26 samples (10%) that had sensitive alleles across all three loci (Table 5.2).

As shown in Table 5.2, the prevalence of drug resistant genotypes (homozygous at any one or more drug resistance mutations) was 15% for *pfdhfr*, 46% for *pfdhps* and 20% for *pfcrt*. Many loci were heterozygous for both sensitive and resistant alleles at any one or more drug resistant SNPs. Thus, populations of different combinations of SNPs (haplotypes) were present in any given sample (Table 5.3). The prevalence of these heterozygous alleles across all the polymorphic positions studied in all three genes is shown in Figure 5.2. As a result of such multi-genome infections (Table 5.2), it is difficult to accurately determine the true nature of drug resistant haplotypes. However, by genotyping each sample with the allele with majority read counts, I was able to infer combinations of alleles for the three loci that confer different levels of drug resistance.

For *pfdhfr*, I observed 3% and 22% of samples with a single mutation at codons 51 (N51I) or 59 (C59R) respectively, and 67% of samples had double mutations (C59R+ S108N). Two triple resistant alleles were found; 14% of samples had the triple resistant alleles (N51I + C59R + S108N) and 0.4% triple resistant at codons 51, 59 and 164 (N51I + C59R + I164L (Table 5.5). The *pfdhfr* triple resistant allele (N51I + C59R + S108N) is known to confer higher levels of pyrimethamine resistance and is strongly correlated with SP treatment failure[149]. This triple resistant was observed across the geographical areas sampled within the KNDs.

I observed single resistant alleles for *pfdhps* at three SNPs (S436A, G437A and K540E). The K540E substitution is comparatively very rare and was found only the Central district. I did not find the *pfdhps* double resistant allele (G437A + K540E) but the *pfdhps* double resistant allele (S436A + G437A) was found at 23% prevalence (Table 5.5). Also evidenced was a new variant not previously described (E189Q), which was highly prevalent in the KNDs and was observed in the background of the either S436A or G437A resistant alleles (Table 5.5).

The prevalence of *pfcrt* mutations were 0.4% double mutations (M75D + K76T), 29% quadruple mutations (M74I + M75D + K76T + Q271E) and 4% with quintuple resistant allele [M74I + M75D + K76T + Q271E+ I326T] (Table 5.6). All inferred haplotypes in *pfcrt* involved the 76T primary mutation known to confer chloroquine resistance. The prevalence of the *pfcrt* 76T resistant allele alone was 23%, a 50% reduction from pre-ACT prevalence estimates.

One advantage of this dataset is to be able to determine the presence of mutations in any or all of the three loci. This is important as it is known that multiple

mutations in *pfdhfr* plus additional mutations in *pfdhps* in the same sample results in higher levels of SP resistance[182,224]. As shown in Table 5.7, I observed parasites with multiple mutations in both *pfdhfr* and *pfdhps*. The most prevalent haplotype was a triplet resistant (51I + 59R / 437A) found in 89 (34%) samples. Also fairly common was a quadruplet resistant allele (51I + 59R + 108N / 437A) which was found in 26 samples, and a much less common quadruplet resistant allele (51I + 59R / 436A+437A), which was slightly more prevalence in the communities south of the KNDs and absent in the east and west.

Allele type	*pfdhfr*	*pfdhps*	*pfcrt*	Any
Sensitive	47	57	159	26
Heterozygous	179	85	33	34
Resistant	40	120	48	164
Discarded samples	11	15	37	53
Total count	277	277	277	277

Table 5.2: **Counts of three groups of parasite genotypes**. Sensitive refers to homozygous for sensitive SNP positions (fully sensitive), Heterozygous refers to the presence of sensitive alleles and one or more drug resistance SNPs, and Resistant refers to homozygous for any one or more drug resistance SNPs.

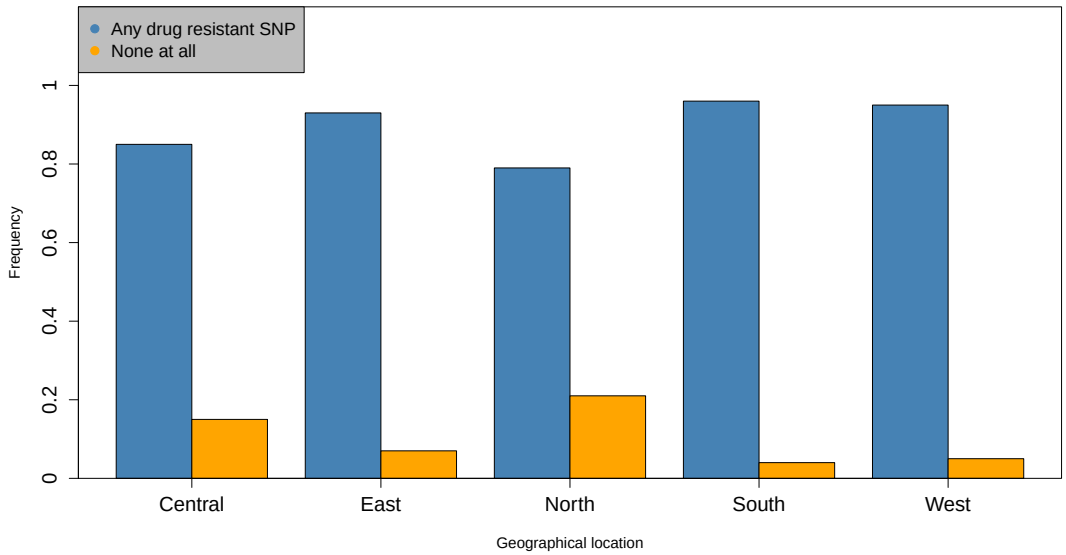


Figure 5.1: **Frequency distribution of samples with any mutation in all three loci *pfdhfr*, *pfdhps* and *pfcrt***

	SNP combinations				
Gene	1	2	3	4	Count
pfdhfr	40 (15%)	142 (53%)	36 (14%)	1 (0.4%)	219
pfdhps	146 (54%)	60 (23%)	3 (1.1%)	-	206

Table 5.3: **Proportion of infections with different numbers of drug resistance SNPs for *pfdhfr* and *pfdhps* loci**. The table does not include homozygous sensitive that are fully sensitive (0 loci)

	SNP combinations			
Gene	1	4	5	Count
pfcrt	1 (0.4%)	70 (29%)	10 (4%)	81

Table 5.4: **Proportion of infections with different combinations of drug resistance SNPs for *pfcrt***. The table does not include homozygous sensitive that are fully sensitive (0 loci)

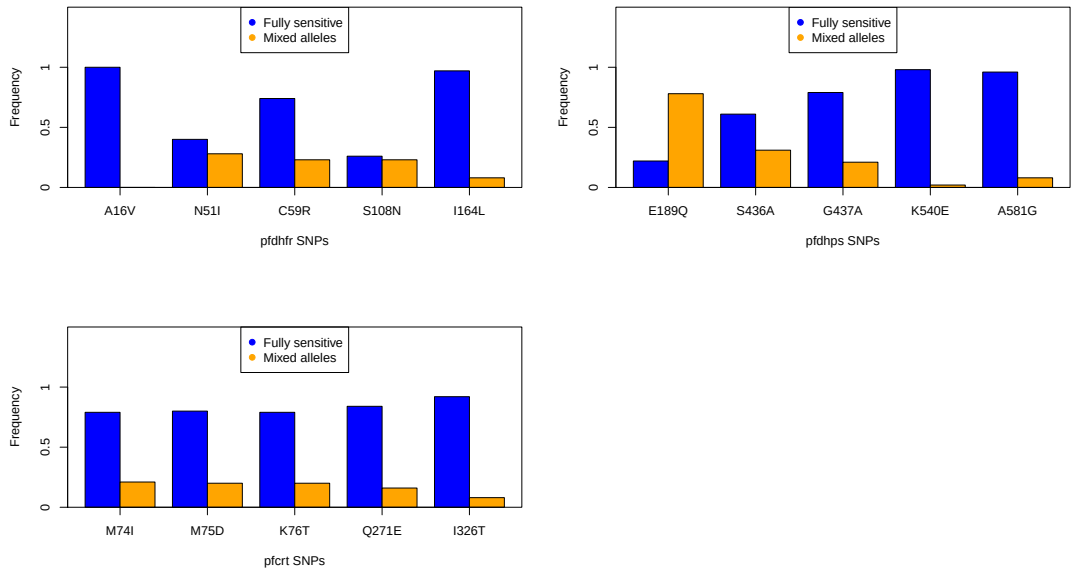


Figure 5.2: Frequency of mixed alleles and homozygous sensitive alleles at five codon positions in *pfdhfr*, *pfdhps* and *pfcrt*

pfdhfr (N=266)						
Type	16	51	59	108	164	n (%)
Nucleotide change	A→T	A→T	T→C	C→A	A→T	
Sensitive	A	A	T	C	A	47 (18)
Resistant	A	**T**	**C**	C	A	160 (60)
Resistant	A	A	**C**	**A**	A	18 (6.8)
Resistant	A	**T**	**C**	**A**	A	36 (14)
Resistant	A	**T**	**C**	C	**T**	1 (0.4)
Codon change	A→V	N→I	C→R	S→N	I→L	
pfdhps (N=262)						
Type	189	436	437	540	581	n (%)
Nucleotide change	G→C	T→G	G→C	A→G	C→G	
Sensitive	G	T	G	A	C	57 (22)
Resistant	**C**	**G**	G	A	C	61 (23)
Resistant	**C**	T	**C**	A	C	81 (31)
Resistant	**C**	**G**	**C**	A	C	59 (23)
Resistant	G	T	G	**G**	C	4 (1.9)
Codon change	E→Q	S→A	G→A	K→E	A→G	

Table 5.5: **Frequency of inferred haplotypes for *pfdhfr* and *pfdhps*.** Amino acid residue positions are shown along the top and the resulting amino acid substitutions are shown underneath. Substitutions that confer resistance are shown in bold.

pfcrt (N=240)						
Type Nucleotide change	74 G→T	75 A→G	76 A→C	271 C→G	326 T→C	n (%)
Sensitive	G	A	A	C	T	159 (66)
Resistant	**T**	**G**	**C**	**G**	T	70 (29)
Resistant	**T**	**G**	**C**	**G**	**C**	10 (4.2)
Resistant	G	**G**	**C**	C	T	1 (0.4)
Codon change	M→I	N→D	K→T	Q→E	I→T	

Table 5.6: **Frequency of inferred haplotypes for *pfcrt*.** Amino acid residues positions are shown along the top and the resulting amino acid substitutions are shown underneath. Substitutions that confer resistance are shown in bold.

	Geographic location					
Resistant allele	Central (100)	East (28)	North (38)	West (44)	South (52)	Total 262
pfdhfr/pfdhps						
None	57	11	21	29	14	132
51I+59R/437A	32	12	12	13	20	89
51I+59R+108N/437A	7	5	4	2	8	26
51I+59R/436A+437A	4	0	1	0	10	15
Total count	43 (43%)	17 (61%)	17 (45%)	15 (34%)	38 (73)	262

Table 5.7: Geographic distribution of parasites with multiple point mutations in both *pfdhfr* and *pfdhps*.

5.5 Discussion

From deep sequencing of 277 Ghanaian field isolates, I identified from the KNDs a subset of the world-wide well-characterized polymorphisms in *pfdhfr*, *pfdhps* and *pfcrt* known to occur in Africa[117,148,149,194]. A high level of heterozygosity (a mixture of sensitive and drug resistant alleles) was observed and these mutli-genome infections can be problematic when inferring haplotypes. However, by assigning one genotype per infection using the allele with majority read counts, and discounting the minority alleles (<5 reads), I was able to infer haplotypes for all three loci. In general, common combinations of SNPs did not exceed three in *pfdhfr* and *pfdhps* (Table 5.3).

The observed prevalence of sensitive alleles for *pfdhfr* was 18%, which suggest a predominance of drug resistant alleles in the KND population. For pyrimethamine (Pyr) resistance, the basal mutation is a Ser-to-Asn substitution at codon 108 of DHFR, which has been demonstrated to result in a 100-fold decrease in sensitivity[175,251]. In the KNDs no single resistants at codon 108 were found. This mutation occurred in a background of other polymorphisms, consistent with the observation that resistant *pfdhfr* alleles are varied and code for a range of tolerance to Pyr from intermediate to high depending on the number of mutations present[149,251]. For *pfdhfr* single resistant alleles were observed for codons 51 and 59. The predominant *pfdhfr* resistant haplotype (60%) was a double resistant at codon 51 and 59 (51I/59R) with a lower proportion (7%) of parasites having the alternative double resistant *pfdhfr* alleles (59R/108N) known to confer intermediate Pyr-resistance[194]. Whereas the more highly resistant *pfdhfr* triple resistant

allele (51I/59R/108N), known to be a significant contributor to SP treatment failure[149], was observed at 14% prevalence. Also evidenced was a rare *pfdhfr* triple resistant allele (51I/59R/164L), however, there is lack of *in vitro* data on resistance perspective of this haplotype.

Parasites with *pfdhps* sensitive alleles had a prevalence of 22%. I found single *pfdhps* resistant alleles at codons 436, 437 and 540 but did not observe the presence of the *pfdhps* double resistant allele (A437G + K540E), which confers higher sulfadoxine resistance. The K540E mutation occurred as a single mutation in a minority of samples (0.4%). However, I found the *pfdhps* double resistant allele (S436A + G437A) to occur at a prevalence of 23% in the KNDs population. Also evidenced was extremely high frequencies (>70%) of the 189Q polymorphisms across the KNDs. This locus has not been previously associated with any resistance phenotype (Figure 5.2). However, it underscores the utility of deep sequencing data for validating as well as discovering new variants that may be population specific and could contribute to the overall sensitivity of the DHPS enzyme.

A high prevalence of drug resistant SNPs in *pfdhfr* and *pfdhps* is indicative of SP resistance at the population-level. Sulfadoxine-pyrimethamine targets the DHPS and DHFR enzymes in the folate biosynthesis pathway of *P. falciparum* , acting in synergy to disrupt folate synthesis and kill the parasite[155,194,224]. It has also been demonstrated that, resistance to SP is exacerbated when parasites have multiple mutations in the *dhfr* gene, and additional mutations in *dhps*[182,224]. Therefore the observed multiple mutations in both genes involving the *pfdhfr* triple resistant (51I/59R/108N) known to correlate with SP treatment failure[149], and

G437A basal mutation in the *pfdhps* locus, is a strong indication of established SP resistance.

These data have direct implications for the continuous use of SP for intermittent preventive treatment of malaria in pregnancy (IPTP) in this populations. The high prevalence of heterozygous genotypes in *pfdhfr* and *pfdhps*, brings into focus the clinical implications of these mixed alleles. Pregnant women who currently receive SP as part of the IPTP programme may be at risk of clinical malaria due to selection of resistant genotypes by SP given as a prophylactic. This data suggests a re-evaluation of the programme in Ghana and Sub-Saharan Africa in general.

On the other hand, the observed high prevalence of *pfcrt* sensitive alleles (quinoline drug sensitive parasites) across the KNDs suggests that prevalence of the *pfcrt* K76T primary mutation that is associated with CQ resistance may be on the decline. Prior to the introduction of ACTs in 2004, SP and CQ were the two commonly used antimalarial drugs in this area. However six years after the withdrawal of CQ as a first-line treatment, overall prevalence of the *pfcrt* 76T resistant alleles was found to be 23%, which is about 50% reduction from the 46% estimates prior to the introduction of ACTs in this populations[50,55,139]. These findings are supported by other studies, which showed that withdrawal of CQ could lead to the re-emergence of CQ sensitive parasites and reduction in the prevalence of the K76T mutation[104,138,150]. However, the current 23% prevalence of the *pfcrt* 76T basal mutation associated with CQ-resistance is still high and may underscore the continuous elicit use of CQ or suggest selection of this polymorphism by amodiaquine (the partner drug for artesunate) for the treatment of

uncomplicated malaria in this population. I observed longer *pfcrt* haplotypes as compared to the other two loci, with the majority haplotype being a quadruplet resistant allele (74I/75D/76K/271E). These long haplotypes may underscore the overall high mutation rate around this drug-transporter locus and may signal ongoing selection.

Taken together, The evidence of haplotype diversity in all three genes, reflects the known sequential acquisition of these mutations[44,182]. Since the incidence of malaria may not be uniform across the KNDs[100,157], the observed variations in the frequencies of basal mutations, and inferred haplotypes in all three genes may reflect varying drug pressure on parasites across the KNDs. The observed multiple mutations in *pfdhfr* and *pfdhps*, especially the appearance of the K540E resistant in the KNDs, suggest established antifolate resistance. Also the extensive mixtures of both sensitive and resistant allele across most loci in all three genes, highlights the advantage of deep sequencing in identifying minor clones in mixed genome infections.

Above all, this population-level analysis brings into perspective, studies that have demonstrated that human host immunity may play a role in the clearance of resistant *P. falciparum* infections[45]. Therefore, such age-dependent host-parasite interactions, could affect the prevalence of these mutations and their combinations in a given population. Also, hepatic drug metabolism is crucial for the elimination of many drugs including antimalarials. Polymorphism in cytochrome P450 (P450), a drug-metabolizing enzyme is reported to affect the rate at which individuals metabolize a given drug[91]. Thus, parasites in different patients could be exposed to different drug concentrations with different contact times depending

on the individual diversity at the P450 locus[128]. Undoubtedly, being a multi-genome disease, drug resistance in malaria needs to be examined in the context of host-parasite interactions- the main thrust of this thesis.

5.6 Conclusion

In this chapter, I used deep sequencing data to described the prevalence of mutations in *pfdhfr*, *pfdhps*, and *pfcrt*, and identified patterns of inferred haplotypes imposed by the sequential acquisition of resistant alleles in these genes. This provides baseline data for examining changes in drug resistance mutation frequencies in this region. The observations support;

- the re-emergence of chloroquine sensitive parasites (50% reduction) and low prevalence of the *pfcrt* 76T mutation.
- a high prevalence of *pfdhfr* (ie pyrimethamine resistance) and *pfdhps* (ie sulfadoxine resistance) mutations within the parasite population in the KNDs.
- established SP resistance across all geographical locations in the KNDs as evidenced by the presence of parasites with multiple mutations in *pfdhfr* and additional mutations in *pfdhps*.

Chapter 6

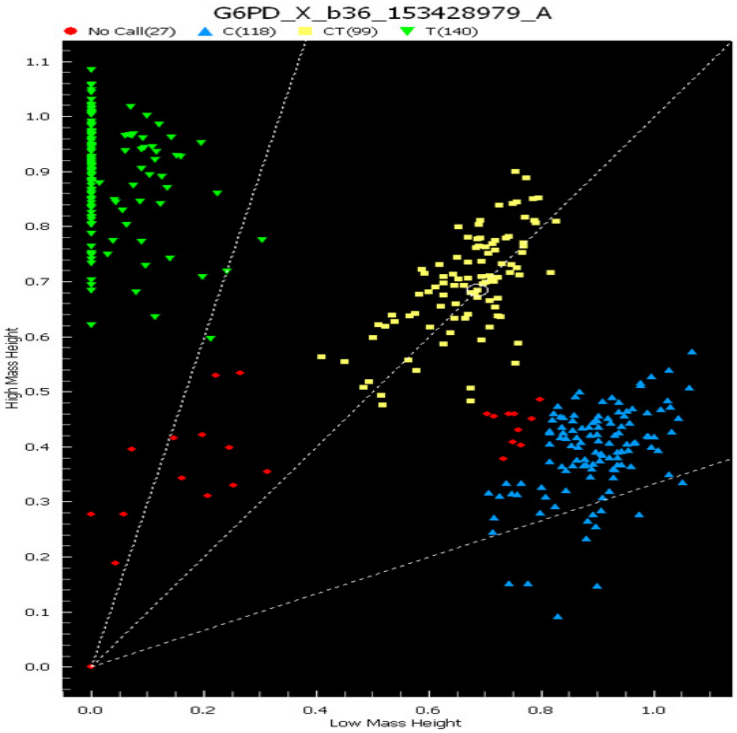
Sequenom iPLEX haploid genotyping

6.1 Introduction

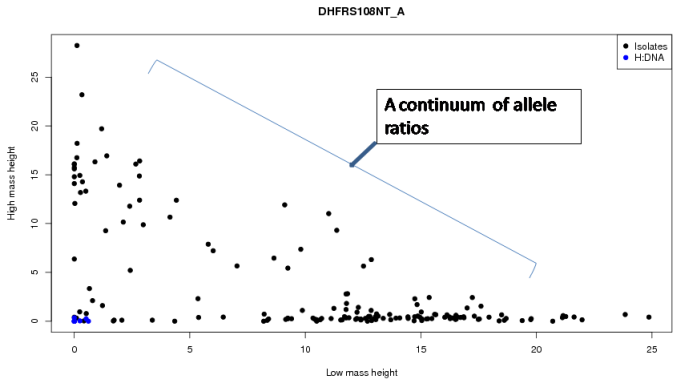
In chapters 4 and 5, I used Illumina whole-genome sequencing data to describe *P. falciparum* genetic diversity across thousands of loci and at specific genomic regions respectively. Thus, demonstrating the value of deep sequencing for variant discovery and for population-level characterization of loci of interest. However, current sequencing cost is high, in the order of $150.00 USD per sample and the quality of parasite DNA required is not always attainable. Therefore, though cost is expected to rapidly decline, these technologies may not be appropriate for follow-up studies of new variants or specific variants of interest in large sample sets with limited parasite DNA. Hence the need to explore alternative genotyping platforms that offer significant sample and SNP throughput for SNP validation at an affordable cost. One such platform is the Sequenom iPLEX platform (up

to 40 reactions per well), which has been extensively used for genotyping human polymorphisms. This platform enables the simultaneous processing of up to 384 samples at a cost of $0.04USD per SNP, allowing for tens of thousands of samples to be processed within a week. The major challenge in using this platform to genotype *P. falciparum* field samples, however, is the occurrence of mixed infection (heterozygosity) that is a continuum of allele ratios at a given SNP (subfigure 6.1(b)). Thus making the existing platform software redundant as it was not designed for this type of allelic distribution, but, an equal representation of alleles as in diploids (subfigure 6.1(a)).

Work undertaken in the Kwiatkowski (DK) laboratory to garner understanding of Sequenom haploid genotyping, which I contributed to, in parallel with my own work, reported a genotype concordance rate between Illumina and Sequenom of 99.9% for homozygous SNPs[121]. The challenge that remained was accurate genotyping of allelic ratios in heterozygous samples (mixed infections), which I sought to address using the samples discussed in Chapter 4 that were observed to contain high levels of heterozygosity. This Chapter therefore, sets out my direct contribution to work in progress aimed at a full understanding of the issues surrounding the use of Sequenom to accurately and reliably genotype clinical samples with mixed population of parasites.



(a) Diploid genotypes resolve into three clusters; homozygous AA or BB or Heterozygous AB



(b) Haploid parasites have a continuum of alleles

Figure 6.1: **Examples of how typical high throughput genotyping softwares resolve genotypes**
Sub-figures 6.1(a) shows the resolution of diploid genotypes into three clusters because the alternate alleles are in equal proportion, and 6.1(b) depicts a continuum of alleles that reflect the composition of haploids in clinical samples, which typically contain billions of genomes in a single infection.

6.2 Objectives

1. To assess the utility of the Sequenom iPLEX high-throughput genotyping platform to measure (or type) *P. falciparum* polymorphisms in field samples,

2. To develop a genotype-calling algorithm for haploid genome mixtures, and

3. To assess the quality of genotype-calls obtained.

6.3 Materials and Methods

6.3.1 SNP Selection

Of particular interest were SNPs in antimalarial drug resistance genes, genes involved in parasite invasion and genes implicated in metabolite (or drug) transport. Published literature sources and online *Plasmodium* genomics resources, notably *PlasmoDB.org release (version 6.0)* were used to assemble SNPs. In general, only bi-allelic SNPs were selected and SNPs that were informative (ie had different alleles for the reference clones 3D7 and Dd2) were preferentially chosen. For testing purposes, 3 polymorphisms were deliberately monomorphic (same alleles) for 3D7 and Dd2. Except for multidrug resistance loci such as the *pfdhfr* and *pfmdr-1* genes, SNPs for assay design were selected on the basis of no proximal potential SNPs within 80 bases of the target SNP, either upstream or downstream. 58 SNPs in total met these selection criteria and were grouped into three priority categories (*High, Medium and Low*) for assay design, details are as

shown in Table 6.1.

Priority	**No. of SNPs**	**Description**
High	39	Antimalarial drug resistance SNPs plus SNPs in candidate vaccine antigens
Medium	11	SNPs in genes expressing surface proteins implicated in parasite invasion
Low	8	SNPs in putative metabolite or drug transporters

Table 6.1: **SNP Categorization by design priority**

6.3.2 Multiplex Design

I used the MassARRAY Assay Design Software (version 3.1) to achieve multiplex design for the iPLEX methodology with the 58 selected SNP. An internally ranked file of a list of the high priority SNPs was first fed to the design software so that a multiplex of the top priority SNPs was obtained. Ranked lists of the medium and low priority SNPs were then sequentially added to the software whilst replexing on the initial multiplex of top priority SNPs. This ensured that the multiplex was designed around the high priority SNPs. A single multiplex containing a total of 35 assays, comprising 19 antimalarial drug resistance SNPs, 11 SNPs in candidate vaccine antigens and 5 SNPs in putative metabolite/drug transporters was obtained.

6.3.3 Sequenom Optimization for haploid genotyping

A series of simple experiments were designed around generating data for understanding the underlisted questions among others, and developing a simple algorithm for resolving *Plasmodium* genotypes using raw mass spectrometry data in the form of peak height intensities.

1. What were the limits of detection of alleles at a given SNP?
2. Can ratios of alleles in samples with mixed infections be determined?
3. How many control clusters are adequate for genotyping?

6.3.3.1 Sample preparation

Control samples used on all genotying plates included water and pooled human DNA (*HapMap CEPH reference panel*) as negative controls. Parasite DNA was diluted in water to a stock concentration of $20\text{ng}\,\mu\text{l}^{-1}$. The DNA from parasite strains were diluted in pooled human DNA at a ratio of 1:31 by mass, approximately equal to 1% parasitaemia (see Appendix C for calculations) in order to mimic the state of the field samples. Parasite mixtures were prepared to be able to confirm the detection of homozygous samples in clones and heterozygous samples in 50:50 mixtures as well as other mixture ratios as shown in Tables 6.3, 6.4, 6.5.

All DNA samples were amplified using Primer-Extension Pre-amplification (PEP)[253]. In brief, 5µl of gDNA from field isolates or parasite mixtures was amplified in 96-

well PCR plate in a final volume of 25µl or 50µl respectively using N15 primers. The PEP material was then stored at -20°C until ready for genotyping.

6.3.3.2 Multiplex Testing

After assay design, the multiplex was tested on a set of six clonal laboratory strains (Table 6.2). Assays showing abnormal behaviour including; poor signals, elevated background noise, positive signals in pooled human DNA and/or water, complete failure with no apparent reasons, were removed from the multiplex before testing. Of the initial 35 assays in the multiplex, 6 assays were removed and 29 assays taken forward. Based on this initial screen, 3D7 and Dd2 were then selected from the six strains for further testing.

Strain	**Origin**
3D7	Parent strain (NF54) isolated from an individual living near Schipol Airport, Amsterdam, The Netherlands
Dd2	Indochina:Derived from W2
HB3	Honduras: Cloned line derived from microscopic selection from the isolate Hondo-1
ITA4	Mozambique:Clone derived from ITO4
T996	Thailand
ITA4_2	Same as ITA4

Table 6.2: List of laboratory strains screened as positive controls

6.3.3.3 MALDI-TOF MS sensitivity

In order to determine the sensitivity of the MALDI-TOF assay for the 29 working SNPs, a 10-fold serial dilution of neat PEP DNA of 3D7 was prepared. The se-

quential concentrations were $(1.0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$ and $10^{-5})$ PEP stock/µl, which was genotyped alongside other samples.

6.3.3.4 Genotyping

All genotyping procedures were carried out as per Sequenom protocol described in detail in Chapter 2 (section 2.2.5). However, PEP DNA was used instead of genomic DNA. Primer sequences are listed in chapter 2 (Table 2.2). All primer (Metabion, Martinsried, Germany) were hydraded to 100 µmol and pools of primers (100 nmol each) were prepared for the multiplex. A master PCR mixture was made for the multiplex prior to first-round PCR containing: 0.68 µl primer mixture, 0.65µl $MgCl_2$ (50mM), 0.2 µl dNTPs (25mM pooled), 1.25 µl Taq buffer (10x), 2.82 µl MilliQ water, 0.4µl HotStar Taq (5U/µl)(Qiagen). 4.5µl of PCR master mixture was plated per well in a 384-well PCR plate format (Thermo Fisher Scientific) and 3µl of 1:10 diluted PEP DNA was added per well. All prepared PCR plates were then sealed with Microseal 'A' lids (Bio-Rad) and cycled on an MJ Tetrad using the following cycling conditions: 94°C for 15 min, 44 cycles of (94°C for 20 sec, 56°C for 30 sec, 72°C for 1 min), and 72°C for 3 min.

Two microlitre of sample from each well of a row (preferrably containing both samples and controls) was run on a 2% agarose gel to confirm the first-round PCR had worked before further processing. Unincorporated dNTPs were removed by adding 2µl of iPLEX shrimp-alkaline phosphatase mixture to 5µl of first-round PCR reaction product and incubated at 37°C for 40min, followed by a denaturation step at 85°C for 5 min and then cooling to 15°C for 15 min.

iPLEX primer extension was done using an extension reaction containing 0.2µl iPLEX termination mixture, 0.041µl extension Taq, 0.2µl extension buffer, 0.5331µl primer mixture and 1.03µl water per well. The final concentration of the extension primers was determined by their individual masses (<3000Da 0.625µM; 3000-5800Da 0.84µM; 5800-7000Da 1.04µM; 7000-10,000Da 1.25µM; >10,000Da 1.5µM).

Extension cycling was undertaken on an MJ Tetrad using the following cycling conditions: 94°C for 30 sec, 40 cycles of (94°C for 5 sec, 5 cycles of (52°C for 5 sec, 80°C for 5 sec)), then 72°C for 3 min abd 15°C for 15 min. Plates were treated with 6mg ion-exchange resin (Sequenom) per well and 16µl MilliQ water. Plates were rotated for 30 min and then centrifuged to pellet the resin before spotting samples onto the SpectroCHIPS and running on the mass-spectrometer.

In order to ensure adequate positive control cluster formation for confirming ratios in mixed samples, 32 replicates of each mixture of 3D7 and Dd2 were genotyped. In the first experiment 9 control clusters (0%, 10%, 25%, 40%, 50%, 60%, 75%, 90%, 100%) **3D7:Dd2** (Table 6.3) were used and these mixtures were done in reversed in order to ascertain any amplification bias that might occur. I next, genotyped 7 control clusters (5%, 10%, 15%, 20%, 30%, 40%, 50%, 100%) **3D7:Dd2** [Table 6.4] alongside field isolates and finally 5 control clusters (0%, 25%, 50%, 75%, 100%) **3D7:Dd2** [Table 6.5].

Strain	**mixtures ratio (%)**								
	Diluted in pooled Human DNA								
3D7	100	90	75	60	50	40	25	10	0
Dd2	0	10	25	40	50	60	75	90	100
	Diluted in water								
3D7	100	90	75	60	50	40	25	10	0
Dd2	0	10	25	40	50	60	75	90	100

Table 6.3: Experiment 1: Nine artificial mixtures of parasite lines prepared using pooled human DNA and water.

Strain	**mixture ratio (%)**						
	Diluted in pooled Human DNA						
3D7	100	75	60	50	40	25	0
Dd2	0	25	40	50	60	75	100
	Diluted in water						
3D7	100	75	60	50	40	25	0
Dd2	0	25	40	50	60	75	100

Table 6.4: Experiment 2: Seven artificial mixtures of parasite lines prepared using pooled human DNA and water.

Strain	**Mixture ratio (%)**				
	Diluted in water				
3D7	100	75	50	25	0
Dd2	0	25	50	75	100

Table 6.5: Experiment 3: Five artificial mixtures of parasite lines prepared in double distilled water.

6.3.3.5 Data Analysis

The Sequenom Typer Software version 4.03 was used for initial assessment of assay quality and overall multiplex performance. The mass spectrometry data in the form of peak heights for the two alleles at each SNP was exported as a Microsoft Excel file (*csv* format) for further analysis in the open source statistical software R[189].

6.4 Results

A total of 550 *P. falciparum* clinical isolates were genotyped in batches alongside artificial mixtures (Tables 6.3,6.4,6.5) of parasite lines with known bi-allelic SNP variations, which were used as positive controls. Despite the inherent problems of *P. falciparum* genome resulting from its high AT content, in this study both wild-type sequences and allelic variants for the SNPs investigated could readily and reproducibly be detected by MALDI-TOF MS, in a single multiplex according to differences in their time of flight, given by the mass to charge ratio (m/z) of the analyte ion. However, two SNPs; $dhfr^{A16V}$ and $Pfcrt^{K76T}$ consistently failed possibly as a result of lack of hybridization of the extension primer due to lack of specificity of the underlining sequence. Overall, the experiments worked very well with the rest of the 27 SNPs returning a good amount of data. Of 'well-behaved' SNPs, 12 SNPs (44%) were outstanding with between 0% to 5% missing data.

In terms of MALDI-TOF sensitivity, the detection thresholds varied depending on the polymorphism. Notwithstanding the possibility of amplification bias between

alleles, at 10^{-3} and 10^{-5} dilutions of PEP stock/µl, both wild-type and mutant alleles for 3D7 were still being detected for 98% of the assays. Therefore, given an initial 1% parasitaemia, the sensitivity of assays was about 0.004% parasitaemia.

6.4.1 Assessment of peak height intensities

Can homozygous samples be genotyped? As illustrated in Figure 6.2, the assignment of genotypes for homozygous samples (single clone infections) is not a problem. The Sequenom software by design is able to accurately work out the linear transformation of these clusters and assign genotype calls as such a distribution of alleles is similar to the allelic distribution of homozygous genotype clusters in diploids. The problem arises when the allele distribution is a continuum and the software has to partition the plotting space for each cluster of allele ratios and if a linear relationship exist, evaluate the angle between clusters to assign a genotype. Therefore, it was important that the artificial mixtures of parasites clones, which I used as positive controls to help partition the space formed distinct none overlaping clusters.

6.4.1.1 Positive control clusters of parasite mixtures

Did the artificial mixtures produce distinct clusters? Figure 6.3 shows four prototypical SNPs for 3D7:Dd2 mixtures that were prepared at nine different ratios to assure a low ratio for the secondary clone (minor allele) to assess cluster boundaries. As illustrated in the figure, three scenarios could be obtained; the top panel shows good cluster separation amongst the different ratios. Figure 6.3(a)

shows nearly ideal clusters that could potentially be used to accurately genotype parasites of mixed genetic types as often found in natural infections. In figure 6.3(b) however, though the control clusters are good, there occurs a significant lift in the low mass cluster along the x-axis. On the other hand, the bottom panel shows different ways in which the outcome can fail to provide adequate clustering to be used for genotyping field isolates. Figure 6.3(c) is a typical case of a poorly designed SNP due to inadequate annotation of the parasite genome, resulting in SNP interference or some other technical reason inherent in the technique that is not apparent. Despite several optimization experiments, some SNPs consistently performed poorly. Of the 27 SNPs I found 12 to consistently provide the level of separation in figure 6.3(a) when using nine control clusters.

