A Population Genetic Analysis of Antifolate Resistance in *Plasmodium falciparum* in southeast Africa.

A thesis presented by

Richard James Pearce

In partial fulfilment of the requirements for admittance to the degree of:

Doctor of Philosophy

London School of Hygiene and Tropical Medicine

2005

Acknowledgements

The work of this thesis involves the analysis of blood samples taken from infected people in South Africa, Mozambique and Tanzania. All samples were collected during the course of regional investigations of drug resistance, drug treatment or malaria diagnosis. I am indebted to the people who consented to take part in the studies and to the large number of people who were involved in the collection of these samples.

Chapter 2: Field collections in Hai were designed and coordinated by Dr Daniel Chandramohan and funded by a DfID grant (R7950). Those in the Pare Mountains were undertaken by Drs. Chris Drakeley and Frank Mosha and conducted under the auspices of the Joint Malaria Programme, a collaborative research initiative between The London School of Hygiene and Tropical Medicine, The Tanzanian National Institute for Medical Research, The Kilimanjaro Christian Medical College and the Centre for Medical Parasitology, University of Copenhagen funded by a grant from MRC UK (G9901439, to Professor BM Greenwood, Professor EM Riley et al). We are grateful for logistical support from Prof. WMMM Nkya and Dr H. Reyburn and for technical support from E Nyale, F. Laizer and M. Mosha.

Chapters 3-6: South African samples from KwaZulu Natal were collected as part of the national malaria research programme and made available for analysis by Professor Brian Sharp, Head of the Malaria Lead Programme of Medical Research Council, Durban. The Mozambican samples were collected under the auspices of Lubombo Spatial Development Initiative (LSDI) again provided for analysis by Prof. Sharp. The South African samples from Mpumalanga were collected by A. Mabuza

and his Mpumalanga Malaria Control programme team in 2001 as part of the Southeast African Combination Antimalarial Therapy (SEACAT) evaluation,which received partial financial support from the UNDP / World Bank / WHO Special Programme for Research and Training in Tropical Diseases (WHO TDR). They were made available for analysis by Professor Karen Barnes.

The samples from southern Tanzania were collected as a part of the Interdisciplinary Monitoring Project for Antimalarial Combination Therapy in Tanzania (IMPACT-Tz), funded by USAID, CDC, and Wellcome Trust, were made available for analysis by co-Principal Investigators Salim Abdulla, S. Patrick Kachur and Peter Bloland.

I am indebted to Mr Allen Malisa and Ms .Alana Keyser for typing the large numbers of bloodspots samples from Tanzania, Mozambique and South Africa, and to Mrs Hirva Pota for technical assistance with the microsatellite analysis of samples from southern Tanzania.

I am grateful to constructive comments and discussion throughout the duration of the research project from other members of the Roper and Conway lab groups (2000 to present), in particular Drs David Conway, Spencer Polley and Kevin Tetteh, Mr Gareth Weedall and Mrs Hirva Pota.

I will be eternally grateful to Dr Cally Roper for giving me the opportunity to combine my role as a research technician with that of research degree student. Her unremitting patience, sense of fair play and encouragement, in addition to allowing me the space to pursue an argument experimentally or discursively (and be wrong), has been a great education.

My own financial support came through an Advanced Training Fellowship from The Wellcome Trust (ref 060714) awarded to Dr C.Roper.

Finally, I am very grateful to my wife, Anne-Marie, for her unerring support and generosity of spirit when I had none.

Summary

This thesis is a description of the effects of Sulphadoxine/Pyrimethamine (SP) selection on the genome of *Plasmodium falciparum.* Antimalarial resistance in *P. falciparum* to the antifolate combination of SP is conferred by a series of substitutions that alter the active sites of the target enzymes dihydrofolate reductase (DHFR) and dihydropteroate synthetase (DHPS).

Different combinations of substitutions result in differing levels of drug insensitivity as shown by both in vitro studies and association with treatment failure. In southeast Africa, where the samples taken for this study originate, the two most highly resistant *dhfr* and *dhps* alleles are the triple mutant (N51I+C59R+S108N) and the double mutant (A437G+K540E) , respectively. A molecular population genetic analytic approach is applied to examine the emergence and spread of SP resistant mutations in Africa. We observe that differences in the frequencies of the resistance alleles between southeast African populations broadly reflect heterogeneity in drug selection history across the region. We find that this exists despite strongly homogenising gene flow.

Selection for a favourable allele can have effects on neutral loci flanking the selected site. A selective sweep occurs when neutral flanking loci hitchhike with the selected allele as it increases in frequency, reducing genetic diversity along the chromosome in the population. The hitchhiking alleles indicate the ancestry of the selected allele and has shown that alleles at *dhfr* and *dhps* comprised of multiples of mutations have emerged rarely in east African parasite populations (Roper et al. 2003).

The thesis is the first description of the full extent of the selective sweeps around three *dhfr* and *dhps* resistance alleles in southeast Africa. The thesis contains analysis of the changes that occur to a selective sweep over time and in populations with contrasting recombination rates and selection histories. Through use of a deterministic model we identify that gene flow plays an important role in establishing the frequency of the resistance allele at values greater than $^{1}/_{Ne}$, the frequency of a de novo mutation.

Contents

Table of Figures

Table of Tables

Chapter 1 *Plasmodium falciparum*,Drug Selection and Selective Sweeps

The apicomplexan parasite, *Plasmodium falciparum* is the most deadly of the four human malarias. The World Health Organisation (WHO) estimated that in 1998 *P. falciparum* globally caused 273 million clinical attacks per year, with over 90% of clinical disease occurring in sub Saharan Africa, of which the majority was in children (Greenwood and Mutabingwa 2002). A recently published estimate for the year 2002 reported that the number of clinical attacks per year globally was 515 million clinical attacks (Snow et al. 2005),hig her than estimated by WHO and revising the burden of malaria in sub Saharan Africa to 70% of the global total, to reflect the under reported burden of malaria in southeast Asian populations

In addition to the cost in human life, *P. falciparum* malaria through its effect on morbidity greatly burdens the economies of African countries. The cost of malaria to the annual GDP of African countries is a 1.9% reduction in economic growth to below that of the global average outside Africa (2.3%) (Sachs and Malaney 2002). The importance of reducing the malaria burden globally is great. The control of this disease and HIV/Aids are arguably the first steps in relieving poverty in developing countries. Organisations such as the WHO and Global Fund have recognised this and have targeted *P. falciparum* malaria as one of three key infections to control in the developing world, putting malaria alongside HIV/Aids and Tuberculosis.

The focus of this literature review will specifically be on drug resistance in *P. falciparum* malaria,conc entrating particularly on the situation in southeast African while making some reference to drug resistance globally.

I *Plasmodium falciparum* basic biology

a The life cycle

The definitive host of *P. falciparum* is the female Anopheline mosquito. Blood feeding on humans (the intermediary host) by the mosquito releases infective sporozoites collected in the mosquito salivary glands into the human bloodstream. Once in the bloodstream, the sporozoites move rapidly to infect hepatocytes in the liver (Figure 1-1). Here sporozoites multiply through several rounds of cell division (schizogony) eventually releasing merozoites into the bloodstream where they infect erythrocytes. Here begins erythrocytic schizogony and the parasite numbers increase through numerous rounds of multiplication via the trophozoite stage and then mature schizonts that rupture, releasing merozoites into the bloodstream, infecting other erythrocytes. A subset of merozoites leaves erythrocytic schizogony and differentiates into gametocytes. The gametocytes are taken up in a blood meal by a female Anopheline mosquito where they further develop into male and female gametes that fuse and form a zygote in the insects' stomach lining. The zygote develops into the ookinete which moves across the mosquito stomach wall forming a sporozoite-filled oocyst. The oocyst bursts, the sporozoites are released into the haemocoele and move to the mosquito salivary glands. There the process can begin again when the mosquito inoculates these into the bloodstream of another host.

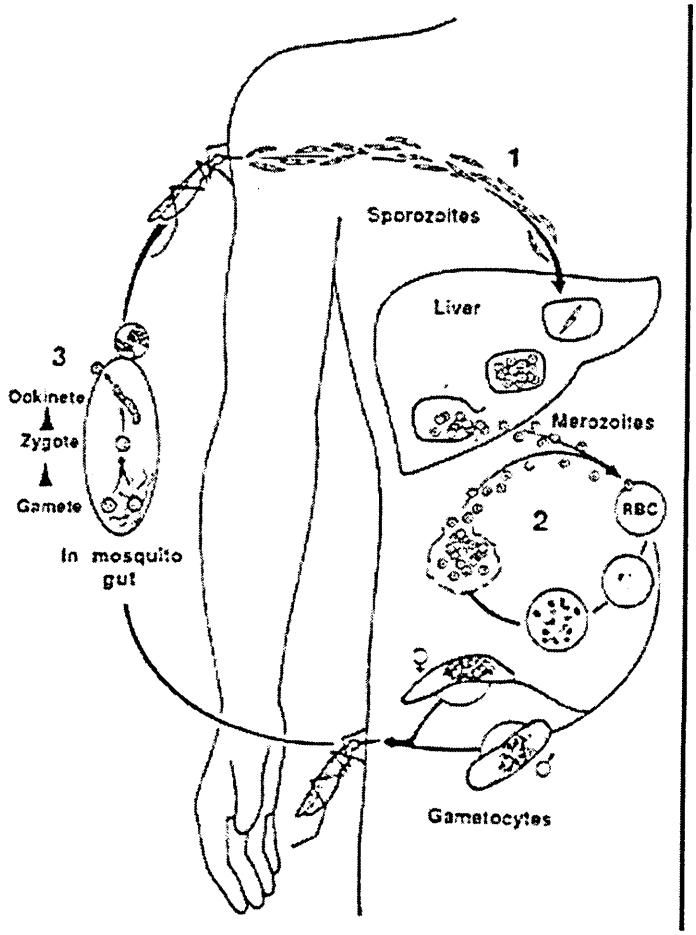


Figure 1-1: The life cycle of *Plasmodium falciparum* between the definitive host, the Anopheline mosquito, and the human host. Details of the lifecycle steps can be found in the text. Taken from (Miller et al. 1986)

b Genetic consequences of the lifecycle

The *P. falciparum* genome is comprised of 14 highly A+T rich (>80%) chromosomes (Gardner et al. 2002), which are haploid throughout the parasites lifecycle in the human host. When the male and female gametes fuse in the mosquito gut, the genome becomes diploid and homologous recombination occurs during meiosis. Recombination occurs in *P. falciparum* at a high rate; breakpoints are estimated to occur at a rate of 1cM per 17kb as determined by analysis of microsatellite genotypes in a genetic cross between two *P. falciparum* strains, Hb3 and Dd2 (Su et al. 1999). This estimated meiotic recombination rate is 20 times higher than that estimated in Drosophila and 40 times higher than in humans (Conway et al. 1999). Estimates of the recombination rate in the field in African *P. falciparum* populations are similarly high. Linkage disequilibrium in antigen genes MSP1 and AMA1 reaches half its maximal value at a distance of about 300 base pairs (Conway et al. 1999; Polley and Conway 2001).

It is important to note that the meiotic recombination rate in a genetic cross is an overestimate of recombination in the field because selfing occurs between gametocytes of the same genotype. In African parasite populations parasite genetic diversity is high, whereas at other locations globally diversity tends to be lower; least diverse populations are in South America (Anderson et al. 2000a). This difference in genetic diversity limits the opportunities for recombination to occur between different genetically distinct parasite lineages.

c *Transmission intensity*

In areas of high transmission intensity infections tend to be composed of multiple genetically distinct parasite lines (Babiker, Ranford-Cartwright, and Walliker 1999). These occur either through super-infection of a host who has received multiple infectious bites, or through the bite of a mosquito carrying a mixture of genotypes within its' inoculating saliva. The multiplicity of an infection is correlated with a measure of transmission intensity, known as the entomological inoculation rate (EIR). This is a measure of the number of infective bites per individual per annum. In a review of 159 distinct sites recorded in sub Saharan Africa, regardless of land use type (EIRs are lower in urban areas), annual EIR ranged from 0 to 884, with a mean of 121 infected bites per annum (Hay et al. 2000) indicating that malaria transmission is both endemic and high. By contrast, in South America and southeast Asia, the annual EIR tends to be low, rarely exceeding 5 infective bites per adult per annum (reviewed by (Greenwood and Mutabingwa 2002). Malaria in such areas is unstable, and rather than endemic in a population, it is prone to epidemics, where the majority of disease occurs during short bursts of intense transmission. It is less likely that multiple infections occur under these circumstances and the opportunities for outcrossing are reduced as a consequence.

d *Transmission intensity and its' effect on host parasite interactions*

In addition to the effect on genetic diversity, the second effect of high malaria transmission is that the development of clinical immunity in the human host occurs quicker than in low transmission areas, as an individuals' immune system receives a far higher number of challenges (Snow et al. 1997). For an individual living in an

area of intense transmission the first few years of childhood are when the risks of severe disease and mortality are highest. With age the risk of mortality and severity of disease decrease as the individual develops immunity against the disease (Snow et al. 1997). Infections in the clinically immune are by definition asymptomatic, although attacks of fever and headache are common. However, premunition is not sterilising immunity and resolution of infections tends to occur through, or is assisted by, use of antimalarial chemotherapy.

