Genetic Profiling of Drug Resistance in *Plasmodium falciparum*

Laura K. Certain

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2007

Program Authorized to Offer Degree:
Department of Genome Sciences

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Laura K. Certain

and have found that it is complete and satisfactory in all respects, and that any and all revisions required by the final examining committee have been made.

Chair of the Supervisory Committee:

Carol H. Sibley

Reading Committee:

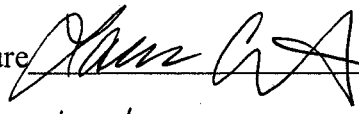
Marilyn C. Roberts

Carol H. Sibley

Wesley C. Van Voorhis

Date: 6/19/2007

In presenting this dissertation in partial fulfillment of the requirements for the doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of the dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, 1-800-521-0600, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature 

Date 6/21/07

University of Washington

Abstract

Genetic Profiling of Drug Resistance in *Plasmodium falciparum*

Laura K. Certain

Chair of the Supervisory Committee:
Professor Carol H. Sibley
Genome Sciences

Malaria is an infectious febrile illness caused by four species of *Plasmodium* parasite, with *Plasmodium falciparum* responsible for the greatest number of fatalities. Unfortunately, drug resistant strains of *P. falciparum* are increasingly prevalent. Resistance to sulfadoxine-pyrimethamine (SP) is due to point mutations in the gene that encodes dihydrofolate reductase (*dhfr*). Patients infected with a parasite carrying three mutations in *dhfr* (N51I/C59R/S108N) are at elevated risk of failing SP treatment. Prior studies of the extended haplotype encompassing *dhfr* suggest that a single triple-mutant allele of *dhfr* emerged in Asia and spread to Africa. However, it is unclear whether this "Asian" strain replaced triple-mutants that had previously evolved in Africa, or simply invaded a population devoid of triple-mutants. In addition, prior studies were hampered by an inability to study infections containing DNA from multiple parasite strains. This dissertation presents a novel method for using yeast to separate haplotypes in mixed malaria infections. Second, it investigates the history of the Asian strain in Africa by analyzing samples collected in Kilifi, Kenya, from 1987-88 and 1993-95, periods immediately before and after the widespread use of SP. We genotyped each sample at *dhfr* and flanking microsatellite loci. All of the triple-mutants had the same haplotype, and it matched the haplotype of the Asian strain from previous studies. The wild-type parasites had a variety of haplotypes, none related to the triple-mutant haplotype. Each double-mutant (N51I/S108N or C59R/S108N) had a single haplotype. Both of the double-mutant haplotypes shared some alleles with wild-type samples and with each other, but neither shared any alleles with the triple-mutant haplotype. In addition, the Asian

triple-mutant was present in 1988, well before the widespread use of SP in Kenya. These results indicate that the Asian triple-mutant arrived in Kenya before any SP use and rose to high frequency as soon as SP was introduced.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation would not have been possible without help from many people. First, let me acknowledge those who provided scientific advice and equipment. Tim Anderson and Shalini Nair let me spend a week with them in San Antonio and introduced me to the wonderful world of *P. falciparum* microsatellites. Ann Riddle did the microsatellite analysis described in Chapter Two, and Joelle Thomas helped me pin all those yeast colonies. Sandra Pennington and M. K. Raghuraman answered innumerable questions about yeast, as did Jenny Madeoy about sequencing, Steve Salipante about microsatellite analysis, and Josh Akey about population genetics. I used equipment belonging to the Swanson, Akey, Green, and Waterston labs, as well as whichever lab owned the 30 °C shaker on the first floor of the J-wing.

I would like to thank my thesis committee for providing challenge and support. In particular, I thank Dr. Wes Van Voorhis and Prof. Marilyn Roberts for reading my dissertation and making valuable comments. I would also like to thank fellow members of the Sibley Lab for creating a positive work environment. In particular I would like to acknowledge my bay-mate in Foege, Marnie Briceño, who – in addition to generating data for this dissertation and forthcoming papers – was extremely tolerant of my tendency to chat while avoiding work. Finally, I would like to acknowledge Carol Sibley as a phenomenal thesis advisor. She was always willing to talk with me, and discussions with her always made me feel more excited about my project and about science in general.

Though the people mentioned above made my graduate school experience a positive one, research always has its ups and downs. The following friends helped see me through the roller coaster: fellow graduate students Julia Jarvis, Liila Woods, Joanna Kelley, and Cindy Desmarais; Maureen Holstad and others in the MSTP; my running partner Kristin Larson; Bayta Maring and the rest of the Esoterics; and my best friends from childhood, Kathy Nagel and Julie Bannerman.

The support of my over-educated family has been invaluable. My brother and his wife fed me dinner far too many times, while their sons helped me keep life in perspective. My sister is my medical school advisor, as well as my life-long role model. My parents, both former graduate students, are the reason I am where I am today. They were the first people I called when things went particularly well (or horribly) in lab. Finally, I thank my husband, Benjamin Goldstein. He not only supported me through this whole process, putting up with a rather tense bride at times, he also read my entire dissertation, catching typos, grammar mistakes, and logical inconsistencies. I am indeed lucky.

The following sources provided financial support: University of Washington Medical Scientist Training Program, University of Washington Department of Genome Sciences, the ARCS Foundation, a Poncin scholarship, the University of Washington Graduate School, a grant from the NIH to Dr. Sibley.

CHAPTER ONE: INTRODUCTION & BACKGROUND

Introduction – Malaria and Drug Resistance

Malaria is an infectious, febrile illness prevalent throughout most of the tropics (**Figure 1.1**). There are over 500 million cases each year and at least one million deaths, with most of these deaths occurring among young children in sub-Saharan Africa [1]. It has been estimated that malaria costs Africa $12 billion every year in lost Gross Domestic Product (GDP), and that it reduces economic growth by 1.3% [2]. In addition, recent modeling studies indicate that malaria infection can increase the spread of HIV; malaria infection increases viral load and thus increases the probability of HIV transmission [3]. Clearly, malaria is one of the most significant public health challenges worldwide. Thanks in large part to the Bill and Melinda Gates Foundation, money to treat, prevent, and study malaria has increased significantly over the past decade [4]. Nevertheless, it remains the most common cause of death in African children under the age of five [2]. In fact, the proportion of childhood deaths in Africa due to malaria increased from 18% in the 1980s to ~30% in the 1990s, presumably due to the parasite's increasing resistance to drugs used to treat malaria [5].

Malaria is caused by four species of blood parasites: *Plasmodium falciparum, P. vivax, P. malariae* and *P. ovale.* Though *P. vivax* is the most prevalent, *P. falciparum* is the most virulent and is therefore the best studied. All four species are transmitted by mosquitoes of the genus *Anopheles*. In the human host, the parasite causes disease by digesting hemoglobin and other red blood cell proteins. The infected erythrocytes lyse or are cleared by the spleen, resulting in anemia. In addition, the erythrocytic lysis causes the release of cytokines that suppress hematopoiesis, exacerbating the anemia. *P. falciparum* causes further harm by inducing the formation of "sticky knobs" on the surface of the erythrocytes; these erythrocytes then bind to the endothelial cells in the microvasculature, blocking blood flow [6].

The classic symptom of malaria infection is intermittent fever, coincident with the rupture of the erythrocytes. In addition to fever, clinical malaria is characterized by a variety of non-specific symptoms, including chills, headache, fatigue, muscle pains,

nausea, vomiting and cough [6,7]. With time, anemia, thrombocytopenia, splenomegaly, hepatomegaly, and jaundice can occur [6]. Left untreated, *P. falciparum* infection can lead to death from cerebral malaria or severe anemia.

Those most at risk for severe complications are non-immune adults (e.g. tourists from non-endemic areas), children under the age of five, and pregnant women. In places where malaria is hyper-endemic, people gradually acquire some immunity to the parasite. Thus, while the young children living in these areas can become severely ill from infection, the adults rarely do. Parasitemia, the proportion of erythrocytes infected, is also lower in adults. Pregnant women, however, appear to lose some immunity (or are subject to infection by a different population of parasites) and are at increased risk for developing high parasitemia and clinical malaria [8]. Depending on the level of immunity of the mother (corresponding to the level of endemicity of malaria), this high parasitemia can cause anemia, hypoglycemia, and other manifestations of clinical malaria. Primigravid women are most at risk [8,9]. In addition to maternal morbidity and mortality, malaria infection during pregnancy can cause spontaneous abortions, pre-term birth, low birth-weight, congenital infection and perinatal death for the infant [8].

There are a variety of methods used to prevent malaria. One simple method is the use of insecticide treated bed nets (ITNs). Malaria is transmitted by female *Anopheles* mosquitoes; the *A. gambiae, A. funestus, A. nili,* and *A. moucheti* species groups are the main vectors in sub-Saharan Africa [10]. In general, these species bite between dusk and dawn. Therefore, sleeping under a net reduces personal exposure to malaria. Though there was initially concern that distributing ITNs would delay children's acquisition of immunity to malaria, and thus increase child mortality, this is not the case; ITNs reduce child mortality [11]. Furthermore, ITNs reduce malaria exposure throughout the community, not just among those sleeping under nets, because they reduce the mosquito population [12]. Unfortunately, not all families – particularly in rural areas, where malaria transmission is often higher – can afford ITNs, and therefore usage is low unless they are provided free of charge [13]. Malaria can also be prevented by spraying insecticides to kill the mosquito vector [14].

In addition to preventing malaria by reducing the population of the mosquito vector, one can prevent the disease through pharmacological prophylaxis. Travelers to malaria-endemic regions (typically tourists from the U.S. and Europe) have long been taking doxycycline, atovaquone-proguanil (Malarone®, Glaxo-Smith-Kline, USA), chloroquine, or mefloquine (Lariam®, Roche Pharmaceuticals, USA) to prevent infection while abroad [15]. Because people living in endemic regions develop some immunity to malaria, chemoprophylaxis is not generally used within those populations. Recently, however, researchers have tried intermittent preventive treatment (IPT) to prevent malaria in those at highest risk for developing a life-threatening infection: infants and pregnant women. A meta-analysis of IPTp (IPT in pregnancy) trials indicated that giving primi- or segundigravida women sulfadoxine-pyrimethamine, an inexpensive and well-tolerated antimalarial drug, during pregnancy led to fewer instances of severe anemia and fewer perinatal deaths [16]. Likewise, administering antimalarials to infants (IPTi; IPT in infants) reduces infant malaria, anemia, and hospital admissions [17].

Though much effort has gone into the development of a vaccine for malaria, thus far there are none available. Part of the difficulty stems from the fact that even a natural malaria infection does not lead to lifelong immunity. Though people living in endemic regions eventually acquire immunity, it comes at the expense of significant childhood mortality, and if they move to a non-endemic region they lose immunity. Furthermore, exactly which of the numerous and variable parasite antigens lead to immunity is not understood [18]. However, there is currently one vaccine, the RTS,S vaccine (GlaxoSmithKline), advancing to Phase III clinical trials [19]. Phase II trials in children aged 1-4 in Mozambique indicated that the vaccine was safe and immunogenic. During the first six months of the study, the vaccine reduced risk of clinical malaria by 30%, time to first infection by 45%, and incidence of severe malaria by 58% [20]. After another twelve months of follow-up, the vaccine continued to provide some protection [21].

When prevention fails, there are various treatments for malaria, including quinine, chloroquine, sulfadoxine-pyrimethamine (SP), mefloquine and artemisinin

derivatives (**Table 1.1**) Quinine, made from the bark of the cinchona tree, was first used by Europeans as a treatment for "ague" in the 1600s, after Jesuit priests in South America learned of the bark from the Incas [22-24]. Though there are some reports of quinine resistant malaria [25], in general it remains effective. However, due to its side effects – tinnitus, nausea, vertigo – and its horrible taste, many patients will not take a full course and therefore other antimalarials are preferred [24,26].

Chloroquine and SP are widely used to treat *P. falciparum* infections. They are safe, well-tolerated, and affordable. Treating a case of malaria with one of these two drugs costs less than $0.20 [24]. Unfortunately, resistance is emerging, such that in many places these drugs are ineffective. Drug resistance, as defined by the World Health Organization, is "the ability of a parasite strain to survive and/or multiply despite the administration and absorption of a drug given in doses equal to or higher than those usually recommended but within tolerance of the subject." [27] Tests for resistance can either monitor therapeutic efficacy in the patient (*in vivo* tests; **Table 1.2**), or test the parasite directly (*in vitro* tests). To test the parasite directly, it is cultured in the presence of varying concentrations of drug and the drug concentration required for growth inhibition is determined. In addition, one can test for the presence of genetic mutations in the parasite DNA known to confer resistance. Though the exact prevalence of resistant parasites will depend on the method used, in general there is good correlation between *in vivo* and *in vitro* tests [28]. *In vivo* tests are likely to show a lower level of resistance because a patient's immune system may clear the parasite regardless of treatment.

Chloroquine was first used to treat malaria in 1934, with widespread use beginning in the 1950s [23,29]. Treatment failure was first documented in Thailand and Colombia in the late 1950s, and was subsequently seen in Papua New Guinea in the 1960s and in East Africa in the late 1970s [29-31]. By the early 1980s, chloroquine resistance was present on the Kenyan coast, the region of interest for this dissertation [32]. Today, chloroquine is only effective in certain countries in Central America [23,24].

SP, also known by the trade name Fansidar, followed a similar pattern on an accelerated schedule. First used to treat malaria in 1967, resistance emerged within

five years in Asia and is now widespread in that part of the world. SP resistance is also prevalent in the Amazon Basin of South America [23]. In East Africa, SP began to lose efficacy in the late 1980s. Currently, resistance levels vary extensively from country to country in Africa, with treatment failures anywhere from 0-35% [30]. In Kilifi, Kenya, the region of interest for this dissertation, SP was first used in standard clinical practice in 1992-1993 [33,34]. A study of parasite *in vitro* susceptibility to sulfadoxine and pyrimethamine in 1987-88 and 1993-95 indicated that parasites resistant to pyrimethamine, but not sulfadoxine, were present even before the introduction of SP. By 1995 a majority of parasites showed increased *in vitro* resistance to both drugs [33].

Due to increasing resistance to the inexpensive antimalarials chloroquine and SP, some countries have switched to other drugs to treat malaria. Mefloquine was introduced in the 1980s to replace chloroquine, and though resistance exists in Southeast Asia [35], it is still effective in most of the world. It costs about twenty times as much as chloroquine or SP [24]. Resistance appears to be due to an increase in copy number of the gene *pfmdr1* (plasmodium falciparum multi-drug resistance), which codes for a transmembrane protein that localizes to the parasite digestive vacuole, and is a homolog of a mammalian protein responsible for drug resistance in tumor cells [36]. *Pfmdr1* is also implicated in resistance to lumefantrine, halofantrine, quinine, and artemisinin [36].

The last of these, artemisinin, is a new drug that has raised hopes that we will stay ahead of the parasite a little longer in the drug resistance arms race. Extracts from the *Artemisia annua* plant have been used for over a thousand years in China for the treatment of fevers [23]. In 1971 the active component, artemisinin, was characterized, and several artemisinin derivatives have been shown to be effective in treating malaria that is resistant to more common drugs [37]. The half-life of these derivatives is very short, which should make it more difficult for the parasite to acquire resistance; pairing an artemisinin derivative with a different antimalarial should further reduce the selection of drug resistance [38]. If two drugs have completely different targets, and either drug alone is sufficient to kill the parasite, then the parasite must simultaneously

acquire resistance to both drugs in order to survive. Because genetic mutations conferring resistance are rare events, it is extremely unlikely that a parasite could acquire all necessary mutations at the same time. However, if there is cross-resistance between the drugs then using a combination will be less effective at preventing the development of drug resistance [39]. Moreover, if the half-life of the partner drug is significantly longer than that of artemisinin – as is typically the case – then the parasites are exposed to low levels of the partner drug for an extended period, which may lead to selection of resistance to the partner drug.

Though affordability and supply are impediments [40], WHO now recommends that artemisinin-based combination therapy (ACT, an artemisinin derivative combined with another antimalarial) be the standard treatment for malaria [41]. The artemisinin derivative is artesunate, artemether, or dihydroartemisinin; partner drugs include amodiaquine, mefloquine, lumefantrine, and piperaquine. Artemether-lumefantrine (Co-Artem®; Novartis International AG, Basel, Switzerland) is currently recommended by WHO, though it will perhaps be replaced by dihydroartemisinin-piperaquine (Artekin®; Holleykin Pharmaceutical Co., Ltd., Guangzhou, Guandong, People's Republic of China). The latter formulation is less expensive and does not need to be taken with fat to ensure bioavailability [38,42]. In addition, Sanofi-Aventis (Paris, France) has recently announced that they will provide a new co-formulation of artemisinin and amodiaquine, ASAQ, at cost to international health agencies [43]. Hopefully these developments mean that ACTs will soon be as readily available – and as affordable – as chloroquine or SP.

Motivation for this research

As mentioned above, many malaria parasites have developed resistance to the standard, inexpensive drugs used to treat malaria in the developing world: chloroquine and sulfadoxine-pyrimethamine (SP). We still have effective, though more expensive, antimalarials; however, the parasite will undoubtedly acquire resistance to those as well. In order to delay the selection of resistance to new antimalarial drugs and the spread of resistance to those in current use, we must first understand how resistance

develops and spreads. In particular, we must determine whether resistance emerges once and then spreads over a large geographic area, or whether resistance emerges many times, on many genetic backgrounds. In the case of SP, resistance emerged rapidly after the introduction of the drug. This dissertation presents a study of the emergence of SP resistance in Kilifi, Kenya; it also documents the prevalence of chloroquine resistance in that same community. By studying the development of SP resistance we may learn how to prevent the spread of resistance to new antimalarial drugs.

The emergence and spread of drug resistance can be investigated by studying the population genetics of *Plasmodium falciparum.* By constructing haplotypes from linked genetic markers, researchers can infer the evolutionary history of a resistance allele. Simply put, if parasites share long stretches of DNA then they are likely to share ancestry. Such studies have indicated that resistance in Africa to both chloroquine and SP is due to immigrant parasites from Southeast Asia [44-47]. Though there are several origins of chloroquine resistance worldwide [47], all chloroquine-resistant parasites in Africa appear to be descended from resistant parasites in Southeast Asia [48]. Likewise, only one origin of resistance to SP was identified for Asia and parts of East Africa [46], though a recent study found some evidence of multiple origins in Kenya [49]. Data from more locations, over a longer period of time, are necessary to determine how and when SP resistance appeared and spread in Africa.

Life cycle and population structure of *P. falciparum*

To understand the population genetics of malaria, one must first understand the life cycle of the parasite (**Figure 1.2**) [50]. The mosquito injects sporozoites, haploid forms of the parasite, into the human host. These sporozoites travel to the liver, where they invade hepatocytes and transform into schizonts. Each schizont then produces thousands of merozoites, which invade the bloodstream. Once inside an erythrocyte, a merozoite becomes a trophozoite, which consumes erythrocytic proteins and eventually becomes a schizont, producing more merozoites which rupture the cell and invade more erythrocytes. Occasionally, trophozoites form gametocytes instead of

schizonts. Gametocytes of both mating types develop in the human bloodstream, and are ingested by the mosquito during a blood meal. Thus, all human stages of the malaria parasite are haploid.

Within minutes of ingestion by the mosquito, the gametocytes become gametes and combine, creating a diploid zygote. This brief diploid phase quickly ends with meiosis, and then many rounds of mitosis and a series of developmental changes produce haploid sporozoites, which migrate to the salivary glands. With the next blood meal, these sporozoites are injected into a person and the cycle continues. Thus, the only opportunity for recombination and re-assortment of genetic material (i.e. meiosis) occurs in the gut of the mosquito, after the mosquito ingests gametocytes from an infected person. Because all gametocytes in a clonal infection are genetically identical, rearrangement of genetic material is only observed when one person is infected with multiple strains [51]. As a consequence, the population genetics of *P. falciparum* may be considerably different in areas of high transmission, where mixed infections are common [52].

For example, if resistance is a multigenic trait (e.g. *dhfr* and *dhps* mutations affecting SP resistance; see below) then independent assortment will separate these loci and retard the emergence of resistant parasites. If resistance is monogenic, but requires multiple mutations (e.g. *dhfr* mutations affecting pyrimethamine resistance), recombination can break down alleles that encode highly resistant enzymes. Recombination can also bring resistant alleles together, but until these alleles become extremely prevalent in the population recombination is more likely to separate them than to bring them together. Numerous modeling studies have attempted to determine the effect of factors such as intensity of transmission, level of drug use, and the number of mutations required for resistance on the origin and spread of antimalarial resistance [53-57]. However, most authors conclude by saying that we do not have enough information about the values of important parameters to apply the models in the field [55,56]. In particular, there remains debate over the relationship between the level of transmission and the rate of emergence of drug resistance [58].

The evolutionary history of *P. falciparum* is also debated, with some favoring a recent common ancestor (5000-10,000 years ago [59]), some favoring an ancient common ancestor (>100,000 years ago [60-62]), and some taking intermediate views [63-65]. For reviews of this debate, please see Hartl *et al.* (2002), Hume, Lyons, and Day (2003), and Hartl (2004) [66-68]. In general, those favoring a recent common ancestor point to the low level of diversity in many *P. falciparum* genes, in particular the lack of any synonymous SNPs (single nucleotide polymorphisms) in many genes [59,65,69]. Studies of the human alleles for glucose-6-phosphate dehydrogenase deficiency and sickle cell anemia also lend support to a recent common ancestor. Though homozygosity for either of these alleles causes severe disease, heterozygotes are resistant to malaria. Therefore, one would expect these alleles to have emerged when malaria became a threat to human health. The estimated age of these alleles supports the hypothesis that malaria underwent a population expansion 3000-10,000 years ago [68,70-72]. Agriculture replaced hunting and gathering at roughly this time period, leading to both an increase in human population density and an increase in breeding sites for *Anopheles* mosquitoes, both of which would have led to an increase in *P. falciparum* population size [68].

On the other hand, other studies have found a relatively high level of diversity in *P. falciparum,* and argue that the population size as been large for hundreds of thousands of years [60-62]. Because the chimpanzee malaria, *P. reichenowi*, diverged from *P. falciparum* roughly 6-10 million years ago, at the same time as the human-chimp split, *P. falciparum* has apparently been a human parasite for millions of years [73]. While most would accept this fact, the debate surrounds the issue of whether or not there was a bottleneck roughly 10,000 years ago, at the time of the advent of agriculture. A study of mitochondrial DNA tried to reconcile the two hypotheses, arguing that the population is ancient but expanded 10,000 years ago [64]. A definitive answer will only come with genome-wide analysis of many isolates from around the world [67]. Such studies are underway, but have not yet resolved the issue [74,75].

While the exact history of *P. falciparum* remains elusive, we do know some things about the current population genetics of the parasite. At the most basic, we

know that the *P. falciparum* genome consists of 14 chromosomes totaling 23 Mb, encoding ~5300 genes; it is 80% A/T [76]. SNPs occur at least every kilobase [75], microsatellites (short nucleotide sequences repeated in tandem) at least every 2-3 kb [77]. Beyond these basic facts, researchers have tried to determine population genetic parameters such as effective population size, population substructure (F_{ST}), inbreeding, heterozygosity, recombination rates, and mutation rates for SNPs and microsatellites.

The first major inroad to determining these values came from a laboratory cross between two strains of *P. falciparum*, Dd2 and HB3 [78,79]. Genotyping 901 markers in the progeny of this cross led to a map unit distance of ~17 kb/cM, or a recombination rate of ~6 x 10^{-4} /kb/generation [80]. Recombination appeared to be relatively uniform across the genome. Of course, in natural populations the effective recombination rate will depend on the level of inbreeding; only if different parasite genomes are present in a single blood meal may recombination be observed. Studies of natural populations estimate that the inbreeding coefficient (the probability that the two gametes creating a given diploid parasite are from the same parent) ranges from approximately 0.4-0.9, but that the level of recombination can still be high [81].

In the genetic cross mentioned above, some of the progeny had microsatellite alleles different from either parent, leading to an estimated mutation rate, μ, for microsatellites of 1.59 x 10^{-4} per locus per generation [82]. A mutation rate for SNPs has been estimated from the *P. falciparum-P. reichenowi* divergence as 1.7 x 10^{-9} substitutions per site per generation [60], or roughly 6 x 10^{-9} substitutions per site per year [64].

The first thorough examination of the population structure of *P. falciparum* was a study of 465 samples from 9 locations worldwide, typed at 12 microsatellites [82,83]. Anderson *et al.* (2000) found that population structure varied from place to place, with an expected heterozygosity as low as 0.3 in places with low transmission, like Colombia, and as high as 0.8 in high transmission areas of Africa; F_{ST} was 0.364 in South America and 0.007 in Africa. Likewise, the effective population size ranged from ~700 in South American populations to ~20,000 in African populations,

depending on the location (and hence the variability) and the mutation model used [82]. A separate study using SNPs estimated the worldwide effective population size as 300,000-800,000 [60]. Since this study used only coding regions and published sequences it will likely be replaced by a better estimate as more data become available [74,75,84]. However, since a given infected person carries ~10^{11} parasites, it is not clear what "effective population size" means for *P. falciparum* [85].

Prior studies of the population genetics of drug resistance

In order to study the population genetics of drug resistance in *P. falciparum*, researchers have used microsatellites for genetic markers [44,45,47,86]. Microsatellites in *P. falciparum* are often $(AT)_n$. Polymerase slippage during DNA replication and unequal crossover during meiosis cause microsatellites to change length over time (e.g. $(AT)_7$ becomes $(AT)_{13}$), so that different individuals may have a different number of repeats at a given microsatellite locus. Because each microsatellite can have any number of repeats, a microsatellite has many more alleles than a SNP, which has only four possible states: A, T, C, or G (**Figure 1.3**). Determining the length of several microsatellites allows one to define a haplotype. A haplotype is a collection of genotypes (alleles) existing in the same haploid genome, typically in the same region of a chromosome. To study resistance to a given drug, researchers construct haplotypes using microsatellites that flank the drug resistance genes. By comparing the number of haplotypes present in sensitive strains to those present in resistant strains, researchers can determine how many times drug resistance has emerged.

The rationale is as follows (**Figure 1.4**): prior to drug use, there are a variety of haplotypes in the parasite population. When parasites are exposed to the drug, most die – taking their haplotypes with them. A few, however, carry mutations that allow them to survive and multiply to become the new population. Their haplotypes are then the only haplotypes seen in the population. If two parasites share a haplotype then they are derived from the same ancestral parasite that survived selection. That is, a shared haplotype indicates shared ancestry. If drug resistance has developed on many genetic backgrounds, then one would expect to see a variety of haplotypes in both the

sensitive and the resistant strains (**Figure 1.5**). If, on the other hand, drug resistance emerged once and then spread, one would expect to see only one or two haplotypes in the resistant strains, and to see those haplotypes over a large area.

Malaria researchers first used this method to study the evolution of chloroquine resistance. Chloroquine is believed to kill malaria parasites by interfering with their ability to digest and detoxify hemoglobin. The mechanism of chloroquine resistance is not well understood, but likely involves either efflux of the drug or an altering of pH in the parasite digestive vacuole such that chloroquine becomes less effective [87]. Genetically, chloroquine resistance is due at least partially to point mutations in the gene *pfcrt* (*P. falciparum* chloroquine resistance transporter) on chromosome 7, which encodes an integral membrane protein of the parasite digestive vacuole [88].

A change from a lysine to a threonine at codon 76 in *pfcrt* is strongly correlated with chloroquine resistance. In addition, there are changes at neighboring codons. In wild-type (chloroquine sensitive) parasites, codons 72-76 have the haplotype CVMNK. Chloroquine resistant parasites from Southeast Asia and South America have the haplotypes CVIET or SVMNT at these codons. As mentioned above, chloroquine resistance was seen first in Asia and South America, and subsequently in Africa [29]. Examining microsatellites flanking *pfcrt*, researchers determined that chloroquine resistance arose relatively rarely, with founder events in Southeast Asia, South America, and Papua New Guinea, and then resistance spread from Asia to Africa [47,89-91]. In Africa, all chloroquine resistant parasites studied prior to 2004 carried the CVIET haplotype for *pfcrt* [48,92]. Examining 238 samples from across the continent at *pfcrt* and a nearby microsatellite, Ariey *et al.* (2006) found that all but two of the 125 resistant parasites matched a Cambodian haplotype, while the sensitive parasites had a variety of haplotypes, confirming that chloroquine resistance in Africa is due to resistant parasites imported from Asia [48].

One bright spot in this story of drug resistance is the re-emergence of chloroquine-sensitive parasites once use of the drug is discontinued. In bacteria, it is sometimes observed that mutations conferring antibiotic resistance are deleterious in the absence of selective pressure from drugs (i.e. drug resistance comes at a fitness

cost). Therefore, when drug pressure is removed the wild-type (drug-susceptible) bacteria might out-compete the resistant strains. Unfortunately, bacteria often acquire compensatory mutations, such that the fitness cost of drug resistance disappears and the drug-resistant bacteria survive even in the absence of selective pressure from drug use [93]. If the same holds true in malaria, then one would expect the resistant strains to thrive even after the drug is discontinued. In 1993, Malawi stopped using chloroquine as treatment for malaria. Over the subsequent decade, the prevalence of parasites bearing the K76T mutation decreased until it dropped below the level of detection in 2001 [94]. A recent trial shows that chloroquine is once again effective at treating malaria in Malawi [95]. Unfortunately, while other countries have sometimes seen a decrease in the prevalence of the K76T mutation after removing chloroquine from use, the prevalence has not dropped to zero anywhere but Malawi [95,96].

As with chloroquine, the mutations conferring resistance to SP are well known [97-102]. Pyrimethamine and sulfadoxine inhibit two enzymes in the folate pathway (**Figure 1.6**), dihydrofolate reductase and dihydropteroate synthase. Without this pathway, the parasite cannot make pyrimidines, and thus cannot replicate. Point mutations at specific sites in the *dhps* (dihydropteroate synthase; chromosome 8) and *dhfr* (dihydrofolate reductase; chromosome 4) genes confer resistance to sulfadoxine and pyrimethamine, respectively. Moreover, these mutations occur in a stepwise fashion, with a higher number of mutations corresponding to a higher level of resistance [103]. A parasite with three mutations in *dhfr* will show substantial *in vitro* resistance to pyrimethamine (at least ten-fold higher than wild-type [104]), and has a higher probability of causing clinical SP treatment failure (**Figure 1.7**) [105,106].

The mutations in the sulfadoxine target, dihydropteroate synthase (*dhps*), are less studied, but also affect resistance. In Africa, changes at codon 437 (alanine to glycine) and codon 540 (lysine to glutamate) are the common mutations associated with SP resistance [105,107,108]. Changes at codons 436, 581, and 613 have also been seen [101,105,109]. The "quintuple mutant," a parasite with three mutations in *dhfr* (N51I/C59R/S108N) and two mutations in *dhps* (A437G/K540E) is strongly

associated with SP treatment failure (odds of failure is ~19 relative to wild-type parasites) [105,110].

Because the mutations are simple point mutations, one might expect them to arise often, on many genetic backgrounds. *In vitro* studies indicate that a single SNP conferring SP resistance occurs once in every 10^9-10^{11} replications [111], and in the typical symptomatic patient there are 10^{10}-10^{12} parasites [112]. Therefore, one might expect a mutation increasing SP resistance to occur in every patient. Given the number of people infected with malaria (>500 million cases per year [1]), even if only a small fraction of those newly resistant parasites were passed on to the mosquito, we would expect SP resistance to emerge multiple times. The rapid emergence of resistance to SP after its introduction, frequently in areas where DHFR inhibitors were used intensively, also lends support to the idea of multiple origins. However, it appears that SP resistance, like chloroquine resistance, emerged relatively rarely and then spread, necessitating a revision of the models [113].

Three studies on three different continents have all found a dramatic decrease in genetic diversity in strains resistant to SP compared with sensitive strains [44,45,86]. In South America, Cortese *et al.* (2002) collected 22 samples from eight countries and examined *dhfr*, *dhps*, and microsatellites adjacent to these genes. For *dhfr*, they found that sensitive (wild-type) parasites had a variety of microsatellite haplotypes, while resistant parasites (three *dhfr* mutations) all had identical haplotypes [86]. Similarly, Roper *et al.* (2003) examined 175 samples in South Africa and found decreasing haplotype diversity with increasing mutations in *dhfr* and *dhps* [44]. Finally, Nair *et al.* (2003) examined 33 microsatellites across the *dhfr* locus and found increased linkage disequilibrium surrounding *dhfr* in resistant, but not sensitive, parasites, indicating a selective sweep [45]. In addition, the microsatellite haplotype for all parasites with three *dhfr* mutations (triple-mutants) in South Africa was identical to the most common haplotype for resistant parasites in Asia, suggesting that the strain first appeared in Asia and then spread to East Africa [46]. More recently, a few other haplotypes have been found for triple-mutants in Kenya and Senegal, but it is not clear if they are independent origins or simply the result of recombination of imported strains with

native strains [49,114]. Resistant parasites from South America are distinct from those in Africa and Southeast Asia [115].

Because resistant alleles of both *pfcrt* and *dhfr* are surrounded by genomic regions of decreased diversity, it was proposed that scanning the *P. falciparum* genome for regions of reduced diversity could lead to the discovery of other genes affecting drug resistance [116]. However, such a method will only work when resistance emerges rarely and spreads widely. While studies of chloroquine and SP tempt us to assume this to be the model [113], recent studies of *pfmdr1* and mefloquine resistance indicate that other models are possible. As stated above, increased copy number of the gene *pfmdr1* is correlated with resistance to mefloquine and other antmalarial drugs [36,117]. A study of microsatellites flanking *pfmdr1* revealed at least five independent amplifications of *pfmdr1* in Southeast Asia alone [118]. Multiple origins of the resistance genotype cause "soft" selective sweeps which are not as easily seen in scans of genomic diversity.

Goals for the present research

The prior studies of SP resistance [44-46,49,86,114], while intriguing, were all hampered by the inability to study mixed infections. In places where malaria transmission is high, people are often infected with multiple strains of *P. falciparum* [52]. However, using standard analysis techniques, in which each microsatellite is amplified and analyzed individually, it is not possible to construct unambiguous haplotypes for mixed infections of *P. falciparum*. Therefore, the previous studies have either used samples from areas of low transmission, where mixed infections are less common [44,86], or have excluded any mixed infections from their microsatellite analysis [45,49]. Nair *et al.* (2003) were able to analyze only 353 of 583 samples collected due to mixed infections. Consequently, a method that allowed the study of mixed infections would increase our ability to characterize the population genetics of SP resistance in areas of intense malaria transmission.

Another problem with existing studies is that all samples were collected after SP resistance was established. That is, they show that most, if not all, of the resistant

parasites currently in Africa are related to the resistant parasites in Southeast Asia, but they do not establish that there were never any other resistant strains in Africa. It is possible that there were triple-mutant parasites that arose in Africa prior to the immigration of the resistant strains from Southeast Asia, but that the Asian parasites out-competed the native parasites. If so, then that would point to the existence of other attributes of the Asian strain that increase its fitness.

This dissertation presents work to address these two problems. First, a novel method for separating mixed infections of *P. falciparum* is described. In this method, large fragments of *P. falciparum* DNA are cloned into yeast, such that complete haplotypes remain intact. Second, we investigate the history of SP resistance in Kilifi, Kenya. By analyzing samples collected at two time points, one significantly before SP was used and one immediately following its introduction, we can reconstruct the emergence of SP resistance. For comparison, we present data on the chloroquine resistance gene, *pfcrt*, as well.

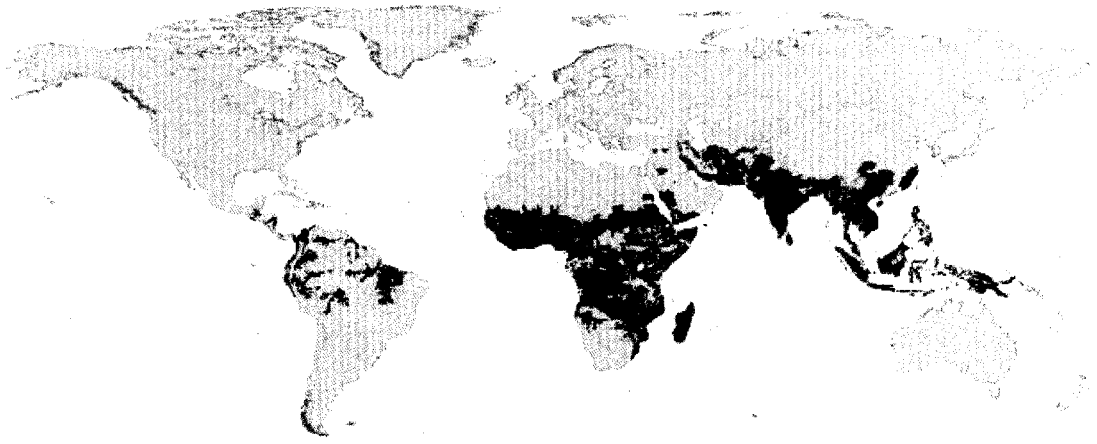


Figure 1.1 – Global Distribution of *P. Falciparum* [119].

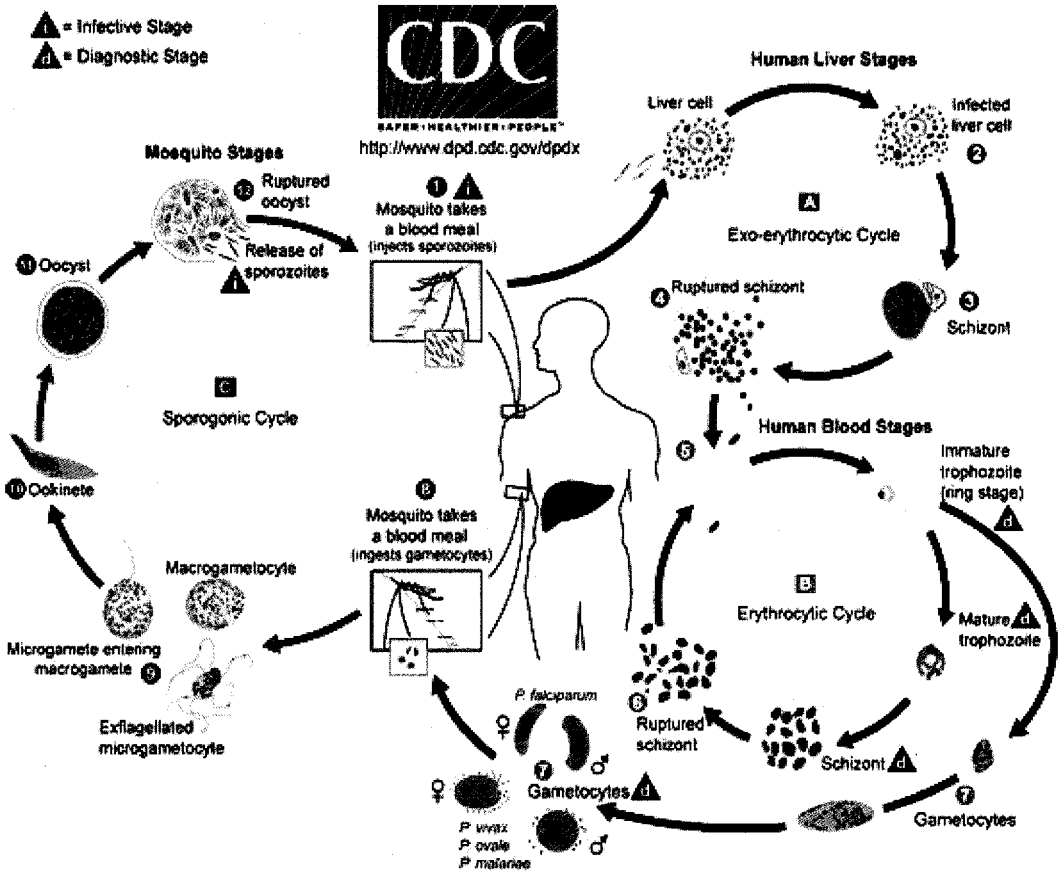


Figure 1.2 – Life Cycle of the Parasite [120]. The malaria parasite life cycle involves two hosts. During a blood meal, a malaria-infected female *Anopheles* mosquito inoculates sporozoites into the human host ❶. Sporozoites infect liver cells ❷and mature into schizonts ❸, which rupture and release merozoites ❹. (Of note, in *P. vivax* and *P. ovale* a dormant stage [hypnozoites] can persist in the liver and cause relapses by invading the bloodstream weeks, or even years later.) After this initial replication in the liver (exo-erythrocytic schizogony A), the parasites undergo asexual multiplication in the erythrocytes (erythrocytic schizogony B). Merozoites infect red blood cells ❺. The ring stage trophozoites mature into schizonts, which rupture releasing merozoites ❻. Some parasites differentiate into sexual erythrocytic stages (gametocytes) ❼. Blood stage parasites are responsible for the clinical manifestations of the disease. The gametocytes, male (microgametocytes) and female (macrogametocytes), are ingested by an *Anopheles* mosquito during a blood meal ❽. The parasites' multiplication in the mosquito is known as the sporogonic cycle C. While in the mosquito's stomach, the microgametes penetrate the macrogametes generating zygotes ❾. The zygotes in turn become motile and elongated (ookinetes) ❿which invade the midgut wall of the mosquito where they develop into oocysts ⓫. The oocysts grow, rupture, and release sporozoites ⓬, which make their way to the mosquito's salivary glands. Inoculation of the sporozoites ❶into a new human host perpetuates the malaria life cycle.

An example of a microsatellite in a DNA sequence:
CGTATAGGCATATATATATATATATCGTTAGAACTAAT
CGTATAGGCATATATATATATATATATATCGTTAGAACTAAT
CGTATAGGCATATATATATATATCGTTAGAACTAAT
CGTATAGGCATATATATATATATATATATATCGTTAGAACTAAT

An example of a SNP (Single Nucleotide Polymorphism) in a DNA sequence:
GTTGTAGTTATGGGAAGAACAAGCTGGGAAAGCATTCCAAA
GTTGTAGTTATGGGAAGAACAAACTGGGAAAGCATTCCAAA

Figure 1.3 – Microsatellites and SNPs.

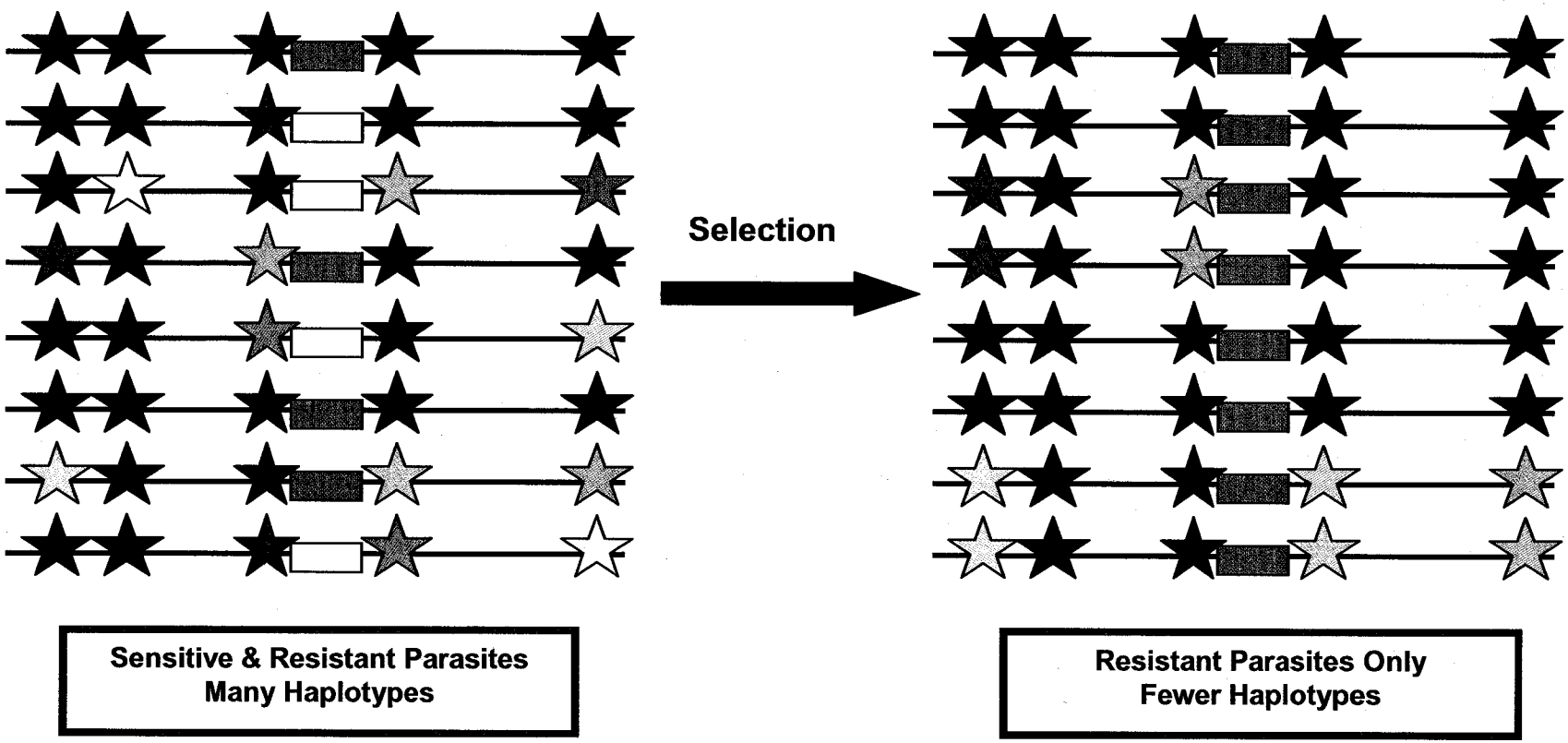
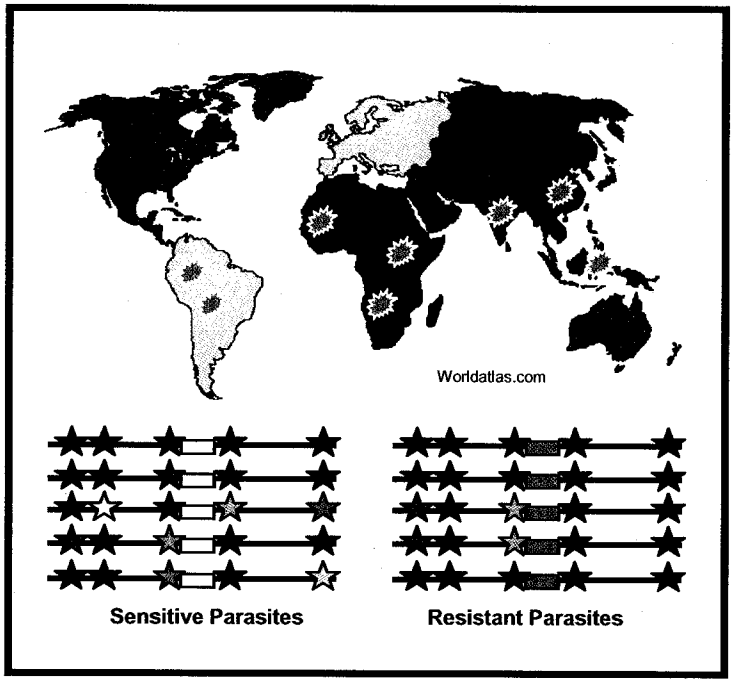
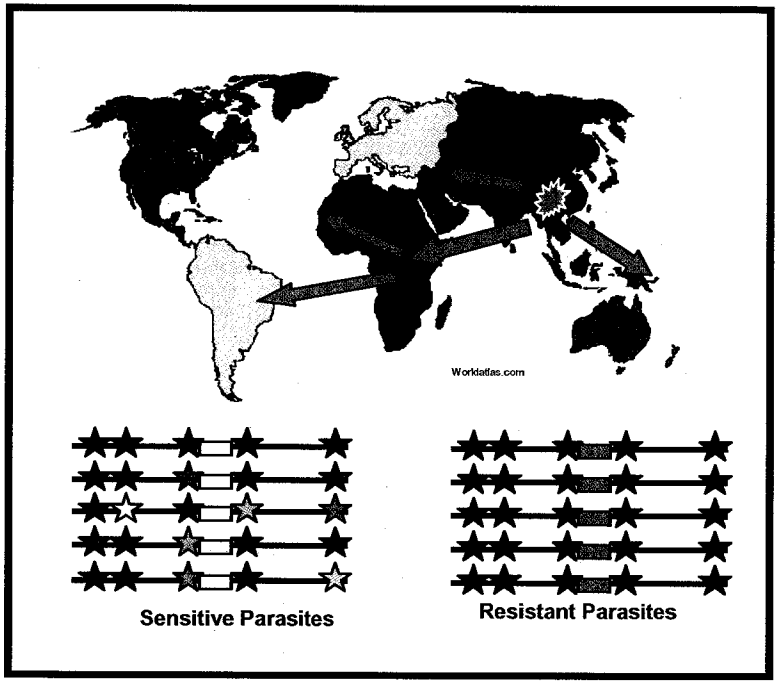


Figure 1.4 – Effect of a Selective Sweep on Haplotypes. Each line represents the region of parasite DNA that contains the gene of interest (rectangle). A wild-type (yellow) allele for this gene means the parasite is sensitive to the drug. A mutant (purple) allele for this gene allows the parasite to survive selection. Stars are microsatellites or other genetic markers; different colors indicate different alleles. The set of alleles (colors) on a given line (chromosome) is a haplotype.



Option A: Resistance emerges frequently.



Option B: Resistance emerges rarely and spreads.

Figure 1.5 – Two Scenarios for the Emergence and Spread of Drug Resistance. Purple bursts on the map indicate foci of resistance. Below each map is a representation of haplotypes for different parasites; each line is a haplotype. Stars are microsatellites, with color indicating allele. The yellow and purple boxes represent sensitive and resistant alleles of *dhfr*, respectively.

.

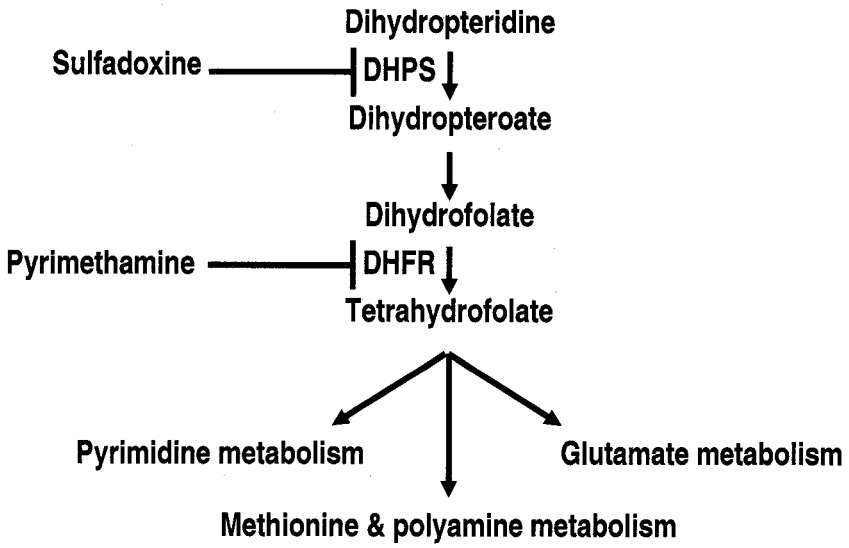


Figure 1.6 – A Simplified Folate Pathway [121]. The enzymes dihydropteroate synthase (DHPS) and dihydrofolate reductase (DHFR) are shown in blue. Sulfadoxine and pyrimethamine, respectively, inhibit these two enzymes.

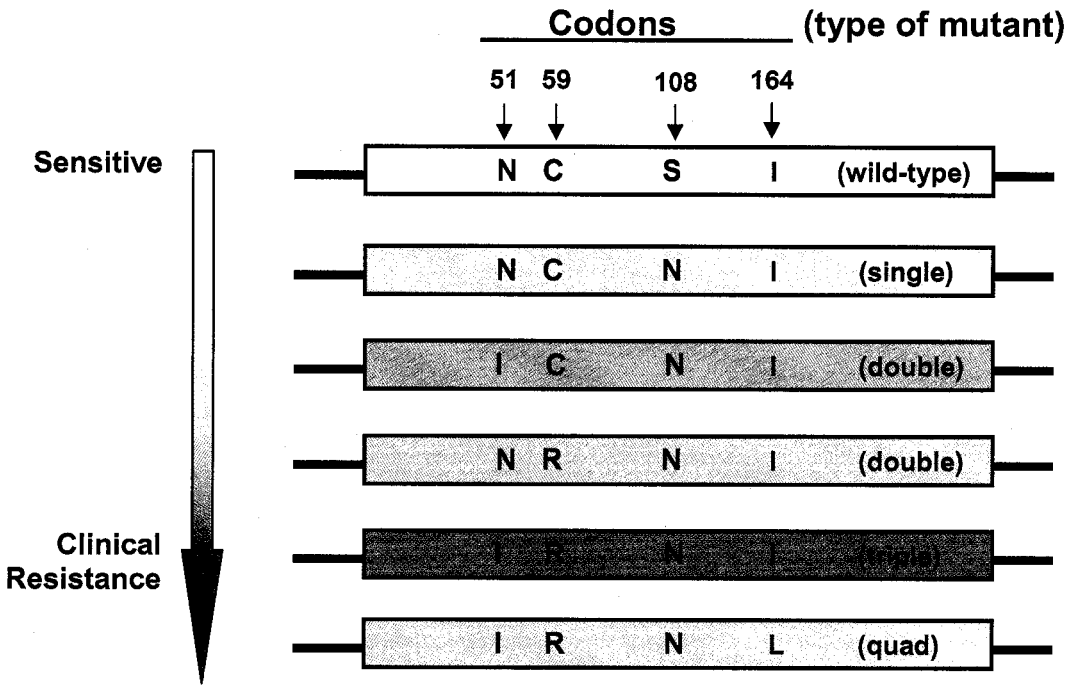


Figure 1.7 – Effect of Point Mutations in *Dhfr* on Resistance to SP.

Table 1.1 – Characteristics of Antimalarial Drugs [24]. Cost is cost/treatment, not per dose.

Drug	Cost ($)	Doses	Side Effects	Severe Side Effects
Chloroquine	0.11	3	GI upset, itching, dizziness	Death from overdose
Sulfadoxine-pyrimethamine	0.14	1	Rash[122]	Stevens-Johnson syndrome
Quinine	0.97	21	Tinnitus, vertigo, headache, fever, syncope, delirium, nausea	Hemolytic anemia, coma, respiratory arrest, renal failure
Mefloquine	2.55	1	Vomiting, headache, insomnia, vivid dreams, anxiety, dizziness	Psychosis
Artemether-lumefantrine	9.12	6	Dizziness, palpitations	Impaired hearing
Artesunate-mefloquine	5.00	6	Vomiting, anorexia, diarrhea	None known

Table 1.2 – Outcomes of *In Vivo* Tests for Resistance. Classification of treatment outcomes is according to WHO protocol, 2005 [27].

Outcome	Characteristics
Early Treatment Failure	Danger signs or severe malaria on day 1, 2 or 3 in the presence of parasitemia. Parasitemia on day 2 higher than on day 0. Parasitemia on day 3 with axillary temperature ≥37.5 °C. Parasitemia on day 3 ≥25% of count on day 0.
Late Clinical Failure	Danger signs or severe malaria in the presence of parasitemia on any day between day 4 and day 28, without the patient previously meeting any of the criteria of early treatment failure. Axillary temperature ≥37.5 °C in the presence of parasitemia on any day between day 4 and day 28, without the patient previously meeting any of the criteria of early treatment failure.
Late Parasitological Failure	Presence of parasitemia between day 4 and day 28 with temperature <37.5 °C, without the patient previously meeting any of the criteria of early treatment failure or late clinical failure.
Adequate clinical and parasitological response	Absence of parasitemia on day 28, irrespective of axillary temperature, without the patient meeting any of the criteria of early treatment failure, late clinical failure or late parasitological failure.

NOTES TO CHAPTER ONE

1. Snow, R.W., Guerra, C.A., Noor, A.M., Myint, H.Y. & Hay, S.I. The global distribution of clinical episodes of Plasmodium falciparum malaria. *Nature* **434**, 214-7 (2005).

2. Malaria in Africa. in *Roll Back Malaria Infosheets* Vol. 2006 (Geneva, 1998).

3. Abu-Raddad, L.J., Patnaik, P. & Kublin, J.G. Dual infection with HIV and malaria fuels the spread of both diseases in sub-Saharan Africa. *Science* **314**, 1603-6 (2006).