Next, I reduced the number of control cluster to five to explore if this will make the control clusters more dense with clear cut boundaries. As shown in figure 6.4, the cluster formation for five control mixture levels for different SNPs was in general more separated as compared to eight or nine mixture levels. All three panels, show adequate cluster formation with clear boundaries sufficient for defining parameters to develop an algorithm for genotyping field isolates. However, as shown in figure 6.4(b) there still occurs a *lift* in the clusters for the high mass ratio. Thus, some amount of rescaling has to be done to correct for the lift (the triangular area marked by the blue line in subfigure 6.4(b)), which underscores the challenge in accurately genotyping polymorphisms in certain regions of the *P. falciparum* genome.

The data presented shows that the artificial mixture ratios of parasite line could potentially be used to genotype mixed infections in clinical samples, but, is there a

linear transformation between clusters that could form the basis for an algorithm for assigning genotypes? In order to ascertain this, a polar transformation of the clusters was undertaken. As illustrated in Figure 6.5, a polar transformation of the data was sufficient to achieve a linear translation of the signal intensities and angle θ between clusters to a valid genotype in many instances. However, because of the compounding problem of the lift, such a simple translation between angle and genotype would be incorrect for certain SNPs clusters. The more reason why a simple algorithm that could rescale the plotting area and provide some level of adjustment for the lift was necessary.

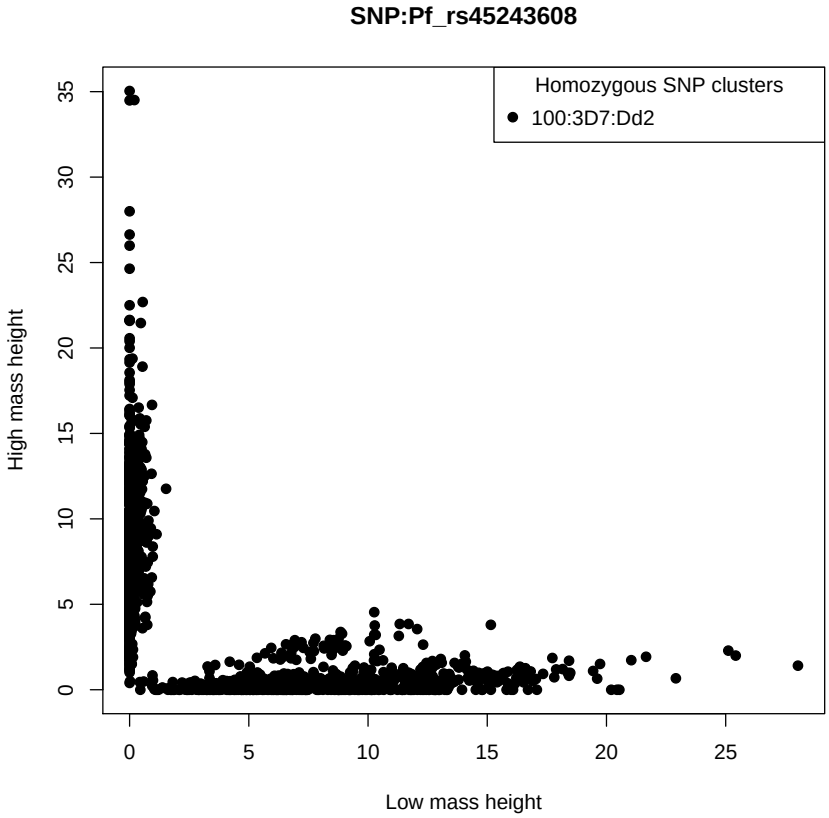
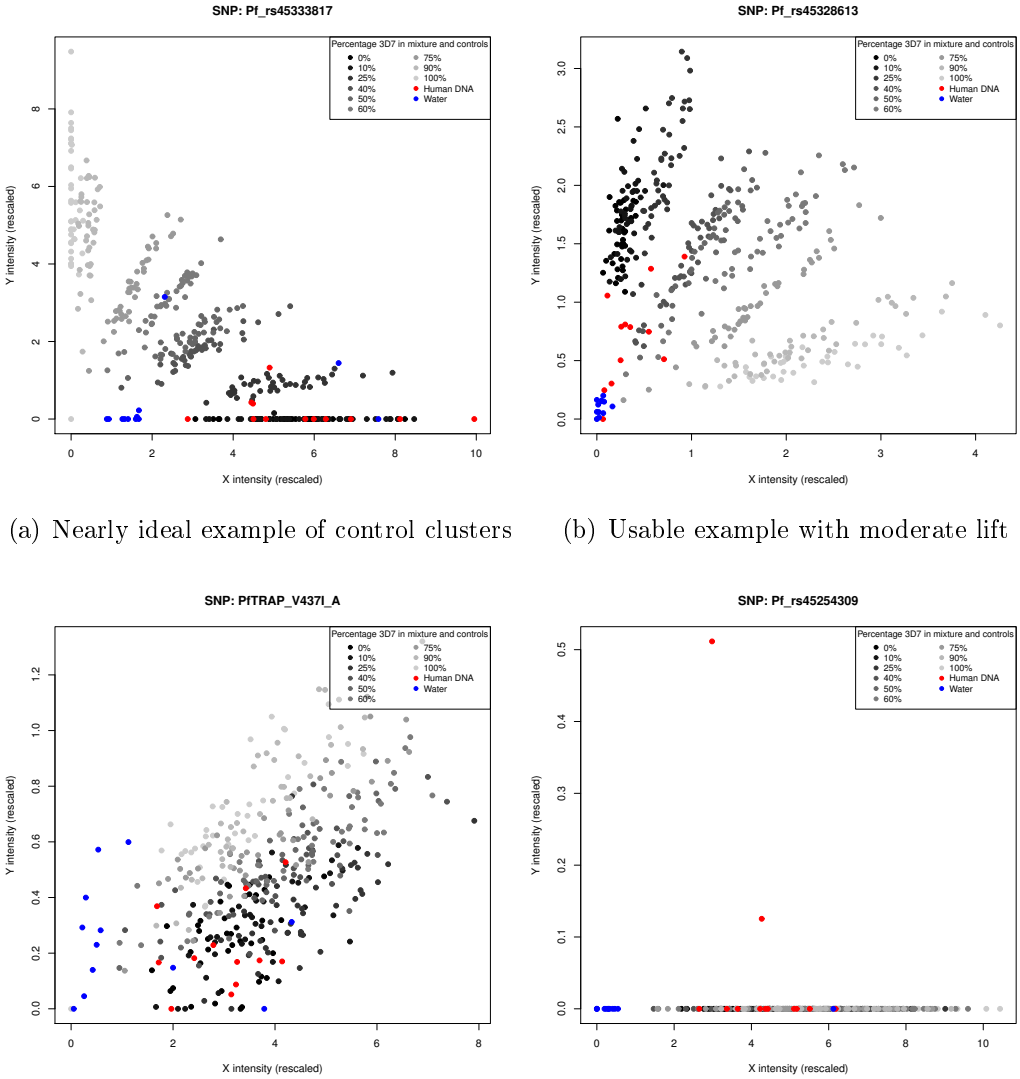


Figure 6.2: Illustration of homozygous clusters for a 100% 3D7. Cluster partitioning similar to diploids



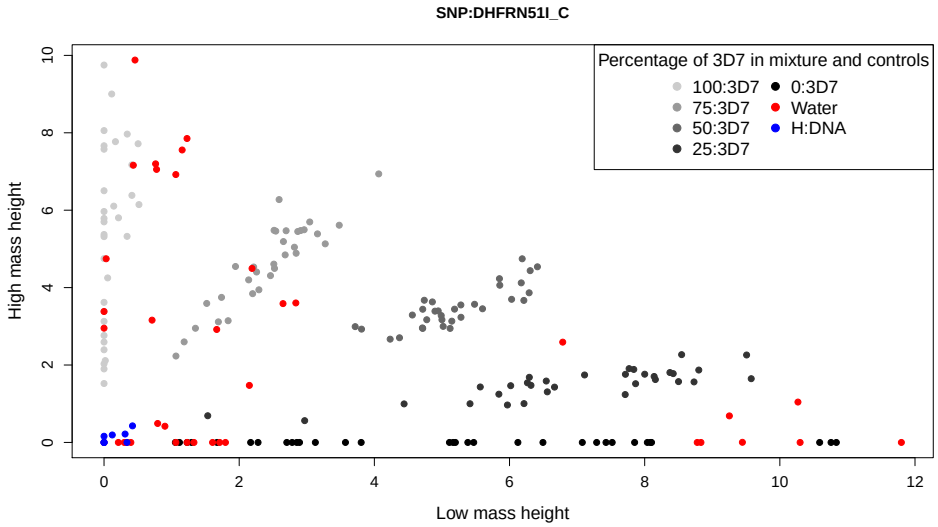
(a) Nearly ideal example of control clusters

(b) Usable example with moderate lift

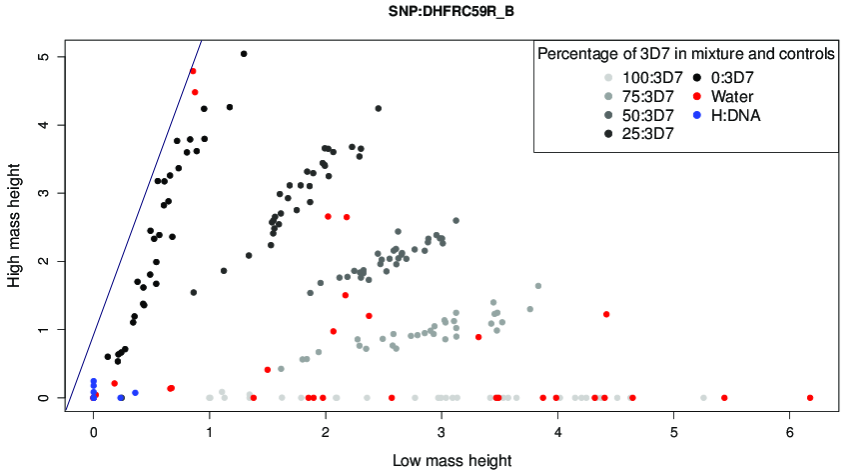
(c) Unusable example due to SNP interference

(d) Failed SNP

Figure 6.3: **Example plots of peak height intensities at two alleles for 9 mixtures of 4 different SNPs. Water and pooled human DNA as negative controls.**
Sub-figure 6.3(a) shows a SNP (rs45333817) in the PFD1155w gene. 6.3(b) shows a SNP (rs45328613) in the PFC0210c gene. 6.3(c) shows a SNP (PfTRAP-V437I) in PF13_0201 gene. 6.3(d) shows a SNP (rs45254309) in PF11_0344 gene.



(a) Ideal clusters with no lift



(b) Adequate clusters with a lift marked out by the blue line

Figure 6.4: **Example plots of peak height intensities at two alleles for 5 mixtures at 2 malaria relevant SNPs.** In the top panel and bottom right, sub-figures 6.4(a), and 6.4(b) show SNPs in the dihydrofolate reductase (PFD0830w) gene.

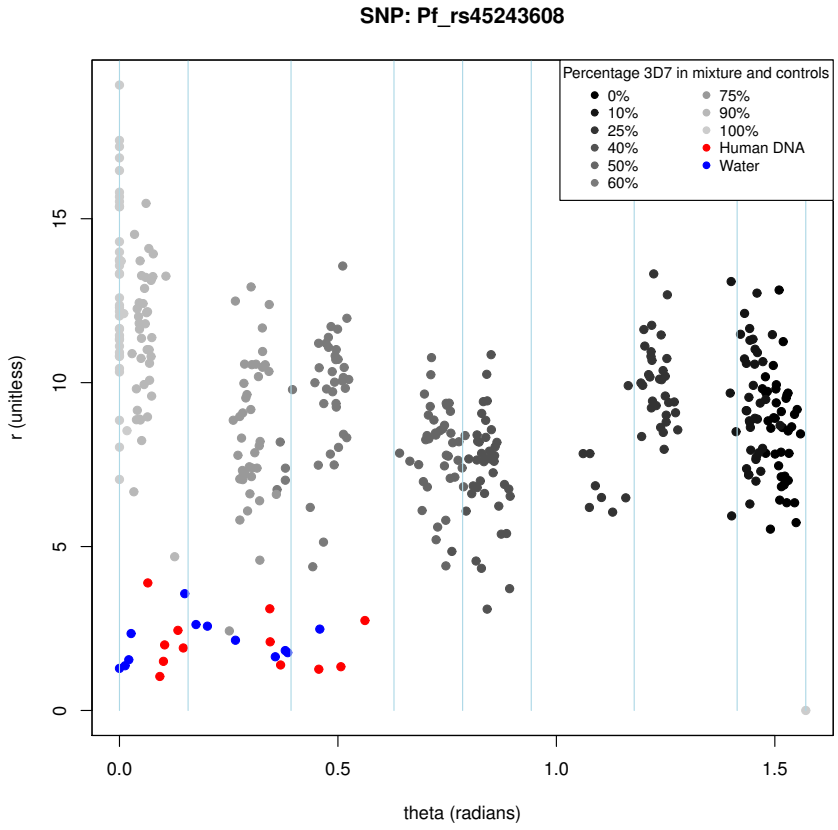


Figure 6.5: **Polar transformation of (X, Y) intensity data for a single SNP.** Light blow lines show where cluster means would occur if linear approximation were correct and there was no lift. Theta values approximate to allele ratio and r values equates to signal intensity. Theta values approximating $<$ 0.3 and $>$ 1.3 for homozygous genotype calls, and a range of theta values from $0.3 \leq \theta \leq 1.3$ for heterozygous genotype calls

6.4.2 Algorithm for resolving haploid genotypes

The data generated by genotyping clonal mixtures was used to explore ways of converting the cluster plots of these intensities (Figures 6.3, 6.4 and 6.5) into accurate genotypes. It was easy to identify the clusters, however, the challenge

was how to effectively deal with within-SNP (or assay) sensitivity in order to make accurate assessments. As shown in the 100% mixture in subfigure 6.4(a) (top panel) of Figure 6.4, the clusters do not always lie flush on an axis, and the pattern was somewhat random. The angle between such a cluster and the nearby axis, I refer to as the *lift*. Due to the presence of the lift for many SNPs, a simple translation between angle and genotype was not reasonable.

We developed a simple algorithm that exploits the observation that the angle between the mean of two adjacent control clusters was approximately linear in their proportions, as evidenced in Figure 6.5, where theta values equate to allele ratio and r values to signal intensity. A diagram of the key quantities used in resolving *Plasmodium* haploid genotypes is as shown in Figure 6.6. Suppose we have two positive control clusters C_o and C_1, with known genotypes, P_o and P_1, and angle separating their means as θ. For a cluster C_i with unknown genotype falling between C_o and C_1 (ie with angle $\rho \leq \theta$), the linear approximation implies that the genotype for C_i is given by

$$P_i \approx \frac{\rho}{\theta}.(P_1 - P_o) + P_o \tag{6.1}$$

This procedure can be generalized to the case where there are n known clusters. Here, we first must find the pair of known genotypes whose angles from the x-axis bookend the unknown one. Of note, therefore, is that this requires that two of the clusters correspond to the homozygous genotypes, that is, two different alleles for two different reference clones at 100% mixture for each clone. Going by the linear approximation, the genotypes value must lie anywhere between these two clusters, and so we have reduced to the case with just two known genotypes.

To assess variation within the genotype estimates, I noted that the within-cluster variance of the known samples strongly correlated with the within-cluster variation of unknown samples (likely due to experimental design). This implied that I could not independently assess my observational error within a sample. In order to avoid this, a bootstrap repeated sampling technique was used to calculate the sample variance[54]

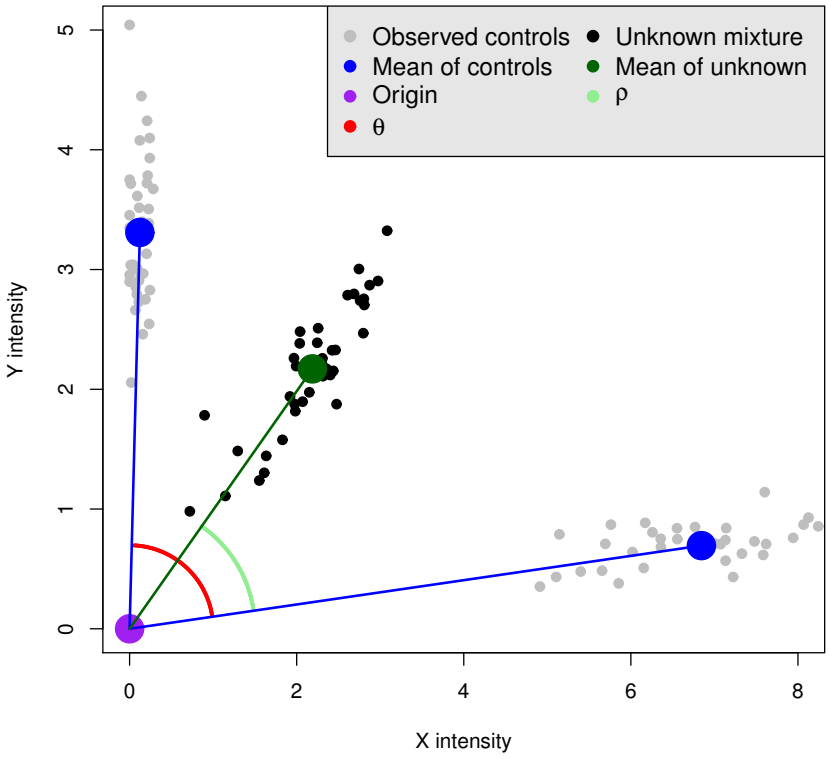
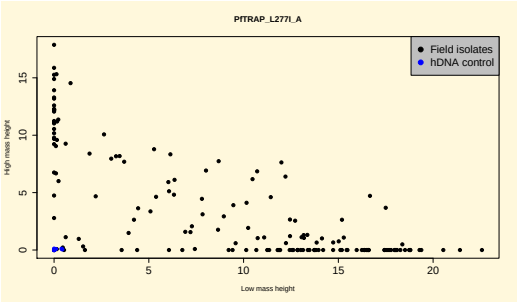


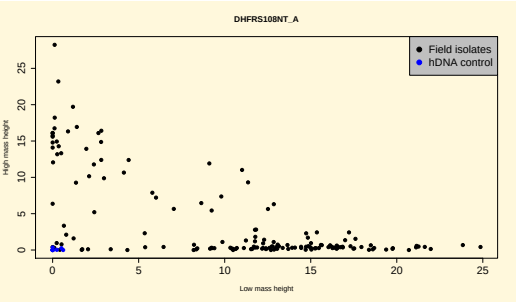
Figure 6.6: Diagram of relationship between the quantities used in the haploid call algorithm

6.4.3 Plots of Peak height intensities for field isolates

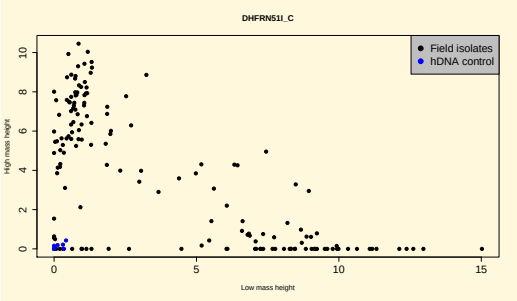
In order to test my genotyping-calling algorithm, field isolates were genotyped in batches alongside the artificial mixtures of parasite clones, pooled human DNA and water as negative controls. As evidenced in Figures 6.7, majority of these infections were of diverse genetic types, no obvious clustering of intensities was observed. Implying that any proportion of alleles was potentially a valid genotype, and this is where the Sequenom Typer v4.03 Software designed for diploid genotyping becomes redundant and fails, as it was not designed for resolving such a distribution of alleles. The Software was, therefore, useful for initial assessment of assay quality and overall multiplex performance, but, could only be of limited use in accurately resolving haploid genotypes as the manual option could be used to assign genotypes to homozygous samples (Figure 6.2). The algorithm discussed in section 6.4.2 was therefore, used to assign genotype calls for all the field samples.



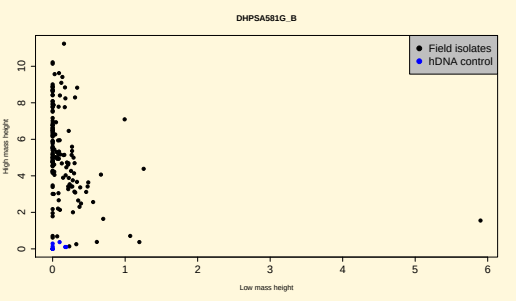
(a) High heterozygosity SNP



(b) Moderate heterozygosity SNP



(c) Much less heterozygosity SNP



(d) A nearly homozygous SNP

Figure 6.7: **Example plots of peak height intensities at two alleles for field isolates showing 4 SNPs.** Sub-figures 6.7(a), and 6.7(b) respectively depict a scatter of peak height intensities that show a moderately mixed infection and a highly mixed infection at SNPs in dihydrofolate reductase (pfdhfr) and thrombospondin-related anonymous protein (TRAP) genes, whilst Sub-figures 6.7(c) and 6.7(d) respectively illustrate a much less heterozygous samples and a nearly homozygous samples at SNPs in pfdhfr and dihydropteroate synthetase (pfdhps) genes.

6.4.4 Assessing the genotyping quality

To assess the quality of genotype calls made by my haploid genotyping algorithm, I inferred the genotype (that is mixture ratio) for four of my known parasite mixture values (or ratios): (10%, 40%, 60% and 90%) where the 3D7 reference genome was the primary component and Dd2 the secondary or minor one. Three

different sets of control clusters were examined;

(1) using two control clusters that bookend the homozygous positions (ie 100% 3D7, 100% Dd2),

(2) using three control clusters that marked the 50:50 ratio (as in diploid) position and the homozygous positions (ie 100% 3D7, 50% 3D7, 100% Dd2), and

(3) using five control clusters that marked the homozygous, heterozygous 50:50 and the three-quarter and one-quarter positions (ie 100% 3D7, 25% Dd2, 50% 3D7, 75% 3D7, 100% Dd2). These clusters were investigated using informative SNPs that had different alleles for the two reference genomes (ie the primary and secondary clones). I measured the unnormalized bias, that is, the inferred value (or proportion) minus the known value for each of the three clusters and used the bootstrap method to determine the spread of these estimates. Figure 6.8 shows the results for two representative SNPs. As evidenced in Figure 6.8 three control clusters exhibited the least bias, although only fractionally better than two control clusters. Five control clusters performed poorly, likely as a result of overlap among control and test (unknown) clusters.

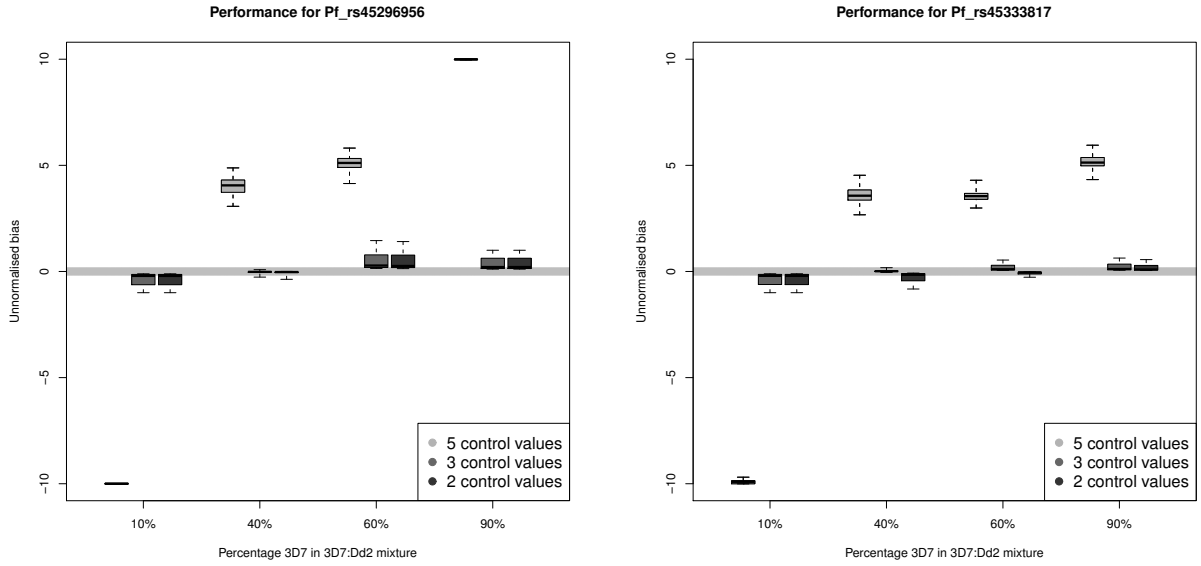


Figure 6.8: The unnormalized bias for two representative SNPs, for three different sets of control clusters

6.5 Discussion

This study was designed to use bi-allelic SNPs to optimize the Sequenom iPLEX platform and establish MALDI-TOF MS based assays for high throughput determination of *Plasmodium* variants from DNA isolated from clinical samples. The study sought to determine the occurrence of the SNPs in *Plasmodium* infections, and to estimate the relative proportion of clones that carry these particular SNPs, as natural infections usually contain billions of parasites of different genetic types.

In general, because the *Plasmodium* genome has over 80% AT content, particularly in non-coding regions[72], assays in these regions of the genome were difficult to design (lack of unique primer sequences due to low complexity), however, the Sequenom iPLEX platform proved feasible for genotyping mixed infections in clinical samples. Both the wild-type and mutant alleles were easily detected for successfully designed assays. Of the 29 SNPs in the multiplex, useful data could be derived for 93% (27/29) of assays. These results attest not only to the feasibility of using the Sequenom iPLEX platform for genotyping field samples, but also show that the genotyping approach used was good. In my protocol, an important modification to the manufacturer's (Sequenom®) instruction, which included use of whole genome PEP amplified DNA instead of genomic DNA was made. I did not observe an apparent indication of amplification bias as a result of PEP. On the contrary, I believe the use of PEP could have increased copy numbers of fragments for minority alleles in these highly mixed field samples for detection.

I observed differences in sensitivity for detecting the wild-type variants or the mutant variants for certain SNPs. Several reasons could account for these differences,

including but not limited to the genomic location of the SNP, initial parasitaemia and instrument background noise associated with certain SNPs. These findings are however, supported by a previous study that performed limited MALDI-TOF MS genotyping of the dihydrofolate reductase (DHFR) variants[125]. I attribute this observation to low parasitaemias of different *P. falciparum* clones occurring in these natural infections. In fact, the use of PEP DNA in our protocol might have mitigated this problem. Notwithstanding these issues, sensitivity of Sequenom was down to 0.004% of parasitaemia, which is within the range of sensitivity for clinical microscopy. Thus, it is reassuring to note that patient samples that reach detectable levels by microscopy in the field are not likely to fail Sequenom genotyping. Overall, there was little interference of human DNA with *P. falciparum* probes. A couple of assays tended to give random positive signals with pooled human DNA controls, a phenomenon that is known to bedeviled most high throughput techniques and merits its own investigation.

The artificial parasite mixes that were genotyped in replicates to serve as positive control clusters (Figures 6.3 and 6.4) proved adequate for making haploid genotype calls. Two approaches to genotyping *Plasmodium* field isolates are potentially useful for malaria control. The first approach is to resolve the genotypes of field isolates at two levels; *homozygous* samples (for allele **a** or **b**) and *heterozygous or mixed* samples (alleles **a**/**b**). The second, is to determine *homozygous* samples and estimate the proportion of clones carrying particular SNPs in heterozygous samples. Each of these approaches have their clinical and epidemiological benefits but, the former may seem more critical to the clinician who is interested in knowing whether a patient has a wild-type, mutant variant or both

alleles for a locus of interest (eg drug resistance gene).

Despite that the control cluster plots for some SNPs did not lie flush on the nearby axis, a polar transformation of the data as evidenced in Figure 6.5 was sufficient to define theta values approximating < 0.3 and > 1.3 as homozygous calls, and a range of theta values approximating $0.3 \leq \theta \leq 1.3$ as heterozygous calls.

The simple haploid genotyping algorithm (Figure 6.6) proved adequate for reliable estimation of proportions of clones in mixed infections that carry particular SNPs. By using the fact that angle θ between the mean of two adjacent control clusters was linear in their proportions, it was possible to accurately assign genotype ratios to heterozygous SNPs (eg 40% A and 60% T). It is worth noting that the data quality did not necessary correlate with increasing number of mixture ratios for controls. The data showed that as the number of control clusters increased, the quality of genotype calls for sample clusters with under 20% allele ratios tended to merge with the homozygous clusters at the flanks. There was an apparent grey area, where minority clone ratios may inevitably be detected as homozygotes. Thus, there was no added quality benefit in increasing the number of control clusters beyond five, rather it increased the chance for overlap of clusters and reduced fidelity of heterozygous genotype calls (Figure 6.8). Among the artificial ratios tested, the two homozygous clusters (ie 100% 3D7 and 100% Dd2) or the three clusters where both the homozygous positions and the 50% ratio (ie 100% 3D7, 50% 3D7 and 100% Dd2) were partitioned proved sufficient for minimizing within-cluster variability and translating the angle between the means of adjacent clusters into reliable genotype proportions. However, three control clusters were

found to be optimal as this adequately bookend the homozygous positions and marked the position of the heterozygote for assigning genotype calls.

This study was limited by the lack of sufficient internal controls, resulting in the possibility that a clone of low parasitaemia may not have been detected, and the fact that currently only known bi-allelic variants could be detected. However, the evidence discussed above strongly indicate that the Sequenom iPLEX platform is a sensitive method for typing polymorphisms in *P. falciparum* , and as part of a broader ongoing effort to fully understand the issues that are key to accurate and reliable genotyping of *P. falciparum* field isolates some progress has been made. As work in progress in the DK laboratory, there remains technical issues to resolve concerning the random occurrence of the lift, and how to effectively implement the algorithm on a larger scale. Obviously more data needs to be generated to shed more light on the problem, but, one solution may be to further develop the algorithm to enable user determined rescaling of the plotting space to account for the lift when necessary.

6.6 Summary

- This Chapter lays out my contribution to ongoing work aimed at validating the Sequenom iPLEX platform for genotyping *P. falciparum* clinical samples.
- I designed a multiplex of 29 SNPs for interrogating the platform to generate mass spectrometry peak height intensity data for analysis. Of these SNPs, 93% (27/29) worked and produced quality data.

- I employed a strategy of genotyping artificial mixes of parasite lines in replicates and plotting these to obtain control clusters that were mapped to test samples for genotyping.

- The angle between the mean of two adjacent control clusters was approximately linear to their mixture ratio and provided the basis for developing an algorithm for assigning genotypes.

- A lift in the control clusters for homozygotes occurred in certain SNPs and added a layer of complication to the linear transformation of clusters for assigning genotypes.

- Three control ratios that marked the 50:50 mixture ratio and also book-end the homozygous positions (ie 100% 3D7, 50% 3D7, 100% Dd2) were found optimal for accurate assignment of genotype ratios for heterozygous samples.

- Despite the major drawback of demand for internal controls to accurately identify all alleles in a given field sample, and inadequate understanding of the basis for a lift in certain SNPs but not others, this experimental approach proved sufficient for assigning mixture ratios for heterozygous samples.

6.7 Concordance between Sequenom and Illumina genotyping

6.7.1 Motivation

Thus far, a subset of the 550 *P. falciparum* clinical isolates that were sampled in Chapter 3 have been genotyped by two high-throughput genotyping methods; Illumina short-read alignment (RA) and Sequenom iPLEX genotyping. As stated earlier previous concordance rate between these two methods was 99.9% for homozygous genotypes (manually called)[121]. It was therefore, of interest to assess the concordance rate between the two approaches for the genotype calls made with my algorithm for a set of SNPs and samples that were genotyped by both methods.

6.7.2 Scope of samples

Concordance rates of genotype calls based on the Sequenom assays and genotype calls based on Illumina sequencing was calculated for 9 SNPs in 147 samples. It should be noted, however, that these 9 SNPs were non-typable and were removed from our *typable* SNP list because of the stringent quality filtering procedures that were implemented.

6.7.3 Results

Figure 6.9 shows the distribution of genotype concordance between the two approaches. Also presented in Figure 6.10 are genotype concordance distributions for all sites from MS and RA (left), as well as that for just sites that were identified as heterozygous by RA (right). In both cases, the results indicate an unbiased average agreement between the methods. Further, as shown on Table (6.6), where Sequenom gave a valid call, heterozygous call based on >5 reads, the concordance rate was 76.8%.

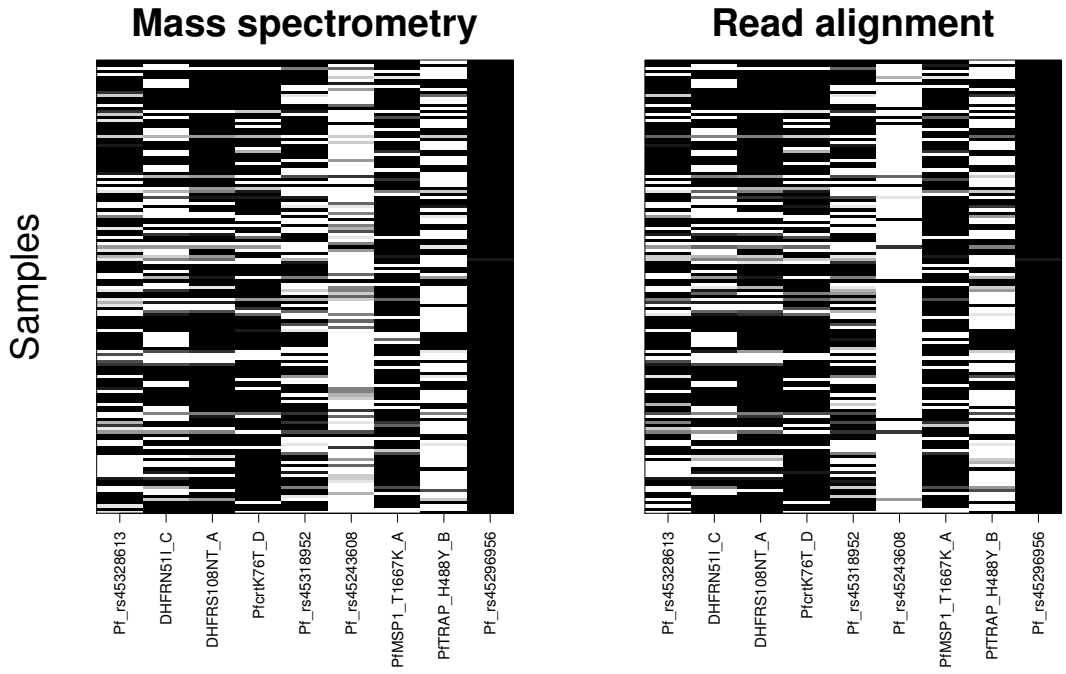


Figure 6.9: Genotype calls by mass spectrometry (left) and read-alignment (right) for 9 SNPs and 147 samples. White indicates reference, black indicates non-reference and shades of gray indicate heterozygous (mixed) samples

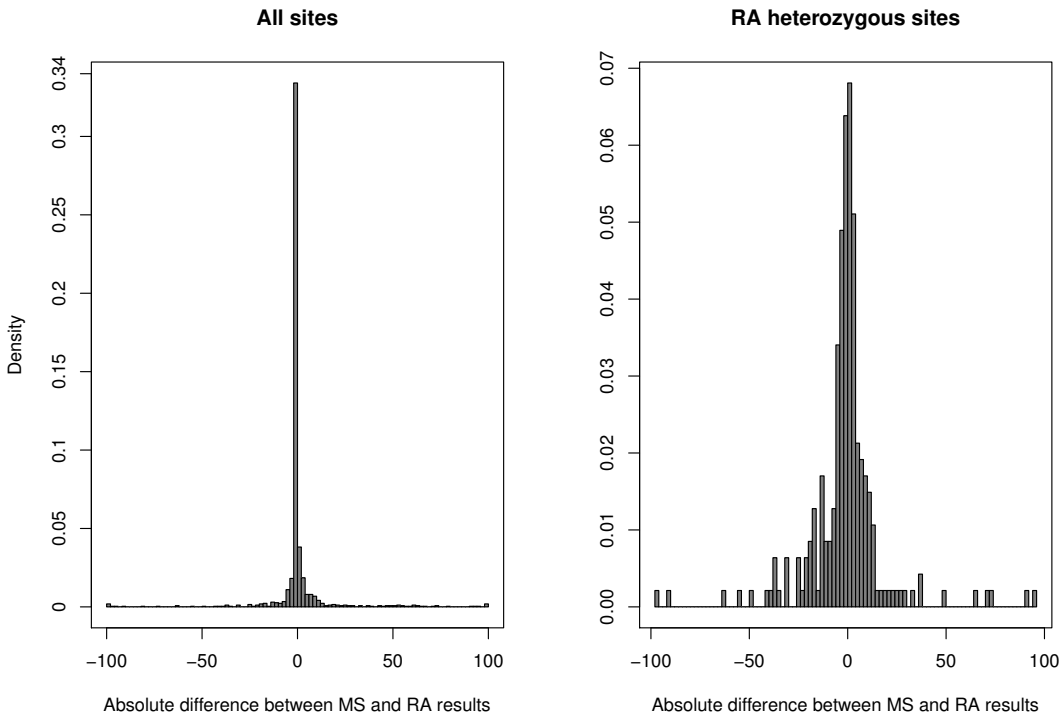


Figure 6.10: Difference between genotype calls between read-alignment (RA) and mass spectrometry (MS), for all sites (left) and for sites called as heterozygous by RA (right)

	Concordance based on 9 SNPs	
	Rate (%)	**Number of genotypes**
Hoomozygotes		
Reference allele	80.5	330
Reference read count	33051	
Non-reference allele	96.8	694
Non-reference read count	51438	
Both alleles	90.9	1024
Heterozygotes		
Reference allele	74.7	59
Reference read count	13060	
Non-reference allele	78.9	60
Non-reference read count	8673	
Both alleles	76.8	119
Overall	89.1	1143

Table 6.6: Concordance between Sequenom and Illumina sequencing

6.7.4 Conclusion

- The genotype concordance rate for homozygous calls was 90.9% overall and 96.8% when Illumina sequencing showed the non-reference allele.
- Where Sequenom gave a valid call, heterozygous call based on >5 reads, the concordance rate was about 90%.
- Concordance for all sites (homozygotes and heterozygotes) was 89.1%.
- The Sequenom iPLEX platform is therefore, sensitive for genotying bi-allelic SNPs in *Plasmodium species*, however, accurate and reliable genotyping of heterozygous sites in clinical samples with mixed infections remain a major challenge.