II Antimalarials and *P. falciparum*

a History of antimalarial usage

Treating malaria infection has been a concern of humans for centuries and it is unsurprising that the earliest antimalarials were isolated from plants. The oldest recorded antimalarial was used in China over 2000 years ago. Artemesinin is extracted from the Chinese herb quinghao (*Artemisia annua*) and had been done so for over 1500 years before the discovery of Quinine, found in the bark of the Peruvian 'fever tree' (*Chinchona spp.*). Artemesinin is undergoing resurgence of interest and its use in Artemesinin-based Combination Therapy (ACT) is the most promising future direction of antimalarial treatments.

Quinine was isolated from the bark of the Chinchona tree in the 19th century (Cowman 1997; Talisuna, Bloland, and D'Alessandro 2004). It has provided the backbone for much of the development of novel antimalarials since shortly after the First World War, as well as current usage as treatment of drug resistant malaria infections. It is not in widespread use as a prophylactic or therapeutic antimalarial

because of severe side effects of intravascular haemolysis and acute renal failure; colloquially called 'black water fever'

Shortly before the Second World War, breakthroughs were made in the manufacture of synthetic quinine analogues known as the 4-aminoquinolones. Of the four discovered to have antimalarial activity, chloroquine and amodiaquine were the two most useful, having the least contraindications, although amodiaquine is not recommended for prophylactic use due to adverse reactions (Cowman 1997).

Use of the highly schizonticidal compound chloroquine as a drug of choice for malaria treatment began in the mid 1940s and it was readily adopted due to its cheap manufacture and minimal side effects. The other great benefit of chloroquine is its antipyretic activity, as reducing fevers quickly aids the perception of a curative effect of the drug whereas other antimalarials that do not have antipyretic action, such as sulphadoxine pyrimethamine, the perceived benefit is lessened. This has latterly become an issue as anecdotal evidence points to its continued use in populations where the prevalence of resistance alleles are high, because the antipyretic benefit remains.

In the 1960s resistance to the 4-aminoquinolones, most importantly chloroquine, began to emerge with the earliest reports of chloroquine resistance coming out of South America in 1960 followed by Southeast Asia in 1962 reviewed by (Payne 1989). Resistance to chloroquine then radiated out from these initial foci eventually reaching Africa in the 1980s (Figure 1-2). From the temporal distribution of reported chloroquine resistance Payne concluded that there was a dispersal of resistance through Southeast Asia, spreading outwards to India and eventually East Africa.

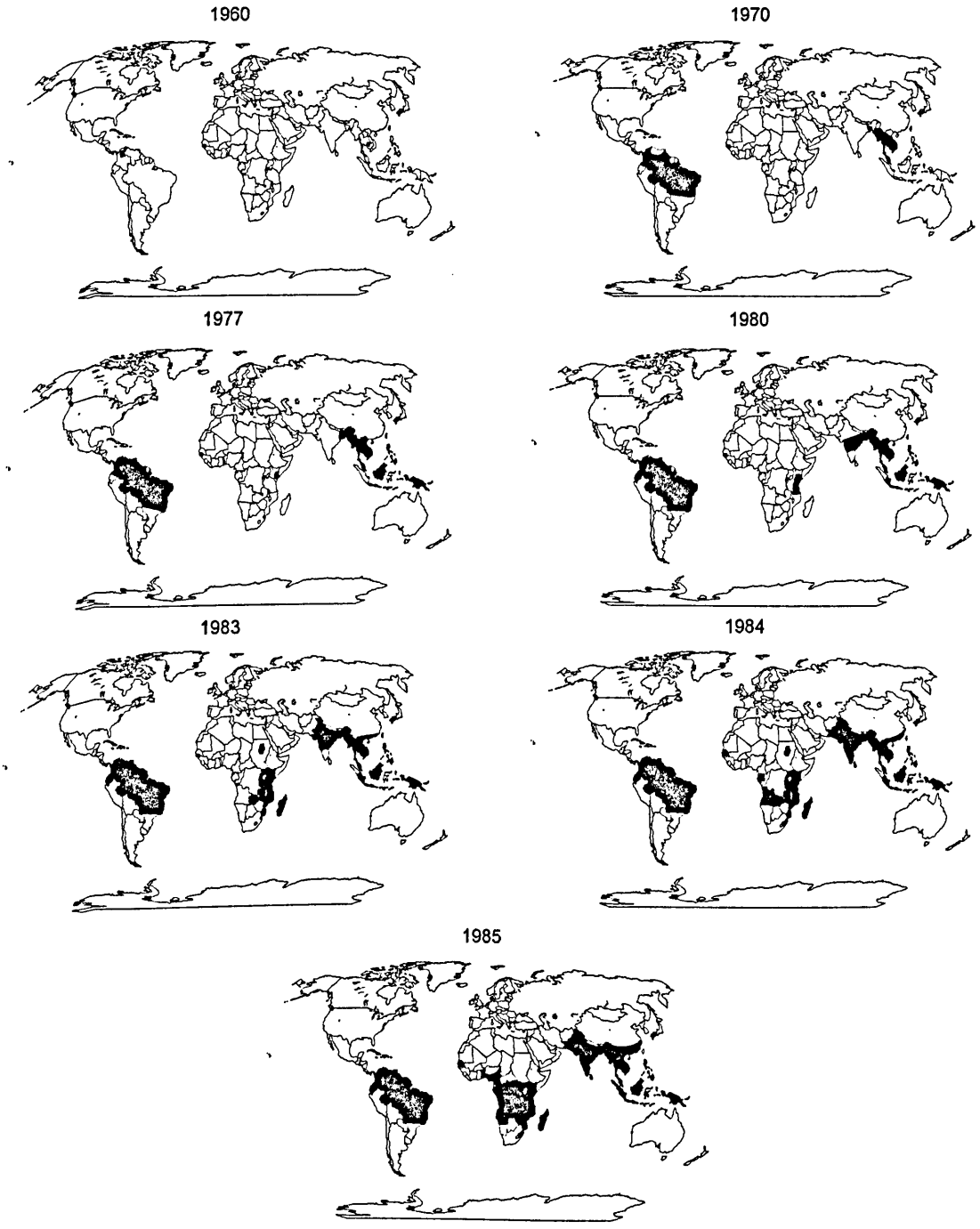


Figure 1-2: The geographical expansion of chloroquine resistance globally from 1960 to 1985. Adapted from (Payne 1989)

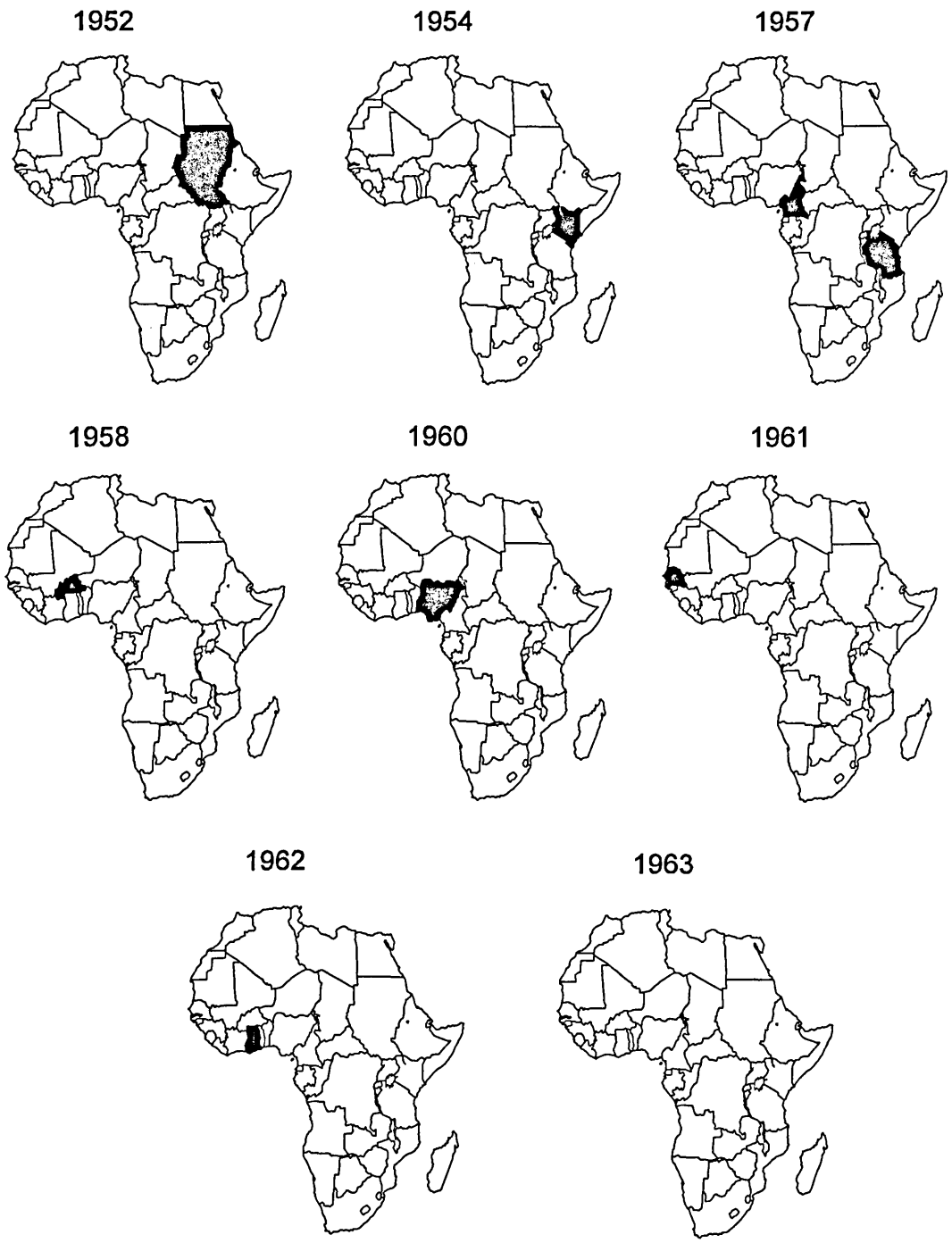


Figure 1-3: A time series of reported in vivo resistance after wide scale prophylactic use of pyrimethamine in Africa from 1952 to 1963, collated from (Peters 1970). Countries shaded in black represent the first report of pyrimethamine resistance in that country. Note the apparent spread of the resistance phenotype in East Africa from Sudan to Kenya to Tanzania but in West Africa the pattern is more (cont'd over)

disjointed. The explanation for the apparently discontinuous spread of a resistance phenotype in West Africa (1957) is either under-reporting or the occurrence of de novo mutation

The conclusion that spread rather than repeated de novo emergence of resistance in each country was later supported by molecular studies of the ancestry of chloroquine resistance alleles (Wootton et al. 2002), and this is reviewed below.

As chloroquine came into widespread usage in the 1940s the chemical company ICI began to investigate the potential of pyrimidine derivatives as antimalarials. In 1945 proguanil was identified as a drug with low toxicity and high activity against avian malaria and in the early 1950s pyrimethamine was developed. Early use of pyrimethamine quickly led to resistance detected within populations across Africa (Figure 1-3) reviewed by (Peters 1970). Notably there is an apparent spread of the pyrimethamine resistance phenotype in east Africa from Sudan to Kenya to Tanzania, similar to the spread chloroquine resistance out of south east Asia, however it also appears that it has emerged de novo independently at several sites, particularly in the disjointed distribution of pyrimethamine resistance in West Africa.

In an attempt to prolong the useful therapeutic life of pyrimethamine, it was combined during the 1960s with a sulphonamide, sulphadoxine, in a drug called SP or Fansidar, as trademarked by Hoffman LaRoche. Sulphadoxine and Pyrimethamine are competitive inhibitors of enzymatic components of the folate biosynthesis pathway, dihydropteroate synthase (DHPS; bifunctionally combined with hydroxymethylpterin pyrophosphokinase, PPPK-DHPS) and dihydrofolate reductase (DHFR; bifunctionally combined with thymidylate synthetase, DHFR-TS) respectively. The two inhibitors work synergistically in disrupting folate synthesis and the parasites lifecycle. They are sometimes considered to be a single 'monotherapy' drug rather than a 'combination therapy' because they target the same pathway (Sibley et al. 2001), the development of resistance to one of the drugs

reversing the synergy. The first reported national use of SP in southeast Africa began in Tanzania in 1982 when it was introduced as second-line treatment for use in cases of chloroquine failure. As early as 1984 it was the policy of the Muheza district hospital to use SP as first line antimalarial (Mutabingwa et al. 2001), although it did not become first line antimalarial throughout Tanzania as national policy until 2001. KwaZulu Natal in South Africa followed shortly after Muheza, changing policy from chloroquine usage to SP in 1988. The first country in Africa to begin using SP as first line treatment was Malawi, in 1993. Four years later Kenya, South Africa (following KwaZulu Natal) and Botswana followed Malawi in adopting SP as first line antimalarial (Bloland et al. 1993; Bloland et al. 1998).