4. Greenwood, B.M., Bhasin, A., Bowler, C.H., Naylor, H. & Targett, G.A. Capacity strengthening in malaria research: the Gates Malaria Partnership. *Trends Parasitol* **22**, 278-84 (2006).

5. Korenromp, E.L., Williams, B.G., Gouws, E., Dye, C. & Snow, R.W. Measurement of trends in childhood malaria mortality in Africa: an assessment of progress toward targets based on verbal autopsy. *Lancet Infect Dis* **3**, 349-58 (2003).

6. Leder, K. & Weller, P.F. Epidemiology, pathogenesis, clinical features, and diagnosis of malaria. (UpToDate 14.3, 2006).

7. Murray, P.R., Rosenthal, K.S., Kobayashi, G.S. & Pfaller, M.A. *Medical Microbiology*, (Mosby, St. Louis, MO, 2002).

8. Chedraui, P.A., Daily, J. & Lockwood, C.J. Malaria in pregnancy. (UpToDate Online 14.3, 2006).

9. Duffy, P.E. & Fried, M. Malaria in the pregnant woman. *Curr Top Microbiol Immunol* **295**, 169-200 (2005).

10. Fontenille, D. & Simard, F. Unravelling complexities in human malaria transmission dynamics in Africa through a comprehensive knowledge of vector populations. *Comp Immunol Microbiol Infect Dis* **27**, 357-75 (2004).

11. Lindblade, K.A. et al. Sustainability of reductions in malaria transmission and infant mortality in western Kenya with use of insecticide-treated bednets: 4 to 6 years of follow-up. *Jama* **291**, 2571-80 (2004).

12. Magesa, S.M. et al. Trial of pyrethroid impregnated bednets in an area of Tanzania holoendemic for malaria. Part 2. Effects on the malaria vector population. *Acta Trop* **49**, 97-108 (1991).

13. Maxwell, C.A., Rwegoshora, R.T., Magesa, S.M. & Curtis, C.F. Comparison of coverage with insecticide-treated nets in a Tanzanian town and villages where nets and insecticide are either marketed or provided free of charge. *Malar J* **5**, 44 (2006).

14. Kleinschmidt, I. et al. Reduction in infection with Plasmodium falciparum one year after the introduction of malaria control interventions on Bioko Island, Equatorial Guinea. *Am J Trop Med Hyg* **74**, 972-8 (2006).

15. Weller, P.F. & Leder, K. Prevention of malaria infection in travelers. (UpToDate, 2006).

16. Garner, P. & Gulmezoglu, A.M. Drugs for preventing malaria in pregnant women. *Cochrane Database Syst Rev*, CD000169 (2006).

17. Schellenberg, D. et al. Intermittent treatment for malaria and anaemia control at time of routine vaccinations in Tanzanian infants: a randomised, placebo-controlled trial. *Lancet* **357**, 1471-7 (2001).

18. Marsh, K. & Kinyanjui, S. Immune effector mechanisms in malaria. *Parasite Immunol* **28**, 51-60 (2006).

19. Matuschewski, K. Vaccine development against malaria. *Curr Opin Immunol* **18**, 449-57 (2006).

20. Alonso, P.L. et al. Efficacy of the RTS,S/AS02A vaccine against Plasmodium falciparum infection and disease in young African children: randomised controlled trial. *Lancet* **364**, 1411-20 (2004).

21. Alonso, P.L. et al. Duration of protection with RTS,S/AS02A malaria vaccine in prevention of Plasmodium falciparum disease in Mozambican children: single-blind extended follow-up of a randomised controlled trial. *Lancet* **366**, 2012-8 (2005).

22. Rocco, F. *The Miraculous Fever-Tree*, (HarperCollins Publishers Inc, New York, 2003).

23. Wongsrichanalai, C., Pickard, A.L., Wernsdorfer, W.H. & Meshnick, S.R. Epidemiology of drug-resistant malaria. *Lancet Infect Dis* **2**, 209-18 (2002).

24. Baird, J.K. Effectiveness of antimalarial drugs. *N Engl J Med* **352**, 1565-77 (2005).

25. Demar, M. & Carme, B. Plasmodium falciparum in vivo resistance to quinine: description of two RIII responses in French Guiana. *Am J Trop Med Hyg* **70**, 125-7 (2004).

26. Daily, J.P. Antimalarial drug therapy: the role of parasite biology and drug resistance. *J Clin Pharmacol* **46**, 1487-97 (2006).

27. Susceptibility of Plasmodium falciparum to antimalalrial drugs: report on global monitoring: 1996-2004. (World Health Organization, Geneva, 2005).

28. Umotong, A.B. et al. Correlation between in vivo and in vitro response of chloroquine-resistant Plasmodium falciparum in Calabar, south-eastern Nigeria. *Acta Trop* **49**, 119-25 (1991).

29. Payne, D. Spread of chloroquine resistance in Plasmodium falciparum. *Parasitol Today* **3**, 241-6 (1987).

30. Talisuna, A.O., Bloland, P. & D'Alessandro, U. History, dynamics, and public health importance of malaria parasite resistance. *Clin Microbiol Rev* **17**, 235-54 (2004).

31. Wellems, T.E. Plasmodium chloroquine resistance and the search for a replacement antimalarial drug. *Science* **298**, 124-6 (2002).

32. Watkins, W.M. et al. Effectiveness of amodiaquine as treatment for chloroquine-resistant Plasmodium falciparum infections in Kenya. *Lancet* **1**, 357-9 (1984).

33. Mberu, E.K. et al. The changing in vitro susceptibility pattern to pyrimethamine/sulfadoxine in Plasmodium falciparum field isolates from Kilifi, Kenya. *Am J Trop Med Hyg* **62**, 396-401 (2000).

34. Watkins, W. Personal communication. (2007).

35. Nosten, F. et al. Mefloquine-resistant falciparum malaria on the Thai-Burmese border. *Lancet* **337**, 1140-3 (1991).

36. Sidhu, A.B. et al. Decreasing pfmdr1 copy number in plasmodium falciparum malaria heightens susceptibility to mefloquine, lumefantrine, halofantrine, quinine, and artemisinin. *J Infect Dis* **194**, 528-35 (2006).

37. Klayman, D.L. Qinghaosu (artemisinin): an antimalarial drug from China. *Science* **228**, 1049-55 (1985).

38. Yeung, S., Pongtavornpinyo, W., Hastings, I.M., Mills, A.J. & White, N.J. Antimalarial drug resistance, artemisinin-based combination therapy, and the contribution of modeling to elucidating policy choices. *Am J Trop Med Hyg* **71**, 179-86 (2004).

39. MMV. Selection of Partners for Antimalarial Drug Combinations. in *Strategic Consultation* (Medicines for Malaria Venture, Geneva, Switzerland, 2005).

40. Mutabingwa, T.K. Artemisinin-based combination therapies (ACTs): best hope for malaria treatment but inaccessible to the needy! *Acta Trop* **95**, 305-15 (2005).

41. Guidelines for the Treatment of Malaria. (World Health Organization, Geneva, 2006).

42. Davis, T.M., Karunajeewa, H.A. & Ilett, K.F. Artemisinin-based combination therapies for uncomplicated malaria. *Med J Aust* **182**, 181-5 (2005).

43. McNeil Jr., D.G. Low Cost Antimalaria Pill Available. in *The New York Times* (New York, NY, 2007).

44. Roper, C. et al. Antifolate antimalarial resistance in southeast Africa: a population-based analysis. *Lancet* **361**, 1174-81 (2003).

45. Nair, S. et al. A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. *Mol Biol Evol* **20**, 1526-36 (2003).

46. Roper, C. et al. Intercontinental spread of pyrimethamine-resistant malaria. *Science* **305**, 1124 (2004).

47. Wootton, J.C. et al. Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. *Nature* **418**, 320-3 (2002).

48. Ariey, F. et al. Invasion of Africa by a single pfcrt allele of South East Asian type. *Malar J* **5**, 34 (2006).

49. McCollum, A.M. et al. Antifolate Resistance in Plasmodium falciparum: Multiple Origins and Identification of Novel dhfr Alleles. *J Infect Dis* **194**, 189-97 (2006).

50. Malaria and the Red Cell. Vol. 2007 (Brigham and Women's Hospital, Harvard University, Boston, 2002).

51. Ranford-Cartwright, L.C., Balfe, P., Carter, R. & Walliker, D. Frequency of cross-fertilization in the human malaria parasite Plasmodium falciparum. *Parasitology* **107 (Pt 1)**, 11-8 (1993).

52. Arnot, D. 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**, 580-5 (1998).

53. Hastings, I.M. A model for the origins and spread of drug-resistant malaria. *Parasitology* **115 (Pt 2)**, 133-41 (1997).

54. Hastings, I.M. Malaria control and the evolution of drug resistance: an intriguing link. *Trends Parasitol* **19**, 70-3 (2003).

55. Hastings, I.M. & D'Alessandro, U. Modelling a predictable disaster: the rise and spread of drug-resistantmalaria. *Parasitol Today* **16**, 340-7 (2000).

56. Koella, J.C. & Antia, R. Epidemiological models for the spread of anti-malarial resistance. *Malar J* **2**, 3 (2003).

57. Molyneux, D.H., Floyd, K., Barnish, G. & Fevre, E.M. Transmission control and drug resistance in malaria: a crucial interaction. *Parasitol Today* **15**, 238-40 (1999).

58. Mharakurwa, S. Plasmodium falciparum transmission rate and selection for drug resistance: a vexed association or a key to successful control? *Int J Parasitol* **34**, 1483-7 (2004).

59. Volkman, S.K. et al. Recent origin of Plasmodium falciparum from a single progenitor. *Science* **293**, 482-4 (2001).

60. Hughes, A.L. & Verra, F. Very large long-term effective population size in the virulent human malaria parasite Plasmodium falciparum. *Proc Biol Sci* **268**, 1855-60 (2001).

61. Hughes, A.L. & Verra, F. Extensive polymorphism and ancient origin of Plasmodium falciparum. *Trends Parasitol* **18**, 348-51 (2002).

62. Mu, J. et al. Chromosome-wide SNPs reveal an ancient origin for Plasmodium falciparum. *Nature* **418**, 323-6 (2002).

63. Conway, D.J. Tracing the dawn of Plasmodium falciparum with mitochondrial genome sequences. *Trends Genet* **19**, 671-4 (2003).

64. Joy, D.A. et al. Early origin and recent expansion of Plasmodium falciparum. *Science* **300**, 318-21 (2003).

65. Rich, S.M., Licht, M.C., Hudson, R.R. & Ayala, F.J. Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of Plasmodium falciparum. *Proc Natl Acad Sci U S A* **95**, 4425-30 (1998).

66. Hartl, D.L. The origin of malaria: mixed messages from genetic diversity. *Nat Rev Microbiol* **2**, 15-22 (2004).

67. Hartl, D.L. et al. The paradoxical population genetics of Plasmodium falciparum. *Trends Parasitol* **18**, 266-72 (2002).

68. Hume, J.C., Lyons, E.J. & Day, K.P. Human migration, mosquitoes and the evolution of Plasmodium falciparum. *Trends Parasitol* **19**, 144-9 (2003).

69. Conway, D.J. et al. Origin of Plasmodium falciparum malaria is traced by mitochondrial DNA. *Mol Biochem Parasitol* **111**, 163-71 (2000).

70. Currat, M. et al. Molecular analysis of the beta-globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the beta(S) Senegal mutation. *Am J Hum Genet* **70**, 207-23 (2002).

71. Flint, J., Harding, R.M., Boyce, A.J. & Clegg, J.B. The population genetics of the haemoglobinopathies. *Baillieres Clin Haematol* **11**, 1-51 (1998).

72. Tishkoff, S.A. et al. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* **293**, 455-62 (2001).

73. Escalante, A.A. & Ayala, F.J. Phylogeny of the malarial genus Plasmodium, derived from rRNA gene sequences. *Proc Natl Acad Sci U S A* **91**, 11373-7 (1994).

74. Jeffares, D.C. et al. Genome variation and evolution of the malaria parasite Plasmodium falciparum. *Nat Genet* **39**, 120-5 (2007).

75. Volkman, S.K. et al. A genome-wide map of diversity in Plasmodium falciparum. *Nat Genet* **39**, 113-9 (2007).

76. Gardner, M.J. et al. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419**, 498-511 (2002).

77. Su, X. & Wellems, T.E. Toward a high-resolution Plasmodium falciparum linkage map: polymorphic markers from hundreds of simple sequence repeats. *Genomics* **33**, 430-44 (1996).

78. Walker-Jonah, A., Dolan, S.A., Gwadz, R.W., Panton, L.J. & Wellems, T.E. An RFLP map of the Plasmodium falciparum genome, recombination rates and favored linkage groups in a genetic cross. *Mol Biochem Parasitol* **51**, 313-20 (1992).

79. Wellems, T.E. et al. Chloroquine resistance not linked to mdr-like genes in a Plasmodium falciparum cross. *Nature* **345**, 253-5 (1990).

80. Su, X. et al. A genetic map and recombination parameters of the human malaria parasite Plasmodium falciparum. *Science* **286**, 1351-3 (1999).

81. Conway, D.J. et al. High recombination rate in natural populations of Plasmodium falciparum. *Proc Natl Acad Sci U S A* **96**, 4506-11 (1999).

82. Anderson, T.J. et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. *Mol Biol Evol* **17**, 1467-82 (2000).

83. Awadalla, P., Walliker, D., Babiker, H. & Mackinnon, M. The question of Plasmodium falciparum population structure. *Trends Parasitol* **17**, 351-3 (2001).

84. Mu, J. et al. Genome-wide variation and identification of vaccine targets in the Plasmodium falciparum genome. *Nat Genet* **39**, 126-30 (2007).

85. Anderson, T. Evolution of Malaria Parasites Under Strong Drug Selection. in *Pests, Plagues & Plants: Genomics & Global Health* (University of Washington, Seattle, 2007).

86. Cortese, J.F., Caraballo, A., Contreras, C.E. & Plowe, C.V. Origin and dissemination of Plasmodium falciparum drug-resistance mutations in South America. *J Infect Dis* **186**, 999-1006 (2002).

87. Ursos, L.M. & Roepe, P.D. Chloroquine resistance in the malarial parasite, Plasmodium falciparum. *Med Res Rev* **22**, 465-91 (2002).

88. Fidock, D.A. et al. Mutations in the P. falciparum digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol Cell* **6**, 861-71 (2000).

89. Chen, N. et al. pfcrt Allelic types with two novel amino acid mutations in chloroquine-resistant Plasmodium falciparum isolates from the Philippines. *Antimicrob Agents Chemother* **47**, 3500-5 (2003).

90. Su, X., Kirkman, L.A., Fujioka, H. & Wellems, T.E. Complex polymorphisms in an approximately 330 kDa protein are linked to chloroquine-resistant P. falciparum in Southeast Asia and Africa. *Cell* **91**, 593-603 (1997).

91. Mehlotra, R.K. et al. Evolution of a unique Plasmodium falciparum chloroquine-resistance phenotype in association with pfcrt polymorphism in

Papua New Guinea and South America. *Proc Natl Acad Sci U S A* **98**, 12689-94 (2001).

92. Alifrangis, M. et al. Occurrence of the Southeast Asian/South American SVMNT haplotype of the chloroquine-resistance transporter gene in Plasmodium falciparum in Tanzania. *J Infect Dis* **193**, 1738-41 (2006).

93. Andersson, D.I. The biological cost of mutational antibiotic resistance: any practical conclusions? *Curr Opin Microbiol* **9**, 461-5 (2006).

94. Kublin, J.G. et al. Reemergence of chloroquine-sensitive Plasmodium falciparum malaria after cessation of chloroquine use in Malawi. *J Infect Dis* **187**, 1870-5 (2003).

95. Laufer, M.K. et al. Return of chloroquine antimalarial efficacy in Malawi. *N Engl J Med* **355**, 1959-66 (2006).

96. Wang, X. et al. Decreased prevalence of the Plasmodium falciparum chloroquine resistance transporter 76T marker associated with cessation of chloroquine use against P. falciparum malaria in Hainan, People's Republic of China. *Am J Trop Med Hyg* **72**, 410-4 (2005).

97. Cowman, A.F., Morry, M.J., Biggs, B.A., Cross, G.A. & Foote, S.J. Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of Plasmodium falciparum. *Proc Natl Acad Sci U S A* **85**, 9109-13 (1988).

98. Sirawaraporn, W., Sathitkul, T., Sirawaraporn, R., Yuthavong, Y. & Santi, D.V. Antifolate-resistant mutants of Plasmodium falciparum dihydrofolate reductase. *Proc Natl Acad Sci U S A* **94**, 1124-9 (1997).

99. Snewin, V.A., England, S.M., Sims, P.F. & Hyde, J.E. Characterisation of the dihydrofolate reductase-thymidylate synthetase gene from human malaria parasites highly resistant to pyrimethamine. *Gene* **76**, 41-52 (1989).

100. Triglia, T. & Cowman, A.F. Primary structure and expression of the dihydropteroate synthetase gene of Plasmodium falciparum. *Proc Natl Acad Sci U S A* **91**, 7149-53 (1994).

101. Brooks, D.R. et al. Sequence variation of the hydroxymethyldihydropterin pyrophosphokinase: dihydropteroate synthase gene in lines of the human malaria parasite, Plasmodium falciparum, with differing resistance to sulfadoxine. *Eur J Biochem* **224**, 397-405 (1994).

102. Reeder, J.C. et al. Point mutations in the dihydrofolate reductase and dihydropteroate synthetase genes and in vitro susceptibility to pyrimethamine and cycloguanil of Plasmodium falciparum isolates from Papua New Guinea. *Am J Trop Med Hyg* **55**, 209-13 (1996).

103. Winstanley, P., Ward, S., Snow, R. & Breckenridge, A. Therapy of falciparum malaria in sub-saharan Africa: from molecule to policy. *Clin Microbiol Rev* **17**, 612-37 (2004).

104. Khalil, I. et al. Dihydrofolate reductase and dihydropteroate synthase genotypes associated with in vitro resistance of Plasmodium falciparum to pyrimethamine, trimethoprim, sulfadoxine, and sulfamethoxazole. *Am J Trop Med Hyg* **68**, 586-9 (2003).

105. Kublin, J.G. et al. Molecular markers for failure of sulfadoxine-pyrimethamine and chlorproguanil-dapsone treatment of Plasmodium falciparum malaria. *J Infect Dis* **185**, 380-8 (2002).

106. Sibley, C.H. et al. Pyrimethamine-sulfadoxine resistance in Plasmodium falciparum: what next? *Trends Parasitol* **17**, 582-8 (2001).

107. Nzila, A.M. et al. 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**, 991-6 (2000).

108. Plowe, C.V. et al. Mutations in Plasmodium falciparum dihydrofolate reductase and dihydropteroate synthase and epidemiologic patterns of pyrimethamine-sulfadoxine use and resistance. *J Infect Dis* **176**, 1590-6 (1997).

109. Wang, P., Read, M., Sims, P.F. & Hyde, J.E. Sulfadoxine resistance in the human malaria parasite Plasmodium falciparum is determined by mutations in dihydropteroate synthetase and an additional factor associated with folate utilization. *Mol Microbiol* **23**, 979-86 (1997).

110. Wernsdorfer, W.H. & Noedl, H. Molecular markers for drug resistance in malaria: use in treatment, diagnosis and epidemiology. *Curr Opin Infect Dis* **16**, 553-8 (2003).

111. Paget-McNicol, S. & Saul, A. Mutation rates in the dihydrofolate reductase gene of Plasmodium falciparum. *Parasitology* **122**, 497-505 (2001).

112. White, N.J. et al. Averting a malaria disaster. *Lancet* **353**, 1965-7 (1999).

113. Hastings, I.M. The origins of antimalarial drug resistance. *Trends Parasitol* **20**, 512-8 (2004).

114. Ndiaye, D. et al. Defining the origin of Plasmodium falciparum resistant dhfr isolates in Senegal. *Acta Trop* **99**, 106-11 (2006).

115. McCollum, A.M., Mueller, K., Villegas, L., Udhayakumar, V. & Escalante, A.A. Common origin and fixation of Plasmodium falciparum dhfr and dhps mutations associated with sulphadoxine-pyrimethamine resistance in a low transmission area in South America. *Antimicrob Agents Chemother* (2007).

116. Anderson, T.J. Mapping drug resistance genes in Plasmodium falciparum by genome-wide association. *Curr Drug Targets Infect Disord* **4**, 65-78 (2004).

117. Price, R.N. et al. Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. *Lancet* **364**, 438-47 (2004).

118. Nair, S. et al. Recurrent Gene Amplification and Soft Selective Sweeps During Evolution of Multidrug Resistance in Malaria Parasites. *Mol Biol Evol* (2006).

119. Guerra, C.A., Snow, R.W. & Hay, S.I. Mapping the global extent of malaria in 2005. *Trends Parasitol* **22**, 353-8 (2006).

120. Schema of the Life Cycle of Malaria. (Centers for Disease Control and Prevention, 2006).

121. Folate Biosynthetic Pathway. in *Malaria Parasite Metabolic Pathways* Vol. 2007 (The Hebrew University of Jerusalem).

122. Van Voorhis, W. Personal communication: dissertation revisions. (2007).

CHAPTER TWO: A NOVEL METHOD FOR SEPARTING DNA FROM MIXED INFECTIONS OF *P. FALCIPARUM*

In areas of low malaria transmission most patients are infected with only one haploid parasite; that is, the infections are mainly clonal. However, in many areas where evolutionary studies would be of interest, a large proportion of patients carry multiple parasites. Current methods require that each genetic marker be typed individually. Therefore, if the initial DNA sample is mixed one cannot easily construct unambiguous haplotypes. One way to construct haplotypes from a mixed infection would be to adapt the patient sample to culture and then dilute the culture to create clonal cultures, but such an option is extremely labor-intensive. Alternatively, one could apply statistical methods to predict the most likely haplotypes (see **Appendix A**), but most available methods were created with human (diploid) data in mind [1,2] and may not extend to mixed samples containing more than two strains of malaria. Recently, a statistical method was developed for mixed infections of malaria, but it was designed for use with pyrosequencing of bi-allelic SNPs, and may not work as well for multi-allelic microsatellite data [3]. For these reasons, most workers have simply excluded from analysis any sample that appears to be polyclonal. This approach precludes efficient use of samples from many of the most interesting endemic regions.

This chapter presents a method for constructing haplotypes from mixed samples of *P. falciparum* DNA. Using basic molecular biology – PCR, yeast transformation, sequencing – it is possible to isolate the linked markers in a mixed sample of *P. falciparum* DNA into separate yeast colonies, thus generating a valid haplotype for that sample. Though we present the method as it applies to the dihydrofolate reductase gene (*dhfr*) and its flanking microsatellite genetic markers, the method is applicable to any gene and any type of genetic marker.

Methods

The overall method is described in **Figure 2.1**. In brief, the protocol uses PCR to amplify a 6 kb region containing *dhfr* and upstream microsatellite markers. These PCR products are co-transformed with a plasmid vector into yeast, and resulting colonies are streaked or pinned onto new plates. DNA extracted from the resulting colonies is then ready for analysis at the microsatellites and *dhfr*. The following paragraphs explain the protocol in detail.

DNA Preparation and Long-range PCR

As described in **Figure 2.1**, the first step of the protocol is to amplify a fragment containing *dhfr* and upstream microsatellites (or any target region). The template for this PCR was *P. falciparum* DNA extracted from cultured strains of 3D7, K1, and Dd2 parasites. For 3D7 and K1, we used the QIAamp DNA Mini Kit from Qiagen (Valencia, CA) to extract the DNA. We followed the printed protocol, with the exception of using inversion rather than vortexing to mix, in order to minimize shearing of the DNA [4]. Dd2 DNA was extracted from cultured parasites using 15% (w/v) saponin to lyse the red blood cells, followed by the DNeasy kit from Qiagen (Valencia, CA; John White, personal communication). DNA concentration was determined by a spectrophotometer.

Primers for this initial PCR amplification are in **Table 2.1**. Either of the two pairs listed may be used, with no appreciable difference in output. The longer pair has a more extensive region of homology with the plasmid vector. We obtained primers from Invitrogen (Carlsbad, CA) and used two different polymerase enzymes, KOD XL Polymerase (Novagen, Madison, WI) and FailSafe (Epicentre, Madison, WI), both of which amplified the target fragment well. For KOD XL, each 20 µl reaction contained 1x KOD XL buffer, 0.2 mM each dNTP, 0.3 µM forward primer, 0.3 µM reverse primer, 1 unit KOD XL Polymerase, and approximately 300 ng template DNA. For FailSafe, each 20 µl reaction contained 1x Buffer A, 1 µM forward primer, 1 µM reverse primer, 1 unit FailSafe Enzyme Mix, and approximately 300 ng template

DNA. Using an adequate amount of template DNA is essential. Cycling parameters for both enzymes were as follows: 94 °C for 2 minutes; 30 cycles of 94 °C for 30 seconds, 50 °C for 45 seconds, 60 °C for 6 minutes; 60 °C for 5 minutes. Products were confirmed using a 0.7% agarose gel.

Transformation

The plasmid used in the transformations was pRSU, derived from pRS424 [5]. pRS424 is a shuttle vector that can be used in either yeast or *E. coli* and contains the following markers: 2 micron sequence that allows autonomous replication in yeast, TRP1 marker for selection in yeast, the T1 ori and pMB1 ori, for replication in *E. coli*, and the β-lactamase gene to confer ampicillin resistance in *E. coli*. To create pRSU, we amplified the URA3 gene from yeast and inserted it into the multiple cloning site in pRS424 using the Sal I and Sma I sites. The resulting plasmid is 6911 basepairs, and is depicted in **Figure 2.2**.

Prior to transformation into yeast, pRSU was cut with Bam H1 to stimulate gap repair of the double-strand break region. The 5' ends of the primers used in the long-range PCR were designed with homology to pRSU upstream of URA3 and downstream of the Bam H1 site. 5-Fluoro-orotate (FOA) is toxic to yeast cells that can metabolize uracil, so transformants that contain the intact plasmid cannot grow on plates containing FOA. With appropriate integration, the PCR product should replace URA3, thus rendering the host cell able to grow without tryptophan and not susceptible to poisoning by FOA [6-9].

For the transformations, any yeast strain that lacks the ability to produce its own tryptophan and uracil is appropriate as a host strain; we used strain BB14-3a (*MATa bar1 his6 leu2-3,112 trp1-289 ura3-52*). We used standard protocols for lithium acetate transformations in yeast [10-12], growing cultures at 30 °C in YEPD to an OD_{660} of 0.4-1.3 (0.5-3 x 10^7 cells/ml). Each transformation used 10^8 cells [11], 50 μg salmon testes DNA (Sigma, St. Louis, MO), 300-900 ng linear plasmid vector, and 1.5-3 μg PCR product (3-6 μl). For a negative control, we used the product from a no-

template PCR in a transformation. We plated cells onto standard media that lacked tryptophan and contained FOA (1 g/L) and uracil (0.05 g/L).

Streaking or Pinning

After three days of growth at 30 °C, colonies were either streaked or pinned onto new plates and returned to 30 °C for 2-4 days. Pinning was done manually with a platinum wire. This step was necessary to remove residual, unincorporated PCR product from the surface of the cells and the plate, so that it did not interfere with the microsatellite analysis (Josh Veatch, personal communication).

DNA Extraction and Analysis

To extract the DNA from the yeast colonies, we suspended each colony in 40 µl of 20 mM NaOH and incubated at 97 °C for 10 minutes. The resulting mixture was used immediately or stored at -20 °C for up to four months. For microsatellite analysis, we followed the protocol of Nair *et al.* (2003). Each 10 µl reaction contained 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.4 µM forward (labeled) primer, 0.4 µM reverse primer, 0.15 units Taq polymerase, and 1 µl DNA. (**Table 2.1**) lists primer sequences for amplifying two microsatellites within the 6 kb fragment, S780 and S784.) Cycling conditions were: 94 °C for 2 minutes; 25 cycles of 94 °C for 30 seconds, 45 °C for 30 seconds, 60 °C for 30 seconds; 60 °C for 2 minutes.[13] A nested PCR was not necessary. After PCR amplification, we diluted the products 20-fold prior to analysis on a MegaBACE 1000 DNA Analysis System (Amersham, Piscataway, NJ). Product length was determined by inspection using Genetic Profiler (Amersham, Piscataway, NJ). Product length at a given microsatellite locus for a single yeast colony was recorded as the allele at that locus for that colony. The combination of lengths (alleles) at the two microsatellite loci was the haplotype for that colony. The most common haplotype among all colonies from a particular *P. falciparum* sample was the haplotype for that sample. For determining the *dhfr* genotype, we used the PCR/RFLP

method of Duraisingh [14]. On the rare occasions when two alleles were detected at a single locus, those alleles were not recorded.

Results

We tested our method by examining two microsatellites upstream of *dhfr,* one immediately upstream (S784; chr.4 position 755,005-755,056; *dhfr* begins at 755,069) and the other 3.8 kb upstream (S780; chr.4, position 751,282-751,329). These correspond to loci d_104_5 and d_100_8 in Nair *et al.* (2003). Both are dinucleotide repeats, $(AT)_n$. We followed the protocol outlined above, using DNA extracted from cultured 3D7, K1, and Dd2 parasites. DNA from one, two, or all three strains of *P. falciparum* was added in equal amounts to the long-range PCR as the template. By starting with DNA from only one strain of *P. falciparum* (3D7, K1, or Dd2) in the initial long-range PCR, we simulated how this method would behave for clonal samples (i.e. samples from patients infected with only one strain of *P. falciparum*); by starting with DNA from multiple strains we simulated mixed infections.

Results at a Single Locus

At the S780 locus, we determined alleles (microsatellite lengths) for 96 3D7 colonies, 83 K1 colonies, and 97 Dd2 colonies (**Table 2.2**). The distribution of alleles for each sample at S780 is shown in **Figure 2.3a-c**. When we started with DNA from only one strain, the majority of the colonies contained an allele of the expected size. However, as is routinely seen in analysis directly from genomic DNA, we did see larger and smaller sizes at each microsatellite, presumably due to polymerase slippage (stutter) in the initial long-range PCR (**Figure 2.4**; see Discussion). Colonies derived from K1 gave the tightest distribution of lengths; those from Dd2 gave the broadest. To verify that this variation was not due to replication in yeast, we pinned 36 colonies a second time, extracted DNA from the new colonies, and compared the alleles (microsatellite lengths) to those observed in the original colonies. Thirteen of the colonies showed a clear peak in fluorescence at a specific length on the capillary gel

readout at one or both loci for both pinnings (first and second); in all cases the allele from the new colony was identical to that from the parent colony. Therefore, the size variation is not due to instability of the microsatellite during propagation in yeast.

To determine whether instability was a problem for other types of genetic markers, we genotyped the SNPs within the *dhfr* coding region in 26 colonies [14] (**Table 2.3**). The twelve colonies from samples that contained only one type of *P. falciparum* DNA in the initial long-range PCR all showed the correct *dhfr* genotype. For the 14 from mixed samples, all had a genotype that matched one of the component strains. Therefore, SNPs are more stable than microsatellites in this system.

The overall goal of the method is to allow analysis of samples that contain more than one parasite genotype. To test this aspect of the method, we mixed equal quantities of purified DNA from the different strains prior to the initial long-range PCR. **Figure 2.3d-f** shows the analysis of the S780 microsatellite for paired mixtures of the strains, and **Figure 2.3g** shows the result when DNA from all three strains was combined. Although the initial long-range PCR started with the same amount of DNA from each strain, far more colonies were derived from one strain in each case. With the three strains that we used, there was a strong bias in favor of colonies derived from the K1 template, and colonies derived from Dd2 were observed far less often than expected. Though alleles derived from each input strain are present, this extreme bias underscores the conclusion that one cannot use this approach to infer the relative proportions of strains in the input sample.

Results at Both Loci – Constructing Haplotypes

We determined haplotypes (lengths for both loci) for 83 3D7 colonies, 71 K1 colonies, and 67 Dd2 colonies. Because some colonies derived from each DNA sample had the true allele and some had stutter alleles, each sample produced a range of haplotypes. For example, the haplotypes found for colonies derived from K1 DNA are shown in **Figure 2.5**. The most frequently observed haplotype was 196/178, the true haplotype. However, there were 13 other haplotypes found in at least one colony.

They were various combinations of the correct allele at one locus paired with a stutter allele at the other locus, or combinations of stutter alleles at both loci. The first and second most common haplotypes for each sample are given in **Table 2.4**.

We constructed haplotypes for the 276 mixed-sample colonies that showed a clear peak in fluorescence at a specific length for both loci. In the mixtures, the most common haplotype observed among the colonies was always the correct haplotype for one of the strains. The second most common haplotypes were stutter versions of the most common haplotypes. For example, in the K1 and Dd2 mixture, the most common haplotype was 196/178 (K1), but the second most common was not 210/178 (Dd2). Rather, it was 194/178, a stutter version of K1. Therefore, we can clearly identify a haplotype from a mixed sample; however, minor haplotypes are not informative. Moreover, the dominant haplotype does not necessarily represent the predominant strain in the sample, because there is a bias in the detection of some microsatellite alleles over others.

To compare our results to the standard protocol used by Nair *et al.* (2003), we used nested PCR on DNA samples from the Malaria Research and Reference Reagent Resource Center (MR4) to determine alleles for 3D7, K1, Dd2, W2, and V1/S *P. falciparum* strains. We also examined mixtures of: 3D7 and K1 DNA; 3D7 and Dd2 DNA; K1 and Dd2 DNA; and 3D7, K1, and Dd2 DNA. In all cases the alleles for the MR4 samples matched the allele sizes we found using our method (**Table 2.5**). In the mixtures, Dd2 had a weaker fluorescent peak than 3D7 or K1 (data not shown), mirroring what we saw using yeast. Therefore, this skew seems not to derive from our method, but rather to be inherent in the DNA sequence. Also of note, though the V1/S haplotype matches that seen for the triple mutant in Nair *et al.* (2003), the Dd2 and W2 haplotypes do not. V1/S has four mutations in *dhfr*; Dd2 and W2 have three (Dd2 is derived from W2). All three strains came from Southeast Asia.

Efficiency of Method

Because all of the colonies that grow on the FOA selection plates have lost the URA3 gene, each should contain our 6 kb insert, but not all of them did. The fraction of colonies that contained the *P. falciparum* insert is shown in **Table 2.2**. We analyzed 3259 colonies at one microsatellite (S780) and 3279 colonies at a second microsatellite (S784); we analyzed 3024 colonies at both loci. However, of the 3259 analyzed at S780, only 687 had a clear peak in fluorescence at a specific length; the remainder were blank (or had a peak that was unclear or double). That is, only 687 of 3259 colonies that grew on the selective plates clearly contained the insert. The corresponding proportion at S784 was 807/3279. Of the 3024 colonies analyzed at both loci, 497 had a clear peak in fluorescence at a specific length for both loci. The average success rate across transformations was 26% for S780 and 30% for S784, with standard deviations of 19% and 22%, respectively. These numbers (26%, 30%) differ from those in **Table 2.2** (21%, 25%), because we did not weight the average to reflect the number of colonies analyzed per transformation. The plates from the control transformations performed with the PCR reaction that contained no template DNA had some colonies; however, none of those analyzed (S780: n=118; S784: n=112) contained either microsatellite.

Because analyzing microsatellite length on a capillary sequencer is expensive, we explored ways to reduce cost by screening for colonies that did have an insert before analyzing the microsatellites. One simple way to screen is to attempt to amplify *dhfr* from the colonies. For one set of transformations (38 colonies), we did the normal microsatellite analysis (using a capillary sequencer) and also determined the presence of *dhfr* by attempting to amplify it with PCR and then visualizing on an agarose gel. Of the 20 colonies for which *dhfr* amplified (band visible on the gel), 19 also had a clear peak in fluorescence at the microsatellite loci (**Table 2.6**). However, many colonies that gave no visible band on the gel *also* had a clear peak in fluorescence, presumably because the microsatellite analysis is more sensitive. To address this lack of sensitivity, one could do a nested PCR for amplifying *dhfr*, or amplify with

fluorescent primers and visualize with an appropriate UV camera system. Alternatively, one could avoid the capillary sequencer altogether and determine microsatellite length by using a high-percentage agarose gel [15].

Translation to the Field

This method will only be useful if it works for samples collected in the field. The most problematic step is likely to be the initial long-range PCR, because a successful amplification of such a long fragment requires a large amount of template DNA. To test the applicability of this protocol to field samples, we attempted to amplify the 6 kb fragment from DNA extracted from three different field sample types: 2 ml of whole blood from patients in Mae Sot, Thailand; 25 µl cultured parasites (50% hematocrit, ~9% parasitemia) spotted onto filter paper; ~100 µl blood from patients in Sri Lanka spotted onto filter paper [16]. Using the same PCR protocol as described in the methods, we successfully amplified the 6 kb fragment from 4/4 whole blood samples, 4/4 mock filter paper samples, and 1/4 Sri Lankan filter paper samples. We did not make any attempt to optimize the protocol for field samples, to use a nested PCR, or to amplify a fragment shorter than 6 kb. However, using a nested protocol would increase the inherent noise in the system due to polymerase slippage, thus reducing efficiency further.

Discussion

Prior haplotype analyses have been limited to monoclonal samples. Since transmission is high in many areas of Africa and Asia, polyclonal samples are the rule rather than the exception. These are also areas where the opportunity for recombination between different genotypes is highest, and it is of great interest to determine the relatedness of parasites from these areas, as well. With this in mind, we have designed a method that will allow haplotype analysis of multiclonal samples.

We have demonstrated that our method can determine a valid haplotype for mixed samples of *P. falciparum* DNA. Though this method can determine only one

haplotype per sample, it marks an improvement over current methods, which exclude mixed infections from analysis. With our method, one can use all samples from areas of intense transmission, rather than discarding most of them. However, because some alleles are amplified preferentially over others, it will be problematic to get accurate prevalence data for a particular haplotype. Rather, one can simply determine presence or absence of haplotypes.

The microsatellites in *P. falciparum* are almost all dinucleotide repeats. As a result, when one types a microsatellite in the capillary sequencer, in addition to the main peak (at the expected allele length) there are flanking peaks at lengths two bases off from the true length (**Figure 2.4a**). This stutter is routinely observed due to polymerase slippage during the PCR amplification of the microsatellites. Because the polymerase in the initial long-range PCR also slips, these small size differences are generated and then are stably maintained when the DNA is incorporated into the yeast plasmid. As a consequence, a few of the yeast colonies produced from a single strain of *P. falciparum* are products of stutter and show allele lengths slightly different from the parent strain. **Figure 2.4b** shows that the distribution of alleles among the yeast colonies reflects the shape of the stutter peaks observed in the microsatellite analysis of DNA from an individual colony. This effect does not preclude accurate identification of a correct haplotype, but multiple colonies need to be analyzed. SNPs or trinucleotide repeats would show less instability, and consequently would require analysis of fewer colonies. None were present in the genomic region we examined, but future studies using this method could use trinucleotide repeats or SNPs.

It is important to emphasize that none of the methods that depend on standard PCR amplification can determine the relative proportions of different alleles in a sample. There is a strong bias toward better amplification of particular alleles, usually the smallest of the size classes, and this was demonstrated in our experiments. Regardless of whether we analyzed the samples in the yeast system or analyzed them directly using the standard protocol, we found that the smaller K1 allele always amplified better than Dd2, even if we began with equal amounts of each. Because we

used a spectrophotometer to determine DNA concentration, it is possible that this skew reflects slight differences in the amount of input DNA. However, the dominance of K1 appeared across many DNA preparations and in experiments using DNA from MR4. Therefore, we believe this skew is not simply a reflection of the proportion of the two haplotypes in the starting templates.

A concern during both the long-range PCR [17] and the homologous recombination in yeast is that there could be crossover between two different strains of DNA, resulting in a plasmid that contained DNA from both strains. From the data in **Table 2.3** one could postulate that some crossover occurred; 6/14 samples had one or two microsatellite alleles that did not match the *dhfr* coding region genotype. In general, the switch from one strain to another occurred between the two microsatellites (the longest stretch between loci). However, in most cases one cannot distinguish between crossover and stutter. For example, one of the 3D7/K1 samples shows the K1 allele at all loci except S784. This haplotype is due either to double-crossover or to stutter at the S784 locus. Though the latter explanation is far more likely, one cannot exclude crossover as the cause. However, crossover is a relatively rare event; therefore, by recording only the most common haplotype for each sample we can be confident of avoiding false haplotypes.

One drawback to this method is its low efficiency; many of the colonies did not contain the target insert. A restriction digest of plasmid DNA extracted from these false-positive colonies indicated that these colonies had lost the URA3 gene without gaining the target insert (the long-range PCR product). Because the no-template-DNA control PCR product produced some (empty) colonies, we suspected that the yeast cells were using primer-dimers for gap repair rather than the 6 kb insert. Therefore, we redesigned the vector so that it had homology only to regions of *P. falciparum* DNA contained at the ends of the PCR product, not within the primers themselves. However, using this plasmid (pBDGR) instead of our original plasmid (pRSU) produced ten-fold fewer colonies for each transformation (~14 vs. ~140 per transformation), and the yield of true positive colonies was similar to or slightly less

than with pRSU. We tried numerous other variations on the protocol to increase efficiency (see **Appendix B**), but none led to an improvement over the protocol described in this chapter.

Another challenge to this protocol, particularly in applying it to samples collected in the field, is the large amount of DNA required for the initial long-range PCR. Though we were able to amplify the 6 kb fragment from blood spotted onto filter paper, the PCR was more successful for DNA extracted from whole blood. One could certainly attempt to optimize the PCR protocol for blood spotted onto filter paper, but it is likely that there would always be some filter paper samples that would fail to amplify. However, for many studies a fragment shorter than 6 kb would be sufficient. For example, the microsatellites within 100 bp upstream and 500 bp downstream of the *dhfr* coding region distinguish the triple mutant allele of *dhfr* that is thought to originate in Asia from other alleles that appear of have arisen in Africa [18-20]. Amplification from a filter paper sample of a fragment that would include both of these markers would be far easier, and would still be informative for many questions of interest. Thus, filter paper samples are likely to be amenable to this kind of analysis in many situations.

In summary, we have developed a method that allows determination of a valid haplotype from mixed samples of *P. falciparum* DNA. Though we tested the method using microsatellites and *dhfr*, it could be used for any type of genetic marker, and any target region of a genome. In fact, it would perform much better for SNPs because polymerase slippage would not affect SNPs. Likewise, trinucleotide repeats would be more stable in PCR than dinucleotide microsatellites, and therefore false "stutter" alleles would be less of a problem for this type of genetic marker. An obvious question is whether or not the method will work on field samples. We believe that it will, as we have successfully amplified the 6 kb fragment from dried blood samples taken in the field, and that is the step most prone to failure. We hope that the method will allow study of the population genetics of *P. falciparum* in regions that have hitherto been neglected due to the high proportion of mixed infections.

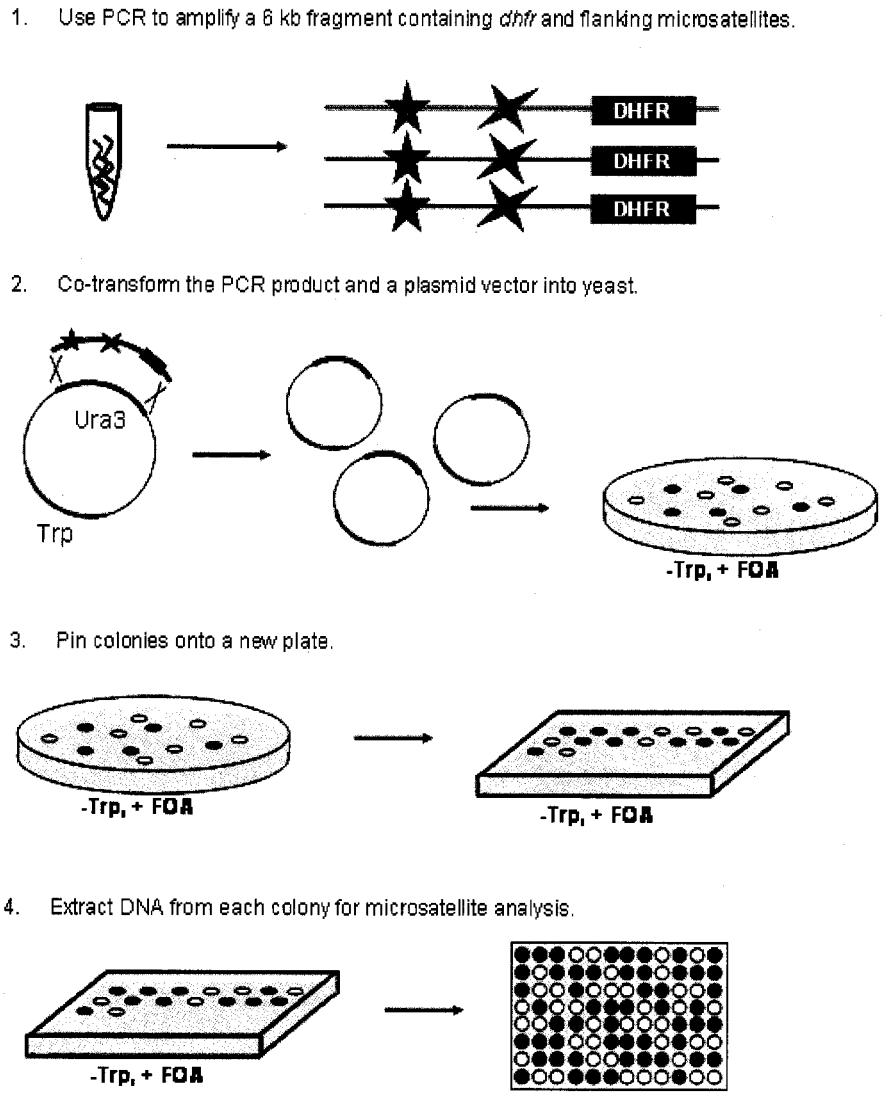


Figure 2.1 – A Schematic of the Method. The first step is to amplify a 6 kb fragment containing *dhfr* and upstream microsatellite markers (shapes). In step two, the PCR products are co-transformed with a cut plasmid vector into yeast and plated on selective media containing 5-fluoro-orotate (FOA). In step three, colonies are pinned onto a new plate to remove residual, unincorporated PCR product. The last step is to extract DNA from the colonies and analyze it at the genetic loci. The different colors indicate the different genotypes/haplotypes of *P. falciparum* present in the sample. The white colonies are those that grow on the selective plates but do not contain the insert (i.e. false positives).

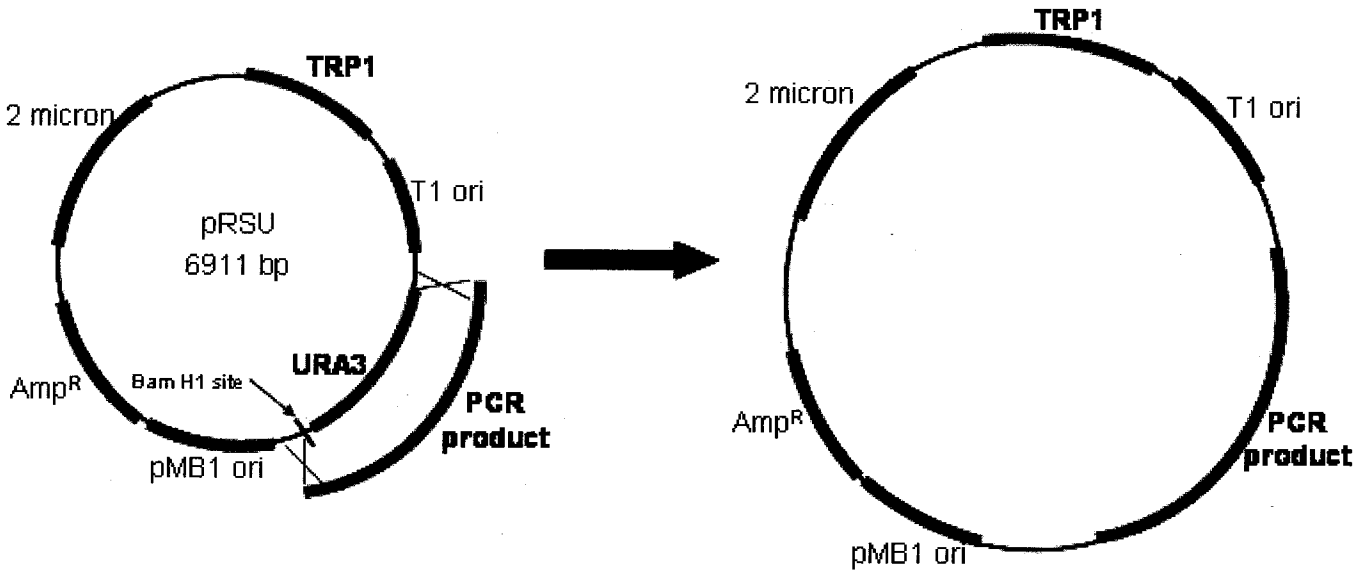


Figure 2.2 – Diagram of the Plasmid pRSU Before and After Transformation.

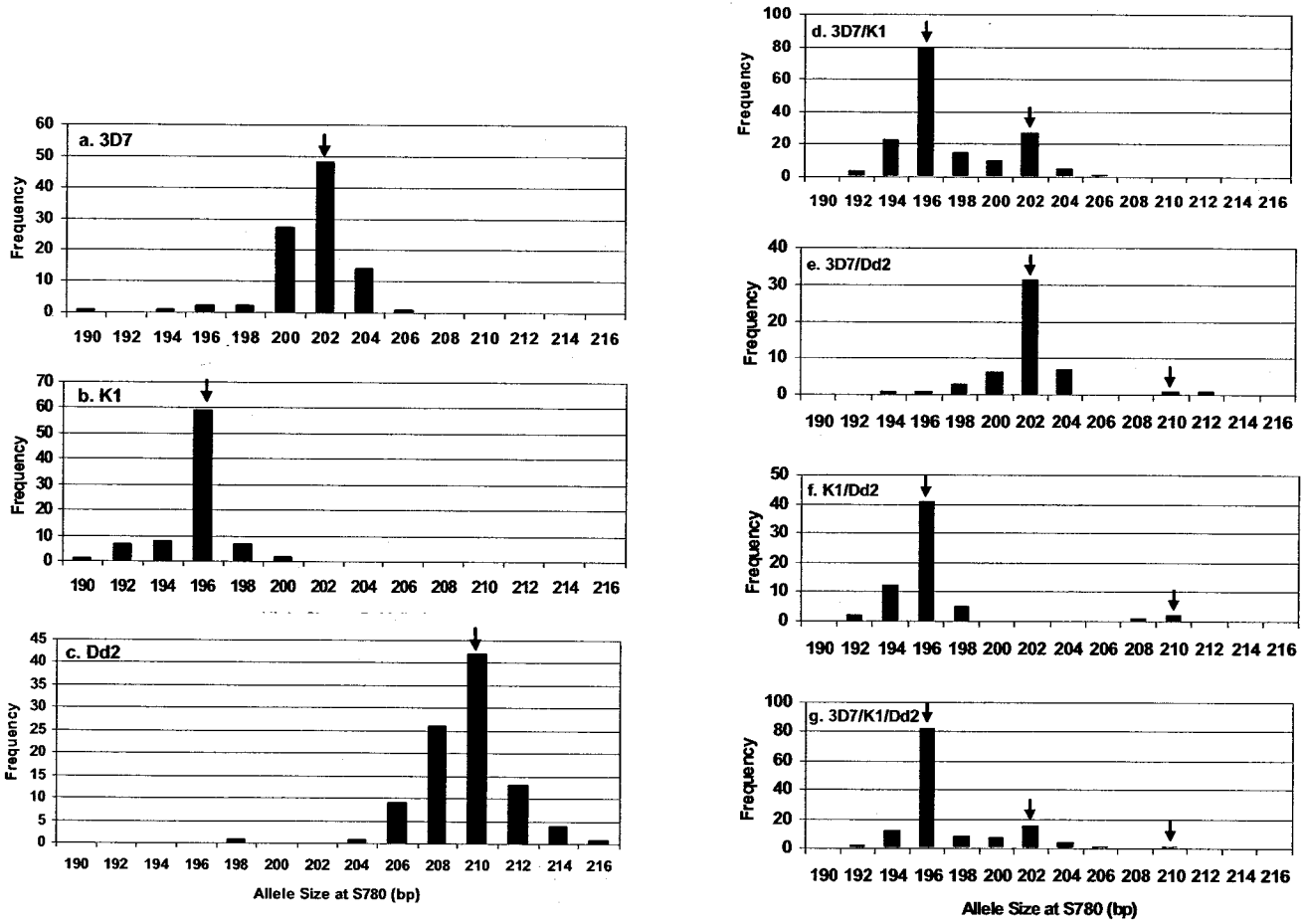


Figure 2.3 – Distribution of Alleles at Locus S780. Distribution of alleles (lengths) at locus S780 in yeast colonies containing the 6 kb fragment amplified from genomic DNA of the indicated strains (a-g). Arrows indicate the size of true alleles.

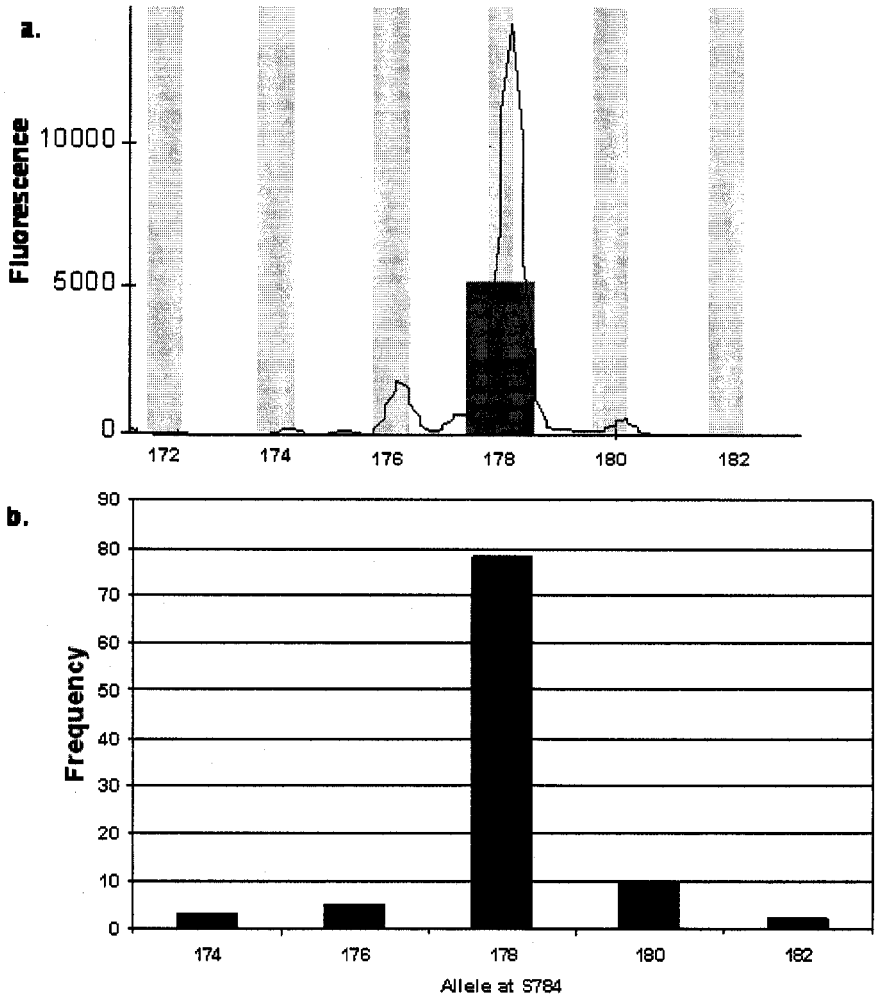


Figure 2.4 – Microsatellite Stutter. a. Sample capillary gel readout of analysis at locus S784 of DNA purified from a single yeast colony containing the 6 kb fragment amplified from K1 DNA; the large grey rectangle indicates the allele called by Genetic Profiler. b. Distribution of alleles (lengths) at locus S784 in all yeast colonies containing the 6 kb fragment amplified from genomic K1 DNA.