- This notwithstanding, the ability to provide some sense of allele ratios in samples from mixed infections hold important benefits for public health interventions in areas of high malaria endemicity such as the KNDs of Northern Ghana.

Chapter 7

Host-parasite genetic interactions

7.1 Introduction

In malaria, it has long been recognized that both host factors[197,205], and parasite factors can influence disease outcome[115]. Indeed, several studies on human malaria have shown classic evidence of host factors associated with resistance (or susceptibility), including the sickle-cell trait[245] and specific Major histocompatibility complex (MHC) alleles[203]. Likewise, parasite encoded phenotypes such as cytoadherence and rosette formation are recognized to contribute to the severity of malaria disease[53,197,205]. It is not known if polymorphisms in specific parasite genes could enhance their survival against host protective mechanisms such as oxidative stress resulting from sickle-cell or Glucose-6-Phosphate dehydrogenase deficiency gene polymorphisms. Antimalarial drug resistant parasites are known to have a fitness advantage under drug pressure, but, may be less fit and vulnerable to host immune attacks in the absence of drug pressure[238]. Indeed,

age-dependent clearance of antimalarial drug resistant parasites was observed in Mali[45], but, these host-parasite interactions are poorly understood.

To begin to explore such host-parasite interactions, in this Chapter, I utilize a well-characterized malariaGEN case-control sample resource previously collected in the KNDs. A case-control analysis of a subset of this data (including the severe malaria cases and half of the healthy controls) was previously undertaken by Anita Ghansah[73,74]. Before exploring host-parasite genetic interactions in this population, I seek to describe the full dataset that includes severe malaria cases, mild malaria controls, and healthy asymptomatic controls.

7.2 Specific objectives

1. To investigate human polymorphisms associated with malaria resistance (or susceptibility) in this population to identify host factors for testing interactions with parasite factors.

2. To assess the association of drug resistance SNPs in *pfdhfr* and *pfdhps* with malaria disease.

3. To assess the interaction between *pfdhfr* resistant alleles and specific human polymorphisms associated with malaria-protection in the KNDs population.

7.3 Material and Methods

The samples discussed in this study are from a frequency-matched case-control study conducted in the KNDs from August 2002 to January 2004. These are a set of about 4000 well-characterized MalariaGEN archived samples comprising human DNA for 700 severe malaria cases, 1642 mild malaria cases and 1674 healthy controls.

7.3.1 Participant Recruitment

Cases consisted of children 6-60 months of age who presented to the NWMH with severe malaria within the study period. Criteria for diagnoses of cases were per standard WHO definitions of severe malaria[240]. Two categories of control subjects were recruited, and two sets of children from each category were frequency-matched to cases based on age category (6-24 months and 25-60 months), sex and location.

Mild (uncomplicated) malaria controls were identified at four satellite outpatient clinics in the KNDs using axillary temperature >37.5°C and the OptiMAL Rapid Malaria Test (DiaMed, FLOW Inc. Portland, Oregon). Microscopy was then used to assess parasite densities and patients with parasitaemias > 2500 parasites µl of blood (and no other obvious cause of fever) were recruited. Further details of subject recruitment for this study may be reviewed in Osafo-Addo *et al.,* 2008[165] and in[157].

7.3.2 Ethical considerations

This study was reviewed by the scientific and ethical review board of the Noguchi Memorial Institute for Medical Research, the Navrongo Health Research Center IRB, the Ghanaian Ministry of Health Ethical review committee, and the review board of U.S. Naval Medical Research Unit 3. Written informed consent was obtained from adults and the parent or legal guardians of all children that were enrolled in this study. All consent forms were translated into the two major languages in the KNDs and administered alongside the English versions.

7.3.3 DNA extraction

Whole blood was collected on Whatman no 1 filter paper (Whatman®, England), air dried and stored at room temperature in air-tight bags containing silica gel. Genomic DNA was extracted from the filter paper blood blots as described in[96]. The dried blood blots were cut out with a sterile scissors and suspended in 150 µl of 10% Saponin (Sigma, St. Louis, MO) in 1X PBS (pH 7.2). Samples were incubated for 2 hours or overnight, vortexed and centrifuged. The pellets were washed in PBS and the supernatant discarded. 150 µl of sterile double distilled water and 50 µl of 20% (wt/vol) solution of Chelex® 100 (Sigma, St. Louis, MO) were added. The samples were boiled at 95°C for 10 min, vortexing every 2 minutes. After centrifugation, the supernatant containing the DNA was collected and stored at -80°C until ready for use.

7.3.4 SNP selection

The human SNPs were selected from a database of SNPs assembled by the Kwiatkowski laboratory for undertaking quality control genotyping of samples contributed by partner sites to the MalariaGEN consortium. SNPs were chosen that had a minor allele frequency (MAF) of $> 1\%$ in African populations, including polymorphisms previously reported to be associated with severe malaria disease. A total of 65 human SNPs were genotyped. To this, I added 5 SNPs in the *pfdhfr*, and 3 SNPs in *pfdhps*, all associated with antifolate drug resistance, and which were found to be highly prevalent in the KNDs (Chapter 5). These were selected from the multiplex of 29 *P. falciparum* SNPs that was used to optimize the Sequenom iPLEX platform for haploid genotyping (Chapter 6).

7.3.5 SNP genotyping

Genotyping of the selected SNPs was undertaken using the Sequenom MassARRAY iPLEX platform according a protocol established in the Kwiatkowski laboratory. Details of Sequenom genotyping was described in Chapter 2 (section 2.2.5). In brief, PEP DNA was used for amplification of each polymorphic region in a first-round PCR, subsequent primer extension and detection was by MALDI-TOF Mass Spectrometry. Genotypes for human SNPs were assigned using the Sequenom Typer software while parasite genotypes were assigned using my '*P. falciparum* genotype calling algorithm' (Chapter 6).

7.3.6 Statistical Analysis

An initial quality check (QC) was performed on both the genotype and clinical data. The extent of missing genotypes and samples (clinical data) was first assessed by plotting the data. Based on this QC, a threshold of $> 10\%$ missing data was applied for all SNPs and samples. Association of case and control genotypes with clinical phenotypes was carried out using either Chi-square (χ^2) or Fisher exact test. In addition, multiple logistic regression analysis was used to look for the direction of significant effects using four conventional models (Additive, homozygous dominant, homozygous recessive and heterozygous models). Disease association was assessed at two broad phenotypic levels - severe malaria and mild malaria, and genotype data were coded with respect to the derived allele in all model comparisons. Here, I tested three levels of disease status; infection, clinical disease, and severe disease as inter-dependent phenotypic manifestations. This was important in order to derive potential candidates for virulence resistance, as these will be useful for the subsequent host-parasite genetic interaction analysis. The 52 SNPs used for this analysis are clustered in 27 candidate genes. Therefore, given the five genetic models and eight phenotypes tested, there was 40 independent comparisons for each SNP, about one thousand comparisons in total for all SNPs. However, some SNPs are not independent of one another because of linkage disequilibrium, which varies between ethnic groups. Also, the genetic models and phenotypes tested are not truely independent. Moreover, given that these are candidate genes with *a priori* high chance of a positive result, a full Bonferroni correction (10^{-5}) may be conservative. Therefore, to minimize the chance of false positive results due to multiple testing and lower the risk of gen-

erating false negatives, I used a threshold for statistical significance at $p \leq 10^{-3}$ (i.e $\alpha/50$) [see Appendix D for details].

I next investigated the main effects of parasite genotype (three levels: wild-type, resistant, mixed) using either Chi-square (χ^2) or Fisher exact test on severe malaria and mild malaria, with host phenotypes as proxy virulence phenotypes.

Finally, I used a generalized linear model fitted to the host genetic, parasite genetic and host phenotype data to look for significant interaction effects.

7.4 Results

7.4.1 Host disease association

Of the 65 host SNPs genotyped, 19% (12/65) were removed from this analysis because of a high rate of missingness (>10%). One additional SNP, rs33950507 (Haemoglobin E) in the HBB gene was found to be monomorphic in this population and was removed from further analyses. No severe violations of the Hardy-Weinberg Equilibrium was observed among controls in the remaining 52 SNPs. Overall, 670 cases of severe malaria (SM), 1400 cases of mild malaria (MM) and 1374 cases of population controls passed the QC threshold of $< 10\%$ missing data. The minor allele frequencies (MAF) for SM cases and population controls, and SM cases versus mild malaria (MM) controls are shown in subfigures 7.1(a) and 7.1(b) respectively.

Previous case-control analysis of a subset of this data reported the known sickle-

cell trait (HbAS) protection from fatal malaria in this population[73]. This finding was replicated in this analysis (see below). Indeed, it was the strongest signal of protection from both severe malaria and mild malaria. Other less evaluated polymorphisms were also found to be associated with severe malaria, mild malaria and malaria infection in this population.

7.4.1.1 Severe malaria association

The protective effect of the sickle-cell trait (marker rs334) was observed in all levels of comparison [Table 7.1 and Figure 7.2]. The strength of the association signal varied depending on the infection status of population controls. SM cases versus all controls (OR 0.42, 95% CI 0.27-0.67,P=0.0003). But, 40% of these population controls were asymptomatic with a positive blood smear by microscopy. SM cases versus uninfected controls (OR 0.33, 95% CI 0.20-0.55,P=0.00002). SM versus asymptomatic controls (OR 0.49, 95% CI 0.29-0.81,P=0.005).

I also found statistical evidence of protection for rs2243250 [an IL4-589 promoter polymorphism] (OR 0.55, 95% CI 0.37-0.83, p = 0.004) and IL-22 (rs2227491) polymorphisms (OR 0.59, 95% CI 0.41-0.85, p = 0.004) against SM (Table 7.1). Whereas, an IL-17 receptor C polymorphism was associated with a risk of SM in a homozygous recessive model (OR 1.44, 95% CI 1.14-1.82, p = 0.002).

7.4.1.2 Mild malaria association

I found a non-significant trend for sickle-cell trait protection when mild malaria cases were compared with asymptomatic controls (OR 0.77, 95% CI 0.54-1.09)[Table

7.2]. However, when I compared mild malaria cases with the uninfected controls, there was a significant protection (OR 0.48,95% CI 0.33-0.77, p = 0.0002)[Table 7.2]. In addition, the rs3092945 (CD40LG) was associated with increased risk of mild malaria (OR:1.50, 95% CI: 1.11-2.04, P = 0.008).

Also, in this analysis, I found another IL-22 variant (rs1012356) to be significantly associated with protection from clinical malaria in a homozygous recessive model (OR 0.70, 95% CI 0.58-0.85, p = 0.0003) Table 7.2]. In addition, a polymorphism in IL-17 receptor C was associated with an increased risk of clinical malaria (OR 1.44, 95% CI 1.14-1.82, p = 0.0022)[Table 7.2].

7.4.1.3 Comparison of severe and mild malaria cases

To assess genetic factors that may play an important role in disease progression, genotype comparisons were carried out between severe and mild malaria cases. A polymorphism in IL-10 gene (rs3024500), an anti-inflammatory cytokine gene, was associated with risk of severe disease (OR 1.38, 95% CI 1.14-1.69, p = 0.0012) (Table 7.3).

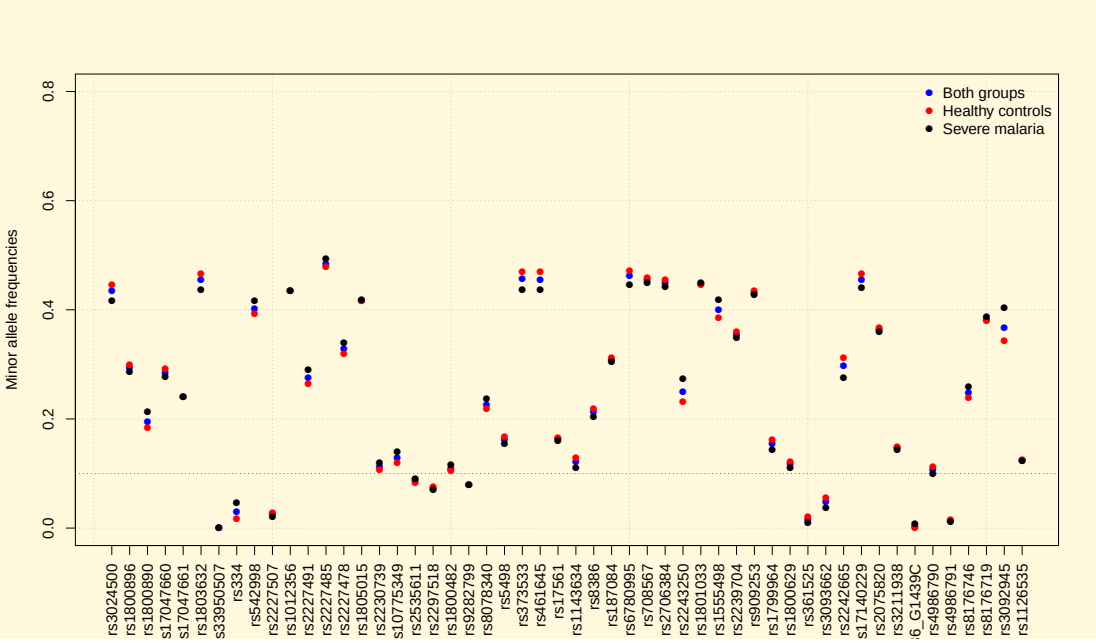
7.4.1.4 Association with malaria infection

When I compared uninfected controls with asymptomatic controls, IL-17 receptor C was found to be associated with an increased risk of malaria infection (OR 1.41,95% CI 1.11-1.78, p = 0.004)[Table 7.4]. In this analysis, I observed a marginal significant association of the sickle-cell trait with protection against malaria infection (OR: 0.65, 95% CI: 0.43-0.90, p = 0.03).

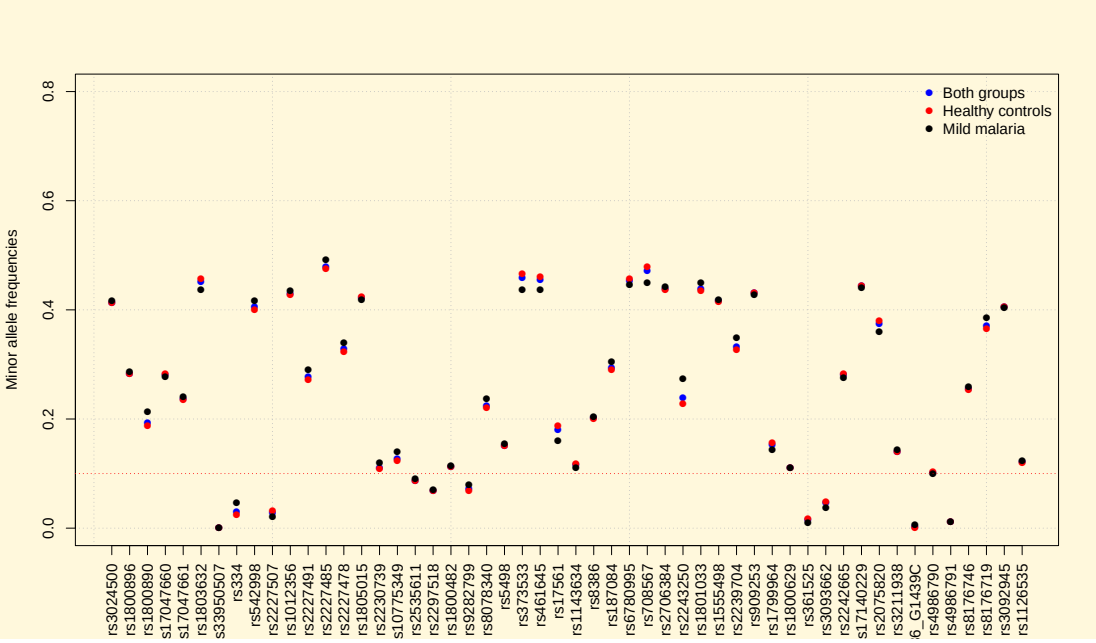
In most of the disease associations undertaken above, the results show that the sickle-cell trait is generally protective against malaria. There was evidence of protection against infection (uninfected vs asymptomatic), however, the protection is more pronounced as infected individuals become febrile and progress to severe malaria (Table 7.5). The signal for the sickle-cell trait protection was affected by the infection status of controls, with the best signal of protection (3-fold) obtained when SM cases were compared with uninfected population controls. A lower signal of sickle cell protection (2-fold) was observed when SM cases were compared with asymptomatic controls and an intermediate signal of protection (2.4-fold) when SM cases were compared with a mixture of both asymptomatic and uninfected population controls (Table 7.5). This observed dampening of the sickle-cell protective signal by infection status of controls highlight the importance of control ascertainment for case-control analysis, especially for malaria disease association-the need for control subjects to not be infected. Also, the general protection of the sickle-cell trait against malaria disease underscores the need to adjust positive signals obtained in disease association studies of severe malaria for the sickle-cell effect.

7.4.1.5 Remaining SNPs

No other significant associations were found for the rest of the SNPs. The data for these comparisons are shown are shown in Figures 7.2 and 7.3. Also, details of all 52 SNPs including MAF is shown in appendix E.



(a) Minor allele frequency for SM cases and healthy controls



(b) Minor allele frequency for SM and MM controls

Figure 7.1: Minor allele frequencies for SM cases and controls. In subfigure 7.1(a) shows the frequencies of severe malaria cases and population controls and that the combined group whilst subfigure 7.1(b) shows that for mild malaria and the same controls. As indicated by the red dotted line majority SNPs were about 10% frequency.

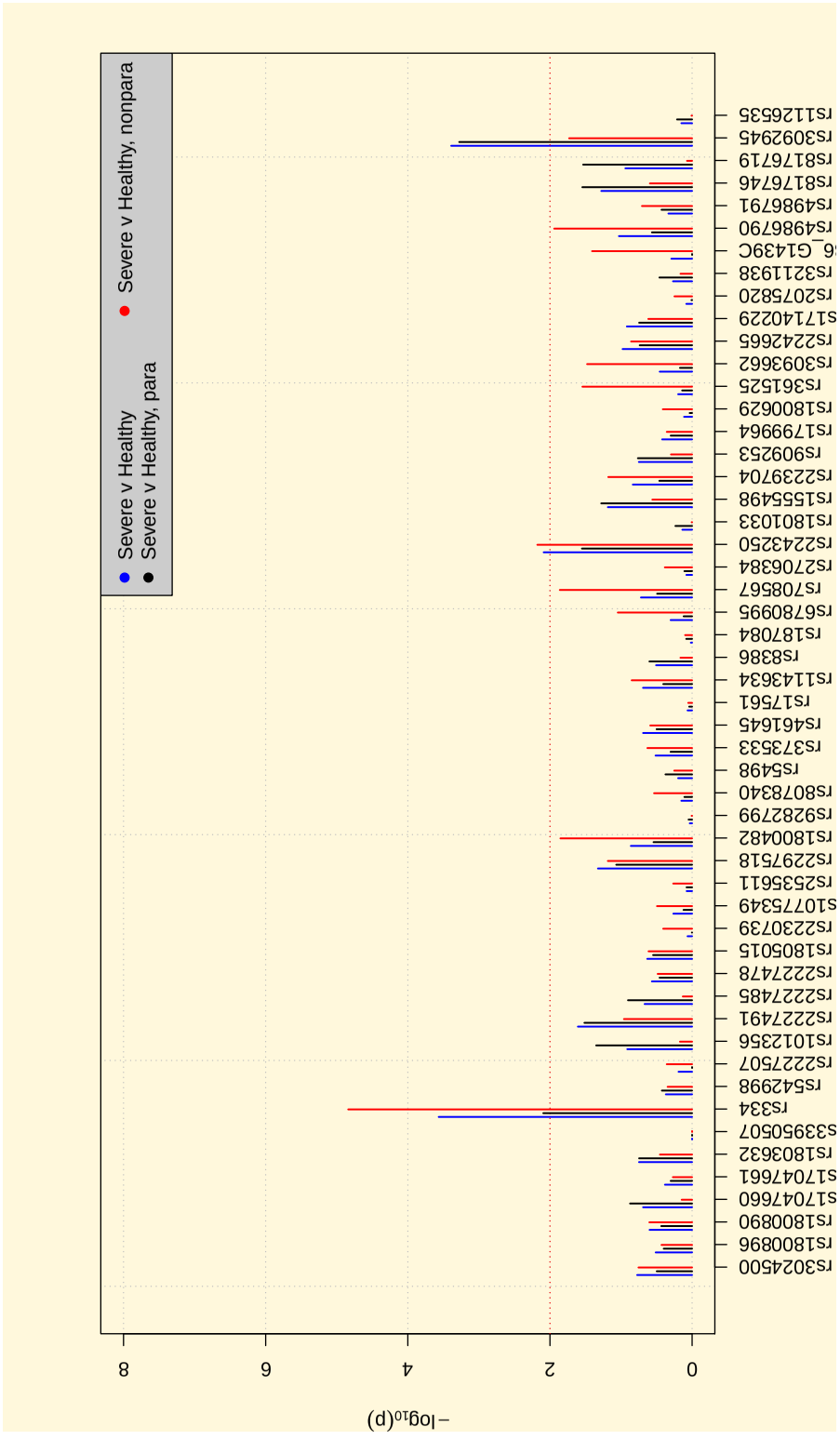


Figure 7.2: **SNP association for SM cases versus population controls.** Fisher exact P-values for the best genetic model in a regression analysis is shown. The red dotted line shows the cut-off for statistical significance

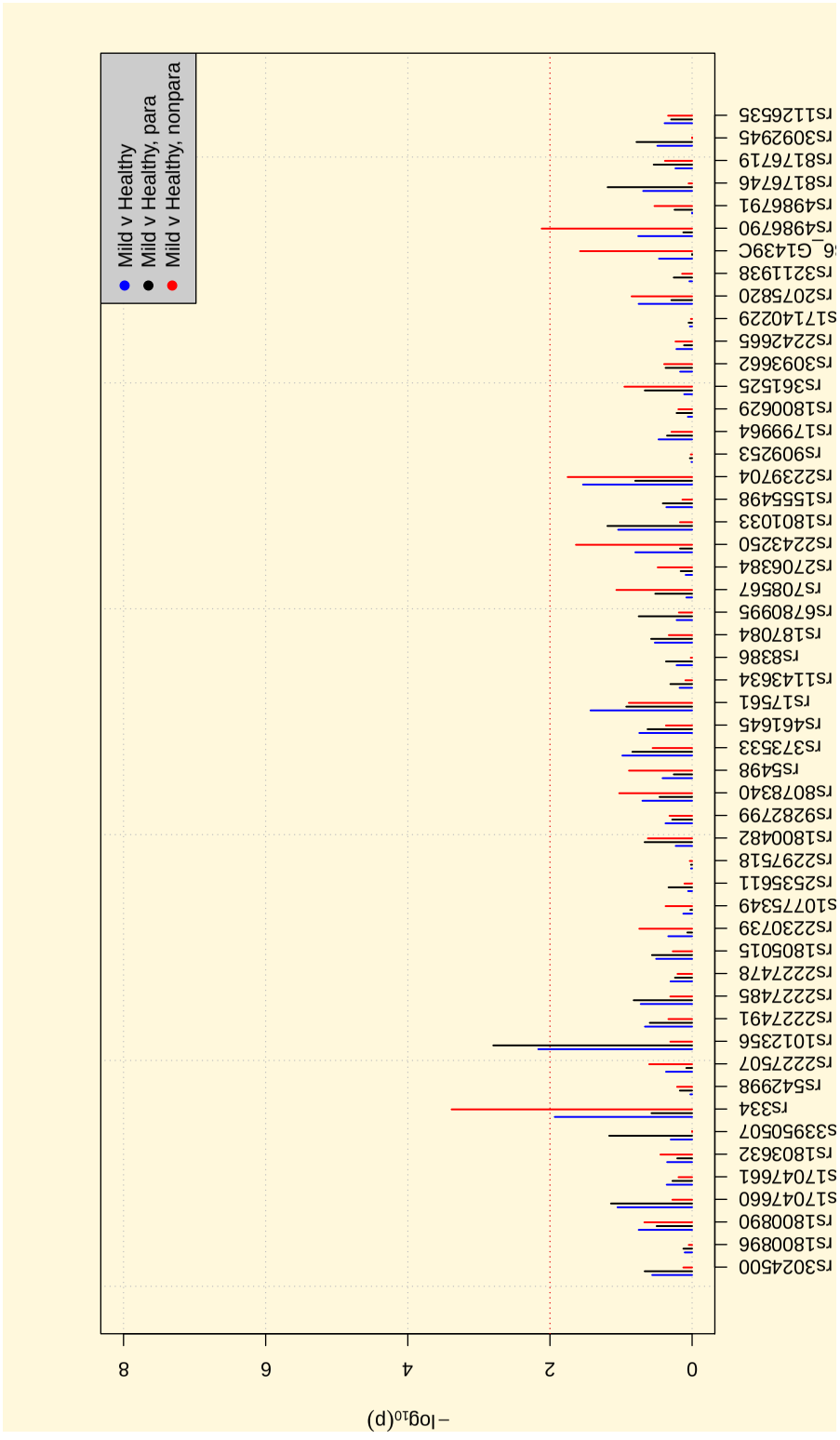


Figure 7.3: **SNP association for mild malaria cases versus population controls.** The best genetic model in a logistic regression analysis is shown. The red dotted line shows the cut-off for statistical significance

Cases vs healthy controls with/without parasitaemia							
		MAF					
SNP	**Gene**	**Cases**	**Controls**	**Model**	**OR**	**95% CI**	**P-value**
		(n=670)	(n=1374)				
rs3092945	CD40LG	0.34	0.40	**T**C/**TT** vs CC	0.86	0.77-0.96	0.0066
rs334	HBB	0.02	0.04	**A**T vs TT/**AA**	0.42	0.27-0.67	0.0003
rs2243250	IL4	0.23	0.25	TT/**C**T vs **CC**	0.55	0.37-0.83	0.0040
rs2227491	IL22	0.27	0.29	CC/C**T** vs **TT**	0.59	0.41-0.85	0.0043
Cases vs healthy controls without parasitaemia							
		MAF					
SNP	**Gene**	**Cases**	**Controls**	**Model**	**OR**	**95% CI**	**P-value**
		(n=670)	(n=524)				
rs708567	IL17RC	0.46	0.45	**GG**/**G**A vs AA	1.44	1.14-1.82	0.0022
rs334	HBB	0.02	0.02	**A**T vs TT/**AA**	0.33	0.20-0.55	0.00002
rs2243250	IL4	0.23	0.25	TT/**C**T vs**CC**	0.46	0.29-0.74	0.0014
rs2227491	IL22	0.27	0.28	CC/C**T** vs**TT**	0.59	0.41-0.85	0.0043
Cases vs healthy controls with parasitaemia							
		MAF					
SNP	**Gene**	**Cases**	**Controls**	**Model**	**OR**	**95% CI**	**P-value**
		(n=670)	(n=850)				
rs708567	IL17RC	0.46	0.47	**GA**/**GG** vs TT	1.44	1.14-1.82	0.0022
rs334	HBB	0.02	0.03	**A**T vs TT/**AA**	0.49	0.29-0.81	0.0054
rs2243250	IL4	0.23	0.24	TT/**C**T vs**CC**	0.61	0.40-0.94	0.0240
rs2227491	IL22	0.28	0.26	CC/C**T** vs**TT**	0.60	0.41-0.88	0.0089
rs3092945	CD40LG	0.38	0.37	C**T**/CC vs**TT**	1.50	1.11-2.04	0.0082
rs8176719	ABO	0.37	0.38	**II**/DI vs DD	0.70	0.07-0.60	0.0080

Table 7.1: **Genotype based test of association with severe malaria.** Results for the best fitted model is shown. The ancestral is shown in **bold** and is taken from dbSNP build 137 and where the ancestral allele is not identified a reference allele was assigned based on the variant in the human reference sequence in Ensembl build 56. The odds ratios are with respect to the derived allele and case subjects.

Mild vs healthy controls with/without parasitaemia							
		MAF					
SNP	**Gene**	**Mild**	**Controls**	**Model**	**OR**	**95% CI**	**P-value**
		(n=1600)	(n=1374)				
rs334	HBB	0.03	0.03	**A**T vs TT/**AA**	0.42	0.27-0.67	0.0046
rs1012356	IL22	0.44	0.43	**TT**/**AT** vs AA	0.75	0.63-0.89	0.0013
Mild vs healthy controls without parasitaemia							
		MAF					
SNP	**Gene**	**Mild**	**Controls**	**Model**	**OR**	**95% CI**	**P-value**
		(n=1600)	(n=524)				
rs708567	IL17RE	0.47	0.48	**G**A/**GG** vs AA	1.44	1.14-1.82	0.0022
rs334	HBB	0.03	0.03	**A**T vs**AA**/TT	0.48	0.33-0.71	0.0002
rs2243250	IL4	0.24	0.25	TT vs **CC**	0.79	0.68-0.93	0.0048
rs2239704	LTA	0.33	0.33	**GG**/**G**T vs TT	0.66	0.49-0.90	0.0087
Mild vs healthy controls with parasitaemia							
		MAF					
SNP	**Gene**	**Mild**	**Controls**	**Model**	**OR**	**95% CI**	**P-value**
		(n=1600)	(n=850)				
rs334	HBB	0.03	0.03	**A**T vs**AA**/TT	0.77	0.54-1.09	0.1440
rs1012356	IL22	0.45	0.43	TT/T**A** vs**AA**	0.70	0.58-0.85	0.0003
rs2227491	IL22	0.27	0.28	CC/C**T** vs**TT**	0.60	0.41-0.88	0.0089
rs3092945	CD40LG	0.40	0.41	C**T**/CC vs**TT**	1.50	1.11-1.04	0.0082
rs8176719	ABO	0.36	0.36	D**I**/**II** vs DD	0.70	0.07-0.60	0.0080

Table 7.2: **Genotype based test of association with mild malaria**. Results for the best fitted model is shown. The ancestral is shown in **bold** and is taken from dbSNP build 137 and where the ancestral allele is not identified a reference allele was assigned based on the variant in the human reference sequence in Ensembl build 56. The odds ratios are with respect to the derived allele and case subjects.

		MAF					
SNP	**Gene**	**Cases**	**Mild**	**Model**	**OR**	**95% CI**	**P-value**
		(n=670)	(n=1600)				
rs334	HBB	0.02	0.02	**A**T vs **AA**/TT	0.65	0.40-1.06	0.0863
rs3024500	IL10	0.42	0.45	**GG**/A**G** vs AA	1.38	1.14-1.69	0.0012
-	-	0.42	0.45	AG/GG vs AA	1.29	1.08-1.55	0.0054
rs3092945	CD40LG	0.34	0.38	**TT**/C**T** vs **TT**	0.65	0.44-0.96	0.0323

Table 7.3: **Genotype based test of association of severe vs mild malaria**. Results for the best fitted model is shown. The ancestral is shown in **bold** and is taken from dbSNP build 137 and where the ancestral allele is not identified a reference allele was assigned based on the variant in the human reference sequence in Ensembl build 56. The odds ratios are with respect to the derived allele and case subjects.

SNP	Gene	MAF INF (n=850)	MAF UNF (n=524)	Model	OR	95% CI	P-value
rs334	HBB	0.03	0.04	**A**T vs **A**A/TT	0.65	0.43-0.96	0.03
rs708567	IL17RE	0.47	0.55	**G**A/TT vs **GG**	1.41	1.11-1.78	0.004

Table 7.4: **Genotype based association of uninfected vs infected controls.** INF- asymptomatic population controls and UNF refers to uninfected population controls. The ancestral is shown in **bold** and is taken from dbSNP build 137 and where the ancestral allele is not identified a reference allele was assigned based on the variant in the human reference sequence in Ensembl build 56. The odds ratios are with respect to the derived allele and case subjects.

Comparison	Marker	Uninfected OR (effect)	Mixed OR (effect)	Asymptomatic OR (effect)
Severe malaria	rs334	0.33 (3.03)	0.42 (2.38)	0.49 (2.04)
Mild malaria	rs334	0.48 (2.08)	0.42 (2.38)	0.77 (1.30)
Asymptomatic	rs334	0.65 (1.54)	-	-

Table 7.5: Summary of the protective effect of sickle (rs334) for three malaria phenotypes (rows) with different infection statuses for controls (columns) [data taken from Table 7.1 to 7.4].