Resistance to SP was relatively quick to develop, although slower than to pyrimethamine alone. The earliest reports of emerging SP resistance in Africa were from Muheza district in Tanzania during 1994 and 1995 (Ronn et al. 1996; Trigg et al. 1997). Studies from the region around Muheza reported that SP was highly effective during the late 1980s but resistance was present in Magoda village near Muheza in 1994 (Ronn et al. 1996). It was subsequently reported in villages in the surrounding area (Jelinek et al. 1997; Trigg et al. 1997; Jelinek et al. 1998).

b Mechanism of resistance to Sulphadoxine Pyrimethamine

The mechanism of resistance to sulphadoxine and pyrimethamine was identified through association in vitro with a series of substitutions within the active site of the target enzymes of the folate biosynthesis pathway, dihydropteroate synthase (DHPS) (Brooks et al. 1994; Triglia and Cowman 1994) and dihydrofolate reductase (DHFR)

(Cowman et al. 1988; Peterson, Walliker, and Wellems 1988; Snewin et al. 1989) respectively. This was demonstrated through laboratory based in vitro sensitivity tests and transfection experiments on DHFR (Zolg et al. 1989; Wu, Kirkman, and Wellems 1996) and DHPS (Triglia et al. 1997; Wang et al. 1997b; Triglia et al. 1998). A Ser to Asn substitution at codon 108 of DHFR decreases sensitivity to pyrimethamine by a 100 fold (Wu, Kirkman, and Wellems 1996). Additional substitutions at codons 51 (N51I), 59 (C59R), and 164 (I164L) progressively increase levels of resistance to pyrimethamine (Hyde 1990; Wu, Kirkman, and Wellems 1996). Isolates containing all four substitutions have been found in South America and Southeast Asia but have yet to have been reported in field studies in Africa (Wang et al. 1997a; Mutabingwa et al. 2001; Kublin et al. 2002). At DHPS 14 substitutions at five sites have been characterised world wide of which six have been recorded in Africa. The A437G and K540E mutations are the most frequently reported in Africa (Wang et al. 1997a; Wang et al. 1997b; Eberl et al. 2001).

The role of the point mutations at each locus in conferring resistance to SP in vivo has been inferred from studies showing predictive association of particular mutations with treatment failure (Omar, Adagu, and Warhurst 2001; Kublin et al. 2002), and from over-representation of mutations in recrudescent infections following treatment failure (Edoh et al. 1997; Jelinek et al. 1997; Khan et al. 1997; Basco, Tahar, and Ringwald 1998; Cortese and Plowe 1998; Curtis, Duraisingh, and Warhurst 1998; Jelinek et al. 1999a; Jelinek et al. 1999b; Basco et al. 2000; Doumbo et al. 2000; Nzila et al. 2000a).

c *Folate salvage and dhps enzyme*

The importance of sulphadoxine in killing parasites in vivo has been under some doubt, because of the ability of the reported parasite to source folate externally (Sibley et al. 2001). Different parasite lines show differing abilities to use exogenous folate and some grow almost normally in very high concentrations of sulphadoxine in vitro,ind icating that *dhps* is non essential in the biosynthesis of folate (Wang et al. 1997b; Wang et al. 1999). The ability of parasites to salvage folate from the host has been postulated as the explanation of why in the majority of populations where SP resistance occurs, resistance alleles at *dhfr* are found in higher frequencies than at *dhps*.

Genetic analysis of folate salvage in *P. falciparum* was performed on the genetic cross between a sulphadoxine sensitive parasite strain, Hb3, and a sulphadoxine resistant strain, Dd2. It was found that there was complete linkage of the folate utilisation phenotype with a 48.6kb region around *dhfr* on chromosome 4 (Wang et al. 2004a). Of the 7 putative open reading frames in this region, the strongest candidate for the folate salvage phenotype was *dhfr*. Low level folate reductase activity has been detected at *dhfr* of other systems such as chickens, bacteria and mammals, and it is plausible that this extends to *P. falciparum* (Wang et al. 2004a). It is argued, but not shown, that different combinations of mutations at *dhfr* may affect the ability of a parasite to utilise exogenous folate, as it is speculated that as Hb3, which displays little folate utilisation effect, is a single S108N mutant, and that the presence of three mutations (N51I+C59R+S108N) at *dhfr* of parasite line Dd2 explains its highly efficient folate utilisation effect phenotype (Wang et al. 2004a).

In contradiction to the assertion that *dhps* is non essential, recent transfection studies have shown that parasites with truncated forms of *dhps* are not viable (Wang et al. 2004b). It was shown that DHPS activity above a low yet critical level is essential for viable parasites, regardless of the availability of salvageable folate. Earlier studies had assumed that high concentrations of sulphadoxine were completely inhibiting all *dhps* activity with negligible effect on parasite viability because of the folate salvage effect. The authors speculate that this contradiction is resolved by considering that *dhps* may be active in more than one compartment in the cell, in which one compartment is perhaps less susceptible to sulphadoxine influx (Wang et al. 2004b). Analysis of *dhps* mutant alleles in field populations of *P. falciparum* have shown that positive selection is operating on *dhps* A437G+K540E double mutant resistance alleles during a period of SP firstline use in South African populations (Roper et al. 2003).

d Evolution and emergence of SP resistance

Compared to chloroquine resistance, SP resistance developed relatively quickly. This is thought to be due to the long half life of the drug in the body of treated people, wherein a long tail of sub inhibitory plasma concentrations can select for allelic forms of the enzymes that do not survive full therapeutic doses, but are mutant stepping stones to fully drug insensitive forms of the enzymes (Hastings, Watkins, and White 2002). For example N51I+S108N *dhfr* double mutant allele can only survive plasma concentrations found in the body 5-7 days after initial dosing with pyrimethamine, whereas the sensitive form of the enzyme is inhibited by pyrimethamine concentrations present in the plasma for more than 52 days after the initial dosing (Watkins et al. 1997). The *dhfr* N51I+C59R+S108N triple mutant is thought to be of

borderline sensitivity to the initial high concentrations of pyrimethamine in the blood plasma (Watkins et al. 1997; Hastings, Watkins, and White 2002).

In addition to the long half life of the drug, the mechanism of resistance to SP is simple when compared to chloroquine, requiring only substitutions in the drug target to preferentially bind the native substrate rather than the inhibitor. By contrast chloroquine targets the polymerisation of toxic haem from digestion of haemoglobin to the inert haemozoin and thus resistance occurs through the prevention of chloroquine from accumulating in the digestive vacuole of the red blood cell (Cowman 1997).

It has long been assumed that the point mutations that confer resistance to SP occur regularly. Indeed, the in vitro mutation rate of codon 108 of *dhfr* has been calculated as $2.5x10^{-9}$ mutations/DHFR gene/infection (Paget-McNicol and Saul 2001). As each patent infection contains between 10^{10}- 10^{12} parasites, this equates to a mutation occurring at *dhfr* during every infection. Recent studies using flanking markers around *dhfr* and *dhps* have thrown the ability of these de novo mutations to survive into doubt (reviewed below).

III The effects of strong selection on chromosomal flanking sequence

Strong positive selection operating on a favourable allele can reduce gene diversity at loci flanking the selected gene as the frequency of a favourable allele increases. The association of neutral flanking alleles with the selected site is termed 'hitchhiking' (Smith and Haigh 1974; Kaplan, Hudson, and Langley 1989). As the hitchhiking alleles increase in frequency in the population, a valley of reduced diversity on the

flanking sequence surrounding the selected site occurs, hence the term selective sweep (Smith and Haigh 1974).

a Determinant parameters of selective sweeps

The persistence of a selective sweep in a population and the length of chromosome affected are dependant on both the strength of selection and the rate of recombination. If selection for a favourable allele is strong and it increases in frequency to fixation rapidly, the selective sweep will be large. Conversely if the rate of recombination is high the associations between selected site and flanking markers are rapidly broken down and the selective sweep is likely to be small.

There are a range of sizes of selective sweeps reported. In humans the region of reduced diversity found around the lactate dehydrogenase extended for more than 1 Mb (Bersaglieri et al. 2004). In organisms where there are higher rates of recombination, selective sweeps are contained to a smaller region of sequence flanking the selected site such as the extended haplotype of ~88kb in a highly recombining region of the genome of *Drosophila simulans* (Quesada et al. 2003).

Interpretation of the cause of selection giving rise to ancient selective sweeps is often speculative. In organisms with long generation times, such as humans, the formation of a selective sweep occurs over a long time, limiting the observations that can be made of the changes occurring in genome in real time. Retrospective data can clarify the nature, strength and duration of selection occurring for an allele in some cases.

By contrast the selective sweeps that arise due to therapeutic drug selection in malaria parasites occur in a defined period, selection beginning when the drug was first used and ending when the drug is no longer used. Changes in the frequency of resistance alleles can be used to directly measure the strength of selection occurring. This combined with a short generation time provides an opportunity to observe the effects of selection on a genome with precision over a precisely defined period of time.

b Drug resistance associated selective sweeps in malaria: Chloroquine resistance

The first of the drug resistance related selective sweeps to be identified was in a sample of chloroquine resistant parasites sampled from Africa, southeast Asia, South America and Papua New Guinea (Wootton et al. 2002). Although the genetic basis for chloroquine resistance is thought to be multigenic, the key determinant is thought to be *Pfcrt* on chromosome 7 (Sidhu, Verdier-Pinard, and Fidock 2002). Wootton et al took samples of chloroquine resistant and sensitive parasites and using 342 highly polymorphic microsatellite markers described patterns of linkage disequilibrium over all 14 haploid chromosomes. Through the comparison of chloroquine resistant and chloroquine sensitive genomes they were able to identify regions of extensive linkage disequilibrium among the resistant isolates. This process identified only one locus, namely *Pfcrt* on chromosome 7 which was flanked by a region of >200kb of reduced allelic diversity.

Four chloroquine resistance founding events were identified (Wootton et al. 2002). The *Pfcrt* allele found in high frequencies in both Asian and African chloroquine resistant populations, shared the same ancestry. A further three lineages were

described in two locations in South America and one in Papua New Guinea, in each case the *Pfcrt* allele present had arisen only once and was of distinct ancestry to alleles sampled at other sites (Wootton et al. 2002). The transcontinental spread of the chloroquine resistance allele across Asia and the African continent is consistent with the history of chloroquine resistance and its spread out of southeast Asia into neighbouring countries (Payne 1989) (Figure 1-2). It is a matter for speculation where, when and how many times the allele came to enter the African parasite population.

The key conclusion from this work is that the widespread occurrence of chloroquine resistance is not due to numerous de novo mutation events, but rather due to the expansion of a limited number of monophyletic resistance lineages.

c Drug resistance selective sweeps in malaria: SP resistance

Genetic mapping of resistance traits in mixed samples from many global locations can be confounded by admixture. It is therefore preferable to limit detection of linkage disequilibrium to samples taken from a single population, where the risks of admixture linkage are much reduced. There have been three studies describing the ancestry of pyrimethamine resistance, but as opposed to the chloroquine study, these were performed in a continent specific manner.

a.South America

Analysis of *dhfr* and *dhps* alleles and closely linked microsatellite loci, one in an intron within *dhps* and another within the 5' UTR of *dhfr*,showed strong association

of flanking alleles with each of the resistance alleles (Cortese et al. 2002). However, matching polymorphism at four microsatellite loci not physically linked to either *dhps* or *dhfr*,wa s also found and there was linkage disequilibrium between resistance alleles at *dhfr*,*dhps* ,*Pfcrt* and *pfmdr1*. The linkage disequilibrium between these four resistance loci indicates that recombination between drug resistant parasites and drug sensitive parasites is very rare, either due to low transmission or assortative mating enforced by intense selection pressure. In populations where transmission is low, diversity at neutral loci is generally low, reflecting the low recombination rate. Thus descriptions of the full extent of a selective sweep around any one of the resistance loci would be impossible.

b.Southeast Africa

Chronologically the second published report of evidence of a selective sweep around SP resistance loci was in populations in southeast Africa. This description was of a region encompassing 3 microsatellites within 8kb of *dhps* on chromosome 8 and 3 microsatellites within 5kb of *dhfr* on chromosome 4 (Roper et al. 2003). In South Africa and Tanzania we found three separate origins of parasites carrying two mutations at *dhfr*, but just a single origin of parasites carrying the triple mutant (N51I+C59R+S108N) *dhfr* alleles. Microsatellite flanking markers were tightly associated with each *dhfr* resistance allele containing 2 or more mutations and the association of different flanking markers with different *dhfr* resistance alleles demonstrates that each has an independent origin (Roper et al. 2003). At *dhps* the double A437G+K540E mutant allele was identified as being monophyletic in South Africa and northern Tanzania. The work presented in this thesis extends this work describing the full extent of the selective sweeps around these two SP resistance loci

in southeast Africa. Here the *P. falciparum* population is one of the most diverse populations with both extremes of global recombination rates.

c. *Southeast Asia*

The full extent of a selective sweep around *dhfr* on chromosome 4 has been described in southeast Asian pyrimethamine resistant parasites collected from clinics on the Thailand – Myanmar border. It was described as a region of 12kb of strongly reduced gene diversity within a wider valley extending from 58kb upstream to over 40kb downstream (Nair et al. 2003). The significance of the reduction of microsatellite variation within the ~100kb region around *dhfr* was determined by comparison of chromosomes carrying the resistance alleles to a deterministic hitchhiking model, as sensitive chromosomes were absent in the region (Nair et al. 2003).