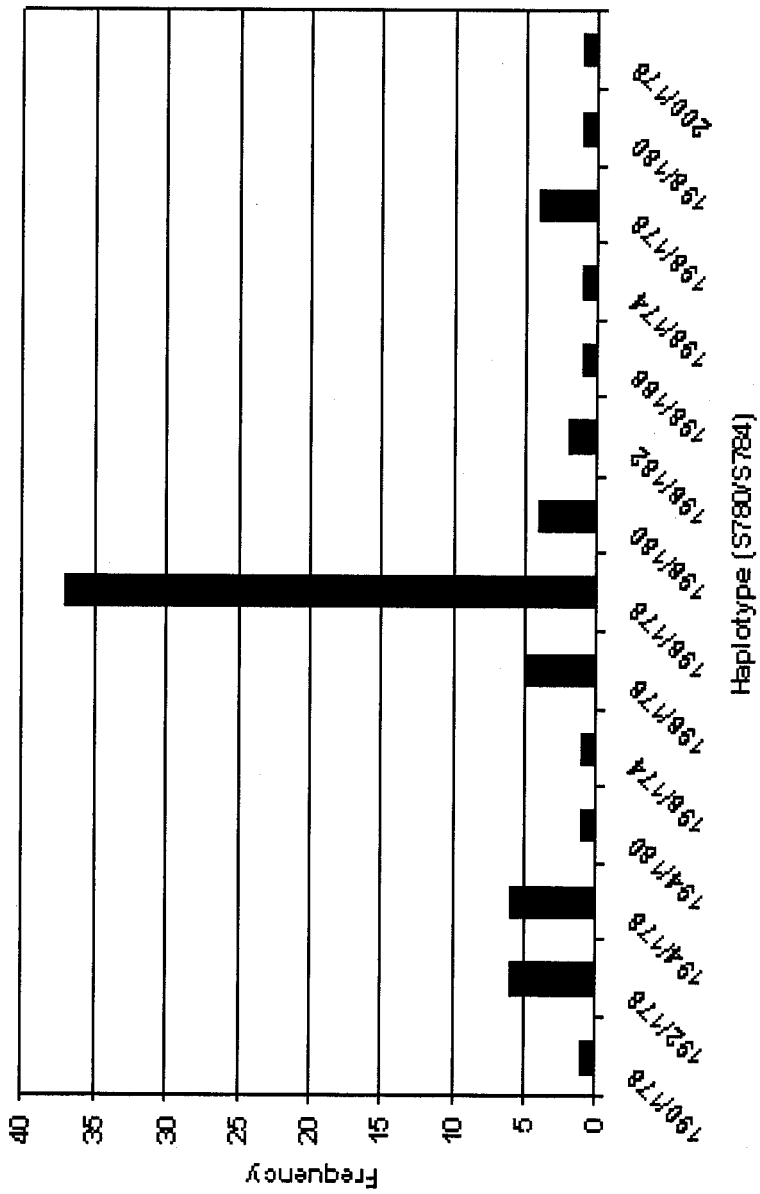


Figure 2.5 – Distribution of Haplotypes. Haplotypes in yeast colonies containing the 6 kb fragment amplified from genomic K1 DNA.

Table 2.1 – PCR Primers Used in the Protocol.

Reaction	Forward Primer	Reverse Primer
Long-range PCR	5'-aca aac cga agt tat ctg atg tag aaa agg att aaa gat gct acg tgg tat gaa atg ccg act cct cca atc tca tgt aga-3'	5'-gag ctc cac cgc ggt ggc ggc cgc tct aga act agt gga tct cat atg aca tgt atc ttt gtc atc att ctt taa agg cat atc-3'
Long-range PCR, greater homology to plasmid vector	5'-tga caa tac aga cga tga taa caa acc gaa gtt atc tga tgt aga aaa gga tta aag atg cta cgt ggt atg aaa tgc cga ctc ctc caa tct cat gta g-3'	5'-cta aag gga aca aaa gct gga gct cca ccg cgg tgg cgg ccg ctc tag aac tag tgg atc tca tat gac atg tat ctt tgt cat cat tct tta aag gca t-3'
S780 (Nair *et al.*, 2003)	FAM-5'-acagttataagatttaatgcaa-3'	5'-actgatgaaattgtaaatga-3'
S784 (Nair *et al.*, 2003)	FAM-5'-atattccaacattttcaaga-3'	5'-atttttgctttcaaccttac-3'

Table 2.2 – Proportion of Colonies Containing the Insert. Insert presence indicated by a clear peak in fluorescence at a specific length on the capillary gel readout for each microsatellite locus.

	Colonies analyzed		Colonies with a clear peak in fluorescence (%)		Analyzed at both loci	Clear peak at both loci (%)
Template	**S780**	**S784**	**S780**	**S784**		
3D7	329	350	96 (29)	114 (33)	314	83 (26)
K1	237	240	83 (35)	98 (41)	233	71 (30)
Dd2	617	606	97 (16)	108 (18)	529	67 (13)
3D7/K1	524	520	164 (31)	193 (37)	499	130 (26)
3D7/Dd2	482	496	51 (11)	71 (14)	441	37 (8)
K1/Dd2	454	458	63 (14)	78 (17)	434	37 (9)
3D7/K1/Dd2	616	609	133 (22)	145 (24)	574	72 (13)
Total	3259	3279	687 (21)	807 (25)	3024	497 (16)
H2O	118	112	0 (0)	0 (0)		

Table 2.3 – Haplotypes and DHFR Genotypes for a Subset of Colonies.

	Alleles (Length in bp)		DHFR Genotype (Codons)		
Template DNA	**S780**	**S784**	**51**	**59**	**108**
K1	196	180	N	R	N
K1	196	176	N	R	N
K1	196	178	N	R	N
K1	196	178	N	R	N
3D7	202	174	N	C	S
3D7	196	174	N	C	S
3D7	200	174	N	C	S
3D7	202	174	N	C	S
Dd2	210	178	I	R	N
Dd2	210	178	I	R	N
Dd2	210	176	I	R	N
Dd2	208	178	I	R	N
3D7K1	202	174	N	C	S
3D7K1	202	174	N	C	S
3D7K1	196	178	N	C	S
3D7K1	196	174	N	R	N
3D7K1	202	178	N	R	N
3D7K1	202	178	N	R	N
3D7K1	196	180	N	R	N
3D7K1	196	172	N	C	S
3D7Dd2	202	174	N	C	S
3D7Dd2	202	172	N	C	S
3D7Dd2	204	174	N	C	S
K1Dd2	210	176	I	R	N
K1Dd2	210	178	N	R	N
3D7K1Dd2	210	180	I	R	N

Table 2.4 – First and Second Most Common Haplotypes for Each Sample.

	Most common haplotype			Second most common haplotype		
Sample	**S780/S784**	**Frequency**	**%**	**S780/S784**	**Frequency**	**%**
3D7	202/174	27/83	33%	200/174	15/83	18%
K1	196/178	37/71	52%	192/178;194/178	6/71	8%
Dd2	210/178	19/67	28%	208/178	11/67	16%
3D7/K1	196/178	35/130	27%	196/174	11/130	8%
3D7/Dd2	202/174	15/37	41%	202/178;204/174	3/37	8%
K1/Dd2	196/178	17/37	46%	194/178	5/37	14%
3D7/K1/Dd2	196/178	27/72	38%	196/176	11/72	15%

Table 2.5 – Haplotypes for MR4 Samples.

	Allele (Length in bp)	
Sample (MR4)	**S780**	**S784**
3D7	202	174
K1	196	178
Dd2	210	178
W2	210	178
V1/S	196	178

Table 2.6 – Testing for the Presence of *Dhfr* vs. Testing for the Presence of Microsatellite Loci.

	S780		S784	
DHFR	Clear peak	No clear peak	Clear peak	No clear peak
Present	19	1	19	1
Absent	9	9	12	6

NOTES TO CHAPTER TWO

1. Stephens, M., Smith, N.J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**, 978-89 (2001).

2. Stephens, M. & Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* **73**, 1162-9 (2003).

3. Takala, S.L. et al. A high-throughput method for quantifying alleles and haplotypes of the malaria vaccine candidate Plasmodium falciparum merozoite surface protein-1 19 kDa. *Malar J* **5**, 31 (2006).

4. Sakihama, N., Mitamura, T., Kaneko, A., Horii, T. & Tanabe, K. Long PCR amplification of Plasmodium falciparum DNA extracted from filter paper blots. *Exp Parasitol* **97**, 50-4 (2001).

5. Christianson, T.W., Sikorski, R.S., Dante, M., Shero, J.H. & Hieter, P. Multifunctional yeast high-copy-number shuttle vectors. *Gene* **110**, 119-22 (1992).

6. Jansen, G., Wu, C., Schade, B., Thomas, D.Y. & Whiteway, M. Drag&Drop cloning in yeast. *Gene* **344**, 43-51 (2005).

7. Hua, S.B., Qiu, M., Chan, E., Zhu, L. & Luo, Y. Minimum length of sequence homology required for in vivo cloning by homologous recombination in yeast. *Plasmid* **38**, 91-6 (1997).

8. Oldenburg, K.R., Vo, K.T., Michaelis, S. & Paddon, C. Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Res* **25**, 451-2 (1997).

9. Raymond, C.K., Sims, E.H. & Olson, M.V. Linker-mediated recombinational subcloning of large DNA fragments using yeast. *Genome Res* **12**, 190-7 (2002).

10. Schiestl, R.H. & Gietz, R.D. High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr Genet* **16**, 339-46 (1989).

11. Gietz, D., St Jean, A., Woods, R.A. & Schiestl, R.H. Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res* **20**, 1425 (1992).

12. Gietz, R.D., Schiestl, R.H., Willems, A.R. & Woods, R.A. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast* **11**, 355-60 (1995).

13. Nair, S. et al. A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. *Mol Biol Evol* **20**, 1526-36 (2003).

14. Duraisingh, M.T., Curtis, J. & Warhurst, D.C. Plasmodium falciparum: detection of polymorphisms in the dihydrofolate reductase and dihydropteroate synthetase genes by PCR and restriction digestion. *Exp Parasitol* **89**, 1-8 (1998).

15. Mwangi, J.M., Omar, S.A. & Ranford-Cartwright, L.C. Comparison of microsatellite and antigen-coding loci for differentiating recrudescing Plasmodium falciparum infections from reinfections in Kenya. *Int J Parasitol* **36**, 329-36 (2006).

16. Hapuarachchi, H.C. et al. Point mutations in the dihydrofolate reductase and dihydropteroate synthase genes of Plasmodium falciparum and resistance to sulfadoxine-pyrimethamine in Sri Lanka. *Am J Trop Med Hyg* **74**, 198-204 (2006).

17. Tanabe, K. et al. In vitro recombination during PCR of Plasmodium falciparum DNA: a potential pitfall in molecular population genetic analysis. *Mol Biochem Parasitol* **122**, 211-6 (2002).

18. Roper, C. et al. Antifolate antimalarial resistance in southeast Africa: a population-based analysis. *Lancet* **361**, 1174-81 (2003).

19. Roper, C. et al. Intercontinental spread of pyrimethamine-resistant malaria. *Science* **305**, 1124 (2004).

20. McCollum, A.M. et al. Antifolate Resistance in Plasmodium falciparum: Multiple Origins and Identification of Novel dhfr Alleles. *J Infect Dis* **194**, 189-97 (2006).

CHAPTER THREE: DRUG RESISTANCE HAPLOTYPES OF SAMPLES COLLECTED IN KILIFI, KENYA, 1993-1995

As described in Chapter 1, prior studies of SP resistance were hampered not only by the inability to include mixed infections but also by the focus on samples collected well after SP resistance emerged. While the reduced diversity of parasites carrying three mutations in *dhfr* ("triple-mutants") in contemporary samples is intriguing [1-3], ideally we would track the parasites' haplotypes as resistance emerged. Only by determining haplotypes for samples collected as SP resistance developed can we determine how often the triple-mutant arose.

Consider the following: though samples from South Africa in 1996 and 1999 indicated that all triple-mutants in South Africa shared the Asian triple-mutant haplotype [3,4], samples from Kenya (Kisumu) and Senegal in 2002-2004 showed multiple haplotypes for the triple-mutant [5,6]. Because earlier samples from Kenya and Senegal were not examined, it is unclear when the alternative triple-mutant haplotypes appeared. Therefore, we cannot choose between several possible explanations for the presence of these haplotypes. They could be the result of recombination of the Asian triple-mutant with local parasites. Or, they could represent mutation and diversification of the Asian haplotype. Finally, they could be local, independent origins of the triple-mutant *dhfr* allele, arising either before or after the arrival of the Asian triple-mutant. Furthermore, even if we determine that the novel haplotypes in Kisumu and Senegal are derived from the Asian haplotype, it remains unknown whether the current triple-mutant found in both Asia and Africa is the only one ever to have existed, or whether it simply out-competed all other triple-mutant strains of *P. falciparum.* To explore this question, we analyzed samples collected in Kilifi, Kenya, in 1993-1995, just after SP was first introduced [7].

Kilifi is a small town on the coast of Kenya (**Figure 3.1**) with a population of ~30,000. It is the capital of Kilifi District, an area of the Coast Province that covers ~2500 km^2 from near Mombasa in the south, up the coast towards Malindi, and ~10

km inland [8]. It is the second poorest district in Kenya, and had a population of ~544,000 in 1999. Though the district hospital in Kilifi serves this entire population, the population within the 700 km^2 nearest the hospital is ~200,000 [8,9]. Malaria transmission is continuous, but the most intense transmission follows the two rainy seasons, April to July and November to December [8]. Severe disease is common, as are multiple infections [7,10]. Kilifi District has an average entomologic inoculation rate (EIR; infective mosquito bites per person) of four per year [11], but can be up to 100 bites per year in the southern part of the district [8]. In 1989 malaria research began in earnest in Kilifi, with collaboration between the Kenya Medical Research Institute and the University of Oxford. Since 1994 the Wellcome Trust has operated a research unit in Kilifi at the District Hospital [12]. Because of this longstanding research collaboration, there are samples from Kilifi from both before and after the introduction of SP as standard treatment for malaria.

Prior to the early 1990s, chloroquine was the standard treatment for uncomplicated malaria infection. Though in the early 1980s chloroquine was completely effective [13], by 1984 a study in Malindi showed 28% chloroquine resistance [14], and by 1990 there was 60% chloroquine resistance near Mombasa [15]. Because of the increasing resistance to chloroquine, the Kilifi District Hospital switched to SP as its first-line treatment in ~1992 [7,16]. Chloroquine remained available in local shops [17], but SP was the standard treatment at Kilifi District Hospital. Prior to the switch to SP, the only patients receiving that antifolate were those participating in research studies. Though these studies indicated some resistance to pyrimethamine [7,18,19], the SP combination was completely effective in this region as of 1990 [15].

With its relatively well-defined drug use history, Kilifi is a prime location in which to study the emergence of SP resistance. We chose to examine samples collected in 1993-1995, reasoning that these samples would capture the very beginning of any SP resistance in Kilifi. To determine the prevalence of SP-resistance alleles, we genotyped *dhfr* and *dhps* for the point mutations that confer resistance to pyrimethamine and sulfadoxine (see below). To compare these resistant strains to

those from prior studies, we determined haplotypes from the microsatellites near *dhfr*. We also determined the prevalence of the chloroquine resistance allele in these samples.

Methods

The samples used for this analysis were collected in Kilifi, Kenya, as part of a clinical trial comparing chlorproguanil-dapsone (LapDap®) with SP [20]. LapDap® is a combination antifolate that targets *dhfr* and *dhps*, as SP does. The mutations in *dhfr* that confer resistance to pyrimethamine likewise confer some resistance to cycloguanil [21-23], but LapDap® remains clinically effective against parasites with three mutations in *dhfr* [24,25]. From July 1993 through April 1995, children with uncomplicated malaria were randomized to treatment arms. Venous blood was collected at enrollment and during 28 days of follow-up. One to three milliliters of venous blood was collected from each patient and cryopreserved [26]. A few samples are from the same patient, but collected at different times. Pairs 37/83, 546/601, 585/632, 686/713, and 696/716 each represent two samples from a single patient (enrollment and follow-up samples). Informed consent for the original study was assured by the Kenya Medical Research Unit, Kilifi, and the present study of anonymous samples is in the exempt category of the University of Washington Human Subjects Review Board. Samples were stored, frozen, in Nairobi. In September 2005, I traveled to Nairobi, thawed a subset of samples (N=85), and extracted DNA from 10-100 µl of thawed whole blood. Samples were chosen based on parasitemia, having a previously determined *dhfr* genotype, and the amount of blood remaining. DNA was extracted using a QIAamp DNA Micro Kit (Qiagen, Valencia, CA) using inversion instead of vortexing to minimize shearing of DNA [27]. DNA was eluted twice from the column using 30 µl of water each time. Eluted DNA was then allowed to dry at room temperature before shipment back to the United States for analysis. To reconstitute the samples, I added 30 µl of water and incubated at 37 °C for approximately 1-3 hours.

To genotype the samples at *dhfr*, I used the method of Duraisingh [28]. This method amplifies *dhfr* in two fragments and then digests each separately with several restriction enzymes (**Figure 3.2**). The banding pattern produced by the restriction enzyme digests indicates the *dhfr* genotype(s) of that sample. The first round of PCR contained the following: 1x Buffer A for FailSafe, 1 μM each primer (**Table 3.1**), 0.5 units FailSafe enzyme mix (Epicentre, Madison, WI), and 5 μl template DNA. Cycling conditions were: 94 °C for 3 min; 35 cycles of 94 °C for 30 sec, 55 °C for 45 sec, 72 °C for 60 sec; 72 °C for 10 min. Second round reactions contained: 1x buffer, 2.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.5 μM each primer (**Table 3.1**), 0.5 units Taq (Fermentas, Hanover, MD), and 1 μl 100x dilution of the first-round product. Cycling conditions for the second round were: 94 °C for 3 min; 30 cycles of 94 °C for 1 min, 45 °C for 1 min, 72 °C for 1 min; 72 °C for 10 min.

Five microliters of the second-round PCR products were digested with *Alu* I, *Tsp509* I, *Xmn* I, and *Dra* I according to the manufacturer's instructions (New England Biolabs, Ipswich, MA). The restriction enzyme digests were visualized on 3% 3:1 agarose gels containing ethidium bromide. If a sample gave any hint of multiple restriction patterns on the gel it was considered mixed. **Figure 3.3** shows gels of the digests used to determine genotypes at codons 51 (*Tsp509* I), 59 (*Xmn* I), and 108 (*Alu* I) of *dhfr*. Recall that pyrimethamine resistance is conferred by point mutations at codons 51 (asparagine to isoleucine), 59 (cysteine to arginine), 108 (serine to asparagine), and 164 (isoleucine to leucine) of *dhfr* confer resistance to pyrimethamine (**see Figure 1.7**). As the number of mutations increases so does resistance [29], such that a patient infected with a parasite carrying three mutations in *dhfr* is more likely to fail treatment with SP [24,30].

The *dhps* and *pfcrt* alleles were identified directly (by Marnie Briceño) by DNA sequence analysis. For dhps, we examined codons 436, 437, 540, 581, and 613, as changes at these codons have been associated with sulfadoxine resistance [24,31,32]. In Africa, changes at codon 437 (alanine to glycine) and codon 540 (lysine to glutamate) are the most common resistance mutations [24,33,34]. For *pfcrt* we sequenced

only the region containing codons 72-76, as a change from lysine to threonine at codon 76 is the main change associated with chloroquine resistance in *P. falciparum* [35]. Wild-type African parasites have the haplotype CVMNK at these codons; chloroquine-resistant parasites have the haplotype CVIET [36,37].

PCR conditions for *dhps* were: 20 µl reaction volume containing 1x buffer, 2mM $MgCl_2$, 0.25 mM dNTP's, 0.25 µM each primer (**Table 3.1**), 0.5 units Taq (Fermentas, Hanover, MD), and 1-4 µl template [28,38]. The product of the S1224/S1225 reaction is 770 bp; the products of the S1224/S1263 and S1262/S1225 reactions are 515 and 287 bp, respectively (**Figure 3.4**). Cycling conditions for the 770 bp product were: 94 °C for 3 min; 35 cycles of 94 °C for 1 min, 51 °C for 2 min, 70 °C for 1 min; 70 °C for 5 min [38]. For samples for which the 770 bp fragment failed to amplify we attempted to amplify *dhps* in two pieces, using S1224/S1263 and S1262/S1225. For samples that still did not amplify and/or sequence properly we used the following cycling conditions: 94 °C for 3 min; 35 cycles of 94 °C for 45 sec, 51 °C for 45 sec, 65 °C for 45 sec; 65 °C for 5 min. Sequencing was done using the same protocol as *pfcrt* (see below), with the exception that the GeneClean step was omitted.

PCR conditions for *pfcrt* were: 20 µl reaction volume containing 1x buffer, 1.5 mM $MgCl_2$, 0.25 mM dNTP's, 0.2 µM each primer, 0.5 units Taq (Fermentas, Hanover, MD) [37]. Cycling conditions were: 94 °C for 3 min; 35 cycles of 94 °C for 30 sec, 56 °C for 30 sec, 62 °C for 1 min; 62 °C for 5 min; 15 °C for 5 min [37,39]. PCR products were cleaned using the GeneClean Kit (Q-Biogene, Morgan Irvine, CA), then sequenced using the following conditions: 10 µl reaction consisting of 1.7 µl 5x BigDye buffer (400 mM Tris-HCl, 10 mM $MgCl_2$, pH = 9.0), 1 µl BigDye (Applied Biosystems, Foster City, CA), 0.8 µM primer (same primers as for PCR amplification), and 1-3 µl PCR product. Products of sequencing reactions were separated on an ABI 3130xl (Applied Biosystems, Foster City, CA) and analyzed using Sequencher (Gene Codes Corporation, Ann Arbor, MI). Samples were

considered mixed if the fluorescent signal for the minor genotype was at least 10% the height of the signal for the predominant genotype.

Each sample was genotyped at five microsatellite loci, located 5.0, 3.8, and 0.1 kb upstream of the start of *dhfr-ts*, and 0.5 and 5.9 kb downstream of the end of *dhfr-ts*. These loci were also used by Roper *et al.* (2003, 2004) and Nair *et al.* (2003). For each microsatellite, I amplified the locus using PCR and analyzed product size on a MegaBACE 1000 DNA Analysis System with Genetic Profiler (Amersham, Piscataway, NJ). Product size was recorded as the allele for that sample. The primers for the -5.0 locus are given in Roper *et al.* (2003); the primers for the remaining loci are given in Nair *et al.* (2003). All primers are also in **Table 4.1**. The names (distances from *dhfr*) of the microsatellites in our study are slightly different from those in the other studies due to differences in how the distances from *dhfr-ts* were calculated (e.g. from the first codon of *dhfr-ts vs.* from codon 108).

PCR amplification of the microsatellites was done in a 96-well plate format, one 10 µl reaction per well, each well containing 1x buffer for Taq, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.4 µM forward (labeled) primer, 0.4 µM reverse primer, 0.15 units Taq polymerase (Promega, Madison, WI), and 1 µl DNA. Cycling conditions were: 94 °C for 2 minutes; 25 cycles of 94 °C for 30 seconds, 45 °C for 30 seconds, 60 °C for 30 seconds; 60 °C for 2 minutes [2]. A nested PCR was not necessary because almost all amplifications were positive with a single step PCR. Those samples that had more than one allele at any microsatellite locus were considered mixed. Data for monoclonal and mixed samples are presented separately, as unambiguous haplotypes could only be assigned for the clonal samples. The designation of allele sizes in **Tables 3.4 and 3.5** does not match exactly those reported in the previous studies; each allele of the triple-mutant haplotype is one or two bases longer in our samples. The differences are likely due to differences in the capillary gel analyzers in the different laboratories.

Dhfr-haplotype data for all clonal samples was analyzed using the program Network (www.fluxus-engineering.com). Microsatellite differences were given the

default weight of 10. Single nucleotide polymorphisms were given a weight of 90, to reflect their presumed slower mutation rate, or a weight of zero, to remove them from the analysis and thus remove any assumption about the frequency of origin of *dhfr* mutants. Epsilon was set at either 0 or 11, to explore the amount of homoplasy in this data set [40]. (See results section for a more complete explanation of Network.)

Genotyping results

Of the 85 samples collected, 22 were wild-type at *dhfr*, 1 was a single-mutant (S108N), 21 were double-mutants (13 N51I/S108N; 8 C59R/S108N), 12 were triple-mutants (N51I/C59R/S108N), and 29 were mixed at *dhfr*. There were no quadruple-mutants: all samples were wild-type (isoleucine) at codon 164. Of the 56 samples with a single *dhfr* genotype, 14 were mixed at one or more microsatellite loci (8 wild-type, 1 single, 2 double, 3 triple), leaving 42 samples with unambiguous haplotypes (**Table 3.2**). Separating the samples into groups by year shows that perhaps the triple-mutant increased in frequency during these years (**Figure 3.5**). However, the proportion of triple-mutants in 1993 (1/15) is not significantly less than in 1994-1995 (8/27; p=0.09, one-sided Fisher's exact test). In addition, the later samples may contain more samples that were collected after treatment with SP or LapDap®.

Drug treatment data was available for 47/85 samples. Of those, 32 were collected at enrollment, seven were collected after treatment with SP, and eight were collected after treatment with LapDap® (**Table 3.3**). A comparison of the two pre-treatment groups indicates no significant difference with respect to *dhfr* genotypes (p=0.6, two-sided Fisher's exact test). The two post-treatment groups are likewise not significantly different (p=0.7, two-sided Fisher's exact test). For SP, the difference in distribution of *dhfr* genotypes between the pre- and post-treatment groups did not reach significance (p=0.12, two-sided Fisher's exact test), nor did the change in proportion of mixed infections before and after treatment (p=0.13; two-sided Fisher's exact test). In contrast, the difference in distribution of *dhfr* genotypes before and after treatment with LapDap® was significant (p=0.003, two-sided Fisher's exact test).

Combining both treatment groups, samples collected after treatment were significantly less likely to be mixed (p=0.0002, two-sided Fisher's exact test).

The results of the microsatellite analysis for the clonal samples are shown in **Table 3.4**. As in previous studies [2-4,41], all of the triple-mutants share the same haplotype, and it appears to be the same haplotype as previously identified in samples from Asia and Africa [4]. However, there is some variation among the triple-mutants, indicating that recombination or mutation has occurred since the introduction of this strain to Kenya. There is little variation among the double-mutants. Each type of double-mutant (51/108 or 59/108) has a single haplotype, and neither double-mutant haplotype shares any alleles with the triple-mutant consensus haplotype. Wild-type samples showed a variety of haplotypes, as expected. Of note, the locus immediately upstream of *dhfr* is extremely informative; all of the triple-mutants had allele 109, whereas all the other samples (with one exception) had allele 89.

A plot of expected heterozygosities at each locus is shown in **Figure 3.6.** Expected heterozygosity is the probability that two alleles drawn at random from the population will be different. A higher value indicates greater population diversity. The wild-type parasites show a significantly higher heterozygosity than any other genotype (p=0.03 for each comparison, one-sided Wilcoxon signed rank test). The triple-mutant shows higher heterozygosity than either double-mutant, but this difference is not significant.

The data for the mixed samples are presented in **Table 3.5.** Though without phasing the haplotypes it is impossible to tell exactly which strains make up each sample, in general one sees combinations of the haplotypes found in the clonal samples. For example, samples 686, 694, and 774 are likely mixtures of the triple-mutant and the 51I/108N double-mutant. One must remember when examining these data that *dhfr* genotypes were determined by RFLP analysis on standard ethidium bromide stained gels, which is less sensitive than capillary electrophoresis. Therefore, a sample that appears clonal at *dhfr* may in fact be mixed, yet below our level of detection. Likewise, certain microsatellite alleles amplify better than others (see

Chapter 2), so the absence of an allele in the data does not prove the absence of that allele in the sample.

The *dhps* genotypes of the samples listed in **Table 3.4** were determined, and the results are presented in **Table 3.6.** Samples 41 and 779 failed to amplify, and sample 629 showed a mixed genotype at *dhps*, leaving 39 samples that showed a single genotype at both *dhfr* and *dhps*. The majority of parasites were wild-type at *dhps*, but there were seven double-mutants (A437G/K540E). All seven of the *dhps* double-mutants were in parasites that also carried a mutant *dhfr*. The frequency of *dhps* double-mutants was not significantly different between the *dhfr* double- and triple-mutants (p=1, two-sided Fisher's exact test). The absence of *dhps* double-mutants among those parasites with a wild-type *dhfr* was marginally significant (p=0.03, one-sided Fisher's exact test).

The *Pfcrt* genotype of all samples was also determined; the genotypes are presented in **Table 3.7**. Of the 82 samples that we sequenced successfully (3/85 failed to amplify), 56 contained a single *pfcrt* genotype: 45 CVIET (chloroquine resistant) and 11 CVMNK (chloroquine sensitive). The remaining 26 samples were mixed; that is, the sequencing output showed both SNPs at codons 74-76. For all but three of these, the CVIET appeared to be the predominant allele. Comparing the last two columns of **Table 3.7** shows that the distribution of *dhfr* genotypes is the same for both *pfcrt* genotypes (p=1, two-sided Fisher's exact test). In other words, there is no linkage disequilibrium between these two loci.

Evolutionary analysis

To examine further the relationship between the *dhfr* haplotypes in this sample set, I used the program Network (www.fluxus-engineering.com) to construct a network of the haplotypes [40]. A network of haplotypes is similar to a phylogenetic tree, except that a network shows all possible evolutionary histories, whereas a tree shows only one. Such networks have been used to examine human Y-chromosome [42] and mitochondrial DNA [43], and are useful for genetic data sets in which multiple

equally likely evolutionary histories exist. A tree would connect all sequences without creating any "cycles;" it would show a specific population history. A network, on the other hand, combines all minimum spanning trees, so that all likely connections (histories) are represented. Each node is a haplotype, with edges joining closely related haplotypes. The length of an edge is proportional to the number of genetic changes between the two haplotypes it connects. In calculating distance, Network assumes a stepwise mutation model for the microsatellites; that is, a change from 89 to 109 at locus -0.1 creates a longer edge than a change from 89 to 97. Though this model may not be completely accurate for *P. falciparum* [44], it mainly affects the length of the edges, not the over topology. Moreover, though perhaps not all allele changes are stepwise, it seems probable that a change from 200 to 202 (for example, at locus -3.8) represents less evolutionary distance than a change from 200 to 192.

The results of the analysis are shown in **Figures 3.7-11**. **Figures 3.7 and 3.8** show the result of the analysis with the SNP weight equal to 90 and epsilon set at zero (see below for an explanation of epsilon). Because Network assumes no recombination, I analyzed the data both with (**Figure 3.7**) and without (**Figure 3.8**) sample 779. This sample could be the product of recombination between the triple-mutant haplotype and the 51I/108N double-mutant haplotype. As expected, the triple-mutant samples cluster together, away from all other samples. When sample 779 is included, it links the triple-mutants to the double-mutants. However, when it is excluded the triple-mutants are connected to the other set of double-mutants, suggesting that any connection is spurious. I excluded sample 779 from all further analyses. (For the curious: if sample 779 is included and epsilon is increased to 11, keeping the SNP weight at 90, it links the triple-mutants to the "blue" double-mutants.)

Figure 3.9 shows the effect of increasing epsilon to a value greater than ten, the weight given microsatellite allele differences. Increasing the value of epsilon has the effect of "fuzzifying" [40] the tree and showing homoplasy. Epsilon is the genetic distance within which median vectors are constructed [45]; increasing epsilon shows

more possible median vectors (inferred haplotypes or nodes) between sampled haplotypes. For these data, setting epsilon >10 produced numerous reticulations among the wild-type population. This result indicates that the underlying assumptions of the model (no recombination; no independent mutations to the same allele) are incorrect, and that there is extensive homoplasy. Each inferred node can be interpreted as a haplotype present in the population that has not yet been sampled, or an ancestral haplotype that was once present but is no longer. The former seems likely, since there are no duplicate haplotypes among the wild-type parasites, and it is therefore unlikely that our sample of the wild-type population captures the full diversity of the population. In addition, the reticulations appear only among the wild-type haplotypes, suggesting that we have adequately sampled the resistant parasite population.

Figure 3.10 shows the effect of setting the SNP weight equal to zero, effectively removing *dhfr* genotypes from the analysis. By removing the SNPs from the Network analysis, we avoid introducing any bias about how often the triple-mutant parasite arose. The overall topology of the network does not change; the triple-mutants still cluster away from the other samples. **Figure 3.11** sets epsilon back to zero, and the basic topology still holds. In summary, the topology of the network remains similar for all analyses, yet the connection of the triple-mutants to the rest of the samples changes with each parameter change. This result supports the hypothesis that the triple-mutant represents an invading haplotype that is unrelated to the other haplotypes in the sample set.

Discussion

The initial observation of a single haplotype for all parasites carrying three mutations in *dhfr* came from two studies in Africa and Southeast Asia [2,3]. The African samples came from two locations: KwaZulu-Natal, South Africa, in 1995-99 and the Kilimanjaro region of Tanzania in 2001 [3]. In both locations, SP had been widely available for some time and its effectiveness was known to be diminishing [3]. The samples from Asia were collected in a variety of locations, but SP had long been

ineffective in that whole region [2,29]. Our goal was to analyze samples from a different location and to determine the haplotype of the triple-mutant alleles present in Kilifi before intensive SP use. Kilifi district is one of the poorest in Kenya, and SP was first used in Kilifi in the District Hospital in 1992-1993 [7,16]. These factors support the contention that there was virtually no use of SP before that time in the local population. To our surprise, we found the same haplotype pattern among the alleles of *dhfr* as in studies of more contemporary samples. Even at this very early time after SP introduction, there was limited genetic diversity among the resistant parasites. Furthermore, the triple-mutant parasites we examined appear to have the same haplotype as triple-mutants previously identified in South Africa and Tanzania[4], and in South East Asia [2].

The triple-mutant we found in our samples is presumed to have arisen in Asia and subsequently spread to Africa [4]. Therefore, either this mutant had already out-competed the "local" Kenyan triple-mutants by 1993, or there were never any "local" triple-mutants. Of note, both the 51I/108N and the 59R/108N double-mutants show little diversity in their haplotypes, comparable to or lower than that of the triple-mutants (**Figure 3.6**). This lack of diversity, while not statistically significant, suggests that these double-mutants emerged more recently than the triple-mutant. In addition, the double-mutants are clearly derived from local wild-type strains, as they have allele 89 at locus -0.1, and share other flanking alleles as well. The Network analyses support this derivation (**Figures 3.7-11**). If we accept that triple-mutants evolve from existing double-mutants, then an African triple-mutant probably had not emerged by the time these samples were collected.

Though the triple-mutant may be explained by immigration, the double-mutants appear local. If so, then the significant levels of double-mutants of both genotypes (51I/108N, 59R/108N) immediately after the introduction of SP for malaria treatment indicates that these alleles were present in the population before widespread SP use. One explanation is the use of the antibiotic combination trimethoprim-sulfamethoxazole (cotrimoxazole). The two drugs in this combination target *dhfr* and

dhps, respectively, and *in vitro* studies demonstrate cross-resistance between trimethoprim and pyrimethamine [46]. Field studies, have not confirmed that short-term cotrimoxazole use contributes to SP resistance [47-49]. However, the long-term effects are unknown, and cotrimoxazole has been used as an antibiotic for decades (FDA approved in 1973 [50]).

Because SP was introduced in Kilifi in the early 1990s [7], it is surprising that this triple-mutant was already so prevalent in 1993-1995. On the other hand, the presence of only one triple-mutant in the samples from 1993 suggests that perhaps the triple-mutant is a relatively new arrival. To determine when the triple-mutant arrived in Kenya, we examined samples from 1987-88, well before any common use of SP (see Chapter 4).

Sequencing of *pfcrt* in these samples showed that in 1993-95 the chloroquine-resistant allele (CVIET) was quite common in Kilifi, Kenya. This result agrees with the high level of treatment failure of chloroquine by this time [14]. Indeed, it was the failure of chloroquine that led clinicians in Kilifi to switch to SP as the first-line treatment of malaria in 1992-1993 [7,16]. (Kenya as a country did not change to SP as its first line antimalarial treatment until 1998 [16].) As is readily apparent from **Table 3.7**, *dhfr* and *pfcrt* are in complete linkage equilibrium. This result is not surprising, as the genes are on different chromosomes. However, it has been proposed that the chloroquine-resistant allele of *pfcrt* and the triple-mutant allele of *dhfr* arrived in Africa in the same parasite genome [4]. While this may have been the case, the linkage between the two genes rapidly probably disappeared rapidly as the invading parasite strain recombined with the local population. (See Chapter 4 for more discussion of this hypothesis.)

Unlike *pfcrt*, *dhps* may be in linkage disequilibrium with *dhfr*, even though they too are on separate chromosomes. In our samples, we only found mutant *dhps* alleles in combination with mutant *dhfr* alleles; all parasites carrying a wild-type *dhfr* also carried a wild-type *dhps*. This may suggest that the *dhps* mutant alleles emerged more recently than the mutant *dhfr* alleles, and that they are only advantageous in

combination with a mutant *dhfr* allele. It may also be the case that all of the parasites carrying *dhfr* mutations, not only the triple-mutants, are immigrants from neighboring regions where SP resistance emerged earlier. The "89" allele at locus -0.1 is common in South Africa as well (see Chapter 5), so its presence suggests only an East African origin, not necessarily a Kenyan one. In this scenario, the SP-resistant immigrants carried two or three mutations in *dhfr* and two mutations in *dhps*. Through independent assortment with wild-type local parasites, a few parasites would have emerged that were wild-type for one gene yet carried mutations in the other. If carrying a mutant *dhps* and a wild-type *dhfr* incurred a fitness cost, or perhaps simply by chance, parasites with that combination would not persist in the population. Studying microsatellites surrounding *dhps* might shed light on this possibility. If the microsatellite haplotypes near *dhps* in the Kilifi samples match the haplotypes in South African samples, then it is likely that the resistance mutations in *dhps* share a common origin for both locations.

Because the samples from 1993-95 contain some samples collected after treatment with SP or LapDap®, the data may be skewed towards finding mutant *dhfr* genotypes. While this precludes making a definitive estimate of the prevalence of mutant *dhfr* genotypes in Kilifi during this time period, it presents some interesting findings. The data in **Table 3.3** indicate that the post-treatment samples were less likely to show multiple *dhfr* genotypes (i.e. to be mixed). Most likely, this difference represents the elimination of non-mutant *dhfr* genotypes present in the initial infection in favor of a resistant, mutant genotype. For example, sample 546 was collected from a patient prior to treatment with SP, and appears to be a mixture of a triple-mutant and a double-mutant (51I/108N). Sample 601 was collected from the same patient within the four weeks following treatment with SP, and shows the triple-mutant genotype only. A reasonable inference is that the treatment with SP caused the triple-mutant parasite population within this patient to expand at the expense of the less-resistant double-mutant. We shall see more evidence of how rapidly the triple-mutant population can expand in Chapter 4.

In conclusion, by 1993 *P. falciparum* parasites carrying three mutations in *dhfr* were present in Kilifi, Kenya. Moreover, these triple-mutants share the same haplotype, and therefore likely the same ancestry, as triple-mutant parasites in Southeast Asia and South Africa. Because SP was first used shortly before these samples were collected, the high prevalence of the triple-mutant suggests that it was already in Kilifi prior to the use of SP. To investigate this possibility, we next examined samples collected in 1987-88.

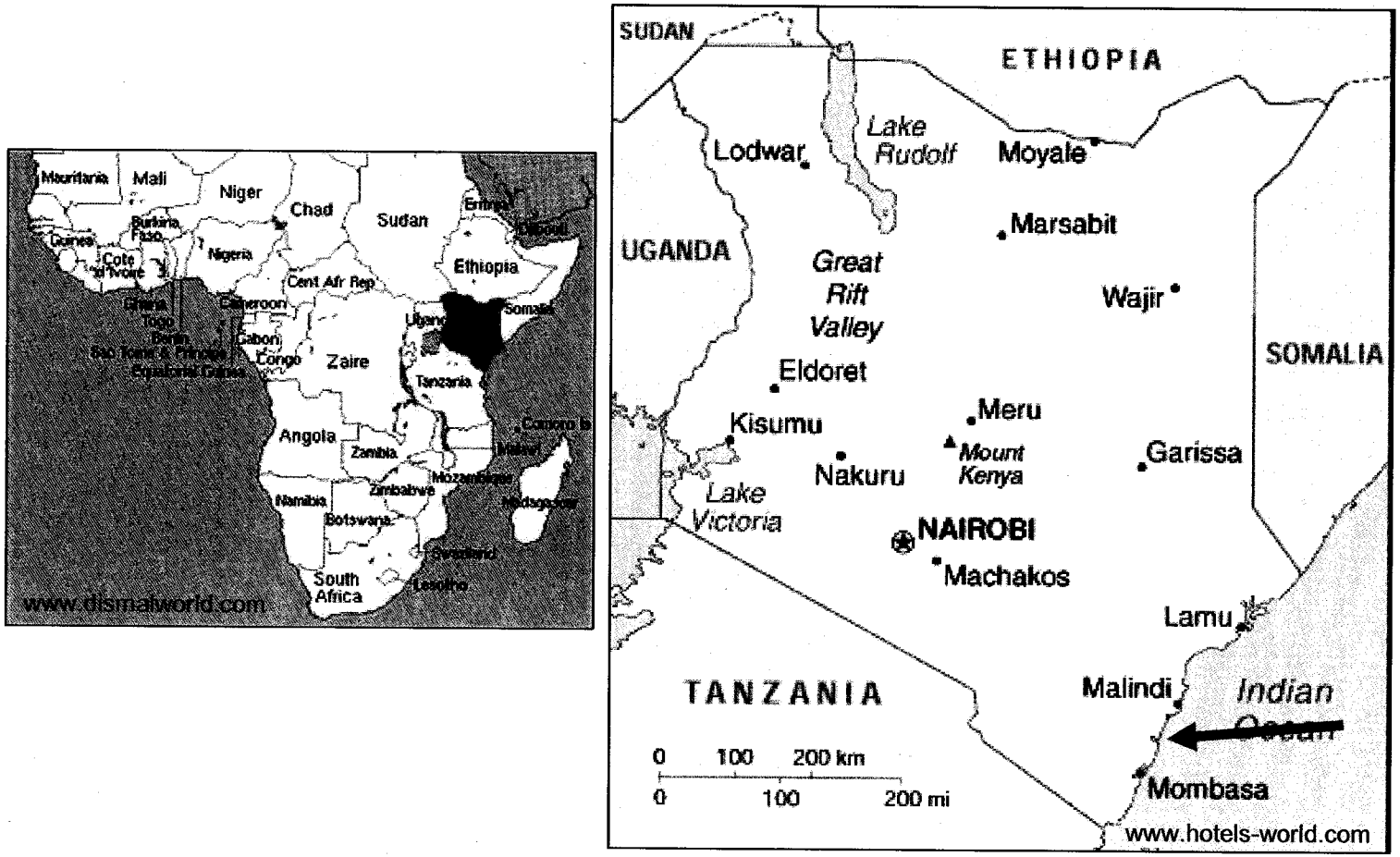


Figure 3.1 – Map of Kilifi, Kenya. The red arrow indicates the location of Kilifi town.

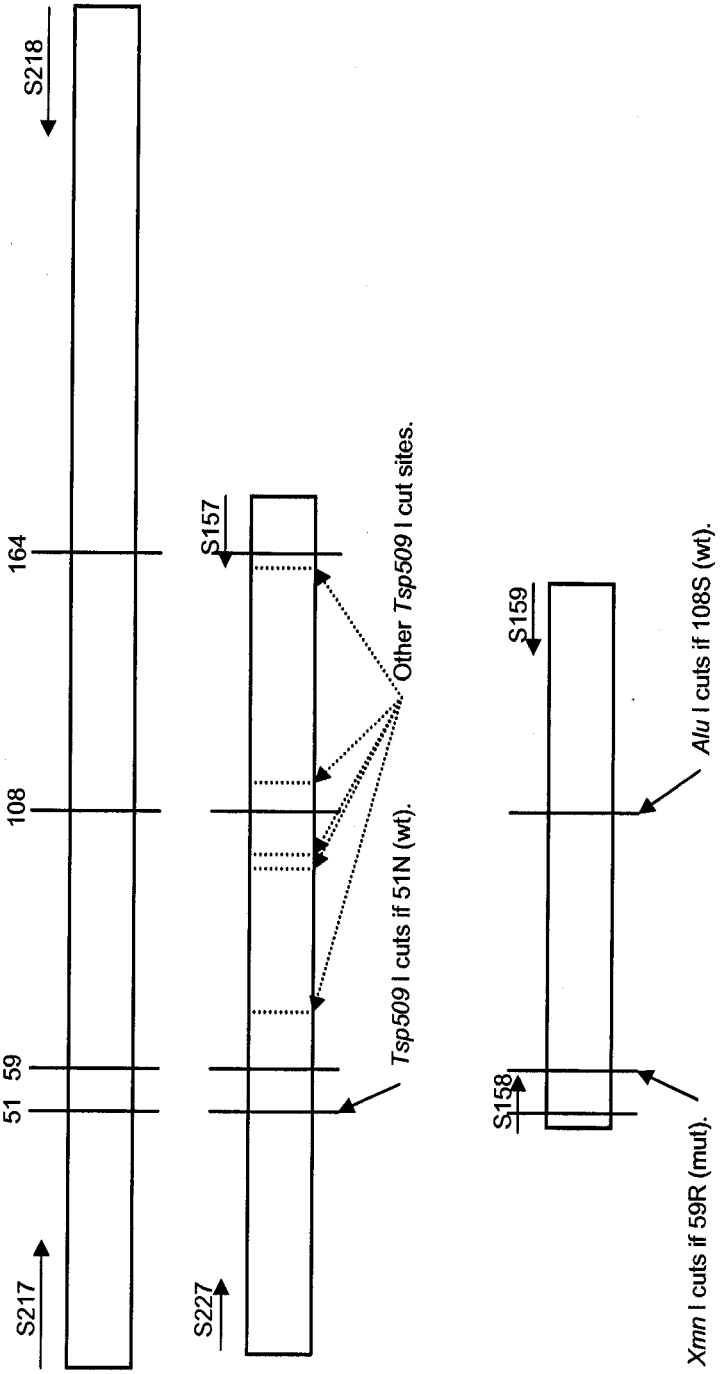


Figure 3.2 – *Dhfr* Genotyping Schematic. This figure shows the relative sizes of the PCR products for the *dhfr* genotyping analysis. Solid vertical lines indicate codons at which mutations confer pyrimethamine resistance.

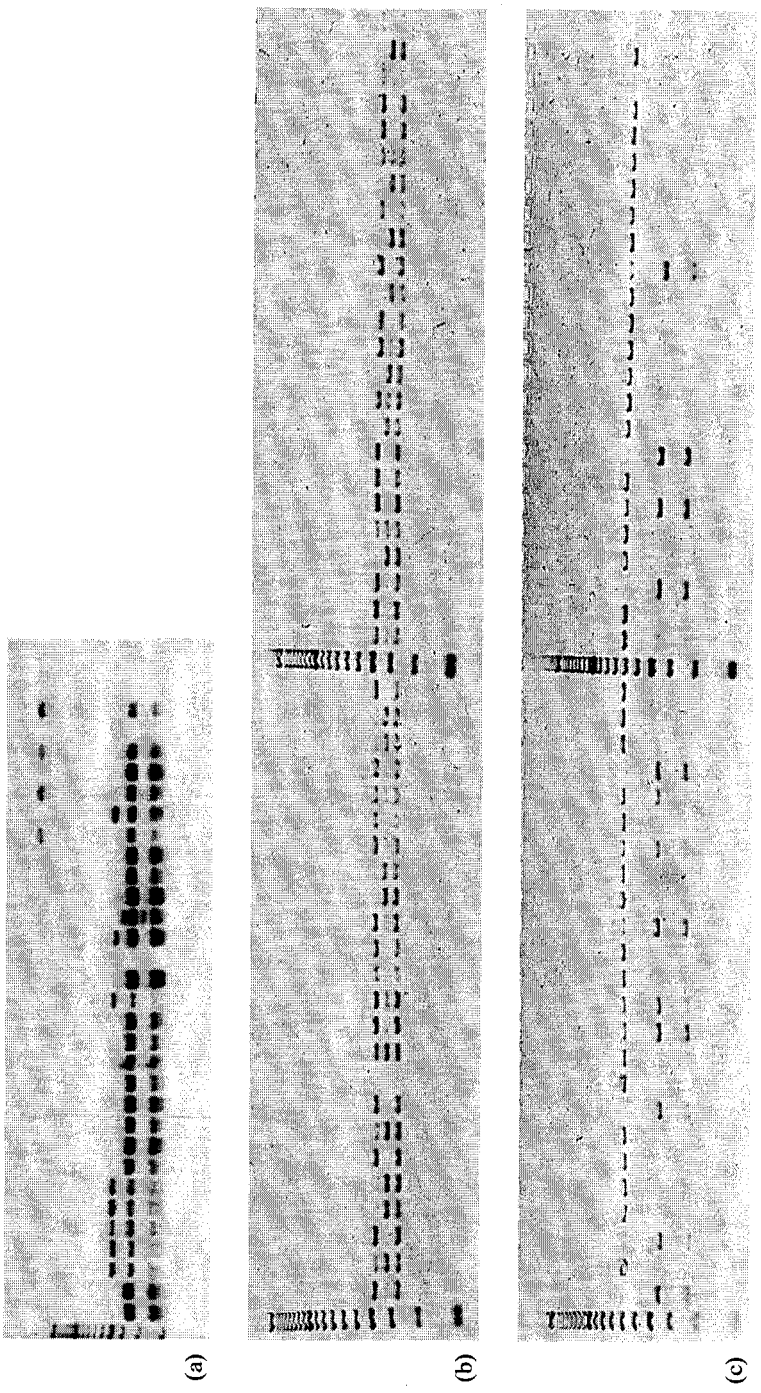


Figure 3.3 – Examples of Gels to Genotype *Dhfr.* (a) *Tsp509* I digest; (b) *Xmn* I digest; (c) *Alu* I digest.

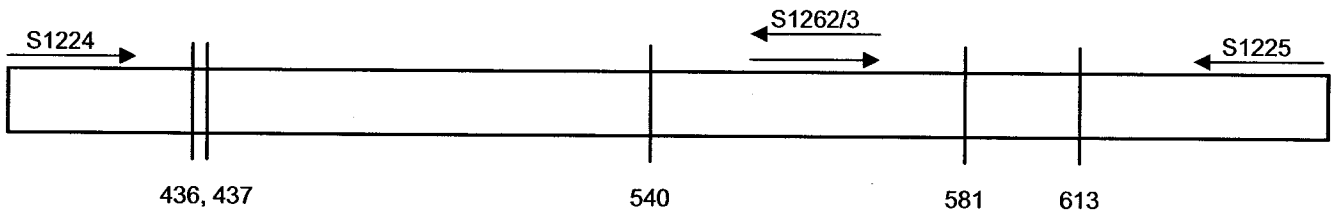


Figure 3.4 – Primers and Resistance Codons of *Dhps*. This figure shows the portion of *dhps* amplified by primers S1224 and S1225. Solid vertical lines indicate codons at which mutations confer sulfadoxine resistance.

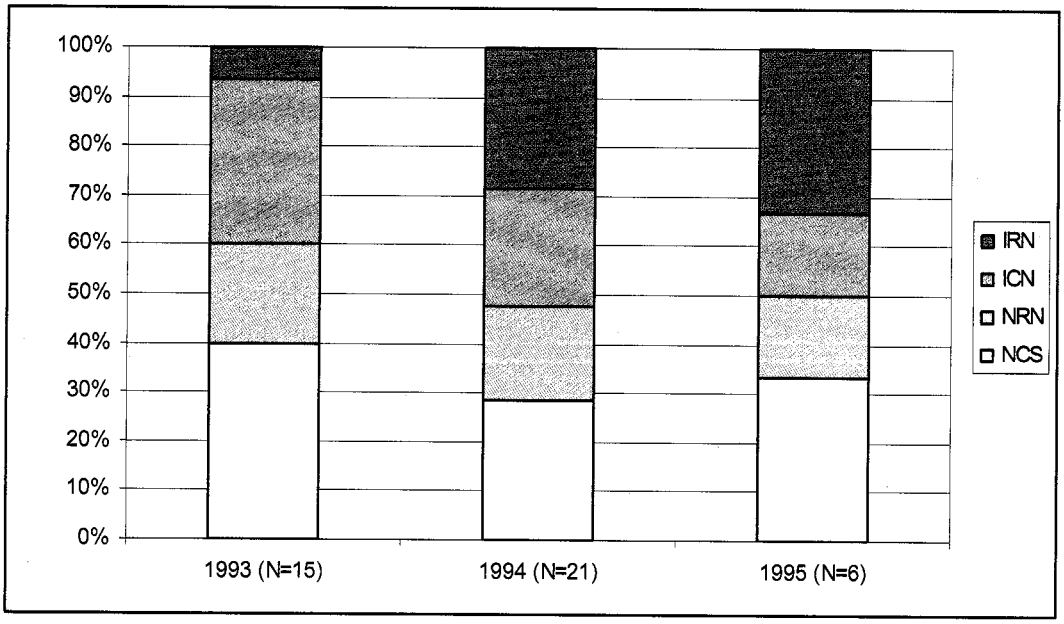


Figure 3.5 – Prevalence of *Dhfr* Genotypes by Year. The legend names genotypes by the amino acids carried at codons 51, 59, and 108.

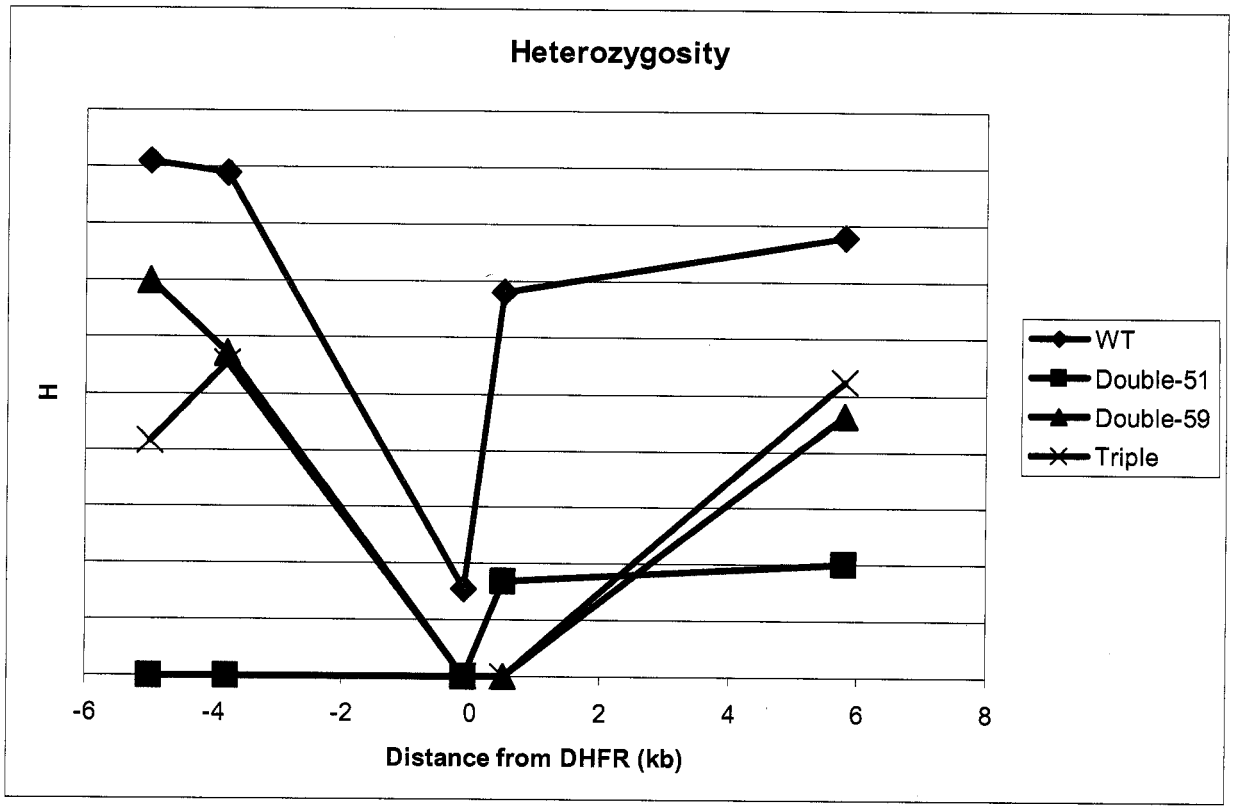


Figure 3.6 – Heterozygosity for the Different *Dhfr* Genotypes. Heterozygosity is the probability that two alleles drawn at random from the population will be different. Double-51 and Double-59 refer to parasites that have genotypes 51I/59C/108N and 51N/59R/108N, respectively.