7.4.2 Host-parasite interactions

7.4.2.1 Drug resistant parasites and host disease phenotypes

Over the last few decades parasites have evolved resistance against a range of popular antimalarial drugs. Several of the parasite genes involved and the base changes and subsequent amino acid substitutions that confer drug resistance have been characterized. This section examined the association of drug resistant mutations in the *pfdhfr* and *pfdhps* genes in relation to malaria disease phenotypes. Here, I assessed two related questions; Is a child who has parasites with known resistance alleles more likely to develop mild or severe malaria? Moreover given that several drug resistant alleles may occur in a single gene, does having parasites with multiple alleles change the risk for developing more severe disease? To test these questions, I begun by carrying out single SNP assessment of drug resistant alleles in *pfdhfr* and *pfdhps* genes in a set of 657 severe malaria cases, 1600 mild malaria cases and 1079 asymptomatic population controls. The frequencies of the resistant alleles (non-reference) are shown in Table 7.6. For the sake of comparison, I included public access allele frequency data from malariaGEN. In general allele frequencies in my disease categories were lower that those observed in the malariaGEN Africa-wide data.

Table 7.7 shows the breakdown of the SNP data into counts of individuals by three genotypes (i.e sensitive (homozygous reference alleles), mixed (both reference and non-reference alleles) and resistant (homozygous non-reference alleles)) and three disease categories analyzed. In order to aid clear interpretation of the data, in my analysis, however, I considered individuals who were either homozy-

gous sensitive or resistant and ignored infections with mixed alleles (heterozygous samples). Notwithstanding the obvious loss of data, I used this approach because individual infections may differ fundamentally in the proportions of alleles. For example, some samples have a low proportion of resistant alleles at a given locus (ie more sensitive) and others have a high proportion of resistant alleles (ie more resistant) and a concomitant range of intermediate proportions. Thus, the current challenge with this type of analysis is how to adequately code such data to permit an accurate partitioning of individuals for analysis. As discussed in Chapter 6 such allele ratios are currently difficult to accurately genotype, particularly when an allele frequency is below 10% in a sample. However, I did undertake analysis of the heterozygous samples by adopting a rather simplistic approach, where I coded mixed infections as equal to explore the data and highlight the issues (results may be reviewed in Appendix F). In some instances, others have genotyped heterozygous samples with the majority allele (Miotto et al., submitted), but in the specific case of drug resistance, this may result in the over representation of sensitive (ie reference) alleles.

I used the Fisher exact test to assess the risk of disease association with resistant alleles. No significant association of parasite single SNPs was observed when severe malaria cases were compared with mild cases (Tables 7.8) or asymptomatic population controls (Table 7.9). Likewise when mild malaria cases were compared with asymptomatic controls, no significant associations were found, except for the unexpected significant association of the *pfdhfr* codons 51 with increased risk of mild malaria (OR: 1.41, 95% CI: 1.11-1.80, P = 0.004) [Table 7.10]. Thus, in general there was no clear pattern of association of drug resistant alleles with

malaria disease phenotypes at the single locus level.

Further, I did not observe any significant difference in the overall distribution of homozygous resistant alleles (i.e for 1, 2, or 3 alleles per infection at any locus) and that of fully sensitive alleles between disease phenotypes (Table 7.11). Given the low frequency of *pfdhps* A581G in my data, the two *pfdhps* SNPs were not included in this analysis and was removed from further analysis.

Meanwhile, it is known that sequential accumulation of point mutations in the *pfdhfr* gene by parasites may significantly increase the individual risk of treatment failure. This propensity of parasites to accumulate these mutations may therefore, make it difficult to appropriately partition individuals based on single loci. It may, therefore, be more appropriate to examine combinations of alleles (haplotypes) at these loci. Therefore, I next used the Cochrane-Armitage test to assessed whether there was a linear trend in *pfdhfr* SNP combinations and the risk of malaria disease.

As evidenced in Table 7.12, there was a significant difference in the distribution of the frequencies of resistant alleles ranging from 1, any single resistant allele to 3- samples with three *pfdhfr* resistant alleles (51I + 59R + 108N) for mild or severe versus asymptomatic controls but not with severe versus mild or severe versus asymptomatic controls. The Cochran-Armitage test for a linear trend in the proportions of these resistant alleles showed a significant trend for the accumulation of these alleles during the progress from asymptomatic malaria to mild febrile disease (χ^2_{trend} = 5.38, P = 0.020). The significant difference (p = 9.33e-05) between the standard χ^2 test and the trend test (χ^2_{trend}) suggests,

however that the linear trend alone does not explain all the relationship between mild malaria and asymptomatic malaria (Table 7.12).

To investigate this further, the association of resistant allele haplotypes (known to be functional) with disease phenotypes was assessed. The frequency of occurrence of these among disease categories are shown in Table 7.13. A linear regression model was fitted to the data using these functional combinations of point mutations as covariates. The presence of the *pfdhfr* double resistant allele (59R +108N) was observed to be associated with 85% increased risk of severe malaria (OR:1.85; 95%CI:1.19-2.87; p = 0.0058) when severe cases were analysed with asymptomatic controls and the triple resistant allele (51I +59R + 108N) was associated 44% increased risk of mild malaria (OR:1.44; 95%CI:1.07-1.93; p = 0.015) when mild malaria cases were analysed with asymptomatic controls (Table 7.14). These *pfdhfr* double and triple resistant alleles have been shown in *in vitro* studies to confer higher levels of sulfadoxine-pyrimethamine resistance. In summary, this analysis did not show any differences in the distribution of single SNPs between the disease phenotypes investigated. Analysis of combinations of SNPs known to confer higher levels of drug resistance showed a significant trend toward increased risk of clinical malaria but not disease severity.

SNPs	NRAF Africa	NRAF SEA	NRAF PNG	NRAF Ghana SM	NRAF Ghana MM	NRAF Ghana Asym	Ghana All	REF Allele	NREF Allele
dhfrN51I	0.71	0.92	0.0	0.28	0.31	0.23	0.28	A	T
dhfrC59R	0.78	0.92	1.0	0.55	0.48	0.45	0.49	T	C
dhfrS108N	-	-	-	0.59	0.53	0.52	0.55	C	A
dhpsG437A	0.5	0.0	0.86	0.22	0.23	0.21	0.22	T	G
dhpsA581G	0.15	0.81	0.0	0.04	0.05	0.06	0.05	A	G

Table 7.6: **Non-Reference Allele frequency (NRAF) of parasite SNPs in different populations.** Populations include Southeast Asia (SEA), Papua New Giunea (PNG), and Africa-wide (MalariaGEN open access data). NRAFs for my data (i.e Ghana) are shown by disease status (SM-severe malaria, MM-mild malaria, Asym-Asymptomatic). Reference alleles are characterized as sensitive and non-reference alleles are resistant

SNP	Genotype	Severe	Mild	Asymptomatic
DHFR51	Sensitive	267	517	331
	Mixed	118	242	179
	resistant	148	338	153
	Total count	533	1,097	663
DHFR59	Sensitive	158	290	175
	Mixed	92	273	189
	resistant	308	526	300
	Total count	558	1,089	664
DHFR108	Sensitive	142	248	136
	Mixed	87	259	189
	resistant	324	587	349
	Total count	553	1,094	674
DHPSG437A	Sensitive	223	670	372
	Mixed	48	41	42
	resistant	12	41	28
	Total count	283	752	442

Table 7.7: **Counts of severe, mild malaria and asymptomatic population controls by parasite genotypes.** This gives a snapshot of the data presented. sensitive - reference alleles at the given locus, mixed - occurrence of both reference and resistant (non-reference) alleles and resistant - homozygous occurrence of non-reference alleles.

Severe vs Mild malaria						
SNP	**Comparison**	**Severe**	**Mild**	**OR**	**95%CI**	**P-value**
DHFR51	Sensitive	267	517	0.85	0.66-1.19	0.196
	resistant	148	338			
	Total count	415	855			
DHFR59	Sensitive	158	290	1.07	0.84-1.37	0.584
	resistant	308	526			
	Total count	466	816			
DHFR108	Sensitive	142	248	0.96	0.75-1.24	0.800
	resistant	324	587			
	Total count	466	835			
DHPS437	Sensitive	223	670			
	resistant	12	41	0.88	0.41-1.74	0.870
	Total count	235	711			

Table 7.8: **Association of parasite genotypes with severity of malaria disease**. Severe cases are coded as 1 and mild malaria cases 0 and sensitive alleles are coded 0 and resistant alleles 1.

	Severe vs Asymptomatic controls					
SNP	**Test**	**Severe**	**Controls**	**OR**	**95%CI**	**P-value**
DHFR51	Sensitive	267	331			
	resistant	148	153	1.21	0.91-1.61	0.202
	Total count	415	484			
DHFR59	Sensitive	158	175			
	resistant	308	300	1.14	0.86-1.51	0.341
	Total count	466	475			
DHFR108	Sensitive	142	136			
	resistant	324	349	0.89	0.67-1.19	0.433
	Total count	466	485			
DHPS437	Sensitive	223	372			
	resistant	12	28	0.71	0.32-1.48	0.399
	Total count	235	400			

Table 7.9: **Association of parasite genotypes with severity of malaria disease**.Severe cases are coded as 1 and asymptomatic cases 0 and sensitive alleles are coded 0 and resistant alleles 1.

Mild malaria vs Asymptomatic controls						
SNP	**Test**	**Mild**	**Controls**	**OR**	**95%CI**	**P-value**
DHFR51	Sensitive	517	331			
	resistant	338	153	1.41	1.11-1.80	0.004
	Total count	855	484			
DHFR59	Sensitive	290	175			
	resistant	526	300	1.06	0.83-1.35	0.674
	Total count	816	475			
DHFR108	Sensitive	248	136	0.92	0.71-1.18	0.922
	resistant	587	349			
	Total count	835	485			
DHPS437	Sensitive	670	372			
	resistant	41	28	0.81	0.48-1.39	0.438
	Total count	711	400			

Table 7.10: **Association of parasite genotypes with mild malaria disease**. Mild cases are coded as 1 and asymptomatic cases 0 and sensitive alleles are coded 0 and resistant alleles 1.

Any fully resistant alleles vs fully sensitive alleles						
Allele type	**severe**	**Mild**	**severe**	**Asym**	**Mild**	**Asym**
Sensitive	118	203	118	89	203	89
Resistant	208	393	208	180	393	180
	OR(95% CI)	**P-value**	**OR**(95% CI)	**P-value**	**OR**(95% CI)	**P-value**
Fisher test	1.10 (0.81-1.47)	0.516	1.15 (0.81-1.64)	0.438	1.04 (0.76-1.44)	0.816

Table 7.11: **Disease association analysis using the Fisher exact test.** Fully sensitive - no non-reference alleles across allele loci: any fully resistant - presence of any 1, 2 or 3 non-reference alleles in a given sample.

A	**Test of homozygous resistant allele combinations**					
Allele count	**severe**	**Mild**	**severe**	**Asym**	**Mild**	**Asym**
0	118	203	118	89	203	89
1	1	4	1	15	4	15
2	77	126	77	52	126	52
3	130	263	130	113	263	113
Total count	208	393	208	180	393	180
B						
Linear trend	**χ^2-value**	**P-value**	**χ^2-value**	**P-value**	**χ^2-value**	**P-value**
χ^2_{Std} test	1.87	0.392	16.35	0.00028	20.65	3.28e-05
C-A χ^2_{Trend}	0.84	0.360	1.70	0.1919	5.38	0.020
χ^2_{diff}	1.03	0.309	14.65	0.00013	15.26	9.33e-05

Table 7.12: **Counts of multiple allele combinations per sample across *pfdhfr* and linear trend analysis.** In Table A, Asym-Asymptomatic, 1-samples with N51I or C59R or S108N resistant alleles, 2-samples with 51I + 108N or 59R + 108N resistant alleles and 3- samples with N51I + C59R + S108N resistant alleles. Table B shows the standard χ^2 test (χ^2_{Std}) and Cochrane-Armitage (C-A) test for a linear trend. When the difference between the χ^2_{trend} and χ^2_{Std} is not significant then the linear trend explains all of the relationship.

Haplotypes	**Severe**	**Mild**	**Asymptomatic**
51I + 59R	0.00	0.00	0.00
51I + 108N	0.02	0.03	0.01
59R + 108N	0.15	0.09	0.06
51I + 59R + 108N	0.28	0.23	0.15

Table 7.13: Frequencies of known functional combinations of point mutations in *pfdhfr*.

resistant alleles	**Comparison**	**OR**	**95% CI**	**P-value**
59R + 108N	Severe vs Mild	1.28	0.88-1.85	0.182
51I + 59R+108N	Severe vs Mild	0.95	0.71-1.27	0.733
59R + 108N	Severe vs Asymptomatic	1.85	1.19-2.87	0.0058
51I + 59R+108N	Severe vs Asymptomatic	1.36	0.97-1.91	0.073
59R + 108N	Mild vs Asymptomatic	1.44	0.97-2.16	0.072
51I + 59R+108N	Mild vs Asymptomatic	1.44	1.07-1.93	0.015

Table 7.14: **Disease association analysis of known functional resistant alleles.** Disease groups on the left are coded as 1 and those on the right as 0. Resistant alleles are coded as 1 and sensitive alleles as 0.

7.4.2.2 Host-parasite genotype interactions

Thus far, I have established that sickle protects against malaria disease in general but its effect is strongest with severe malaria. Also, clinical malaria cases (either severe or mild) are more likely to have parasites that have drug resistant alleles that confer higher levels of resistance (i.e a fitness advantage) compared to population controls. However, despite the sickle trait protection, it is known that some children with the trait still develop severe malaria. Therefore, it makes sense to ask if the fitness advantage conferred by drug resistant alleles modifies the protective effect of sickle.

I assessed this by examining whether the effect of sickle on disease status depends on drug resistant alleles. In other words, I wanted to assess whether the protective effect of sickle (i.e either strength or direction) is modified by the type of drug resistant allele the parasites carry. I tested the *pfdhfr* codon 108 Serine to Asparagine substitution characterized as the basal mutation that initiates antifolate drug resistance and the double (51I + 108N or 59R + 108N) and triple (51I + 59R + 108N) resistant alleles (haplotypes) that were found to significantly increase the risk of clinical malaria (see section 7.4.2.1 Table 7.14). I fitted a logistic regression model to my data for severe and mild malaria, using the *pfdhfr* locus and the sickle locus (rs334 or HbS) as predictor variables (that capture their main effects) and two-way sickle by resistant allele interaction effects (equation 7.2). The alleles for *pfdhfr* were coded as 0 for reference alleles (sensitive) and 1 for presence of the resistant alleles (non-reference) and for the sickle locus, normal haemoglobin type (AA) was coded as 1 and sickle trait (AT) as 0. The interaction term for normal haemoglobin and drug sensitive alleles was omitted

from the model, so that the odds ratios refer to this as baseline level. The full model may be written as;

$$Y \sim \text{Bernoulli}(p_i) \tag{7.1}$$

$$\log\left(\frac{p_i}{1-p_i}\right) = b_o + b_1 H_i + b_2 G_i + b_3 H_i \times G_i \tag{7.2}$$

Where Y is the disease status of individual i with probability p (severe malaria coded as 1 and mild malaria as 0), b_o is the intercept, b_1, b_2 and b_3 are the regression coefficients. H_i is the sickle locus, G_i is the drug resistant allele.

I observed a significant interaction between the *pfdhfr* resistant alleles and the sickle-cell trait (OR:0.16, 95% CI: 0.03-0.86, P = 0.032) [Table 7.16]. These results signify departure from the simple multiplicative (or independent effects) model. The negative coefficients indicate that the joint effects of these resistant alleles and the sickle-cell trait is less than that obtained by simply multiplying the odds. The effects of these interactions can be visualized from the interaction plot (Figure 7.4). The plot shows four related findings;

(1) that there is no significant difference in the risk of severe malaria for children with normal haemoglobin who are infected with parasites with *pfdhfr* sensitive or resistant alleles,

(2) that children with the sickle-cell trait (HbAS) are protected against severe malaria when infected with parasites with *pfdhfr* sensitive alleles as compared to children who have normal haemoglobin (HbAA) and are infected with parasites with *pfdhfr* sensitive alleles,

(3) that when children with the sickle cell trait are infected with parasites that

carry these resistant alleles, the protective effect of the sickle trait is modified, and

(4) that the sickle trait protective effect against severe malaria is modified to the extent that the risk of severe malaria is not different between children with normal haemoglobin and those with the sickle-cell trait who are infected with parasite carrying *dhfr* resistant alleles.

	Normal haemoglobin		**Sickle-cell trait**	
Phenotype	**HbAA/ Sen**	**HbAA/ Res**	**HbAS/ Sen**	**HbAS/ Res**
Severe malaria	11	197	2	116
Mild malaria	19	370	18	184

Table 7.15: Counts of *pfdhfr* sensitive (Sen) and resistant (Res) alleles among individuals with sickle-cell trait (HbAS) and normal haemoglobin (HbAA).

	Host-by-parasite genotype interaction				
Model terms	**Coeff**	**Std.Err**	**OR**	**95% CI**	**P-value**
DHFR	1.6507	0.8361	5.21	1.01-26.83	0.048
HbS	1.7359	0.7547	5.67	1.29-24.91	0.021
DHFR by HbS	-1.8196	0.8490	0.16	0.03-0.86	0.032

Table 7.16: **Model for severe malaria with interactions terms for the sickle-cell locus (HbS) and *pfdhfr* resistant alleles (DHFR).** Coeff- coefficients of the terms; StErr - standard error of the terms.

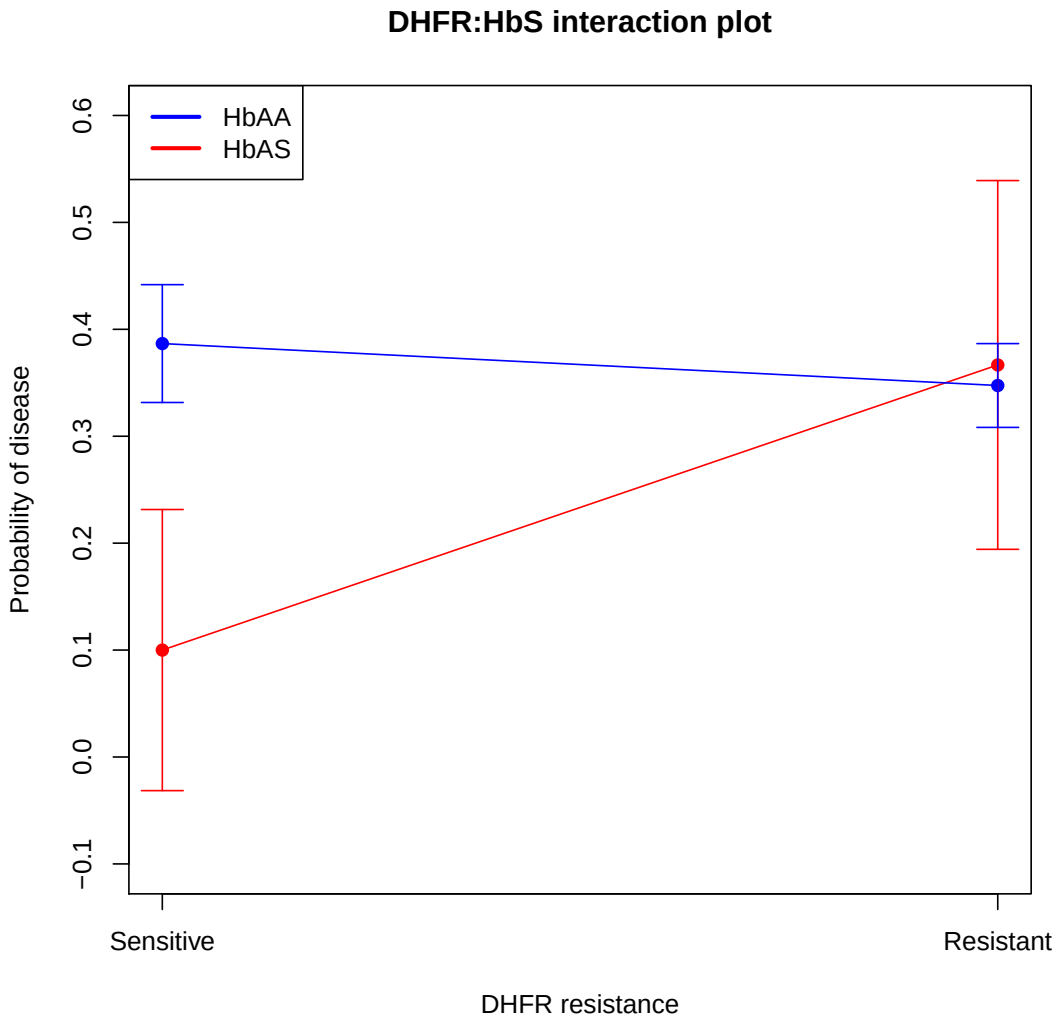


Figure 7.4: **Risk estimates for sickle and *pfdhfr* interactions.** *pfdhfr* resistant alleles included the single (108N), double (108N +51I or 59R) and triple (108N + 51I + 59R) resistant allele. The baseline is normal haemoglobin (HbAA) + sensitive parasites.

7.5 Discussion

7.5.1 Host genetic factors

As proof-of-principle and consistent with previous case-control analysis undertaken in this population[73], the evidence for sickle-cell trait (HbAS) protection against severe malaria was the strongest signal in this data (Table 7.1). The sickle-cell locus (rs334) is located on chromosome 11 of the human β-globulin (HBB) gene. Though the mutation is lethal, balancing selection has maintained the recessive allele in malaria endemic populations in the heterozygous form for the survival advantage it confers against death from severe malaria[4,246]. To date, this polymorphism is the most reproducible in population studies across different malaria endemic populations in Africa with different transmission intensities[92,246]. My data supports a general protection by the sickle-cell trait against malaria infection, but there appears to be a gradient of increasing protection as febrile symptoms progress from mild to severe disease. Indeed, the strength of the association signal varied with the infection status of population controls, with the strongest effect (3-fold protection) observed when severe cases were compared with uninfected population controls (Table 7.5).

These observations, suggest strong disease protection for the sickle-cell carrier state and marginal protection against symptomless parasitaemia in this population. Findings that are consistent with previous studies in Kenya, that showed that the sickle-cell carrier state had 50% protection against mild malaria, and about 90% protection against severe or complicated malaria. However, no effect was observed in the Kenyan study on prevalence of symptomless parasitaemia[246].

Besides the sickle-cell locus, the evidence supporting other associations, mainly in cytokine or cytokine related gene polymorphisms are not far fetched. Over the last few decades, a body of literature has emerged that contend that the effects of cytokines combine to make falciparum malaria primarily an inflammatory cytokine driven disease[37,180,181]. The balance between pro- and anti-inflammatory cytokines being critical for the onset of systemic inflammation[34]. An excessive cytokine-mediated inflammatory response could result in severe malaria[97,137]. Therefore, the association of IL-4 promoter polymorphism (IL-4-589C) with 55% reduced risk of SM (Table 7.1), and IL-22 polymorphism with 59% reduced risk of SM both underscore the role of cytokines in malaria pathology[113,217,230]. Conversely, IL-17 receptor C polymorphism (rs708567) was associated with increased risk of SM in this population.

The role of IL-10 in mediating malaria-induced inflammation is well established (reviewed in[37]). High TNFα levels and low IL-10 levels have been linked to severe malarial anaemia (SMA), with the ratio between these two opposing cytokines being the most distinctive predictor of clinical presentation[41,46,106,166]. I did not find any association of TNFα polymorphisms with severe malaria. This could be because the TNF SNPs genotyped in this study did not appropriately tag the causal variant in this population. However, to date, the association of TNFα genetic polymorphisms with severe malaria has been difficult to replicate, as analysis with larger sample sizes have failed to produce any signal of protection (Kirk Rockett, personal communication). IL-10 polymorphisms have been assiciated with severe malaria in different populations[118,167,252] Therefore, the observed association of IL-10 with increased risk of disease progression (Table

7.3), in the KNDs population, where SMA is the most prevalent phenotype[157], may corroborate the role of IL-10 signaling in severe malarial anaemia in this population. Overall, my analysis approach and the results obtained highlights the importance of appropriate ascertainment of controls for case-control analysis of disease susceptibility. This is evidenced by the dampening of the association signal for sickle-cell trait protection when asymptomatic controls were used.

7.5.2 Parasite genetic factors

It is thought that the fitness conferred by antimalarial drug resistance may not contribute to virulence *per se*, drug resistant parasites may be less fit in the absence of drug pressure and much less able to multiply as compared to their drug-sensitive counterparts[238]. Going by this assertion, one would expect two phenomena at the population level; first, since drug resistant alleles may only contribute to virulence under drug pressure, in the absence of drug pressure, the prevalence of resistant alleles among asymptomatic individuals should reflect the population-level frequencies of alleles. Second, that variants that confer higher levels of resistance should be more prevalent in clinical cases as compared to asymptomatic infections. The evidence from my data support a general trend towards high prevalence of multiple drug resistance alleles in the *pfdhfr* gene among children with clinical malaria (mild and severe malaria) as compared to asymptomatic infections (Table 7.14). The *pfdhfr* double resistant allele (59R + 108N), and the triple allele (51I + 59R + 108N) were over represented in clinical cases as compared to asymptomatic controls. These alleles, however, did not differ significantly between severe and mild malaria cases. This may

highlight the importance of other factors (host genetic and/or environmental) in the progression from mild febrile disease to severe life-threatening disease.

Though, it is not possible to establish the sequential acquisition of these resistant alleles with this data, the results from looking at combinations of alleles per sample are consistent with *in vitro* studies that show that sequential acquisition of point mutations in these genes contribute to higher levels of resistance[194,224]. The *pfdhfr* triple resistant allele (51I + 59R + 108N) is characterized to result in 1600-fold increased tolerance of parasites to pyrimethamine (pyr), whereas the S108N single mutation increases pyr tolerance 100-fold[194]. Whilst parasites with the double resistant alleles (51I + 108N or 59R + 108N) have a similar tolerance between the single and the triple resistant alleles[194]. The double and the triple resistant alleles are known to predict sulfadoxine-pyrimethamine (SP) treatment failure[148]. The wide deployment of SP in Ghana in the early 1980s as first line treatment for uncomplicated malaria may account for the trends observed, since these data were collected before the introduction of artemisinin combination treatments in Ghana. In summary, these findings underscore the fitness advantage conferred by acquisition of antimalarial drug resistant alleles, but the lack of association at the single SNP level may highlight host clearance of such drug resistant alleles that confer lower levels of resistance (i.e lower fitness) as supported by one study in Mali[45].

7.5.3 Host-parasite interactions

I explored host-parasite interactions using the sickle-cell and *pfdhfr* loci, two genetic factors that have opposing effects on the risk of severe malaria. I observed significant interactions between *pfdhfr* resistant alleles and the sickle cell trait that signify that children who have the sickle-cell trait (HbAS) have about 3-fold protection against severe malaria when infected with parasites that have *pfdhfr* sensitive alleles as compared to children who have normal haemoglobin (HbAA) and are infected with parasites that have *pfdhfr* sensitive alleles. This is consistent with the effect of the sickle-cell trait observed in this population in subsection 7.4.1, Table 7.5. In addition, the interaction between these loci indicate that when children who have the sickle cell trait are infected with parasites that carry *dhfr* resistant alleles, the protective effect of the sickle trait is abrogated and the risk of severe malaria is not different between children with normal haemoglobin and those with the sickle-cell trait. Also, there was no significant difference in the risk of severe malaria for children with normal haemoglobin who are infected with parasites with *pfdhfr* sensitive or resistant alleles.

These findings raise the possibility that parasites that have alleles that confer higher levels of drug resistance may be better adapted to the sickle condition and could grow relatively better and increase the individual risk of severe disease. These results bring to the fore, an observation that was made several decades ago, that certain strains of parasites may be better adapted to some host polymorphisms[176]. Also, Williams *et al.,* (2005), observed that the protection afforded by the sickle-cell trait was not independent of malaria exposure as the age-specific protection increased with age[246]. This study estimated a 20% in-

crease in protection by age 2 years. Thus, some children who have the sickle trait may be less able to naturally clear certain parasites as compared to those with normal haemoglobin due to difference in the frequency of exposure. Despite the possibility that these findings could be due to chance or an underlining selection bias, perhaps, this is the first evidence that drug resistance alleles may be better adapted to the sickle condition. However, this assertion should be tempered by the fact that the sickle frequency in this data is low, as reflected in the small counts observed in Table 7.15.

Overall, this data are interesting, and to the best of my knowledge, this is the first time that children who bear the brunt of malaria disease ($\leq$ 5 years of age) have been genotyped for both their own polymorphisms and polymorphisms in the parasites that they carried for exploring host-parasite interactions. Though, there is a paucity of literature to aid interpretation of such analysis, the results suggest that drug resistant alleles apart from having a fitness advantage under drug pressure, may also survive better under the oxidative stress and general inflammatory conditions imposed by the sickle-cell trait. As evidenced, the sickle-cell trait protects against severe life-threatening malaria in this population, but, yet some children who have the sickle-cell trait still get severe malaria. Therefore, it goes without saying that certain parasites will have evolved mechanisms to survive and multiply in sickle red blood cells to cause severe disease. The question is, could such parasites have drug resistance related fitness? The evidence from this exploratory analysis is suggestive of that, but is it real? My plan is to validate these findings by carrying out replication studies in other malariaGEN sample sets from Sub-Saharan Africa, particularly in settings where the frequency

of sickle is high. I also intend to further investigate the region of this gene by using my parasite sequencing data to compare the genomes of fully sensitive and fully resistant parasites in order to rule out the possibility of this finding being a "bystander effect" (i.e a proxy for a functional variant not directly assayed).

7.6 Summary

- The sickle-cell trait (HbAS) was observed to confer protection against both severe and mild malaria disease with marginal protection against infection in the northern Ghana.
- An IL-4 promoter polymorphism (rs2243250), IL-22 (variant rs2227491) and variant rs1012356 were associated with increased risk of severe malaria.
- A variant of IL-10 (rs3024500) was associated with increased risk of progression from mild to severe disease when mild cases were compared to severe cases
- The IL-17 receptor C was also associated with increased risk of SM. The derived-allele homozygous state (AA) of this gene was associated with 44% increased the risk of SM.
- The Cochran-Armitage test for a linear trend in the proportions of *pfdhfr* resistant alleles between mild malaria cases and asymptomatic controls showed a significant increased risk of mild malaria as the number of resistant alleles per parasite increased from 1 to 3 ($\chi^2 = 5.381$, P = 0.020).

- The *pfdhfr* functional double resistant alleles (59R + 108N) known to confer higher levels of sulfadoxine-pyrimethamine resistance was found to generally confer an increased risk of clinical malaria (either severe or mild malaria).

- An interaction between *pfdhfr* resistant alleles and the sickle-cell trait indicates that the joint occurrence of these genetic factors may modify the protection for sickle individuals who have *pfdhfr* resistant alleles and increase their risk of severe malaria to about the same level as individuals with normal haemoglobin who have these resistant alleles.

7.7 Conclusion

In this chapter, I undertook a host-parasite genetic interaction analysis of known antimalarial drug resistance SNPs in the *pfdhfr* gene and the sickle-cell trait. First, re-evaluating the evidence for malaria susceptibility in an expanded dataset of human genetic polymorphism from Ghana. This analysis confirmed the association of the sickle-cell trait with a 3-fold protection against severe malaria in Ghana and highlighted a role for cytokine genes in malaria susceptibility in this region. Second, host-parasite interaction analysis between the sickle locus (rs334) and the *pfdhfr* locus known to confer antifolate resistance revealed that children with the sickle-cell trait (HbAS) are protected against severe malaria only when infected with parasites with *pfdhfr* sensitive alleles as compared to children who have normal haemoglobin (HbAA) and are infected with parasites with *pfdhfr* sensitive alleles, but, when children with the sickle cell trait are infected with

parasites that carry these resistant alleles, the protective effect of the sickle trait is lost to *dhfr* resistant alleles. However, there was no significant difference in the risk of severe malaria for children with normal haemoglobin who are infected with parasites with *pfdhfr* sensitive or resistant alleles. The crucial point of this exploratory analysis is the potential for useful insight to be gained, particularly in this era of high-throughput genetics and whole genome sequencing. With rapid decline in whole genome sequencing cost, host-parasite interactions can now be conducted at several thousands of loci to look for trade-offs in the host-parasite interaction that could hold potential benefit for malaria control.

Chapter 8

Discussion and Conclusion

8.1 General discussion

8.1.1 Setting up the framework

The introduction to this work highlights one important fact, that malaria is a multi-genome disease involving the mosquito genome, the human genome and the parasite genome. Each of these genomes have their unique levels of complexity and challenges that they each pose to malaria researchers. In the past, it proved difficult to integrate research across these three genomes, but, the advent of high-throughput platforms, powerful computers, informatic tools and above all, international collaborations in malaria research such as the MalariaGEN[151], has provided a platform for integrating host (both human and mosquito) and parasite genetics. In this context, the focus of my thesis was to undertake sampling and purification of *P. falciparum* DNA from rural communities in northern Ghana for next-generation sequencing. The main motivation was to characterize

P. falciparum genetic diversity and to extend existing case-control samples (collected to investigate host genetics of resistance to severe malaria) for exploring host-parasite genetic interactions. Therefore, the sampling sites were purposely selected to represent the rural study population in its socio-cultural, demographic and ecological diversity. This strategy and the field methods used proved adequate in purifying high quality parasite DNA suitable for Illumina deep sequencing. Over 500 field isolates were obtained and to date, 300 of these have been sequenced. There were, however, some limitations. My sampling framework did not include asymptomatic individuals who form the majority of infections in these communities. As stated in chapter 2 (section 2.1.1), the sampling strategy adopted in this study was aimed at maximizing the chance for purifying parasite DNA suitable for sequencing. This necessitated the sampling of individuals presenting at community clinics with a malarial fever, though such clinical samples may suffer from sampling bias due to health seeking behaviours that have traditional underpinnings and poor access to facilities. However, the use of front-line community clinics with well defined catchment areas was to limit potential bias in patient selection that may be due to access. Though sampling asymptomatic individuals may have an added benefit, I was unable to obtain ethical approval to sample such individuals. The field protocols I used required 2-5ml of whole blood and there was no evidence at the time to justify blood draws from such apparently healthy individuals. My sample set may therefore, be limited as a representative population sample, since it does not include the asymptomatic sub-population of parasites. However, as discussed in chapter 7, I successfully used the Sequenom iPLEX platform to genotype parasite polymorphisms in our previous case-control sample sets (2002/2004 collections) which provides an opportunity to validate any

interesting polymorphisms among asymptomatic individuals in the control set. As sequencing chemistry improves, significant reductions in the quantity of parasite DNA required for sequencing should allow sampling of asymptomatic infections in cross-sectional surveys in the future. However, even then a substantial proportion of either low density or submicroscopic infections may be missed[162]. Further, sampling was not representative across all age groups. The majority of the data was from young children under the age of 10 years; reflective of the general burden of malaria disease among young children in areas of intense transmission.

Despite these shortcomings, I believe that the data generated are characteristic of malaria epidemiology in this part of northern Ghana and are suitable for not only addressing the aims of this work, but also particularly valuable for contributing to the ongoing characterization of the natural diversity of *P. falciparum* worldwide.