There was a diversity of resistant alleles at *dhfr* with mutations at codons 51, 59, 108 and 164. Importantly all alleles carrying 2-4 mutations had identical or very similar flanking microsatellites indicating a single ancestral origin. The selective sweep is effectively shared between multiple alleles. Selection was not concurrent with the time period that the samples were taken (Nosten et al. 2000). Since the cessation of drug pressure in the early 1980s, the selective sweep has survived despite recombination. The size of this selective sweep around *dhfr* was smaller than that described around *Pfcrt*. Factors which may explain this are the global sampling strategy for the *Pfcrt* analysis and the age of the *dhfr* sweep in Asia. With regard to a description of a selective sweep around Southeast Asian sulphadoxine resistant *dhps* alleles, this work is ongoing (T.J.C.Anderson, personal communication).

d The importance of dispersal in resistance

Numerous models describing the evolution of antimalarial drug resistance have made the assumption that one mutation should arise de novo per infection. It is clear from the above descriptions of the ancestry of SP resistance alleles in South America, Asia and Africa, that contrary to expectations highly resistant *dhfr* and *dhps* alleles arise very rarely indeed (Cortese et al. 2002; Nair et al. 2003; Roper et al. 2003). Comparing the flanking haplotype at 6 markers over a 30kb region around the *dhfr* triple mutant resistance alleles from Africa and *dhfr* resistance alleles from southeast Asia, it was shown that the triple mutant allele from southeast Asia had been introduced into Africa and has subsequently introgressed into the population over considerable distances (>4000km) (Roper et al. 2003; Roper et al. 2004). The spread of the *dhfr* triple mutant resistance allele from southeast Asia to Africa echoes the spread of the chloroquine resistance *Pfcrt* allele between the two populations (Wootton et al. 2002). This is a strong indication that the spread of resistance alleles is a more important factor in determining the useful therapeutic life of an antimalarial, rather than the de novo mutation rate (Hastings 2004). The differences seen between sweeps around genes such as *dhfr* and *Pfcrt* and the remarkable effect of epidemiological context such as in South America or southeast Asia means that further research is needed to explore the African situation.

IV Thesis Outline

Chapter 2: We describe a high throughput sequence specific oligonucleotide probing (SSOP) dot blot methodology for detection of point mutations at *dhfr* and *dhps* in

field collections of blood from *P. falciparum* infected people. The occurrence of single genotype infections allowed us to infer allelic haplotypes consisting of different combinations of point mutations present in three northern Tanzania populations. We compared the level of resistance between sites by measuring the frequency of these allelic haplotypes in each population and made inferences about the recent history of SP selection based on allele frequencies and linkage disequilibrium between *dhfr* and *dhps* resistance alleles.

Chapter 3: We applied the SSOP-dot blot and allelic haplotype methodology to three populations in southeast Africa which have different histories of SP use. We found heterogeneity in the frequencies of resistance alleles between the three populations. By typing the populations at eight neutral microsatellite markers we showed that this heterogeneity exists in the face of strongly homogenising gene flow. In one of the populations, South Africa, we found evidence of epidemic expansion of genetically similar parasites and high degrees of inbreeding.

Chapter 4: We describe, for the first time,the f ull extent of the selective sweep around *dhfr* triple mutant alleles in Africa *P. falciparum.* We compared diversity on resistant and sensitive chromosomes. By using triple mutant chromosomes taken during a longitudinal study in South Africa, we observe the effects of time on the size of a selective sweep. Sampling in other southeast African populations with different selection histories and transmission intensities we examine the impact on diversity around the triple mutant allele. Using a deterministic model to predict the extent of the selective sweep under given selection and recombination conditions we show that gene flow was almost certainly important in establishing an initial starting frequency greater than $^{1}/_{Ne}$ in all these populations.

Chapter 5: We describe the extent of the selective sweep around a *dhfr* C59R+S108N double mutant allele. The *dhfr* double mutant allele is less resistant to pyrimethamine than the *dhfr* triple mutant and was displaced by the triple mutant allele in South African populations where SP was the primary treatment for malaria (Roper et al. 2003). We find evidence of reduced gene diversity on double mutant chromosomes sampled from Tanzania and South Africa. The extent of the selective sweep was smaller than that present on triple mutant chromosomes for the same population but was not as small as predicted by the deterministic model for the population specific selection and recombination parameters. Once again it appears that gene flow in determining initial starting frequencies played an important role in defining the size and shape of the selective sweep.

Chapter 6: We describe the extent of the selection sweep around the *dhps* A437G+K540E double mutant allele. This allele is a more recent arrival in southeast Africa than resistant *dhfr* alleles. Chromosomes sampled from two South African populations, had significantly reduced gene diversity around the *dhps* double mutant allele when compared with sensitive chromosomes. The extent of the selective sweep was greater than that around the *dhfr* triple mutant in the same sample. We discuss the effects of time, migration, drug use, and selection pressure on the size and shape of the selective sweeps around *dhfr* and *dhps*.

Chapter 7: We draw broad conclusions relating to the effect of SP selection of the genome of southeast African *P. falciparum*.

Chapter 2 Characterising Point Mutation Haplotypes in the Dihydrofolate Reductase and Dihydropteroate Synthase Genes of *Plasmodium falciparum*

I Abstract

The genetic determinants of in vitro resistance to the two drugs individually are shown to be point mutations at seven sites in dihydrofolate reductase (*dhfr*) conferring resistance to pyrimethamine and five sites in dihydropteroate synthase (*dhps*) conferring resistance to sulphadoxine. Different combinations of mutations within each gene confer differing degrees of drug sensitivity but information about the haplotypic conformations of point mutations and the frequency with which they occur has been lacking because of the complicating effects of multiple infection. We developed a novel high throughput sequence specific oligonucleotide probing based approach to screen for and infer haplotype frequencies. We then demonstrate its practical use in an analysis of three *P. falciparum* populations in northern Tanzania. This chapter describes the method and haplotype scoring approach and illustrates its application in Tanzania. Screening surveys of asymptomatic infections for the presence of all known point mutations in *dhfr* and *dhps* genes we showed that just five *dhfr* and three *dhps* allelic haplotypes are present. A high frequency of both triple mutant *dhfr* and double mutant *dhps* mutant alleles were found in addition to significant inter-regional heterogeneity in allele frequency. In vivo studies have shown that the co-occurrence of three *dhfr* mutations and two *dhps* mutations in an infection prior to treatment are statistically predictive of treatment failure. We combined data for both loci to determine the frequency of two locus genotypes. The triple *dhfr*/double *dhps* genotype was present in all three regions with frequencies ranging between 30% and 63% predicting that treatment failure rates will be high.

This work has been published as Pearce et al 2003 (Appendix 8).

II Introduction

Sulphadoxine – Pyrimethamine (SP) has now replaced chloroquine as the first line curative antimalarial in much of East Africa. Resistance to Sulphadoxine and Pyrimethamine are associated in vitro with a series of substitutions within the active site of target enzymes of the folate biosynthesis pathway, dihydropteroate synthase (DHPS) (Brooks et al. 1994; Triglia and Cowman 1994) and dihydrofolate reductase (DHFR) (Cowman et al. 1988; Peterson, Walliker, and Wellems 1988; Snewin et al. 1989) respectively, and this has been demonstrated through laboratory based in vitro sensitivity tests and transfection experiments on DHFR with respect to pyrimethamine (Zolg et al. 1989; Wu, Kirkman, and Wellems 1996) and DHPS with respect to sulphadoxine (Triglia et al. 1997; Wang et al. 1997b; Triglia et al. 1998). The sequence changes coding for substitutions that are naturally occurring worldwide are summarised in Table 2-1. A Ser to Asn substitution at codon 108 of DHFR decreases sensitivity to pyrimethamine by a 100 fold (Wu, Kirkman, and Wellems 1996). Additional substitutions at codons 51 (N51I), 59 (C59R), and 164 (I164L) progressively increase levels of resistance to pyrimethamine (Hyde 1990; Wu, Kirkman, and Wellems 1996). Isolates containing all four substitutions have been found in South America and Southeast Asia but have yet to have been reported in Africa (Wang et al. 1997a; Mutabingwa et al. 2001; Kublin et al. 2002). At DHPS 14 substitutions at five sites have been characterised world wide of which six have been recorded in Africa, with A437G and K540E mutations being the most frequently reported (Wang et al. 1997a; Wang et al. 1997b; Eberl et al. 2001).

DHFR Probes	Probe Name	Amino Acid	Probe Sequence
CODON 16	A*	ALA	CC ATA TGT **GCA** TG T TGT A
	S	SER	CC ATA TGT **TCA** TG T TGT A
	V	VAL	CC ATA TGT **GTA** TG T TGT A
CODON 50 &	CN*	CYS ASN	TGG AAA **TGT AAT** TCC CTA
CODON 51	CN2*	CYS ASN	TGG AAA **TGT AAC** TCC CTA
	RN	ARG ASN	TGG AAA **CGT AAT** TCC CTA
	RN2	ARG ASN	TGG AAA **CGT AAC** TCC CTA
	RI	ARG ILE	TGG AAA **CGT ATT** TCC CTA
	CI	CYS ILE	TGG AAA **TGT ATT** TCC CTA
CODON 59	C*	CYS	AA TAT TTT **TGT** GCA GTT A
	R	ARG	AA TAT TTT **CGT** GCA GTT A
CODON 108	N	ASN	A AGA ACA **AAC** TGG GAA AG
	S*	SER	A AGA ACA **AGC** TGG GAA AG
	T	THR	A AGA ACA **ACC** TGG GAA AG
CODON 140	V*	VAL	AT GAA GAT **GTT** TAT ATC A
	L	LEU	AT GAA GAT **CTT** TAT ATC A
CODON 164	I*	ILE	GT TTT ATT **ATA** GGA GGT T
	L	LEU	GT TTT ATT **TTA** GGA GGT T
DHPS Probes			
CODON 436 &	SA*	SER ALA	GAA TCC **TCT GCT** CCT TTT
CODON 437	SG	SER GLY	GAA TCC **TCT GGT** CCT TTT
	FA	PHE ALA	GAA TCC **TTT GCT** CCT TTT
	FG	PHE GLY	GAA TCC **TTT GGT** CCT TTT
	AA	ALA ALA	GAA TCC **GCT GCT** CCT TTT
	AG	ALA GLY	GAA TCC **GCT GGT** CCT TTT
	CA	CYS ALA	GAA TCC **TGT GCT** CCT T TT
CODON 540	K*	LYS	ACA ATG GAT **AAA** CTA ACA
	E	GLU	ACA ATG GAT **GAA** CTA ACA
CODON 581	A*	ALA	A GGA TTT **GCG** AAG AAA CA
	G	GLY	A GGA TTT **GGG** AAG AAA CA
CODON 613	A*	ALA	GA TTT ATT **GCC** CAT TGC
	T	THR	GA TTT ATT **ACC** CAT TGC
	S	SER	GA TTT ATT **TCC** CAT TGC

Table 2-1: Summary of the sites at which a SNP known to be associated with SP resistance occurs, and the oligonucleotide probe designed to detect it. The sequence in bold represents the codon within which the point mutation occurs. The asterisk following the Probe Name indicates the wildtype sensitive codon. Column 3 indicates the amino acid change that occurs at the relevant codon following the point mutation

The role of the point mutations at each locus in conferring resistance to SP in vivo has been inferred from studies showing predictive association of particular mutations with treatment failure (Omar, Adagu, and Warhurst 2001; Kublin et al. 2002), and from over-representation of mutations in recrudescent infections after treatment (Edoh et al. 1997; Jelinek et al. 1997; Khan et al. 1997; Basco, Tahar, and Ringwald 1998; Cortese and Plowe 1998; Curtis, Duraisingh, and Warhurst 1998; Jelinek et al. 1999a; Jelinek et al. 1999b; Basco et al. 2000; Doumbo et al. 2000; Nzila et al. 2000a). Such studies are complicated by mixed infections, although the blood stage parasites are haploid the co-occurrence of more than one genotype in an infection means that variation at multiple sites cannot be assigned to an individual parasite line within the infection and accordingly the predictive association with treatment outcome is complicated.