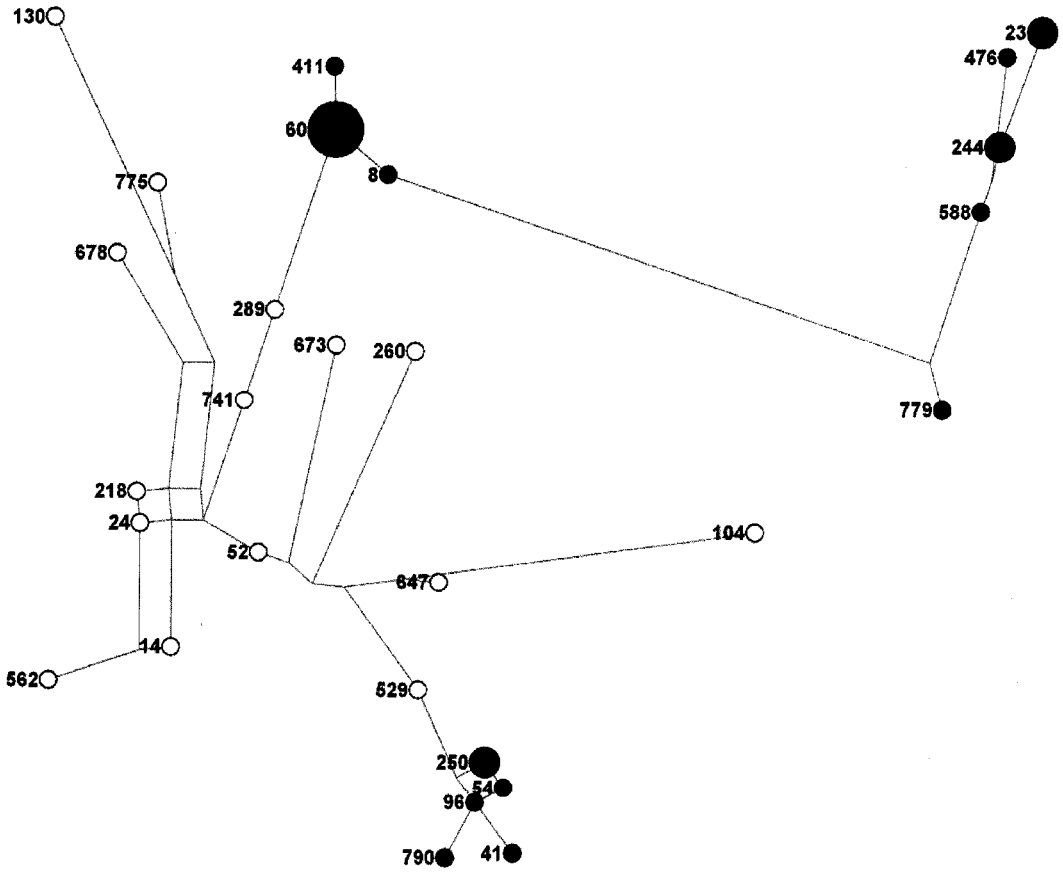


Figure 3.7 – Median-Joining Network of Clonal Samples at *Dhfr* and Flanking Microsatellites, 1993-95; SNP Weight = 90; ε = 0. Alleles are color coded as follows: purple = triple-mutant; blue = double-mutant (51/108); green = double-mutant (59/108); yellow = wild-type. Each node represents a haplotype, with the size of the node proportional to the number of samples with that haplotype. Haplotypes are numbered according to the first sample with that haplotype (e.g. the node labeled 60 represents samples 60, 83, 102, 113, 457, 632, 696, 713, and 802).

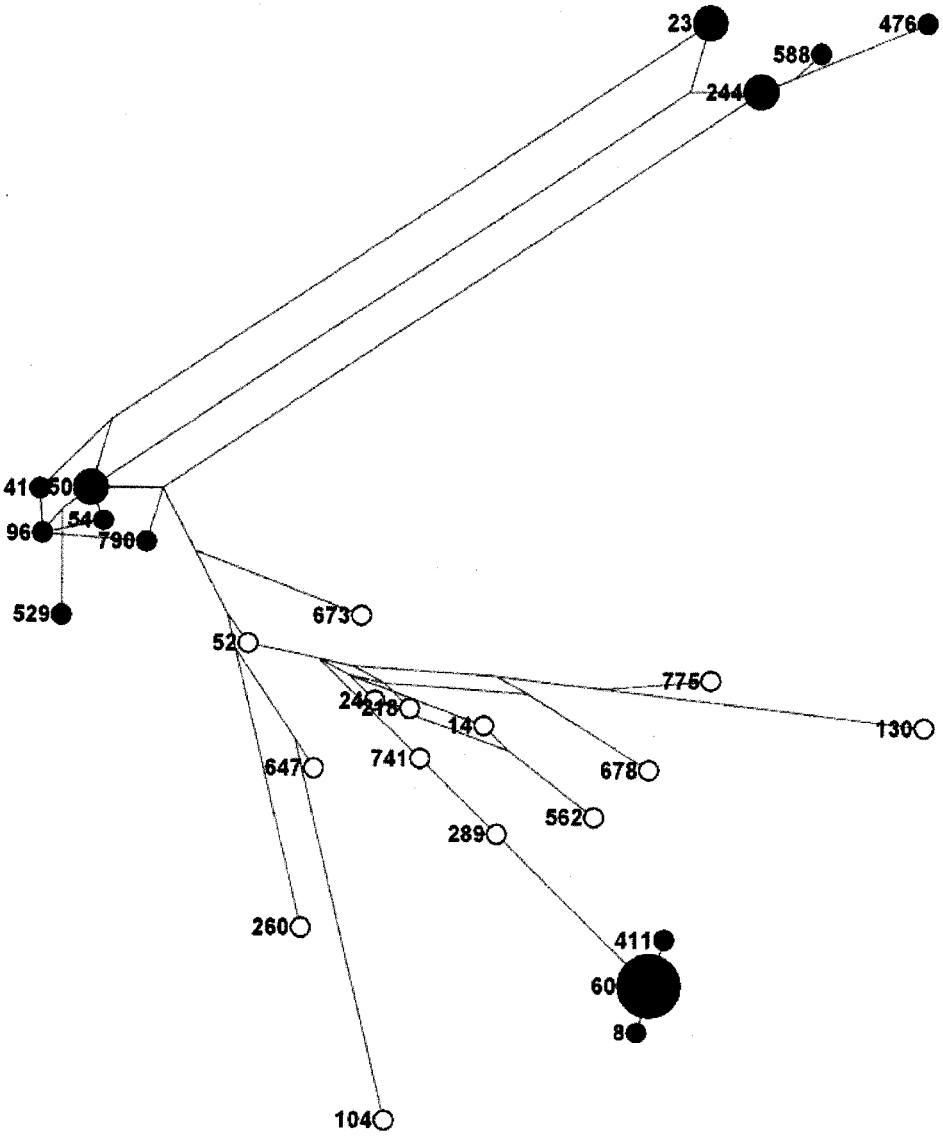


Figure 3.8 – Median-Joining Network of Clonal Samples at *Dhfr* and Flanking Microsatellites, 1993-95; SNP Weight = 90; ε = 0; No Sample 779. Alleles are color coded as follows: purple = triple-mutant; blue = double-mutant (51/108); green = double-mutant (59/108); yellow = wild-type. Each node represents a haplotype, with the size of the node proportional to the number of samples with that haplotype. Haplotypes are numbered according to the first sample with that haplotype (e.g. the node labeled 60 represents samples 60, 83, 102, 113, 457, 632, 696, 713, and 802).

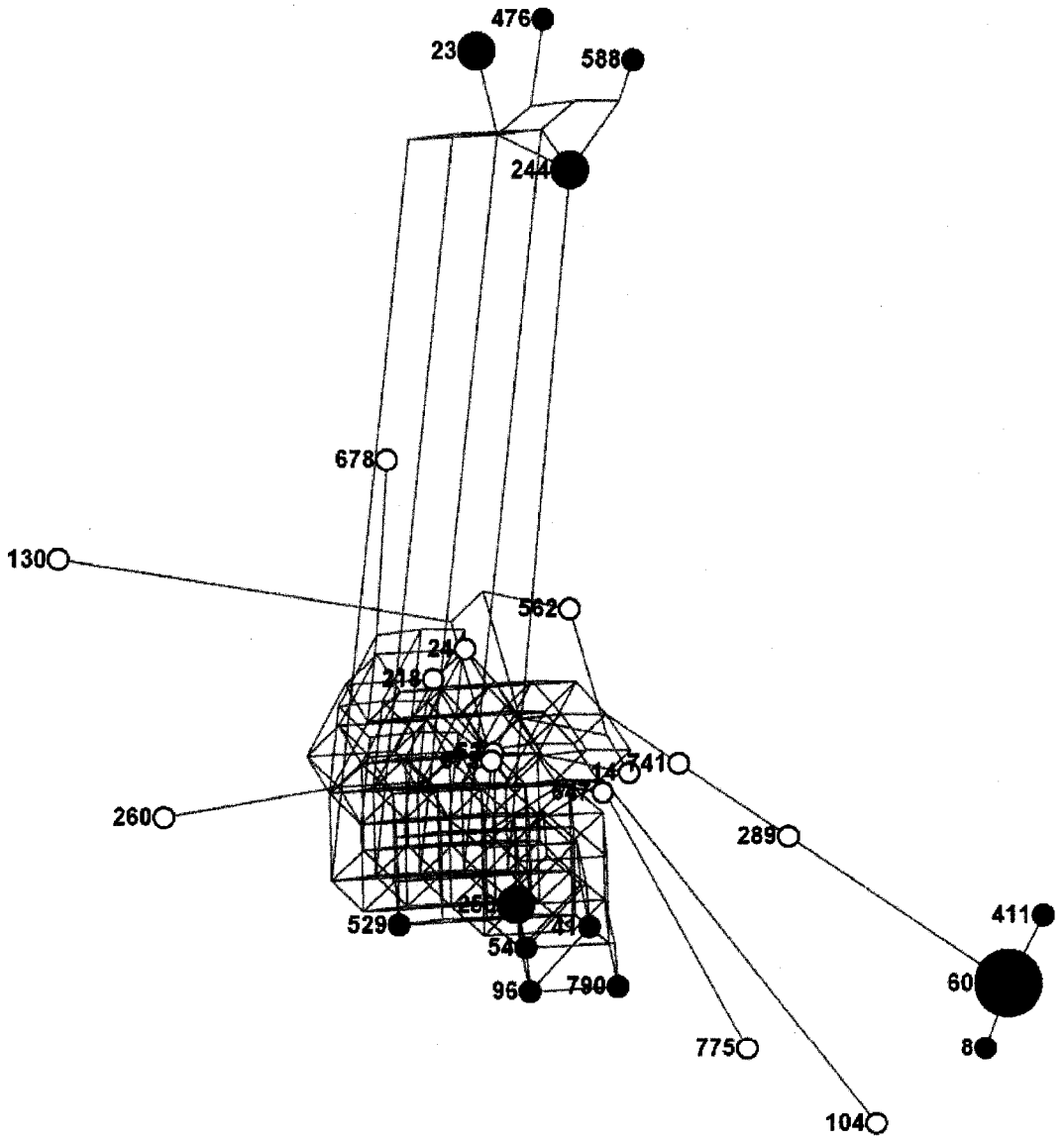


Figure 3.9 – Median-Joining Network of Clonal Samples at *Dhfr* and Flanking Microsatellites, 1993-95; SNP Weight = 90; ε = 11; No Sample 779. Alleles are color coded as follows: purple = triple-mutant; blue = double-mutant (51/108); green = double-mutant (59/108); yellow = wild-type. Each node represents a haplotype, with the size of the node proportional to the number of samples with that haplotype. Haplotypes are numbered according to the first sample with that haplotype (e.g. the node labeled 60 represents samples 60, 83, 102, 113, 457, 632, 696, 713, and 802).

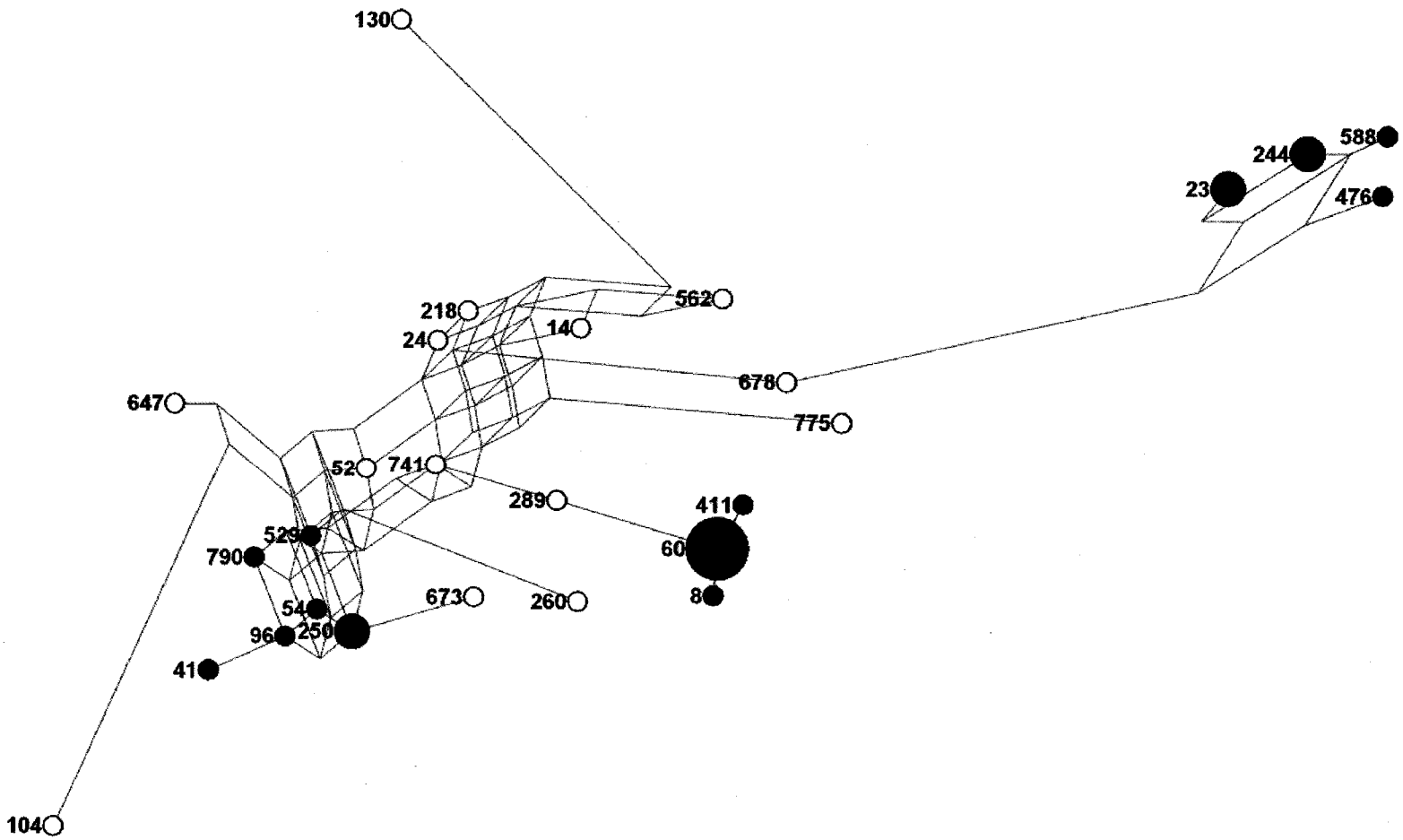


Figure 3.10 – Median-Joining Network of Clonal Samples at *Dhfr* and Flanking Microsatellites, 1993-95; SNP Weight = 0; ε = 11; No Sample 779. Alleles are color coded as follows: purple = triple-mutant; blue = double-mutant (51/108); green = double-mutant (59/108); yellow = wild-type. Each node represents a haplotype, with the size of the node proportional to the number of samples with that haplotype. Haplotypes are numbered according to the first sample with that haplotype (e.g. the node labeled 60 represents samples 60, 83, 102, 113, 457, 632, 696, 713, and 802).

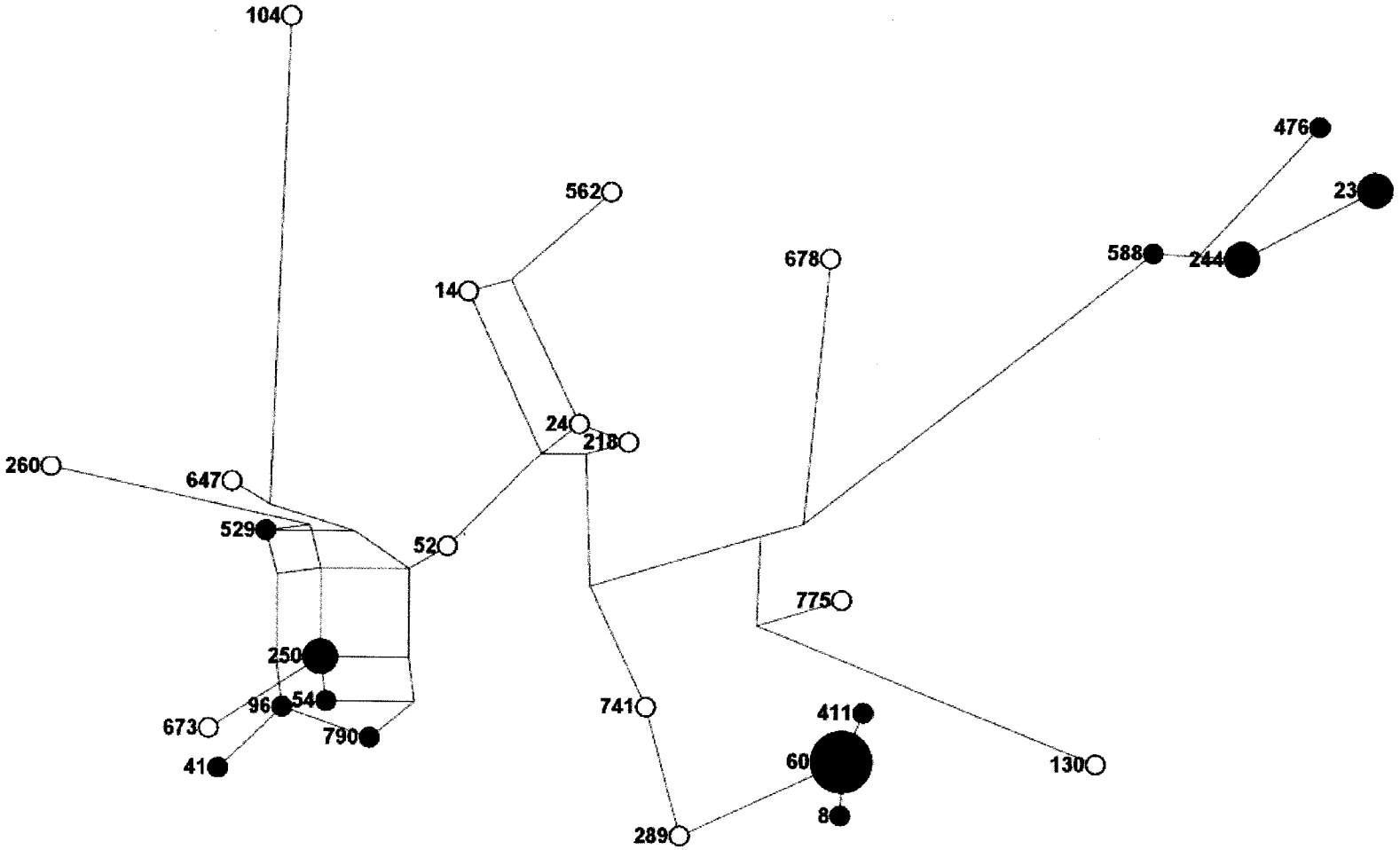


Figure 3.11 – Median-Joining Network of Clonal Samples at *Dhfr* and Flanking Microsatellites, 1993-95; SNP Weight = 0; ε = 0; No Sample 779. Alleles are color coded as follows: purple = triple-mutant; blue = double-mutant (51/108); green = double-mutant (59/108); yellow = wild-type. Each node represents a haplotype, with the size of the node proportional to the number of samples with that haplotype. Haplotypes are numbered according to the first sample with that haplotype (e.g. the node labeled 60 represents samples 60, 83, 102, 113, 457, 632, 696, 713, and 802).

Table 3.1 – Primers for Genotyping and Sequencing of *Dhfr*, *Dhps*, and *Pfcrt*.

	Forward primer	Reverse primer
DHFR		
Round 1 PCR	5'-ctccttttatgatggaacaagtctgcgacgttttcg (S217)	5'-tcatatgacatgtatctttgtcatcattctttaaaggc (S218)
Round 2 PCR		
For Tsp 509 I digest	5'-gacgttttcgatatttatgccata (S227)	5'-aaattcttgataaacaacggaacctttta (S157)
For Alu I and Xmn I digests	5'-gaaatgtaattccctagatatggaatatt (S158)	5'-ttaatttcccaagtaaaactattagagcttc (S159)
DHPS		
Larger fragment	5'-gattctttttcagatggagg (S1224)	5'-ttcctcatgtaattcatctga (S1225)
Smaller fragments	5'-gtattaaatggaatacctcgttatag (S1262)	5'-ctataacgaggtattccatttaatac (S1263)
PFCRT		
	5'-ggtggaggttcttgtcttgg (S1236)	5'-attaaagttgtgagtttcggatg (S1237)

Table 3.2 – *Dhfr* Genotypes and Prevalence of Mixed Infections, 1993-95.

DHFR Genotype	N	Mixed at ≥1 Microsatellite Locus	Used in Analysis
Wildtype	22	8	14
Single (S108N)	1	1	0
Double-51 (N51I/S108N)	13	2	11
Double-59 (C59R/S108N)	8	0	8
Triple (N51I/C59R/S108N)	12	3	9
Mixed	29	28	0
Total	85	42	42

Table 3.3 – *Dhfr* Genotypes Before and After Treatment with SP or LapDap®.

	SP Group		LapDap® Group		Combined Groups	
DHFR Genotype	Pre-Treatment	Post-Treatment	Pre-Treatment	Post-Treatment	Pre-Treatment	Post-Treatment
Mixed	7	0	10	0	17	0
Wild-type	3	0	1	1	4	1
ICN	3	2	1	4	4	6
NRN	2	1	1	1	3	2
Triple-mutant	3	4	1	2	4	6
Total	18	7	14	8	32	15

Table 3.4 – Haplotypes of Clonal Samples, 1993-95. Microsatellite alleles (PCR product lengths) and *dhfr* genotypes for each clonal sample at each locus. Parentheses indicate that the fluorescent peak on the capillary gel readout was weak, and therefore the allelic designation is provisional. Alleles are color-coded as follows: purple = triple-mutant; blue = double-mutant (51/108); green = double-mutant (59/108); yellow = wild-type; grey = 89 at locus -0.1. Blanks indicate missing data.

	Microsatellites			DHFR Genotype			Microsat's	
Sample	-5.0	-3.8	-0.1	N51I	C59R	S108N	0.5	5.8
23								
244								
476	213	194						
588		192						
629								(120)
698								
716								
779	195	192						(110)
800								
8	195	192	89	I	C	N	99	
60	195	192	89	I	C	N	97	118
83	195	192	89	I	C	N	97	118
102	195	192	89	I	C	N	97	118
113	195	192	89	I	C	N	97	118
411	195	192	89	I	C	N	97	122
457	195	192	89	I	C	N	97	(118)
632	195	192	89	I	C	N	97	118
696	195	192	89	I	C	N	97	118
713	195	192	89	I	C	N	97	118
802	195	192	89	I	C	N	97	
41	201	202	89	N	R	N	95	120
54	(201)	200	89	N	R	N	95	116
96	201	202	89	N	R	N	95	116
250		200	89	N	R	N	95	116
255		200	89	N	R	N	95	116
413	203	200	89	N	R	N	95	116
529	209	202	89	N	R	N	95	116
790	(200)	202	89	N	R	N	(95)	
14	209	192	89	N	C	S	95	106
24	211	194	89	N	C	S	95	
52	207	198	89	N	C	S	95	
104	207	206	107	N	C	S	95	110
130	217	190	89	N	C	S	112	110
218		194	89	N	C	S	97	
260	217	200	89	N	C	S	97	118
289	195	192	89	N	C	S	95	
562	217	192	89	N	C	S	95	106
647	209	208	89	N	C	S	95	
673	203	200		N	C	S	101	116
678	209	(194)	89	N	C	S		118
741	201	192	89	N	C	S	95	
775		186	89	N	C	S	(105)	110

Table 3.5 – Haplotypes of Non-Clonal Samples, 1993-95. Microsatellite alleles (PCR product lengths) and *dhfr* genotypes for each non-clonal sample at each locus. Same colors as Table 3.4.

	Microsatellites			DHFR			Microsatellites	
Sample	-5.0	-3.8	-0.1	51	59	108	0.5	5.8
76	(201, 205)	196, 202		I	C	N	95, 107	112, 116
224			101, 105		C	N	95, 107	
601								(114, 128)
65	201	192, 200		N/I	R	N	95	116
374				I	C/R	N	97, 107	
498	195, 205			I	C/R	N	95, 107	112, 116
546	(195, 205)	192, 196, 202		I	C/R	N	(95, 97, 107)	(112, 116)
686	195, 205	192, 196		I	C/R	N	97	112, 116
694	(195, 205, 219)	192, 196, 204		I	C/R	N	97, 107	112, 116
774	195, 205	192, 196		I	C/R	N	97, 107	118
280	205, 207	196, 198		I	C/R	S/N	93, 107	110, 112
77	195, 201, 203	192, 200, 202	89, 103	N/I	C/R	S/N	95, 101	116
414	(205, 207)	194		N/I	C/R	S/N	95, 107	112, 116
37	207	192, 196		N/I	C/R	N	95, 107	112, 122
182	201, 205	196, 202		N/I	C/R	N	95, 107	116
203	201	202		N/I	C/R	N	95	116
445	201, 205	196, 196, 202		N/I	C/R	N	95, 107	(101, 112, 116)
761	195, 201, 207	192, 202		N/I	C/R	N	95, 97	116, 118
262	195	192		I	C	N	97	(112), 118
585	195, 209	192		I	C	N	97	118
272	195, 211	192, 194	89, 112	I	C	S/N	97, 109	112, 118
28	(200, 205)	(192, 200)		N	C/R	N	95	116
523	195, 201	196, 196		N	C/R	N	95, 105	
53	207	192, 202		N	C	N	95	116
751	(205, 217)	190, 204	101, 105	N/I	C	N	95, 103	102
6		192, 202		N	C	S/N	95, 101	112, (116)
10	207	190, (198)		N	C	S/N	99	
64		200		N	C	S/N	93	110, 112, 116, 129
93	201, 221	200, 210		N	C	S/N	93	106, 116, 129
397	211	188, 198	89, 111	N	C	S/N	95	(100, 110)
657	209	192, 208		N/I	C	S/N	99	110, 118
328	201, 215	190, 200		N	C/R	S/N	95	108, 116
425	197, 209	194, 204		N	C/R	S/N	95, 97	110, 116
521	195, 201	192, 198, 202			C/R	S/N	95, 97, 107	116, 118
47	199, 207	194, 198, 202, 206		N	C	S	97, 99	110, 116
115	205, 207	(194, 212)		N	C	S	97, 109	110, 112
119	215, 217	194, 208	89, 101	N	C	S	95, 116	110
127	205, 207	190, 198	89, 97	N	C	S	97, 108, 112	114
338	209, 217	188, 192, 198		N	C	S	95	106, 112, 116
416	207, 211	204, 208		N	C	S	95	110
542	209	198, 200		N	C	S	(95, 97)	110, 112
544	205, 209	192, 196	89, 108	N	C	S	95, 112	116
350	217, 221	208		N		S	(97)	110

Table 3.6 – *Dhps* Genotypes for Samples Collected in Kilifi, Kenya, 1993-95.

	DHPS Genotype at Codons 437 and 540		
DHFR Genotype	Mixed	437A/540K (wt)	437G/540E
Wild-type	0	14	0
Double-mutant	0	13	5
Triple-mutant	1	5	2
Total	1	32	7

Table 3.7 – *Pfcrt* Genotypes for Samples Collected in Kilifi, Kenya, 1993-95.

	PfCRT Genotype at Codons 72-76		
DHFR Genotype	Mixed	CVMNK	CVIET
Mixed	13	2	13
Wild-type	5	4	13
Single-mutant	0	0	1
Double-mutant	4	4	12
Triple-mutant	4	1	6
Total	26	11	45

NOTES TO CHAPTER THREE

1. Cortese, J.F., Caraballo, A., Contreras, C.E. & Plowe, C.V. Origin and dissemination of Plasmodium falciparum drug-resistance mutations in South America. *J Infect Dis* **186**, 999-1006 (2002).

2. Nair, S. et al. A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. *Mol Biol Evol* **20**, 1526-36 (2003).

3. Roper, C. et al. Antifolate antimalarial resistance in southeast Africa: a population-based analysis. *Lancet* **361**, 1174-81 (2003).

4. Roper, C. et al. Intercontinental spread of pyrimethamine-resistant malaria. *Science* **305**, 1124 (2004).

5. McCollum, A.M. et al. Antifolate Resistance in Plasmodium falciparum: Multiple Origins and Identification of Novel dhfr Alleles. *J Infect Dis* **194**, 189-97 (2006).

6. Ndiaye, D. et al. Defining the origin of Plasmodium falciparum resistant dhfr isolates in Senegal. *Acta Trop* **99**, 106-11 (2006).

7. Mberu, E.K. et al. The changing in vitro susceptibility pattern to pyrimethamine/sulfadoxine in Plasmodium falciparum field isolates from Kilifi, Kenya. *Am J Trop Med Hyg* **62**, 396-401 (2000).

8. Williams, T. KEMRI CGMRC Kilifi, Kenya. in *INDEPTH Demographic Surveillance Sites* Vol. 2007 (INDEPTH Network, 2006).

9. Ndiritu, M. et al. Immunization coverage and risk factors for failure to immunize within the Expanded Programme on Immunization in Kenya after introduction of new Haemophilus influenzae type b and hepatitis b virus antigens. *BMC Public Health* **6**, 132 (2006).

10. Kyes, S. et al. Limited spatial clustering of individual Plasmodium falciparum alleles in field isolates from coastal Kenya. *Am J Trop Med Hyg* **57**, 205-15 (1997).

11. Mbogo, C.N. et al. Relationships between Plasmodium falciparum transmission by vector populations and the incidence of severe disease at nine sites on the Kenyan coast. *Am J Trop Med Hyg* **52**, 201-6 (1995).

12. KEMRI-Wellcome Trust Collaborative Research Programme. Vol. 2007 (Wellcome Trust Research Laboratories/KEMRI, 2002).

13. Spencer, H.C., Masaba, S.C. & Kiaraho, D. Sensitivity of Plasmodium falciparum isolates to chloroquine in Kisumu and Malindi, Kenya. *Am J Trop Med Hyg* **31**, 902-6 (1982).

14. Watkins, W.M. et al. Effectiveness of amodiaquine as treatment for chloroquine-resistant Plasmodium falciparum infections in Kenya. *Lancet* **1**, 357-9 (1984).

15. Hagos, B., Khan, B., Ofulla, A.V., Kariuki, D. & Martin, S.K. Response of falciparum malaria to chloroquine and three second line antimalarial drugs in a Kenyan coastal school age population. *East Afr Med J* **70**, 620-3 (1993).

16. Watkins, W. Personal communication. (2007).

17. Snow, R.W., Peshu, N., Forster, D., Mwenesi, H. & Marsh, K. The role of shops in the treatment and prevention of childhood malaria on the coast of Kenya. *Trans R Soc Trop Med Hyg* **86**, 237-9 (1992).

18. Spencer, H.C., Watkins, W.W., Sixsmith, D.G. & Koech, D.K. Response of Plasmodium falciparum to dihydrofolate reductase inhibitors in Malindi, Kenya. *Trans R Soc Trop Med Hyg* **80**, 201-3 (1986).

19. Watkins, W.M. & Mosobo, M. Treatment of Plasmodium falciparum malaria with pyrimethamine-sulfadoxine: selective pressure for resistance is a function of long elimination half-life. *Trans R Soc Trop Med Hyg* **87**, 75-8 (1993).

20. Amukoye, E. et al. Chlorproguanil-dapsone: effective treatment for uncomplicated falciparum malaria. *Antimicrob Agents Chemother* **41**, 2261-4 (1997).

21. Ferlan, J.T., Mookherjee, S., Okezie, I.N., Fulgence, L. & Sibley, C.H. Mutagenesis of dihydrofolate reductase from Plasmodium falciparum: analysis

in Saccharomyces cerevisiae of triple mutant alleles resistant to pyrimethamine or WR99210. *Mol Biochem Parasitol* **113**, 139-50 (2001).

22. Hankins, E.G., Warhurst, D.C. & Sibley, C.H. Novel alleles of the Plasmodium falciparum dhfr highly resistant to pyrimethamine and chlorcycloguanil, but not WR99210. *Mol Biochem Parasitol* **117**, 91-102 (2001).

23. Hunt, S.Y., Rezvani, B.B. & Sibley, C.H. Novel alleles of Plasmodium falciparum dhfr that confer resistance to chlorcycloguanil. *Mol Biochem Parasitol* **139**, 25-32 (2005).

24. Kublin, J.G. et al. Molecular markers for failure of sulfadoxine-pyrimethamine and chlorproguanil-dapsone treatment of Plasmodium falciparum malaria. *J Infect Dis* **185**, 380-8 (2002).

25. Mutabingwa, T. et al. Chlorproguanil-dapsone for treatment of drug-resistant falciparum malaria in Tanzania. *Lancet* **358**, 1218-23 (2001).

26. Nzila-Mounda, A. et al. Kenyan Plasmodium falciparum field isolates: correlation between pyrimethamine and chlorcycloguanil activity in vitro and point mutations in the dihydrofolate reductase domain. *Antimicrob Agents Chemother* **42**, 164-9 (1998).

27. Sakihama, N., Mitamura, T., Kaneko, A., Horii, T. & Tanabe, K. Long PCR amplification of Plasmodium falciparum DNA extracted from filter paper blots. *Exp Parasitol* **97**, 50-4 (2001).

28. Duraisingh, M.T., Curtis, J. & Warhurst, D.C. Plasmodium falciparum: detection of polymorphisms in the dihydrofolate reductase and dihydropteroate synthetase genes by PCR and restriction digestion. *Exp Parasitol* **89**, 1-8 (1998).

29. Wongsrichanalai, C., Pickard, A.L., Wernsdorfer, W.H. & Meshnick, S.R. Epidemiology of drug-resistant malaria. *Lancet Infect Dis* **2**, 209-18 (2002).

30. Sibley, C.H. et al. Pyrimethamine-sulfadoxine resistance in Plasmodium falciparum: what next? *Trends Parasitol* **17**, 582-8 (2001).

31. Brooks, D.R. et al. Sequence variation of the hydroxymethyldihydropterin pyrophosphokinase: dihydropteroate synthase gene in lines of the human malaria parasite, Plasmodium falciparum, with differing resistance to sulfadoxine. *Eur J Biochem* **224**, 397-405 (1994).

32. Wang, P., Read, M., Sims, P.F. & Hyde, J.E. Sulfadoxine resistance in the human malaria parasite Plasmodium falciparum is determined by mutations in dihydropteroate synthetase and an additional factor associated with folate utilization. *Mol Microbiol* **23**, 979-86 (1997).

33. Nzila, A.M. et al. 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**, 991-6 (2000).

34. Plowe, C.V. et al. Mutations in Plasmodium falciparum dihydrofolate reductase and dihydropteroate synthase and epidemiologic patterns of pyrimethamine-sulfadoxine use and resistance. *J Infect Dis* **176**, 1590-6 (1997).

35. Fidock, D.A. et al. Mutations in the P. falciparum digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol Cell* **6**, 861-71 (2000).

36. Alifrangis, M. et al. Occurrence of the Southeast Asian/South American SVMNT haplotype of the chloroquine-resistance transporter gene in Plasmodium falciparum in Tanzania. *J Infect Dis* **193**, 1738-41 (2006).

37. Ariey, F. et al. Invasion of Africa by a single pfcrt allele of South East Asian type. *Malar J* **5**, 34 (2006).

38. Pearce, R.J., Drakeley, C., Chandramohan, D., Mosha, F. & Roper, C. 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**, 1347-54 (2003).

39. Humphreys, G.S. et al. Amodiaquine and Artemether-Lumefantrine Select Distinct Alleles of the Plasmodium falciparum mdr1 Gene in Tanzanian Children Treated for Uncomplicated Malaria. *Antimicrob Agents Chemother* **51**, 991-7 (2007).

40. Bandelt, H.J., Forster, P. & Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**, 37-48 (1999).

41. Pearce, R. et al. Reduced variation around drug-resistant dhfr alleles in African Plasmodium falciparum. *Mol Biol Evol* **22**, 1834-44 (2005).

42. Cooper, G., Amos, W., Hoffman, D. & Rubinsztein, D.C. Network analysis of human Y microsatellite haplotypes. *Hum Mol Genet* **5**, 1759-66 (1996).

43. Bandelt, H.J., Forster, P., Sykes, B.C. & Richards, M.B. Mitochondrial portraits of human populations using median networks. *Genetics* **141**, 743-53 (1995).

44. Anderson, T.J., Su, X.Z., Roddam, A. & Day, K.P. Complex mutations in a high proportion of microsatellite loci from the protozoan parasite Plasmodium falciparum. *Mol Ecol* **9**, 1599-608 (2000).

45. Network 4.2.0.1 User Guide. (Fluxus Technology Ltd., 2007).

46. Iyer, J.K., Milhous, W.K., Cortese, J.F., Kublin, J.G. & Plowe, C.V. Plasmodium falciparum cross-resistance between trimethoprim and pyrimethamine. *Lancet* **358**, 1066-7 (2001).

47. Jelinek, T. et al. Plasmodium falciparum: selection of serine 108 of dihydrofolate reductase during treatment of uncomplicated malaria with co-trimoxazole in Ugandan children. *Am J Trop Med Hyg* **61**, 125-30 (1999).

48. Thera, M.A. et al. Impact of trimethoprim-sulfamethoxazole prophylaxis on falciparum malaria infection and disease. *J Infect Dis* **192**, 1823-9 (2005).

49. Malamba, S.S. et al. Effect of cotrimoxazole prophylaxis taken by human immunodeficiency virus (HIV)-infected persons on the selection of sulfadoxine-pyrimethamine-resistant malaria parasites among HIV-uninfected household members. *Am J Trop Med Hyg* **75**, 375-80 (2006).

50. answers.com. Trimethoprim. Vol. 2007 (Answers Corporation, 2007).

CHAPTER FOUR: DRUG RESISTANCE HAPLOTYPES OF SAMPLES COLLECTED IN KILIFI, KENYA, 1987-1988

The data presented in Chapter 3 show that by 1993 a strain of *P. falciparum* carrying three mutations in *dhfr* and a microsatellite haplotype identical to that of Asian SP-resistant parasites was present in Kilifi, Kenya. This "triple-mutant" parasite causes a ten-fold increase in the likelihood of SP treatment failure [1]. The relatively high prevalence of this strain in 1993-1995 was somewhat surprising, since SP was first used in Kilifi in 1992-1993 [2,3]. To try to identify when this resistant strain first appeared in Kilifi, we examined the earliest samples available, collected in 1987 and 1988.

As described in the previous chapter, in the late 1980's malaria research was just beginning in Kilifi. Therefore, the information about the prevalence of resistance to chloroquine and SP during these early years is limited. A study of pregnant women in Kilifi in 1988 showed ~50% clinical treatment failure of chloroquine [4], and a study in nearby Malindi several years earlier showed 28% clinical failure (19/69 children treated with chloroquine had persistent or recurrent parasitemia within 14 days) [5]. Despite its poor efficacy, chloroquine remained the recommended treatment for uncomplicated *P. falciparum* infection in Kilifi until ~1992, and in Kenya as a whole until 1998 [3]. While SP was not used routinely to treat malaria in the district hospital in the 1980's, SP and the related compound, chlorproguanil-dapsone (LapDap®), were used in a series of research studies. The samples used in the present analysis are from one such investigation [6].

In addition to the use of SP in research, sulfadoxine and pyrimethamine had both been available for decades. In fact, in 1953 pyrimethamine was tested for use as malaria prophylaxis [7]. For six months, residents of a village in Tanzania received pyrimethamine once a month, and at each visit a blood smear was examined for the presence of parasitemia. Prior to any pyrimethamine administration, 75% of people had parasites in their blood. After one month, the parasitemia rate had dropped to 35%, but after five months it was back up to 73%, and after six months to 80% [7].

These results indicated that pyrimethamine is ineffective for long-term prophylaxis. The assumption is that the exposure to pyrimethamine selected for parasites carrying one or more mutations in *dhfr*, rendering the drug ineffective for prophylaxis. Even a single change to 108N causes an increase in pyrimethamine resistance [8], and recall from Chapter 1 that *dhfr* resistance mutations occur at a rate of $\sim 10^{-10}$ *in vitro* [9]. In response to the outcome of the 1953 study, pyrimethamine alone was withdrawn and later introduced as a part of a fixed combination with sulfadoxine, the SP that has been used ever since.

As detailed in the introduction, using a combination of drugs is predicted to delay selection for resistance to either component, but only if both drugs are effective when used alone and there is no cross-resistance between them [10]. Sulfadoxine and pyrimethamine target the same pathway, and sulfadoxine alone is not effective against malaria; rather, it enhances the effectiveness of pyrimethamine [11]. Therefore, it is possible that even using the SP combination would rapidly select for mutations that confer resistance and SP would lose effectiveness within a short time. In addition, it is important to distinguish between the initial selection of a mutant allele and the increase in the proportion of that mutant allele in the parasite population once it gains a foothold, seen as a selective sweep. That is, the 1953 study of pyrimethamine may represent the rapid initial selection of new mutations in *dhfr*, or it may show the result of a rapid expansion of a resistance allele already present in the population at low levels, one that had undergone its initial selection elsewhere. One way to distinguish between the initial selection and the expansion of a resistance allele is to examine haplotypes. As described in Chapter 1, a shared haplotype indicates a shared ancestry. The number of different haplotypes carrying the resistance allele indicates the number of times selection for that allele occurred; the increase in prevalence of a particular haplotype indicates the rate of expansion of that allele.

Two more recent studies from Mali [12] and Senegal [13] indicate that what we witness is the rapid local expansion of resistance alleles already present in the population. In 1996, pyrimethamine was again tested as malaria prophylaxis, this time

in Mali [12]. In contrast to the 1953 study, this study included genetic analysis of the samples, allowing researchers to monitor the effect of pyrimethamine prophylaxis on the prevalence of *dhfr* mutations. The study was conducted in a rural village in Mali, where malaria is holoendemic with seasonal peaks. Chloroquine and quinine were the standard treatments for malaria in this village; SP was not available. For the study, children ages 2-12 received pyrimethamine weekly for 5 weeks during the 1996 rainy season (June-October). Blood samples were taken at enrollment and every week thereafter, as well as at the time of any "breakthrough" infections ("new infection after initial absence or clearance of parasitemia"). *Dhfr* genotypes were determined for all samples. Prior to the first dose of pyrimethamine, the prevalence of the 108N, 51I and 59R mutations were 12.8%, 4.2%, and 10.9%, respectively. Among the breakthrough infections, 100% were 108N, 50% 51I, and 90% 59R, illustrating the rapid expansion of mutant *dhfr* alleles in the presence of pyrimethamine. Of eight samples that failed to clear within the first week, all were wild-type for *dhfr* at enrollment but double- or triple-mutants (51I/108N, 59R/108N, or 51I/59R/108N) at day seven. Microsatellite analysis indicated that the mutant strains detected at day 7 were not the same as the wild-type strains at day 0, suggesting that the mutant strains were not *de novo* derivatives of the wild-type strain (which would have the same microsatellite alleles as the wild-type) but were simply below the level of detection of PCR at day 0.

A recent longitudinal survey from Senegal indicates that use of SP, not just pyrimethamine alone, likewise causes an increase in prevalence of mutant *dhfr* genotypes almost immediately after its introduction [13]. This study monitored the prevalence of mutant *dhfr* genotypes in the village of Dielmo from 1990-1999; SP was first used in 1995. Prior to 1995 mutant *dhfr* genotypes were present at less than 20%; triple-mutants were present, but only observed sporadically (1.1% of infections from 1990-1994). From 1995 onwards the prevalence of the triple-mutant was at least 15%, increasing to 32% by the end of the study in 1999 [13]. The microsatellite haplotype of this triple-mutant appeared identical to the Asian haplotype, indicating that these triple-mutants are related to those seen previously in South Africa, Kenya, and Asia.

The interpretation is that this triple-mutant was originally selected in Asia, spread widely around the world at low levels, and undergoes a rapid expansion in any location as soon as drug pressure is applied.

Because SP use appears to cause rapid expansion of resistance alleles, it is possible that the research conducted in Kilifi in the late 1980's was enough to reveal underlying drug resistant genotypes present in the parasite population. With this in mind, we examined samples from a SP resistance trial conducted in Kilifi in 1987-88.

Methods

Samples

In 1987-88, Watkins and Mosobo (1993) studied *in vitro* resistance to sulfadoxine and pyrimethamine [6]. Children with uncomplicated malaria were recruited to the study in July 1987 and followed every two weeks until December 1988. If parasites were detected at follow-up visits, blood was taken for *in vitro* chemosensitivity assays and the child was retreated with SP. Children were treated with SP throughout the study, though chloroquine was the standard treatment in Kilifi at this time for other children treated for uncomplicated malaria in the outpatient clinic. The study examined whether parasites isolated when blood concentrations of the drugs were below the effective therapeutic range were more likely to show resistance to sulfadoxine and/or pyrimethamine *in vitro*.

For the *in vitro* studies, parasites from infected children were cultured in the presence of various concentrations of sulfadoxine or pyrimethamine for 48 hours and then spotted onto glass slides for quantification by microscope. Some of these slides were stained and viewed; others were left unstained. These slides were stored at room temperature in Nairobi until 1997, when they were brought to Seattle. In 2005, the slides were inundated by flooding from the floor above. After this, they were placed at -20 °C for several months to prevent the formation of mold. They were then thawed, allowed to dry, and stored once again at room temperature. Informed consent for the original study was assured by the Kenya Medical Research Institute, Nairobi and the

present study of anonymous samples is in the exempt category of the University of Washington Human Subjects Review Board. Slides were labeled with the following information: drug (SD=sulfadoxine, PYR=pyrimethamine); identification number of study participant; date.

Because those samples that were cultured in pyrimethamine might be biased towards showing a greater prevalence of mutant *dhfr* genotypes, our primary outcome of interest, only samples cultured in sulfadoxine were used for this analysis. Each slide had eight spots, each from culture in a different concentration of sulfadoxine, from 0 to 161 μM. For most of the slides DNA was extracted from eight spots; however, if the blood spots were particularly dark (i.e. full of hemoglobin) only four or six were used to avoid overwhelming the purification kit. To extract DNA I used the protocol for forensic case samples in the QIAamp DNA Micro Kit (Qiagen, Valencia, CA). To get the dried blood off each glass slide, I mixed 20 μl Proteinase K with 300 μl Buffer ATL, drew that mixture up into a pipette tip, and then slowly expelled the liquid onto the slide while scraping the spots with the pipette tip. In this manner, I washed the slides into a 1.5 ml tube; I then followed the printed protocol. DNA was eluted twice from the column using 40 μl of water each time, incubating at room temperature for five minutes prior to the elution centrifugation. To minimize the risk of contamination, DNA extractions were performed in a room separate from the rest of the analysis.

Because the cohort was followed every two weeks in the original study, some of the samples used were from the same patient at different visits. If these visits were not separated by a significant amount of time, then the two samples may represent the same infection. Therefore, these samples cannot be considered completely independent.

Genotyping

The *dhfr* genotype of each sample was determined by DNA sequencing. Attempts to amplify the entire 700 bp *dhfr* gene using PCR were unsuccessful, presumably because the DNA from these twenty-year-old slides is of low quality

(i.e. fragmented). Therefore, I amplified the gene in three pieces, using two rounds of PCR for each, and sequenced each piece individually (**Figure 4.1**). Primers for both rounds of PCR and for sequencing were the same, and are given in **Table 4.1**. Both rounds of PCR amplification were done with a 20 µl reaction volume containing 1x PCR PreMix A and 0.5 units of the FailSafe enzyme mix (Epicentre, Madison, WI). The first round had 2 µl of the eluted DNA as template and 0.2 µM each primer; the second round had 1 µl of the first round product as template and 0.5 µM each primer. Cycling conditions for both rounds were as follows: 94 °C for 2 minutes; 30 cycles of 94 °C for 30 seconds, 55 °C for 45 seconds, 60 °C for 60 seconds; 60 °C for 2 minutes. Sequencing was done in a 10 µl volume, using 1 µl of the second round PCR product, 1.7 µl 5x sequencing buffer for Big Dye (400 mM Tris-HCl, 10 mM $MgCl_2$, pH = 9.0), 0.8 µM primer, and 1 µl Big Dye (Applied Biosystems, Foster City, CA). Cycling conditions and analysis of the sequencing reaction were the same as in Chapter 3. Though most of *dhfr* was sequenced, only codons 51, 59, 108, and 164 were specifically examined for mutations.

Each sample that gave a clean sequencing read for *dhfr* was then genotyped at six microsatellite loci, located 5.0, 4.1, 3.8, and 0.1 kb upstream of the start of *dhfr-ts*, and 0.5 and 5.8 kb downstream of the end of *dhfr-ts*. These loci were also used by Roper *et al.* (2003, 2004) and Nair *et al.* (2003), permitting a comparison between my results and their published work (see Chapter 5). For each microsatellite, I amplified the locus using semi-nested PCR and analyzed product size on an ABI 3100 (Applied Biosystems, Foster City, CA). Product size was recorded as the allele for that sample. If a sample showed multiple alleles at any locus it was considered mixed and all alleles were recorded. The primers for the -5.0 and -4.1 loci are given in Roper *et al.* (2003) and in **Table 4.1**; the primers for the remaining loci are given in Nair et al. (2003) and in **Table 4.1**. As in the previous chapter, the names of the microsatellites in our study are slightly different from those in the other studies due to differences in the method by which distances from *dhfr-ts* were calculated (e.g. from the first codon of *dhfr-ts vs.* from codon 108).

The first round of PCR amplification was done in a 25 µl reaction volume, containing 1x buffer for Taq, 3 mM $MgCl_2$, 0.2 mM dNTPs, 0.1 µM each primer, and 0.75 U Taq (Invitrogen, Carlsbad, CA). Cycling conditions were: 94 °C for 2 minutes; 5 cycles of 94 °C for 30 seconds, 50 °C for 30 seconds, 60 °C for 30 seconds; 60 °C for 2 minutes; 25 cycles of 94 °C for 30 seconds, 45 °C for 30 seconds, 60 °C for 30 seconds; 60 °C for 2 minutes [14]. Second round reactions were 10 µl, containing 1x buffer for Taq, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.4 µM forward (labeled) primer, 0.4 µM reverse primer, 0.3 U Taq polymerase (Invitrogen, Carlsbad, CA), and 1 µl first-round PCR product. Cycling conditions were: 94 °C for 2 minutes; 25 cycles of 94 °C for 30 seconds, 45 °C for 30 seconds, 60 °C for 30 seconds; 60 °C for 2 minutes [14].

For allele determination, 0.125-1.5 µl of the second-round PCR product was combined with 12.1 µl HiDi formamide and 0.5 µl GeneScan HD 400 ladder (Applied Biosystems, Foster City, CA), then run on an ABI 3100 capillary electrophoresis machine and analyzed with GenoTyper software (Applied Biosystems, Foster City, CA). Those samples that had more than one allele at any microsatellite locus were considered mixed. A sample with only one allele at every locus was considered clonal, though it is certainly possible that the sample contained multiple strains below our level of detection by PCR. The designation of allele sizes does not match exactly those reported in the previous studies nor in the previous chapter. In this case, each allele of the triple-mutant haplotype is detected as four bases shorter than it was when analyzed on the Amersham machine (see Chapter 3). For reference, DNA from the *P. falciparum* strain 3D7 gave an allele size of 101 when analyzed at the -0.1 locus on the ABI 3100.

For the subset of samples that I successfully genotyped for *dhfr,* I sequenced regions of *pfcrt* and *dhps* that contain known SNPs associated with resistance to chloroquine or to sulfadoxine (codons 72-76 of *pfcrt*; codons 437 and 540 of *dhps*). Each region was amplified using two rounds of PCR. Reaction conditions were the

same as for *dhfr*; primers are in **Table 4.1.** Cycling conditions for *pfcrt* were the same as for *dhfr*; for *dhps*, the annealing temperature was lowered to 50 °C.

Results

To investigate the presence of parasites carrying mutant *dhfr* alleles prior to the widespread use of SP, I examined samples collected in Kilifi, Kenya, between July 1987 and November 1988 (i.e. nearly 20 years ago). I extracted DNA from 57 slides, 21 of which were stained. Early results indicated that samples extracted from the stained slides failed to amplify, and therefore these samples were discarded from further analysis. Of the 36 samples from unstained slides, 13 were wild-type at all four *dhfr* codons, 7 were double-mutants (3 N51I/S108N; 4 C59R/S108N), 2 were triple-mutants (N51I/C59R/S108N; one was missing sequence at 108, but was inferred to be a triple-mutant), 6 carried more than one allele at *dhfr*, and 8 failed to amplify for one or more fragments (**Table 4.2**). Of the 22 samples with a single, complete *dhfr* genotype, 9 had multiple alleles at one or more microsatellite loci (5 wild-type, 3 double, 1 triple), leaving 13 samples with unambiguous haplotypes (**Table 4.3**). The data for all samples, including mixed samples, are shown in **Table 4.4**. Clearly, mutant *dhfr* genotypes are far more prevalent in the 1988 samples than in the 1987 samples. Even with the small sample size of the clonal samples (n=13) the increase in mutant *dhfr* genotypes from 1987 to 1988 was marginally significant (p=0.04, one-sided Fisher's exact test).

In 1993-95, 28% of parasites carrying mutant *dhfr* alleles also carried a mutant *dhps* allele (A437G/K540E; Chapter 3). Therefore, I determined the genotypes for three common mutants of *dhps* in the samples from 1987-88. The results of sequencing codons 436, 437 and 540 of *dhps* are shown in **Table 4.5**. Almost all of the samples were wild-type at both codons 437 and 540, the codons at which mutations were found in the samples from 1993-1995. However, three samples had a single nucleotide change from A to G at the second position of codon 540, causing a change from lysine to arginine (AAA → AGA). This change is different from that

normally seen at codon 540; sulfadoxine resistance is associated with a change from lysine to glutamate (AAA → GAA). In addition, five samples had changes in codon 436. One showed a change from serine to cysteine (TCT → TGT), three carried a change to alanine (TCT → GCT), and one was a mixture of serine and alanine (TCT and GCT).

The *dhfr* and *dhps* results indicate the prevalence of mutations conferring resistance to SP, but there had been no previous determination of the genotypes associated with resistance to chloroquine, the drug that was in routine use in Kilifi at the time. Therefore, I looked for the signature change from lysine to threonine at codon 76 of *pfcrt* and its associated SNP haplotype (a change from CVMNK to CVIET at codons 72-76). Because chloroquine was widely used in the 1980's, but SP was not, we would expect mutations in *pfcrt* to be much more common. All but one of the samples sequenced at *dhfr* were successfully sequenced at *pfcrt* (**Table 4.6**). Of those 27 samples, one was mixed, twelve were wild-type (CVMNK at codons 72-76), and fourteen were mutant (CVIET). Taking both years together, the proportion of parasites carrying chloroquine resistance alleles (14/26) was not significantly different from the proportion carrying mutations in *dhfr* (9/22; p=0.4, two-sided Fisher's exact test). However, in 1987 the proportion of parasites with mutant *pfcrt* (5/8) was significantly different from the lack of mutations in *dhfr* (0/8; p=0.03, two-sided Fisher's exact test). In 1988 the proportions were not significantly different (9/18 mutant at *pfcrt* vs. 9/13 mutant at *dhfr*; p=0.5, two-sided Fisher's exact test).

As mentioned in the previous chapter, one hypothesis about the presence in Africa of parasites carrying three mutations in *dhfr* and an Asian microsatellite haplotype is that the triple-mutant allele arrived in the same parasite as the chloroquine-resistance allele of *pfcrt* [15]. It has been shown that chloroquine-resistance arose in Asia and then that resistant strain spread to Africa [16]. Because resistance to both chloroquine and SP were widespread in Asia before resistance to either drug appeared in Africa [17], any Asian immigrant parasite with a mutant *pfcrt* would likely carry a mutant *dhfr* as well. Immediately after this immigrant arrived in Africa, before

independent assortment separated the two genes, we would see a higher prevalence of mutant *pfcrt* alleles in parasites with mutant *dhfr* alleles. In other words, if the parasites carrying mutations in *dhfr* in 1988 represent a recent immigration of drug resistant parasites from Asia, then we could find linkage disequilibrium between *dhfr* and *pfcrt*, even though they are on different chromosomes. However, the distribution of *dhfr* genotypes was the same for both *pfcrt* genotypes (p=1, two-sided Fisher's exact test), indicating no linkage disequilibrium.