8.1.2 *P. falciparum* genetic diversity

Plasmodium falciparum genetic diversity was first investigated using Illumina NGS (Chapter 4). Currently, PCR-free DNA library preparation for Illumina NGS requires $\geq$ 5µg DNA and <60% human DNA contamination. Most of the samples that had lower coverage had human DNA contamination of >50% and tended to have a lower parasite density. These samples were among the initial group of samples collected when I was still optimizing the WBC separation protocols. In general, these earlier collections were also sequenced at a time when the Illumina PCR was undergoing optimization for the *P. falciparum* AT rich genome[171]. This notwithstanding, a conservative genotyping of my samples to-

gether with all other samples in the PGV pipeline yielded 400,000 'typable' (high confidence) SNPs for 277 of my samples. The analysis of 363,330 of these SNPs, which were variable in my Ghanaian population, showed a high prevalence of low frequency variants in this region of northern Ghana. Analysis of genetic relatedness between parasites using all the typed loci revealed no population structure in the parasite population in this region of northern Ghana. When I integrated my data with similar published data from Burkina Faso and Mali, I could still not see any significant population structure between the sites sampled in Ghana, Burkina Faso and Mali. Multiple lines of evidence suggest lack of population structure (or random mating) in areas of intense malaria transmission[7,11,121]. Consistent with these previous evidences, recent overall annual EIR estimates in the KNDs was 140 infective bites per person-years (Asoala et al., in preparation), and range between 10-25 infective bites per month in the areas sampled in Mali (Djimde, personal communication)[83]. On the contrary parasite populations sampled in Cambodia and Thailand, where EIRs are generally around 1 infective bites per person-years were found to show greater population structure[121]. Therefore, the extreme within-host diversity observed in the Ghanian samples and those of West Africa in general may be attributed to a high prevalence of super-infection as a direct consequences of intense transmission. However, demography and the pattern of human migration may also play an important role in shaping the diversity observed in West Africa as opposed to southeast Asia[11,121].

Our current understanding of the complexity of natural infections and epidemiological factors which shape within-host diversity remain poor. Conventionally, within-host diversity has been defined in terms of multiplicity of infection (MOI)

scored by PCR-based genotyping of small numbers of loci (typically <10)[15]. Deep sequencing enables individual read counts to be derived from genome-wide datasets and offers an opportunity for genome-wide description of within-host diversity. One such genome-wide metric is F_{WS}, a metric recently introduced by Manske *et al* 2012[121]. The F_{WS} metric characterizes within-host diversity and its relationship to population-level diversity[15,121]. Essentially, the F_{WS} metric provides an estimate of the risk of out-crossing between parasites within an individual to generate new genotypes during recombination in the mosquito host[15].

Utilizing the F_{WS} metric, a high prevalence of within-host diversity and risk of out-crossing was observed across the different ecological areas of the KNDs (Chapter 4, cf: Figure 4.6). However, there were a small number of infections with low genetic diversity. Consistent with previous reports of low genetic diversity (high risk of inbreeding) in malaria endemic populations with low stable transmission[7,13,18,121,143], all such infections with low genetic diversity (high risk of inbreeding) occurred during the low transmission season and were predominant in communities in the west of the KNDs, where the ecology is dry and rocky with highlands (ranging 200-400M) and the human settlements are comparatively more isolated. Estimated EIRs were comparatively lower in communities in the west compared to the rest of the KNDs[12], particularly during the dry season.

Further, I observed a significant (F = 5.13, adjusted p = 0.01) shift towards infections with low genetic diversity during the dry season (low transmission). This highlights how rapid (down to a couple weeks) natural *falciparum* populations change in response to environmental and other factors. The observed low genetic diversity did not attain the levels of low genetic diversity observed in the

Southeast Asia populations[121], mainly because the dry season is short (January through April) and this does not allow for the expansion of that subpopulation with low genetic diversity before the rainy season (high transmission) sets in. The high transmission season is characterized by a high risk of out-crossing leading to the break down of any incipient low genetic diversity. Such a perennial cycle of population-level mixing of parasite genomes and gene flow has implications for malaria control, as mutations that confer a fitness advantage such as drug resistance may arise during low transmission, and then quickly spread through the parasite population during high transmission.

In areas of very high malaria transmission, conventional MOI is reported to be age-dependent, with highest values reached in young children[19,99,210]. Conversely, studies from areas with low to moderate endemicity either reported little or no influence of age on infection complexity[17,145,254]. Within-host diversity as measured by the F_{WS} metric was found to be negatively correlated with age. Even though I did not sample all age categories in this study, the evidence support a higher prevalence of genetically diverse infections in young children as compared to older semi-immune individuals, which is consistent with the age distribution of conventional MOI in areas of intense malaria transmission[19,99].

In summary, deep sequencing of field isolates from the KNDs has provided useful insight into the genetic diversity of parasite populations in this highly malarious area. The F_{WS} metric has been evaluated in a specific epidemiological setting for the first time. By and large, from this data, the metric is a useful statistic for estimating within-host diversity and the risk of out-crossing (or inbreeding) in a given parasite population. The immense variability of the F_{WS} scores in time

and space and the predominance of low estimates (F_{WS} estimates ≤ 0.6) either reflects the genuine epidemiology of this statistic or may underscore a weakness of it in areas of intense malaria transmission. The accuracy of the F_{WS} metric as gauge for assessing within-host diversity in relation to population-level diversity can only be established with more data from different epidemiological settings. However, its accuracy is evident in this analysis as the F_{WS} metric was able to spot the population flux in within-host diversity and risk of out-crossing (or inbreeding) in the parasite population in Ghana.

Antimalarial drug resistance is an important challenge to national malaria control programmes. The use of certain drugs either as prophylaxis to prevent malaria infection or as treatment for clinical malaria for several decades has led to the adaptation of the parasite to these drugs in what is known as antimalarial drug resistance. One utility of deep sequencing data is variant discovery (Chapter 4) and the opportunity it offers for describing the diversity of specific genes.

Using these data I have identified several known drug resistance mutations and also one novel non-synonymous substitution (glutamate to glutamine) that results in a change from a highly charged side chain to an uncharged one. This mutation is very high in frequency in my data and it is possible that such a change in polarity as a result of this mutation might affect enzyme activity, but to the best of my knowledge this mutation has not been characterized.

I have shown in chapter 5 that antifolate drug resistance conferred by known point mutations in the *P. falciparum* dihydrofolate reductase (*pfdhfr*) and dihydropteroate synthase (*pfdhps*) genes is highly prevalent across the KNDs. Individ-

ual samples with multiple mutations in *pfdhfr* and additional mutations in *pfdhps* were observed, suggesting that sulfadoxine-pyrimethamine (SP) resistance is well established in the general population. A recent phase IV study monitoring the effectiveness and safety of ACTs in Ghana, reported 29% and 31.6% continuous use of SP and chloroquine (CQ) respectively in the KNDs (Atuguba *et al 2012* personal communication). Such illicit use of these drugs may account for the population prevalence of the observed mutations. The results of this population level analysis challenges the current use of SP for intermittent preventive treatment of malaria in pregnancy (IPTP) in the KNDs, and suggest a re-evaluation of the programme.

My analysis found a relatively low (50%) prevalence of *pfcrt* 76T mutation associated with CQ resistance, suggesting a re-emergence of CQ sensitive parasites consistent with the discontinued use of CQ as first-line treatment for malaria in Ghana. This reversal of resistance following the removal of drug pressure, particularly for the *pfcrt* locus has been observed in other endemic populations[104,138], but the 32% reported (Atuguba *et al 2012*) continuous illicit use of CQ in Ghana and elsewhere is worrying.

8.1.3 Sequenom® genotyping of multi-genome infections

Though the value of Next-generation sequencing (NGS) can not be over emphasized, the current high cost of sequencing limits their use in a number of ways; including a survey of new variants in large collections of clinical samples with limited parasite DNA. The Sequenom iPLEX MALDI-TOF mass spectrometry (MS)

based genotyping platform offers throughput on the order of thousands of samples per week at a moderate cost, and is ideal for large-scale validation of variants in samples with limited DNA. We previously demonstrated 99.9% concordance rate between Illumina short-read genotyping and Sequenom for homozygous (single genome infections) haploid calls[121]. However, one important drawback was that the Sequenom platform software could not be relied upon to make accurate genotype calls for mixed infections of *P. falciparum* , which give a considerably more complex pattern of allele distribution than human diploid genotypes for which the standard software was designed. This problem was the main focus of chapter 6, where I begun by preparing artificial mixture ratios of clonal 3D7 and Dd2 laboratory strains with known variants and used these to characterize the Sequenom platform. The MS peak height intensity data generated by genotyping these mixtures was used to develop a genotype-calling algorithm for measuring genotype ratios in multi-genome infections and typing single infections. I then tested my algorithm by genotyping DNA from the sequenced samples (highly mixed) and used my algorithm to measure the ratio of alleles for these samples.

This experiment was limited, firstly, by availability of internal controls (known ratios of alleles) to genotype the field isolates, and secondly only known SNPs could be designed for validation. This notwithstanding, genotype concordance rates between Illumina short-read genotyping and Sequenom was 90% for heterozygous calls (ratio of alleles) when Sequenom gave a valid call. There remain issues with expanding the positive controls set to account for most of the variation in allele ratios in field isolates, and gaining an in-depth understanding of assay specific behaviour that at this point remian poorly understood. Also, evidenced

were borderline cases where my genotype-calling algorithm could not distinguish homozygotes from heterozygotes, in which case heterozygous samples with minority alleles ($< 10\%$) were typed as homozygotes. It is within this small region of ambiguity that Ilumina short-read genotyping is more sensitive than Sequenom genotyping.

The Sequenom iPLEX platform therefore, was successful in typing parasites even within clinical samples with limited DNA (chapter 7). The algorithm does work and the validation was good, a full grasp of the remaining issues is an ongoing process that will be resolved by the broader initiative on *Plasmodium* genotyping in the DK laboratory.

8.1.4 Host-parasite genetic interactions

By using a study design approach to purify *P. falciparum* DNA from rural communities in the KNDs, I was able to characterize *P. falciparum* genetic diversity in this population and extend previous collections of clinical samples, which were mainly for human genetics to begin to explore host-parasite genetic interactions.

A re-evaluation of the expanded human dataset replicated the protective effect of the sickle-cell trait (OR:0.42, 95% CI: 0.27-0.67,P=0.0003). As reported in other studies[246], the sickle-cell trait was associated with disease protection and a marginal protection from infection (OR:0.77, 95% CI: 0.54-1.09). Other polymorphisms, including IL-4 (589C promoter polymorphism), IL-22 (variant rs2227491 and rs1012356), and IL-17 receptor C (IL-17RC) were also associated with malaria disease in this population. Curiously, IL-17RC was associated with increased risk

of malaria infection in homozygous recessive model (OR:1.41, 95% CI: 1.11-1.78). Kuestner *et al.,* (2007) demonstrated that the proinflammatory cytokines IL-17A and IL-17F implicated in the progression of inflammatory diseases both bind IL-17RC with high affinity though IL-17RC is the primary receptor for IL-17F[105]. Thus, polymorphisms in IL-17RC that lead to enhanced signaling in response to these cytokines could result in severe disease. There is growing consensus that malaria is an inflammatory cytokine driven disease, and excessive inflammatory response could lead to severe pathology[98,180,181]. Therefore, it is not surprising that many cytokine genes and some cytokine receptor gene polymorphisms were found as candidate genes in this analysis. However, to date, for most of the candidate genes studied, it has been difficult to ascertain genetic control, thus the role of these gene polymorphisms in malaria resistance (or susceptibility) remain an open debate. Even the mechanism of protection for classical polymorphisms such as the sickle-cell trait is not yet fully understood.

With respect to associations of *pfdhfr* and *pfdhps* genetic polymorphisms with malaria, no significant associations of single SNPs were observed for severe or mild malaria comparisons with asymptomatic controls. The Cochram-Armitage test of linear trend showed a significant increased risk of mild and severe malaria from asymptomatic infection as children become infected with parasites that acquire increasing numbers (1 to 3) of *pfdhfr* polymorphisms. The *pfdhfr* double (51I + 108N or 59R + 108N) and triple (51I + 59R + 108N) resistant alleles were associated with clinical malaria (both mild and severe) [7.14].

To explore host-by-parasite genotype interactions using these resistant alleles, I used the sickle-cell locus (rs334), which was found to protect against severe

malaria in this population. I found significant statistical interactions between the *pfdhfr* resistant alleles and the sickle-cell trait (Table 7.16 and Figure 7.4). The import of this interaction is that children who have the sickle-cell trait (HbAS) have a 3-fold protection against severe malaria when infected with parasites that have *pfdhfr* sensitive alleles as compared to children who have normal haemoglobin (HbAA) and are infected with parasites that carry *pfdhfr* sensitive alleles. In addition, the interaction between these loci show that when children who have the sickle cell trait are infected with parasites that carry *dhfr* resistant alleles, the protective effect of the sickle trait becomes abrogated and the risk of severe malaria no longer differ between children with normal haemoglobin and those with the sickle-cell trait. The question is, are these findings merely due chance? One could speculate that drug resistant forms of *P. falciparum* might have a biological fitness profile that makes them better adapted to specific host genotypes (i.e sickle trait) or more resistant to immune clearance by children who have the sickle trait compared to their counterparts with normal haemoglobin.

The malariaGEN has well-characterized samples sets from different malaria endemic settings that will be used to validate these findings in replication studies, the outcome of which will inform the design of *in vitro* studies to investigate the response of field isolates that carry *pfdfhr* alleles to varying oxidant stress levels akin to the sickle condition. More important overall, however, is that this limited exploratory host-parasite interaction analysis underscores what could be achieved in the near future with host and parasite whole genome sequence data. Barring the ethical issues involved, the increasing availability of deep sequencing data, should provide the needed boost for host-parasite genetic interaction stud-

ies; particularly in examining the interaction of rare human and parasite variants in the epidemiology of severe malaria. The availability of large, international collections of samples such as those of the MalariaGEN provides the platform for such analysis to be undertaken.

8.2 Summary of findings

- In chapter 3 a community-based sampling framework for sampling *P. falciparum* over time and space in the Kassena-Nankana districts (reference population for this thesis) of northern Ghana was established to provide the basis for characterizing the parasite genetic diversity.

- The WBC separation protocol and the use of Plasmodipur® filters and CF11-based filters was adequate for depleting whole blood of WBCs and minimizing human DNA contamination.

- The cheaper CF11-based filters were however, more efficacious than the Plasmodipur® filters in removing WBCs and concentrating iRBCs for parasite DNA extraction

- A minimum blood draw of 1.0 ml (average of 3.0 ml), and parasitaemia of 2000 trophozoites μl^{-1} of blood (50 trophozoites/200 WBCs) is adequate for obtaining quality DNA for suitable for sequencing.

- *Plasmodium falciparum* genetic diversity was described in chapter 4; the allelic spectrum of *P. falciparum* population in the KNDs of Northern Ghana was dominated by low frequency variants.

- No evidence of population structure was observed among parasites samples from different locations in this area.

- A high degree of genome-wide within-host diversity and risk of out-crossing estimated by the $\hat{F}_{WS}$ metric was observed, but, with occasional risk of inbreeding (particularly within communities in the west of KNDs).

- An overall mean $\hat{F}_{WS}$ score of 0.25 was observed, and a high proportion, 88% of infections had $\hat{F}_{WS}$ scores < 0.6, and 51% were of extremely high genetic diversity ($\hat{F}_{WS}$ scores ≤ 0.25), showing over representation of high genetic diversity in this population.

- There was high variability of $\hat{F}_{WS}$ estimates between different geographical location sampled in the KNDs.

- Samples with low genetic diversity were significantly more prevalent in the low transmission season

- No correlation of $\hat{F}_{WS}$ estimates with mild or severe malaria was observed, though low estimates (high genetic diversity) was more prevalent in both disease phenotypes, suggesting an inverse relationship with clinical malaria.

- A negative correlation was observed between $\hat{F}_{WS}$ estimates and age at infection and higher prevalence of infections with low genetic diversity was found in older children and adults.

- In chapter 5, I profiled the prevalence of mutations in *pfdhfr*, *pfdhps*, and *pfcrt*, and identified patterns of inferred haplotypes imposed by the sequential acquisition of resistant alleles in these genes.

- A high prevalence of *pfdhfr* (ie pyrimethamine resistance) and *pfdhps* (ie sulfadoxine resistance) mutations within the parasite population was observed.

- Incipient, but, wide spread sulfadoxine-pyrimethamine resistance across all geographical locations in the KNDs as evidenced by multiple mutations in both genes

- I observed 23% prevalence of the *pfcrt* 76T basal mutation associated with CQ resistance, evidence that suggests the re-emergence of CQ sensitive parasites (50% reduction from pre-ACT estimates) five years after replacement of CQ with ACTs as first-line treatment for uncomplicated malaria.

- In chapter 6, the Sequenom iPLEX platform was utilized to measure (or type) *P. falciparum* polymorphisms in field isolates.

- The strategy of preparing parasite mixes proved adequate for generating mass spectrometry peak height intensities that were used to as positive controls for determining *P. falciparum* genotypes.

- A haploid genotype-calling algorithm was developed for using control cluster plots from parasite mixes to assign *P. falciparum* genotypes and measure ratios for heterozygous infections

- The genotype concordance rate between Ilumina short-reads genotypes and Sequenom was 90% for heterozygous call based on >5 reads.

- Host-parasite genetic interactions was explored in chapter 7. The sickle-cell trait (HbAS) was observed to confer protection against both severe and

mild malaria disease in the northern Ghana.

- An IL-4 promoter polymorphism (rs2243250), IL-10, IL-22 (variant rs2227491) and variant rs1012356 were associated with severe malaria and mild malaria.

- The IL-17 receptor C was also associated with increased risk of SM. The derived-allele homozygous state (AA) of this gene was associated with 44% increased the risk of SM.

- There was evidence of an increased risk of clinical malaria when patients acquired the *pfdhfr* double and triple resistant alleles.

- The Cochran-Armitage test for a linear trend in the proportions of *pfdhfr* resistant alleles between mild malaria cases and asymptomatic controls showed a significant increased risk of mild malaria as the number of resistant alleles per parasite increased from 1 to 3 ($\chi^2 = 5.381$, P = 0.020).

- The *pfdhfr* functional double (59R + 108N) and triple (51I + 59R + 108N) resistant alleles known to confer higher levels of sulfadoxine-pyrimethamine resistance was found to generally confer an increased risk of clinical malaria (either severe or mild malaria).

- An interaction between *pfdhfr* resistant alleles and the sickle-cell locus indicates

  – that there is no significant difference in the risk of severe malaria for children with normal haemoglobin who are infected with parasites with *pfdhfr* sensitive or resistant alleles,

- that children with the sickle-cell trait (HbAS) are protected against severe malaria when infected with parasites with *pfdhfr* sensitive alleles as compared to children who have normal haemoglobin (HbAA) and are infected with parasites with *pfdhfr* sensitive alleles,
- that when children with the sickle cell trait are infected with parasites that carry these resistant alleles, the protective effect of the sickle trait is lost and the risk of severe malaria is no longer different between children with normal haemoglobin and those with the sickle-cell trait who are infected with parasite carrying *dhfr* resistant alleles.

8.3 Challenges

My research work had a range of challenges; foremost was the ethics, engaging rural communities with the science of pathogen genomics in setting up the community-based sampling framework. By drawing on my previous experience, I put together a simplified message and working through the traditional leaders who are gate-keepers in these communities, I was able to conveyed some understanding of malaria parasite genomics and the need to take whole blood for DNA extraction. Further complicating issues was the conduct of my research within the principles of open data access, which was challenging to convey to this largely rural population with little knowledge of computers and a language that lacks scientific vocabulary. However, owing to the goodwill and trust that the Navrongo Health Research Centre, where my work was coordinated, had developed over the years with these communities, I was able to legitimize this project

by bringing on board the community leadership of Chiefs and their Elders and the District Health Management teams of the KNDs. Thus, paving the way for the documentation of individual informed consent at enrollment.

Beyond the field data collection, working with mixed (multi-genome) infections, a common feature of *P. falciparum* malaria in areas of high disease endemicity[129,130], presented a formidable challenge to high-throughput genotyping. In particular, Sequenom® iPLEX genotyping, which was used for this work could not detect the occurrence of low frequency allele in any given parasite sample ($<$ 10%). In such cases, the frequency of the variant with the lower detection limit tended to be overestimated. Hence it remains challenging to accurately measure the relative proportions of bi-allelic SNP in multi-genome infections that border homozygosity. In addition, one fundamental challenge is obtaining parasite clones with known alleles that adequately serve as internal controls that capture the global allelic variation and could be used to estimate relative proportions of clones in clinical samples (its like building a HapMap type of database for *Plasmodium*).

Direct sampling of *P. falciparum* from whole blood taken from infected patients without a culture step for NGS has been achieved. The challenge, however, is to reduce the current 2-5ml of blood required to about 1ml of blood either conveniently spotted on filter paper or an alternative device that will make for easy field application.

In terms of sequencing, the last couple of years have seen revolutionary achievements in the scale and efficiency of next-generation sequencing. In spite of this

achievements, sequencing the extremely base biased genome of *P. falciparum* poses a great challenge[171]. The *P. falciparum* genome is AT-rich, coding regions have a mean AT content of more than 57%, with up to 100% in intergenic and non-coding regions[71]. Consequently, there is uneven coverage of the genome, with coding regions currently sequenced at a greater read depth than the other regions of the genome. Thus, restricting variant discovery and subsequent analysis to under 50% of the genome[121]. Additionally, within coding regions of the genome, highly polymorphic genes such as the *var* gene clusters, are currently very challenging to accurately map to the reference genome and derive sequence variation[121]. These challenges severely limit the amount of genomic information that could be gleaned to study patterns of gene flow at local scales between villages in endemic communities.

In summary, despite these challenges, this work demonstrates the feasibility of purifying parasite DNA from clinical samples for deep sequencing in rural endemic populations. The literature contains multiple-lines of evidence suggesting high within-host diversity, random mating and increased out-crossing in areas of high malaria transmission[7,11,121,142,190]. Consistent with these studies, this work utilized deep sequencing data from natural infections to demonstrate high within-host diversity of *P. falciparum* and a high risk of out-crossing in a well-characterized area with intense malaria transmission. The lack of population structure at local and regional levels reflects the general panmixis in parasite populations in West Africa, which is in stack contrast with parasite populations in areas of Cambodia and Thailand (traditional associated with emergence of antimalarial drug resistance) that show a greater degree of mating isolation[121]

(Olivo Miotto *et al.,* 2012 manuscript submitted).

8.3.1 Future

There is emerging consensus on the global population structure of *P. falciparum*, but, the fine-scale population structure and gene flow in endemic populations requires further characterization. Current informatic and data analysis tools are inadequate in addressing in this problem, particularly in areas of intense malaria transmission where multi-genome infection is common and defining haplotypes for tracking gene flow complicated and highly challenging. This study has highlighted the occurrence of small group of parasites that appear to be under diversification from the general population in northern Ghana. The question of the occurrence of parasite subpopulation in the fringes of the general population needs further investigation. Moreover, the ongoing artemisinin selection events in Cambodia and Thailand warrants monitoring of parasite populations in endemic countries, but, is real-time monitoring of these populations feasible at the moment? Yes, but not with deep sequencing platforms as yet. High-throughput genotyping platforms such as the Sequenom iPLEX platform may be more suitable in this regard on a time-scale of a week. However, I believe the second generation sequencing platforms will continue to play a leading role in variant discovery and the tracking of global parasite migration. Undoubtedly, a lot of insight has been gained from the deep sequencing of global parasite populations, but, many questions remain. How could deep sequencing data provide the much needed insight into host-parasite genetic interactions?

Moreover, could it provide a deeper insight into the role of rare human and par-

asite variants in the epidemiology of severe malaria?.

More eminent, though, is antimalarial drug resistance that remains a major threat to malaria control and elimination worldwide. The current absence of a causal molecular marker for the artemisinin resistance phenotype is of great concern. What role could deep sequencing of contemporary parasite populations play in identifying this causal variant?

These questions in my view will continue to stimulate malaria genomics in the near future. By and large, there are positive signs that deep sequencing data when combined with good phenotypic data could prove vital in identifying parasite subpopulations that harbour a fitness phenotype such as drug resistance. Such data are beginning to impact positively on the tracking of the genetic trait responsible for delayed artemisinin clearance in Cambodia and Thailand (Olivo Miotto et al., 2012 manuscript submitted for publication).

How could the sequence data from field isolates support the evaluation of candidate vaccine trials? Recent reports of reduced efficacy of the leading malaria vaccine candidate RTS,S/AS01 among infants after 12 months of follow up[221], reaffirms the need for population specific sequence data to inform vaccine evaluation. Population specific sequence analysis of the diversity of candidate vaccine epitopes, such as the CSP epitopes used to formulate RTS,S/AS01 could shed some light on the potential reasons for vaccine failure or reduced protection.

In this thesis, I have considered both parasite and human genomes in a specific epidemiological setting. However, with the ongoing sequencing of the mosquito (*Anopheles spp*) genome high resolution studies involving all three genomes in the near future are feasible. At present, the limiting factor is not the technol-

ogy to generate the genome-wide data, particularly as sequencing costs decline. The challenge resides in the development of new improved informatic tools and computing power to handle the large volumes of data generated especially for host-parasite genome interactions (human and/or mosquito hosts). Currently when we purify parasitized RBCs we discard the leukocytes, and we are obliged to discard/mask any human reads when analysing parasite DNA sequences. This practice can be adapted for host-parasite interaction studies by always isolating leukocytes, to allow human DNA recovery, before purifying the RBCs for parasite DNA. This will require work to overcome current ethical challenges to include human sequence data, in order to maximize sample and data utilization for deepening our understanding of the malaria problem. Similarly improving DNA recovery from mosquitoes will also allow us to survey parasite populations in this host. The future therefore is really promising for studying genomic interactions between the parasite, human and the mosquito.

Appendix A

Reagents and Equipment

This appendix contains the reagents and equipment used for the thesis. Tables A.1 and A.2 list the reagents and equipment respectively.

Reagent	**Product code**	**Manufacturer**
QIAamp DNA Blood Midi Kit	51185	Qiagen
QIAamp DNA Blood Maxi Kit	51194	Qiagen
Dulbecco's Phosphate Buffered Saline	D8537-1L	Sigma
Lymphoprep	1114545	Axis-Shield
CF11 cellulose powder	4021-50	Whatman
99.7-100% ethanol	3221	Sigma
Giemsa	G4507	Sigma-Aldrich
Agarose	A6929	Sigma
Phenol red	P4758	Sigma
dNTPs	Bio39028	Bioline
$MgCl_2$	Bio21040	Bioline
Primers	-	Metabion
PCR Buffer	Bio21040	Bioline
Biotaq DNA polymerase	Bio21040	Bioline
Ethidium bromide 10ml	E1510	Sigma

Table A.1: Reagents

Equipment	Product code	Manufacturer
2ml skirted screw-cap tubes	72.694.007	Greiner Bio-one
50ml Falcon tubes	210261	Greiner Bio-one
15ml Falcon tubes	188261	Greiner Bio-one
25µL Plasmodipur filters	8011	EuroProxima
10ml Syringe, no needle		GHS supply chain
blood slides	S8400	Sigma
Paracheck	30302025	Orchid Biomedical Systems
10ml Serological pipette	16.1254.001	Sigma
50ml centrifuge tubes	227261	Bioline
50ml centrifuge tubes	227261	Bioline
10ml EDTA vacutainer	366401	Becton-Dickinson
Benchtop centrifuge	-	Beckman Instruments
15ml centrifuge tubes	188261	Sigma
MJ tetrad thermocycler	-	MJ Research
Thermofast 96-well plates	AB-0980	ABgene
Thermofast 384-well plates	TF-0384	ABgene
PCR foil seal	AB-0626	ABgene
PCR film seal	AB-0558	ABgene
Microseal A film	MSA5001	Bio-Rad
Gilson pipette P1000	FA10006M	Gilson
Gilson pipette P200	FA10005M	Gilson
Gilson pipette P20	F144058P	Gilson
Gilson pipette P10	F144055P	Gilson
1000µL tips	1111-2831	USA Scientific
1-200µL filter tips	S1120-8810	USA Scientific
10µL tips	SS-L10	Bioclean
Gel tray	-	Biorad
Gel combs	-	Biorad
Gel tanks	-	Biorad

Table A.2: Equipment

Appendix B

Abreviations

SM: $\longrightarrow$ severe malaria

MM: $\longrightarrow$ mild malaria

CM: $\longrightarrow$ cerebral malaria

SMA:$\longrightarrow$ severe malarial anaemia

SSA:$\longrightarrow$ Sub-Saharan Africa

MOI:$\longrightarrow$ multiplicity of infection

EIR:$\longrightarrow$ entomological inoculation rate

ACT:$\longrightarrow$ artemisinin combination therapy

MSP:$\longrightarrow$ merozoite surface protein

GLURP:$\longrightarrow$ glutamate rich protein

CQ: $\longrightarrow$ cholroquine

SP:$\longrightarrow$ sulfadoxine-pyrimethamine

MFQ:$\longrightarrow$ mefloquine

QN: $\longrightarrow$ quinine

P (pyr) : $\longrightarrow$ pyrimethamine

CY: $\longrightarrow$ cycloguanil

WHO:$\longrightarrow$ World Health Organizatio

TNF:$\longrightarrow$ tumour necrosis factor

HBB:$\longrightarrow$ β -globulin gene

CR: $\longrightarrow$ complement receptor

G6PD:$\longrightarrow$ glucose-6-phosphate dehydrogenase deficiency gene

RBC: $\longrightarrow$ red blood cell

pRBC:$\longrightarrow$ parasitized red blood cell

IRBC:$\longrightarrow$ infected red blood cell

IL: $\longrightarrow$ interleukin

TGF:$\longrightarrow$ transforming growth factor

IEs:$\longrightarrow$ infected erythrocytes

VSA:$\longrightarrow$ variant surface antigen

bp: $\longrightarrow$ base pairs

kb: $\longrightarrow$ kilobase

HLA:$\longrightarrow$ Human leukocyte antigen

LTA:$\longrightarrow$ lymphotoxin alpha

LD: $\longrightarrow$ Linkage disequilibrium

NGS:$\longrightarrow$ next-generation sequencing

MALDI-TOF:$\longrightarrow$ Matrix-Assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometary

SNP:$\longrightarrow$ single nucleotide polymorphism

PCR:$\longrightarrow$ polymerase chain reaction

QRT-PCR:$\longrightarrow$ quantitative real-time PCR

$MgCL_2$:$\longrightarrow$ magesium chloride

PEP:⟶ Primer Extension Pre-amplification

GM: ⟶ geometric mean

WBCs:⟶ white blood cells

MAF:⟶ minor allele frequency

Appendix C

Calculation of parasite to-human DNA proportions

Calculation of parasite to-human DNA proportions to achieve the equivalent of 1% parasitaemia.

Assumptions

1% parasitized red blood cells (pRBCs)

No anaemia and 5×10^9 RBC ml^{-1} and 6×10^6WBC ml^{-1}

1 human genome copy (haploid) = 3.33pg (using genome size of 3.25 giga-base pairs)

1 human genome (diploid) = 2 copies = 6.67pg

1 falciparum genome = 1 copy (haploid) = 25.6fg (using genome size of 25 mega-base pairs)

Calculation

$\approx$ 40µg HuDNA (1.2×10^7 copies [ie 6×10^6 genomes]) per ml of blood.

1.28 µg parasite DNA (5×10^7 genomes) for a 1% parasitaemia per ml of blood

Approximate Genome ratio at 1% parasitaemia: 4 parasite: 1 human

Approximate DNA mass ratio at 1% parasitaemia: 1 parasite: 31 human

Appendix D

Calculation of threshold for statistical significance

	Gene	No. SNPs	Genetic model	Comparisons Phenotypes	Per Gene	Per SNP
1	IL10	3	5	8	120	40
2	IL17RE	1	5	8	40	40
3	IL17RD	1	5	8	40	40
4	IL4	1	5	8	40	40
5	IL13	1	5	8	40	40
6	IL1A	1	5	8	40	40
7	IL1B	1	5	8	40	40
8	IL20RA	1	5	8	40	40
9	IL22	5	5	8	200	40
10	IL4R	1	5	8	40	40
11	CR1	2	5	8	80	40
12	TRIM5	1	5	8	40	40
13	TLR4	2	5	8	80	40
14	TLR1	2	5	8	80	40
15	TLR6	2	5	8	80	40
16	TLR9	1	5	8	40	40
17	LTA	2	5	8	80	40
18	TNF	5	5	8	200	40
19	CD36	2	5	8	80	40
20	ABO	2	5	8	80	40
21	HBB	2	5	8	80	40
22	SPTB	1	5	8	40	40
23	NOS2	4	5	8	160	40
24	EMR1	2	5	8	80	40
25	ICAM1	2	5	8	80	40
26	CD4OLG	2	5	8	80	40
27	G6PD	2	5	8	80	40
	Total	52				1080

(Genetic models: Additive, Dominant, Recessive, heterozygous and General)
(Phenotypic comparisons: SM vs MM, SM vs HC, MM vs HC, SM vs HC,para, SM vs HC, nonpara, MM vs HC, para, MM vs HC, nonpara, HC, para vs HC, nonpara)

Appendix E

Full set of 52 human SNPs analyzed in chapter 7.

Table E.1: 52 human SNPs analyzed in chapter 7

Number	Polymorphism	Gene	Chromosome	Coordinate	MAF
1	hCD36_G1439C	CD36	7	80300449	0.01
2	rs1012356	IL22	12	68644618	0.45
3	rs10775349	ADCY9	16	4079823	0.13
4	rs1126535	CD40LG	X	135730555	0.12
5	rs1143634	IL1B	2	113590390	0.12
6	rs1555498	IL20RA	6	137325847	0.41
7	rs17047660	CR1	1	207782889	0.27
8	rs17047661	CR1	1	207782889	0.24
9	rs17140229	CFTR	7	117230283	0.45

Continued on next page

Table E.1 – continued from previous page

Number	Polymorphism	Gene	Chromosome	Coordinate	MAF
10	rs17561	IL1A	2	113537223	0.17
11	rs1799964	TNF	6	31542308	0.15
12	rs1800482	NOS2	17	26128509	0.12
13	rs1800629	TNF	6	31543031	0.12
14	rs1800890	IL10	1	206949365	0.20
15	rs1800896	IL10	1	206946897	0.29
16	rs1801033	C6	5	41199959	0.45
17	rs1803632	GBP2	1	89582690	0.45
18	rs1805015	IL4R	16	27374180	0.41
19	rs187084	TLR9	3	52261031	0.31
20	rs2075820	NOD1	7	30492237	0.37
21	rs2227478	IL22	12	68648622	0.33
22	rs2227485	IL22	12	68647713	0.49
23	rs2227491	IL22	12	68646521	0.28
24	rs2227507	IL22	12	68642647	0.03
25	rs2230739	ADCY9	16	4033436	0.11
26	rs2239704	LTA	6	31540141	0.35
27	rs2242665	CTL4	6	31839309	0.29
28	rs2243250	IL4	5	132009154	0.24
29	rs2297518	NOS2	17	26096597	0.07
30	rs2535611	ADORA2B	17	15861332	0.08

Continued on next page

Table E.1 – continued from previous page

Number	Polymorphism	Gene	Chromosome	Coordinate	MAF
31	rs2706384	IRF1	5	131826880	0.45
32	rs3024500	IL10	1	206940831	0.43
33	rs3092945	CD40LG	X	135729609	0.38
34	rs3093662	TNF	6	31544189	0.05
35	rs3211938	CD36	7	80300449	0.14
36	rs334	HBB	11	5248232	0.04
37	rs33950507	HBB	11	5248173	0.01
38	rs361525	TNF	6	31543101	0.02
39	rs373533	EMR1	19	6919624	0.45
40	rs461645	EMR1	19	6919753	0.45
41	rs4986790	TLR4	9	120475302	0.10
42	rs4986791	TLR4	9	120475602	0.01
43	rs542998	RTN3	11	63487386	0.40
44	rs5498	ICAM1	19	10395683	0.16
45	rs6780995	IL17RD	3	57138419	0.47
46	rs708567	IL17RE	3	9960070	0.47
47	rs8078340	NOS2	17	26129212	0.22
48	rs8176719	ABO	9	136132909	0.37
49	rs8176746	ABO	9	136131322	0.24
50	rs8386	GNAS	20	57485812	0.21
51	rs909253	LTA	6	31540313	0.43

Continued on next page

Table E.1 – continued from previous page

Number	Polymorphism	Gene	Chromosome	Coordinate	MAF
52	rs9282799	NOS2	17	26128728	0.08

Appendix F

Analysis of samples with mixed genotypes

Any fully resistant alleles vs fully sensitive alleles						
Allele type	**severe**	**Mild**	**severe**	**Asym**	**Mild**	**Asym**
Sensitive	125	206	125	92	206	92
Resistant	327	649	327	382	649	382
	OR(95% CI)	**P-value**	**OR**(95% CI)	**P-value**	**OR**(95% CI)	**P-value**
Fisher test	1.20 (0.92-1.57)	0.161	1.59 (1.15-2.19)	0.0032	1.32 (0.99-1.76)	0.054

Table F.1: **Disease association analysis using the Fisher exact test.** Fully sensitive - no non-reference alleles across all loci; any fully resistant (homozygous/heterozygous) - presence of any 1, 2 or 3 non-reference allele.