To understand the natural history of the point mutations that occur at *dhfr* and *dhps* one needs to know how they occur in populations. In order to study the frequency of alleles at a population level, we have designed a PCR-SSOP high throughput approach for detection of known single nucleotide polymorphisms (SNP) in order to identify and construct haplotypes. Haplotypes are combinations of SNPs that are in the same gene, in the same parasite: as distinct from associations of point mutations that co-occur because there is a mixture of parasites of different genotypes within a single infection. Haplotypes are biologically meaningful since they determine the resistance properties of parasites that are exposed to drug at the time of treatment. For example a triple mutant *dhfr* haplotype of N51I+C59R+S108N has a 1.5 – 3 fold higher pyrimethamine resistance in vitro than either of N51I+S108N or C59R+S108N double mutant haplotypes (Sirawaraporn et al. 1997). A mixed infection containing these two double mutant alleles is less resistant to pyrimethamine than an infection

containing the triple mutant allele despite all three mutations being present in either case. It is important when comparing populations, to measure the frequency of haplotypes, rather than prevalence of each point mutation separately, because haplotypes are the determinants of the drug resistance levels.

The method we have employed involves the PCR amplification of sequence from the coding regions of *dhfr* and *dhps* genes, which is fixed on to membranes and probed with sequence specific oligonucleotide probes (SSOP) (Conway et al. 1999), designed to detect each of the single base pair substitutions at all positions summarised in Table 2-1. Figure 2-1 is an example of a pair of dot blots probed for the polymorphism at codon 59 of *dhfr*. The SSOP method has advantages for high throughput, while retaining the equivalent sensitivity and specificity of other methods used for detection of *dhfr* and *dhps* SNPs (Abdel-Muhsin et al. 2002; Ranford-Cartwright et al. 2002). We used tetramethylammonium chloride (TMAC) to standardise the melting temperature (Tm) of dioxigenin labelled oligonucleotide probes, so enabling duplicate membranes to be probed and washed at a standard temperature and sequence variants at all SNP sites to be detected simultaneously. In addition to these improvements over other variations on SSOP-dot blotting (Abdel-Muhsin et al. 2002), we modified the methodology described in (Conway et al. 1999) in the following ways:

1. DNA extractions from bloodspots were performed in 96 well arrays in plates designed to take at least one ml capacity. By performing DNA extractions in this format, many more extractions could be performed in a shorter period of time removing a bottleneck to subsequent high throughput steps.
2. The second modification was to switch the substrate used in the detection of bound probe sequence. In the original methodology the detection substrate

was CSPD, a chemiluminescent substrate which was used to expose the blots to x-ray film. The range of detection of film is limited and a strong signal can quickly exceed the upper limit of the sensitivity range. The length of exposure of the blot to the film was variable as the strength of signal, a function of the quantity of target sequence on the blot, varied between populations analysed. To eliminate this, a chemifluorescent signal was used instead. The substrate ECF (Enhanced Chemi-Fluorescence) is broken down by alkaline phosphatase and then using the Storm phosphoimager the strength of signal following excitation with blue laser light (nm=440) is digitised and quantified. The intensity of the excitatory light can be raised or lowered to adjust for the quantity of target sequence on the blots, thereby decreasing or increasing the emission over all spots to fit within the range of sensitivity of the Storm Phosphoimager.

3. The final modification is a benefit of being able to quantify signal from the blots as a function of the maximum intensity of signal within a defined area, termed volume, as defined by the software Image Quant. Scoring the absence, presence and relative abundance of signal was previously performed by eye and was open to subjective bias. Quantification of signal has removed this subjectivity. A number of rules for scoring were developed and are described below. Essentially the volume of signal for a single sample probe for one SNP can be compared to the other SNPs for that locus through their relative signal strengths.

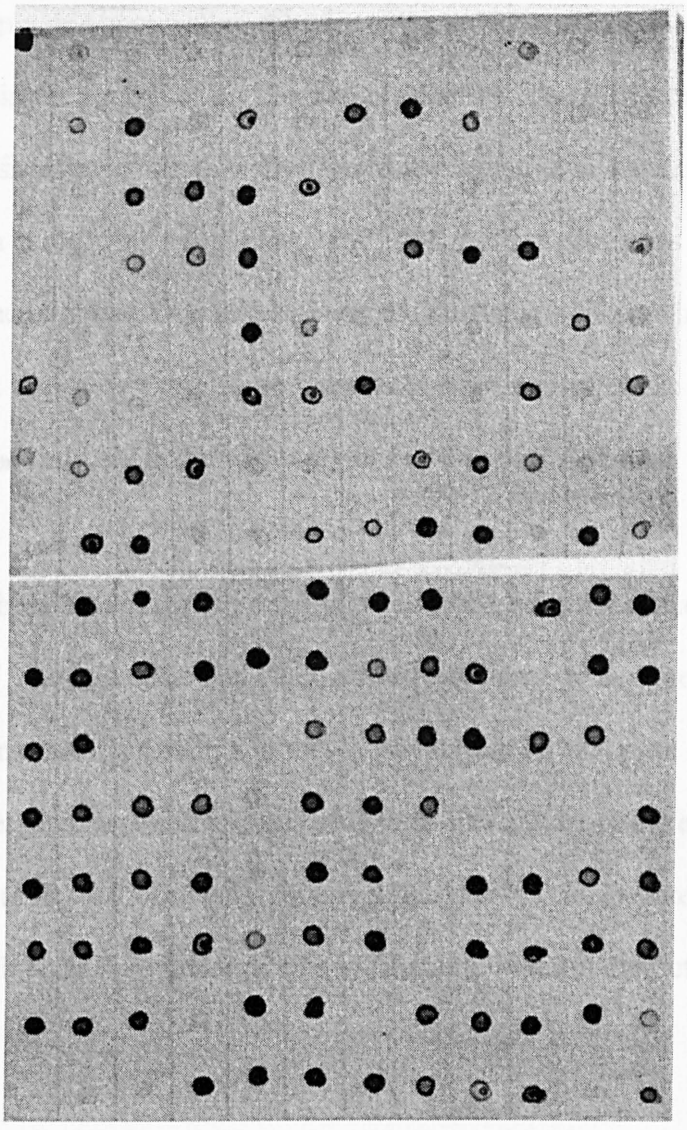


Figure 2-1: Example of a pair of dot blots detecting polymorphism at codon 59 of *dhfr*. The presence or absence of point mutations and relative amounts of signal in each of the samples can be determined by eye or quantitation of digitised signal.

We chose to apply this approach to samples collected in northern Tanzania to evaluate the resistance status of the parasite population. The earliest reports of emerging SP resistance in Africa were from Muheza district in Tanzania during 1994 and 1995 (Ronn et al. 1996; Trigg et al. 1997). SP was introduced as first line treatment for uncomplicated malaria in Tanzania during 2001 following 18 years of second line use. It is now a priority to know how widespread genetic determinants of SP resistance currently are in the wider regions of northern Tanzania.

To investigate this we have carried out a population-based genetic analysis of *P. falciparum* of N. Pare, S. Pare and Hai districts, which are distinct geographically. There has been no malaria research in these districts since 1965, until recent work showing relatively low levels of transmission with an estimate of the entomological inoculation rate (EIR) at 24 (infective bites/person/year) in Hai (Drakeley et al. 2005), as compared with an EIR in the range of 34-405 in the Muheza district (Ellman et al. 1998).

It is widely understood that people self-treat with antimalarial drugs, which can be freely purchased (McCombie 1996), as a consequence, a proportion of people attending health facilities with signs and symptoms of malaria may have had recent prior exposure to drug. In this study we have analysed material from community surveys of asymptomatic infections. This, we believe, will be less subject to bias due to prior drug selection and therefore representative of the parasite population at large.

III Methods and Materials

a The study area and the samples:

Plasmodium falciparum positive samples were collected from two separate studies within the same area of north east Tanzania; one was a study of infants and young children in Hai District and the second was a study of people up to 45yrs in age in the North/South Pare districts. Both studies were cross sectional malariometric surveys across an altitude band of 550-1600m and most of the study subjects were asymptomatic.

For the villages in the Hai District samples were collected in May 2001 by D. Chandramohan. All <5 year-old children from 16 randomly selected villages were invited to attend the survey clinic at a central clinic. A finger prick blood sample for blood slide, and filter paper blood sample was collected from <5 year-old children. The filter paper blood samples were air dried and stored at 4°C with desiccant. Bloodspots from blood film positive children were selected retrospectively for genotyping.

Samples from the North and South Pare mountains were collected in November 2001 during malariometric cross sectional surveys by C Drakeley and F Mosha. A random sample of 1250 individuals (250 per village) under 45 years of age were recruited and a finger blood sample taken into an EDTA microtainer. Filter paper blood spots were made with 10μl of packed cells from samples of individuals found to be parasite positive.

b DNA extraction:

DNA extraction from bloodspots on filter paper was carried out in a 96 well plate format. A segment of the bloodspot was first soaked in 0.5% Saponin in 1xPBS overnight, and then washed twice in 1ml 1xPBS. The segment was then boiled for 8 minutes in 100µl PCR quality water plus 50µl 20% chelex suspension in distilled water (pH 9.5).

c PCR amplification of dhfr and dhps:

A 711bp fragment of *dhps* and a 594bp fragment of *dhfr* containing the polymorphic codons were independently amplified by nested PCR in a 96 well plate format. PCR primer sequences and reaction conditions are indicated in Table 2-2. The 25µl PCR reaction mix contained primers at 0.25µM final concentration, 2mM $MgCl_2$, 250µM of each dNTP, 1x Bioline Taq Polymerase. 1µl of template DNA was introduced to outer reactions. 1µl of *dhps* outer PCR product was introduced into a 25µl inner amplification. Aliquots of 1µl of three fold diluted *dhfr* outer PCR product were introduced into a 25µl inner amplification reaction.

d Molecular genotyping of point mutations using SSOP

Final round PCR products were heat denatured (95°C for 2 minutes), cooled and then spotted onto nylon membranes in 1µl volumes in a 12x8 grid. A panel of 4 PCR samples of known sequence representing all common sequence variants was spotted on every blot to act as positive/negative controls for probe specificity. Replicate blots were made of each array so that probing with the full panel of oligonucleotide probes

for that gene could be conducted simultaneously. After drying, cross-linking was performed with 1200 J ultraviolet light. Sequence specific 18bp oligonucleotide probes 3'-end labelled with dioxigenin (Roche Boehinger Mannheim, Germany) were each designed to compliment the known sequence polymorphisms in *dhfr* and *dhps* listed in Table 2-1. SNP specific hybridisation was followed by high stringency TMAC washes and detection of DIG labelled probes using Alkaline Phosphatase conjugated Anti-DIG Fab fragments (Roche Boehinger Mannheim, Germany) as described in (Conway et al. 1999). Visualisation was performed through the alkaline phosphatase catalysed breakdown of the fluorogenic substrate ECF (Amersham Pharmacia Biotech, UK) and scanned on the Molecular Dynamics Storm 840 Phosphoimager (Amersham Pharmacia Biotech, UK).

e Scoring

We scored the presence, absence or relative abundance of the variant sequence polymorphism at each site separately. Images of blots probed with variant sequences for a single locus were transferred as .tif files to ImageMaster Total Lab (Amersham Pharmacia Biotech, UK). In the 'array analysis' subsection of the software a standard area of each spot was defined and the intensity of chemifluorescence in that area measured. Background was adjusted for by subtraction of the volume of the negative controls from the volume data. Thus volume of chemifluorescence for each spot was calculated as Volume = (MaxIntensity X SpotArea) – Background. To determine the threshold of detection per se the presence flagging option was employed. By this method the faintest spot considered present and not background was selected to set the flagging threshold value.