As with the samples from 1993-95, I used the program Network to construct a median-joining network of the 13 unmixed samples from 1987-88. Networks indicate likely relationships between haplotypes, and can therefore lend support to assumptions about the relatedness of different haplotypes. The results of the analysis are shown in **Figures 4.2-3. Figure 4.2** shows the result of the analysis with the SNP weight equal to 90 and epsilon set at zero. **Figure 4.3** shows the effect of setting the SNP weight equal to zero, effectively removing *dhfr* genotypes from the analysis. The topology does not change significantly. In addition, increasing epsilon to a value greater than ten ($\varepsilon = 11$) has no effect (data not shown). Though the small number of samples makes this network less informative than those in Chapters 3 and 5, we can see that the triple-mutant appears only distantly related to the other samples.

Discussion

There are several points of interest in these data. First, the triple-mutant haplotype found in all prior studies was found again, though at a very low level. In fact, it was found only in a single patient. The patient labeled Kilifi 9 presented to the clinic a total of nine times between July 15, 1987, and November 25, 1988. The three samples from this patient in the present analysis are from October and November, 1988, and it is these three samples that contain the triple-mutant, either alone or combined with another haplotype. It is likely that the reason this patient was parasitemic so often, and therefore had so many blood samples taken, was because s/he was carrying an SP resistant parasite.

A second key point is that all of the samples from 1987 were wild-type (with the exception of one mixed infection), while half of those from 1988 were double- or triple-mutants. This difference is likely due to the fact that the samples from 1987 were collected at recruitment, before any SP administration. By the end of the study, however, these children had been treated with SP for 18 months. In other words, it seems the study itself is responsible for the appearance of so many resistant genotypes among the samples from 1988. One explanation is that the mutant *dhfr* genotypes were present in the population in 1987, but at a frequency so low they were not detected. Once the study participants received SP, however, selection pressure revealed the lurking mutant genotypes within this group of children. This model is also supported by the results of the pyrimethamine prophylaxis studies and longitudinal survey discussed in detail above, in which resistance appeared almost immediately after the drug was first used [7,12,13]. It appears that *any* use of SP will rapidly reveal resistant parasites in the population.

If so, did these studies – and others like them – contribute to the emergence of SP resistance in Kilifi, Kenya, and elsewhere in Africa? Perhaps the answer is irrelevant because the use of SP was inevitable once chloroquine failed, but the question raises an important issue for researchers studying drug resistance: by conducting their study they may be affecting the outcome. This rather unsettling possibility for malaria drug resistance research is discussed further in Chapter 6.

Regardless of any selection that may have occurred during the study, it is clear that the triple-mutant with the Asian haplotype was present in Kilifi in 1988. Therefore, it was present significantly before any widespread SP use. Is it possible to determine exactly when the triple-mutant arrived in Kilifi? Because resistant parasites may be present at low level, only detectable once drug pressure is applied, early samples may not give the answer unless collected during a time of drug pressure. One sample set that would be extremely interesting to genotype is that from the pyrimethamine prophylaxis study conducted in Tanzania in 1953 [7], or sets from similar studies conducted at about the same time [18,19]. As discussed above, we assume

that the rapid loss of clinical effectiveness of pyrimethamine was due to the selection of local parasites carrying one or more mutations in *dhfr*. In other words, we would predict to see local resistant alleles of *dhfr* gaining a foothold in this population. It seems unlikely that the loss of clinical effectiveness was due to the Asian triple-mutant, since SP was not widely used in Asia until 1967. Nonetheless, we have been surprised before, and it would be interesting to track any *dhfr* mutations present in this early study to see if their haplotypes continued to appear in later samples. For example, if the samples in the 1953 study carried two mutations in *dhfr*, do they share a microsatellite haplotype with either of the double-mutants seen in more recent samples (blue or green)? Or, did these resistance alleles disappear once the drug pressure was removed? Unfortunately, the samples from the 1953 study are no longer available, but perhaps other samples from similar studies can be found.

A third interesting finding was the presence of non-synonymous SNPs in codons 436 and 540 of *dhps*. The most common SNP at codon 436 changed serine to alanine in three samples; there was also one sample with a change to cysteine at codon 436. Both of these changes have been observed previously and associated with sulfadoxine resistance [20-22]. The SNP at codon 540 changed that amino acid from lysine to arginine in three different samples. While a change to glutamate at codon 540 is common, and associated with sulfadoxine resistance, to our knowledge this study represents the first observation of a change to arginine [23-25]. Because a change from lysine to arginine is conservative, it is unclear whether or not this change would have any effect on enzyme function or sulfadoxine resistance. The *in vitro* study of these samples revealed a few that were resistant to sulfadoxine [2,6], but these could be the samples with mutations at codon 436. It is of course theoretically possible that these changes arose and became the predominant allele *in vitro* during the 48-hour exposure to sulfadoxine. However, because 48-hours is only one cell cycle, and because the majority of the samples were wild-type at *dhps*, it seems unlikely that the culture had any effect on the genotypes detected.

Examining *pfcrt* in these samples indicated that roughly half of the samples carried the resistant genotype. This value corresponds well with the level of *in vivo* chloroquine resistance seen in pregnant women in Kilifi at this time [4]. Furthermore, the prevalence of the mutant genotype was significantly less than in 1993-1995 (14/26 in 1987-88 vs. 45/56 in 1993-95, p=0.01, one-sided Fisher's exact test; p=0.03, χ^2), supporting an increase in parasites that carried resistant genotypes with continued use of chloroquine. Looking only at samples collected at enrollment to the 1987-88 study, the prevalence of the chloroquine resistance allele was significantly greater than that of any SP-resistance allele. This result is expected, given that chloroquine was the standard treatment of malaria while SP had not been used broadly. However, by the end of the study SP-resistance alleles were just as common as chloroquine-resistance alleles, further underscoring the rapidity with which the prevalence of SP-resistance alleles grows.

There was no linkage disequilibrium between *pfcrt* and *dhfr*. Therefore, either the resistance mutations in *pfcrt* arrived (or arose) in Kilifi separately from the mutations in *dhfr*, or the parasite strain carrying both resistance mutations arrived long enough before 1987-88 that independent assortment has already completely broken down any linkage between the two genes. Because the only chance for independent assortment occurs in the mosquito phase of the malaria life cycle (see Chapter 1), it was thought that in areas of high transmission, where mixed infections are common, the parasite populations were largely outbred. However, recent studies of malaria oocysts in mosquitoes from western Kenya indicate that even in areas of high transmission, where mixed infections are common, the inbreeding coefficient is high, with significant linkage disequilibrium even between genetic markers on separate chromosomes [26]. Therefore, if resistance alleles to both drugs arrived in Africa in the same parasite genome, then they arrived many parasite generations prior to the present analysis.

By 1988 the triple-mutant with the Asian haplotype was present in Kilifi, Kenya. Because the samples from 1987 are almost exclusively wild-type, the selection

pressure in this case appears to be the drug resistance trial itself. The rapidity with which mutant genotypes were detected suggests that the mutant genotypes were present at low levels before the trial and needed only the selection pressure of SP administration to increase in prevalence. Earlier samples might confirm the presence of mutant genotypes prior to SP use, but they are scarce. An alternative approach is to use the genetic diversity of current samples to estimate when the parasite arrived. As a first step, the next chapter combines all data published on microsatellites near *dhfr* thus far to show the relationships between haplotypes found at various times and places in Africa.

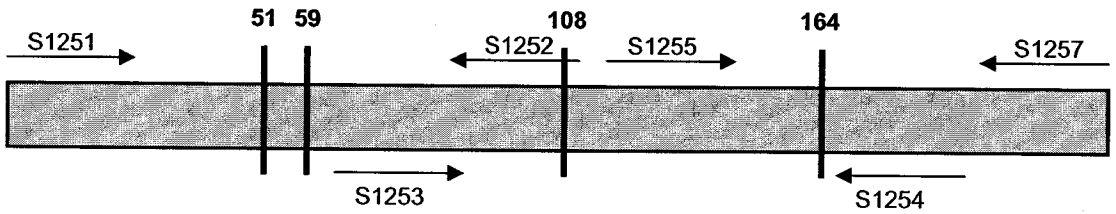


Figure 4.1 – Location of PCR Primers for Piecewise *Dhfr* Amplification. This figure shows the relative locations of the PCR primers for amplifying *dhfr* in three small pieces. The grey rectangle is *dhfr*; solid vertical lines indicate codons at which mutations confer pyrimethamine resistance.

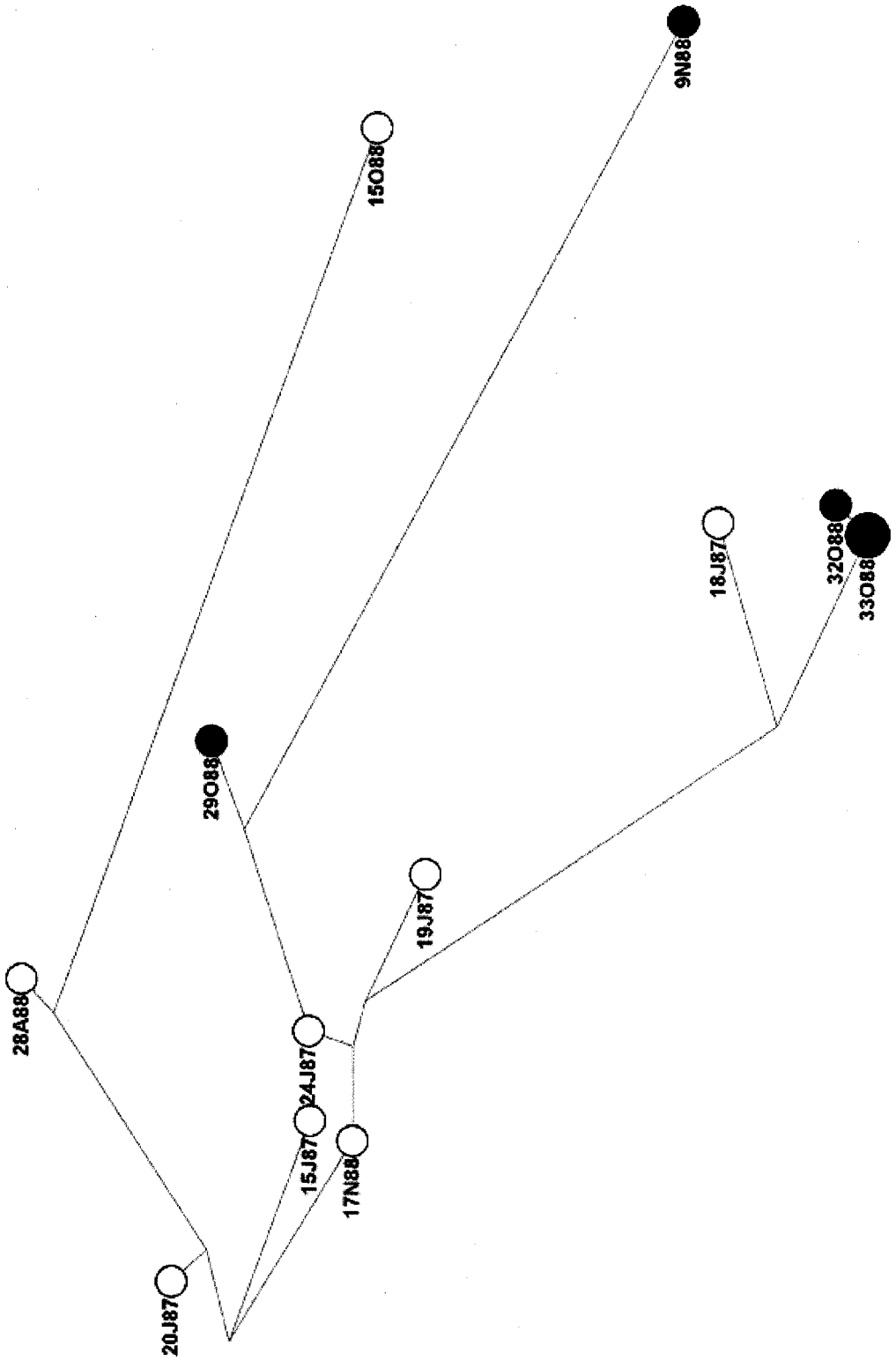


Figure 4.2 – Median-Joining Network of Clonal Samples at *Dhfr* and Flanking Microsatellites, 1987-88; SNP=90, ε=0.

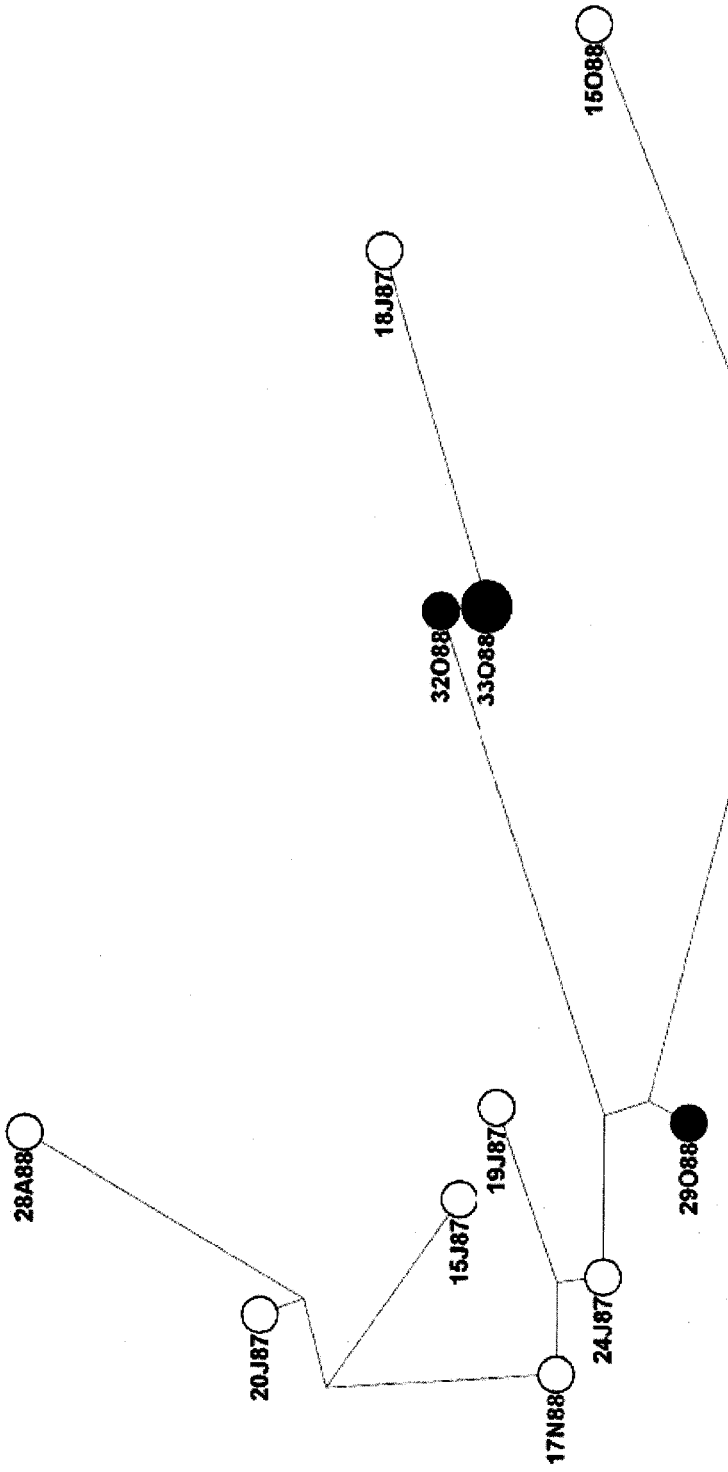


Figure 4.3 – Median-Joining Network of Clonal Samples at *Dhfr* and Flanking Microsatellites, 1987-88; SNP=0, ε=0.

Table 4.1 – Primers for PCR, Genotyping, and Sequencing of *Dhfr*, *Dhps*, and *Pfcrt*

Gene/Codon	Forward Primer		Reverse Primer	
Dhfr				
51 & 59	gccatatgtgcatgttgtaaggttg (S1251)		ctttcccagcttgttcttccca (S1252)	
108	aaatatttaaacaaagaaactgtgg (S1253)		cttgataaacaacggaacctcc (S1254)	
164	cctttaagcaataggataaatgtt (S1255)		actagtatatacatcgctaacaga (S1257)	
Dhps				
436 & 437	gattctttttcagatggagg (S1224)		cttataattggtttcgcatcac (S1264)	
540	gaatgtgttgataatgatttagttg (S1265)		ctataacgaggtattccatttaatac (S1263)	
Pfcrt	ggtggaggttcttgtcttgg (S1236)		attaaagttgtgagtttcggatg (S1237)	
Microsatellites	**Round One**	**Round Two**	**Round One**	**Round Two**
-5.0	gaacttgtttatacctgcat (S1004)	Fam-cctgcatttgcaagaagta (S632)	cacatattatacaggacga (S576)	Same as Round 1
-4.1	gtcaataatttctgcatcat (S1003)	Fam-taccatagcagtctttgca (S631)	cgatatatctgatgggtga (S573)	Same as Round 1
-3.8	gtaataaaatatgcacgctca (S1002)	Fam-acagttataagatttaatgcaa (S780)	actgatgaaattgtaaatga (S781)	Same as Round 1
-0.1	atttttgtattcccaaatag (S1001)	attccaacattttcaaga (S570)	ggcataaatatcgaaaac (S569)	Fam-tccatcataaaaaggaga (S650)
+0.5	attttacaatttcggattttac (S1005)	Fam-taaagaaggcataattttca (S1006)	cattgagataaataagtgttca (S719)	Same as Round 1
+5.8	tttaattaaaggacaatatc (S1007)	Fam-ttgtttttgttaagttgttt (S1009)	gaatatgacacaaattagtagg (S1008)	Same as Round 1

Table 4.2 – *Dhfr* Genotypes and Prevalence of Mixed Infections, 1987-88.

DHFR Genotype	N	Mixed at ≥ Microsatellite Locus	Shown in Table 4.3
Wildtype	13	5	8
Single (S108N)	0	0	0
Double-51 (N51I/S108N)	3	2	1
Double-59 (C59R/S108N)	4	1	3
Triple (N51I/C59R/S108N)	2	1	1
Mixed	6	4	0
Total	28	13	13

Table 4.3 – Haplotypes of Clonal Samples, 1987-88. Microsatellite alleles (PCR product lengths) and *dhfr* genotypes for each clonal sample at each locus. Parentheses indicate that the fluorescent peak on the capillary gel readout was weak, and therefore the allelic designation is provisional. An asterisk indicates that there was a tall stutter peak one base larger than the recorded allele; this one-base difference is ignored for the purposes of color coding. Alleles are color coded as follows: purple = triple-mutant; blue = double-mutant (51/108); green = double-mutant (59/108); yellow = wild-type; grey = 85 at locus -0.1. Alleles for locus -4.1 are not colored because that locus was not typed in the samples from 1993-1995.

Pt.	Date	-5.0	-4.1	-3.8	-0.1	N51I	C59R	S108N	0.5	5.8
15	15-Jul-87	215	167	192	85	N	C	S	89.5	109
18	16-Jul-87	191	185	196	85	N	C	S	92.5	109
19	16-Jul-87	211	---	186	85	N	C	S	89.5*	107
20	16-Jul-87	209	169	196	85	N	C	S	89.5*	105
24	16-Jul-87	205	173	190	85	N	C	S	---	107
9	11-Nov-88	201	173	194	85	I	R	N	102.5	116
15	5-Oct-88	207	174	208	109	N	C	S	94.5	112
17	16-Nov-88	203	169	190	85	N	C	S	91.5*	107
24	24-Nov-88	197	182	198	85	N	R	N	89.5*	112
28	6-Aug-88	---	171	200	85	N	C	S	(96.5)	109
29	19-Oct-88	---	175	188	85	I	C	N	91.5*	114
32	14-Oct-88	197	180	198	85	N	R	N	---	112
33	29-Oct-88	197	182	198	85	N	R	N	(90.5)	112

Table 4.4 – Haplotypes of Non-Clonal Samples, 1987-88. Microsatellite alleles (PCR product lengths) and *dhfr* genotypes for each sample at each locus. Parentheses indicate that the fluorescent peak on the capillary gel readout was weak, and therefore the allelic designation is provisional. An asterisk indicates that there was a tall stutter peak one base larger than the recorded allele; this one-base difference is ignored for the purposes of color coding. Alleles are color coded as follows: purple = triple-mutant; blue = double-mutant (51/108); green = double-mutant (59/108); yellow = wild-type; grey = 85 at locus -0.1. Alleles for locus -4.1 are not colored because that locus was not typed in the samples from 1993-1995.

Pt.	Date	-5.0	-4.1	-3.8	-0.1	N51I	C59R	S108N	0.5	5.8
15	15-Jul-87	215	167	[illegible]	85	N	C	S	89.5	109
17	16-Jul-87	---	179	198	85	N	mix	mix	89.5	112, 116
18	16-Jul-87	191	185	196	85	N	C	S	92.5	109
19	16-Jul-87	[illegible]	---	186	85	N	C	S	89.5*	107
20	16-Jul-87	209	169	196	85	N	C	S	89.5*	105
21	16-Jul-87	203	---	202	85	N	C	S	89.5*	101, 107
22	16-Jul-87	209	172, 174	190	85	N	C	S	91.5*	112
23	16-Jul-87	205, 209	169, 177	(190), 196, 204	85	N	C	S	---	107
24	16-Jul-87	205	173	190	85	N	C	S	---	107
9	14-Oct-88	191, 197	173	188, 192, 198	[illegible]	I	C	[illegible]	90.5, 102.5	105, 112
9	11-Nov-88	[illegible]	173	[illegible]	[illegible]	I	C	N	[illegible]	116
9	25-Nov-88	197	173, 182	192, 198	85	mix	mix	mix	90.5	112
11	24-Nov-88	191, 197, 217	160, 178	190, (196)	85, 93	N	C	S	---	103, 107, 112
15	5-Oct-88	207	174	208	109	N	C	S	94.5	112
17	16-Nov-88	203	169	190	85	N	C	S	91.5*	107
23	26-Oct-88	191, 209	179	188, 208	85	mix	mix	N	90.5	114
24	6-Oct-88	205	167, 173	196	85	N	C	S	(90.5)	105, 107
24	24-Nov-88	197	182	198	85	N	R	N	89.5*	112
28	6-Aug-88	---	171	200	85	N	C	S	(96.5)	109
29	19-Oct-88	---	175	188	85	I	C	N	91.5*	114
29	9-Nov-88	191	175	188	85	mix	C	N	92.5	114
31	8-Oct-88	191, 197	179	188, 198	85	I	C	N	92.5	112, 114
31	16-Nov-88	191, 209	179	(188), 208	85	N	R	N	90.5, 92.5	112, 114
32	14-Oct-88	197	180	198	85	N	R	N	---	112
33	29-Oct-88	197	182	198	85	N	R	N	(90.5)	112
39	9-Nov-88	203	158	194	85	N	mix	mix	---	118
41	12-Oct-88	191	179	188	85	I	C	N	92.5	109, 114
41	29-Oct-88	191	179	188	85	mix	C	mix	(92.5)	109, 114

Table 4.5 – *Dhps* Genotypes for Samples Collected in Kilifi, Kenya, 1987-88.

DHFR Genotype	Mixed	436S/437A/540K (wt)	436A/437A/540K	436S/437A/540R
Mixed	0	5	1	0
Wild-type	0	11	0	2
Single-mutant	0	0	0	0
Double-mutant	0	4	2*	1
Triple-mutant	1	0	1	0
Total	1	20	4	3

* One of these samples was 436**C**/437A/540K.

Table 4.6 – *Pfcrt* Genotypes for Samples Collected in Kilifi, Kenya, 1987-88.

	PfCRT Genotype at Codons 72-76		
DHFR Genotype	Mixed	CVMNK	CVIET
Mixed	0	3	3
Wild-type	0	5	7
Single-mutant	0	0	0
Double-mutant	0	3	4
Triple-mutant	1	1	0
Total	1	12	14

NOTES TO CHAPTER FOUR

1. Kublin, J.G. et al. Molecular markers for failure of sulfadoxine-pyrimethamine and chlorproguanil-dapsone treatment of Plasmodium falciparum malaria. *J Infect Dis* **185**, 380-8 (2002).

2. Mberu, E.K. et al. The changing in vitro susceptibility pattern to pyrimethamine/sulfadoxine in Plasmodium falciparum field isolates from Kilifi, Kenya. *Am J Trop Med Hyg* **62**, 396-401 (2000).

3. Watkins, W. Personal communication. (2007).

4. Rukaria, R.M., Ojwang, S.B., Oyieke, J.B. & Kigondu, C.B. In vivo and in vitro response of Plasmodium falciparum to chloroquine in pregnant women in Kilifi district, Kenya. *East Afr Med J* **69**, 306-10 (1992).

5. Watkins, W.M. et al. Effectiveness of amodiaquine as treatment for chloroquine-resistant Plasmodium falciparum infections in Kenya. *Lancet* **1**, 357-9 (1984).

6. Watkins, W.M. & Mosobo, M. Treatment of Plasmodium falciparum malaria with pyrimethamine-sulfadoxine: selective pressure for resistance is a function of long elimination half-life. *Trans R Soc Trop Med Hyg* **87**, 75-8 (1993).

7. Clyde, D.F. Observations on monthly pyrimethamine (daraprim) prophylaxis in an East African village. *East Afr Med J* **31**, 41-6 (1954).

8. Wu, Y., Kirkman, L.A. & Wellems, T.E. Transformation of Plasmodium falciparum malaria parasites by homologous integration of plasmids that confer resistance to pyrimethamine. *Proc Natl Acad Sci U S A* **93**, 1130-4 (1996).

9. Paget-McNicol, S. & Saul, A. Mutation rates in the dihydrofolate reductase gene of Plasmodium falciparum. *Parasitology* **122**, 497-505 (2001).

10. MMV. Selection of Partners for Antimalarial Drug Combinations. in *Strategic Consultation* (Medicines for Malaria Venture, Geneva, Switzerland, 2005).

11. Chulay, J.D., Watkins, W.M. & Sixsmith, D.G. Synergistic antimalarial activity of pyrimethamine and sulfadoxine against Plasmodium falciparum in vitro. *Am J Trop Med Hyg* **33**, 325-30 (1984).

12. Doumbo, O.K. et al. Rapid selection of Plasmodium falciparum dihydrofolate reductase mutants by pyrimethamine prophylaxis. *J Infect Dis* **182**, 993-6 (2000).

13. Noranate, N. et al. Rapid Dissemination of Plasmodium falciparum Drug Resistance Despite Strictly Controlled Antimalarial Use. *PLoS ONE* **2**, e139 (2007).

14. Nair, S. et al. A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. *Mol Biol Evol* **20**, 1526-36 (2003).

15. Roper, C. et al. Intercontinental spread of pyrimethamine-resistant malaria. *Science* **305**, 1124 (2004).

16. Ariey, F. et al. Invasion of Africa by a single pfcrt allele of South East Asian type. *Malar J* **5**, 34 (2006).

17. White, N.J. Antimalarial drug resistance: the pace quickens. *J Antimicrob Chemother* **30**, 571-85 (1992).

18. Clyde, D.F. & Shute, G.T. Resistance of East African varieties of Plasmodium falciparum to pyrimethamine. *Trans R Soc Trop Med Hyg* **48**, 495-500 (1954).

19. Jones, S.A. Resistance of P. falciparum and P. malariae to pyrimethamine (daraprim) following mass treatment with this drug; a preliminary note. *East Afr Med J* **31**, 47-9 (1954).

20. Pearce, R.J., Drakeley, C., Chandramohan, D., Mosha, F. & Roper, C. 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**, 1347-54 (2003).

21. Eberl, K.J. et al. Prevalence of polymorphisms in the dihydrofolate reductase and dihydropteroate synthetase genes of Plasmodium falciparum isolates from southern Mauritania. *Trop Med Int Health* **6**, 756-60 (2001).

22. Wang, P., Read, M., Sims, P.F. & Hyde, J.E. Sulfadoxine resistance in the human malaria parasite Plasmodium falciparum is determined by mutations in

dihydropteroate synthetase and an additional factor associated with folate utilization. *Mol Microbiol* **23**, 979-86 (1997).

23. Nyamwange, C.I. & Nyamogoba, H. Molecular markers in epidemiological monitoring of the spread of resistance to antimalarials: a review. *East Afr Med J* **79**, 480-4 (2002).

24. Plowe, C.V. et al. Mutations in Plasmodium falciparum dihydrofolate reductase and dihydropteroate synthase and epidemiologic patterns of pyrimethamine-sulfadoxine use and resistance. *J Infect Dis* **176**, 1590-6 (1997).

25. Wang, P. et al. 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**, 161-77 (1997).

26. Razakandrainibe, F.G. et al. "Clonal" population structure of the malaria agent Plasmodium falciparum in high-infection regions. *Proc Natl Acad Sci U S A* **102**, 17388-93 (2005).

CHAPTER FIVE: A COMPARISON OF THE DATA FROM KILIFI, KENYA, WITH DATA FROM OTHER LOCATIONS AND TIMES

From the data presented in the previous two chapters, it appears that a strain of *P. falciparum* carrying three mutations in *dhfr* ("triple-mutant") and the same microsatellite haplotype as triple-mutant parasites in Southeast Asia and South Africa was present in Kilifi, Kenya, before any widespread use of SP in Kilifi District. One explanation is that the triple-mutant arose in Southeast Asia, spread through Africa at low levels, and underwent rapid population expansion as soon as drug pressure was applied in a given location. Samples collected immediately after the proposed immigration of the Asian strain could support this hypothesis, but we do not know when that occurred. Even if the approximate date were known, it is presumably prior to 1987, and samples from before that time are scarce or nonexistent. Therefore, we must study samples collected from a variety of African locations at a variety of more recent times to try to reconstruct the evolution of SP resistance in Africa.

As described in Chapter 1, numerous groups have found limited genetic diversity among parasites with three mutations in *dhfr*, suggesting a single origin of that *dhfr* allele. The first studies on this subject showed that triple-mutant parasites from both Southeast Asia and South Africa had the same haplotype of microsatellites flanking *dhfr* [1-3], leading the authors to conclude that all triple-mutant parasites in Africa are descended from an invading Asian strain. This hypothesis is supported by my study of parasites from Kilifi, Kenya; all of the triple-mutants I found had that same Asian haplotype. However, studies from Senegal and Kisumu, Kenya, have found multiple microsatellite haplotypes among parasites carrying three mutations in *dhfr* [4,5]. Because the studies from Senegal and Kisumu only examined samples collected after SP resistance was well-established, it is unclear whether these alternative haplotypes represent recombination of the Asian triple-mutant with local (non-triple) strains, or are independent origins of the triple-mutant. Moreover, even if they represent independent origins, it is not possible to determine when they emerged – before or after the arrival of

the Asian triple-mutant. The goal of this chapter is to present all the *dhfr* haplotype data collected thus far in Africa as a first step towards reconstructing the evolution of SP resistance in Africa.

Methods

This chapter presents my data from Kilifi with data from four published papers: [2-5] (**Table 5.1**). For clarity, I have named each dataset according to location and years of samples collection. Therefore, the datasets are: Kilifi 1987-88 (Chapter 4), Kilifi 1993-95 (Chapter 3), Kisumu 2002-3 [4], South Africa 1996/9a [2], South Africa 1996/9b [3], and Senegal 2003-4 [5]. All of the papers created haplotypes from microsatellite loci near *dhfr* (**Figure 5.1**), and there is at least some overlap of loci among all of the studies (**Table 5.2**). The name of each locus reflects its distance, in kilobases, upstream of the start of *dhfr* (negative loci) or downstream of the end of *dhfr-ts* (positive loci). (DHFR is part of the bifunctional dihydrofolate reductase-thymidylate synthase protein, coded for by the gene *dhfr-ts*.) In addition to the studies mentioned above, the longitudinal study from Dielmo, Senegal, described in Chapter 4 determined microsatellite haplotypes for some of its samples [6]. However, because they only looked at two microsatellite loci the data are not included here. In addition, I did not include the Thai samples in Roper *et al.* (2004), as my focus is on Africa.

The basic characteristics of each study included are given in **Table 5.1**. A variety of transmission intensities are represented, which is of interest for two reasons. First, recall from Chapter 1 that the recombination rate of *P. falciparum* depends on the prevalence of mixed infections, which is higher in areas with intense transmission. Therefore, shared haplotype blocks will degrade more quickly in regions with a high level of transmission; the parasite population will be less inbred. Second, there is debate over the relationship between the level of transmission and the rate of emergence of resistance [7]. Some modeling studies predict that resistance will be higher in areas with low levels of transmission [8,9], while others predict the opposite [10-12]. Field studies indicate that the effect of transmission intensity on drug resistance may differ between

drugs [13]. Because malaria control programs will lower transmission, it is important to gather information on how drug resistance dynamics vary with transmission intensity.

There are several ways to measure level of transmission. First, one can determine whether transmission is continuous or seasonal. Second, one can use the Entomological Inoculation Rate (EIR), which is the number of infectious mosquito bites per person per year. Third, one can measure the prevalence of parasitemia among children (e.g. by collecting blood from a random sample of children and determining the presence of parasites) and classify an area as hypoendemic (≤10% prevalence), mesoendemic (11-50% prevalence), hyperendemic (50-75%) or holoendemic (>75%) [14]. Because different methods have been used to measure transmission intensity in the places discussed in this chapter, for simplicity I classified each site according to the third scheme based on the data available (**Table 5.1**). Kilifi, Kenya, has continuous transmission with marked seasonal variability. The mean EIR in Kilifi District is four, with most locations falling within the range 0-18 [15]. Kilifi District is classified as mesoendemic for *P. falciparum* [14]. Western Senegal is likewise mesoendemic, with a similar EIR to Kilifi [5]. KwaZulu Natal, South Africa, is hypoendemic for *P. falciparum* (≤10% prevalence), with seasonal transmission January to June [2]. Both Kisumu, Kenya, and Eastern Senegal show intense transmission (holo- or hyperendemic; prevalence >50%); the EIR in Eastern Senegal is over 100 [5].

The timing of SP use versus the time of sample collection varies somewhat between locations. As described previously, the Kilifi 1987-88 samples were collected prior to extensive use of SP, while the Kilifi 1993-95 samples were collected immediately after SP became the standard first line treatment at Kilifi District Hospital. The rest of Kenya changed to SP in 1998, so the Kisumu 2002-3 samples were collected well after SP use began. In KwaZulu-Natal, South Africa, chloroquine was replaced with SP in 1988. By 1996, there was 20% treatment failure with SP. By 2000 the failure rate was 70%, prompting a switch from SP to Coartem (artemether-lumefantrine) [16]. Therefore, the South Africa 1996/9 samples were collected as SP was nearing the end of its useful life. In Senegal, chloroquine remained the official first-line treatment until

2004, at which point the country switched to SP-amodiaquine [17]. The Senegal 2003-4 samples were therefore collected at roughly the time that chloroquine was replaced.

As a first step in synthesizing the data from these studies, I simply placed all data into tables and color-coded according to the scheme I used for my own data: wild-type *dhfr* is yellow; double-mutant 51I/108N is blue; double-mutant 59R/108N is green; triple-mutant (51I/59R/108N) is purple. I additionally designated single-mutants (108N) as orange and any *dhfr* allele containing the 164L mutation as pink. The mutation at codon 164 typically occurs only in the presence of the other mutations and confers extreme resistance to pyrimethamine [18,19]. For the microsatellite alleles, colors reflect the most common haplotype for that *dhfr* genotype in the Kenya 1993-95 dataset.

Next, I used Network to construct median-joining networks of the available data. Ideally, we would create a network that included all the samples. However, because there were only two microsatellite loci shared by all datasets, I chose to analyze subsets of the data and thereby include more loci. I analyzed the following subsets: "Kenya" (Kilifi 1987-88, Kilifi 1993-95, and Kisumu 2002-3); "East Africa" (Kilifi 1987-88, Kilifi 1993-95, and South Africa 1996/9b); and "All Africa" (Kilifi 1987-88, Kisumu 2002-3, South Africa 1996/9a & b, Senegal 2003-4).

As mentioned previously, the recorded size of a given microsatellite allele depends both on the PCR primers used and on the capillary electrophoresis system. In consequence, the specific allele sizes that make up a given haplotype vary between the different studies. For example, at locus -0.1 the canonical triple-mutant haplotype has allele 109 in the Kilifi 1993-95 and Senegal 2003-4 datasets, but allele 110 in South Africa 1996/9a, 108 in South Africa 1996/9b, and 105 in Kilifi 1987-88 and Kisumu 2002-3 (see Results). To compare the datasets I therefore had to standardize the allele sizes. I did so by declaring the canonical triple-mutant to have alleles 20, 40, 20, 24, 30, 20 at loci -5.0, -4.1, -3.8, -0.1, +0.5, +5.8, respectively. (These allele sizes were chosen so that after standardization all alleles would still be positive.) For each dataset, I then calculated what value would need to be subtracted from the measured allele of the triple-mutant at a given locus to produce the desired haplotype for the triple-mutant, and

subtracted that value from all samples at that locus. **Table 5.3** indicates the values used to standardize the alleles. For example, in the Kilifi 1993-95 dataset the triple-mutant haplotype is [205, 196, 109, 107, 112] at loci [-5.0, -3.8, -0.1, +0.5, +5.8]. Therefore, I subtracted [185, 176, 85, 77, 92] from every haplotype in the Kenya 1993-95 dataset.

Though reasonable, this standardization technique is not perfect. For example, the samples in the two papers presenting data from South Africa (1996/9a and b [2,3]) are mostly the same, yet after standardization they did not show identical haplotypes. I entered the data from these two papers separately in Network, and therefore these samples are over-represented in the network. When a node representing a sample from South Africa 1996/9a is separated from South Africa 1996/9b sample by a short distance it is likely that they are in fact a single sample, and the discrepancy in haplotypes is due to the imperfection of standardization. However, because Network uses a stepwise mutation model for microsatellites the effect of this imperfection will be minimal. An allele difference of small absolute value will appear as a small evolutionary distance in the network, so haplotypes that should be identical – but are not due to problems with the standardization – will be very close together in the network.

The size of a given node in the network is proportional to the number of times that haplotype appeared in the combined dataset. When entering data into the combined dataset, I entered it as it appeared in the original paper. For most of the data this meant that each sample was entered once. However, for the data from South Africa 1996/9a [2] it meant that each observed haplotype was entered once, regardless of how often it appeared. For example, the canonical triple-mutant was observed 43 times in that study; all 43 are represented as a single sample in the combined dataset. Therefore, the relative sizes of the nodes are approximations only. The benefit to this is that the node for the triple-mutant is not so large as to over shadow all others.

Results

To synthesize the data thus far on the evolution of SP resistance in Africa, I first constructed tables of all published data (**Tables 5.4-9**). Taken together, the tables clearly show that most triple-mutants share the same haplotype (purple). In addition, parasites from Kilifi, Kisumu, and South Africa with mutations at codons 51 and 108 of *dhfr* all share the same microsatellite haplotype (blue). In contrast, there does not appear to be a widespread haplotype shared by parasites with mutations at codons 59 and 108 (the green *dhfr* genotype). One interesting finding in the data from Senegal is the presence of many microsatellite alleles characteristic of the triple-mutant haplotype among the parasites that are wild-type at *dhfr*.

After constructing these tables, I used the program Network (described in Chapter 3) to create median-joining networks of various subsets of data (**Figures 5.2-7**). Subsets were chosen according to geography and shared loci. The parameter epsilon was set at zero for all analyses (see Chapter 3); SNP differences were given a weight of 90 or zero. Giving the SNP differences a weight of 90 reflects their presumed lower mutation rate compared to microsatellites, which had a weight of 10. Giving the SNP differences a weight of zero shows how the haplotypes would be grouped if *dhfr* genotype were ignored, thus avoiding any assumptions about the number of times each *dhfr* mutation arose. The labeling scheme for the different sample sets is given in **Table 5.10**. Each node is labeled with the name of the first sample with that haplotype in the dataset. A list of which samples are within each node is given in **Tables 5.11-14**.

First, I combined my data (Kilifi 1987-88 and 1993-95) with that from McCollum *et al.* (2006) (Kisumu 2002-3) to see the network for samples collected in Kenya. The results are shown in **Figures 5.2 and 5.3**. The first point of interest is that the wild-type parasites from Kilifi 1987-88 cluster among the wild-type parasites from Kilifi 1993-95, indicating that they are part of the same population. Second, the parasites carrying mutations at codons 51 and 108 of *dhfr* (the blue double-mutants) cluster together, whether from Kilifi or Kisumu, and this clustering remains intact even when the *dhfr* genotype is ignored (SNP weight = 0; **Figure 5.3**). The Kisumu 2003-4

sample with the genotype 51I/108N/164L likewise clusters with the 51I/108N double-mutants. As for the triple-mutants (51I/59R/108N), there is some variety of haplotypes among the Kisumu 2002-3 samples, but in general they cluster together.

The second subset analyzed was the "East Africa" subset, including the Kilifi 1987-88, Kilifi 1993-95, and South Africa 1996/9b datasets. Again, I analyzed this combined dataset twice, once with a SNP weight of 90 (**Figure 5.4**) and once with a SNP weight of zero (**Figure 5.5**). The two networks are very similar. The wild-type parasites from both locations (Kilifi, Kenya, and KwaZulu-Natal, South Africa) are intermixed, suggesting little if any population subdivision. The two types of double-mutants (51I/108N, blue and 59R/108N, green) cluster according to *dhfr* genotype (**Figure 5.4**). Notice, however, that the samples from South Africa 1996/9b are either wild-type or triple-mutant for *dhfr*; all of the double-mutants in the East Africa combined dataset are from Kilifi. When *dhfr* genotype is ignored (**Figure 5.5**), the clustering holds for the 51I/108N (blue) samples but is somewhat disrupted for the 59R/108N (green) samples. Triple-mutants from Kilifi and South Africa cluster together regardless of whether or not *dhfr* genotype is included in constructing the network.

Finally, I constructed a network using data from West Africa as well as East, combining Kilifi1987-88, Kisumu 2002-3, South Africa 1996/9a, South Africa 1996/9b, and Senegal 2003-4. The results are shown in **Figures 5.6 and 5.7**. Looking first at **Figure 5.6**, we see once again that the wild-type parasites from all locations are intermixed. The single-mutants (108N, orange) are dispersed into three clusters, suggesting at least three independent origins of this mutation in these regions. For the 51I/108N double-mutants (blue), there is one main cluster, and it is close to the main cluster of 59R/108N double-mutants (green). The largest node in the figure is that of the canonical triple-mutant allele (node "9N88"). All of the triple-mutants are contained within this node or are only a short distance from it.

When the *dhfr* genotypes are ignored, and only microsatellite haplotypes are used to construct the network of the "All Africa" dataset, the picture changes somewhat (**Figure 5.7**). The most notable feature is that now there are nodes that contain samples

of different *dhfr* genotypes. For example, in the large node of the canonical triple-mutant haplotype (node "9N88"), there are a few samples that carried a wild-type *dhfr*. In addition, there now appear to be numerous independent origins for the single mutation at codon 108 (108N, orange). Finally, the samples cluster into two groups, one on the right with the triple-mutant, and one on the left with most of the double-mutants. Compared to wild-type parasites from Kilifi and South Africa, there are more wild-type parasites from Senegal 2003-4 on the right side of the network ($p<10^{-5}$, two-sided Fisher's exact test), suggesting that there perhaps is some population subdivision.

Discussion

There are a few general comments that can be made about all of the network diagrams. First, there is far more genetic diversity among wild-type parasites than among parasites carrying two or three mutations in *dhfr*, consistent with the recent (on an evolutionary scale) emergence of the SP resistant genotypes. Second, the wild-type parasites from various places and times are mostly intermixed, consistent with earlier findings of little population subdivision among *P. falciparum* in Africa [20]. Finally, the largest feature in every network is the node containing all samples with the canonical triple-mutant haplotype, reflecting the common ancestry of most, and possibly all, triple-mutants in Africa. Though there are some interesting exceptions to these patterns, discussed further below, in general the network diagrams confirm the hypothesis that the triple-mutant parasites in Africa are descended from an Asian triple-mutant immigrant.

Considering first the Kenya dataset (**Figures 5.2 and 5.3**), we see one main origin for parasites carrying three mutations in *dhfr*. Nearly all of the triple-mutants are contained within node KIL23 (**Table 5.11**), and most of those that do not share this exact haplotype are only a short distance away in the network. This topology suggests that the triple-mutants in Kilifi and Kisumu are descended from an Asian progenitor. Some mutation or recombination has occurred since the immigration of the Asian triple-mutant, creating several similar, but not identical, microsatellite haplotypes among the

triple-mutants, seen as smaller purple nodes near the main triple-mutant node. There are, however, a few triple-mutants that do not seem to fit this pattern, namely MCC46 and MCC49. When *dhfr* genotype is ignored (**Figure 5.3**), these triple-mutants from Kisumu appear more closely related to the wild-type parasites from Kilifi than to the other triple-mutants. Unfortunately, there is not enough information to determine whether these samples are independent origins of the triple-mutant genotype, or whether they simply result from recombination of the invading triple-mutant with local wild-type parasites. Typing more samples at more loci, or examining archived samples from earlier time points, could resolve the issue. If they are independent origins of the triple-mutant, their low frequency suggests that they arose more recently than the immigration and local expansion of the Asian triple-mutant.

Another interesting finding in the Kenya dataset is the location of the two samples carrying the *dhfr* genotype 51I/108N/164L (node MCC3G, pink). The networks clearly place these samples with the 51I/108N double-mutants (blue), even when *dhfr* genotype is ignored (**Figure 5.3**). Therefore, it seems that this genotype arose out of the local double-mutant population. This result is of interest because the change to leucine at codon 164 is associated with intense resistance to pyrimethamine and cycloguanil [18,19], but has been seen only rarely in Africa [21,22]. Given how broadly the Asian triple-mutant appears to have spread across Africa, it seems likely that this 164L strain will spread easily too, further reducing the efficacy of antifolate drugs in Africa. If this strain does *not* spread beyond Kisumu, then it suggests that the Asian triple-mutant has some other attribute that drives its spread and expansion. Favoring this second interpretation is the fact that mutations at codon 164 have already been detected at low levels elsewhere in Africa [23,24], but have yet to become widespread.

In the East Africa dataset the picture is extremely clear: there is one origin for each *dhfr* genotype, with the possible exception of the 59R/108N (green) double-mutants. Regardless of whether or not *dhfr* genotypes are included in constructing the network, all parasites carrying three mutations in *dhfr* cluster together, away from the majority of the rest of the samples. Furthermore, they are not closely related to either

type of double-mutant (51I/108N, blue or 59R/108N, green), supporting the hypothesis that the triple-mutants currently found in East Africa are not derived from local double-mutants, but rather result from invasion of an Asian immigrant. Another finding in this dataset is the lack of subdivision between wild-type parasites from Kilifi, Kenya, and those from KwaZulu-Natal, South Africa. As stated above, earlier population genetics studies of *P. falciparum* in Africa found little population subdivision [20], so this finding is not unexpected. It points to the continuous flow of people and parasites along the east coast of Africa.

Finally, let us consider the origin of parasites carrying three mutations in *dhfr* as seen in the All Africa dataset (**Figures 5.6 and 5.7**). Once again, there is one large node that contains almost all of the triple-mutant samples, suggesting a common origin for all triple-mutants. However, there are a few other nodes of triple-mutant parasites whose microsatellite haplotypes are not particularly close to the common triple-mutant haplotype (e.g. nodes NDI S20, NDI P19, and MCC 46 in **Figure 5.6**). While these nodes may represent independent origins of the triple-mutant *dhfr* allele, it seems more likely that they represent recombination between the triple-mutant and local parasites. If they were independent origins, we would expect to see these triple-mutants arising from one of the local double-mutant populations (51I/108N, blue or 59R/108N, green). Instead, these "rogue" triple-mutants are dispersed among the larger, wild-type population.

Figure 5.7, which shows the network created if *dhfr* genotypes are ignored, further supports the hypothesis that the triple-mutants have recombined with local wild-type parasites. There are wild-type parasites from Senegal (e.g. NDI H4, NDI V28, and NDI V48) that have the same microsatellite haplotype as the canonical triple-mutant (**Table 5.14**). Looking at the data from the original study of the Senegalese samples (**Table 5.9**; [5]), we see many shared alleles between the wild-type parasites and those with mutations in *dhfr*. These shared alleles may be the result of recombination between the wild-type and SP-resistant parasites. Of course, recombination is not the only way to explain the shared alleles. They could be the result of homoplasy: multiple evolutionary

paths to the same allele or haplotype. Or both processes could be occurring. One way to judge would be to study earlier samples from Senegal, to see if these alleles were present among the wild-type parasites prior to the appearance of parasites with three mutations in *dhfr*. In addition to a retrospective study, looking at more microsatellites, further from *dhfr*, could shed light on the issue. If these shared alleles stem from recombination, then there should be a clear haplotype block on one side of the breakpoint and a different haplotype on the other side. If, on the other hand, these samples simply represent independent mutations of the microsatellites to the same alleles as are found in the triple-mutant haplotype, then we would not see such haplotype blocks. Though from **Table 5.9** it appears that there are no haplotype blocks, and therefore the shared alleles are due to convergent evolution (homoplasy), it is difficult to judge from only three loci.

One possible effect of the shared alleles between wild-type Senegalese parasites and the triple-mutant is the appearance of population subdivision. Though the wild-type parasites are again intermixed, when *dhfr* genotypes are ignored (**Figure 5.7**) there seems to be a tendency for the samples from Senegal to appear on the right side of the network, nearer the triple-mutant node. This may be due to a difference in microsatellite allele frequencies between East and West African parasites regardless of any recombination with the Asian triple-mutant. For example, almost all samples on the left side of the network have the same (gray) allele at locus -0.1. This locus has low diversity in samples from East Africa, but not from Senegal. Alternatively, the location of the Senegalese samples in the network could be the result of more recombination between wild-type (local) and triple-mutant (immigrant) parasites in Senegal than in East Africa.

As described above, the rate of recombination is directly related to the level of transmission of *P. falciparum*. Therefore, the amount of recombination between the immigrant triple-mutant and local wild-type parasites depends both on transmission intensity and time. Since the history of resistance indicates that the proposed Asian immigrant arrived in East Africa and then spread to West Africa [25], the parasites in East

Africa have had more time to recombine with the Asian haplotype. Furthermore, Kisumu has intense transmission, so we would predict the highest level of recombination between immigrant and local in Kisumu. However, the network in **Figure 5.7** indicates that the parasites from Senegal, not Kisumu, are most closely related to the immigrant triple-mutant. While perhaps Eastern Senegal has a higher level of transmission than Kisumu, Western Senegal does not, and samples from both Senegalese locations appeared mostly on the right side of the network. This discrepancy between prediction and observation may simply reflect the collection times of the various sample sets (the East African samples were collected earlier than the Senegal samples), but it warrants further exploration.

The All Africa network is the only one to include samples with a single mutation in *dhfr* (108N, orange). From both networks (**Figures 5.6 and 5.7**) we see that each one of these single-mutants has its own haplotype. This genetic diversity implies multiple independent origins of the 108N mutation, supporting earlier evidence that mutations in *dhfr* occur relatively frequently [26]. The fact that there are parasites carrying the 108N mutation that have the same microsatellite haplotype as parasites carrying a wild-type *dhfr* (e.g. NDI H12/NDI H6, NDI S18/NDI T7) further supports this hypothesis. If so, then why are there apparently so few origins of *dhfr* genotypes with more than one mutation? The simplest explanation is that – even if mutations occur relatively frequently – their survival in the population is rare. Most mutations, even if advantageous, are lost [27]. Furthermore, if there are already advantageous mutations in the population, such as 108N, then the relative selective advantage of an additional mutation is reduced [27].

Within the All Africa networks we see evidence of the imperfection of my standardization method. As stated in the methods section, the samples in the South Africa 1996/9a and South Africa 1996/9b datasets are mostly, if not entirely, the same. In **Figures 5.6 and 5.7**, we see multiple closely related pairs of nodes in which one is from the South Africa 1996/9a dataset and the other is from the South Africa 1996/9b dataset (e.g. ROP25/AF40, ROP28/AF41, ROP31/AF32, and ROP41/AF26). I believe

that each of these pairs reflects a single sample, but that the standardization process left them with slightly different haplotypes. One reason this imprecision arises is that the repeat unit of these microsatellites is not always exactly two when measured by capillary electrophoresis, even though each is an $(AT)_n$ repeat. For example, when I determined allele sizes at locus +5.8, I found that the allele sizes called by the software were spaced at intervals slightly greater than two. Therefore, when I rounded to the nearest base the alleles switched from odd to even at allele 109. That is, though the alleles were evenly spaced, in the data they are presented as: 101, 103, 105, 107, 109, 112, 114, 116, etc. Assuming that other researchers follow a similar pattern for calling alleles, subtracting the same value from every allele at a given locus will lead to inaccuracies in the standardized dataset. However, because Network uses a stepwise mutation model this inaccuracy will not affect the overall topology of the network.

The overall picture remains one of a single, Asian origin for parasites carrying three mutations in *dhfr*. In Kisumu there may be independent origins of triple-mutants, but it seems likely that they arose more recently than the immigration of the Asian triple-mutant. A longitudinal, retrospective study would be needed to be sure. Likewise, a retrospective study of samples from Senegal would determine whether the wild-type and triple-mutant parasites share alleles due to recombination or due to multiple origins of the same microsatellite alleles (homoplasy). Such a study might also be able to shed light on the recombination and mutation rates of *P. falciparum*, knowledge that would greatly aid our efforts to understand the evolution of drug resistance.

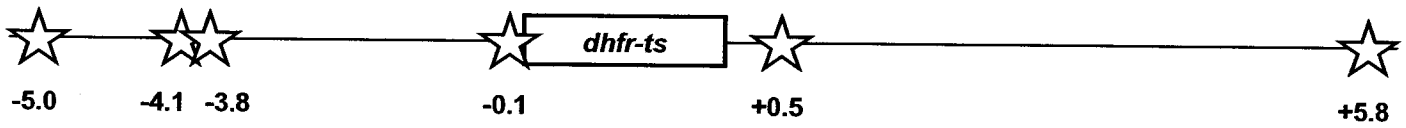


Figure 5.1 – Location of Microsatellites Relative to the Gene *Dhfr-ts*. Names of loci indicate distance, in kilobases, upstream from the start (negative numbers) or downstream from the end (positive numbers) of the gene. The gene itself is about 1.8 kb long.