A	**Test of resistant alleles combinations**					
A						
Allele count	**severe**	**Mild**	**severe**	**Asym**	**Mild**	**Asym**
0	125	206	125	92	206	92
1	6	16	6	33	16	33
2	101	197	101	110	197	110
3	220	436	220	239	436	239
Total count	327	649	327	382	649	382
B						
Linear trend	**χ^2-value**	**P-value**	**χ^2-value**	**P-value**	**χ^2-value**	**P-value**
χ^2_{Std} test	0.40	0.817	15.69	0.00039	20.26	3.98e-05
C-A χ^2_{Trend}	0.04	0.837	6.65	0.0099	8.39	0.0037
χ^2_{diff}	0.36	0.547	9.04	0.0026	11.87	0.00057

Table F.2: **Counts of multiple resistant allele (homozygous and heterozygous) combinations per sample across *pfdhfr* and linear trend analysis.** In Table A, Asym-Asymptomatic, 1-samples with N51I or C59R or S108N resistant alleles, 2-samples with 51I + 108N or 59R + 108N resistant alleles and 3-samples with N51I + C59R + S108N resistant alleles. Table B shows the standard χ^2 test (χ^2_{Std}) and Cochrane-Armitage (C-A) test for a linear trend. When the difference between the χ^2_{trend} and χ^2_{Std} is not significant then the linear trend explains all of the relationship.

References

1. LJ Williams Abdi H. Principal component analysis. *Wiley interdisciplinary reviews:Computational Statistics*, 2:433–459, 2010. 97
2. Uhlemann AC, Yuthavong Y, and Fidock D. *Mechanisms of antimalarial drug action and resistance. In Molecular approaches to malaria.* ASM Press. Washington DC, USA, 229âĂŞ261., 2005. 17, 19, 20
3. A. Agarwal, A. Guindo, Y. Cissoko, J. G. Taylor, D. Coulibaly, A. KonÃĺ, K. Kayentao, A. Djimde, C. V. Plowe, O. Doumbo, T. E. Wellems, and D. Diallo. Hemoglobin c associated with protection from severe malaria in the dogon of mali, a west african population with a low prevalence of hemoglobin s. *Blood*, 96(7):2358–2363, Oct 2000. 13, 14
4. Michael Aidoo, Dianne J Terlouw, Margarette S Kolczak, Peter D McElroy, Feiko O ter Kuile, Simon Kariuki, Bernard L Nahlen, Altaf A Lal, and Venkatachalam Udhayakumar. Protective effects of the sickle cell gene against malaria morbidity and mortality. *Lancet*, 359(9314):1311–1312, Apr 2002. doi: 10.1016/S0140-6736(02)08273-9. URL http://dx.doi.org/10.1016/S0140-6736(02)08273-9. 13, 206
5. F. al Yaman, B. Genton, J. C. Reeder, R. F. Anders, T. Smith, and M. P. Alpers. Reduced risk of clinical malaria in children infected with multiple

clones of plasmodium falciparum in a highly endemic area: a prospective community study. *Trans R Soc Trop Med Hyg*, 91(5):602–605, 1997. 117

6. P. L. Alonso, S. W. Lindsay, J. R. Armstrong Schellenberg, P. Gomez, A. G. Hill, P. H. David, G. Fegan, K. Cham, and B. M. Greenwood. A malaria control trial using insecticide-treated bed nets and targeted chemoprophylaxis in a rural area of the gambia, west africa. 2. mortality and morbidity from malaria in the study area. *Trans R Soc Trop Med Hyg*, 87 Suppl 2: 13–17, Jun 1993. 8, 9
7. T. J. Anderson, B. Haubold, J. T. Williams, J. G. Estrada-Franco, L. Richardson, R. Mollinedo, M. Bockarie, J. Mokili, S. Mharakurwa, N. French, J. Whitworth, I. D. Velez, A. H. Brockman, F. Nosten, M. U. Ferreira, and K. P. Day. Microsatellite markers reveal a spectrum of population structures in the malaria parasite plasmodium falciparum. *Mol Biol Evol*, 17(10):1467–1482, Oct 2000. 14, 15, 80, 107, 114, 117, 218, 219, 233
8. T. J. Anderson, X. Z. Su, A. Roddam, and K. P. Day. Complex mutations in a high proportion of microsatellite loci from the protozoan parasite plasmodium falciparum. *Mol Ecol*, 9(10):1599–1608, Oct 2000. 14
9. Tim J C Anderson, Shalini Nair, Standwell Nkhoma, Jeff T Williams, Mallika Imwong, Poravuth Yi, Duong Socheat, Debashish Das, Kesinee Chotivanich, Nicholas P J Day, Nicholas J White, and Arjen M Dondorp. High heritability of malaria parasite clearance rate indicates a genetic basis for artemisinin resistance in western cambodia. *J Infect Dis*, 201(9):1326–1330, May 2010. doi: 10.1086/651562. URL `http://dx.doi.org/10.1086/651562`. 11, 20, 29
10. Wilhelm J. Ansorge. Next-generation dna sequencing techniques. *New*

Biotechnology, 25:195 – 203, 2009. 21

11. Thomas G Anthony, David J Conway, Janet Cox-Singh, Asmad Matusop, Shanmuga Ratnam, Sunita Shamsul, and Balbir Singh. Fragmented population structure of plasmodium falciparum in a region of declining endemicity. *J Infect Dis*, 191(9):1558–1564, May 2005. doi: 10.1086/429338. URL http://dx.doi.org/10.1086/429338. 15, 107, 114, 117, 218, 233

12. Maxwell Appawu, Seth Owusu-Agyei, Samuel Dadzie, Victor Asoala, Francis Anto, Kwadwo Koram, William Rogers, Francis Nkrumah, Stephen L Hoffman, and David J Fryauff. Malaria transmission dynamics at a site in northern ghana proposed for testing malaria vaccines. *Trop Med Int Health*, 9(1):164–170, Jan 2004. 33, 74, 82, 107, 116, 219

13. D. Arnot. Unstable malaria in sudan: the influence of the dry season. clone multiplicity of plasmodium falciparum infections in individuals exposed to variable levels of disease transmission. *Trans R Soc Trop Med Hyg*, 92(6): 580–585, 1998. 6, 219

14. Sarah Auburn, Susana Campino, Taane G Clark, Abdoulaye A Djimde, Issaka Zongo, Robert Pinches, Magnus Manske, Valentina Mangano, Daniel Alcock, Elisa Anastasi, Gareth Maslen, Bronwyn Macinnis, Kirk Rockett, David Modiano, Christopher I Newbold, Ogobara K Doumbo, Jean Bosco OuÃľdraogo, and Dominic P Kwiatkowski. An effective method to purify plasmodium falciparum dna directly from clinical blood samples for whole genome high-throughput sequencing. *PLoS One*, 6(7):e22213, 2011. doi: 10.1371/journal.pone.0022213. URL http://dx.doi.org/10.1371/journal.pone.0022213. 52, 76

15. Sarah Auburn, Susana Campino, Olivo Miotto, Abdoulaye A Djimde, Issaka

Zongo, Magnus Manske, Gareth Maslen, Valentina Mangano, Daniel Alcock, Bronwyn Macinnis, Kirk A Rockett, Taane G Clark, Ogobara K Doumbo, Jean Bosco OuÃľdraogo, and Dominic P Kwiatkowski. Characterization of within-host plasmodium falciparum diversity using next-generation sequence data. *PLoS One*, 7(2):e32891, 2012. doi: 10.1371/journal.pone.0032891. URL http://dx.doi.org/10.1371/journal.pone.0032891. 6, 85, 114, 115, 118, 219

16. Kodjo Ayi, Franco Turrini, Antonio Piga, and Paolo Arese. Enhanced phagocytosis of ring-parasitized mutant erythrocytes: a common mechanism that may explain protection against falciparum malaria in sickle trait and beta-thalassemia trait. *Blood*, 104(10):3364–3371, Nov 2004. doi: 10.1182/blood-2003-11-3820. URL http://dx.doi.org/10.1182/blood-2003-11-3820. 13

17. H. A. Babiker. Unstable malaria in sudan: the influence of the dry season. plasmodium falciparum population in the unstable malaria area of eastern sudan is stable and genetically complex. *Trans R Soc Trop Med Hyg*, 92(6): 585–589, 1998. 6, 220

18. H. A. Babiker, L. C. Ranford-Cartwright, D. Currie, J. D. Charlwood, P. Billingsley, T. Teuscher, and D. Walliker. Random mating in a natural population of the malaria parasite plasmodium falciparum. *Parasitology*, 109 (Pt 4):413–421, Nov 1994. 219

19. H. A. Babiker, L. C. Ranford-Cartwright, and D. Walliker. Genetic structure and dynamics of plasmodium falciparum infections in the kilombero region of tanzania. *Trans R Soc Trop Med Hyg*, 93 Suppl 1:11–14, Feb 1999. 6, 220

20. H. A. Babiker, S. J. Pringle, A. Abdel-Muhsin, M. Mackinnon, P. Hunt,

and D. Walliker. High-level chloroquine resistance in sudanese isolates of plasmodium falciparum is associated with mutations in the chloroquine resistance transporter gene pfcrt and the multidrug resistance gene pfmdr1. *J Infect Dis*, 183(10):1535–1538, May 2001. doi: 10.1086/320195. URL http://dx.doi.org/10.1086/320195. 18

21. J. Kevin Baird, S. Owusu Agyei, Gregory C Utz, Kwadwo Koram, Mazie J Barcus, Trevor R Jones, David J Fryauff, Fred N Binka, Stephen L Hoffman, and Francis N Nkrumah. Seasonal malaria attack rates in infants and young children in northern ghana. *Am J Trop Med Hyg*, 66(3):280–286, Mar 2002. 33, 82

22. Karen I Barnes, David N Durrheim, Francesca Little, Amanda Jackson, Ushma Mehta, Elizabeth Allen, Sicelo S Dlamini, Joyce Tsoka, Barry Bredenkamp, D. Jotham Mthembu, Nicholas J White, and Brian L Sharp. Effect of artemether-lumefantrine policy and improved vector control on malaria burden in kwazulu-natal, south africa. *PLoS Med*, 2(11):e330, Nov 2005. doi: 10.1371/journal.pmed.0020330. URL http://dx.doi.org/10.1371/journal.pmed.0020330. 9

23. David R Bentley, Shankar Balasubramanian, Harold P Swerdlow, Geoffrey P Smith, John Milton, Clive G Brown, Kevin P Hall, Dirk J Evers, Colin L Barnes, Helen R Bignell, Jonathan M Boutell, Jason Bryant, Richard J Carter, R. Keira Cheetham, Anthony J Cox, Darren J Ellis, Michael R Flatbush, Niall A Gormley, Sean J Humphray, Leslie J Irving, Mirian S Karbelashvili, Scott M Kirk, Heng Li, Xiaohai Liu, Klaus S Maisinger, Lisa J Murray, Bojan Obradovic, Tobias Ost, Michael L Parkinson, Mark R Pratt, Isabelle M J Rasolonjatovo, Mark T Reed, Roberto Rigatti, Chiara

Rodighiero, Mark T Ross, Andrea Sabot, Subramanian V Sankar, Aylwyn Scally, Gary P Schroth, Mark E Smith, Vincent P Smith, Anastassia Spiridou, Peta E Torrance, Svilen S Tzonev, Eric H Vermaas, Klaudia Walter, Xiaolin Wu, Lu Zhang, Mohammed D Alam, Carole Anastasi, Ify C Aniebo, David M D Bailey, Iain R Bancarz, Saibal Banerjee, Selena G Barbour, Primo A Baybayan, Vincent A Benoit, Kevin F Benson, Claire Bevis, Phillip J Black, Asha Boodhun, Joe S Brennan, John A Bridgham, Rob C Brown, Andrew A Brown, Dale H Buermann, Abass A Bundu, James C Burrows, Nigel P Carter, Nestor Castillo, Maria Chiara E Catenazzi, Simon Chang, R. Neil Cooley, Natasha R Crake, Olubunmi O Dada, Konstantinos D Diakoumakos, Belen Dominguez-Fernandez, David J Earnshaw, Ugonna C Egbujor, David W Elmore, Sergey S Etchin, Mark R Ewan, Milan Fedurco, Louise J Fraser, Karin V Fuentes Fajardo, W. Scott Furey, David George, Kimberley J Gietzen, Colin P Goddard, George S Golda, Philip A Granieri, David E Green, David L Gustafson, Nancy F Hansen, Kevin Harnish, Christian D Haudenschild, Narinder I Heyer, Matthew M Hims, Johnny T Ho, Adrian M Horgan, Katya Hoschler, Steve Hurwitz, Denis V Ivanov, Maria Q Johnson, Terena James, T. A. Huw Jones, Gyoung-Dong Kang, Tzvetana H Kerelska, Alan D Kersey, Irina Khrebtukova, Alex P Kindwall, Zoya Kingsbury, Paula I Kokko-Gonzales, Anil Kumar, Marc A Laurent, Cynthia T Lawley, Sarah E Lee, Xavier Lee, Arnold K Liao, Jennifer A Loch, Mitch Lok, Shujun Luo, Radhika M Mammen, John W Martin, Patrick G McCauley, Paul McNitt, Parul Mehta, Keith W Moon, Joe W Mullens, Taksina Newington, Zemin Ning, Bee Ling Ng, Sonia M Novo, Michael J O'Neill, Mark A Osborne, Andrew Osnowski, Omead

Ostadan, Lambros L Paraschos, Lea Pickering, Andrew C Pike, Alger C Pike, D. Chris Pinkard, Daniel P Pliskin, Joe Podhasky, Victor J Quijano, Come Raczy, Vicki H Rae, Stephen R Rawlings, Ana Chiva Rodriguez, Phyllida M Roe, John Rogers, Maria C Rogert Bacigalupo, Nikolai Romanov, Anthony Romieu, Rithy K Roth, Natalie J Rourke, Silke T Ruediger, Eli Rusman, Raquel M Sanches-Kuiper, Martin R Schenker, Josefina M Seoane, Richard J Shaw, Mitch K Shiver, Steven W Short, Ning L Sizto, Johannes P Sluis, Melanie A Smith, Jean Ernest Sohna Sohna, Eric J Spence, Kim Stevens, Neil Sutton, Lukasz Szajkowski, Carolyn L Tregidgo, Gerardo Turcatti, Stephanie Vandevondele, Yuli Verhovsky, Selene M Virk, Suzanne Wakelin, Gregory C Walcott, Jingwen Wang, Graham J Worsley, Juying Yan, Ling Yau, Mike Zuerlein, Jane Rogers, James C Mullikin, Matthew E Hurles, Nick J McCooke, John S West, Frank L Oaks, Peter L Lundberg, David Klenerman, Richard Durbin, and Anthony J Smith. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218):53–59, Nov 2008. doi: 10.1038/nature07517. URL http://dx.doi.org/10.1038/nature07517. 15, 22, 23, 24, 85

24. F. N. Binka, A. Kubaje, M. Adjuik, L. A. Williams, C. Lengeler, G. H. Maude, G. E. Armah, B. Kajihara, J. H. Adiamah, and P. G. Smith. Impact of permethrin impregnated bednets on child mortality in kassena-nankana district, ghana: a randomized controlled trial. *Trop Med Int Health*, 1(2): 147–154, Apr 1996. 8, 9, 107

25. P. B. Bloland, E. M. Lackritz, P. N. Kazembe, J. B. Were, R. Steketee, and C. C. Campbell. Beyond chloroquine: implications of drug resistance for evaluating malaria therapy efficacy and treatment policy in africa. *J Infect*

Dis, 167(4):932–937, Apr 1993. 10

26. Silayuv E Bongfen, Patricia M Ntsama, Sandra Offner, Thomas Smith, Ingrid Felger, Marcel Tanner, Pedro Alonso, Issa Nebie, Jackeline F Romero, Olivier Silvie, Ralph Torgler, and Giampietro Corradin. The n-terminal domain of plasmodium falciparum circumsporozoite protein represents a target of protective immunity. *Vaccine*, 27(2):328–335, Jan 2009. doi: 10.1016/j.vaccine.2008.09.097. URL http://dx.doi.org/10.1016/j.vaccine.2008.09.097. xiii, 3, 4, 5, 7
27. P. Borst, W. Bitter, R. McCulloch, F. Van Leeuwen, and G. Rudenko. Antigenic variation in malaria. *Cell*, 82(1):1–4, Jul 1995. 16, 17
28. O. H. Branch, S. Takala, S. Kariuki, B. L. Nahlen, M. Kolczak, W. Hawley, and A. A. Lal. Plasmodium falciparum genotypes, low complexity of infection, and resistance to subsequent malaria in participants in the asembo bay cohort project. *Infect Immun*, 69(12):7783–7792, Dec 2001. 117
29. P. C. Bull, B. S. Lowe, M. Kortok, C. S. Molyneux, C. I. Newbold, and K. Marsh. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nat Med*, 4(3):358–360, Mar 1998. 16
30. Peter C Bull, Brett S Lowe, Ngumbao Kaleli, Fixton Njuga, Moses Kortok, Amanda Ross, Francis Ndungu, Robert W Snow, and Kevin Marsh. Plasmodium falciparum infections are associated with agglutinating antibodies to parasite-infected erythrocyte surface antigens among healthy kenyan children. *J Infect Dis*, 185(11):1688–1691, Jun 2002. doi: 10.1086/340420. URL http://dx.doi.org/10.1086/340420. 16
31. Gerardo Cabrera, Michel Cot, Florence Migot-Nabias, Peter G Kremsner,

Philippe Deloron, and Adrian J F Luty. The sickle cell trait is associated with enhanced immunoglobulin g antibody responses to plasmodium falciparum variant surface antigens. *J Infect Dis*, 191(10):1631–1638, May 2005. doi: 10.1086/429832. URL http://dx.doi.org/10.1086/429832. 13

32. C. C. Campbell, W. Chin, W. E. Collins, S. M. Teutsch, and D. M. Moss. Chloroquine-resistant plasmodium falciparum from east africa: cultivation and drug sensitivity of the tanzanian i/cdc strain from an american tourist. *Lancet*, 2(8153):1151–1154, Dec 1979. 10

33. Ian H Cheeseman, Becky A Miller, Shalini Nair, Standwell Nkhoma, Asako Tan, John C Tan, Salma Al Saai, Aung Pyae Phyo, Carit Ler Moo, Khin Maung Lwin, Rose McGready, Elizabeth Ashley, Mallika Imwong, Kasia Stepniewska, Poravuth Yi, Arjen M Dondorp, Mayfong Mayxay, Paul N Newton, Nicholas J White, FranÃğois Nosten, Michael T Ferdig, and Timothy J C Anderson. A major genome region underlying artemisinin resistance in malaria. *Science*, 336(6077):79–82, Apr 2012. doi: 10.1126/science.1215966. URL http://dx.doi.org/10.1126/science.1215966. 11, 20, 29, 81, 121

34. Ian A Clark, Alison C Budd, Lisa M Alleva, and William B Cowden. Human malarial disease: a consequence of inflammatory cytokine release. *Malar J*, 5:85, 2006. doi: 10.1186/1475-2875-5-85. URL http://dx.doi.org/10.1186/1475-2875-5-85. 207

35. D. J. Conway, D. R. Cavanagh, K. Tanabe, C. Roper, Z. S. Mikes, N. Sakihama, K. A. Bojang, A. M. Oduola, P. G. Kremsner, D. E. Arnot, B. M. Greenwood, and J. S. McBride. A principal target of human immunity to malaria identified by molecular population genetic and immunological

analyses. *Nat Med*, 6(6):689–692, Jun 2000. doi: 10.1038/76272. URL http://dx.doi.org/10.1038/76272. 80

36. D. J. Conway, C. Fanello, J. M. Lloyd, B. M. Al-Joubori, A. H. Baloch, S. D. Somanath, C. Roper, A. M. Oduola, B. Mulder, M. M. Povoa, B. Singh, and A. W. Thomas. Origin of plasmodium falciparum malaria is traced by mitochondrial dna. *Mol Biochem Parasitol*, 111(1):163–171, Nov 2000. 14, 15, 80

37. Kevin N Couper, Daniel G Blount, and Eleanor M Riley. Il-10: the master regulator of immunity to infection. *J Immunol*, 180(9):5771–5777, May 2008. 12, 207

38. A. Craig, I. Hastings, A. Pain, and D. J. Roberts. Genetics and malaria–more questions than answers. *Trends Parasitol*, 17(2):55–56, Feb 2001. 11, 30

39. Peter D Crompton, Matthew A Kayala, Boubacar Traore, Kassoum Kayentao, Aissata Ongoiba, Greta E Weiss, Douglas M Molina, Chad R Burk, Michael Waisberg, Algis Jasinskas, Xiaolin Tan, Safiatou Doumbo, Didier Doumtabe, Younoussou Kone, David L Narum, Xiaowu Liang, Ogobara K Doumbo, Louis H Miller, Denise L Doolan, Pierre Baldi, Philip L Felgner, and Susan K Pierce. A prospective analysis of the ab response to plasmodium falciparum before and after a malaria season by protein microarray. *Proc Natl Acad Sci U S A*, 107(15):6958–6963, Apr 2010. doi: 10.1073/pnas.1001323107. URL http://dx.doi.org/10.1073/pnas.1001323107. 6

40. U. D'Alessandro and H. ButtiÃńns. History and importance of antimalarial drug resistance. *Trop Med Int Health*, 6(11):845–848, Nov 2001. 10, 11

41. N. P. Day, T. T. Hien, T. Schollaardt, P. P. Loc, L. V. Chuong, T. T. Chau, N. T. Mai, N. H. Phu, D. X. Sinh, N. J. White, and M. Ho. The prognostic and pathophysiologic role of pro- and antiinflammatory cytokines in severe malaria. *J Infect Dis*, 180(4):1288–1297, Oct 1999. doi: 10.1086/315016. URL http://dx.doi.org/10.1086/315016. 207

42. Meghna Desai, Feiko O ter Kuile, FranÃğois Nosten, Rose McGready, Kwame Asamoa, Bernard Brabin, and Robert D Newman. Epidemiology and burden of malaria in pregnancy. *Lancet Infect Dis*, 7(2):93–104, Feb 2007. doi: 10.1016/S1473-3099(07)70021-X. URL http://dx.doi.org/10.1016/S1473-3099(07)70021-X. 3

43. Neekesh V Dharia, Amar Bir Singh Sidhu, MarÃŋa BelÃĺn Cassera, Scott J Westenberger, Selina Er Bopp, Rich T Eastman, David Plouffe, Serge Batalov, Daniel J Park, Sarah K Volkman, Dyann F Wirth, Yingyao Zhou, David A Fidock, and Elizabeth A Winzeler. Use of high-density tiling microarrays to identify mutations globally and elucidate mechanisms of drug resistance in plasmodium falciparum. *Genome Biol*, 10(2):R21, 2009. doi: 10.1186/gb-2009-10-2-r21. URL http://dx.doi.org/10.1186/gb-2009-10-2-r21. 14, 17

44. A. DjimdÃĺ, O. K. Doumbo, J. F. Cortese, K. Kayentao, S. Doumbo, Y. DiourtÃĺ, A. Dicko, X. Z. Su, T. Nomura, D. A. Fidock, T. E. Wellems, and C. V. Plowe. A molecular marker for chloroquine-resistant falciparum malaria. *N Engl J Med*, 344(4):257–263, Jan 2001. doi: 10.1056/NEJM200101253440403. URL http://dx.doi.org/10.1056/NEJM200101253440403. 10, 18, 116, 121, 134

45. Abdoulaye A DjimdÃĺ, Ogobara K Doumbo, Ousmane Traore, Ando B

Guindo, Kassoum Kayentao, Yacouba Diourte, Safiatou Niare-Doumbo, Drissa Coulibaly, Abdoulaye K Kone, Yacouba Cissoko, Mamadou Tekete, Bakary Fofana, Alassane Dicko, Dapa A Diallo, Thomas E Wellems, Dominic Kwiatkowski, and Christopher V Plowe. Clearance of drug-resistant parasites as a model for protective immunity in plasmodium falciparum malaria. *Am J Trop Med Hyg*, 69(5):558–563, Nov 2003. 134, 172, 209

46. D. Dodoo, F. M. Omer, J. Todd, B. D. Akanmori, K. A. Koram, and E. M. Riley. Absolute levels and ratios of proinflammatory and anti-inflammatory cytokine production in vitro predict clinical immunity to plasmodium falciparum malaria. *J Infect Dis*, 185(7):971–979, Apr 2002. doi: 10.1086/339408. URL http://dx.doi.org/10.1086/339408. 207

47. Arjen M Dondorp, FranÃğois Nosten, Poravuth Yi, Debashish Das, Aung Phae Phyo, Joel Tarning, Khin Maung Lwin, Frederic Ariey, Warunee Hanpithakpong, Sue J Lee, Pascal Ringwald, Kamolrat Silamut, Mallika Imwong, Kesinee Chotivanich, Pharath Lim, Trent Herdman, Sen Sam An, Shunmay Yeung, Pratap Singhasivanon, Nicholas P J Day, Niklas Lindegardh, Duong Socheat, and Nicholas J White. Artemisinin resistance in plasmodium falciparum malaria. *N Engl J Med*, 361(5):455–467, Jul 2009. doi: 10.1056/NEJMoa0808859. URL http://dx.doi.org/10.1056/NEJMoa0808859. 11, 20, 29, 121

48. Devin Dressman, Hai Yan, Giovanni Traverso, Kenneth W Kinzler, and Bert Vogelstein. Transforming single dna molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci U S A*, 100(15):8817–8822, Jul 2003. doi: 10.1073/pnas.1133470100. URL http://dx.doi.org/10.1073/pnas.1133470100. 22

49. P. Druilhe, P. Daubersies, J. Patarapotikul, C. Gentil, L. Chene, T. Chongsuphajaisiddhi, S. Mellouk, and G. Langsley. A primary malarial infection is composed of a very wide range of genetically diverse but related parasites. *J Clin Invest*, 101(9):2008–2016, May 1998. doi: 10.1172/JCI119890. URL http://dx.doi.org/10.1172/JCI119890. 6

50. Nancy O Duah, Michael D Wilson, Anita Ghansah, Ben Abuaku, Dominic Edoh, Neils B Quashie, and Kwadwo A Koram. Mutations in plasmodium falciparum chloroquine resistance transporter and multidrug resistance genes, and treatment outcomes in ghanaian children with uncomplicated malaria. *J Trop Pediatr*, 53(1):27–31, Feb 2007. doi: 10.1093/tropej/fml076. URL http://dx.doi.org/10.1093/tropej/fml076. 116, 120, 133

51. Patrick E Duffy. Plasmodium in the placenta: parasites, parity, protection, prevention and possibly preeclampsia. *Parasitology*, 134(Pt 13):1877–1881, 2007. doi: 10.1017/S0031182007000170. URL http://dx.doi.org/10.1017/S0031182007000170. 17

52. Patrick E Duffy and Michal Fried. Plasmodium falciparum adhesion in the placenta. *Curr Opin Microbiol*, 6(4):371–376, Aug 2003. 17

53. Shigetoshi Eda and Irwin W Sherman. Cytoadherence of malaria-infected red blood cells involves exposure of phosphatidylserine. *Cell Physiol Biochem*, 12(5-6):373–384, 2002. 171

54. B Efron. Boostrap methods: Another look at the jackknife. *Ann. Statist*, 7: 1–26, 1979. 155

55. S. Ehrhardt, F. P. Mockenhaupt, P. Agana-Nsiire, A. Mathieu, S. D. Anemana, K. Stark, R. N. Otchwemah, and U. Bienzle. Efficacy of chloroquine in the treatment of uncomplicated, plasmodium falciparum malaria

in northern ghana. *Ann Trop Med Parasitol*, 96(3):239–247, Apr 2002. doi: 10.1179/000349802125000772. URL http://dx.doi.org/10.1179/000349802125000772. 10, 120, 133

56. John Eid, Adrian Fehr, Jeremy Gray, Khai Luong, John Lyle, Geoff Otto, Paul Peluso, David Rank, Primo Baybayan, Brad Bettman, Arkadiusz Bibillo, Keith Bjornson, Bidhan Chaudhuri, Frederick Christians, Ronald Cicero, Sonya Clark, Ravindra Dalal, Alex Dewinter, John Dixon, Mathieu Foquet, Alfred Gaertner, Paul Hardenbol, Cheryl Heiner, Kevin Hester, David Holden, Gregory Kearns, Xiangxu Kong, Ronald Kuse, Yves Lacroix, Steven Lin, Paul Lundquist, Congcong Ma, Patrick Marks, Mark Maxham, Devon Murphy, Insil Park, Thang Pham, Michael Phillips, Joy Roy, Robert Sebra, Gene Shen, Jon Sorenson, Austin Tomaney, Kevin Travers, Mark Trulson, John Vieceli, Jeffrey Wegener, Dawn Wu, Alicia Yang, Denis Zaccarin, Peter Zhao, Frank Zhong, Jonas Korlach, and Stephen Turner. Real-time dna sequencing from single polymerase molecules. *Science*, 323(5910):133–138, Jan 2009. doi: 10.1126/science.1162986. URL http://dx.doi.org/10.1126/science.1162986. 23

57. Anders Enevold, Michael Alifrangis, Juan J Sanchez, Ilona Carneiro, Cally Roper, Claus BÃÿrsting, John Lusingu, Lasse S Vestergaard, Martha M Lemnge, Niels Morling, Eleanor Riley, and Chris J Drakeley. Associations between alpha+-thalassemia and plasmodium falciparum malarial infection in northeastern tanzania. *J Infect Dis*, 196(3):451–459, Aug 2007. doi: 10.1086/519390. URL http://dx.doi.org/10.1086/519390. 11, 12

58. Milan Fedurco, Anthony Romieu, Scott Williams, Isabelle Lawrence, and Gerardo Turcatti. Bta, a novel reagent for dna attachment on glass and

efficient generation of solid-phase amplified dna colonies. *Nucleic Acids Res*, 34(3):e22, 2006. doi: 10.1093/nar/gnj023. URL http://dx.doi.org/10.1093/nar/gnj023. 22

59. Greg W Fegan, Abdisalan M Noor, Willis S Akhwale, Simon Cousens, and Robert W Snow. Effect of expanded insecticide-treated bednet coverage on child survival in rural kenya: a longitudinal study. *Lancet*, 370(9592): 1035–1039, Sep 2007. doi: 10.1016/S0140-6736(07)61477-9. URL http://dx.doi.org/10.1016/S0140-6736(07)61477-9. 8, 9
60. D. Fernandez-Reyes, A. G. Craig, S. A. Kyes, N. Peshu, R. W. Snow, A. R. Berendt, K. Marsh, and C. I. Newbold. A high frequency african coding polymorphism in the n-terminal domain of icam-1 predisposing to cerebral malaria in kenya. *Hum Mol Genet*, 6(8):1357–1360, Aug 1997. 11
61. D. A. Fidock, T. Nomura, A. K. Talley, R. A. Cooper, S. M. Dzekunov, M. T. Ferdig, L. M. Ursos, A. B. Sidhu, B. NaudÃľ, K. W. Deitsch, X. Z. Su, J. C. Wootton, P. D. Roepe, and T. E. Wellems. Mutations in the p. falciparum digestive vacuole transmembrane protein pfcrt and evidence for their role in chloroquine resistance. *Mol Cell*, 6(4):861–871, Oct 2000. 18, 121
62. T. P. FlÃijeck, T. Jelinek, A. H. Kilian, I. S. Adagu, G. Kabagambe, F. Sonnenburg, and D. C. Warhurst. Correlation of in vivo-resistance to chloroquine and allelic polymorphisms in plasmodium falciparum isolates from uganda. *Trop Med Int Health*, 5(3):174–178, Mar 2000. 18
63. S. Fogh, S. Jepsen, and P. EffersÃÿe. Chloroquine-resistant plasmodium falciparum malaria in kenya. *Trans R Soc Trop Med Hyg*, 73(2):228–229, 1979. 10

64. S. J. Foote, J. K. Thompson, A. F. Cowman, and D. J. Kemp. Amplification of the multidrug resistance gene in some chloroquine-resistant isolates of p. falciparum. *Cell*, 57(6):921–930, Jun 1989. 18

65. S. J. Foote, D. Galatis, and A. F. Cowman. Amino acids in the dihydrofolate reductase-thymidylate synthase gene of plasmodium falciparum involved in cycloguanil resistance differ from those involved in pyrimethamine resistance. *Proc Natl Acad Sci U S A*, 87(8):3014–3017, Apr 1990. 19

66. S. J. Foote, D. E. Kyle, R. K. Martin, A. M. Oduola, K. Forsyth, D. J. Kemp, and A. F. Cowman. Several alleles of the multidrug-resistance gene are closely linked to chloroquine resistance in plasmodium falciparum. *Nature*, 345(6272):255–258, May 1990. doi: 10.1038/345255a0. URL http://dx.doi.org/10.1038/345255a0. 18

67. M. Fried and P. E. Duffy. Adherence of plasmodium falciparum to chondroitin sulfate a in the human placenta. *Science*, 272(5267):1502–1504, Jun 1996. 17

68. M. Fried, F. Nosten, A. Brockman, B. J. Brabin, and P. E. Duffy. Maternal antibodies block malaria. *Nature*, 395(6705):851–852, Oct 1998. doi: 10.1038/27570. URL http://dx.doi.org/10.1038/27570. 17

69. A. FÃď rnert, I. Rooth, Svensson, G. Snounou, and A. BjÃűrkman. Complexity of plasmodium falciparum infections is consistent over time and protects against clinical disease in tanzanian children. *J Infect Dis*, 179(4):989–995, Apr 1999. doi: 10.1086/314652. URL http://dx.doi.org/10.1086/314652. 117

70. Stacey Gabriel, Liuda Ziaugra, and Diana Tabbaa. Snp genotyping using the sequenom massarray iplex platform. *Curr Protoc Hum Genet*, Chapter

2:Unit 2.12, Jan 2009. doi: 10.1002/0471142905.hg0212s60. URL http://dx.doi.org/10.1002/0471142905.hg0212s60. 27, 44

71. Malcolm J Gardner, Neil Hall, Eula Fung, Owen White, Matthew Berriman, Richard W Hyman, Jane M Carlton, Arnab Pain, Karen E Nelson, Sharen Bowman, Ian T Paulsen, Keith James, Jonathan A Eisen, Kim Rutherford, Steven L Salzberg, Alister Craig, Sue Kyes, Man-Suen Chan, Vishvanath Nene, Shamira J Shallom, Bernard Suh, Jeremy Peterson, Sam Angiuoli, Mihaela Pertea, Jonathan Allen, Jeremy Selengut, Daniel Haft, Michael W Mather, Akhil B Vaidya, David M A Martin, Alan H Fairlamb, Martin J Fraunholz, David S Roos, Stuart A Ralph, Geoffrey I McFadden, Leda M Cummings, G. Mani Subramanian, Chris Mungall, J. Craig Venter, Daniel J Carucci, Stephen L Hoffman, Chris Newbold, Ronald W Davis, Claire M Fraser, and Bart Barrell. Genome sequence of the human malaria parasite plasmodium falciparum. *Nature*, 419(6906):498–511, Oct 2002. doi: 10.1038/nature01097. URL http://dx.doi.org/10.1038/nature01097. 85, 233