	Primer Sequence	PCR reaction conditions
DHFR Outer 650bp	M1 5' TTTATGATGGAACAAGTCTGC 3' M7 5' CTAGTATATACATCGCTAACA 3'	94°C for 3 min 94°C x 1 min, 52°C x 2 min, 72°C x 1 min, 40X 72°C for 10 min
Inner 594bp	M3b 5'TGATGGAACAAGTCTGCGACGTT 3' M9 5' CTGGAAAAAATACATCACATTCATATG 3'	94°C for 3 min 94°C x 1 min, 44°C x 2 min, 72°C x 1 min, 4X 94°C x 1 min, 44°C x 1 min, 72°C x 1 min, 34X 72°C for 10 min
DHPS Outer 770bp	N1 5'GATTCTTTTTCAGATGGAGG 3' N2 5' TTCCTCATGTAATTCATCTGA 3'	94°C for 3 min 94°C x 1 min, 51°C x 2 min, 72°C x 1 min, 40X 72°C for 10 min
Inner 711bp	R2 5'AACCTAAACGTGCTGTTCAA 3' R/ 5' AATTGTGTGATTTGTCCACAA 3'	As for Outer

Table 2-2: Table of PCR primer sequences and reaction conditions for the nested amplification of *dhfr* and *dhps*

To compare SNP specific probes at a single site Microsoft Excel bar charts were drawn comparing the volume data for each probe on every sample and the presence flagging result in each case. The following rules were used to determine whether a SNP was present or absent at each site.

a) A SNP was considered present in a PCR product when volume value with a particular probe is greater than background. If volume values were low, presence flagging provided an internal control to avoid possible biasing between probes or blots.

b) Absent: All volume values below the first gridline on the chart were rejected. No set value can be given for this criterion as volume value comparison is relative and varies depending on the strength of the probe labelling and binding.

Samples were categorised into single, majority or mixed at each site as follows: Samples were considered to be of mixed haplotypes if the volume value of the minority SNP was greater than half the volume value of the majority SNP. Samples were considered to be mixed but containing a majority SNP if the minority SNP was less than half of the majority value, but greater than the first gridline on the chart. Samples were considered to be single if there was only one SNP present at a site given the above rules.

To combine data from all sites in a gene and construct haplotypes it was necessary to discard samples in which a mixture was found, but which did not contain a majority SNP. Thus for the purpose of generating frequency data one haplotype was scored

from each sample, this was of either a single or majority type, because haplotypes from mixed infections cannot be constructed.

Following the same principle, for measuring the frequency of two locus genotypes a subset of samples in which a single or majority allelic haplotype was found at both *dhfr* and *dhps* was used.

f Statistical Analysis

Statistical analysis of population differences in haplotype frequencies (Wright's F_{ST}) (Cockerham and Weir 1984) and linkage disequilibrium were carried out using Arlequin software (Schneider, Roessli, and Excoffier 2000). Statistical analysis of contingency tables of the association of haplotypes within two locus combinations was performed using a Chi Squared test.

IV Results and Interpretation.

Of the 165 bloodspots that yielded PCR products, 10.3% were mixed at *dhps* and 1.2% mixed at *dhfr*, no 'majority' haplotypes were identified. The low number of mixed infections was a reflection of the low level of malaria transmission in the three regions. On stratification of the populations of N. Pare and S. Pare into the age ranges of 0-4yrs and 5-45yrs, no significant difference was found in the frequency of *dhfr* and *dhps* allelic haplotypes, allowing comparisons to be made between the samples from those sites with those from the Hai district where samples were taken exclusively from <5yrs.

Using single genotype infections we were able to determine the point mutation haplotypes that were present in the three districts surveyed. The *dhfr* and *dhps* allelic haplotypes present in each region are shown in Figure 2-2. Three point mutations in *dhfr* were found, occurring at codons 51, 59 and 108. Of the eight possible haplotypic conformations of these three SNPs, five were found. This represents only a subset of point mutations reported globally, but matches well with previous reports of point mutations present in single genotype infections that have been described elsewhere in East Africa, namely Kenya and Tanzania (Wang et al. 1997a; Nzila et al. 2000a). Likewise, at *dhps* three point mutations were found and of the eight possible haplotypes only three were found, namely the sensitive allele, single mutant allele S436A and the double mutant allele A437G, K540E haplotype, which has been widely recorded in East Africa (Jelinek et al. 1997; Wang et al. 1997a; Jelinek et al. 1998; Nzila et al. 2000a; Nzila et al. 2000b; Mutabingwa et al. 2001; Kublin et al. 2002) but not so far in West Africa (Plowe et al. 1997; Wang et al. 1997a; Eberl et al. 2001).

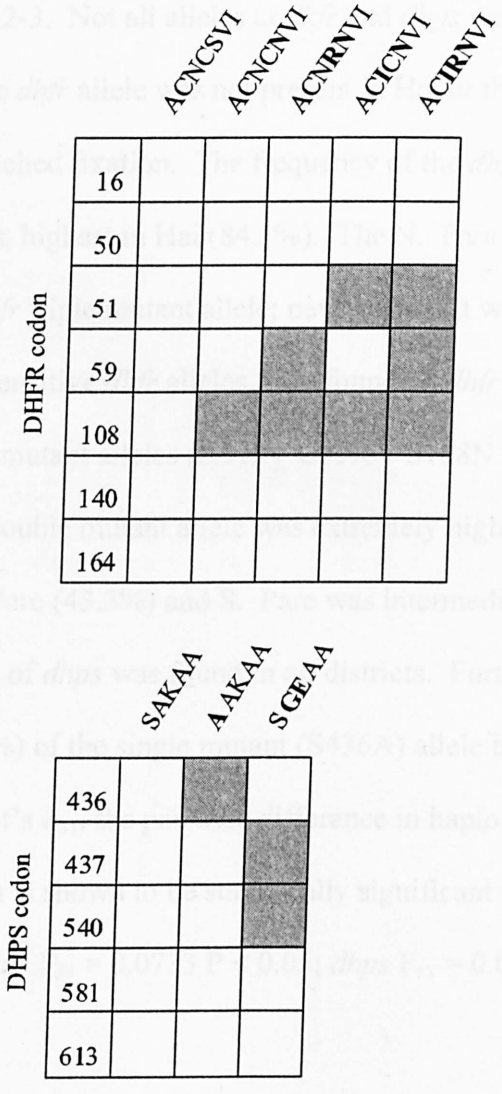
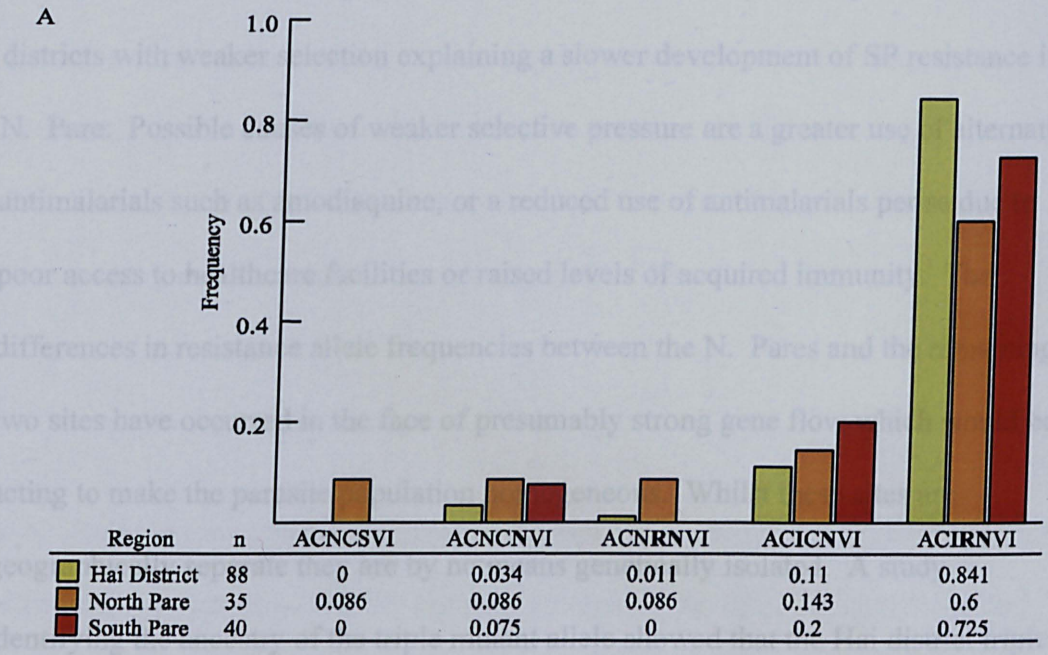


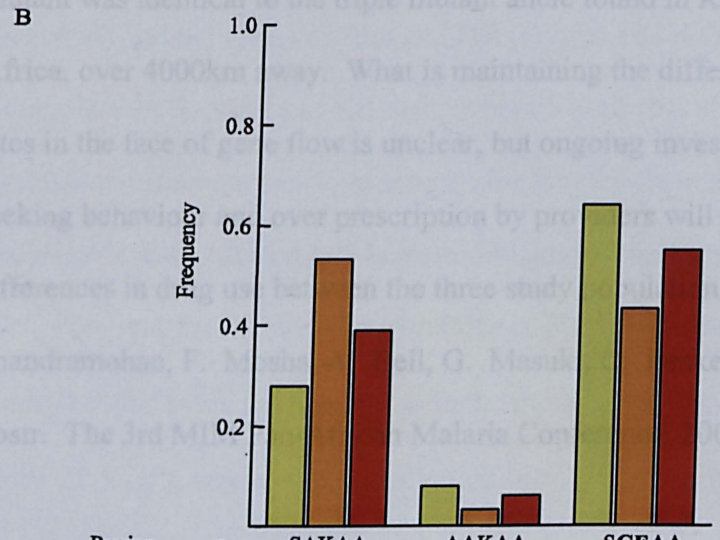
Figure 2-2: Scheme of *dhfr* and *dhps* alleles found in this study. Shaded blocks indicate the site that the constituent SNPs of each haplotype arise. Names of the alleles are composed of the amino acids present at each of the sites, in consecutive order of the codon number, described as having a role in SP resistance

a Regional Variation in Population Resistance

Allelic haplotype frequencies in the three regions differed significantly and are summarised in Figure 2-3. Not all alleles at *dhfr* and *dhps* were present in each of the districts. The sensitive *dhfr* allele was not present in Hai or the S. Pares, the S108N substitution having reached fixation. The frequency of the *dhfr* triple mutant allele was high in all districts, highest in Hai (84.1%). The N. Pare sample set had the lowest frequency of *dhfr* triple mutant allele; nevertheless it was five times greater than the frequency of sensitive *dhfr* alleles. We found no *dhfr* I164L at all and therefore no quadruple mutant alleles (N51I + C59R + S108N + I164L). The frequency of the *dhps* double mutant allele was extremely high being greatest in Hai (64.2%), lowest in N. Pare (43.3%) and S. Pare was intermediate (54.5%). Unlike *dhfr*, the sensitive allele of *dhps* was found in all districts. Furthermore there was a low frequency (4%-7.4%) of the single mutant (S436A) allele in all districts. On calculation of the Wright's F_{ST}, the pairwise difference in haplotype frequencies at both loci between region is shown to be statistically significant when comparing Hai District with N. Pare (*dhfr* $F_{ST} = 0.0733$ $P < 0.05$; *dhps* $F_{ST} = 0.0748$ $P < 0.05$).



Region	n	ACNCSVI	ACNCNVI	ACNRNVI	ACICNVI	ACIRNVI
Hai District	88	0	0.034	0.011	0.11	0.841
North Pare	35	0.086	0.086	0.086	0.143	0.6
South Pare	40	0	0.075	0	0.2	0.725



Region	n	SAKAA	AAKAA	SGEAA
Hai District	81	0.284	0.074	0.642
North Pare	30	0.533	0.033	0.433
South Pare	33	0.394	0.061	0.545

Figure 2-3: Frequency of a) *dhfr* alleles and b) *dhps* alleles found in the three districts of Hai, North Pare, and South Pare

Interestingly there was no significant difference in the remaining pairwise comparisons, despite the differences in haplotype frequencies between North and South Pares. These findings point to some difference in selection pressure between districts with weaker selection explaining a slower development of SP resistance in N. Pare. Possible causes of weaker selective pressure are a greater use of alternative antimalarials such as amodiaquine, or a reduced use of antimalarials per se due to poor access to healthcare facilities or raised levels of acquired immunity. The differences in resistance allele frequencies between the N. Pares and the remaining two sites have occurred in the face of presumably strong gene flow which would be acting to make the parasite population homogeneous. Whilst these sites are geographically separate they are by no means genetically isolated. A study identifying the ancestry of the triple mutant allele showed that the Hai district triple mutant was identical to the triple mutant allele found in KwaZulu Natal, South Africa, over 4000km away. What is maintaining the differences between the study sites in the face of gene flow is unclear, but ongoing investigations into treatment seeking behaviour and over prescription by providers will aim to further illuminate differences in drug use between the three study populations (Swarthout, T., D. Chandramohan, F. Mosha, A. Bell, G. Masuki, C. Drakeley, and H. Reyburn. Abstr. The 3rd MIM Pan-African Malaria Conference, 2002).