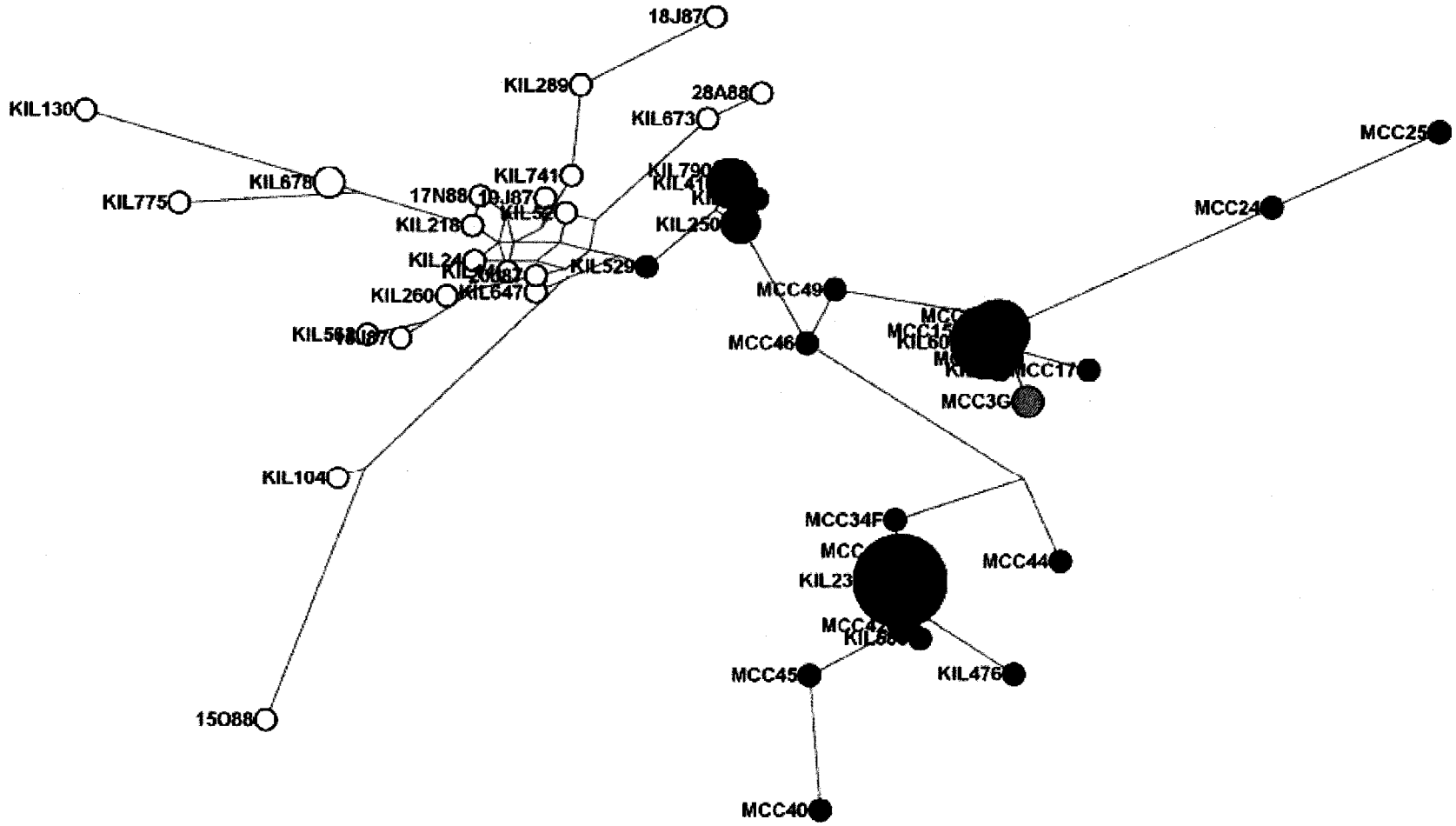


Figure 5.2 – Median-Joining Network of Kilifi 1987-88, Kilifi 1993-95, and Kisumu 2002-3; SNP Weight = 90. Alleles are color coded as follows: purple = triple-mutant (51I/59R/108N); pink = alternate triple (51I/108N/164L); blue = double-mutant (51I/108N); green = double-mutant (59R/108N); yellow = wild-type. Each node represents a haplotype, with the size of the node proportional to the number of samples with that haplotype. Haplotypes are labeled according to the first sample with that haplotype (**see Table 5.11**).

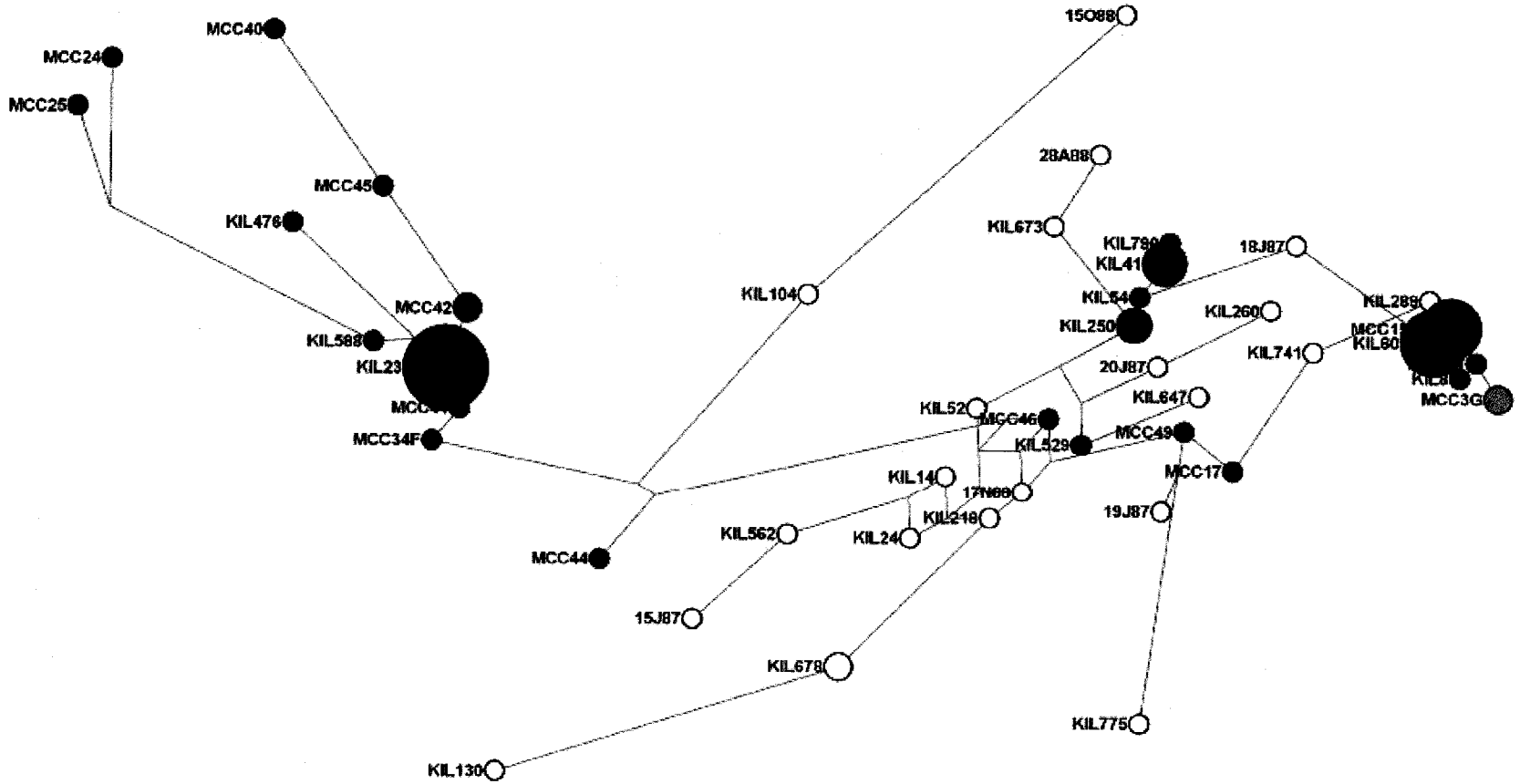


Figure 5.3 – Median-Joining Network of Kilifi 1987-88, Kilifi 1993-95, and Kisumu 2002-3; SNP Weight = 0. Alleles are color coded as follows: purple = triple-mutant (51I/59R/108N); pink = alternate triple (51I/108N/164L); blue = double-mutant (51I/108N); green = double-mutant (59R/108N); yellow = wild-type. Each node represents a haplotype, with the size of the node proportional to the number of samples with that haplotype. Haplotypes are labeled according to the first sample with that haplotype **(see Table 5.11)**.

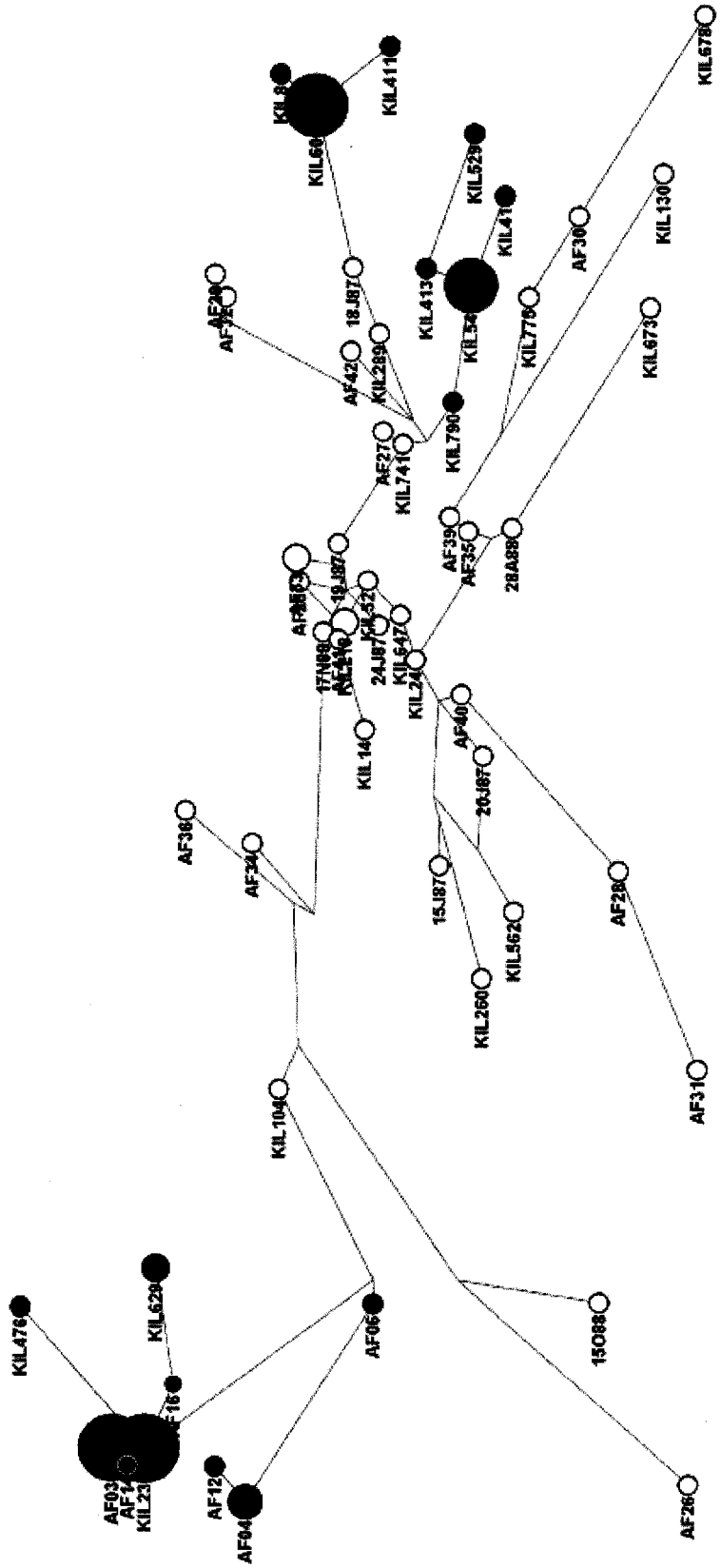


Figure 5.4 – Median-Joining Network of Kilifi 1987-88, Kilifi 1993-95, and South Africa 1996/9b; SNP Weight = 90.

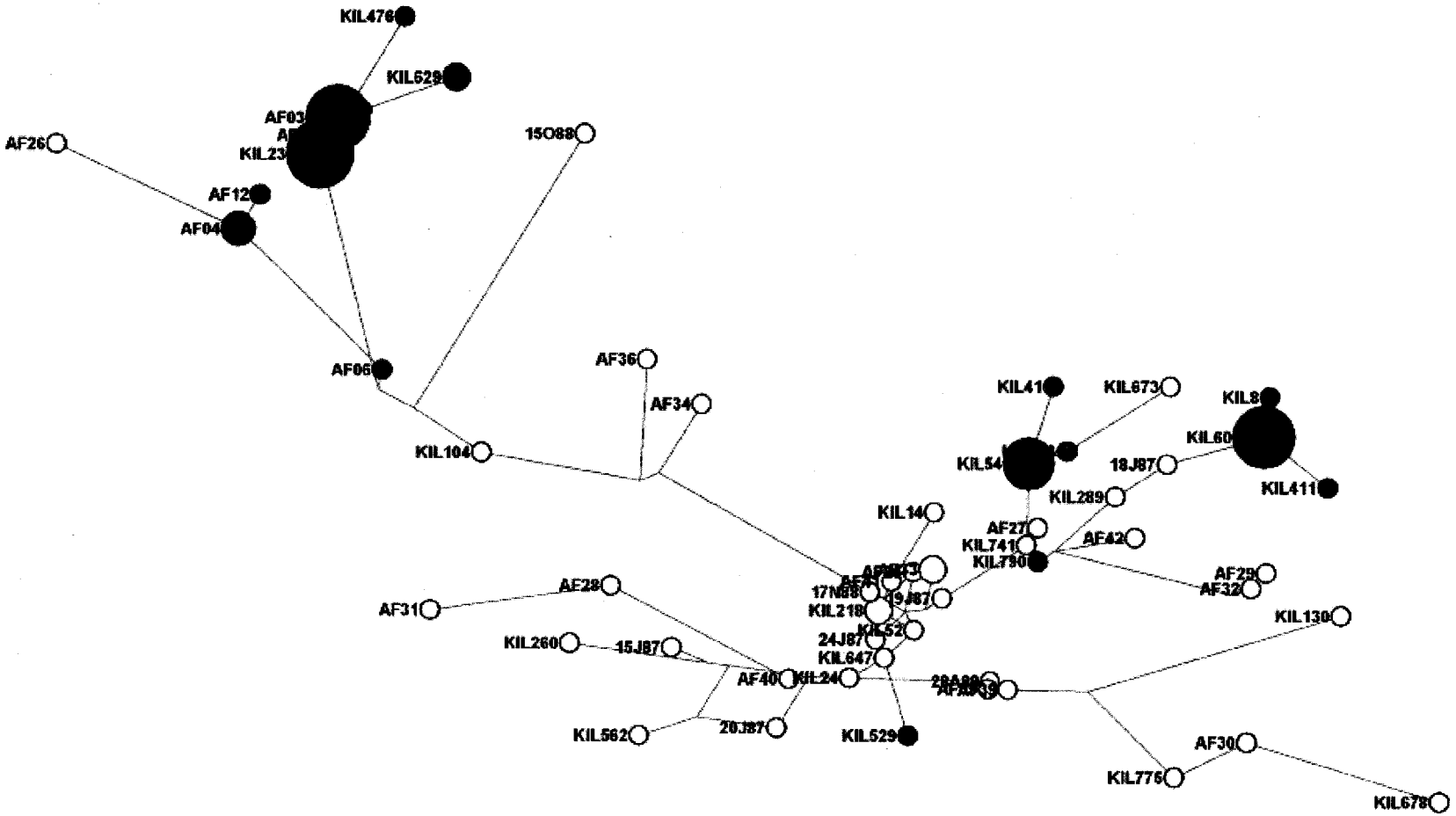


Figure 5.5 – Median-Joining Network of Kilifi 1987-88, Kilifi 1993-95, and South Africa 1996/9b; SNP Weight = 0.

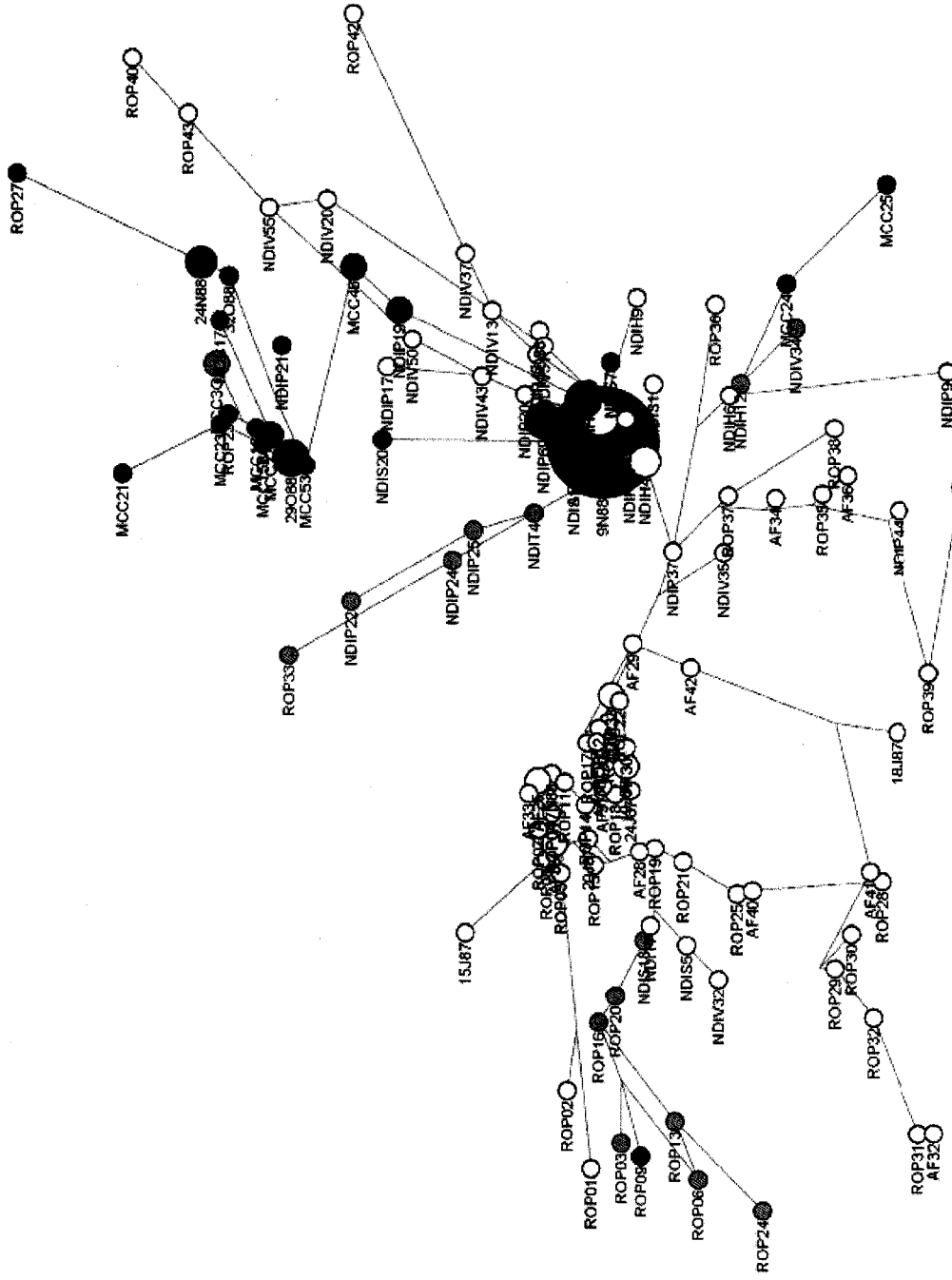


Figure 5.6 – Median-Joining Network of Kilifi 1987-88, Kisumu 2002-3, South Africa 1996/9a & B, Senegal 2003-4; SNP Weight = 90.

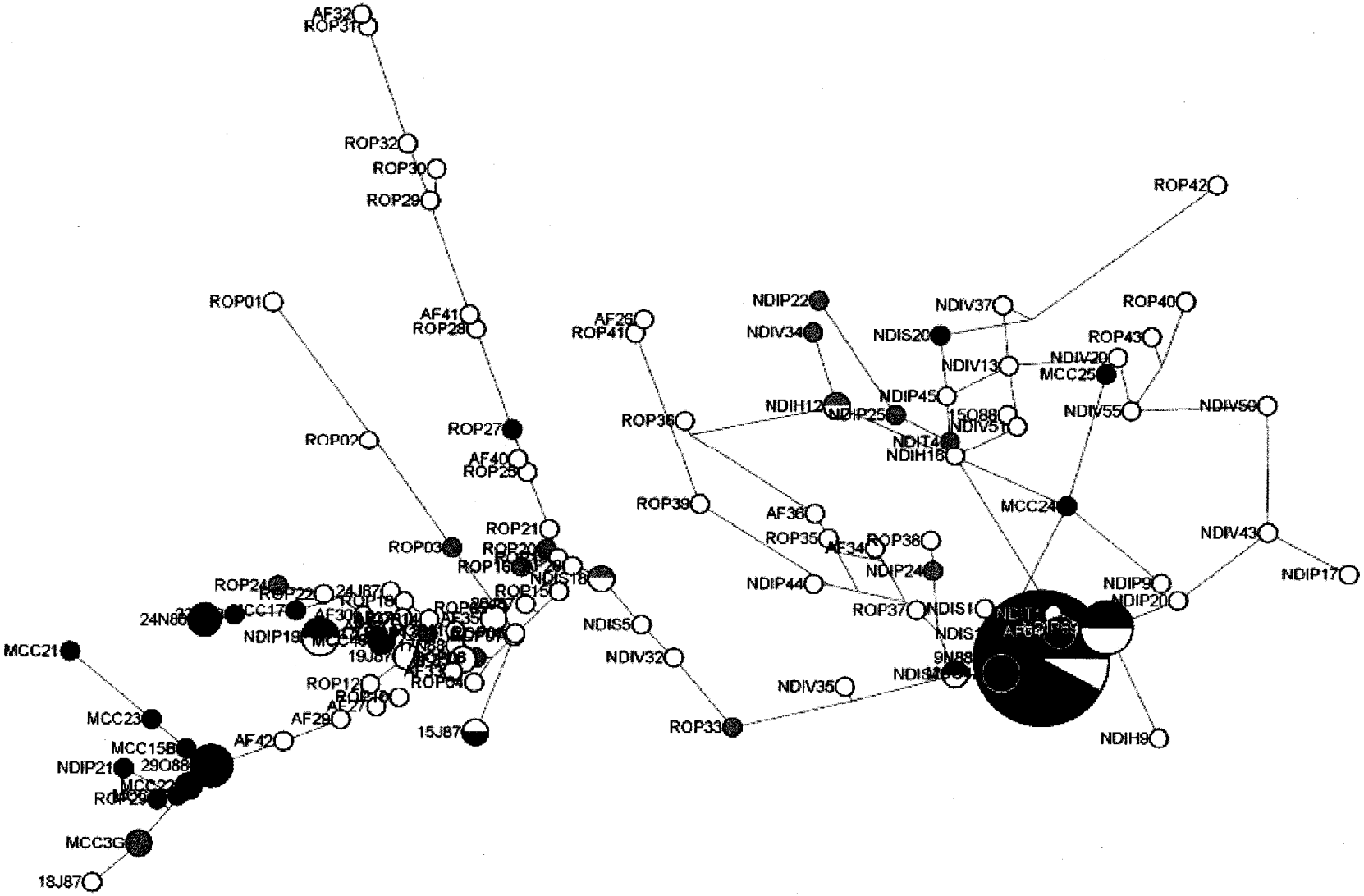


Figure 5.7 – Median-Joining Network of Kilifi 1987-88, Kisumu 2002-3, South Africa 1996/9a & B, Senegal 2003-4; SNP Weight = 0.

Table 5.1 – Summary of Studies Included.

Study	Reference	Location	Years of Collection	Transmission Intensity [14]	Number of Samples	Type of Study
Kilifi 1987-88	Chapter 4	Kilifi, Kenya	1987-88	Mesoendemic	13	Drug resistance
Kilifi 1993-95	Chapter 3	Kilifi, Kenya	1993-95	Mesoendemic	42	Drug resistance
Kisumu 2002-3	[4]	Kisumu, Kenya	2002-3	Holo/Hyperendemic	35	Cotrimoxazole prophylaxis
South Africa 1996-9a & b	[2,3]	KwaZulu Natal, South Africa	1996,1999	Hypoendemic	~43*	
Senegal 2003-4	[5]	Eastern Senegal	2003-4	Holo/Hyperendemic	29	Survey of infected patients
Senegal 2003-4	[5]	Western Senegal	2003-4	Mesoendemic	30	Survey of infected patients

* The samples from these two papers are the same set. However, because of differences in experimental technique it is not possible to determine which samples from the 2003 paper correspond to which haplotypes in the 2004 paper. Therefore, it is unclear how many independent samples are actually represented in the combined dataset.

Table 5.2 – Microsatellite Loci Included in Each Study.

	Microsatellite Loci					
Study	-5.0	-4.1	-3.8	-0.1	+0.5	+5.8
Kilifi 1987-88	x	x	x	x	x	x
Kilifi 1993-95	x		x	x	x	x
Kisumu 2002-3	x	x	x	x	x	
South Africa 1996-9a	x	x		x		
South Africa 1996-9b	x	x		x	x	x
Senegal 2003-4	x	x		x		

Table 5.3 – Values Subtracted from Each Locus to Standardize the Data across Samples.

	Microsatellite Loci					
Study	-5.0	-4.1	-3.8	-0.1	+0.5	+5.8
Kilifi 1987-88	181	133	172	81	72.5	88
Kilifi 1993-95	185	...	176	85	77	92
Kisumu 2002-3	182	134	172	81	73	...
South Africa 1996-9a	183	137	...	86	...	...
South Africa 1996-9b	183	137	...	84	76	91
Senegal 2003-4	191	143	...	85	...	...

Table 5.4 – Microsatellite Haplotypes from Kilifi, Kenya, 1987-88.

		Microsatellites				DHFR Codons			Microsatellites	
Pt.	Date	-5.0	-4.1	-3.8	-0.1	51	59	108	+0.5	+5.8
15	15-Jul-87	215	167	192	85	N	C	S	89.5	109
18	16-Jul-87	191	185	196	85	N	C	S	92.5	109
19	16-Jul-87		---	186	85	N	C	S	89.5*	107
20	16-Jul-87	209	169	196	85	N	C	S	89.5*	105
24	16-Jul-87	205		190	85	N	C	S	---	107
9	11-Nov-88									116
15	5-Oct-88	207	174	208	109	N	C	S	94.5	112
17	16-Nov-88	203	169	190	85	N	C	S	91.5*	107
24	24-Nov-88	197	182	198	85	N	R	N	89.5*	112
28	6-Aug-88	---	171	200	85	N	C	S	(96.5)	109
29	19-Oct-88	---	175	188	85	I	C	N	91.5*	114
32	14-Oct-88	197	180	198	85	N	R	N	---	112
33	29-Oct-88	197	182	198	85	N	R	N	(90.5)	112

Table 5.5 – Microsatellite Haplotypes from Kilifi, Kenya, 1993-95.

	Microsatellites			DHFR Codons			Microsatellites	
Sample	-5.0	-3.8	-0.1	51	59	108	+0.5	+5.8
23								
244								
476	213	194						
588		192						
629								(120)
698								
716								
779	195	192						(110)
800								
8	195	192		I	C	N	99	
60	195	192		I	C	N	97	118
83	195	192		I	C	N	97	118
102	195	192		I	C	N	97	118
113	195	192		I	C	N	97	118
411	195	192		I	C	N	97	122
457	195	192		I	C	N	97	(118)
632	195	192		I	C	N	97	118
696	195	192		I	C	N	97	118
713	195	192		I	C	N	97	118
802	195	192		I	C	N	97	
41	201	202		N	R	N	95	120
54	(201)	200		N	R	N	95	116
96	201	202		N	R	N	95	116
250		200		N	R	N	95	116
255		200		N	R	N	95	116
413	203	200		N	R	N	95	116
529	209	202		N	R	N	95	116
790	(200)	202		N	R	N	(95)	
14	209	192		N	C	S	95	106
24	211	194		N	C	S	95	
52	207	198		N	C	S	95	
104	207	206	107	N	C	S	95	110
130	217	190		N	C	S	112	110
218		194		N	C	S	97	
260	217	200		N	C	S	97	118
289	195	192		N	C	S	95	
562	217	192		N	C	S	95	106
647	209	208		N	C	S	95	
673	203	200		N	C	S	101	116
678	209	(194)		N	C	S		118
741	201	192		N	C	S	95	
775		186		N	C	S	(105)	110

Table 5.6 – Microsatellite Haplotypes from Kisumu, Kenya, 2002-3.

	Microsatellites				DHFR Codons				
Sample	**-5.3**	**-4.1**	**-3.8**	**-0.1**	**51**	**59**	**108**	**164**	**+0.5**
15b	191		188	85	I	C	N	I	93
16	191	179	188	85	I	C	N	I	93
17	199	178	188	85	I	C	N	I	93
18	191	176	188	85	I	C	N	I	93
19	191	176	188	85	I	C	N	I	93
20	191	176	188	85	I	C	N	I	93
21	191	164	188	85	I	C	N	I	95
22	191	178	188	85	I	C	N	I	93
23	191	171	188	85	I	C	N	I	93
24	191		188		I	C	N	I	93
25	191	178	188	110	I	C	N	I	100
26	191	178	188	85	I	C	N	I	93
29									
30									
31									
32									
33									
34d									
34e									
35									
36									
37									
38									
34f									99
40									116
41									101
42				102					
43				102					
44				102					87
45				102					110
46				85					93
49			188	85					93
53	191	176	188	85					93
3g	191	183	188	85	I	C	N	L	97
3h	191	183	188	85	I	C	N	L	97

Table 5.7 – Microsatellite Haplotypes from KwaZulu-Natal, South Africa, 1996/9 – Dataset A.

	Microsatellites			DHFR Codons			
Haplotype	**-5.3**	**-4.1**	**-0.1**	**51**	**59**	**108**	**164**
34 n=43							
26 n=11	199	186	90	N	R	N	I
27	212	186	90	N	R	N	I
(n=2) 9	217	171	90	I	C	N	I
(n=4) 23	193	183	90	I	C	N	I
3	213	167	90	N	C	N	I
6	205	171	90	N	C	N	I
(n=2) 13	205	175	90	N	C	N	I
16	213	175	90	N	C	N	I
20	213	177	90	N	C	N	I
24	205	183	90	N	C	N	I
33	208	181	98	N	C	N	I
1	208	157	90	N	C	S	I
2	214	159	90	N	C	S	I
4	206	169	90	N	C	S	I
5	210	169	90	N	C	S	I
7	208	171	90	N	C	S	I
8	210	171	90	N	C	S	I
10	200	173	90	N	C	S	I
(n=2) 11	206	173	90	N	C	S	I
12	200	175	90	N	C	S	I
(n=2) 14	206	175	90	N	C	S	I
15	210	175	90	N	C	S	I
17	202	177	90	N	C	S	I
18	206	177	90	N	C	S	I
19	212	177	90	N	C	S	I
21	212	179	90	N	C	S	I
22	204	181	90	N	C	S	I
25	212	183	90	N	C	S	I
28	208	189	90	N	C	S	I
29	206	196	90	N	C	S	I
30	210	196	90	N	C	S	I
31	198	200	90	N	C	S	I
32	206	200	90	N	C	S	I
35	204	171	102	N	C	S	I
36	195	169	104	N	C	S	I
37	206	175	104	N	C	S	I
38	210	179	104	N	C	S	I
39	210	165	106	N	C	S	I
40	208	189	108	N	C	S	I
41	210	159	112	N	C	S	I
42	226	183	112	N	C	S	I
43	205	189	112	N	C	S	I

Table 5.8 – Microsatellite Haplotypes from KwaZulu-Natal, South Africa, 1996/9 – Dataset B.

	Microsatellites			DHFR Codons				Microsatellites	
Sample	-5.0	-4.1	-0.1	51	59	108	164	+0.5	+5.8
1									
2									
3									109
4		175						83	
5		175							109
6								93	
7									
8								83	
9									109
10									109
11									109
12								83	109
13									109
14									110
15									109
16									114
17									109
18									
19		175							
20									109
21									109
22									109
23									
24								83	
25	205	172	88	N	C	S	I	93	109
26	209	159	110	N	C	S	I	83	113
27	199	174	88	N	C	S	I	93	
28	211		88	N	C	S	I	83	
29	197	175	88	N	C	S	I	85	109
30	205	179	88	N	C	S	I	106	111
31	205	172	88	N	C	S	I	83	114
32	197	200	88	N	C	S	I	85	112
33	204	172	88	N	C	S	I	93	109
34	205	174	99	N	C	S	I	100	109
35	209	170	88	N	C	S	I	100	110
36	203	170	100	N	C	S	I	96	105
37	205	167	88	N	C	S	I	96	108
38	204	170	88	N	C	S	I	93	109
39	209	170	88	N	C	S	I	100	109
40	211	183	88	N	C	S	I	93	
41	207	189	88	N	C	S	I	93	108
42	197	179	88	N	C	S	I	96	109

Table 5.9 – Microsatellite Haplotypes from Senegal, 2003-5. Bold samples are from eastern Senegal, non-bolded from western.

	Microsatellites			DHFR Codons			
Sample	**-5.0**	**-4.1**	**-0.1**	**51**	**59**	**108**	**164**
P19.03			93				
P51.03			93				
P13.03							
P14.03							
P15.03							
P35.03							
P42.03							
P46.03							
P53.03							
P69.03			113				
Th11.03			113				
S17.03							
V21.04							
V52.04							
S20.03	225						
V54.04			113				
V57.04			113				
P21.03	200	188	93	N	R	N	I
S7.03			103	N	R	N	I
S13.03	208			N	R	N	I
Th12.03	200		103	N	C	N	I
P24.03	218		103	N	C	N	I
P25.03	218	179		N	C	N	I
Th10.03				N	C	N	I
P22.03	218	179	118	N	C	N	I
S18.03	221		93	N	C	N	I
V34.04	200	188	103	N	C	N	I
T4.03	218			N	C	N	I
P33.03			93	N	C	S	I
Th6.03	200		103	N	C	S	I
P37.03			103	N	C	S	I
P44.03	214	177	106	N	C	S	I
Th4.03				N	C	S	I
P9.03	187			N	C	S	I
Th16.03	217			N	C	S	I
P45.03	221			N	C	S	I
Th9.03		175	113	N	C	S	I
P49.03			113	N	C	S	I
P78.03			113	N	C	S	I
P17.03	204		116	N	C	S	I
P20.03			118	N	C	S	I
T6.03			93	N	C	S	I
S5.03	217		93	N	C	S	I
T7.03	221		93	N	C	S	I
V32.04	217	186	93	N	C	S	I
V35.04		188	100	N	C	S	I
V28.04				N	C	S	I
V48.04				N	C	S	I
T1.03	214			N	C	S	I
S1.03	208	185		N	C	S	I
V25.04			113	N	C	S	I
V40.04			113	N	C	S	I
V51.04	217		113	N	C	S	I
V13.04	221		113	N	C	S	I
V37.04	225		113	N	C	S	I
V55.04	214	193	116	N	C	S	I
V20.04	221	193	116	N	C	S	I
V43.04	208		118	N	C	S	I
V50.04	208	188	118	N	C	S	I

Table 5.10 – Sample Labels in the Network Diagrams.

Study	Prefix	Example	Full Name of Example	In Table
Kilifi 1987-88	<none>	15O88	Patient 15, October 1988	5.4
Kilifi 1993-95	KIL	KIL23	23	5.5
Kisumu 2002-3	MCC	MCC3G	3g	5.6
South Africa 1996-9a	ROP	ROP36	36	5.7
South Africa 1996-9b	AF	AF12	12	5.8
Senegal 2003-4	NDI	NDIP20 NDIH11	P20.03 Th11.03	5.9 5.9

Table 5.11 – Nodes Containing Multiple Samples in Figures 5.2 and 5.3.

Node:	Also Contains:
KIL 41	KIL 96 24N88 32O88 33O88
KIL 60	KIL 83 KIL 102 KIL 113 KIL 411 KIL 457 KIL 632 KIL 696 KIL 713 KIL 802 29O88
KIL 250	KIL 255 KIL 413
KIL 678	24J87
MCC 3g	MCC 3h
MCC 15b	MCC 16 MCC 18 MCC 19 MCC 20 MCC 22 MCC 23 MCC 26

Table 5.12 – Nodes Containing Multiple Samples in Figures 5.4 and 5.5.

Node:	Also contains:
KIL 54	KIL 96 KIL 250 KIL 255 24N88 32O88 33O88
KIL 60	KIL 83 KIL 102 KIL 113 KIL 457 KIL 632 KIL 696 KIL 713 KIL 802 29O88

Table 5.13 – Nodes Containing Multiple Samples in Figure 5.6.

Node:	Also contains:
[redacted]	[redacted]
15J87	ROP09
24N88	33O88 ROP26
29O66	MCC 18 MCC 19 MCC 20
[redacted]	[redacted]
AF 25	AF 31
AF 35	AF 39
MCC 3g	MCC 3h
MCC 22	MCC 26
[redacted]	[redacted]
NDI H4	NDI V28 NDI V48
[redacted]	[redacted]
NDI P49	NDI P78 NDI V25 NDI V40
[redacted]	[redacted]

Table 5.14 – Nodes Containing Multiple Samples in Figure 5.7.

Node:	Also Contains: Same as Table 5.13	New
[redacted]	[redacted]	NDI H10 NDI H4 NDI V28 NDI V48
19J87		28A88
24N88	33O88 ROP26	
29O88	MCC 18 MCC 19 MCC 20	[redacted]
[redacted]	[redacted]	
AF 25 **AF 35**	AF 31 AF 39	
MCC 3g	MCC 3h	
MCC 22	MCC 26	
[redacted]	[redacted]	
[redacted]	[redacted]	NDI P33 NDI T6
[redacted]	[redacted]	NDI P49 NDI P78 NDI V25 NDI V40
NDI H12		NDI H6
NDI S7		NDI P37
NDI S18		NDI T7

NOTES TO CHAPTER FIVE

1. Nair, S. et al. A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. *Mol Biol Evol* **20**, 1526-36 (2003).

2. Roper, C. et al. Antifolate antimalarial resistance in southeast Africa: a population-based analysis. *Lancet* **361**, 1174-81 (2003).

3. Roper, C. et al. Intercontinental spread of pyrimethamine-resistant malaria. *Science* **305**, 1124 (2004).

4. McCollum, A.M. et al. Antifolate Resistance in Plasmodium falciparum: Multiple Origins and Identification of Novel dhfr Alleles. *J Infect Dis* **194**, 189-97 (2006).

5. Ndiaye, D. et al. Defining the origin of Plasmodium falciparum resistant dhfr isolates in Senegal. *Acta Trop* **99**, 106-11 (2006).

6. Noranate, N. et al. Rapid Dissemination of Plasmodium falciparum Drug Resistance Despite Strictly Controlled Antimalarial Use. *PLoS ONE* **2**, e139 (2007).

7. Mharakurwa, S. Plasmodium falciparum transmission rate and selection for drug resistance: a vexed association or a key to successful control? *Int J Parasitol* **34**, 1483-7 (2004).

8. Ariey, F. & Robert, V. The puzzling links between malaria transmission and drug resistance. *Trends Parasitol* **19**, 158-160 (2003).

9. Schmidt, K.F. Malaria research. Inbred parasites may spur resistance. *Science* **269**, 1670 (1995).

10. Hastings, I.M. & Mackinnon, M.J. The emergence of drug-resistant malaria. *Parasitology* **117 (Pt 5)**, 411-7 (1998).

11. Mackinnon, M.J. & Hastings, I.M. The evolution of multiple drug resistance in malaria parasites. *Trans R Soc Trop Med Hyg* **92**, 188-95 (1998).

12. Molyneux, D.H., Floyd, K., Barnish, G. & Fevre, E.M. Transmission control and drug resistance in malaria: a crucial interaction. *Parasitol Today* **15**, 238-40 (1999).

13. Talisuna, A.O. et al. Intensity of transmission and spread of gene mutations linked to chloroquine and sulphadoxine-pyrimethamine resistance in falciparum malaria. *Int J Parasitol* **33**, 1051-8 (2003).

14. Snow, R.W., Guerra, C.A., Noor, A.M., Myint, H.Y. & Hay, S.I. The global distribution of clinical episodes of Plasmodium falciparum malaria. *Nature* **434**, 214-7 (2005).

15. Mbogo, C.N. et al. Relationships between Plasmodium falciparum transmission by vector populations and the incidence of severe disease at nine sites on the Kenyan coast. *Am J Trop Med Hyg* **52**, 201-6 (1995).

16. Bredenkamp, B.L., Sharp, B.L., Mthembu, S.D., Durrheim, D.N. & Barnes, K.I. Failure of sulphadoxine-pyrimethamine in treating Plasmodium falciparum malaria in KwaZulu-Natal. *S Afr Med J* **91**, 970, 972 (2001).

17. Ndiaye, D. et al. Mutations in Plasmodium falciparum dihydrofolate reductase and dihydropteroate synthase genes in Senegal. *Trop Med Int Health* **10**, 1176-9 (2005).

18. Cortese, J.F. & Plowe, C.V. Antifolate resistance due to new and known Plasmodium falciparum dihydrofolate reductase mutations expressed in yeast. *Mol Biochem Parasitol* **94**, 205-14 (1998).

19. Watkins, W.M., Mberu, E.K., Winstanley, P.A. & Plowe, C.V. The efficacy of antifolate antimalarial combinations in Africa: a predictive model based on pharmacodynamic and pharmacokinetic analyses. *Parasitol Today* **13**, 459-64 (1997).

20. Anderson, T.J. et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. *Mol Biol Evol* **17**, 1467-82 (2000).

21. Fernandes, N., Figueiredo, P., do Rosario, V.E. & Cravo, P. Analysis of sulphadoxine/pyrimethamine resistance-conferring mutations of Plasmodium falciparum from Mozambique reveals the absence of the dihydrofolate reductase 164L mutant. *Malar J* **6**, 35 (2007).

22. Nzila, A. et al. Why has the dihydrofolate reductase 164 mutation not consistently been found in Africa yet? *Trans R Soc Trop Med Hyg* **99**, 341-6 (2005).

23. Bates, S.J. et al. Rare, highly pyrimethamine-resistant alleles of the Plasmodium falciparum dihydrofolate reductase gene from 5 African sites. *J Infect Dis* **190**, 1783-92 (2004).

24. Hastings, M.D. et al. Highly pyrimethamine-resistant alleles of dihydrofolate reductase in isolates of Plasmodium falciparum from Tanzania. *Trans R Soc Trop Med Hyg* **96**, 674-6 (2002).

25. White, N.J. Antimalarial drug resistance: the pace quickens. *J Antimicrob Chemother* **30**, 571-85 (1992).

26. Paget-McNicol, S. & Saul, A. Mutation rates in the dihydrofolate reductase gene of Plasmodium falciparum. *Parasitology* **122**, 497-505 (2001).

27. Hastings, I.M. The origins of antimalarial drug resistance. *Trends Parasitol* **20**, 512-8 (2004).

CHAPTER SIX: CLOSING THOUGHTS & FUTURE DIRECTIONS

The overall conclusion from the preceding analyses is that parasites carrying three mutations in *dhfr* were initially selected in Southeast Asia and then spread across Asia to Africa, remaining at low levels as it spread. However, as soon as the antifolate sulfadoxine-pyrimethamine (SP) was used in any location, the population of these lurking triple-mutants rapidly expanded, leading to the observed loss in clinical efficacy of SP. While the data clearly support the hypothesis that a triple-mutant invaded Africa from Asia long before SP was widely used, they do not indicate when this immigrant arrived, or how it was able to spread and persist in the absence of drug pressure from SP use. This final chapter will discuss lingering questions and future directions of research.

The origins of the triple-mutant allele of *dhfr*

As stated above, the data presented here and previously [1] indicate that African parasites carrying three mutations in *dhfr* are derived from an Asian immigrant. Moreover, this immigrant was present prior to any widespread use of SP. As we saw in the Kilifi 1987-88 dataset (Chapter 4), the use of SP in a small group of children led to the rapid appearance of many mutant *dhfr* alleles in subsequent samples from those children, including – in one child – a parasite carrying the canonical Asian haplotype. A similar result was found in a longitudinal study in Dielmo, Senegal [2]. Before the use of SP, triple-mutants were present at low levels. After SP became the first-line treatment for uncomplicated malaria, the prevalence of triple-mutants rapidly increased.

There are, however, triple-mutant alleles in Africa that do not have the microsatellite haplotype of Asian samples. These alternate haplotypes could result from recombination of the invading triple-mutant with local parasites, from mutation of microsatellites within the triple-mutant haplotype, or from independent, African

origins of parasites carrying three mutations in *dhfr*. From the data in Chapter 5, it appears that each of these is contributing to the lack of complete uniformity of haplotype among the triple-mutant parasites in Africa.

The evidence for recombination is seen most clearly in the data from Kilifi 1993-95 (Chapter 3) and in the data from Senegal 2003-4 [3]. As mentioned in Chapter 3, sample 779 of the Kilifi 1993-95 dataset appears to be the result of recombination between a parasite carrying the Asian triple-mutant haplotype (purple) and one carrying the common haplotype for double-mutants (51I/108N, blue) (**Table 6.1**). The recombination breakpoint is located somewhere between loci -3.8 and -0.1, as alleles upstream of this point are of the triple-mutant haplotype and alleles downstream are of the double-mutant haplotype (**Table 5.5**). In the data from Senegal (**Table 5.9**), we see numerous shared alleles between wild-type parasites and those with three mutations in *dhfr*. Though the lack of information on microsatellite loci downstream of *dhfr* makes the case less clear, the fact that there are wild-type and triple-mutant samples with the same microsatellite haplotypes suggests recombination. This interpretation assumes that if two microsatellite alleles are identical then they are identical by descent; that is, the probability of two different alleles mutating to become identical is very low.

The allelic diversity of triple-mutant parasites from Senegal is also suggestive of mutation of the immigrant Asian haplotype after its arrival in Africa. For example, the change from 109 to 113 at locus -0.1 in samples P69.03, Th11.03, V54.04, and V57.04 likely result from mutation of that locus (**Table 5.9**), rather than recombination. The loci on either side of that microsatellite still carry the alleles characteristic of the triple-mutant haplotype, so a double recombination would be required to produce that haplotype. Likewise, samples 42-45 from Kisumu 2002-3 (**Table 5.6**) probably indicate mutation of locus -0.1 within the population of invading Asian triple-mutants. If we knew the mutation rate of this locus we could perhaps use these data to calculate when the Asian triple-mutant arrived in Kisumu and in Senegal. Unfortunately, we know very little about the mutation rate of microsatellites in natural

populations of *P. falciparum.* The only data come from a single genetic cross [4,5], so any estimation will await further data from natural populations.

Most intriguing is the evidence for an independent, African origin of parasites carrying the same three mutations in *dhfr* as the Asian allele at the codons 51/59/108. Notably, in the samples from Kisumu 2002-3 there were a handful of parasites that carried three mutations in *dhfr* yet a haplotype similar to that of local parasites carrying only two mutations in *dhfr* (samples 49 and 53 in **Table 5.6**). Though these parasites could simply result from recombination between local double-mutants and immigrant triple-mutants, the fact that the alleles at the microsatellite loci on both sides of *dhfr* have a local (non-triple) allele argues that these samples represent independent origins of the triple-mutant. If they were the result of recombination, we would expect to find the triple-mutant haplotype on one side of the breakpoint and the double-mutant haplotype on the other.

Nevertheless, the vast majority of triple-mutants in Africa are derived from an Asian immigrant. Moreover, I would argue that any triple-mutant parasites in Kisumu that arose locally did so long after the arrival of the Asian triple-mutant. As we see in the samples from Kilifi and South Africa, a substantial proportion of samples are double-mutants of African origin. With the continued use of SP, it is not surprising to find the emergence of another favorable *dhfr* mutation on this double-mutant haplotype. While looking at more loci could resolve whether samples 49 and 53 in the Kisumu 2002-3 dataset are an independent origin or the result of recombination, looking at earlier samples would be even more informative. Such a sample set would indicate when the local triple-mutant haplotype appeared in Kisumu relative to the Asian triple-mutant haplotype. Future work in our lab will focus on creating a longitudinal dataset from Kilifi (1997-2006). If we see the emergence of triple-mutants from the local double-mutant population it will support the hypothesis that the novel triple-mutants in Kisumu arose from the local population after the invading triple-mutant had established itself.

The discussion above assumes that the parasites with two mutations in *dhfr*, whether 51I/108N (blue) or 59R/108N (green), arose independently in Africa. That is, the double-mutant parasites in Africa are assumed to be local. Because they share microsatellite alleles with wild-type African parasites, it seems likely that they arose out of this population. In particular, most wild-type and double-mutant parasites in East Africa share a common allele at locus -0.1, colored gray in the haplotype tables in this thesis. Specifically, there is a 20 bp difference between the allele of the triple-mutant haplotype at locus -0.1 and the most common allele of parasites that do not carry three mutations in *dhfr*. Therefore, we infer that the double-mutants are derived from local wild-type parasites, while the triple-mutants are immigrants. Of course, it is possible that this (gray) allele is found worldwide at locus -0.1, and in fact double-mutants (59R/108N) from Papua New Guinea carry this allele at locus -0.1 [6]. However, wild-type and double-mutant parasites from Kilifi share alleles at loci other than -0.1, supporting the local origin of the double-mutants in Kilifi.

Though the identity of the microsatellite at -0.1 supports that contention that the double-mutants are likely of local origin, they increase in prevalence just as quickly as the triple-mutants when drug pressure is applied. Looking at the data from the Kilifi 1987-88 dataset (**Table 4.3**), we see that samples collected in 1987, prior to any SP use, are all wild-type, while those collected in 1988, after SP use, show numerous double-mutants. Therefore, it seems that the double-mutants were already present in the population – just like the triple-mutants. Since the double-mutants arose locally, what led to their emergence prior to SP use? One likely explanation is the use of cotrimoxazole, an antibiotic combination of trimethoprim and sulfamethoxazole that targets dihydrofolate reductase and dihydropteroate synthase, just like SP. Cotrimoxazole is sometimes promoted for treatment of childhood fevers in Africa, especially in places that lack diagnostic facilities [7-9]. If the fevers are due to malaria, then the parasites are exposed to trimethoprim. Cross-resistance between pyrimethamine and trimethoprim has been demonstrated [10], so use of cotrimoxazole could select for the *dhfr* double-mutants. Alternatively, early use of pyrimethamine as

a monotherapy could have selected for these mutations [11]. As described in Chapter 4, pyrimethamine was first used alone to treat malaria. Because it rapidly lost efficacy, it was withdrawn as a monotherapy and later reintroduced in combination with sulfadoxine. Presumably, it lost efficacy because the parasites acquired mutations in *dhfr*. These mutations may have persisted in the population at low levels, and then rapidly increased in prevalence when use of SP began.

As discussed in previous chapters, it has been proposed that the triple-mutant *dhfr* allele arrived in Africa in the same parasite as the chloroquine-resistant allele of *pfcrt* [1]. This hypothesis is based on the following: chloroquine resistance appeared first in East Africa in 1978 and then spread west, following trade routes [12]; chloroquine resistant parasites in Africa are all descended from an Asian immigrant [13]; and SP resistance was well-established in Asia when chloroquine resistance first appeared in Africa [14]. If the triple-mutant did indeed arrive at this point, then it persisted in the absence of selective pressure from SP. There are several possible reasons for this persistence. First, the mutations conferring SP resistance might not have a fitness cost for the parasite. Such "no-cost" resistance mutations are seen in bacteria [15], and *in vitro* studies indicate that the enzyme coded by the triple-mutant *dhfr* of *P. falciparum* functions as well as wild-type [16]. Therefore, parasites carrying triple-mutant *dhfr* alleles may not have had reduced fitness relative to local wild-type parasites, and therefore they persisted even in the absence of SP use. Use of cotrimoxazole (discussed above) would also have helped maintain the triple-mutant allele. A second possible reason for the maintenance of mutant *dhfr* alleles in the absence of SP use is continuous immigration. That is, there may be a steady inflow of Asian parasites into the African population, and most of these immigrants would carry a mutant *dhfr*. If the level of immigration is high enough, then it could maintain the presence of mutant *dhfr* alleles even if they do incur a fitness cost for the parasite.

Unfortunately, it is not yet possible to determine whether the chloroquine-resistant *pfcrt* allele and the triple-mutant *dhfr* allele arrived in Africa at the same time, in the same parasite. This hypothesis could be supported in two ways. First, let us

assume that the mutant *pfcrt* allele arrived in East Africa when the first cases of chloroquine treatment failure were seen (1978 [12]). We could examine samples collected at this time in East Africa and look for linkage disequilibrium between the Asian chloroquine-resistant allele of *pfcrt* and the Asian triple-mutant *dhfr* allele. Because these genes are on different chromosomes, and therefore will quickly be shuffled by independent assortment, linkage disequilibrium would indicate that the mutant alleles share a very recent common ancestry. Unfortunately, finding samples from the late 1970s may be impossible. An alternative approach is more indirect. If the triple-mutant *dhfr* allele were already at a high prevalence in Southeast Asia when the mutant *pfcrt* allele immigrated to Africa, then a parasite carrying the mutant *pfcrt* allele probably carried the mutant *dhfr* allele as well. Demonstrating that at the time the mutant *pfcrt* allele arrived in Africa the triple-mutant *dhfr* allele was very common in the Southeast Asian population from which the immigrant came would support the hypothesis that both alleles arrived in Africa at the same time.

Other questions

Beyond the questions still surrounding the origin of SP-resistance in Africa, this thesis raises other questions. For one, how can we resolve whether the small amount of diversity within African parasites carrying three mutations in *dhfr* results from independent origins of the triple-mutant allele or from recombination with local, non-triple-mutant alleles? In all datasets except South Africa 1996/9a, there is some diversity of haplotypes among the parasites carrying three mutations in *dhfr*. I believe that most of these represent recombination with local parasites or mutation of microsatellite loci after the Asian haplotype arrived in Africa. However, I believe that the alternative triple-mutant haplotypes found by McCollum *et al.* (2006) represent an independent origin (albeit one that arose much more recently and has not spread). One way to settle the issue would be to examine more microsatellite markers, spanning a greater distance around *dhfr*. Recombination will appear as haplotype blocks with specific breakpoints. Mutation will appear as differences at one or a few loci within a

haplotype block. An independent origin of the *dhfr* triple-mutant would be confirmed by the presence of an extended haplotype block that shared alleles with local, non-triple parasites.

A second question is what happened to the mutant *dhps* alleles seen in the samples from Kilifi 1987-88 (Chapter 4). Sulfadoxine targets the DHPS enzyme, and mutations in *dhps* confer resistance to sulfadoxine. In Africa, mutations at codons 437 (alanine to glycine) and 540 (lysine to glutamate) are the most common [17-19]. We found these mutations in the samples collected in Kilifi in 1993-95, as other analyses have [20], but not in the 1987-88 samples. The Kilifi 1987-88 mutant *dhps* alleles were single-mutants, 436A and 540R; the Kilifi 1993-95 mutant alleles were double-mutants, 437G/540E. Because the Kilifi 1987-88 samples were cultured in sulfadoxine for 48 hours, it is possible that these single mutations arose *in vitro*, and were not actually present in the general parasite population. I think this is unlikely for several reasons. First, even if wild-type parasites died during the *in vitro* test, their DNA would still be on the slide. Second, the *in vitro* test spans only a single cell cycle so it is difficult to imagine that the culture caused any change in allele frequencies within the sample. Further, the fact that most of the *dhps* genotypes were wild-type, and that the mutant *dhps* genotypes increased in frequency between 1987 and 1988, as the *dhfr* genotypes did, indicates that the mutations detected in *dhps* represent mutations that were present in the parasite population in Kilifi in 1987-88.

Taking these *dhps* single-mutant alleles to be true, did they persist in the Kilifi parasite population? Though we found several single-mutant *dhps* alleles (436A, 540R) in the Kilifi 1987-88 dataset, we did not find them in the Kilifi 1993-95 dataset. One explanation is that these single-mutant *dhps* alleles incur a fitness cost for the parasite. If so, then in the absence of sulfadoxine use these alleles would have been lost or reduced to a frequency below our level of detection. Alternatively, the single-mutant *dhps* alleles could have been replaced by immigrant double-mutant alleles. As we examine more recent samples from Kilifi, the 436A allele appears again (data not shown). This could be a resurgence of the strain seen in 1988, an independent origin,

or due to immigration from another location. By using microsatellites near *dhps* we can determine which of these three possibilities is true. In addition, we can determine whether the double-mutant *dhps* alleles are local or immigrant. Future work in our lab will analyze microsatellites near *dhps* and hopefully shed light on these issues.

A third, rather troubling, issue raised by the findings of this thesis is the effect of drug resistance studies on the evolution of drug resistance. The data from Kilifi 1987-88 (Chapter 4) indicate that within a small cohort of children (fewer than 100), use of SP will rapidly reveal resistant alleles of *dhfr*. Because these alleles appear so quickly, we assume that they were present prior to SP, just at a level below detection by this small sample size. The concern, then, is that the study itself could have led to an increase in the prevalence of mutant *dhfr* alleles within the Kilifi population, explaining the high prevalence of mutant *dhfr* alleles in 1988 and in 1993-95. However, this explanation is unlikely. Estimating the population of children under five living near the Kilifi District Hospital to be about 30,000 in 1987 (data not shown; [21-23]), less than 1% of the children in the population were exposed to SP during the course of this study. Moreover, the adults in Kilifi probably harbor the bulk of the parasite population; however, because they have acquired immunity they are not clinically ill and are not treated. Therefore, any significant effect on the general population of parasites is unlikely. A reasonable conclusion is that rather than significantly altering the parasite population in Kilifi, treatment with SP caused this cohort of children to be infected subsequently with strains that were resistant to SP. Pyrimethamine and sulfadoxine have half-lives of 100 and 200 hours, respectively [24]; they remain in the body for weeks. If a child were exposed to malaria parasites during this time, parasites carrying mutant *dhfr* alleles would be more likely to infect successfully and produce a clinical infection. In other words, the children in this study were preferentially infected with parasites carrying mutant *dhfr* alleles. Therefore, a cross-sectional survey of *P. falciparum* in Kilifi in 1988 would probably not show nearly as high a prevalence of resistant *dhfr* alleles as I found.