72. Malcolm J Gardner, Shamira J Shallom, Jane M Carlton, Steven L Salzberg, Vishvanath Nene, Azadeh Shoaibi, Anne Ciecko, Jeffery Lynn, Michael Rizzo, Bruce Weaver, Behnam Jarrahi, Michael Brenner, Babak Parvizi, Luke Tallon, Azita Moazzez, David Granger, Claire Fujii, Cheryl Hansen, James Pederson, Tamara Feldblyum, Jeremy Peterson, Bernard Suh, Sam Angiuoli, Mihaela Pertea, Jonathan Allen, Jeremy Selengut, Owen White, Leda M Cummings, Hamilton O Smith, Mark D Adams, J. Craig Venter, Daniel J Carucci, Stephen L Hoffman, and Claire M Fraser. Sequence of plasmodium falciparum chromosomes 2, 10, 11 and 14. *Na-*

ture, 419(6906):531–534, Oct 2002. doi: 10.1038/nature01094. URL http://dx.doi.org/10.1038/nature01094. 160

73. Anita Ghansah. *Analysis of Disease Association and Evolutionary Selection of Human Beta Globin Gene Variants in Ghana*. PhD thesis, London School of Hygiene and Tropical Medicine, 2008. 172, 178, 206

74. Anita Ghansah, Kirk A Rockett, Taane G Clark, Michael D Wilson, Kwadwo A Koram, Abraham R Oduro, Lucas Amenga-Etego, Thomas Anyorigiya, Abraham Hodgson, Paul Milligan, William O Rogers, and Dominic P Kwiatkowski. Haplotype analyses of haemoglobin c and haemoglobin s and the dynamics of the evolutionary response to malaria in kassena-nankana district of ghana. *PLoS One*, 7(4):e34565, 2012. doi: 10.1371/journal.pone.0034565. URL http://dx.doi.org/10.1371/journal.pone.0034565. 172

75. Michael A Gilchrist and Akira Sasaki. Modeling host-parasite coevolution: a nested approach based on mechanistic models. *J Theor Biol*, 218(3):289–308, Oct 2002. 30

76. Louis Clement Gouagna, Germana Bancone, Frank Yao, Bienvenue Yameogo, Kounbobr Roch DabirÃĭ, Carlo Costantini, Jacques SimporÃĭ, Jean Bosco Ouedraogo, and David Modiano. Genetic variation in human hbb is associated with plasmodium falciparum transmission. *Nat Genet*, 42 (4):328–331, Apr 2010. doi: 10.1038/ng.554. URL http://dx.doi.org/10.1038/ng.554. 14

77. B. M. Greenwood. The epidemiology of malaria. *Ann Trop Med Parasitol*, 91(7):763–769, Oct 1997. 3

78. Brian Greenwood and Theonest Mutabingwa. Malaria in 2002. *Nature*, 415

(6872):670–672, Feb 2002. doi: 10.1038/415670a. URL http://dx.doi.org/10.1038/415670a. 9

79. Brian Greenwood, Pedro Alonso, Feiko O ter Kuile, Jenny Hill, and Richard W Steketee. Malaria in pregnancy: priorities for research. *Lancet Infect Dis*, 7(2):169–174, Feb 2007. doi: 10.1016/S1473-3099(07)70028-2. URL http://dx.doi.org/10.1016/S1473-3099(07)70028-2. 120

80. Brian M Greenwood, Kalifa Bojang, Christopher J M Whitty, and Geoffrey A T Targett. Malaria. *Lancet*, 365(9469):1487–1498, 2005. doi: 10.1016/S0140-6736(05)66420-3. URL http://dx.doi.org/10.1016/S0140-6736(05)66420-3. 2

81. Brian M Greenwood, David A Fidock, Dennis E Kyle, Stefan H I Kappe, Pedro L Alonso, Frank H Collins, and Patrick E Duffy. Malaria: progress, perils, and prospects for eradication. *J Clin Invest*, 118(4):1266–1276, Apr 2008. doi: 10.1172/JCI33996. URL http://dx.doi.org/10.1172/JCI33996. 1, 2, 4, 7, 8, 10

82. Aric Gregson and Christopher V Plowe. Mechanisms of resistance of malaria parasites to antifolates. *Pharmacol Rev*, 57(1):117–145, Mar 2005. doi: 10.1124/pr.57.1.4. URL http://dx.doi.org/10.1124/pr.57.1.4. 19

83. Sekou F Traore Mahamoudou B Toure Mohamed M Traore Sekou Bah Seydou Doumbia GÃijnter C MÃijller, John C Beier and Yosef Schlein. Successful field trial of attractive toxic sugar bait (atsb) plant-spraying methods against malaria vectors in the anopheles gambiae complex in mali, west africa. *Malaria Journal*, 9:2–7, 2010. 114, 218

84. Timothy D Harris, Phillip R Buzby, Hazen Babcock, Eric Beer, Jayson Bowers, Ido Braslavsky, Marie Causey, Jennifer Colonell, James Dimeo,

J. William Efcavitch, Eldar Giladi, Jaime Gill, John Healy, Mirna Jarosz, Dan Lapen, Keith Moulton, Stephen R Quake, Kathleen Steinmann, Edward Thayer, Anastasia Tyurina, Rebecca Ward, Howard Weiss, and Zheng Xie. Single-molecule dna sequencing of a viral genome. *Science*, 320 (5872):106–109, Apr 2008. doi: 10.1126/science.1150427. URL http://dx.doi.org/10.1126/science.1150427. 23

85. D. Hartl and A. G Clark. *Principles of population genetics*. Sinauer Associates, Sunderland, Massachusetts, 2007. 91

86. I. M. Hastings and U. D'Alessandro. Modelling a predictable disaster: the rise and spread of drug-resistantmalaria. *Parasitol Today*, 16(8):340–347, Aug 2000. 10

87. J. B. Herrick. Peculiar elongated and sickle-shaped red blood corpuscles in a case of severe anemia. 1910. *Yale J Biol Med*, 74(3):179–184, 2001. 12, 13

88. A. V. Hill, C. E. Allsopp, D. Kwiatkowski, N. M. Anstey, P. Twumasi, P. A. Rowe, S. Bennett, D. Brewster, A. J. McMichael, and B. M. Greenwood. Common west african hla antigens are associated with protection from severe malaria. *Nature*, 352(6336):595–600, Aug 1991. doi: 10.1038/352595a0. URL http://dx.doi.org/10.1038/352595a0. 11

89. E. Hugosson, S. M. Montgomery, Z. Premji, M. Troye-Blomberg, and A. BjÃűrkman. Higher il-10 levels are associated with less effective clearance of plasmodium falciparum parasites. *Parasite Immunol*, 26(3):111–117, Mar 2004. doi: 10.1111/j.0141-9838.2004.00678.x. URL http://dx.doi.org/10.1111/j.0141-9838.2004.00678.x. 12

90. Lars Hviid and Trine Staalsoe. Malaria immunity in infants: a special case of a general phenomenon? *Trends Parasitol*, 20(2):66–72, Feb 2004. 110

91. M. Ingelman-Sundberg, M. Oscarson, and R. A. McLellan. Polymorphic human cytochrome p450 enzymes: an opportunity for individualized drug treatment. *Trends Pharmacol Sci*, 20(8):342–349, Aug 1999. 134

92. Muminatou Jallow, Yik Ying Teo, Kerrin S Small, Kirk A Rockett, Panos Deloukas, Taane G Clark, Katja Kivinen, Kalifa A Bojang, David J Conway, Margaret Pinder, Giorgio Sirugo, Fatou Sisay-Joof, Stanley Usen, Sarah Auburn, Suzannah J Bumpstead, Susana Campino, Alison Coffey, Andrew Dunham, Andrew E Fry, Angela Green, Rhian Gwilliam, Sarah E Hunt, Michael Inouye, Anna E Jeffreys, Alieu Mendy, Aarno Palotie, Simon Potter, Jiannis Ragoussis, Jane Rogers, Kate Rowlands, Elilan Somaskantharajah, Pamela Whittaker, Claire Widden, Peter Donnelly, Bryan Howie, Jonathan Marchini, Andrew Morris, Miguel SanJoaquin, Eric Akum Achidi, Tsiri Agbenyega, Angela Allen, Olukemi Amodu, Patrick Corran, Abdoulaye Djimde, Amagana Dolo, Ogobara K Doumbo, Chris Drakeley, Sarah Dunstan, Jennifer Evans, Jeremy Farrar, Deepika Fernando, Tran Tinh Hien, Rolf D Horstmann, Muntaser Ibrahim, Nadira Karunaweera, Gilbert Kokwaro, Kwadwo A Koram, Martha Lemnge, Julie Makani, Kevin Marsh, Pascal Michon, David Modiano, Malcolm E Molyneux, Ivo Mueller, Michael Parker, Norbert Peshu, Christopher V Plowe, Odile Puijalon, John Reeder, Hugh Reyburn, Eleanor M Riley, Anavaj Sakuntabhai, Pratap Singhasivanon, Sodiomon Sirima, Adama Tall, Terrie E Taylor, Mahamadou Thera, Marita Troye-Blomberg, Thomas N Williams, Michael Wilson, Dominic P Kwiatkowski, Wellcome Trust Case Control Consortium, and Malaria Genomic Epidemiology Network. Genome-wide and fine-resolution association analysis of malaria in west africa. *Nat Genet*, 41(6):657–665, Jun 2009. doi:

10.1038/ng.388. URL http://dx.doi.org/10.1038/ng.388. 11, 206

93. Daniel C Jeffares, Arnab Pain, Andrew Berry, Anthony V Cox, James Stalker, Catherine E Ingle, Alan Thomas, Michael A Quail, Kyle Siebenthall, Anne-Catrin Uhlemann, Sue Kyes, Sanjeev Krishna, Chris Newbold, Emmanouil T Dermitzakis, and Matthew Berriman. Genome variation and evolution of the malaria parasite plasmodium falciparum. *Nat Genet*, 39(1): 120–125, Jan 2007. doi: 10.1038/ng1931. URL http://dx.doi.org/10.1038/ng1931. 14

94. Christian Jurinke, Dirk van den Boom, Charles R Cantor, and Hubert KÃűster. The use of massarray technology for high throughput genotyping. *Adv Biochem Eng Biotechnol*, 77:57–74, 2002. 27

95. Christian Jurinke, Dirk van den Boom, Charles R Cantor, and Hubert KÃűster. Automated genotyping using the dna massarray technology. *Methods Mol Biol*, 187:179–192, 2002. doi: 10.1385/1-59259-273-2:179. URL http://dx.doi.org/10.1385/1-59259-273-2:179. 27

96. K. C. Kain and D. E. Lanar. Determination of genetic variation within plasmodium falciparum by using enzymatically amplified dna from filter paper disks impregnated with whole blood. *J Clin Microbiol*, 29(6):1171–1174, Jun 1991. 174

97. Chiea C Khor, Stephen J Chapman, Fredrik O Vannberg, Aisling Dunne, Caroline Murphy, Edmund Y Ling, Angela J Frodsham, Andrew J Walley, Otto Kyrieleis, Amir Khan, Christophe Aucan, Shelley Segal, Catrin E Moore, Kyle Knox, Sarah J Campbell, Christian Lienhardt, Anthony Scott, Peter Aaby, Oumou Y Sow, Robert T Grignani, Jackson Sillah, Giorgio Sirugo, Nobert Peshu, Thomas N Williams, Kathryn Maitland, Robert

J O Davies, Dominic P Kwiatkowski, Nicholas P Day, Djamel Yala, Derrick W Crook, Kevin Marsh, James A Berkley, Luke A J O'Neill, and Adrian V S Hill. A mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis. *Nat Genet*, 39(4):523–528, Apr 2007. doi: 10.1038/ng1976. URL http://dx.doi.org/10.1038/ng1976. 207

98. Chiea C Khor, Fredrik O Vannberg, Stephen J Chapman, Haiyan Guo, Sunny H Wong, Andrew J Walley, Damjan Vukcevic, Anna Rautanen, Tara C Mills, Kwok-Chiu Chang, Kai-Man Kam, Amelia C Crampin, Bagrey Ngwira, Chi-Chiu Leung, Cheuk-Ming Tam, Chiu-Yeung Chan, Joseph J Y Sung, Wing-Wai Yew, Kai-Yee Toh, Stacey K H Tay, Dominic Kwiatkowski, Christian Lienhardt, Tran-Tinh Hien, Nicholas P Day, Nobert Peshu, Kevin Marsh, Kathryn Maitland, J. Anthony Scott, Thomas N Williams, James A Berkley, Sian Floyd, Nelson L S Tang, Paul E M Fine, Denise L M Goh, and Adrian V S Hill. Cish and susceptibility to infectious diseases. *N Engl J Med*, 362(22):2092–2101, Jun 2010. doi: 10.1056/NEJMoa0905606. URL http://dx.doi.org/10.1056/NEJMoa0905606. 12, 225

99. L. KonatÃĭ, J. Zwetyenga, C. Rogier, E. Bischoff, D. Fontenille, A. Tall, A. Spiegel, J. F. Trape, and O. Mercereau-Puijalon. Variation of plasmodium falciparum msp1 block 2 and msp2 allele prevalence and of infection complexity in two neighbouring senegalese villages with different transmission conditions. *Trans R Soc Trop Med Hyg*, 93 Suppl 1:21–28, Feb 1999. 6, 220

100. K. A. Koram, S. Owusu-Agyei, G. Utz, F. N. Binka, J. K. Baird, S. L. Hoffman, and F. K. Nkrumah. Severe anemia in young children after high

and low malaria transmission seasons in the kassena-nankana district of northern ghana. *Am J Trop Med Hyg*, 62(6):670–674, Jun 2000. 33, 77, 82, 134

101. Kwadwo A Koram, Seth Owusu-Agyei, David J Fryauff, Francis Anto, Frank Atuguba, Abraham Hodgson, Stephen L Hoffman, and Francis K Nkrumah. Seasonal profiles of malaria infection, anaemia, and bednet use among age groups and communities in northern ghana. *Trop Med Int Health*, 8(9): 793–802, Sep 2003. 33, 77, 82

102. Kwadwo A. Koram, Benjamin Abuaku, Nancy Duah, and Neils Quashie. Comparative efficacy of antimalarial drugs including acts in the treatment of uncomplicated malaria among children under 5 years in ghana. *Acta Tropica*, 95:194 – 203, 2005. 10

103. Eline L Korenromp, Brian G Williams, Eleanor Gouws, Christopher Dye, and Robert W Snow. Measurement of trends in childhood malaria mortality in africa: an assessment of progress toward targets based on verbal autopsy. *Lancet Infect Dis*, 3(6):349–358, Jun 2003. 1

104. James G Kublin, Joseph F Cortese, Eric Mbindo Njunju, Rabia A G Mukadam, Jack J Wirima, Peter N Kazembe, Abdoulaye A DjimdÃľ, Bourema Kouriba, Terrie E Taylor, and Christopher V Plowe. Reemergence of chloroquine-sensitive plasmodium falciparum malaria after cessation of chloroquine use in malawi. *J Infect Dis*, 187(12):1870–1875, Jun 2003. doi: 10.1086/375419. URL http://dx.doi.org/10.1086/375419. 133, 222

105. Rolf E Kuestner, David W Taft, Aaron Haran, Cameron S Brandt, Ty Brender, Karen Lum, Brandon Harder, Shannon Okada, Craig D Ostrander, James L Kreindler, Shean J Aujla, Brian Reardon, Margaret Moore, Pamela

Shea, Randall Schreckhise, Thomas R Bukowski, Scott Presnell, Patricia Guerra-Lewis, Julia Parrish-Novak, Jeff L Ellsworth, Stephen Jaspers, Katherine E Lewis, Mark Appleby, Jay K Kolls, Mark Rixon, James W West, Zeren Gao, and Steven D Levin. Identification of the il-17 receptor related molecule il-17rc as the receptor for il-17f. *J Immunol*, 179(8): 5462–5473, Oct 2007. 225

106. J. A. Kurtzhals, V. Adabayeri, B. Q. Goka, B. D. Akanmori, J. O. Oliver-Commey, F. K. Nkrumah, C. Behr, and L. Hviid. Low plasma concentrations of interleukin 10 in severe malarial anaemia compared with cerebral and uncomplicated malaria. *Lancet*, 351(9118):1768–1772, Jun 1998. doi: 10.1016/S0140-6736(97)09439-7. URL http://dx.doi.org/10.1016/S0140-6736(97)09439-7. 207

107. Dominic P Kwiatkowski. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet*, 77(2): 171–192, Aug 2005. doi: 10.1086/432519. URL http://dx.doi.org/10.1086/432519. 30

108. Dominic P Kwiatkowski. The complexity of genetic variation in a simple immune system. *Trends Genet*, 21(4):197–199, Apr 2005. doi: 10.1016/j.tig.2005.02.002. URL http://dx.doi.org/10.1016/j.tig.2005.02.002. 11

109. S. Kyes, P. Horrocks, and C. Newbold. Antigenic variation at the infected red cell surface in malaria. *Annu Rev Microbiol*, 55:673–707, 2001. doi: 10.1146/annurev.micro.55.1.673. URL http://dx.doi.org/10.1146/annurev.micro.55.1.673. 16

110. John H Leamon, William L Lee, Karrie R Tartaro, Janna R Lanza, Gary J Sarkis, Alex D deWinter, Jan Berka, Michael Weiner, Jonathan M Rothberg,

and Kenton L Lohman. A massively parallel picotiterplate based platform for discrete picoliter-scale polymerase chain reactions. *Electrophoresis*, 24 (21):3769–3777, Nov 2003. doi: 10.1002/elps.200305646. URL http://dx.doi.org/10.1002/elps.200305646. 22

111. J. H. Leech, J. W. Barnwell, L. H. Miller, and R. J. Howard. Identification of a strain-specific malarial antigen exposed on the surface of plasmodium falciparum-infected erythrocytes. *J Exp Med*, 159(6):1567–1575, Jun 1984. 16

112. Heng Li and Richard Durbin. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25(14):1754–1760, Jul 2009. doi: 10.1093/bioinformatics/btp324. URL http://dx.doi.org/10.1093/bioinformatics/btp324. 86

113. G. Luoni, F. Verra, B. ArcÃă, B. S. Sirima, M. Troye-Blomberg, M. Coluzzi, D. Kwiatkowski, and D. Modiano. Antimalarial antibody levels and il4 polymorphism in the fulani of west africa. *Genes Immun*, 2(7):411–414, Nov 2001. doi: 10.1038/sj.gene.6363797. URL http://dx.doi.org/10.1038/sj.gene.6363797. 207

114. K. E. Lyke, R. Burges, Y. Cissoko, L. Sangare, M. Dao, I. Diarra, A. Kone, R. Harley, C. V. Plowe, O. K. Doumbo, and M. B. Sztein. Serum levels of the proinflammatory cytokines interleukin-1 beta (il-1beta), il-6, il-8, il-10, tumor necrosis factor alpha, and il-12(p70) in malian children with severe plasmodium falciparum malaria and matched uncomplicated malaria or healthy controls. *Infect Immun*, 72(10):5630–5637, Oct 2004. doi: 10.1128/IAI.72.10.5630-5637.2004. URL http://dx.doi.org/10.1128/IAI.72.10.5630-5637.2004. 12

115. Margaret J Mackinnon and Andrew F Read. Virulence in malaria: an evolutionary viewpoint. *Philos Trans R Soc Lond B Biol Sci*, 359(1446): 965–986, Jun 2004. doi: 10.1098/rstb.2003.1414. URL http://dx.doi.org/10.1098/rstb.2003.1414. 171

116. Claire L Mackintosh, James G Beeson, and Kevin Marsh. Clinical features and pathogenesis of severe malaria. *Trends Parasitol*, 20(12):597–603, Dec 2004. doi: 10.1016/j.pt.2004.09.006. URL http://dx.doi.org/10.1016/j.pt.2004.09.006. 2

117. Allen L Malisa, Richard J Pearce, Salim Abdulla, Hassan Mshinda, Patrick S Kachur, Peter Bloland, and Cally Roper. Drug coverage in treatment of malaria and the consequences for resistance evolution–evidence from the use of sulphadoxine/pyrimethamine. *Malar J*, 9:190, 2010. doi: 10.1186/1475-2875-9-190. URL http://dx.doi.org/10.1186/1475-2875-9-190. 131

118. Alphaxard Manjurano, Taane G. Clark, Behzad Nadjm, George Mtove, Hannah Wangai, Nuno Sepulveda, Susana G. Campino, Caroline Maxwell, Raimos Olomi, Kirk R. Rockett, Anna Jeffreys, Eleanor M. Riley, Hugh Reyburn, Christopher Drakeley, and MalariaGen Consortium. Candidate human genetic polymorphisms and severe malaria in a tanzanian population. *PLoS ONE*, 7:e47463, 2012. 207

119. Heinrich Magnus Manske and Dominic P Kwiatkowski. Lookseq: a browser-based viewer for deep sequencing data. *Genome Res*, 19(11):2125–2132, Nov 2009. doi: 10.1101/gr.093443.109. URL http://dx.doi.org/10.1101/gr.093443.109. 15

120. Heinrich Magnus Manske and Dominic P Kwiatkowski. Snp-o-matic. *Bioin-*

formatics, 25(18):2434–2435, Sep 2009. doi: 10.1093/bioinformatics/btp403. URL http://dx.doi.org/10.1093/bioinformatics/btp403. 15

121. Magnus Manske, Olivo Miotto, Susana Campino, Sarah Auburn, Jacob Almagro-Garcia, Gareth Maslen, Jack O'Brien, Abdoulaye Djimde, Ogobara Doumbo, Issaka Zongo, Jean-Bosco Ouedraogo, Pascal Michon, Ivo Mueller, Peter Siba, Alexis Nzila, Steffen Borrmann, Steven M Kiara, Kevin Marsh, Hongying Jiang, Xin-Zhuan Su, Chanaki Amaratunga, Rick Fairhurst, Duong Socheat, Francois Nosten, Mallika Imwong, Nicholas J White, Mandy Sanders, Elisa Anastasi, Dan Alcock, Eleanor Drury, Samuel Oyola, Michael A Quail, Daniel J Turner, Valentin Ruano-Rubio, Dushyanth Jyothi, Lucas Amenga-Etego, Christina Hubbart, Anna Jeffreys, Kate Rowlands, Colin Sutherland, Cally Roper, Valentina Mangano, David Modiano, John C Tan, Michael T Ferdig, Alfred Amambua-Ngwa, David J Conway, Shannon Takala-Harrison, Christopher V Plowe, Julian C Rayner, Kirk A Rockett, Taane G Clark, Chris I Newbold, Matthew Berriman, Bronwyn Macinnis, and Dominic P Kwiatkowski. Analysis of plasmodium falciparum diversity in natural infections by deep sequencing. *Nature*, Jun 2012. doi: 10.1038/nature11174. URL http://dx.doi.org/10.1038/nature11174. 15, 16, 29, 52, 80, 85, 86, 93, 94, 98, 99, 107, 113, 114, 115, 117, 118, 137, 165, 218, 219, 220, 223, 233

122. Elaine R Mardis. Next-generation dna sequencing methods. *Annu Rev Genomics Hum Genet*, 9:387–402, 2008. doi: 10.1146/annurev.genom.9.081307.164359. URL http://dx.doi.org/10.1146/annurev.genom.9.081307.164359. 20, 21, 24, 52

123. Elaine R Mardis. The impact of next-generation sequencing technology on

genetics. *Trends Genet*, 24(3):133–141, Mar 2008. doi: 10.1016/j.tig.2007.12.007. URL http://dx.doi.org/10.1016/j.tig.2007.12.007. 21

124. Marcel Margulies, Michael Egholm, William E Altman, Said Attiya, Joel S Bader, Lisa A Bemben, Jan Berka, Michael S Braverman, Yi-Ju Chen, Zhoutao Chen, Scott B Dewell, Lei Du, Joseph M Fierro, Xavier V Gomes, Brian C Godwin, Wen He, Scott Helgesen, Chun Heen Ho, Chun He Ho, Gerard P Irzyk, Szilveszter C Jando, Maria L I Alenquer, Thomas P Jarvie, Kshama B Jirage, Jong-Bum Kim, James R Knight, Janna R Lanza, John H Leamon, Steven M Lefkowitz, Ming Lei, Jing Li, Kenton L Lohman, Hong Lu, Vinod B Makhijani, Keith E McDade, Michael P McKenna, Eugene W Myers, Elizabeth Nickerson, John R Nobile, Ramona Plant, Bernard P Puc, Michael T Ronan, George T Roth, Gary J Sarkis, Jan Fredrik Simons, John W Simpson, Maithreyan Srinivasan, Karrie R Tartaro, Alexander Tomasz, Kari A Vogt, Greg A Volkmer, Shally H Wang, Yong Wang, Michael P Weiner, Pengguang Yu, Richard F Begley, and Jonathan M Rothberg. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057):376–380, Sep 2005. doi: 10.1038/nature03959. URL http://dx.doi.org/10.1038/nature03959. 15

125. Florian Marks, Christian G Meyer, JÃijrgen Sievertsen, Christian Timmann, Jennifer Evans, Rolf D Horstmann, and JÃijrgen May. Genotyping of plasmodium falciparum pyrimethamine resistance by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry. *Antimicrob Agents Chemother*, 48(2):466–472, Feb 2004. 161

126. K. Marsh and R. J. Howard. Antigens induced on erythrocytes by p. falciparum: expression of diverse and conserved determinants. *Science*, 231

(4734):150–153, Jan 1986. 3, 16

127. K. Marsh and R. W. Snow. Host-parasite interaction and morbidity in malaria endemic areas. *Philos Trans R Soc Lond B Biol Sci*, 352(1359): 1385–1394, Sep 1997. doi: 10.1098/rstb.1997.0124. URL http://dx.doi.org/10.1098/rstb.1997.0124. 2

128. C. M. Masimirembwa, J. A. Hasler, and I. Johansson. Inhibitory effects of antiparasitic drugs on cytochrome p450 2d6. *Eur J Clin Pharmacol*, 48(1): 35–38, 1995. 135

129. J. May, F. P. Mockenhaupt, O. G. Ademowo, A. G. Falusi, P. E. Olumese, U. Bienzle, and C. G. Meyer. High rate of mixed and subpatent malarial infections in southwest nigeria. *Am J Trop Med Hyg*, 61(2):339–343, Aug 1999. 232

130. JÃijrgen May and Christian G Meyer. Association of plasmodium falciparum chloroquine resistance transporter variant t76 with age-related plasma chloroquine levels. *Am J Trop Med Hyg*, 68(2):143–146, Feb 2003. 232

131. Alfredo Mayor, Francisco Saute, John J Aponte, JesÃžs Almeda, F. Xavier GÃşmez-OlivÃľ, Martinho Dgedge, and Pedro L Alonso. Plasmodium falciparum multiple infections in mozambique, its relation to other malariological indices and to prospective risk of malaria morbidity. *Trop Med Int Health*, 8(1):3–11, Jan 2003. 117

132. McCormack CP. Human ecology and behaviour in malaria control in tropical Africa . *Bulletin, World Health Organization*, 62:81–87, 1984. 8

133. W. McGuire, A. V. Hill, C. E. Allsopp, B. M. Greenwood, and D. Kwiatkowski. Variation in the tnf-alpha promoter region associated with

susceptibility to cerebral malaria. *Nature*, 371(6497):508–510, Oct 1994. doi: 10.1038/371508a0. URL http://dx.doi.org/10.1038/371508a0. 11

134. C. Menendez, A. F. Fleming, and P. L. Alonso. Malaria-related anaemia. *Parasitol Today*, 16(11):469–476, Nov 2000. 3

135. Michael L Metzker. Emerging technologies in dna sequencing. *Genome Res*, 15(12):1767–1776, Dec 2005. doi: 10.1101/gr.3770505. URL http://dx.doi.org/10.1101/gr.3770505. 23

136. Michael L Metzker. Sequencing technologies - the next generation. *Nat Rev Genet*, 11(1):31–46, Jan 2010. doi: 10.1038/nrg2626. URL http://dx.doi.org/10.1038/nrg2626. xiii, 21, 22, 23, 24, 25, 26

137. Louis H Miller, Dror I Baruch, Kevin Marsh, and Ogobara K Doumbo. The pathogenic basis of malaria. *Nature*, 415(6872):673–679, Feb 2002. doi: 10.1038/415673a. URL http://dx.doi.org/10.1038/415673a. 4, 207

138. Toshihiro Mita, Akira Kaneko, J. Koji Lum, Bwijo Bwijo, Miho Takechi, Innocent L Zungu, Takahiro Tsukahara, Kazuyuki Tanabe, Takatoshi Kobayakawa, and Anders BjÃűrkman. Recovery of chloroquine sensitivity and low prevalence of the plasmodium falciparum chloroquine resistance transporter gene mutation k76t following the discontinuance of chloroquine use in malawi. *Am J Trop Med Hyg*, 68(4):413–415, Apr 2003. 133, 222

139. F. P. Mockenhaupt, S. Ehrhardt, T. A. Eggelte, P. Agana-Nsiire, K. Stollberg, A. Mathieu, M. Markert, R. N. Otchwemah, and U. Bienzle. Chloroquine-treatment failure in northern ghana: roles of pfcrt t76 and pfmdr1 y86. *Ann Trop Med Parasitol*, 99(8):723–732, Dec 2005. doi: 10.1179/136485905X75395. URL http://dx.doi.org/10.1179/136485905X75395. 120, 121, 133

140. Frank P Mockenhaupt, Stephan Ehrhardt, Jakob P Cramer, Rowland N Otchwemah, Sylvester D Anemana, Katharina Goltz, Franziska Mylius, Ekkehart Dietz, Teunis A Eggelte, and Ulrich Bienzle. Hemoglobin c and resistance to severe malaria in ghanaian children. *J Infect Dis*, 190(5):1006–1009, Sep 2004. doi: 10.1086/422847. URL http://dx.doi.org/10.1086/422847. 13

141. D. Modiano, G. Luoni, B. S. Sirima, J. SimporÃľ, F. Verra, A. KonatÃľ, E. Rastrelli, A. Olivieri, C. Calissano, G. M. Paganotti, L. D'Urbano, I. Sanou, A. Sawadogo, G. Modiano, and M. Coluzzi. Haemoglobin c protects against clinical plasmodium falciparum malaria. *Nature*, 414(6861):305–308, Nov 2001. doi: 10.1038/35104556. URL http://dx.doi.org/10.1038/35104556. 13

142. Jianbing Mu, Philip Awadalla, Junhui Duan, Kate M McGee, Deirdre A Joy, Gilean A T McVean, and Xin zhuan Su. Recombination hotspots and population structure in plasmodium falciparum. *PLoS Biol*, 3(10):e335, Oct 2005. doi: 10.1371/journal.pbio.0030335. URL http://dx.doi.org/10.1371/journal.pbio.0030335. 233

143. Jianbing Mu, Philip Awadalla, Junhui Duan, Kate M McGee, Jon Keebler, Karl Seydel, Gilean A T McVean, and Xin zhuan Su. Genome-wide variation and identification of vaccine targets in the plasmodium falciparum genome. *Nat Genet*, 39(1):126–130, Jan 2007. doi: 10.1038/ng1924. URL http://dx.doi.org/10.1038/ng1924. 14, 219

144. Jianbing Mu, Rachel A Myers, Hongying Jiang, Shengfa Liu, Stacy Ricklefs, Michael Waisberg, Kesinee Chotivanich, Polrat Wilairatana, Srivicha Krudsood, Nicholas J White, Rachanee Udomsangpetch, Liwang Cui, May Ho,

Fengzhen Ou, Haibo Li, Jianping Song, Guoqiao Li, Xinhua Wang, Suon Seila, Sreng Sokunthea, Duong Socheat, Daniel E Sturdevant, Stephen F Porcella, Rick M Fairhurst, Thomas E Wellems, Philip Awadalla, and Xin zhuan Su. Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nat Genet*, 42(3):268–271, Mar 2010. doi: 10.1038/ng.528. URL http://dx.doi.org/10.1038/ng.528. 18, 29

145. D. A. MÃijller, J. D. Charlwood, I. Felger, C. Ferreira, V. do Rosario, and T. Smith. Prospective risk of morbidity in relation to multiplicity of infection with plasmodium falciparum in sÃčo tomÃĺ. *Acta Trop*, 78(2):155–162, Feb 2001. 6, 117, 220

146. Bakyaita N, Langi P, Talisuna A, and Edmonson J. *Getting research into policy: using antimalarial drug resistance study results for policy change in Uganda. [abstract]. In: Moving Targets: Parasites, Resistance and Access to Drugs; International Colloquium, Antwerp.* Prince Leopold Institute of Tropical Medicine, Antwerp, Belgium, 50 p., 2000. 10

147. H. S. Nagesha, Din-Syafruddin, G. J. Casey, A. I. Susanti, D. J. Fryauff, J. C. Reeder, and A. F. Cowman. Mutations in the pfmdr1, dhfr and dhps genes of plasmodium falciparum are associated with in-vivo drug resistance in west papua, indonesia. *Trans R Soc Trop Med Hyg*, 95(1):43–49, 2001. 18

148. Inbarani Naidoo and Cally Roper. Following the path of most resistance: dhps k540e dispersal in african plasmodium falciparum. *Trends Parasitol*, 26(9):447–456, Sep 2010. doi: 10.1016/j.pt.2010.05.001. URL http://dx.doi.org/10.1016/j.pt.2010.05.001. 120, 131, 209

149. Inbarani Naidoo and Cally Roper. Drug resistance maps to guide intermittent preventive treatment of malaria in african infants. *Parasitology*, 138(12):1469–1479, Oct 2011. doi: 10.1017/S0031182011000746. URL http://dx.doi.org/10.1017/S0031182011000746. 125, 131, 132

150. Magatte Ndiaye, Babacar Faye, Roger Tine, Jean Louis Ndiaye, Aminata Lo, Annie Abiola, Yemou Dieng, Daouda Ndiaye, Rachel Hallett, Michael Alifrangis, and Oumar Gaye. Assessment of the molecular marker of plasmodium falciparum chloroquine resistance (pfcrt) in senegal after several years of chloroquine withdrawal. *Am J Trop Med Hyg*, 87(4):640–645, Oct 2012. doi: 10.4269/ajtmh.2012.11-0709. URL http://dx.doi.org/10.4269/ajtmh.2012.11-0709. 133

151. Malaria Genomic Epidemiology Network. A global network for investigating the genomic epidemiology of malaria. *Nature*, 456(7223):732–737, Dec 2008. doi: 10.1038/nature07632. URL http://dx.doi.org/10.1038/nature07632. 7, 215

152. C. Newbold, P. Warn, G. Black, A. Berendt, A. Craig, B. Snow, M. Msobo, N. Peshu, and K. Marsh. Receptor-specific adhesion and clinical disease in plasmodium falciparum. *Am J Trop Med Hyg*, 57(4):389–398, Oct 1997. 17

153. C. I. Newbold, A. G. Craig, S. Kyes, A. R. Berendt, R. W. Snow, N. Peshu, and K. Marsh. Pfemp1, polymorphism and pathogenesis. *Ann Trop Med Parasitol*, 91(5):551–557, Jul 1997. 16

154. Harald Noedl, Youry Se, Kurt Schaecher, Bryan L Smith, Duong Socheat, Mark M Fukuda, and Artemisinin Resistance in Cambodia 1 (ARC1) Study Consortium. Evidence of artemisinin-resistant malaria in western cambodia. *N Engl J Med*, 359(24):2619–2620, Dec 2008. 11