During an in vivo study in Malawi, a statistical association was found between the presence of all three *dhfr* mutations and both *dhps* mutations with failure to clear parasitaemia after SP treatment (Kublin et al. 2002). It is probable that in many cases this is due to the presence of the two most highly resistant alleles, triple *dhfr* and the double *dhps* – and this is supported by analysis of recrudescent infections following SP treatment (Kun et al. 1999). Two locus combinations were derived from 'single'

infections where only one haplotype was recorded at both *dhfr* and at *dhps*. The map in Figure 2-4 shows the spatial distribution of frequencies of two locus combinations. The reduction in sample size 'n' reflects the loss due to mixed infections. Here we directly measured the frequency of the highly resistant two locus genotype consisting of the ACIRNVI / SGEAA and found it to be extreme. In Hai it was 63.2%, in S Pare 50%, and in N Pare 22%. Hence there is a clear and testable prediction that SP treatment failure rates in these three regions will differ.

b Selection on dhfr and dhps by SP use

In N. Pare there was more diversity at both loci and consequently a greater diversity of two locus genotypes. The Wright's F_{ST} comparing the three populations at both loci further confirmed this difference between N. Pare and S. Pare (F_{ST} = 0.0583 p=0.0054) and between N. Pare and Hai (F_{ST} = 0.149 p<0.00001). Pair-wise comparison of S. Pare and Hai district populations (F_{ST} = 0.011) showed no significant difference and were merged for subsequent analysis. Statistical analysis of observed and expected two locus combinations was performed on the population of N. Pare and the combined populations of Hai District and S. Pare. We found a significant departure from expected in the merged Hai /S. Pare population (χ^2 test p=0.0018, 6d.f.), whereas the distribution in N. Pare was non-significant (χ^2 test p=0.835, 8d.f.). Linkage disequilibrium analysis was performed on the combined Hai and S. Pare data set, and we found three two locus combinations to be in linkage disequilibrium, namely ACICNVI- SAKAA (D'=0.277, p=0.01), ACNRNVI-AAKAA (D'=1.0, p<0.00001) and ACIRNVI-SGEAA (D'=0.229, p=0.031). No other pair of alleles was found to be in LD.

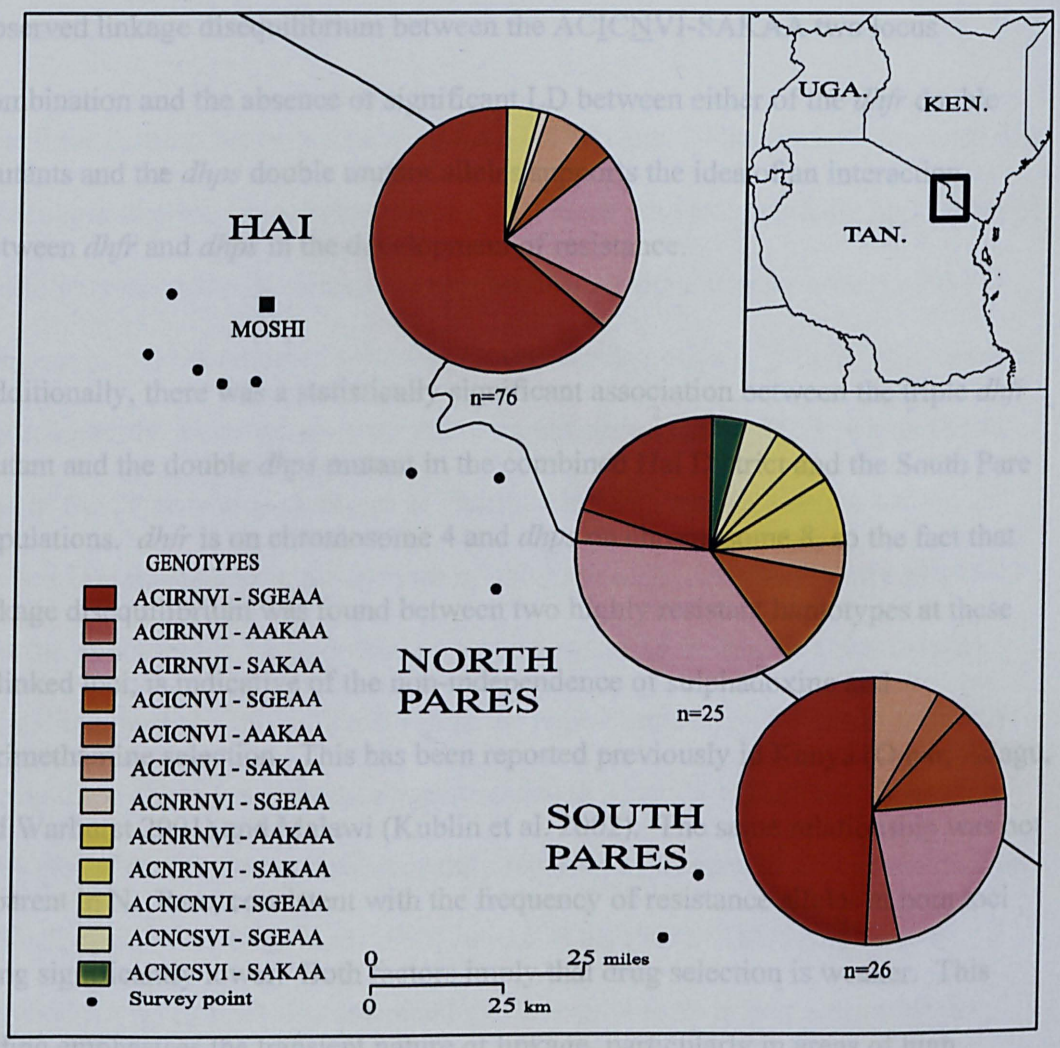


Figure 2-4: Map of the three districts of northern Tanzania showing the frequencies and distribution of *dhfr*/*dhps* two locus genotypes

Within populations in Africa it has been observed that *dhfr* mutations appear earlier in the development of SP resistance (Mberu et al. 2000; Nzila et al. 2000a; Sibley et al. 2001). Our own data suggests the same sequence of events, with resistant *dhfr* being fixed in Hai and South Pare yet sensitive *dhps* relatively common. The observed linkage disequilibrium between the ACICNVI-SAKAA two locus combination and the absence of significant LD between either of the *dhfr* double mutants and the *dhps* double mutant alleles supports the idea of an interaction, between *dhfr* and *dhps* in the development of resistance.

Additionally, there was a statistically significant association between the triple *dhfr* mutant and the double *dhps* mutant in the combined Hai District and the South Pare populations. *dhfr* is on chromosome 4 and *dhps* on chromosome 8, so the fact that linkage disequilibrium was found between two highly resistant haplotypes at these unlinked loci, is indicative of the non-independence of sulphadoxine and pyrimethamine selection. This has been reported previously in Kenya (Omar, Adagu, and Warhurst 2001) and Malawi (Kublin et al. 2002). The same relationship was not apparent in N. Pare, consistent with the frequency of resistance alleles at both loci being significantly lower. Both factors imply that drug selection is weaker. This finding emphasises the transient nature of linkage, particularly in areas of high transmission intensity where recombination rapidly breaks down the linkage between *dhfr* and *dhps*, and argues for caution in use of 'indicator' mutations as a proxy for resistance genotyping. While the co-occurrence of all five mutations in an infection in S. Pare and Hai was in fact a reliable indicator of the presence of a quintuple genotype, this was not the case in N. Pare. The wide spread use of the antibiotic Septrin which contains trimethoprim and sulphamethoxazole to treat other infections

may indirectly select on *dhfr* and *dhps* resistance mutations and further complicate the relationship.

c Past, present and future of Pyrimethamine and Sulphadoxine use in Northern Tanzania

The three districts described here are found at between 200km and 400km northwest of Muheza district, a region historically associated with exceptionally high levels of antifolate resistance. Resistance to pyrimethamine alone was reported in 1954 in Mngeza in Muheza district, following mass administration of prophylactic doses of pyrimethamine monotherapy over a five month period during 1953 (Clyde 1954). Use of the SP combination began in Tanzania in 1982 when it was introduced as second-line treatment for use in cases of chloroquine failure and as early as 1984 it was the policy of the Muheza district hospital to use SP as first line antimalarial (Mutabingwa et al. 2001). Studies from the region report that SP was highly effective during the eighties but resistance was recorded in Magoda village near Muheza in 1994 (Ronn et al. 1996) and subsequently reported in villages in the surrounding area (Jelinek et al. 1997; Trigg et al. 1997; Jelinek et al. 1998). The emergence of resistance to SP in 1994 was attributed by Ronn et al to be in part a result of the prophylactic intervention of weekly dapsone pyrimethamine to all children <10 years old. Parasitological failure rates in children 7 days post treatment with SP in Muheza district hospital was most recently reported to be as high as 45% (Mutabingwa et al. 2001).

Two explanations for the high frequency of resistance alleles in the three districts described here are the widespread use of SP or related drugs, and the movement of resistance from Muheza. However levels of resistance do not show a simple decline

with distance from Muheza to Hai (Fig. 2-4), and clearly local drug use is a very important factor. What is striking from the data presented here is the significant inter-population differences, and it is probable these differences have arisen from differing patterns of drug use in these communities. The relationships between populations to determine whether they are genetically distinct or homogeneous due to gene flow can be elucidated by analysis of marker loci such as microsatellites not subject to drug selection. Unfortunately these samples were of poor quality and it was not possible to perform this additional step. However this approach is applied in the next chapter.

V Concluding Remarks

In this chapter we have described a new approach by which to determine the frequency of point mutation haplotypes in *P. falciparum* populations using blood survey material. It allows quantitation of resistance at the population level and enables direct comparison of population resistance levels even when they differ widely in the proportion of multiply infected individuals. The issue of multiple infections can be problematic when genotyping blood stage parasites, because it causes haplotypic conformations of point mutations to be obscured and rare mutations to be over-sampled. By recording one genotype per infection and discounting minority genotypes we avoid over-sampling of rare genotypes, and estimate the frequency of mutation haplotypes in the population in a measure which is standardised over all populations of different transmission intensity. With this consideration in mind, the SSOP method employed is designed for high throughput screening of blood stage infections, to derive haplotype frequencies from survey material.

Modifications that we made to the original methodology have improved and increased the throughput of the technique. A further modification made after this work was performed was the use of pre-labelled probe sequences. These are labelled and then purified which ensures that only labelled probe sequence is present. The quantity of labelled probe used in the experiment is then standardised between probes, whereas previously unlabelled probe was also present causing variation in the intensity of replicate blots screened with the different probes.

In the published work describing these findings (Pearce et al, 2003; Appendix 8) we recommended that in vivo studies be performed to confirm that the high frequency of resistance alleles is indicative of treatment failure as would be predicted by studies elsewhere in Africa. Surprisingly,in vivo studies of samples taken at a nearby site in Moshi, northern Tanzania (Alifrangis et al. 2005), identified a rate of only 20% parasitological 14 – 28 day treatment failure despite a frequency of 56% for the *dhfr* triple mutant – *dhps* double mutant combination, similar to that reported for the Hai district here (63.2%) (C. Drakeley, Personal Communication). Work is currently underway to establish the role of clinical immunity in reconciling this difference between expected and observed treatment failure rates.

We have shown that heterogeneity exists in the frequencies of resistance alleles between populations in northern Tanzania. We postulate that this heterogeneity exists despite extensive gene flow and we address this question in the following chapter.

Chapter 3 Population Differentiation between Three *Plasmodium falciparum* Populations in Southeast Africa at Positively Selected and Selectively Neutral Markers

I Abstract

We identify significant heterogeneity in the frequency of resistance alleles typed at sulphadoxine/pyrimethamine resistance genes, dihydrofolate reductase and dihydropteroate synthetase, in three southeast African populations; Tanzania, Mozambique and South Africa. There are only a small number of alleles present in each of the populations, five at *dhfr* and three at *dhps* and these were found throughout the region. Within countries we observed significant differences in the allele frequencies between sites, particularly in the frequency of the two most highly resistant alleles; the *dhfr* N51I+C59R+S108N triple mutant and the *dhps* A437G+K540E double mutant. The differences between countries in resistance allele frequencies broadly conforms to our expectations given the known selection histories, with the exception of very high resistance allele frequencies in the supposedly drug naïve Mozambican populations. However, using linkage disequilibrium between the *dhfr* triple mutant and *dhps* double mutant allele as a proxy for ongoing selection, casts doubt on whether selection was actually occurring in Mozambique at the time of sampling. Gene flow homogenises allele frequencies across a region and may explain the high frequencies of resistance alleles in Mozambique. Using eight selectively neutral microsatellites markers we show that there is no population differentiation in the region despite samples being taken over a 1250mile long area. In South Africa we identify a number of epidemic expansions, particularly in the population of Steenbok that may be the cause or a result of the notably high frequencies of the *dhfr*

triple mutant and the *dhps* double mutant allele compared to the other South African populations.