Though I have argued in the preceding paragraph that small drug resistance studies are unlikely to have a large effect on the general parasite population, it is nevertheless the case that treating with SP clearly selects for mutant *dhfr* alleles within individual patients. For example, in the Kilifi 1993-95 dataset (Chapter 3) a sample taken prior to SP treatment was a mixture of triple-mutant and double-mutant parasites (sample 546), while a sample taken from the same patient after treatment with SP contained a triple-mutant strain only (sample 601). Therefore, administering SP to even a small percentage of a population will increase the prevalence of mutant *dhfr* (and *dhps*) alleles slightly. Over time, this is of course precisely how drug resistance increases in a given area: treating enough patients with SP causes the resistant parasite population to expand. But what happens if drug pressure is removed? For chloroquine, it appears that – under some circumstances at least – removal of the drug causes a decrease in prevalence of the mutant *pfcrt* allele. In Malawi in 1993, chloroquine had less than 50% clinical efficacy and was replaced with SP. Eight years later there were no mutant *pfcrt* alleles detectable in the population [25], and chloroquine is now almost 100% effective for treating malaria in Malawi [26]. Would removal of SP cause a similar decrease in mutant *dhfr* and *dhps* allele frequencies?

One way to answer this question would be to look at samples collected after the pyrimethamine prophylaxis studies (described in Chapter 4) ended. Both studies, in Tanzania and Mali, found that pyrimethamine rapidly lost efficacy, presumably due to an increase in prevalence of mutant *dhfr* alleles [11,27]. Did these mutant alleles maintain their high prevalence after pyrimethamine was removed? For how long? If these mutant alleles disappear as quickly as they appear then the early drug resistance studies likely played no role in the evolution of drug resistance. However, since the mutant alleles apparently persist in the population in the absence of selection, there is no reason why their prevalence should decrease. In Thailand, mefloquine replaced SP in 1985 [28], yet mutant *dhfr* alleles remained at high frequency a decade later [29,30]. Moreover, cotrimoxazole is now recommended for prophylaxis in people living with HIV/AIDS in Africa [31], which may continue to select for mutant *dhfr* alleles.

Therefore, I doubt that removal of SP will lead to the disappearance of mutant *dhfr* alleles, unlike chloroquine and *pfcrt*.

Future directions

As described above, several questions remain surrounding the history and evolution of SP-resistant parasites in Africa. Do triple-mutants of African origin eventually arise out of the local double-mutants? Do the mutant *dhps* alleles in Kilifi have the same haplotype as those in South Africa? Do the mutant *dhps* alleles found in Kilifi after 1995 have the same haplotype as those found in 1987-88? In our laboratory we hope to answer these questions by constructing a longitudinal dataset of samples from Kilifi, Kenya. These samples, collected from 1997-2006, will indicate how the prevalence of various alleles of *pfcrt*, *dhfr*, and *dhps* has changed over time. Studying the microsatellites flanking these genes will indicate how various mutant alleles have fared in the population over time. We may also choose to examine the gene that codes for GTP-cyclohydrolase (*PFL1155w*) in our longitudinal study. This protein lies upstream of DHFR and DHPS in the folate pathway and may be the rate-limiting enzyme of the process [32]. A recent study found that amplification of this gene correlated with sulfadoxine resistance [33], and increased copy-number of *PFL1155w* is in linkage disequilibrium with the *dhfr* 164L mutation in samples from Thailand [34]. Therefore, it is speculated that gene amplification of GTP-cyclohydrolase may play a role in SP resistance.

The longitudinal study will provide useful information about the population dynamics of antimalarial resistance, and may address some of the persistent methodological problems encountered when studying the evolution of drug resistance in malaria. For one, all population genetics studies of *P. falciparum* are hindered by the fact that we do not know recombination rates or microsatellite mutation rates *in vivo*. Furthermore, these rates likely vary extensively from place to place with the transmission intensity. By looking at our longitudinal data from Kilifi, we might be able to estimate recombination and mutation rates in this location. For example, there

is an intron within *pfcrt* that has been shown to mutate quickly [35]. By monitoring the number of alleles we see at this locus over time within chloroquine resistant (mutant *pfcrt*) samples, and making the assumptions (a) that there was only one allele for this microsatellite when the mutant *pfcrt* allele first arrived in Africa and (b) that this mutant first arrived in Kenya when chloroquine resistance was first seen in Kenya (1978) [12], we could calculate a mutation rate for this one particular microsatellite.

The longitudinal data could also generate a recombination rate for parasites in Kilifi. When the triple-mutant parasite first appeared and multiplied, presumably all of chromosome 4 (on which *dhfr* resides) was identical, because all of the triple-mutants were descended from a single common ancestor. Over time, recombination will reduce the span of the conserved region. If we genotyped microsatellites spanning a larger genetic distance around *dhfr* in the triple-mutant parasites, we could monitor how quickly recombination decreases the width of the conserved haplotype, thus giving a rough recombination rate. Calculating the mutation and recombination rates for Kilifi would be a significant step, but unfortunately these estimates will not be readily applicable to other locations.

A second methodological problem was the focus of Chapter 2: we do not yet have an efficient way to phase haplotypes from mixed infections of *P. falciparum*. The yeast method I developed certainly works, but it is inefficient. Efficiency might improve with smaller fragments (<6 kb), and current work in the laboratory addresses this possibility. Computational methods are another possibility, and one has been developed for haplotypes created from SNPs rather than microsatellites [36]. As detailed in Appendix A, however, it is not computationally feasible at this point in time to develop a similar algorithm for microsatellites. In addition, our method for detecting microsatellites (essentially PCR with fluorescent primers) is not consistent enough to detect all alleles present in a sample at all loci. In fact, PCR does not even amplify all alleles at all loci for a consistent subset of strains present in a sample, because some alleles amplify more efficiently than others (Chapter 2). Thus, if a strain is present at low level in a sample its alleles will be detected at some loci but not at others.

Advances in molecular biology techniques could help. In recent years, various groups have developed molecular methods for determining the two haplotypes within a sample of human DNA [37-40], and perhaps one of these could be adapted for use with malaria samples. One method is sequencing using polymerase colonies (polonies), in which single molecules are sequenced while fixed in an acrylamide gel [41]. It has been shown that this method can identify haplotypes from pooled samples of human DNA, i.e. DNA containing more than two haplotypes [42]. Therefore, it should be able to identify haplotypes for mixed infections of malaria. The current method is for SNPs, not microsatellites, and probably begins with high quality DNA, but it should be possible to adapt the technique to field samples of *P. falciparum*.

A final methodological hurdle is standardizing microsatellite allele calls between research groups. As demonstrated in Chapter 5, the measured length of a given allele depends on the primers used to amplify it, the capillary electrophoresis machine that measures the amplified product, and rounding the location of the fluorescent peak to the nearest base. Currently, the only way to conclude definitively that two alleles are the same is to analyze them in the same laboratory. Clearly, this solution is not practical. An alternative would be to report all allele sizes relative to the calculated (from the published sequence) allele size of a given locus in strain 3D7. For example, the length of microsatellite -0.1 in the published sequence of 3D7 is 30, or $(AT)_{15}$. I measure the allele size at that locus as 101 (Chapter 4). Therefore, I would report my allele 85 as 14 = 85 – (101 – 30), and my 105 allele as 34. This method of standardization would not, however, solve the problem that sometimes the measured repeat unit is not a whole number of bases, leading to rounding error (see Chapter 5). To solve this aspect, one would need to record (to several decimal places) the computer-called length of numerous alleles that differ by approximately two bases at a given dinucleotide microsatellite locus. The mean of the difference is the repeat unit size, and one can then report the number of repeats of each allele relative to 3D7. However, since microsatellite alleles in *P. falciparum* often undergo complex mutational events [43], this method will likely have problems as well.

Despite all these hurdles, continued study of the population genetics of drug resistance in *P. falciparum* is necessary if we are to learn how resistance emerges and spreads, and thus prevent the development of resistance to new antimalarial drugs. During the time I have studied malaria, roughly 4 million children have died from *P. falciparum* infections, and the rate is increasing due to drug resistance [44]. If new artemisinin-based therapies lose efficacy because we do not understand how to prevent the spread of drug resistance, there will be little left in the pharmacy with which to treat drug resistant malaria.

Table 6.1 – Evidence of Recombination. Microsatellite alleles (PCR product lengths) and *dhfr* genotypes for double- and triple-mutant parasites from Kilifi 1993-95. Alleles are color coded as follows: purple = triple-mutant; blue = double-mutant (51/108); grey = 89 at locus -0.1

	Upstream Microsatellites			DHFR Genotype		
Sample	-5.0	-3.8	-0.1	N51I	C59R	S108N
244	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
698	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
716	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
779	195	192	[illegible]	[illegible]	[illegible]	[illegible]
60	195	192	89	I	C	N
83	195	192	89	I	C	N
102	195	192	89	I	C	N
113	195	192	89	I	C	N
457	195	192	89	I	C	N
632	195	192	89	I	C	N
696	195	192	89	I	C	N
713	195	192	89	I	C	N

NOTES TO CHAPTER SIX

1. Roper, C. et al. Intercontinental spread of pyrimethamine-resistant malaria. *Science* **305**, 1124 (2004).

2. Noranate, N. et al. Rapid Dissemination of Plasmodium falciparum Drug Resistance Despite Strictly Controlled Antimalarial Use. *PLoS ONE* **2**, e139 (2007).

3. Ndiaye, D. et al. Defining the origin of Plasmodium falciparum resistant dhfr isolates in Senegal. *Acta Trop* **99**, 106-11 (2006).

4. Anderson, T.J. et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. *Mol Biol Evol* **17**, 1467-82 (2000).

5. Su, X. et al. A genetic map and recombination parameters of the human malaria parasite Plasmodium falciparum. *Science* **286**, 1351-3 (1999).

6. Mita, T. et al. Independent evolution of pyrimethamine resistance in Plasmodium falciparum isolates in Melanesia. *Antimicrob Agents Chemother* **51**, 1071-7 (2007).

7. Bloland, P.B. et al. Co-trimoxazole for childhood febrile illness in malaria-endemic regions. *Lancet* **337**, 518-20 (1991).

8. Salako, L.A. et al. Treatment of childhood fevers and other illnesses in three rural Nigerian communities. *J Trop Pediatr* **47**, 230-8 (2001).

9. Patrick Kachur, S. et al. Prevalence of malaria parasitemia among clients seeking treatment for fever or malaria at drug stores in rural Tanzania 2004. *Trop Med Int Health* **11**, 441-51 (2006).

10. Iyer, J.K., Milhous, W.K., Cortese, J.F., Kublin, J.G. & Plowe, C.V. Plasmodium falciparum cross-resistance between trimethoprim and pyrimethamine. *Lancet* **358**, 1066-7 (2001).

11. Clyde, D.F. Observations on monthly pyrimethamine (daraprim) prophylaxis in an East African village. *East Afr Med J* **31**, 41-6 (1954).

12. Charmot, G., Amat-Roze, J.M., Rodhain, F., Le Bras, J. & Coulaud, J.P. [Geographic approach to the epidemiology of chloroquine-resistance of Plasmodium falciparum in tropical Africa]. *Ann Soc Belg Med Trop* **71**, 187-97 (1991).

13. Ariey, F. et al. Invasion of Africa by a single pfcrt allele of South East Asian type. *Malar J* **5**, 34 (2006).

14. White, N.J. Antimalarial drug resistance: the pace quickens. *J Antimicrob Chemother* **30**, 571-85 (1992).

15. Andersson, D.I. The biological cost of mutational antibiotic resistance: any practical conclusions? *Curr Opin Microbiol* **9**, 461-5 (2006).

16. Sandefur, C.I., Wooden, J.M., Quaye, I.K., Sirawaraporn, W. & Sibley, C.H. Pyrimethamine-resistant dihydrofolate reductase enzymes of Plasmodium falciparum are not enzymatically compromised in vitro. *Mol Biochem Parasitol* (2007).

17. Kublin, J.G. et al. Molecular markers for failure of sulfadoxine-pyrimethamine and chlorproguanil-dapsone treatment of Plasmodium falciparum malaria. *J Infect Dis* **185**, 380-8 (2002).

18. Nzila, A.M. et al. 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**, 991-6 (2000).

19. Plowe, C.V. et al. Mutations in Plasmodium falciparum dihydrofolate reductase and dihydropteroate synthase and epidemiologic patterns of pyrimethamine-sulfadoxine use and resistance. *J Infect Dis* **176**, 1590-6 (1997).

20. Nzila, A.M. et al. Molecular evidence of greater selective pressure for drug resistance exerted by the long-acting antifolate Pyrimethamine/Sulfadoxine compared with the shorter-acting chlorproguanil/dapsone on Kenyan Plasmodium falciparum. *J Infect Dis* **181**, 2023-8 (2000).

21. Certain, P.R. & Certain, M.W. Personal communication: Laura's Population Problem. (2007).

22. Korenromp, E.L., Williams, B.G., Gouws, E., Dye, C. & Snow, R.W. Measurement of trends in childhood malaria mortality in Africa: an assessment of progress toward targets based on verbal autopsy. *Lancet Infect Dis* **3**, 349-58 (2003).

23. Ndiritu, M. et al. Immunization coverage and risk factors for failure to immunize within the Expanded Programme on Immunization in Kenya after introduction of new Haemophilus influenzae type b and hepatitis b virus antigens. *BMC Public Health* **6**, 132 (2006).

24. White, N.J. Clinical pharmacokinetics of antimalarial drugs. *Clin Pharmacokinet* **10**, 187-215 (1985).

25. Kublin, J.G. et al. Reemergence of chloroquine-sensitive Plasmodium falciparum malaria after cessation of chloroquine use in Malawi. *J Infect Dis* **187**, 1870-5 (2003).

26. Laufer, M.K. et al. Return of chloroquine antimalarial efficacy in Malawi. *N Engl J Med* **355**, 1959-66 (2006).

27. Doumbo, O.K. et al. Rapid selection of Plasmodium falciparum dihydrofolate reductase mutants by pyrimethamine prophylaxis. *J Infect Dis* **182**, 993-6 (2000).

28. Farooq, U. & Mahajan, R.C. Drug resistance in malaria. *J Vector Borne Dis* **41**, 45-53 (2004).

29. Nair, S. et al. A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. *Mol Biol Evol* **20**, 1526-36 (2003).

30. Biswas, S., Escalante, A., Chaiyaroj, S., Angkasekwinai, P. & Lal, A.A. Prevalence of point mutations in the dihydrofolate reductase and dihydropteroate synthetase genes of Plasmodium falciparum isolates from India and Thailand: a molecular epidemiologic study. *Trop Med Int Health* **5**, 737-43 (2000).

31. UNAIDS. Provisional WHO/UNAIDS Secretariat Recommendations on the Use of Cotrimoxazole Prophylaxis in Adults and Children Living With HIV/AIDS in Africa. (2004).

32. Hossain, T. et al. Enhancement of folates in plants through metabolic engineering. *Proc Natl Acad Sci U S A* **101**, 5158-63 (2004).

33. Kidgell, C. et al. A systematic map of genetic variation in Plasmodium falciparum. *PLoS Pathog* **2**, e57 (2006).

34. Anderson, T. Evolution of Malaria Parasites Under Strong Drug Selection. in *Pests, Plagues & Plants: Genomics & Global Health* (University of Washington, Seattle, 2007).

35. Vinayak, S., Mittra, P. & Sharma, Y.D. Wide variation in microsatellite sequences within each Pfcrt mutant haplotype. *Mol Biochem Parasitol* **147**, 101-8 (2006).

36. Takala, S.L. et al. A high-throughput method for quantifying alleles and haplotypes of the malaria vaccine candidate Plasmodium falciparum merozoite surface protein-1 19 kDa. *Malar J* **5**, 31 (2006).

37. Inbar, E., Yakir, B. & Darvasi, A. An efficient haplotyping method with DNA pools. *Nucleic Acids Res* **30**, e76 (2002).

38. Konfortov, B.A., Bankier, A.T. & Dear, P.H. An efficient method for multi-locus molecular haplotyping. *Nucleic Acids Res* **35**, e6 (2007).

39. Yu, C.E., Devlin, B., Galloway, N., Loomis, E. & Schellenberg, G.D. ADLAPH: A molecular haplotyping method based on allele-discriminating long-range PCR. *Genomics* **84**, 600-12 (2004).

40. Wu, W.M. et al. Linear allele-specific long-range amplification: a novel method of long-range molecular haplotyping. *Hum Mutat* **26**, 393-4 (2005).

41. Mitra, R.D. & Church, G.M. In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res* **27**, e34 (1999).

42. Zhang, K. et al. Long-range polony haplotyping of individual human chromosome molecules. *Nat Genet* **38**, 382-7 (2006).

43. Anderson, T.J., Su, X.Z., Roddam, A. & Day, K.P. Complex mutations in a high proportion of microsatellite loci from the protozoan parasite Plasmodium falciparum. *Mol Ecol* **9**, 1599-608 (2000).

44. Trape, J.F. The public health impact of chloroquine resistance in Africa. *Am J Trop Med Hyg* **64**, 12-7 (2001).

BIBLIOGRAPHY

Folate Biosynthetic Pathway. *Malaria Parasite Metabolic Pathways*, 2007, from http://sites.huji.ac.il/malaria/maps/folatebiopath.html

Malaria in Africa. (1998). *Roll Back Malaria Infosheets* Retrieved December 12, 2006, from http://www.rbm.who.int/cmc_upload/0/000/015/370/RBMInfosheet_3.htm

Roll Back Malaria Infosheet. (1998). Retrieved July 29, 2004, from http://www.rbm.who.int/cmc_upload/0/000/015/372/RBMInfosheet_1.htm

KEMRI-Wellcome Trust Collaborative Research Programme. (2002). Retrieved March 30, 2007, from http://www.kemri-wellcome.org/history.php

Malaria and the Red Cell. (2002). Retrieved January 5, 2007

The efficacy of antimalarial monotherapies, sulphadoxine-pyrimethamine and amodiaquine in East Africa: implications for sub-regional policy. (2003). *Trop Med Int Health, 8*(10), 860-867.

Kenya Demographic and Health Survey 2003: Preliminary Report. (2003).). Nairobi, Kenya: Central Bureau of Statistics, Ministry of Planning and National Development.

Susceptibility of Plasmodium falciparum to antimalalrial drugs: report on global monitoring: 1996-2004. (2005).). Geneva: World Health Organization.

Guidelines for the Treatment of Malaria. (No. 978 92 4 154694 2)(2006). No. 978 92 4 154694 2). Geneva: World Health Organization.

Schema of the Life Cycle of Malaria. (2006, February 17, 2006). from http://www.cdc.gov/malaria/biology/life_cycle.htm

Network 4.2.0.1 User Guide. (2007). Fluxus Technology Ltd.

Abdel-Muhsin, A. A., Mackinnon, M. J., Awadalla, P., Ali, E., Suleiman, S., Ahmed, S., et al. (2003). Local differentiation in Plasmodium falciparum drug resistance genes in Sudan. *Parasitology, 126*(Pt 5), 391-400.

Abu-Raddad, L. J., Patnaik, P., & Kublin, J. G. (2006). Dual infection with HIV and malaria fuels the spread of both diseases in sub-Saharan Africa. *Science, 314*(5805), 1603-1606.

Alifrangis, M., Dalgaard, M. B., Lusingu, J. P., Vestergaard, L. S., Staalsoe, T., Jensen, A. T., et al. (2006). Occurrence of the Southeast Asian/South American SVMNT haplotype of the chloroquine-resistance transporter gene in Plasmodium falciparum in Tanzania. *J Infect Dis, 193*(12), 1738-1741.

Alonso, P. L., Sacarlal, J., Aponte, J. J., Leach, A., Macete, E., Aide, P., et al. (2005). Duration of protection with RTS,S/AS02A malaria vaccine in prevention of

Plasmodium falciparum disease in Mozambican children: single-blind extended follow-up of a randomised controlled trial. *Lancet, 366*(9502), 2012-2018.

Alonso, P. L., Sacarlal, J., Aponte, J. J., Leach, A., Macete, E., Milman, J., et al. (2004). Efficacy of the RTS,S/AS02A vaccine against Plasmodium falciparum infection and disease in young African children: randomised controlled trial. *Lancet, 364*(9443), 1411-1420.

Amukoye, E., Winstanley, P. A., Watkins, W. M., Snow, R. W., Hatcher, J., Mosobo, M., et al. (1997). Chlorproguanil-dapsone: effective treatment for uncomplicated falciparum malaria. *Antimicrob Agents Chemother, 41*(10), 2261-2264.

Anderson, T. (2007). Evolution of Malaria Parasites Under Strong Drug Selection, *Pests, Plagues & Plants: Genomics & Global Health*. University of Washington, Seattle.

Anderson, T. J. (2004). Mapping drug resistance genes in Plasmodium falciparum by genome-wide association. *Curr Drug Targets Infect Disord, 4*(1), 65-78.

Anderson, T. J., Haubold, B., Williams, J. T., Estrada-Franco, J. G., Richardson, L., Mollinedo, R., et al. (2000). Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. *Mol Biol Evol, 17*(10), 1467-1482.

Anderson, T. J., Paul, R. E., Donnelly, C. A., & Day, K. P. (2000). Do malaria parasites mate non-randomly in the mosquito midgut? *Genet Res, 75*(3), 285-296.

Anderson, T. J., Su, X. Z., Bockarie, M., Lagog, M., & Day, K. P. (1999). Twelve microsatellite markers for characterization of Plasmodium falciparum from finger-prick blood samples. *Parasitology, 119 (Pt 2)*, 113-125.

Anderson, T. J., Su, X. Z., Roddam, A., & Day, K. P. (2000). Complex mutations in a high proportion of microsatellite loci from the protozoan parasite Plasmodium falciparum. *Mol Ecol, 9*(10), 1599-1608.

Andersson, D. I. (2006). The biological cost of mutational antibiotic resistance: any practical conclusions? *Curr Opin Microbiol, 9*(5), 461-465.

answers.com. (2007). Trimethoprim. Retrieved March 31, 2007, from http://www.answers.com/topic/trimethoprim

Ariey, F., Fandeur, T., Durand, R., Randrianarivelojosia, M., Jambou, R., Legrand, E., et al. (2006). Invasion of Africa by a single pfcrt allele of South East Asian type. *Malar J, 5*, 34.

Ariey, F., & Robert, V. (2003). The puzzling links between malaria transmission and drug resistance. *Trends Parasitol, 19*(4), 158-160.

Arnot, D. (1998). 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.

Awadalla, P., Walliker, D., Babiker, H., & Mackinnon, M. (2001). The question of Plasmodium falciparum population structure. *Trends Parasitol, 17*(8), 351-353.

Babiker, H. A., Abdel-Muhsin, A. A., Hamad, A., Mackinnon, M. J., Hill, W. G., & Walliker, D. (2000). Population dynamics of Plasmodium falciparum in an unstable malaria area of eastern Sudan. *Parasitology, 120 (Pt 2)*, 105-111.

Babiker, H. A., Lines, J., Hill, W. G., & Walliker, D. (1997). Population structure of Plasmodium falciparum in villages with different malaria endemicity in east Africa. *Am J Trop Med Hyg, 56*(2), 141-147.

Baird, J. K. (2005). Effectiveness of antimalarial drugs. *N Engl J Med, 352*(15), 1565-1577.

Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol, 16*(1), 37-48.

Bandelt, H. J., Forster, P., Sykes, B. C., & Richards, M. B. (1995). Mitochondrial portraits of human populations using median networks. *Genetics, 141*(2), 743-753.

Bates, S. J., Winstanley, P. A., Watkins, W. M., Alloueche, A., Bwika, J., Happi, T. C., et al. (2004). Rare, highly pyrimethamine-resistant alleles of the Plasmodium falciparum dihydrofolate reductase gene from 5 African sites. *J Infect Dis, 190*(10), 1783-1792.

Binks, R. H., Baum, J., Oduola, A. M., Arnot, D. E., Babiker, H. A., Kremsner, P. G., et al. (2001). Population genetic analysis of the Plasmodium falciparum erythrocyte binding antigen-175 (eba-175) gene. *Mol Biochem Parasitol, 114*(1), 63-70.

Biswas, S., Escalante, A., Chaiyaroj, S., Angkasekwinai, P., & Lal, A. A. (2000). Prevalence of point mutations in the dihydrofolate reductase and dihydropteroate synthetase genes of Plasmodium falciparum isolates from India and Thailand: a molecular epidemiologic study. *Trop Med Int Health, 5*(10), 737-743.

Bloland, P. B., Redd, S. C., Kazembe, P., Tembenu, R., Wirima, J. J., & Campbell, C. C. (1991). Co-trimoxazole for childhood febrile illness in malaria-endemic regions. *Lancet, 337*(8740), 518-520.

Bohannon, J. (2006). Arata Kochi profile. Fighting words from WHO's new malaria chief. *Science, 311*(5761), 599.

Bottius, E., Guanzirolli, A., Trape, J. F., Rogier, C., Konate, L., & Druilhe, P. (1996). Malaria: even more chronic in nature than previously thought; evidence for subpatent parasitaemia detectable by the polymerase chain reaction. *Trans R Soc Trop Med Hyg, 90*(1), 15-19.

Bredenkamp, B. L., Sharp, B. L., Mthembu, S. D., Durrheim, D. N., & Barnes, K. I. (2001). Failure of sulphadoxine-pyrimethamine in treating Plasmodium falciparum malaria in KwaZulu-Natal. *S Afr Med J, 91*(11), 970, 972.

Brooks, D. R., Wang, P., Read, M., Watkins, W. M., Sims, P. F., & Hyde, J. E. (1994). Sequence variation of the hydroxymethyldihydropterin pyrophosphokinase: dihydropteroate synthase gene in lines of the human malaria parasite, Plasmodium falciparum, with differing resistance to sulfadoxine. *Eur J Biochem, 224*(2), 397-405.

Certain, L. K., & Sibley, C. H. (2007). Plasmodium falciparum: a novel method for analyzing haplotypes in mixed infections. *Exp Parasitol, 115*(3), 233-241.

Certain, P. R., & Certain, M. W. (2007). Personal communication: Laura's Population Problem.

Charmot, G., Amat-Roze, J. M., Rodhain, F., Le Bras, J., & Coulaud, J. P. (1991). [Geographic approach to the epidemiology of chloroquine-resistance of Plasmodium falciparum in tropical Africa]. *Ann Soc Belg Med Trop, 71*(3), 187-197.

Chedraui, P. A., Daily, J., & Lockwood, C. J. (2006). Malaria in pregnancy: UpToDate Online 14.3.

Chen, N., Kyle, D. E., Pasay, C., Fowler, E. V., Baker, J., Peters, J. M., et al. (2003). pfcrt Allelic types with two novel amino acid mutations in chloroquine-resistant Plasmodium falciparum isolates from the Philippines. *Antimicrob Agents Chemother, 47*(11), 3500-3505.

Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H., & Hieter, P. (1992). Multifunctional yeast high-copy-number shuttle vectors. *Gene, 110*(1), 119-122.

Chulay, J. D., Watkins, W. M., & Sixsmith, D. G. (1984). Synergistic antimalarial activity of pyrimethamine and sulfadoxine against Plasmodium falciparum in vitro. *Am J Trop Med Hyg, 33*(3), 325-330.

Clark, A. G. (1990). Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol Biol Evol, 7*(2), 111-122.

Clyde, D. F. (1954). Observations on monthly pyrimethamine (daraprim) prophylaxis in an East African village. *East Afr Med J, 31*(2), 41-46.

Clyde, D. F., & Shute, G. T. (1954). Resistance of East African varieties of Plasmodium falciparum to pyrimethamine. *Trans R Soc Trop Med Hyg, 48*(6), 495-500.

Conway, D. J. (2003). Tracing the dawn of Plasmodium falciparum with mitochondrial genome sequences. *Trends Genet, 19*(12), 671-674.

Conway, D. J., Cavanagh, D. R., Tanabe, K., Roper, C., Mikes, Z. S., Sakihama, N., et al. (2000). A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses. *Nat Med, 6*(6), 689-692.

Conway, D. J., Fanello, C., Lloyd, J. M., Al-Joubori, B. M., Baloch, A. H., Somanath, S. D., et al. (2000). Origin of Plasmodium falciparum malaria is traced by mitochondrial DNA. *Mol Biochem Parasitol, 111*(1), 163-171.

Conway, D. J., Machado, R. L., Singh, B., Dessert, P., Mikes, Z. S., Povoa, M. M., et al. (2001). Extreme geographical fixation of variation in the Plasmodium falciparum gamete surface protein gene Pfs48/45 compared with microsatellite loci. *Mol Biochem Parasitol, 115*(2), 145-156.

Conway, D. J., & Roper, C. (2000). Micro-evolution and emergence of pathogens. *Int J Parasitol, 30*(12-13), 1423-1430.

Conway, D. J., Roper, C., Oduola, A. M., Arnot, D. E., Kremsner, P. G., Grobusch, M. P., et al. (1999). High recombination rate in natural populations of Plasmodium falciparum. *Proc Natl Acad Sci U S A, 96*(8), 4506-4511.

Cooper, G., Amos, W., Hoffman, D., & Rubinsztein, D. C. (1996). Network analysis of human Y microsatellite haplotypes. *Hum Mol Genet, 5*(11), 1759-1766.

Cooper, R. A., Hartwig, C. L., & Ferdig, M. T. (2005). pfcrt is more than the Plasmodium falciparum chloroquine resistance gene: a functional and evolutionary perspective. *Acta Trop, 94*(3), 170-180.

Cortese, J. F., Caraballo, A., Contreras, C. E., & Plowe, C. V. (2002). Origin and dissemination of Plasmodium falciparum drug-resistance mutations in South America. *J Infect Dis, 186*(7), 999-1006.

Cortese, J. F., & Plowe, C. V. (1998). Antifolate resistance due to new and known Plasmodium falciparum dihydrofolate reductase mutations expressed in yeast. *Mol Biochem Parasitol, 94*(2), 205-214.

Cowman, A. F., & Kappe, S. H. (2006). Microbiology. Malaria's stealth shuttle. *Science, 313*(5791), 1245-1246.

Cowman, A. F., Morry, M. J., Biggs, B. A., Cross, G. A., & Foote, S. J. (1988). Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of Plasmodium falciparum. *Proc Natl Acad Sci U S A, 85*(23), 9109-9113.

Currat, M., Trabuchet, G., Rees, D., Perrin, P., Harding, R. M., Clegg, J. B., et al. (2002). Molecular analysis of the beta-globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the beta(S) Senegal mutation. *Am J Hum Genet, 70*(1), 207-223.

Daily, J. P. (2006). Antimalarial drug therapy: the role of parasite biology and drug resistance. *J Clin Pharmacol, 46*(12), 1487-1497.

Davis, T. M., Karunajeewa, H. A., & Ilett, K. F. (2005). Artemisinin-based combination therapies for uncomplicated malaria. *Med J Aust, 182*(4), 181-185.

Demar, M., & Carme, B. (2004). Plasmodium falciparum in vivo resistance to quinine: description of two RIII responses in French Guiana. *Am J Trop Med Hyg, 70*(2), 125-127.

Diourte, Y., Djimde, A., Doumbo, O. K., Sagara, I., Coulibaly, Y., Dicko, A., et al. (1999). Pyrimethamine-sulfadoxine efficacy and selection for mutations in Plasmodium falciparum dihydrofolate reductase and dihydropteroate synthase in Mali. *Am J Trop Med Hyg, 60*(3), 475-478.

Djimde, A. A., Dolo, A., Ouattara, A., Diakite, S., Plowe, C. V., & Doumbo, O. K. (2004). Molecular Diagnosis of Resistance to Antimalarial Drugs during Epidemics and in War Zones. *J Infect Dis, 190*(4), 853-855.

Doumbo, O. K., Kayentao, K., Djimde, A., Cortese, J. F., Diourte, Y., Konare, A., et al. (2000). Rapid selection of Plasmodium falciparum dihydrofolate reductase mutants by pyrimethamine prophylaxis. *J Infect Dis, 182*(3), 993-996.

Druilhe, P., Daubersies, P., Patarapotikul, J., Gentil, C., Chene, L., Chongsuphajaisiddhi, T., et al. (1998). A primary malarial infection is composed of a very wide range of genetically diverse but related parasites. *J Clin Invest, 101*(9), 2008-2016.

Duffy, P. E., & Fried, M. (2005). Malaria in the pregnant woman. *Curr Top Microbiol Immunol, 295*, 169-200.

Duraisingh, M., Ferdig, M. T., Stoeckert, C. J., Volkman, S. K., & McGovern, V. P. (2006). Plasmodium research in the postgenomic era. *Trends Parasitol, 22*(1), 1-4.

Duraisingh, M. T., Curtis, J., & Warhurst, D. C. (1998). Plasmodium falciparum: detection of polymorphisms in the dihydrofolate reductase and dihydropteroate synthetase genes by PCR and restriction digestion. *Exp Parasitol, 89*(1), 1-8.

Duraisingh, M. T., Roper, C., Walliker, D., & Warhurst, D. C. (2000). Increased sensitivity to the antimalarials mefloquine and artemisinin is conferred by mutations in the pfmdr1 gene of Plasmodium falciparum. *Mol Microbiol, 36*(4), 955-961.

Durand, R., Jafari, S., Vauzelle, J., Delabre, J. F., Jesic, Z., & Le Bras, J. (2001). Analysis of pfcrt point mutations and chloroquine susceptibility in isolates of Plasmodium falciparum. *Mol Biochem Parasitol, 114*(1), 95-102.

Eberl, K. J., Jelinek, T., Aida, A. O., Peyerl-Hoffmann, G., Heuschkel, C., el Valy, A. O., et al. (2001). Prevalence of polymorphisms in the dihydrofolate reductase and dihydropteroate synthetase genes of Plasmodium falciparum isolates from southern Mauritania. *Trop Med Int Health, 6*(10), 756-760.

Enserink, M. (2006). Genetics. Parasite-resistant mosquitoes: a natural weapon against malaria? *Science, 312*(5773), 514.

Escalante, A. A., & Ayala, F. J. (1994). Phylogeny of the malarial genus Plasmodium, derived from rRNA gene sequences. *Proc Natl Acad Sci U S A, 91*(24), 11373-11377.

Excoffier, L., & Slatkin, M. (1995). Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol, 12*(5), 921-927.

Fangman, W.

Farooq, U., & Mahajan, R. C. (2004). Drug resistance in malaria. *J Vector Borne Dis, 41*(3-4), 45-53.

Feng, X., Carlton, J. M., Joy, D. A., Mu, J., Furuya, T., Suh, B. B., et al. (2003). Single-nucleotide polymorphisms and genome diversity in Plasmodium vivax. *Proc Natl Acad Sci U S A, 100*(14), 8502-8507.

Fenwick, A. (2006). Waterborne infectious diseases--could they be consigned to history? *Science, 313*(5790), 1077-1081.

Ferdig, M. T., & Su, X. Z. (2000). Microsatellite markers and genetic mapping in Plasmodium falciparum. *Parasitol Today, 16*(7), 307-312.

Ferlan, J. T., Mookherjee, S., Okezie, I. N., Fulgence, L., & Sibley, C. H. (2001). Mutagenesis of dihydrofolate reductase from Plasmodium falciparum: analysis in Saccharomyces cerevisiae of triple mutant alleles resistant to pyrimethamine or WR99210. *Mol Biochem Parasitol, 113*(1), 139-150.

Fernandes, N., Figueiredo, P., do Rosario, V. E., & Cravo, P. (2007). Analysis of sulphadoxine/pyrimethamine resistance-conferring mutations of Plasmodium falciparum from Mozambique reveals the absence of the dihydrofolate reductase 164L mutant. *Malar J, 6*(1), 35.

Ferreira, M. U., Nair, S., Hyunh, T. V., Kawamoto, F., & Anderson, T. J. (2002). Microsatellite characterization of Plasmodium falciparum from cerebral and uncomplicated malaria patients in southern Vietnam. *J Clin Microbiol, 40*(5), 1854-1857.

Fidock, D. A., Nomura, T., Talley, A. K., Cooper, R. A., Dzekunov, S. M., Ferdig, M. T., et al. (2000). Mutations in the P. falciparum digestive vacuole

transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol Cell, 6*(4), 861-871.

Flint, J., Harding, R. M., Boyce, A. J., & Clegg, J. B. (1998). The population genetics of the haemoglobinopathies. *Baillieres Clin Haematol, 11*(1), 1-51.

Fontenille, D., & Simard, F. (2004). Unravelling complexities in human malaria transmission dynamics in Africa through a comprehensive knowledge of vector populations. *Comp Immunol Microbiol Infect Dis, 27*(5), 357-375.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., et al. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature, 419*(6906), 498-511.

Garner, P., & Gulmezoglu, A. M. (2006). Drugs for preventing malaria in pregnant women. *Cochrane Database Syst Rev*(4), CD000169.

Gietz, D., St Jean, A., Woods, R. A., & Schiestl, R. H. (1992). Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res, 20*(6), 1425.

Gietz, R. D., Schiestl, R. H., Willems, A. R., & Woods, R. A. (1995). Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast, 11*(4), 355-360.

Giha, H. A., Staalsoe, T., Dodoo, D., Elhassan, I. M., Roper, C., Satti, G. M., et al. (1999). Overlapping antigenic repertoires of variant antigens expressed on the surface of erythrocytes infected by Plasmodium falciparum. *Parasitology, 119 (Pt 1)*, 7-17.

Giha, H. A., Staalsoe, T., Dodoo, D., Elhassan, I. M., Roper, C., Satti, G. M., et al. (1999). Nine-year longitudinal study of antibodies to variant antigens on the surface of Plasmodium falciparum-infected erythrocytes. *Infect Immun, 67*(8), 4092-4098.

Giha, H. A., Staalsoe, T., Dodoo, D., Roper, C., Satti, G. M., Arnot, D. E., et al. (2000). Antibodies to variable Plasmodium falciparum-infected erythrocyte surface antigens are associated with protection from novel malaria infections. *Immunol Lett, 71*(2), 117-126.

Greenwood, B. M., Bhasin, A., Bowler, C. H., Naylor, H., & Targett, G. A. (2006). Capacity strengthening in malaria research: the Gates Malaria Partnership. *Trends Parasitol, 22*(7), 278-284.

Gregson, A., & Plowe, C. V. (2005). Mechanisms of resistance of malaria parasites to antifolates. *Pharmacol Rev, 57*(1), 117-145.

Guerra, C. A., Snow, R. W., & Hay, S. I. (2006). Mapping the global extent of malaria in 2005. *Trends Parasitol, 22*(8), 353-358.

Hagos, B., Khan, B., Ofulla, A. V., Kariuki, D., & Martin, S. K. (1993). Response of falciparum malaria to chloroquine and three second line antimalarial drugs in a Kenyan coastal school age population. *East Afr Med J, 70*(10), 620-623.

Hankins, E. G., Warhurst, D. C., & Sibley, C. H. (2001). Novel alleles of the Plasmodium falciparum dhfr highly resistant to pyrimethamine and chlorcycloguanil, but not WR99210. *Mol Biochem Parasitol, 117*(1), 91-102.

Hapuarachchi, H. C., Dayanath, M. Y., Bandara, K. B., Abeysundara, S., Abeyewickreme, W., de Silva, N. R., et al. (2006). Point mutations in the dihydrofolate reductase and dihydropteroate synthase genes of Plasmodium falciparum and resistance to sulfadoxine-pyrimethamine in Sri Lanka. *Am J Trop Med Hyg, 74*(2), 198-204.

Hartl, D. L. (2000). *A Primer of Population Genetics* (Third ed.). Sunderland, MA: Sinauer Associates, Inc.

Hartl, D. L. (2004). The origin of malaria: mixed messages from genetic diversity. *Nat Rev Microbiol, 2*(1), 15-22.

Hartl, D. L., Volkman, S. K., Nielsen, K. M., Barry, A. E., Day, K. P., Wirth, D. F., et al. (2002). The paradoxical population genetics of Plasmodium falciparum. *Trends Parasitol, 18*(6), 266-272.

Hastings, I. M. (1997). A model for the origins and spread of drug-resistant malaria. *Parasitology, 115 (Pt 2)*, 133-141.

Hastings, I. M. (2003). Malaria control and the evolution of drug resistance: an intriguing link. *Trends Parasitol, 19*(2), 70-73.

Hastings, I. M. (2004). The origins of antimalarial drug resistance. *Trends Parasitol, 20*(11), 512-518.

Hastings, I. M. (2007). MalHaplotypeFreq: Liverpool School of Tropical Medicine.

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

Hastings, I. M., & Mackinnon, M. J. (1998). The emergence of drug-resistant malaria. *Parasitology, 117 (Pt 5)*, 411-417.

Hastings, M. D., Bates, S. J., Blackstone, E. A., Monks, S. M., Mutabingwa, T. K., & Sibley, C. H. (2002). Highly pyrimethamine-resistant alleles of dihydrofolate reductase in isolates of Plasmodium falciparum from Tanzania. *Trans R Soc Trop Med Hyg, 96*(6), 674-676.

Hill, K., Bicego, G., & Mahy, M. *Childhood Mortality in Kenya: An examination of trends and determinants in the late 1980s to mid 1990s*: Johns Hopkins School of Public Health.

Hill, W. G., & Babiker, H. A. (1995). Estimation of numbers of malaria clones in blood samples. *Proc R Soc Lond B Biol Sci, 262*(1365), 249-257.

Hill, W. G., Babiker, H. A., Ranford-Cartwright, L. C., & Walliker, D. (1995). Estimation of inbreeding coefficients from genotypic data on multiple alleles, and application to estimation of clonality in malaria parasites. *Genet Res, 65*(1), 53-61.

Hinterberg, K., Mattei, D., Wellems, T. E., & Scherf, A. (1994). Interchromosomal exchange of a large subtelomeric segment in a Plasmodium falciparum cross. *Embo J, 13*(17), 4174-4180.

Hossain, T., Rosenberg, I., Selhub, J., Kishore, G., Beachy, R., & Schubert, K. (2004). Enhancement of folates in plants through metabolic engineering. *Proc Natl Acad Sci U S A, 101*(14), 5158-5163.

Hougard, J. M., Fontenille, D., Chandre, F., Darriet, F., Carnevale, P., & Guillet, P. (2002). Combating malaria vectors in Africa: current directions of research. *Trends Parasitol, 18*(7), 283-286.

Hua, S. B., Qiu, M., Chan, E., Zhu, L., & Luo, Y. (1997). Minimum length of sequence homology required for in vivo cloning by homologous recombination in yeast. *Plasmid, 38*(2), 91-96.

Hughes, A. L., & Verra, F. (2001). Very large long-term effective population size in the virulent human malaria parasite Plasmodium falciparum. *Proc Biol Sci, 268*(1478), 1855-1860.

Hughes, A. L., & Verra, F. (2002). Extensive polymorphism and ancient origin of Plasmodium falciparum. *Trends Parasitol, 18*(8), 348-351.

Hume, J. C., Lyons, E. J., & Day, K. P. (2003). Human migration, mosquitoes and the evolution of Plasmodium falciparum. *Trends Parasitol, 19*(3), 144-149.

Humphreys, G. S., Merinopoulos, I., Ahmed, J., Whitty, C. J., Mutabingwa, T. K., Sutherland, C. J., et al. (2007). Amodiaquine and Artemether-Lumefantrine Select Distinct Alleles of the Plasmodium falciparum mdr1 Gene in Tanzanian Children Treated for Uncomplicated Malaria. *Antimicrob Agents Chemother, 51*(3), 991-997.

Hunt, S. Y., Rezvani, B. B., & Sibley, C. H. (2005). Novel alleles of Plasmodium falciparum dhfr that confer resistance to chlorcycloguanil. *Mol Biochem Parasitol, 139*(1), 25-32.

Inbar, E., Yakir, B., & Darvasi, A. (2002). An efficient haplotyping method with DNA pools. *Nucleic Acids Res, 30*(15), e76.

Iyer, J. K., Milhous, W. K., Cortese, J. F., Kublin, J. G., & Plowe, C. V. (2001). Plasmodium falciparum cross-resistance between trimethoprim and pyrimethamine. *Lancet, 358*(9287), 1066-1067.

Jansen, G., Wu, C., Schade, B., Thomas, D. Y., & Whiteway, M. (2005). Drag&Drop cloning in yeast. *Gene, 344*, 43-51.

Jeffares, D. C., Pain, A., Berry, A., Cox, A. V., Stalker, J., Ingle, C. E., et al. (2006). Genome variation and evolution of the malaria parasite Plasmodium falciparum. *Nat Genet.*

Jeffares, D. C., Pain, A., Berry, A., Cox, A. V., Stalker, J., Ingle, C. E., et al. (2007). Genome variation and evolution of the malaria parasite Plasmodium falciparum. *Nat Genet, 39*(1), 120-125.

Jelinek, T., Kilian, A. H., Curtis, J., Duraisingh, M. T., Kabagambe, G., von Sonnenburg, F., et al. (1999). Plasmodium falciparum: selection of serine 108 of dihydrofolate reductase during treatment of uncomplicated malaria with co-trimoxazole in Ugandan children. *Am J Trop Med Hyg, 61*(1), 125-130.

Jones, S. A. (1954). Resistance of P. falciparum and P. malariae to pyrimethamine (daraprim) following mass treatment with this drug; a preliminary note. *East Afr Med J, 31*(2), 47-49.

Joy, D. A., Feng, X., Mu, J., Furuya, T., Chotivanich, K., Krettli, A. U., et al. (2003). Early origin and recent expansion of Plasmodium falciparum. *Science, 300*(5617), 318-321.

Khalil, I., Ronn, A. M., Alifrangis, M., Gabar, H. A., Satti, G. M., & Bygbjerg, I. C. (2003). Dihydrofolate reductase and dihydropteroate synthase genotypes associated with in vitro resistance of Plasmodium falciparum to pyrimethamine, trimethoprim, sulfadoxine, and sulfamethoxazole. *Am J Trop Med Hyg, 68*(5), 586-589.

Kidgell, C., Volkman, S. K., Daily, J., Borevitz, J. O., Plouffe, D., Zhou, Y., et al. (2006). A systematic map of genetic variation in Plasmodium falciparum. *PLoS Pathog, 2*(6), e57.

Klayman, D. L. (1985). Qinghaosu (artemisinin): an antimalarial drug from China. *Science, 228*(4703), 1049-1055.

Kleinschmidt, I., Sharp, B., Benavente, L. E., Schwabe, C., Torrez, M., Kuklinski, J., et al. (2006). Reduction in infection with Plasmodium falciparum one year after the introduction of malaria control interventions on Bioko Island, Equatorial Guinea. *Am J Trop Med Hyg, 74*(6), 972-978.

Koella, J. C., & Antia, R. (2003). Epidemiological models for the spread of anti-malarial resistance. *Malar J, 2*(1), 3.

Konfortov, B. A., Bankier, A. T., & Dear, P. H. (2007). An efficient method for multi-locus molecular haplotyping. *Nucleic Acids Res, 35*(1), e6.

Korenromp, E. L., Williams, B. G., Gouws, E., Dye, C., & Snow, R. W. (2003). 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.

Kublin, J. G., Cortese, J. F., Njunju, E. M., Mukadam, R. A., Wirima, J. J., Kazembe, P. N., et al. (2003). Reemergence of chloroquine-sensitive Plasmodium falciparum malaria after cessation of chloroquine use in Malawi. *J Infect Dis, 187*(12), 1870-1875.

Kublin, J. G., Dzinjalamala, F. K., Kamwendo, D. D., Malkin, E. M., Cortese, J. F., Martino, L. M., et al. (2002). Molecular markers for failure of sulfadoxine-pyrimethamine and chlorproguanil-dapsone treatment of Plasmodium falciparum malaria. *J Infect Dis, 185*(3), 380-388.

Kublin, J. G., & Steketee, R. W. (2006). HIV infection and malaria--understanding the interactions. *J Infect Dis, 193*(1), 1-3.

Kyes, S., Harding, R., Black, G., Craig, A., Peshu, N., Newbold, C., et al. (1997). Limited spatial clustering of individual Plasmodium falciparum alleles in field isolates from coastal Kenya. *Am J Trop Med Hyg, 57*(2), 205-215.

Laufer, M. K., Thesing, P. C., Eddington, N. D., Masonga, R., Dzinjalamala, F. K., Takala, S. L., et al. (2006). Return of chloroquine antimalarial efficacy in Malawi. *N Engl J Med, 355*(19), 1959-1966.

Leder, K., & Weller, P. F. (2006). Epidemiology, pathogenesis, clinical features, and diagnosis of malaria: UpToDate 14.3.

Lin, S., Cutler, D. J., Zwick, M. E., & Chakravarti, A. (2002). Haplotype inference in random population samples. *Am J Hum Genet, 71*(5), 1129-1137.

Lindblade, K. A., Eisele, T. P., Gimnig, J. E., Alaii, J. A., Odhiambo, F., ter Kuile, F. O., et al. (2004). Sustainability of reductions in malaria transmission and infant mortality in western Kenya with use of insecticide-treated bednets: 4 to 6 years of follow-up. *Jama, 291*(21), 2571-2580.

Mackinnon, M. J., & Hastings, I. M. (1998). The evolution of multiple drug resistance in malaria parasites. *Trans R Soc Trop Med Hyg, 92*(2), 188-195.

Magesa, S. M., Wilkes, T. J., Mnzava, A. E., Njunwa, K. J., Myamba, J., Kivuyo, M. D., et al. (1991). Trial of pyrethroid impregnated bednets in an area of Tanzania holoendemic for malaria. Part 2. Effects on the malaria vector population. *Acta Trop, 49*(2), 97-108.

Malamba, S. S., Mermin, J., Reingold, A., Lule, J. R., Downing, R., Ransom, R., et al. (2006). Effect of cotrimoxazole prophylaxis taken by human immunodeficiency virus (HIV)-infected persons on the selection of sulfadoxine-pyrimethamine-resistant malaria parasites among HIV-uninfected household members. *Am J Trop Med Hyg, 75*(3), 375-380.

MARA/ARMA (Cartographer). (2001). *Mali: Distribution of Endemic Malaria*

Marsh, K., & Kinyanjui, S. (2006). Immune effector mechanisms in malaria. *Parasite Immunol, 28*(1-2), 51-60.

Matuschewski, K. (2006). Vaccine development against malaria. *Curr Opin Immunol, 18*(4), 449-457.

Maxwell, C. A., Rwegoshora, R. T., Magesa, S. M., & Curtis, C. F. (2006). Comparison of coverage with insecticide-treated nets in a Tanzanian town and villages where nets and insecticide are either marketed or provided free of charge. *Malar J, 5*, 44.

Mberu, E. K., Mosobo, M. K., Nzila, A. M., Kokwaro, G. O., Sibley, C. H., & Watkins, W. M. (2000). The changing in vitro susceptibility pattern to pyrimethamine/sulfadoxine in Plasmodium falciparum field isolates from Kilifi, Kenya. *Am J Trop Med Hyg, 62*(3), 396-401.

Mbogo, C. N., Snow, R. W., Khamala, C. P., Kabiru, E. W., Ouma, J. H., Githure, J. I., et al. (1995). Relationships between Plasmodium falciparum transmission by vector populations and the incidence of severe disease at nine sites on the Kenyan coast. *Am J Trop Med Hyg, 52*(3), 201-206.

McCollum, A. M., Mueller, K., Villegas, L., Udhayakumar, V., & Escalante, A. A. (2007). Common origin and fixation of Plasmodium falciparum dhfr and dhps mutations associated with sulphadoxine-pyrimethamine resistance in a low transmission area in South America. *Antimicrob Agents Chemother.*

McCollum, A. M., Poe, A. C., Hamel, M., Huber, C., Zhou, Z., Shi, Y. P., et al. (2006). Antifolate Resistance in Plasmodium falciparum: Multiple Origins and Identification of Novel dhfr Alleles. *J Infect Dis, 194*(2), 189-197.

McNeil Jr., D. G. (2007, March 1, 2007). Low Cost Antimalaria Pill Available. *The New York Times.*

Mehlotra, R. K., Fujioka, H., Roepe, P. D., Janneh, O., Ursos, L. M., Jacobs-Lorena, V., et al. (2001). Evolution of a unique Plasmodium falciparum chloroquine-resistance phenotype in association with pfcrt polymorphism in Papua New Guinea and South America. *Proc Natl Acad Sci U S A, 98*(22), 12689-12694.

Mharakurwa, S. (2004). Plasmodium falciparum transmission rate and selection for drug resistance: a vexed association or a key to successful control? *Int J Parasitol, 34*(13-14), 1483-1487.

Mita, T., Tanabe, K., Takahashi, N., Tsukahara, T., Eto, H., Dysoley, L., et al. (2007). Independent evolution of pyrimethamine resistance in Plasmodium falciparum isolates in Melanesia. *Antimicrob Agents Chemother, 51*(3), 1071-1077.

Mitra, R. D., & Church, G. M. (1999). In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res, 27*(24), e34.

MMV. (2005). Selection of Partners for Antimalarial Drug Combinations, *Strategic Consultation*. Geneva, Switzerland: Medicines for Malaria Venture.

Molyneux, D. H., Floyd, K., Barnish, G., & Fevre, E. M. (1999). Transmission control and drug resistance in malaria: a crucial interaction. *Parasitol Today, 15*(6), 238-240.

Mountain, J. L., Knight, A., Jobin, M., Gignoux, C., Miller, A., Lin, A. A., et al. (2002). SNPSTRs: empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. *Genome Res, 12*(11), 1766-1772.

Mu, J., Awadalla, P., Duan, J., McGee, K. M., Keebler, J., Seydel, K., et al. (2006). Genome-wide variation and identification of vaccine targets in the Plasmodium falciparum genome. *Nat Genet.*

Mu, J., Awadalla, P., Duan, J., McGee, K. M., Keebler, J., Seydel, K., et al. (2007). Genome-wide variation and identification of vaccine targets in the Plasmodium falciparum genome. *Nat Genet, 39*(1), 126-130.

Mu, J., Duan, J., Makova, K. D., Joy, D. A., Huynh, C. Q., Branch, O. H., et al. (2002). Chromosome-wide SNPs reveal an ancient origin for Plasmodium falciparum. *Nature, 418*(6895), 323-326.

Murray, P. R., Rosenthal, K. S., Kobayashi, G. S., & Pfaller, M. A. (2002). *Medical Microbiology* (4 ed.). St. Louis, MO: Mosby.

Mutabingwa, T., Nzila, A., Mberu, E., Nduati, E., Winstanley, P., Hills, E., et al. (2001). Chlorproguanil-dapsone for treatment of drug-resistant falciparum malaria in Tanzania. *Lancet, 358*(9289), 1218-1223.

Mutabingwa, T. K. (2005). Artemisinin-based combination therapies (ACTs): best hope for malaria treatment but inaccessible to the needy! *Acta Trop, 95*(3), 305-315.

Mwangi, J. M., Omar, S. A., & Ranford-Cartwright, L. C. (2006). Comparison of microsatellite and antigen-coding loci for differentiating recrudescing Plasmodium falciparum infections from reinfections in Kenya. *Int J Parasitol, 36*(3), 329-336.

Nair, S., Brockman, A., Paiphun, L., Nosten, F., & Anderson, T. J. (2002). Rapid genotyping of loci involved in antifolate drug resistance in Plasmodium falciparum by primer extension. *Int J Parasitol, 32*(7), 852-858.

Nair, S., Nash, D., Sudimack, D., Jaidee, A., Barends, M., Uhlemann, A. C., et al. (2006). Recurrent Gene Amplification and Soft Selective Sweeps During Evolution of Multidrug Resistance in Malaria Parasites. *Mol Biol Evol.*

Nair, S., Williams, J. T., Brockman, A., Paiphun, L., Mayxay, M., Newton, P. N., et al. (2003). A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. *Mol Biol Evol, 20*(9), 1526-1536.

Ndiaye, D., Daily, J. P., Sarr, O., Ndir, O., Gaye, O., Mboup, S., et al. (2006). Defining the origin of Plasmodium falciparum resistant dhfr isolates in Senegal. *Acta Trop, 99*(1), 106-111.

Ndiaye, D., Daily, J. P., Sarr, O., Ndir, O., Gaye, O., Mboup, S., et al. (2005). Mutations in Plasmodium falciparum dihydrofolate reductase and dihydropteroate synthase genes in Senegal. *Trop Med Int Health, 10*(11), 1176-1179.