155. A. M. Nzila, E. K. Mberu, J. Sulo, H. Dayo, P. A. Winstanley, C. H. Sibley, and W. M. Watkins. Towards an understanding of the mechanism of pyrimethamine-sulfadoxine resistance in plasmodium falciparum: genotyping of dihydrofolate reductase and dihydropteroate synthase of kenyan parasites. *Antimicrob Agents Chemother*, 44(4):991–996, Apr 2000. 121, 132

156. Abraham R Oduro, Thomas Anyorigiya, Abraham Hodgson, Patrick Ansah, Francis Anto, Nana A Ansah, Frank Atuguba, George Mumuni, and Joseph Amankwa. A randomized comparative study of chloroquine, amodiaquine and sulphadoxine-pyrimethamine for the treatment of uncomplicated malaria in ghana. *Trop Med Int Health*, 10(3):279–284, Mar 2005. doi: 10.1111/j.1365-3156.2004.01382.x. URL http://dx.doi.org/10.1111/j.1365-3156.2004.01382.x. 120

157. Abraham R Oduro, Kwadwo A Koram, William Rogers, Frank Atuguba, Patrick Ansah, Thomas Anyorigiya, Akosua Ansah, Francis Anto, Nathan Mensah, Abraham Hodgson, and Francis Nkrumah. Severe falciparum malaria in young children of the kassena-nankana district of northern ghana. *Malar J*, 6:96, 2007. doi: 10.1186/1475-2875-6-96. URL http://dx.doi.org/10.1186/1475-2875-6-96. 3, 32, 33, 74, 77, 82, 115, 134, 173, 208

158. Abraham Rexford Oduro, George Wak, Daniel Azongo, Cornelius Debpuur, Peter Wontuo, Felix Kondayire, Paul Welaga, Ayaga Bawah, Alex Nazzar, John Williams, Abraham Hodgson, and Fred Binka. Profile of the navrongo health and demographic surveillance system. *Int J Epidemiol*, 41(4):968–976, Aug 2012. doi: 10.1093/ije/dys111. URL http://dx.doi.org/10.1093/ije/dys111. xiii, 32, 34, 81, 82

159. A. Ofosu-Okyere, M. J. Mackinnon, M. P. Sowa, K. A. Koram, F. Nkrumah, Y. D. Osei, W. G. Hill, M. D. Wilson, and D. E. Arnot. Novel plasmodium falciparum clones and rising clone multiplicities are associated with the increase in malaria morbidity in ghanaian children during the transition into the high transmission season. *Parasitology*, 123(Pt 2):113–123, Aug 2001. 6, 7, 117

160. Mary Chiaka Oguike, Martha Betson, Martina Burke, Debbie Nolder, J. Russell Stothard, Immo Kleinschmidt, Carla Proietti, Teun Bousema, Mathieu Ndounga, Kazuyuki Tanabe, Edward Ntege, Richard Culleton, and Colin J Sutherland. Plasmodium ovale curtisi and plasmodium ovale wallikeri circulate simultaneously in african communities. *Int J Parasitol*, 41(6):677–683, May 2011. doi: 10.1016/j.ijpara.2011.01.004. URL http://dx.doi.org/10.1016/j.ijpara.2011.01.004. 2

161. K. Ohene-Frempong and E. Schwartz. Clinical features of thalassemia. *Pediatr Clin North Am*, 27(2):403–420, May 1980. 12

162. Lucy C Okell, Azra C Ghani, Emily Lyons, and Chris J Drakeley. Submicroscopic infection in plasmodium falciparum-endemic populations: a systematic review and meta-analysis. *J Infect Dis*, 200(10):1509–1517, Nov 2009. doi: 10.1086/644781. URL http://dx.doi.org/10.1086/644781. 217

163. Emelda A Okiro, Abdullah Al-Taiar, Hugh Reyburn, Richard Idro, James A Berkley, and Robert W Snow. Age patterns of severe paediatric malaria and their relationship to plasmodium falciparum transmission intensity. *Malar J*, 8:4, 2009. doi: 10.1186/1475-2875-8-4. URL http://dx.doi.org/10.1186/1475-2875-8-4. 73

164. Wendy Prudhomme O'Meara, Judith Nekesa Mangeni, Rick Steketee, and

Brian Greenwood. Changes in the burden of malaria in sub-saharan africa. *Lancet Infect Dis*, 10(8):545–555, Aug 2010. doi: 10.1016/S1473-3099(10) 70096-7. URL http://dx.doi.org/10.1016/S1473-3099(10)70096-7. 33, 82

165. Awo D Osafo-Addo, Kwadwo A Koram, Abraham R Oduro, Michael Wilson, Abraham Hodgson, and William O Rogers. Hla-drb1*04 allele is associated with severe malaria in northern ghana. *Am J Trop Med Hyg*, 78(2):251–255, Feb 2008. 173

166. C. Othoro, A. A. Lal, B. Nahlen, D. Koech, A. S. Orago, and V. Udhayakumar. A low interleukin-10 tumor necrosis factor-alpha ratio is associated with malaria anemia in children residing in a holoendemic malaria region in western kenya. *J Infect Dis*, 179(1):279–282, Jan 1999. doi: 10.1086/314548. URL http://dx.doi.org/10.1086/314548. 207

167. Collins Ouma, Gregory C Davenport, Tom Were, Michael F Otieno, James B Hittner, John M Vulule, Jeremy Martinson, John M Ong'echa, Robert E Ferrell, and Douglas J Perkins. Haplotypes of il-10 promoter variants are associated with susceptibility to severe malarial anemia and functional changes in il-10 production. *Hum Genet*, 124(5):515–524, Dec 2008. doi: 10.1007/s00439-008-0578-5. URL http://dx.doi.org/10.1007/s00439-008-0578-5. 207

168. S. Owusu-Agyei, K. A. Koram, J. K. Baird, G. C. Utz, F. N. Binka, F. K. Nkrumah, D. J. Fryauff, and S. L. Hoffman. Incidence of symptomatic and asymptomatic plasmodium falciparum infection following curative therapy in adult residents of northern ghana. *Am J Trop Med Hyg*, 65(3):197–203, Sep 2001. 33, 82

169. S. Owusu-Agyei, T. Smith, H-P. Beck, L. Amenga-Etego, and I. Felger. Molecular epidemiology of plasmodium falciparum infections among asymptomatic inhabitants of a holoendemic malarious area in northern ghana. *Trop Med Int Health*, 7(5):421–428, May 2002. 6, 107

170. Seth Owusu-Agyei, David J Fryauff, Daniel Chandramohan, Kwadwo A Koram, Fred N Binka, Francis K Nkrumah, Greg C Utz, and Stephen L Hoffman. Characteristics of severe anemia and its association with malaria in young children of the kassena-nankana district of northern ghana. *Am J Trop Med Hyg*, 67(4):371–377, Oct 2002. 33, 82, 110, 114, 117

171. Samuel O Oyola, Thomas D Otto, Yong Gu, Gareth Maslen, Magnus Manske, Susana Campino, Daniel J Turner, Bronwyn Macinnis, Dominic P Kwiatkowski, Harold P Swerdlow, and Michael A Quail. Optimizing illumina next-generation sequencing library preparation for extremely at-biased genomes. *BMC Genomics*, 13:1, 2012. doi: 10.1186/1471-2164-13-1. URL http://dx.doi.org/10.1186/1471-2164-13-1. 217, 233

172. A. Pain, D. J. Ferguson, O. Kai, B. C. Urban, B. Lowe, K. Marsh, and D. J. Roberts. Platelet-mediated clumping of plasmodium falciparum-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. *Proc Natl Acad Sci U S A*, 98(4):1805–1810, Feb 2001. doi: 10.1073/pnas.98.4.1805. URL http://dx.doi.org/10.1073/pnas.98.4.1805. 17

173. E. Paradis, J. Claude, and K. Strimmer. Ape: analyses of phylogenetics and evolution in r language. *Bioinformatics*, 20:289–290, 2004. 97

174. R. E. Paul, M. J. Packer, M. Walmsley, M. Lagog, L. C. Ranford-Cartwright, R. Paru, and K. P. Day. Mating patterns in malaria parasite populations of papua new guinea. *Science*, 269(5231):1709–1711, Sep 1995. 114, 117

175. Richard J Pearce, Chris Drakeley, Daniel Chandramohan, Frank Mosha, and Cally Roper. Molecular determination of point mutation haplotypes in the dihydrofolate reductase and dihydropteroate synthase of plasmodium falciparum in three districts of northern tanzania. *Antimicrob Agents Chemother*, 47(4):1347–1354, Apr 2003. 121, 131

176. Raymond Pearl. *Human Biology*. Wayne State University Press, 1960. 30, 210

177. D. S. Peterson, D. Walliker, and T. E. Wellems. Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in falciparum malaria. *Proc Natl Acad Sci U S A*, 85(23): 9114–9118, Dec 1988. 19

178. Aung Pyae Phyo, Standwell Nkhoma, Kasia Stepniewska, Elizabeth A Ashley, Shalini Nair, Rose McGready, Carit Ler Moo, Salma Al-Saai, Arjen M Dondorp, Khin Maung Lwin, Pratap Singhasivanon, Nicholas Pj Day, Nicholas J White, Tim Jc Anderson, and FranÃğois Nosten. Emergence of artemisinin-resistant malaria on the western border of thailand: a longitudinal study. *Lancet*, Apr 2012. doi: 10.1016/S0140-6736(12)60484-X. URL http://dx.doi.org/10.1016/S0140-6736(12)60484-X. 11

179. Amy L Pickard, Chansuda Wongsrichanalai, Anne Purfield, Deborah Kamwendo, Kathryn Emery, Christy Zalewski, Fumihiko Kawamoto, R. Scott Miller, and Steven R Meshnick. Resistance to antimalarials in southeast asia and genetic polymorphisms in pfmdr1. *Antimicrob Agents Chemother*, 47(8):2418–2423, Aug 2003. 18

180. T. Planche and S. Krishna. Severe malaria: metabolic complications. *Curr Mol Med*, 6(2):141–153, Mar 2006. 207, 225

181. T. Planche, A. Dzeing, E. Ngou-Milama, M. Kombila, and P. W. Stacpoole. Metabolic complications of severe malaria. *Curr Top Microbiol Immunol*, 295:105–136, 2005. 207, 225

182. C. V. Plowe, J. F. Cortese, A. Djimde, O. C. Nwanyanwu, W. M. Watkins, P. A. Winstanley, J. G. Estrada-Franco, R. E. Mollinedo, J. C. Avila, J. L. Cespedes, D. Carter, and O. K. Doumbo. Mutations in plasmodium falciparum dihydrofolate reductase and dihydropteroate synthase and epidemiologic patterns of pyrimethamine-sulfadoxine use and resistance. *J Infect Dis*, 176(6):1590–1596, Dec 1997. 19, 126, 132, 134

183. S. D. Polley and D. J. Conway. Strong diversifying selection on domains of the plasmodium falciparum apical membrane antigen 1 gene. *Genetics*, 158 (4):1505–1512, Aug 2001. 30

184. Z. Premji, Y. Hamisi, C. Shiff, J. Minjas, P. Lubega, and C. Makwaya. Anaemia and plasmodium falciparum infections among young children in an holoendemic area, bagamoyo, tanzania. *Acta Trop*, 59(1):55–64, Mar 1995. 3

185. R. N. Price, C. Cassar, A. Brockman, M. Duraisingh, M. van Vugt, N. J. White, F. Nosten, and S. Krishna. The pfmdr1 gene is associated with a multidrug-resistant phenotype in plasmodium falciparum from the western border of thailand. *Antimicrob Agents Chemother*, 43(12):2943–2949, Dec 1999. 18

186. Ric N Price, Anne-Catrin Uhlemann, Alan Brockman, Rose McGready, Elizabeth Ashley, Lucy Phaipun, Rina Patel, Kenneth Laing, Sornchai Looareesuwan, Nicholas J White, FranÃğois Nosten, and Sanjeev Krishna. Mefloquine resistance in plasmodium falciparum and increased pfmdr1 gene copy

number. *Lancet*, 364(9432):438–447, 2004. doi: 10.1016/S0140-6736(04)16767-6. URL http://dx.doi.org/10.1016/S0140-6736(04)16767-6. 18

187. Ric N Price, Anne-Catrin Uhlemann, Michele van Vugt, Al Brockman, Robert Hutagalung, Shalini Nair, Denae Nash, Pratap Singhasivanon, Tim J C Anderson, Sanjeev Krishna, Nicholas J White, and FranÃğois Nosten. Molecular and pharmacological determinants of the therapeutic response to artemether-lumefantrine in multidrug-resistant plasmodium falciparum malaria. *Clin Infect Dis*, 42(11):1570–1577, Jun 2006. doi: 10.1086/503423. URL http://dx.doi.org/10.1086/503423. 18

188. S. Proux, L. Hkirijareon, C. Ngamngonkiri, S. McConnell, and F. Nosten. Paracheck-pf: a new, inexpensive and reliable rapid test for p. falciparum malaria. *Trop Med Int Health*, 6(2):99–101, Feb 2001. 35, 82

189. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, ISBN 3-900051-07-0, http://www.R-project.org, 2010. 51, 123, 146

190. F. G. Razakandrainibe, P. Durand, J. C. Koella, T. De MeeÃijs, F. Rousset, F. J. Ayala, and F. Renaud. "clonal" population structure of the malaria agent plasmodium falciparum in high-infection regions. *Proc Natl Acad Sci U S A*, 102(48):17388–17393, Nov 2005. doi: 10.1073/pnas.0508871102. URL http://dx.doi.org/10.1073/pnas.0508871102. 233

191. J. C. Reeder, A. F. Cowman, K. M. Davern, J. G. Beeson, J. K. Thompson, S. J. Rogerson, and G. V. Brown. The adhesion of plasmodium falciparum-infected erythrocytes to chondroitin sulfate a is mediated by p. falciparum erythrocyte membrane protein 1. *Proc Natl Acad Sci U S A*, 96(9):5198–5202, Apr 1999. 17

192. Mario H Rodriguez and Fidel de la C HernÃąndez-HernÃąndez. Insect-malaria parasites interactions: the salivary gland. *Insect Biochem Mol Biol*, 34(7):615–624, Jul 2004. doi: 10.1016/j.ibmb.2004.03.014. URL http://dx.doi.org/10.1016/j.ibmb.2004.03.014. 3, 4

193. S. J. Rogerson, R. Tembenu, C. DobaÃśo, S. Plitt, T. E. Taylor, and M. E. Molyneux. Cytoadherence characteristics of plasmodium falciparum-infected erythrocytes from malawian children with severe and uncomplicated malaria. *Am J Trop Med Hyg*, 61(3):467–472, Sep 1999. 17

194. Cally Roper, Richard Pearce, Barry Bredenkamp, Jonathan Gumede, Chris Drakeley, Frank Mosha, Daniel Chandramohan, and Brian Sharp. Antifolate antimalarial resistance in southeast africa: a population-based analysis. *Lancet*, 361(9364):1174–1181, Apr 2003. doi: 10.1016/S0140-6736(03)12951-0. URL http://dx.doi.org/10.1016/S0140-6736(03)12951-0. 10, 120, 121, 131, 132, 209

195. A. Rowe, J. Obeiro, C. I. Newbold, and K. Marsh. Plasmodium falciparum rosetting is associated with malaria severity in kenya. *Infect Immun*, 63(6):2323–2326, Jun 1995. 17

196. J. A. Rowe, J. M. Moulds, C. I. Newbold, and L. H. Miller. P. falciparum rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature*, 388(6639):292–295, Jul 1997. doi: 10.1038/40888. URL http://dx.doi.org/10.1038/40888. 16, 17

197. J. Alexandra Rowe, Jack Obiero, Kevin Marsh, and Ahmed Raza. Short report: Positive correlation between rosetting and parasitemia in plasmodium falciparum clinical isolates. *Am J Trop Med Hyg*, 66(5):458–460, May 2002. 171

198. J. Alexandra Rowe, Juma Shafi, Oscar K Kai, Kevin Marsh, and Ahmed Raza. Nonimmune igm, but not igg binds to the surface of plasmodium falciparum-infected erythrocytes and correlates with rosetting and severe malaria. *Am J Trop Med Hyg*, 66(6):692–699, Jun 2002. 17

199. J. Alexandra Rowe, D. Herbert Opi, and Thomas N Williams. Blood groups and malaria: fresh insights into pathogenesis and identification of targets for intervention. *Curr Opin Hematol*, 16(6):480–487, Nov 2009. doi: 10.1097/MOH.0b013e3283313de0. URL http://dx.doi.org/10.1097/MOH.0b013e3283313de0. 11

200. Pardis C Sabeti, David E Reich, John M Higgins, Haninah Z P Levine, Daniel J Richter, Stephen F Schaffner, Stacey B Gabriel, Jill V Platko, Nick J Patterson, Gavin J McDonald, Hans C Ackerman, Sarah J Campbell, David Altshuler, Richard Cooper, Dominic Kwiatkowski, Ryk Ward, and Eric S Lander. Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419(6909):832–837, Oct 2002. doi: 10.1038/nature01140. URL http://dx.doi.org/10.1038/nature01140. 30

201. J. R. Schellenberg, S. Abdulla, R. Nathan, O. Mukasa, T. J. Marchant, N. Kikumbih, A. K. Mushi, H. Mponda, H. Minja, H. Mshinda, M. Tanner, and C. Lengeler. Effect of large-scale social marketing of insecticide-treated nets on child survival in rural tanzania. *Lancet*, 357(9264):1241–1247, Apr 2001. doi: 10.1016/S0140-6736(00)04404-4. URL http://dx.doi.org/10.1016/S0140-6736(00)04404-4. 8, 9

202. Achim G Schneider, Zulfikar Premji, Ingrid Felger, Tom Smith, Salim Abdulla, Hans-Peter Beck, and Hassan Mshinda. A point mutation in codon 76 of pfcrt of p. falciparum is positively selected for by chloroquine treatment

in tanzania. *Infect Genet Evol*, 1(3):183–189, May 2002. 6

203. Shelley Segal and Adrian V S Hill. Genetic susceptibility to infectious disease. *Trends Microbiol*, 11(9):445–448, Sep 2003. 171

204. Jay Shendure and Hanlee Ji. Next-generation dna sequencing. *Nat Biotechnol*, 26(10):1135–1145, Oct 2008. doi: 10.1038/nbt1486. URL http://dx.doi.org/10.1038/nbt1486. 22, 23

205. Irwin W Sherman, Shigetoshi Eda, and Enrique Winograd. Cytoadherence and sequestration in plasmodium falciparum: defining the ties that bind. *Microbes Infect*, 5(10):897–909, Aug 2003. 171

206. Balbir Singh, Lee Kim Sung, Asmad Matusop, Anand Radhakrishnan, Sunita S G Shamsul, Janet Cox-Singh, Alan Thomas, and David J Conway. A large focus of naturally acquired plasmodium knowlesi infections in human beings. *Lancet*, 363(9414):1017–1024, Mar 2004. doi: 10.1016/S0140-6736(04)15836-4. URL http://dx.doi.org/10.1016/S0140-6736(04)15836-4. 2

207. W. Sirawaraporn, T. Sathitkul, R. Sirawaraporn, Y. Yuthavong, and D. V. Santi. Antifolate-resistant mutants of plasmodium falciparum dihydrofolate reductase. *Proc Natl Acad Sci U S A*, 94(4):1124–1129, Feb 1997. 19

208. W. Sirawaraporn, S. Yongkiettrakul, R. Sirawaraporn, Y. Yuthavong, and D. V. Santi. Plasmodium falciparum: asparagine mutant at residue 108 of dihydrofolate reductase is an optimal antifolate-resistant single mutant. *Exp Parasitol*, 87(3):245–252, Nov 1997. doi: 10.1006/expr.1997.4221. URL http://dx.doi.org/10.1006/expr.1997.4221. 19

209. Sodiomon Bienvenu Sirima and Adama GansanÃĺ. Artesunate-amodiaquine for the treatment of uncomplicated malaria. *Expert Opin Investig Drugs*,

16(7):1079–1085, Jul 2007. doi: 10.1517/13543784.16.7.1079. URL http://dx.doi.org/10.1517/13543784.16.7.1079. 75

210. T. Smith, H. P. Beck, A. Kitua, S. Mwankusye, I. Felger, N. Fraser-Hurt, A. Irion, P. Alonso, T. Teuscher, and M. Tanner. Age dependence of the multiplicity of plasmodium falciparum infections and of other malariological indices in an area of high endemicity. *Trans R Soc Trop Med Hyg*, 93 Suppl 1:15–20, Feb 1999. 6, 220

211. G. Snounou and H. P. Beck. The use of pcr genotyping in the assessment of recrudescence or reinfection after antimalarial drug treatment. *Parasitol Today*, 14(11):462–467, Nov 1998. 114

212. R. W. Snow, M. Craig, U. Deichmann, and K. Marsh. Estimating mortality, morbidity and disability due to malaria among africa's non-pregnant population. *Bull World Health Organ*, 77(8):624–640, 1999. 2

213. Robert W Snow, Carlos A Guerra, Abdisalan M Noor, Hla Y Myint, and Simon I Hay. The global distribution of clinical episodes of plasmodium falciparum malaria. *Nature*, 434(7030):214–217, Mar 2005. doi: 10.1038/nature03342. URL http://dx.doi.org/10.1038/nature03342. 7

214. R.W. Snow and K. Marsh. Will reducing plasmodium falciparum transmission alter malaria mortality among african children? *Parasitol. Today*, 11: 188–190, 1995. 3

215. Kanlaya Sriprawat, Supaporn Kaewpongsri, Rossarin Suwanarusk, Mara L Leimanis, Usa Lek-Uthai, Aung Pyae Phyo, Georges Snounou, Bruce Russell, Laurent Renia, and FranÃğois Nosten. Effective and cheap removal of leukocytes and platelets from plasmodium vivax infected blood. *Malar J*, 8:115, 2009. doi: 10.1186/1475-2875-8-115. URL http://dx.doi.org/10.

1186/1475-2875-8-115. 38, 76

216. R. W. Steketee, B. L. Nahlen, M. E. Parise, and C. Menendez. The burden of malaria in pregnancy in malaria-endemic areas. *Am J Trop Med Hyg*, 64 (1-2 Suppl):28–35, 2001. 10

217. Mary M Stevenson and Eleanor M Riley. Innate immunity to malaria. *Nat Rev Immunol*, 4(3):169–180, Mar 2004. doi: 10.1038/nri1311. URL http://dx.doi.org/10.1038/nri1311. 207

218. Xinzhuan Su, Karen Hayton, and Thomas E Wellems. Genetic linkage and association analyses for trait mapping in plasmodium falciparum. *Nat Rev Genet*, 8(7):497–506, Jul 2007. doi: 10.1038/nrg2126. URL http://dx.doi.org/10.1038/nrg2126. 2, 3, 4

219. Colin J Sutherland, Naowarat Tanomsing, Debbie Nolder, Mary Oguike, Charlie Jennison, Sasithon Pukrittayakamee, Christiane Dolecek, Tran Tinh Hien, Virgilio E do RosÃąrio, Ana Paula Arez, JoÃčo Pinto, Pascal Michon, Ananias A Escalante, Francois Nosten, Martina Burke, Rogan Lee, Marie Blaze, Thomas Dan Otto, John W Barnwell, Arnab Pain, John Williams, Nicholas J White, Nicholas P J Day, Georges Snounou, Peter J Lockhart, Peter L Chiodini, Mallika Imwong, and Spencer D Polley. Two nonrecombining sympatric forms of the human malaria parasite plasmodium ovale occur globally. *J Infect Dis*, 201(10):1544–1550, May 2010. doi: 10.1086/652240. URL http://dx.doi.org/10.1086/652240. 2

220. Ambrose O Talisuna, Peter Bloland, and Umberto D'Alessandro. History, dynamics, and public health importance of malaria parasite resistance. *Clin Microbiol Rev*, 17(1):235–254, Jan 2004. 8

221. S Clinical Trials Partnership The RTS. A phase 3 trial of rts,s/as01

malaria vaccine in african infants. *New England Journal of Medicine*, 0: doi:10.1056/NEJMoa1208394, 2012. 235

222. Paulina Tindana, Susan Bull, Lucas Amenga-Etego, Jantina de Vries, Raymond Aborigo, Kwadwo Koram, Dominic Kwiatkowski, and Michael Parker. Seeking consent to genetic and genomic research in a rural ghanaian setting: A qualitative study of the malariagen experience. *BMC Med Ethics*, 13(1):15, Jul 2012. doi: 10.1186/1472-6939-13-15. URL http://dx.doi.org/10.1186/1472-6939-13-15. 30, 52

223. H. Tinto, E. B. Zoungrana, S. O. Coulibaly, J. B. Ouedraogo, M. TraorÃĭ, T. R. Guiguemde, E. Van Marck, and U. D'Alessandro. Chloroquine and sulphadoxine-pyrimethamine efficacy for uncomplicated malaria treatment and haematological recovery in children in bobo-dioulasso, burkina faso during a 3-year period 1998-2000. *Trop Med Int Health*, 7(11):925–930, Nov 2002. 10, 18

224. T. Triglia, J. G. Menting, C. Wilson, and A. F. Cowman. Mutations in dihydropteroate synthase are responsible for sulfone and sulfonamide resistance in plasmodium falciparum. *Proc Natl Acad Sci U S A*, 94(25):13944–13949, Dec 1997. 19, 20, 121, 126, 132, 209

225. Daria Van Tyne, Daniel J Park, Stephen F Schaffner, Daniel E Neafsey, Elaine Angelino, Joseph F Cortese, Kayla G Barnes, David M Rosen, Amanda K Lukens, Rachel F Daniels, Danny A Milner, Charles A Johnson, Ilya Shlyakhter, Sharon R Grossman, Justin S Becker, Daniel Yamins, Elinor K Karlsson, Daouda Ndiaye, Ousmane Sarr, Souleymane Mboup, Christian Happi, Nicholas A Furlotte, Eleazar Eskin, Hyun Min Kang, Daniel L Hartl, Bruce W Birren, Roger C Wiegand, Eric S Lander, Dyann F Wirth,

Sarah K Volkman, and Pardis C Sabeti. Identification and functional validation of the novel antimalarial resistance locus $pf10_0355$ in plasmodium falciparum. *PLoS Genet*, 7(4):e1001383, Apr 2011. doi: 10.1371/journal.pgen.1001383. URL http://dx.doi.org/10.1371/journal.pgen.1001383. 17

226. Lyann M B Ursos and Paul D Roepe. Chloroquine resistance in the malarial parasite, plasmodium falciparum. *Med Res Rev*, 22(5):465–491, Sep 2002. doi: 10.1002/med.10016. URL http://dx.doi.org/10.1002/med.10016. 18

227. Manijeh Vafa, Marita Troye-Blomberg, Judith Anchang, AndrÃľ Garcia, and Florence Migot-Nabias. Multiplicity of plasmodium falciparum infection in asymptomatic children in senegal: relation to transmission, age and erythrocyte variants. *Malar J*, 7:17, 2008. doi: 10.1186/1475-2875-7-17. URL http://dx.doi.org/10.1186/1475-2875-7-17. 6

228. Tom van Gool, Marlies E van Wolfswinkel, Rob Koelewijn, Pieter P A M van Thiel, Jan Jacobs, Jaap J van Hellemond, and Perry J J van Genderen. A simple and fast method to exclude high plasmodium falciparum parasitaemia in travellers with imported malaria. *Malar J*, 10:300, 2011. doi: 10.1186/1475-2875-10-300. URL http://dx.doi.org/10.1186/1475-2875-10-300. 6

229. F. Verra, G. Bancone, P. Avellino, I. Blot, J. SimporÃľ, and D. Modiano. Haemoglobin c and s in natural selection against plasmodium falciparum malaria: a plethora or a single shared adaptive mechanism? *Parassitologia*, 49(4):209–213, Dec 2007. 30

230. Federica Verra, Gaia Luoni, Carlo Calissano, Marita Troye-Blomberg, Peter Perlmann, Hedvig Perlmann, Bruno ArcÃă, Bienvenu Sodiomon Sir-

ima, Amadou KonatÃĺ, Mario Coluzzi, Dominic Kwiatkowski, and David Modiano. Il4-589c/t polymorphism and ige levels in severe malaria. *Acta Trop*, 90(2):205–209, Apr 2004. doi: 10.1016/j.actatropica.2003.11.014. URL http://dx.doi.org/10.1016/j.actatropica.2003.11.014. 207

231. S. Viriyakosol, N. Siripoon, C. Petcharapirat, P. Petcharapirat, W. Jarra, S. Thaithong, K. N. Brown, and G. Snounou. Genotyping of plasmodium falciparum isolates by the polymerase chain reaction and potential uses in epidemiological studies. *Bull World Health Organ*, 73(1):85–95, 1995. 114

232. Sarah K Volkman, Pardis C Sabeti, David DeCaprio, Daniel E Neafsey, Stephen F Schaffner, Danny A Milner, Johanna P Daily, Ousmane Sarr, Daouda Ndiaye, Omar Ndir, Soulyemane Mboup, Manoj T Duraisingh, Amanda Lukens, Alan Derr, Nicole Stange-Thomann, Skye Waggoner, Robert Onofrio, Liuda Ziaugra, Evan Mauceli, Sante Gnerre, David B Jaffe, Joanne Zainoun, Roger C Wiegand, Bruce W Birren, Daniel L Hartl, James E Galagan, Eric S Lander, and Dyann F Wirth. A genome-wide map of diversity in plasmodium falciparum. *Nat Genet*, 39(1):113–119, Jan 2007. doi: 10.1038/ng1930. URL http://dx.doi.org/10.1038/ng1930. 14, 15, 29, 80

233. P. Wang, C. S. Lee, R. Bayoumi, A. Djimde, O. Doumbo, G. Swedberg, L. D. Dao, H. Mshinda, M. Tanner, W. M. Watkins, P. F. Sims, and J. E. Hyde. Resistance to antifolates in plasmodium falciparum monitored by sequence analysis of dihydropteroate synthetase and dihydrofolate reductase alleles in a large number of field samples of diverse origins. *Mol Biochem Parasitol*, 89(2):161–177, Nov 1997. 121

234. Ping Wang, Niroshini Nirmalan, Qi Wang, Paul F G Sims, and John E

Hyde. Genetic and metabolic analysis of folate salvage in the human malaria parasite plasmodium falciparum. *Mol Biochem Parasitol*, 135(1):77–87, May 2004. 19

235. D. J. Weatherall and J. B. Clegg. Inherited haemoglobin disorders: an increasing global health problem. *Bull World Health Organ*, 79(8):704–712, 2001. 12

236. T. E. Wellems, A. Walker-Jonah, and L. J. Panton. Genetic mapping of the chloroquine-resistance locus on plasmodium falciparum chromosome 7. *Proc Natl Acad Sci U S A*, 88(8):3382–3386, Apr 1991. 17

237. W. H. Wernsdorfer and D. Payne. The dynamics of drug resistance in plasmodium falciparum. *Pharmacol Ther*, 50(1):95–121, 1991. 10

238. Nicholas J White. Antimalarial drug resistance. *J Clin Invest*, 113(8):1084–1092, Apr 2004. doi: 10.1172/JCI21682. URL http://dx.doi.org/10.1172/JCI21682. 75, 171, 208

239. Nicholas J White, Wirichada Pongtavornpinyo, Richard J Maude, Sompob Saralamba, Ricardo Aguas, Kasia Stepniewska, Sue J Lee, Arjen M Dondorp, Lisa J White, and Nicholas P J Day. Hyperparasitaemia and low dosing are an important source of anti-malarial drug resistance. *Malar J*, 8:253, 2009. doi: 10.1186/1475-2875-8-253. URL http://dx.doi.org/10.1186/1475-2875-8-253. 75

240. WHO. Severe plasmodium falciparum malaria. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 94 (supplement 1):135–228, 2000. 173

241. WHO. World malaria report, 2008. 2

242. WHO. World malaria report 2011, 2011. 2

243. WHO and RBM. Position of who's roll back malaria department on malaria treatment policy. page 10âĂŞ44, 2000. 10

244. Thomas N Williams. Human red blood cell polymorphisms and malaria. *Curr Opin Microbiol*, 9(4):388–394, Aug 2006. doi: 10.1016/j.mib.2006.06.009. URL http://dx.doi.org/10.1016/j.mib.2006.06.009. 11, 12, 13

245. Thomas N Williams, Tabitha W Mwangi, David J Roberts, Neal D Alexander, David J Weatherall, Sammy Wambua, Moses Kortok, Robert W Snow, and Kevin Marsh. An immune basis for malaria protection by the sickle cell trait. *PLoS Med*, 2(5):e128, May 2005. doi: 10.1371/journal.pmed.0020128. URL http://dx.doi.org/10.1371/journal.pmed.0020128. 13, 171

246. Thomas N Williams, Tabitha W Mwangi, Sammy Wambua, Neal D Alexander, Moses Kortok, Robert W Snow, and Kevin Marsh. Sickle cell trait and the risk of plasmodium falciparum malaria and other childhood diseases. *J Infect Dis*, 192(1):178–186, Jul 2005. doi: 10.1086/430744. URL http://dx.doi.org/10.1086/430744. 13, 206, 210, 224

247. C. M. Wilson, A. E. Serrano, A. Wasley, M. P. Bogenschutz, A. H. Shankar, and D. F. Wirth. Amplification of a gene related to mammalian mdr genes in drug-resistant plasmodium falciparum. *Science*, 244(4909):1184–1186, Jun 1989. 18

248. C. J. Woodrow, R. K. Haynes, and S. Krishna. Artemisinins. *Postgrad Med J*, 81(952):71–78, Feb 2005. doi: 10.1136/pgmj.2004.028399. URL http://dx.doi.org/10.1136/pgmj.2004.028399. 11

249. World Health Organization. Monitoring antimalarial drug resistance , 2000. 10

250. World Health Organization. World Malaria Report, 2009. 2

251. Y. Wu, L. A. Kirkman, and T. E. Wellems. Transformation of plasmodium falciparum malaria parasites by homologous integration of plasmids that confer resistance to pyrimethamine. *Proc Natl Acad Sci U S A*, 93(3):1130–1134, Feb 1996. 131

252. Guicheng Zhang, Maria Nelia Manaca, Michelle McNamara-Smith, Alfredo Mayor, Augusto Nhabomba, Tamara Katherine Berthoud, Siew-Kim Khoo, Selma Wiertsema, Ruth Aguilar, Arnoldo Barbosa, LlorenÃğ QuintÃş, Pierre Candelaria, En Nee Schultz, Catherine M Hayden, Jack Goldblatt, Caterina Guinovart, Pedro L Alonso, Peter N LesouÃńf, and Carlota DobaÃśo. Interleukin-10 (il-10) polymorphisms are associated with il-10 production and clinical malaria in young children. *Infect Immun*, 80(7): 2316–2322, Jul 2012. doi: 10.1128/IAI.00261-12. URL http://dx.doi.org/10.1128/IAI.00261-12. 207

253. L. Zhang, X. Cui, K. Schmitt, R. Hubert, W. Navidi, and N. Arnheim. Whole genome amplification from a single cell: implications for genetic analysis. *Proc Natl Acad Sci U S A*, 89(13):5847–5851, Jul 1992. 43, 141

254. J. Zwetyenga, C. Rogier, A. Tall, D. Fontenille, G. Snounou, J. F. Trape, and O. Mercereau-Puijalon. No influence of age on infection complexity and allelic distribution in plasmodium falciparum infections in ndiop, a senegalese village with seasonal, mesoendemic malaria. *Am J Trop Med Hyg*, 59 (5):726–735, Nov 1998. 6, 220