Ndiritu, M., Cowgill, K. D., Ismail, A., Chiphatsi, S., Kamau, T., Fegan, G., et al. (2006). Immunization coverage and risk factors for failure to immunize within the Expanded Programme on Immunization in Kenya after introduction of new Haemophilus influenzae type b and hepatitis b virus antigens. *BMC Public Health, 6*, 132.

Newton, C. R., Winstanley, P. A., Watkins, W. M., Mwangi, I. N., Waruiru, C. M., Mberu, E. K., et al. (1993). A single dose of intramuscular sulfadoxine-pyrimethamine as an adjunct to quinine in the treatment of severe malaria: pharmacokinetics and efficacy. *Trans R Soc Trop Med Hyg, 87*(2), 207-210.

Niu, T., Qin, Z. S., Xu, X., & Liu, J. S. (2002). Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. *Am J Hum Genet, 70*(1), 157-169.

Noranate, N., Durand, R., Tall, A., Marrama, L., Spiegel, A., Sokhna, C., et al. (2007). Rapid Dissemination of Plasmodium falciparum Drug Resistance Despite Strictly Controlled Antimalarial Use. *PLoS ONE, 2*, e139.

Nosten, F., ter Kuile, F., Chongsuphajaisiddhi, T., Luxemburger, C., Webster, H. K., Edstein, M., et al. (1991). Mefloquine-resistant falciparum malaria on the Thai-Burmese border. *Lancet, 337*(8750), 1140-1143.

Nyamwange, C. I., & Nyamogoba, H. (2002). Molecular markers in epidemiological monitoring of the spread of resistance to antimalarials: a review. *East Afr Med J, 79*(9), 480-484.

Nyren, P. (2006). The history of pyrosequencing(r). *Methods Mol Biol, 373*, 1-14.

Nzila, A., Ochong, E., Nduati, E., Gilbert, K., Winstanley, P., Ward, S., et al. (2005). Why has the dihydrofolate reductase 164 mutation not consistently been found in Africa yet? *Trans R Soc Trop Med Hyg, 99*(5), 341-346.

Nzila, A. M., Mberu, E. K., Sulo, J., Dayo, H., Winstanley, P. A., Sibley, C. H., et al. (2000). 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.

Nzila, A. M., Nduati, E., Mberu, E. K., Hopkins Sibley, C., Monks, S. A., Winstanley, P. A., et al. (2000). Molecular evidence of greater selective pressure for drug resistance exerted by the long-acting antifolate Pyrimethamine/Sulfadoxine compared with the shorter-acting chlorproguanil/dapsone on Kenyan Plasmodium falciparum. *J Infect Dis, 181*(6), 2023-2028.

Nzila-Mounda, A., Mberu, E. K., Sibley, C. H., Plowe, C. V., Winstanley, P. A., & Watkins, W. M. (1998). Kenyan Plasmodium falciparum field isolates: correlation between pyrimethamine and chlorcycloguanil activity in vitro and point mutations in the dihydrofolate reductase domain. *Antimicrob Agents Chemother, 42*(1), 164-169.

Ofosu-Okyere, A., Mackinnon, M. J., Sowa, M. P., Koram, K. A., Nkrumah, F., Osei, Y. D., et al. (2001). 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.

Oldenburg, K. R., Vo, K. T., Michaelis, S., & Paddon, C. (1997). Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Res, 25*(2), 451-452.

Paget-McNicol, S., & Saul, A. (2001). Mutation rates in the dihydrofolate reductase gene of Plasmodium falciparum. *Parasitology, 122*(Pt 5), 497-505.

Pasvol, G., Newton, C. R., Winstanley, P. A., Watkins, W. M., Peshu, N. M., Were, J. B., et al. (1991). Quinine treatment of severe falciparum malaria in African children: a randomized comparison of three regimens. *Am J Trop Med Hyg, 45*(6), 702-713.

Patrick Kachur, S., Schulden, J., Goodman, C. A., Kassala, H., Elling, B. F., Khatib, R. A., et al. (2006). Prevalence of malaria parasitemia among clients seeking treatment for fever or malaria at drug stores in rural Tanzania 2004. *Trop Med Int Health, 11*(4), 441-451.

Payne, D. (1987). Spread of chloroquine resistance in Plasmodium falciparum. *Parasitol Today, 3*(8), 241-246.

Pearce, R., Malisa, A., Kachur, S. P., Barnes, K., Sharp, B., & Roper, C. (2005). Reduced variation around drug-resistant dhfr alleles in African Plasmodium falciparum. *Mol Biol Evol, 22*(9), 1834-1844.

Pearce, R. J., Drakeley, C., Chandramohan, D., Mosha, F., & Roper, C. (2003). 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.

Plowe, C. V., Cortese, J. F., Djimde, A., Nwanyanwu, O. C., Watkins, W. M., Winstanley, P. A., et al. (1997). 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.

Plowe, C. V., Cortese, J. F., Djimde, A., Nwanyanwu, O. C., Watkins, W. M., Winstanley, P. A., et al. (1997). 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.

Price, R. N., Uhlemann, A. C., Brockman, A., McGready, R., Ashley, E., Phaipun, L., et al. (2004). Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. *Lancet, 364*(9432), 438-447.

Price, R. N., Uhlemann, A. C., van Vugt, M., Brockman, A., Hutagalung, R., Nair, S., et al. (2006). Molecular and pharmacological determinants of the therapeutic response to artemether-lumefantrine in multidrug-resistant Plasmodium falciparum malaria. *Clin Infect Dis, 42*(11), 1570-1577.

Qin, Z. S., Niu, T., & Liu, J. S. (2002). Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. *Am J Hum Genet, 71*(5), 1242-1247.

Ranford-Cartwright, L. C., Balfe, P., Carter, R., & Walliker, D. (1993). Frequency of cross-fertilization in the human malaria parasite Plasmodium falciparum. *Parasitology, 107 (Pt 1)*, 11-18.

Raymond, C. K., Sims, E. H., & Olson, M. V. (2002). Linker-mediated recombinational subcloning of large DNA fragments using yeast. *Genome Res, 12*(1), 190-197.

Razakandrainibe, F. G., Durand, P., Koella, J. C., De Meeus, T., Rousset, F., Ayala, F. J., et al. (2005). "Clonal" population structure of the malaria agent Plasmodium falciparum in high-infection regions. *Proc Natl Acad Sci U S A, 102*(48), 17388-17393.

Reeder, J. C., Rieckmann, K. H., Genton, B., Lorry, K., Wines, B., & Cowman, A. F. (1996). Point mutations in the dihydrofolate reductase and dihydropteroate synthetase genes and in vitro susceptibility to pyrimethamine and cycloguanil of Plasmodium falciparum isolates from Papua New Guinea. *Am J Trop Med Hyg, 55*(2), 209-213.

Rich, S. M., Licht, M. C., Hudson, R. R., & Ayala, F. J. (1998). Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of Plasmodium falciparum. *Proc Natl Acad Sci U S A, 95*(8), 4425-4430.

Riehle, M. M., Markianos, K., Niare, O., Xu, J., Li, J., Toure, A. M., et al. (2006). Natural malaria infection in Anopheles gambiae is regulated by a single genomic control region. *Science, 312*(5773), 577-579.

Rocco, F. (2003). *The Miraculous Fever-Tree*. New York: HarperCollins Publishers Inc.

Roper, C., Pearce, R., Bredenkamp, B., Gumede, J., Drakeley, C., Mosha, F., et al. (2003). Antifolate antimalarial resistance in southeast Africa: a population-based analysis. *Lancet, 361*(9364), 1174-1181.

Roper, C., Pearce, R., Nair, S., Sharp, B., Nosten, F., & Anderson, T. (2004). Intercontinental spread of pyrimethamine-resistant malaria. *Science, 305*(5687), 1124.

Roper, C., Richardson, W., Elhassan, I. M., Giha, H., Hviid, L., Satti, G. M., et al. (1998). Seasonal changes in the Plasmodium falciparum population in individuals and their relationship to clinical malaria: a longitudinal study in a Sudanese village. *Parasitology, 116 (Pt 6)*, 501-510.

Rosner, B. (2000). *Fundamentals of Biostatistics* (5th ed.). Pacific Grove, CA: Duxbury.

Rukaria, R. M., Ojwang, S. B., Oyieke, J. B., & Kigondu, C. B. (1992). In vivo and in vitro response of Plasmodium falciparum to chloroquine in pregnant women in Kilifi district, Kenya. *East Afr Med J, 69*(6), 306-310.

Sachs, J. D., & Hotez, P. J. (2006). Fighting tropical diseases. *Science, 311*(5767), 1521.

Sakihama, N., Mitamura, T., Kaneko, A., Horii, T., & Tanabe, K. (2001). Long PCR amplification of Plasmodium falciparum DNA extracted from filter paper blots. *Exp Parasitol, 97*(1), 50-54.

Salako, L. A., Brieger, W. R., Afolabi, B. M., Umeh, R. E., Agomo, P. U., Asa, S., et al. (2001). Treatment of childhood fevers and other illnesses in three rural Nigerian communities. *J Trop Pediatr, 47*(4), 230-238.

Sandefur, C. I., Wooden, J. M., Quaye, I. K., Sirawaraporn, W., & Sibley, C. H. (2007). Pyrimethamine-resistant dihydrofolate reductase enzymes of Plasmodium falciparum are not enzymatically compromised in vitro. *Mol Biochem Parasitol.*

Schellenberg, D., Menendez, C., Kahigwa, E., Aponte, J., Vidal, J., Tanner, M., et al. (2001). Intermittent treatment for malaria and anaemia control at time of

routine vaccinations in Tanzanian infants: a randomised, placebo-controlled trial. *Lancet, 357*(9267), 1471-1477.

Schiestl, R. H., & Gietz, R. D. (1989). High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr Genet, 16*(5-6), 339-346.

Schmidt, K. F. (1995). Malaria research. Inbred parasites may spur resistance. *Science, 269*(5231), 1670.

Schofield, C. J., & Dujardin, J. P. (1999). Malaria: growth and selection. *Parasitol Today, 15*(11), 466.

Sestak, K. (2006). Loss of grants hurts the vulnerable. *Science, 311*(5758), 176-177.

Sia, E. A., Butler, C. A., Dominska, M., Greenwell, P., Fox, T. D., & Petes, T. D. (2000). Analysis of microsatellite mutations in the mitochondrial DNA of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A, 97*(1), 250-255.

Sia, E. A., Kokoska, R. J., Dominska, M., Greenwell, P., & Petes, T. D. (1997). Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Mol Cell Biol, 17*(5), 2851-2858.

Sibley, C. H., Hyde, J. E., Sims, P. F., Plowe, C. V., Kublin, J. G., Mberu, E. K., et al. (2001). Pyrimethamine-sulfadoxine resistance in Plasmodium falciparum: what next? *Trends Parasitol, 17*(12), 582-588.

Sidhu, A. B., Uhlemann, A. C., Valderramos, S. G., Valderramos, J. C., Krishna, S., & Fidock, D. A. (2006). Decreasing pfmdr1 copy number in plasmodium falciparum malaria heightens susceptibility to mefloquine, lumefantrine, halofantrine, quinine, and artemisinin. *J Infect Dis, 194*(4), 528-535.

Sikorski, R. S., & Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics, 122*(1), 19-27.

Sirawaraporn, W., Sathitkul, T., Sirawaraporn, R., Yuthavong, Y., & Santi, D. V. (1997). Antifolate-resistant mutants of Plasmodium falciparum dihydrofolate reductase. *Proc Natl Acad Sci U S A, 94*(4), 1124-1129.

Snewin, V. A., England, S. M., Sims, P. F., & Hyde, J. E. (1989). Characterisation of the dihydrofolate reductase-thymidylate synthetase gene from human malaria parasites highly resistant to pyrimethamine. *Gene, 76*(1), 41-52.

Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y., & Hay, S. I. (2005). The global distribution of clinical episodes of Plasmodium falciparum malaria. *Nature, 434*(7030), 214-217.

Snow, R. W., Peshu, N., Forster, D., Mwenesi, H., & Marsh, K. (1992). The role of shops in the treatment and prevention of childhood malaria on the coast of Kenya. *Trans R Soc Trop Med Hyg, 86*(3), 237-239.

Spencer, H. C., Masaba, S. C., & Kiaraho, D. (1982). Sensitivity of Plasmodium falciparum isolates to chloroquine in Kisumu and Malindi, Kenya. *Am J Trop Med Hyg, 31*(5), 902-906.

Spencer, H. C., Watkins, W. W., Sixsmith, D. G., & Koech, D. K. (1986). Response of Plasmodium falciparum to dihydrofolate reductase inhibitors in Malindi, Kenya. *Trans R Soc Trop Med Hyg, 80*(2), 201-203.

Stephens, M., & Donnelly, P. (2003). A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet, 73*(6), 1162-1169.

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet, 68*(4), 978-989.

Sturm, A., Amino, R., van de Sand, C., Regen, T., Retzlaff, S., Rennenberg, A., et al. (2006). Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science, 313*(5791), 1287-1290.

Su, X., Ferdig, M. T., Huang, Y., Huynh, C. Q., Liu, A., You, J., et al. (1999). A genetic map and recombination parameters of the human malaria parasite Plasmodium falciparum. *Science, 286*(5443), 1351-1353.

Su, X., Kirkman, L. A., Fujioka, H., & Wellems, T. E. (1997). Complex polymorphisms in an approximately 330 kDa protein are linked to chloroquine-resistant P. falciparum in Southeast Asia and Africa. *Cell, 91*(5), 593-603.

Su, X., & Wellems, T. E. (1996). Toward a high-resolution Plasmodium falciparum linkage map: polymorphic markers from hundreds of simple sequence repeats. *Genomics, 33*(3), 430-444.

Su, X. Z., Mu, J., & Joy, D. A. (2003). The "Malaria's Eve" hypothesis and the debate concerning the origin of the human malaria parasite Plasmodium falciparum. *Microbes Infect, 5*(10), 891-896.

Takala, S. L., Smith, D. L., Stine, O. C., Coulibaly, D., Thera, M. A., Doumbo, O. K., et al. (2006). A high-throughput method for quantifying alleles and haplotypes of the malaria vaccine candidate Plasmodium falciparum merozoite surface protein-1 19 kDa. *Malar J, 5*, 31.

Talisuna, A. O., Bloland, P., & D'Alessandro, U. (2004). History, dynamics, and public health importance of malaria parasite resistance. *Clin Microbiol Rev, 17*(1), 235-254.

Talisuna, A. O., Langi, P., Mutabingwa, T. K., Van Marck, E., Speybroeck, N., Egwang, T. G., et al. (2003). Intensity of transmission and spread of gene mutations linked to chloroquine and sulphadoxine-pyrimethamine resistance in falciparum malaria. *Int J Parasitol, 33*(10), 1051-1058.

Tanabe, K., Sakihama, N., Farnert, A., Rooth, I., Bjorkman, A., Walliker, D., et al. (2002). In vitro recombination during PCR of Plasmodium falciparum DNA: a potential pitfall in molecular population genetic analysis. *Mol Biochem Parasitol, 122*(2), 211-216.

Thera, M. A., Sehdev, P. S., Coulibaly, D., Traore, K., Garba, M. N., Cissoko, Y., et al. (2005). Impact of trimethoprim-sulfamethoxazole prophylaxis on falciparum malaria infection and disease. *J Infect Dis, 192*(10), 1823-1829.

Tishkoff, S. A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., et al. (2001). Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science, 293*(5529), 455-462.

Trape, J. F. (2001). The public health impact of chloroquine resistance in Africa. *Am J Trop Med Hyg, 64*(1-2 Suppl), 12-17.

Triglia, T., & Cowman, A. F. (1994). Primary structure and expression of the dihydropteroate synthetase gene of Plasmodium falciparum. *Proc Natl Acad Sci U S A, 91*(15), 7149-7153.

Umotong, A. B., Ezedinachi, E. N., Okerengwo, A. A., Usanga, E. A., Udo, J. J., & Williams, A. I. (1991). Correlation between in vivo and in vitro response of chloroquine-resistant Plasmodium falciparum in Calabar, south-eastern Nigeria. *Acta Trop, 49*(2), 119-125.

UNAIDS. (2004). Provisional WHO/UNAIDS Secretariat Recommendations on the Use of Cotrimoxazole Prophylaxis in Adults and Children Living With HIV/AIDS in Africa.

Urdaneta, L., Lal, A., Barnabe, C., Oury, B., Goldman, I., Ayala, F. J., et al. (2001). Evidence for clonal propagation in natural isolates of Plasmodium falciparum from Venezuela. *Proc Natl Acad Sci U S A, 98*(12), 6725-6729.

Ursos, L. M., & Roepe, P. D. (2002). Chloroquine resistance in the malarial parasite, Plasmodium falciparum. *Med Res Rev, 22*(5), 465-491.

van den Broek, I. V., van der Wardt, S., Talukder, L., Chakma, S., Brockman, A., Nair, S., et al. (2004). Drug resistance in Plasmodium falciparum from the Chittagong Hill Tracts, Bangladesh. *Trop Med Int Health, 9*(6), 680-687.

Van Voorhis, W. (2007). Personal communication: dissertation revisions.

Veatch, J. (2005). Personal communication.

Vieira, P. P., Ferreira, M. U., Alecrim, M. G., Alecrim, W. D., da Silva, L. H., Sihuincha, M. M., et al. (2004). pfcrt Polymorphism and the spread of chloroquine resistance in Plasmodium falciparum populations across the Amazon Basin. *J Infect Dis, 190*(2), 417-424.

Vinayak, S., Mittra, P., & Sharma, Y. D. (2006). Wide variation in microsatellite sequences within each Pfcrt mutant haplotype. *Mol Biochem Parasitol, 147*(1), 101-108.

Vogel, G. (2006). Infectious diseases. Tackling neglected diseases could offer more bang for the buck. *Science, 311*(5761), 592-593.

Vogel, G. (2006). Malaria. Chloroquine makes a comeback. *Science, 314*(5801), 904.

Volkman, S. K., Barry, A. E., Lyons, E. J., Nielsen, K. M., Thomas, S. M., Choi, M., et al. (2001). Recent origin of Plasmodium falciparum from a single progenitor. *Science, 293*(5529), 482-484.

Volkman, S. K., Sabeti, P. C., Decaprio, D., Neafsey, D. E., Schaffner, S. F., Milner, D. A., Jr., et al. (2006). A genome-wide map of diversity in Plasmodium falciparum. *Nat Genet.*

Volkman, S. K., Sabeti, P. C., Decaprio, D., Neafsey, D. E., Schaffner, S. F., Milner, D. A., Jr., et al. (2007). A genome-wide map of diversity in Plasmodium falciparum. *Nat Genet, 39*(1), 113-119.

Walker-Jonah, A., Dolan, S. A., Gwadz, R. W., Panton, L. J., & Wellems, T. E. (1992). An RFLP map of the Plasmodium falciparum genome, recombination rates and favored linkage groups in a genetic cross. *Mol Biochem Parasitol, 51*(2), 313-320.

Wang, P., Lee, C. S., Bayoumi, R., Djimde, A., Doumbo, O., Swedberg, G., et al. (1997). 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.

Wang, P., Read, M., Sims, P. F., & Hyde, J. E. (1997). Sulfadoxine resistance in the human malaria parasite Plasmodium falciparum is determined by mutations in dihydropteroate synthetase and an additional factor associated with folate utilization. *Mol Microbiol, 23*(5), 979-986.

Wang, X., Mu, J., Li, G., Chen, P., Guo, X., Fu, L., et al. (2005). Decreased prevalence of the Plasmodium falciparum chloroquine resistance transporter 76T marker associated with cessation of chloroquine use against P. falciparum malaria in Hainan, People's Republic of China. *Am J Trop Med Hyg, 72*(4), 410-414.

Wanji, S., Tanke, T., Atanga, S. N., Ajonina, C., Nicholas, T., & Fontenille, D. (2003). Anopheles species of the mount Cameroon region: biting habits, feeding behaviour and entomological inoculation rates. *Trop Med Int Health, 8*(7), 643-649.

Watkins, W. (2007). Personal communication.

Watkins, W. M., Mberu, E. K., Winstanley, P. A., & Plowe, C. V. (1997). The efficacy of antifolate antimalarial combinations in Africa: a predictive model based on pharmacodynamic and pharmacokinetic analyses. *Parasitol Today, 13*(12), 459-464.

Watkins, W. M., & Mosobo, M. (1993). Treatment of Plasmodium falciparum malaria with pyrimethamine-sulfadoxine: selective pressure for resistance is a function of long elimination half-life. *Trans R Soc Trop Med Hyg, 87*(1), 75-78.

Watkins, W. M., Sixsmith, D. G., Spencer, H. C., Boriga, D. A., Kariuki, D. M., Kipingor, T., et al. (1984). Effectiveness of amodiaquine as treatment for chloroquine-resistant Plasmodium falciparum infections in Kenya. *Lancet, 1*(8373), 357-359.

Watkins, W. M., Woodrow, C., & Marsh, K. (1993). Falciparum malaria: differential effects of antimalarial drugs on ex vivo parasite viability during the critical early phase of therapy. *Am J Trop Med Hyg, 49*(1), 106-112.

Wellems, T. E. (2002). Plasmodium chloroquine resistance and the search for a replacement antimalarial drug. *Science, 298*(5591), 124-126.

Wellems, T. E., Panton, L. J., Gluzman, I. Y., do Rosario, V. E., Gwadz, R. W., Walker-Jonah, A., et al. (1990). Chloroquine resistance not linked to mdr-like genes in a Plasmodium falciparum cross. *Nature, 345*(6272), 253-255.

Wellems, T. E., Su, X. Z., Ferdig, M., & Fidock, D. A. (1999). Genome projects, genetic analysis, and the changing landscape of malaria research. *Curr Opin Microbiol, 2*(4), 415-419.

Weller, P. F., & Leder, K. (2006). Prevention of malaria infection in travelers: UpToDate.

Wernsdorfer, W. H., & Noedl, H. (2003). Molecular markers for drug resistance in malaria: use in treatment, diagnosis and epidemiology. *Curr Opin Infect Dis, 16*(6), 553-558.

White, N. J. (1985). Clinical pharmacokinetics of antimalarial drugs. *Clin Pharmacokinet, 10*(3), 187-215.

White, N. J. (1992). Antimalarial drug resistance: the pace quickens. *J Antimicrob Chemother, 30*(5), 571-585.

White, N. J. (2004). Antimalarial drug resistance. *J Clin Invest, 113*(8), 1084-1092.

White, N. J., Nosten, F., Looareesuwan, S., Watkins, W. M., Marsh, K., Snow, R. W., et al. (1999). Averting a malaria disaster. *Lancet, 353*(9168), 1965-1967.

Wierdl, M., Dominska, M., & Petes, T. D. (1997). Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics, 146*(3), 769-779.

Williams, T. (2006). KEMRI CGMRC Kilifi, Kenya. *INDEPTH Demographic Surveillance Sites* Retrieved April 12, 2007, from http://www.indepth-network.org/dss_site_profiles/KILIFI%20PROFILE%202006.pdf

Winstanley, P., Newton, C., Watkins, W., Mberu, E., Ward, S., Warn, P., et al. (1993). Towards optimal regimens of parenteral quinine for young African children with cerebral malaria: the importance of unbound quinine concentration. *Trans R Soc Trop Med Hyg, 87*(2), 201-206.

Winstanley, P., Ward, S., Snow, R., & Breckenridge, A. (2004). Therapy of falciparum malaria in sub-saharan Africa: from molecule to policy. *Clin Microbiol Rev, 17*(3), 612-637.

Wongsrichanalai, C., Pickard, A. L., Wernsdorfer, W. H., & Meshnick, S. R. (2002). Epidemiology of drug-resistant malaria. *Lancet Infect Dis, 2*(4), 209-218.

Woodrow, C. J., & Krishna, S. (2006). Antimalarial drugs: recent advances in molecular determinants of resistance and their clinical significance. *Cell Mol Life Sci, 63*(14), 1586-1596.

Wootton, J. C., Feng, X., Ferdig, M. T., Cooper, R. A., Mu, J., Baruch, D. I., et al. (2002). Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. *Nature, 418*(6895), 320-323.

Wu, W. M., Tsai, H. J., Pang, J. H., Wang, T. H., Wang, H. S., Hong, H. S., et al. (2005). Linear allele-specific long-range amplification: a novel method of long-range molecular haplotyping. *Hum Mutat, 26*(4), 393-394.

Wu, Y., Kirkman, L. A., & Wellems, T. E. (1996). 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.

Yeung, S., Pongtavornpinyo, W., Hastings, I. M., Mills, A. J., & White, N. J. (2004). Antimalarial drug resistance, artemisinin-based combination therapy, and the contribution of modeling to elucidating policy choices. *Am J Trop Med Hyg, 71*(2 Suppl), 179-186.

Yu, C. E., Devlin, B., Galloway, N., Loomis, E., & Schellenberg, G. D. (2004). ADLAPH: A molecular haplotyping method based on allele-discriminating long-range PCR. *Genomics, 84*(3), 600-612.

Zhang, K., Zhu, J., Shendure, J., Porreca, G. J., Aach, J. D., Mitra, R. D., et al. (2006). Long-range polony haplotyping of individual human chromosome molecules. *Nat Genet, 38*(3), 382-387.

APPENDIX A: COMPUTATIONAL METHODS FOR PHASING HAPLOTYPES FROM MIXED DNA SAMPLES

As discussed in the introduction to Chapter 2, a significant challenge in studying the evolutionary history of *P. falciparum* is our inability to construct haplotypes from mixed samples of DNA. For our purposes, a haplotype is simply a collection of alleles that exists on a single molecule of DNA. If DNA from only one (haploid) parasite is present in a sample, then one can type each genetic locus individually and reconstruct a haplotype. However, if DNA from multiple different strains of *P. falciparum* is present, then genotyping each locus individually will give the alleles present at each locus, but will not combine them into haplotypes. In other words, one cannot "phase" the alleles into the constituent haplotypes. Because a single patient may be infected with multiple strains of *P. falciparum*, many samples must be discarded from any haplotype analysis.

As shown in Chapter 2, we can use molecular biology and yeast to separate a mixed sample of DNA into its component haplotypes. However, this method is limited by our ability to use PCR to amplify a fragment of *P. falciparum* DNA that contains the entire haplotype of interest. This long-range PCR may not be possible for all types of field samples. Therefore, there is considerable interest in developing a computational method to solve the problem of constructing haplotypes for mixed infections. Such a method would take the allele data for each locus and use maximum likelihood, parsimony, or some other statistical algorithm to phase the data into probable haplotypes.

Existing algorithms for phasing haplotypes

The problem of phasing haplotypes was first addressed in humans. Because humans are diploid, every sample is a "mixed" sample. That is, in any given region of the genome most people will have two haplotypes, one on the maternal chromosome and one on the paternal. Several computational methods have been created for human

(diploid) data [1-7]. The first developed was a simple parsimony method created by Clark in 1990 [1]. Essentially, this algorithm uses homozygous samples (or samples heterozygous at a single locus) to determine known haplotypes within a given population. It then searches through the heterozygous samples to see if they could be explained as a combination of a known haplotype and some other haplotype. If so, then the "other" haplotype is recorded as a known haplotype and the un-phased heterozygous samples are searched again. The process iterates until all samples are phased or no more haplotypes can be found.

Though appealing in its simplicity, this parsimony method can encounter various problems. For one, it will fail if there are no homozygous samples and therefore no way to determine any known haplotypes. Second, there may be "orphan" samples that remain unresolved because they do not match any known haplotypes. Finally, the result will depend on the order in which the samples are considered. Though Clark argues that one simply needs to run the algorithm multiple times and choose the answer that phases the greatest number of samples [1], various researchers have attempted to improve upon his parsimony method by using more advanced modeling methods.

The most promising of these methods is a Bayesian method that uses the coalescent to determine the prior [6,7]. Essentially, Bayesian methods estimate the probability of a hypothesis given the data (the posterior), using the *a priori* probability of the hypothesis (the prior) and the probability of the data given the hypothesis (the likelihood). For the problem of mixed infections of *P. falciparum*, the "hypothesis" is the collection of haplotypes in the sample set, and the data are the genotypes for each sample.

The algorithm developed by Stephens *et al.* (2001, 2003) estimates which haplotype configuration has the greatest likelihood of producing the observed data [6,7]. The algorithm starts with an initial set of haplotypes that might explain the data. Next, the algorithm considers one mixed sample and determines its haplotypes by choosing the ones with the highest likelihood, assuming that all the other haplotypes are

currently correct. The algorithm then replaces this sample (with its newly determined haplotype), considers another, and again determines its haplotypes by assuming that all the others are correct. By repeating this process *ad infinitum* the algorithm eventually comes up with a good guess for what the haplotypes are given the data.

Though multiple methods use this general approach, the advantage of the algorithm developed by Stephens *et al.* (2001, 2003) is that it incorporates information about similarities between haplotypes [6,7]. For example, in considering a sample the algorithm may need to create a novel haplotype to explain the data, one that is not present in any other sample. In this situation, the algorithm is more likely to choose a haplotype that is similar to an existing one. A haplotype that is one SNP away from an existing haplotype is more likely than a haplotype that is five SNPs away. By showing this preference in haplotypes, the algorithm of Stephens *et al.* (2001, 2003) uses information about how haplotypes evolve in determining the set of haplotypes most likely to give the observed genotypes [6,7].

The difficulty in using this method to disentangle the genotype data from mixed infections of malaria arises from the fact that a mixed infection often contains more than two strains of *P. falciparum*, while human DNA samples are almost never more than diploid. Therefore, the program would need to be altered before it could be used. On the other hand, the clonal infections provide an ample source of "homozygous" samples, allowing the algorithm to start with a better idea of what haplotypes are in the sample set.

In addition to methods developed for human data, there is one method that has been developed specifically for mixed infections of *P. falciparum* [8]. This method uses pyrosequencing data for the gene *msp1*. Pyrosequencing is a "sequencing-by-synthesis" technique; incorporation of a base into the growing DNA strand causes chemiluminescence that can be detected [9]. It is useful for detecting SNPs, but because it can only sequence short strands of DNA (50-60 bases, www.pyrosequencing.com) it does not work well for microsatellites. One advantage for phasing, however, is that it gives the relative amounts of each allele at each locus. Therefore, if there are two

alleles at each locus then one can easily phase the haplotypes simply by grouping all of the predominant alleles together.

Takala *et al.* (2006) developed a maximum likelihood algorithm that uses allele frequency data from pyrosequencing to phase *msp1* haplotypes from mixed samples of *P. falciparum* DNA. Essentially, their algorithm finds the set of haplotypes that is most likely to produce the observed allele frequencies. While it is encouraging to see a statistical phasing method for *P. falciparum* DNA, this method is limited by several factors. For one, it can only be used on SNPs, and most haplotype data is for microsatellites. In addition, it uses pyrosequencing, a technology that is not yet widely available. Finally, it only works for samples containing three or fewer haplotypes.

More recently, Ian Hastings has begun to develop a maximum likelihood algorithm for phasing *dhfr* genotypes from mixed infections of *P. falciparum* [10]. Unlike Takala's method, Hastings' method does not rely on pyrosequencing data. For each codon in *dhfr*, a sample can have one or the other allele, or both (e.g. serine or asparagine or both at codon 108). Taking this information and information about the prevalence of particular *dhfr* genotypes in the sample area, the algorithm determines the most likely *dhfr* genotypes in a given sample. While again it is encouraging to see people working on the problem of phasing haplotypes for *P. falciparum*, this program will not help to phase microsatellite haplotypes. For one thing, Hastings argues against increasing the number of loci to more than three because the program would then take too long to run. Increasing the number of alleles at each locus would likewise increase the run time. The total number of possible haplotypes is the product of the number of alleles at each locus, so run time increases exponentially with the number of loci and the number of alleles. Because the haplotypes used to study the evolutionary history of SP resistance consist of three to six microsatellites and three to four codons in *dhfr*, the number of possible haplotypes is far larger than any algorithm thus far developed can handle.

Towards a parsimony program for *P. falciparum*

Because of the limitations of other methods, and as an exercise in programming, I attempted to create a parsimony method for phasing data from mixed samples of *P. falciparum* DNA. I modeled my algorithm after Clark's parsimony method [1], using data from the samples collected in Kenya between 1993 and 1995 (Chapter 3). I did not succeed in creating a complete program, but I learned much about the particular challenges presented by mixed data from *P. falciparum*. The complete R script is given below, but essentially the algorithm proceeds as follows:

1. All clonal samples are identified and their haplotypes marked as "known."
2. All mixed samples are examined to see if they contain a known haplotype. If so, then that haplotype is recorded as present in that sample.
3. All mixed samples are examined to see if they contain a haplotype that differs at only one locus from a known haplotype.

At this point I encountered the limits of my computer's processing power. One reason using microsatellites in phasing algorithms is more challenging than SNPs is that microsatellites have many more alleles than a SNP. Therefore, simply permuting through all the different haplotypes that could make up the set of observed alleles for each sample in my dataset was beyond reasonable time limits. Therefore, the code needs to be made more elegant and efficient before it is of practical use. Processing time considerations may be one reason why Ian Hastings limited his algorithm to two alleles at each of three loci.

Quite apart from computational issues, however, using a parsimony method to phase mixed data presented several interesting problems that will need to be addressed before a phasing algorithm for microsatellites will be successful. In general, these problems arose from the fact that not all alleles present in the sample at a given locus will appear in the data. Because not all alleles amplify to the same degree (see Chapter 2 and the skew towards finding K1 alleles), the lack of an allele at a given locus does not mean that that allele was not present in that sample. Therefore, for some

haplotypes one may have extensive missing data problems, such that only some alleles from the haplotype appear in the data.

Another problem is that one cannot determine how many haplotypes are present in a given sample. The maximum number of alleles present at a single locus gives the *minimum* number of haplotypes present in the sample. Assuming no missing data, the maximum number of haplotypes is the product of the number of alleles at each locus. With missing data, the maximum number of haplotypes is unknowable. A molecular technique with a known, consistent level of sensitivity is needed before implementing any algorithm. That is, we need to be able to say with confidence that if we detect an allele from a component strain of a mixed sample at one locus that we will detect an allele from that strain at all loci.

Given these problems, it seems unlikely that any algorithm could identify all haplotypes present (or detected) in a given sample. Therefore, what should the goal of such an algorithm be? Should it give the single most likely haplotype for each sample? The three most likely? Each user of the algorithm may have a different need. If the goal is to be able to use all samples from an area where the majority of infections are mixed, then simply identifying one haplotype from each sample is enough. However, in most cases an accurate guess could probably be made by eye, and any algorithm would have to demonstrate its superiority to simply looking for known haplotypes in the mixed data. For example, in the data from Kenya 1993-95 it seems very likely that sample 76 contains the canonical triple-mutant haplotype (**Table 3.5**). Because many samples lend themselves to this type of educated guessing, an algorithm may be unnecessary for this type of application, or may offer only a marginal improvement.

If, however, the goal is to define each haplotype within a sample, then a computer algorithm is necessary. Before creating a successful algorithm, the programmer will have to decide how to answer the questions outlined above and how to overcome the exponential growth of possible haplotypes as new loci and alleles are added to a dataset. At this point in time an algorithm to phase microsatellite haplotypes from mixed samples of *P. falciparum* DNA may be beyond our

computational abilities. Nevertheless, it is likely that a successful algorithm will be developed eventually (just not by me).

Complete R Script of the Algorithm Thus Far

```
#Reading in the data
rawdata = read.table("kenya9395.txt", header=TRUE, sep="\t",
     row.names=1, as.is=TRUE)

#Counting the number of samples
nsamp<-(nrow(rawdata)/3)

#Pulling out the clonal samples and setting their marker to 7.
finaldata = matrix(data = c(0,0,0,0,0,0,0,0,-9), nrow=(nsamp*3),
     ncol=9, byrow=TRUE)
 nullhap = c(0,0,0,0,0,0,0,0)
 for (i in 3*1:nsamp){
  test=as.vector(rawdata[(i-1),1:8], mode="numeric")
  if (identical(nullhap, test)) finaldata[(i-2),1:8]<-
     as.vector(rawdata[(i-2),1:8], mode="numeric") else finaldata[(i-
     2),1:8]=finaldata[(i-2),1:8]
  if (identical(nullhap, test)) finaldata[(i-2),9]<-c(7) else
     finaldata[(i-2),9]=finaldata[(i-2),9]}
finaldata #Displays the final data matrix

#Creating a dummy matrix of haplotypes
knhaps = matrix(data=c(3,3,3,3,3,3,3,3), nrow=1, ncol=8)

#Creating the "done" marker
done<- c(7)

#Collecting the known haplotypes (NOT setting their marker to 4).
     Trying not to collect duplicate haplotypes.
#So far it's taken 25 minutes to run on my Kenyan data...how much
     longer? Is it stuck in a loop? Probably.. because of "for (j in
     1:nrow(knhaps))"
gatherhaps <- function(x) {
  oldhaps<-knhaps
  for (i in 3*1:(nrow(x)/3)) { #For each sample
        if (identical(x[(i-2),9], c(7))) for (j in 1:nrow(knhaps)){if
     (identical(as.vector(x[(i-2),1:8],mode="numeric"),
     as.vector(knhaps[j,1:8],mode="numeric"))) knhaps<-knhaps else
     {oldhaps<-knhaps
                    numoldhaps<-nrow(oldhaps)
                    knhaps = matrix(data=c(3,3,3,3,3,3,3,3),
     nrow=(numoldhaps+1), ncol=8, byrow=TRUE)
                    knhaps[1:numoldhaps, 1:8]<- oldhaps[1:numoldhaps,
     1:8]
```

```
                    knhaps[(numoldhaps+1),1:8]<-as.vector(x[(i-
    2),1:8],mode="numeric")
                    }
        }
        else knhaps<-knhaps
   }
knhaps}

knhaps=gatherhaps(finaldata)

finaldata #Displays the final data matrix

#Now, filling in the zeros so that every row is a potential
     haplotype.

workdata<-rawdata

 for (i in 3*1:nsamp){
#  if (identical(finaldata[(i-2),9], c(4))) finaldata[(i-2),9]<-
     finaldata[(i-2),9] else    #If clonal, do nothing.
   for (j in 1:8){                              #If not clonal, for
     each column...
    allele2=as.vector(workdata[(i-1),j], mode="numeric")
    allele3=as.vector(workdata[i,j], mode="numeric")
    zero=c(0)
    if (identical(allele2, zero)) workdata[(i-1),j]<-workdata[(i-
    2),j] else workdata[(i-1),j]<-workdata[(i-1),j]
    if (identical(allele3, zero)) workdata[i,j]<-workdata[(i-1),j]
     else workdata[i,j]<-workdata[i,j]
   }
 }


#Code that will go through all the permutations
permloc = function(j){ #This permutes just the j locus.
 for (i in 3*1:nsamp){
    hold<-workdata[(i-2),j]
    workdata[(i-2),j]<<-workdata[(i-1),j]
    workdata[(i-1),j]<<-workdata[i,j]
    workdata[i,j]<<-hold
    }
 }

testhap = function(x){
 for (i in 3*1:nsamp){
   for (j in 1:nrow(knhaps)){
    samp1=as.vector(x[(i-2),1:8], mode="numeric")
    samp2=as.vector(x[(i-1),1:8], mode="numeric")
    samp3=as.vector(x[i,1:8], mode="numeric")
    hap=as.vector(knhaps[j,1:8], mode="numeric")
if (identical(hap,samp1)) if (identical(as.vector(finaldata[(i-
    2),1:8], mode="numeric"),nullhap)) finaldata[(i-2),1:8]<<- samp1
    else if (identical(samp1,as.vector(finaldata[(i-2),1:8],
```

```
    mode="numeric"))) hap<-hap else if
    (identical(as.vector(finaldata[(i-1),1:8], mode="numeric"),
    nullhap)) finaldata[(i-1),1:8]<<-samp1 else if
    (identical(samp1,as.vector(finaldata[(i-1),1:8],
    mode="numeric"))) hap<-hap else if
    (identical(as.vector(finaldata[i,1:8], mode="numeric"), nullhap))
    finaldata[i,1:8]<<-samp1 else hap<-hap else hap<-hap
if (identical(hap,samp2)) if (identical(as.vector(finaldata[(i-
    2),1:8], mode="numeric"),nullhap)) finaldata[(i-2),1:8]<<- samp2
    else if (identical(samp2,as.vector(finaldata[(i-2),1:8],
    mode="numeric"))) hap<-hap else if
    (identical(as.vector(finaldata[(i-1),1:8], mode="numeric"),
    nullhap)) finaldata[(i-1),1:8]<<-samp2 else if
    (identical(samp2,as.vector(finaldata[(i-1),1:8],
    mode="numeric"))) hap<-hap else if
    (identical(as.vector(finaldata[i,1:8], mode="numeric"), nullhap))
    finaldata[i,1:8]<<-samp2 else hap<-hap else hap<-hap
if (identical(hap,samp3)) if (identical(as.vector(finaldata[(i-
    2),1:8], mode="numeric"),nullhap)) finaldata[(i-2),1:8]<<- samp3
    else if (identical(samp3,as.vector(finaldata[(i-2),1:8],
    mode="numeric"))) hap<-hap else if
    (identical(as.vector(finaldata[(i-1),1:8], mode="numeric"),
    nullhap)) finaldata[(i-1),1:8]<<-samp3 else if
    (identical(samp3,as.vector(finaldata[(i-1),1:8],
    mode="numeric"))) hap<-hap else if
    (identical(as.vector(finaldata[i,1:8], mode="numeric"), nullhap))
    finaldata[i,1:8]<<-samp3 else hap<-hap else hap<-hap
   }
 }
}

#I want permloc(8) every n, permloc(7) every 3n, permloc(6) every 9n,
    permloc(5) every 27n, and so on.
#counter<-0 #Don't really need the counter.
#This takes about 3 min for 5 samples, 7 min for 10, 15 for 15.
#So it's far too slow to be useful.
for (i in 1:6561){
 testhap(workdata)
# counter<-counter+1
 permloc(8)
 if (i/3 - floor(i/3) == 0) permloc(7) else workdata<-workdata
 if (i/3^2 - floor(i/3^2) == 0) permloc(6) else workdata<-workdata
 if (i/3^3 - floor(i/3^3) == 0) permloc(5) else workdata<-workdata
 if (i/3^4 - floor(i/3^4) == 0) permloc(4) else workdata<-workdata
 if (i/3^5 - floor(i/3^5) == 0) permloc(3) else workdata<-workdata
 if (i/3^6 - floor(i/3^6) == 0) permloc(2) else workdata<-workdata
 if (i/3^7 - floor(i/3^7) == 0) permloc(1) else workdata<-workdata
}

#******END OF PROGRAM*******BEGINNING OF WORK SPACE*******

#Clearing out the solved samples:
for (i in 3*1:5){
```

```
  if (identical (as.vector(workdata[(i-2),1:8], mode="numeric"),
     as.vector(workdata[(i-1),1:8], mode="numeric")) & identical
     (as.vector(workdata[(i-2),1:8], mode="numeric"),
     as.vector(workdata[i,1:8], mode="numeric"))) workdata [(i-
     2):i,1:8]<-nullhap
  if (identical (as.vector(finaldata[i,1:8], mode="numeric"),
     nullhap)) nullhap<-nullhap else workdata [(i-2):i,1:8]<-nullhap
}

#Now workdata is empty except for those that are not made up only of
     known haplotypes.
#S338 is still all there in workdata.

#Next the program should look for haplotypes that are only one allele
     off from a known.
testhap2 = function(x){
 for (i in 3*1:5){
   for (j in 1:nrow(knhaps)){
    samp1=as.vector(x[(i-2),1:8], mode="numeric")
    trunc1=as.vector(x[(i-2),2:8], mode="numeric")
    samp2=as.vector(x[(i-1),1:8], mode="numeric")
    trunc2=as.vector(x[(i-1),2:8], mode="numeric")
    samp3=as.vector(x[i,1:8], mode="numeric")
    trunc3=as.vector(x[i,2:8], mode="numeric")
    hap=as.vector(knhaps[j,2:8], mode="numeric")
if (identical(hap,trunc1)) if (identical(as.vector(finaldata[(i-
     2),1:8], mode="numeric"),nullhap)) finaldata[(i-2),1:8]<<- samp1
     else if (identical(samp1,as.vector(finaldata[(i-2),1:8],
     mode="numeric"))) hap<-hap else if
     (identical(as.vector(finaldata[(i-1),1:8], mode="numeric"),
     nullhap)) finaldata[(i-1),1:8]<<-samp1 else if
     (identical(samp1,as.vector(finaldata[(i-1),1:8],
     mode="numeric"))) hap<-hap else if
     (identical(as.vector(finaldata[i,1:8], mode="numeric"), nullhap))
     finaldata[i,1:8]<<-samp1 else hap<-hap else hap<-hap
if (identical(hap,trunc2)) if (identical(as.vector(finaldata[(i-
     2),1:8], mode="numeric"),nullhap)) finaldata[(i-2),1:8]<<- samp2
     else if (identical(samp2,as.vector(finaldata[(i-2),1:8],
     mode="numeric"))) hap<-hap else if
     (identical(as.vector(finaldata[(i-1),1:8], mode="numeric"),
     nullhap)) finaldata[(i-1),1:8]<<-samp2 else if
     (identical(samp2,as.vector(finaldata[(i-1),1:8],
     mode="numeric"))) hap<-hap else if
     (identical(as.vector(finaldata[i,1:8], mode="numeric"), nullhap))
     finaldata[i,1:8]<<-samp2 else hap<-hap else hap<-hap
if (identical(hap,trunc3)) if (identical(as.vector(finaldata[(i-
     2),1:8], mode="numeric"),nullhap)) finaldata[(i-2),1:8]<<- samp3
     else if (identical(samp3,as.vector(finaldata[(i-2),1:8],
     mode="numeric"))) hap<-hap else if
     (identical(as.vector(finaldata[(i-1),1:8], mode="numeric"),
     nullhap)) finaldata[(i-1),1:8]<<-samp3 else if
     (identical(samp3,as.vector(finaldata[(i-1),1:8],
     mode="numeric"))) hap<-hap else if
```

```
    (identical(as.vector(finaldata[i,1:8], mode="numeric"), nullhap))
    finaldata[i,1:8]<<-samp3 else hap<-hap else hap<-hap
   }
 }
}

for (i in 1:6561){
 testhap2(workdata)
# counter<-counter+1
 permloc(8)
 if (i/3 - floor(i/3) == 0) permloc(7) else workdata<-workdata
 if (i/3^2 - floor(i/3^2) == 0) permloc(6) else workdata<-workdata
 if (i/3^3 - floor(i/3^3) == 0) permloc(5) else workdata<-workdata
 if (i/3^4 - floor(i/3^4) == 0) permloc(4) else workdata<-workdata
 if (i/3^5 - floor(i/3^5) == 0) permloc(3) else workdata<-workdata
 if (i/3^6 - floor(i/3^6) == 0) permloc(2) else workdata<-workdata
 if (i/3^7 - floor(i/3^7) == 0) permloc(1) else workdata<-workdata
}

#Clearing out the samples that now have 3 haplotypes:
for (i in 3*1:5){
  if (identical (as.vector(finaldata[i,1:8], mode="numeric"),
    nullhap)) nullhap<-nullhap else workdata [(i-2):i,1:8]<-nullhap
}
#So that's one way to solve my little sample dataset. The problem,
    though, is that it doesn't
#account for all alleles present in the data. Does that matter?

remhap = function(x){
 for (i in 3*1:5){                                    #for each batch of
    three haplotypes
   for (j in 1:nrow(knhaps)){                         #for each known
    haplotype
    samp1=as.vector(x[(i-2),1:8], mode="numeric") #samp1-3 are the
    three possible haplotypes
    samp2=as.vector(x[(i-1),1:8], mode="numeric")
    samp3=as.vector(x[i,1:8], mode="numeric")
    hap=as.vector(knhaps[j,1:8], mode="numeric")         #hap is the
    known haplotype
    if (identical(hap,samp1)) workdata[(i-2),1:8] <<- nullhap else
    samp1<-samp1   #if samp1 matches the known haplotype
    if (identical(hap,samp2)) workdata[(i-1),1:8] <<- nullhap else
    samp2<-samp2   #if samp1 matches the known haplotype
    if (identical(hap,samp3)) workdata[i,1:8] <<- nullhap else
    samp3<-samp3   #if samp1 matches the known haplotype
   }
 }
}
for (i in 1:6561){
 remhap(workdata)
# counter<-counter+1
 permloc(8)
 if (i/3 - floor(i/3) == 0) permloc(7) else workdata<-workdata
```

```
 if (i/3^2 - floor(i/3^2) == 0) permloc(6) else workdata<-workdata
 if (i/3^3 - floor(i/3^3) == 0) permloc(5) else workdata<-workdata
 if (i/3^4 - floor(i/3^4) == 0) permloc(4) else workdata<-workdata
 if (i/3^5 - floor(i/3^5) == 0) permloc(3) else workdata<-workdata
 if (i/3^6 - floor(i/3^6) == 0) permloc(2) else workdata<-workdata
 if (i/3^7 - floor(i/3^7) == 0) permloc(1) else workdata<-workdata
}
```

NOTES TO APPENDIX A

1. Clark, A.G. Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol Biol Evol* **7**, 111-22 (1990).

2. Excoffier, L. & Slatkin, M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* **12**, 921-7 (1995).

3. Lin, S., Cutler, D.J., Zwick, M.E. & Chakravarti, A. Haplotype inference in random population samples. *Am J Hum Genet* **71**, 1129-37 (2002).

4. Niu, T., Qin, Z.S., Xu, X. & Liu, J.S. Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. *Am J Hum Genet* **70**, 157-69 (2002).

5. Qin, Z.S., Niu, T. & Liu, J.S. Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. *Am J Hum Genet* **71**, 1242-7 (2002).

6. Stephens, M., Smith, N.J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**, 978-89 (2001).

7. Stephens, M. & Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* **73**, 1162-9 (2003).

8. Takala, S.L. et al. A high-throughput method for quantifying alleles and haplotypes of the malaria vaccine candidate Plasmodium falciparum merozoite surface protein-1 19 kDa. *Malar J* **5**, 31 (2006).

9. Nyren, P. The history of pyrosequencing(r). *Methods Mol Biol* **373**, 1-14 (2006).

10. Hastings, I.M. MalHaplotypeFreq. (Liverpool School of Tropical Medicine, 2007).

APPENDIX B: SUPPLEMENTARY INFORMATION FOR CHAPTER TWO

As stated in Chapter 2, the first step in the method to separate mixed DNA is inefficient due to the presence of colonies that grow on the selective plates but do not, as far as we can tell, contain the desired insert. To rule out contamination, we changed reagents, pipetmen, locations, and thermalcyclers. None of these had any effect. We observed that the negative-control PCR (containing no template DNA) produced colonies, but only if put through the thermalcycler program. We concluded that the primers were priming each other enough to create a fragment that had homology to the plasmid. However, when we redesigned the plasmid so that it only had homology to the desired PCR product, we continued to see some empty colonies, though the total number of colonies from a transformation was fewer. Therefore, probably the "primer-dimers" are only part of the problem. We tried various other things in an effort to reduce the number of colonies; they are listed below. None led to an improvement, and most decreased efficiency.

Variations on the protocol

To avoid re-ligation of the cut plasmid:

Double-cut vector – Cut with *Bam H1* and *Xho I.*

Gel purification of cut vector – Used Freeze 'N Squeeze Spin Columns (Bio-Rad Laboratories, Hercules, CA) to purify the cut vector (both singly and doubly cut).

To increase yield of positive colonies:

No counter-selection – Did not require the long-range PCR product to replace URA3; plates lacked tryptophan but did not contain FOA.

Transformation protocol – Using an older method.[1]

Increasing homology – Using PCR primers with longer tails.

To remove primer-dimers:

GENECLEAN® – Cleaned the long-range PCR products with GENECLEAN® *Turbo* (Qbiogene, Morgan Irvine, CA) prior to transformation.

ExoSAP – Treated the long-range PCR products with exonuclease and shrimp alkaline phosphatase prior to transformation.

NOTE TO APPENDIX B

1. Schiestl, R.H. & Gietz, R.D. High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr Genet* **16**, 339-46 (1989).

VITA FOR LAURA CERTAIN

EDUCATION

Princeton University, Princeton, New Jersey
BA, Chemistry, 2000, *Summa cum laude*

University of St. Andrews, Scotland
Study abroad through Butler University, Indiana, Sept. 1998 – Jan. 1999

University of Washington School of Medicine, Seattle, Washington
MD-PhD candidate, entering class of 2001

AWARDS RECEIVED

President's Award for Academic Achievement, 1997-98, Princeton University
First-Year Physics Medal, 1998-99, University of St. Andrews, Scotland
Evertt S. Wallis Prize in Organic Chemistry, 2000, Princeton University
Phi Beta Kappa, elected 2000
Sigma Xi, elected 2000
Achievement Rewards for College Scientists Fellowship Award, 2004
Poncin Scholarship, 2004-2006
University of Washington Graduate Student Medal, 2006

RESEARCH EXPERIENCE

University of Wisconsin, Madison, June – August 1998
Mentor: Philip Farrell, MD, PhD
Topic: Differences in genetic counseling practices among the 50 states.

Princeton University Chemistry Department, February 1999 – April 2000
Mentor: Michael Hecht, PhD
Topic: Enzymatic activity of proteins designed from first principles.

Children's Hospital Medical Center, Cincinnati, Ohio, July 2000 – June 2001
Mentor: Robert Kahn, MD, MPH
Topics: Peripartum smoking trends; television viewing by toddlers.

Fred Hutchinson Cancer Research Center, July – September 2001
Mentor: Julie Overbaugh, PhD
Topic: HIV sub-typing.

University of Washington, Seattle, June – August 2002
Mentor: Carol Sibley, PhD
Topic: Mechanism of DHFR regulation in *Plasmodium falciparum.*

Malaria Research and Training Center, Bamako, Mali, July – September 2003
Coordinated by: University of Maryland, Center for Vaccine Development
Topic: Malaria field research.

University of Washington, Seattle, Genome Sciences Dept. PhD Program, October 2003 – June 2007
Mentor: Carol Sibley, PhD
Topic: Population genetics of drug resistance in *Plasmodium falciparum.*

PUBLICATIONS & PRESENTATIONS

Moffet DA, Certain LK, *et al.* Peroxidase activity in heme proteins derived from a designed combinatorial library. *J. Am. Chem. Soc.* 2000; 122(31): 7612-7613.

Farrell MH, Certain LK, Farrell PM. Genetic counseling and risk communication services of newborn screening programs. *Arch. Pediatr. Adolesc. Med.* 2001; 155: 120-126.

"Television viewing among very young children." Certain LK and Kahn RS. Presented at the Pediatric Academic Societies' Annual Meeting, April 30, 2001, Baltimore, Maryland.

Certain LK, Kahn RS. Prevalence, correlates and trajectory of television viewing among infants and toddlers. *Pediatrics* 2002; 109(4): 634-642.

Kahn RS, Certain LK, Whitaker RC. A reexamination of smoking before, during, and after pregnancy. *Am. J. Public Health* 2002; 92(11): 1801-1808.

"A novel method for studying mixed infections of *Plasmodium falciparum.*" Certain LK, Sibley CH. Presented at the 16th Annual Seattle Protozoology Conference, May 2004, Seattle, Washington.

"A novel method for studying mixed infections of *Plasmodium falciparum.*" Certain LK, Sibley CH. Poster presented at the Annual Meeting of the American Society of Tropical Medicine and Hygiene, November 2004, Miami Beach, Florida.

"Pilot study of sulfadoxine-pyrimethamine resistance using a novel method to analyze mixed infections." Certain LK, Sibley CH. Poster presented at the Annual Meeting of the American Society of Tropical Medicine and Hygiene, December 2005, Washington, DC.

"Minimum genetic diversity in resistant strains of Plasmodium falciparum from Kenya." Certain LK, Sibley CH. Presented at the 18th Annual Seattle Parasitology Conference, May 2006, Seattle, Washington.

"Limited genetic diversity in resistant strains of *Plasmodium falciparum* from Kenya." Certain LK, Nzila AM, Sibley CH. Poster presented at the Annual Meeting of the American Society of Tropical Medicine and Hygiene, November 2006, Atlanta, GA.

Certain LK, Sibley CH. *Plasmodium falciparum:* A novel method for analyzing haplotypes in mixed infections. *Experimental Parasitology* 2007; 115(3): 233-41.

WORK EXPERIENCE (NON-RESEARCH)

Camp counselor, Concordia Language Villages (French), 1995, 1996

Inter-library services employee, Princeton University, Firestone Library, 1996-2000

MCAT instructor, The Princeton Review, 2000

Organic chemistry TA and tutor, Princeton University, 2000

Medical school tutor, University of Washington, 2002-2007