II Introduction

Gene flow is a homogenising force acting to equilibrate allele frequencies between populations. Positive directional selection, such as operating during selection of drug resistance alleles, acts to increase to fixation the frequency of a favourable allele in a population, and if selection is absent in a second population this generates heterogeneity in allele frequencies between them. It is known that in southeast Africa each of the SP resistance alleles comprised of two or more mutations have arisen once de novo (with the exception of the *dhfr* double mutant N51I+S108N allele, which has two separate ancestries) and have spread widely (Roper et al 2003); in the case of the *dhfr* triple mutant (N51I+C59R+S108N) allele the spread includes migration from a single origin in southeast Asia (Roper et al 2004). Clearly this shows that whilst within populations drug selection is increasing the frequency of the resistance alleles, gene flow is moving these alleles between populations.

The importance of gene flow and understanding its extent is two fold. Firstly, the spread of drug resistance is a public health policy concern. If the parasite populations in southeast Africa are freely mixing with no regard for 'nationality' and border control, decisions taken by policy makers on which antimalarial to use would be best made on an international basis rather than nationally. If the parasite population in the region is panmictic, using an antimalarial already abandoned by your national neighbour will become a false economy as the useful therapeutic life of the drug will arguably be short. However, if the parasite populations are more discrete and not panmictic, the spread of resistance may be prevented by well applied control measures. By looking at the extent of gene flow in south east Africa we can assess how easily the resistance alleles have become disseminated.

The second reason to determine the extent of gene flow between populations will become more apparent in subsequent chapters where we describe the effects of selection on the chromosome sequence flanking the selected site. Whether the parasite population in southeast Africa is panmictic or not has implications for this and will be discussed where appropriate.

In this chapter we address the question of how much gene flow occurs between three countries in southeast Africa with different drug selection histories. In South Africa malaria is a problem limited to north of the country and different provinces have taken different approaches to antimalarial treatment policy. The northeast province of Mpumalanga, where the South African samples were taken, began using SP as first line treatment of *P. falciparum* malaria infections in 1997.

The samples taken in Tanzania are from the south of the country, specifically from three 'census' regions: Morogoro, Rufigi and Kilombero/Ulanga. Nationally Tanzania only began using SP in 2001 following 18 years of second line usage, although as outlined in the previous chapter there is a longer selection history in the northern regions.

Finally, at the time of sampling in 2001, Mozambique had not officially adopted SP as first or second line antimalarial. The only recorded exception is the brief use of SP as first line antimalarial in Maputo after flooding in the region in 2000 (Alifrangis et al. 2003). The samples taken in Mozambique were specifically from populations in the south of the country close to the South Africa border. As would be expected, this

heterogeneity between populations in drug selection history has generated significant differences in the frequencies of the SP resistance alleles between the countries.

The geographical locations of the sample collection sites in these three countries allows us to observe gene flow over short distances (South Africa to Mozambique ~75 miles) and over considerably larger distances (South Africa to Tanzania >1250 miles). It is possible that population differentiation may correlate with distance as we would expect if parasite populations were discrete.

In order to perform the population differentiation we have genotyped unlinked selectively neutral microsatellite markers in a cross-section of the parasite populations. Microsatellites are short polymorphic tandem repeat sequences that change in size through strand slippage increasing or decreasing the number of repeats (Kruglyak et al. 1998). As they can be considered selectively neutral, microsatellites are informative of the background level of genetic drift as it is assumed that any difference between two populations in the distributions of alleles is not due to selection acting on the locus.

Microsatellite loci in *P. falciparum* sampled from Africa tend to be highly polymorphic in the population (Anderson et al. 1999; Anderson et al. 2000a). Using simple assumption free parameters such as Wrights fixation index (F_{ST}) to determine population differentiation can result in an underestimation if applied to highly polymorphic markers (Hedrick 1999). F_{ST} is a descriptor of the proportion of genetic variation found between populations proportional to that within each population, such that a low F_{ST} value indicates little difference in allele frequencies between the two populations. F_{ST} are calculated from the expected heterozygosity in the total

population and the weighted expected heterozygosity in the subpopulations. Therefore, if diversity is high these two values can approach unity, regardless of whether the populations have overlapping or non-overlapping sets of alleles (Paetkau et al. 1997; Hedrick 1999). To minimise the effects of any underestimation of population differentiation, we include a test of individual assignment that takes allele identity into account in addition to allele frequency distributions (Pritchard, Stephens, and Donnelly 2000).

III Methods and Materials

a Sample material

Details of the sample used in this study are given in Table 3-1. DNA extraction from bloodspots on filter paper was carried out in a 96 well plate format. A segment of the bloodspot was first soaked in 0.5% Saponin in 1x phosphate buffered saline (PBS) overnight, and then washed twice in 1ml 1xPBS. The segment was then boiled for 8 minutes in 100µl PCR quality water plus 50µl 20% chelex suspension in distilled water (pH 9.5).

b dhfr and dhps point mutation detection

We typed the point mutations that encode amino acid substitutions at codons 51, 59, 108, 164 of *dhfr* and codons 436/437, 540, 581 and 613 of *dhps* using the sequence specific oligonucleotide probing dot blotting technique described in (Pearce et al. 2003) and in the previous section. The point mutations at *dhfr* and *dhps* in samples taken from Mozambique and South Africa were typed by A. Keyser.

Country	District/Province	Date	Collection method	Study Information
South Africa	Mpumalanga	2001	Passive case detection at 3 clinics during the transmission season. Steenbok(n=103) Komatipoort (n=153) Mangweni (n=92)	Southeast African Combination Antimalarial Therapy (SEACAT). http://www.malaria.org.za/Scacat/ Karen Barnes (Uni. of Cape Town) Brian Sharp (Medical Research Council (MRC), Durban) *Molecular technologist* Alana Keyser (MRC, Durban)
Tanzania	Kilombero/Ulanga (n=549)* Rufiji (n=583)* Morogoro (n=245)*	2000	Household Surveys	Interdisciplinary Monitoring Project for Antimalarial Combination Therapy in Tanzania (IMPACT-Tz). http://www.mimcom.org.uk/ifakara/impact.htm Salim Abdulla, Hassan Mshinda, (Ifakara Health Research and Development Centre (IHRDC)) Patrick Kachur, Peter Bloland (Center for Disease Control, Atlanta.) *Molecular technologist:* Allen Malisa (IHRDC, & later Sokoine University)
Mozambique	Maputo Southeast (n= 105) Southwest (n= 128) Peri urban (n= 182)	2001	Active detection Community surveys in: 1.Catuane and Boane (E. Mozambique) 2. BelaVista and Salamanga (W. Mozambique) Two periurban sites Matola Rio and Belulane were included in the resistance allele typing.	LSDI Lubombo Spatial Development Initiative http://www.malaria.org.za/lsdi/home.html Brian Sharp(MRC, Durban) *Molecular technologist:* Alana Keyser (MRC Durban

Table 3-1: Details of samples used in this study. *Number of samples with complete haplotype at either *dhfr* or *dhps*

The Tanzanian samples were typed by A. Malisa. All typing was performed under the supervision of the author. The significance and association of point mutations at these codons with treatment failure are reviewed in previous chapters of this thesis. However, by way of revision, in Africa the highest form of pyrimethamine resistance found common place in field populations is a combination of three substitutions in *dhfr* at codons N51I+C59R+S108N (Wang et al. 1997a; Pearce et al. 2003; Roper et al. 2003; Alifrangis et al. 2005), referred to in this thesis as the triple mutant allele. Combinations of two point mutations occurring at codon 108 with an additional mutation either at codon 51 or 59, are both referred to here as double mutant alleles. At *dhps* the key sulphadoxine resistance allele is composed of substitutions at codons A437G and K540E (Omar, Adagu, and Warhurst 2001; Alifrangis et al. 2003; Pearce et al. 2003; Roper et al. 2003). A combination of three mutations at *dhfr* and two at *dhps* is statistically associated with treatment failure (Kublin et al. 2002).

c *Microsatellite PCR amplifications*

The microsatellites ARAII, G377, PfPk2, Poly-a, TA109, TA42, TA87 and TA102, chosen for use in this study are well described (Anderson et al. 1999; Anderson et al. 2000b) and have been used before in descriptive studies of *P. falciparum* population structure (Anderson et al. 2000a). They are dispersed throughout the genome and therefore will not be biased by linkage to a selective event on any one chromosome.

The primer sequences for the microsatellites are described by (Anderson et al. 1999). The microsatellites were amplified in a semi-nested manner. The primary reaction comprised: 1µl template, 3.0mmol/l Mg^{2+}, 0.75pmol/L primer and 1 unit of *Taq* polymerase. The reaction was cycled as follows: 2 min at 94°C and then 25 repeated

cycles of 30s at 94°C, 30s at 42°C 30s 40°C and 40s at 65°C followed by 2 min at 65°C.

A third fluorescently labelled primer (Applied Biosystems, Warrington, Cheshire, UK) is incorporated into a second round PCR of total volume 11µl containing 2.5mmol/l Mg^{2+}, 2pmol/L primer, 1 unit of Taq polymerase and 1µl of outer nest template. Cycling conditions were: 2min 94°C then 25 cycles of 20s at 94°C 20s at 45°C 30s at 65°C final step of 2 min at 65°C.

Samples were diluted 1 in 100 and run with LIZ-500 size standard on an ABi 3730 genetic analyzer (Applied Biosystems, Warrington, Cheshire, UK). Fragments were sized using the GeneMapper software (Applied Biosystems, Warrington, Cheshire, UK). In the event of a two or more alleles being detected, the majority allele was used if the minority peaks were less than 50% of the height of the majority. If peaks were of equivalent height, the representative allele was allocated at random. The microsatellite typing was performed by the author with the assistance of H. Pota.

d Statistics

Gene diversity values were calculated as $He = [n/(n-1)][1-\sum p_i^2]$ where n is the number of samples and p_i is the frequency of the *i*th allele. The variance of the gene diversity was calculated using Nei and Roychoudhury's formula (Nei and Roychoudhury 1974):

$$Var = \frac{2}{n(n-1)}\{(n-1)[\sum p_i^3 - (\sum p_i^2)^2] + \sum p_i^2 - (\sum p_i^2)^2\}$$

Wright fixation index (F_{ST}) values were calculated in Arlequin (Schneider, Roessli, and Excoffier 2000) and PowerMarker (Liu and Muse 2004). The number of migrants per pairwise comparison calculated using the frequencies of rare alleles (Barton and Slatkin 1986) was performed using Genepop (Raymond and Rousset 1995). The software "Structure" was used in the assignment of individuals to populations in a probabilistic manner (Pritchard, Stephens, and Donnelly 2000). Priors were set as default for the 'admixture' model and burnin and run length were 10^5 and 10^6 Markov chain Monte Carlo (MCMC) iterations respectively. Further details of the software are found below.

The index of association was calculated using the software LIAN (Haubold and Hudson 2000). The distance measure 1-*Ps* (Bowcock et al. 1994) was calculated using the software Microsat (Minch et al. 1997). Neighbour joining trees were calculated using the PHYLIP v3.6 (Felsenstein 2004) package "neighbor". Consensus trees were constructed using "consense" (Felsenstein 2004) and visualised using the software TreeView (Page 1996).

IV Results and Interpretation

a Frequency of dhfr and dhps alleles in each country

The point mutations present at *dhfr* and *dhps* in parasites sampled from three southeast African countries were typed in each of the three countries using the sequence specific oligonucleotide probing (SSOP) dot blot hybridisation method as described by (Pearce et al. 2003) and in chapter 2 of this thesis.

		dhfr							*dhps*				
	Locality	CNCS	CNCN	CNRN	CICN	CIRN	Rare alleles*	N=	SAKAA	AAKAA	SGEAA	Rare alleles*	N=
Mozambique	Matola Rio	0.08	0.06	0.16	0.02	0.68	0.00	50	0.70	0.04	0.26	0.00	66
	Beluluane	0.07	0.03	0.17	0.03	0.70	0.00	59	0.68	0.06	0.25	0.01	68
	BelaVista	0.14	0.00	0.17	0.00	0.69	0.00	36	0.79	0.00	0.21	0.00	42
	Salamanga	0.12	0.00	0.37	0.05	0.46	0.00	43	0.75	0.04	0.19	0.02	47
	Catuane	0.15	0.02	0.22	0.07	0.54	0.00	46	0.80	0.04	0.14	0.02	50
	Boane	0.29	0.00	0.19	0.00	0.52	0.00	31	0.75	0.00	0.22	0.03	32
	Average	0.14	0.02	0.21	0.03	0.60	0.00		0.75	0.03	0.21	0.01	
Mpumalanga	Steenbok	0.19	0.02	0.16	0.00	0.63	0.00	83	0.73	0.03	0.22	0.01	98
	Koomatipoort	0.28	0.02	0.11	0.06	0.53	0.00	104	0.90	0.07	0.02	0.02	122
	Mangweni	0.35	0.03	0.16	0.06	0.41	0.00	69	0.85	0.08	0.07	0.00	89
	Average	0.27	0.02	0.14	0.04	0.52	0.00		0.83	0.06	0.10	0.01	
Tanzania	Morogoro	0.43	0.00	0.16	0.09	0.32	0.01	180	0.56	0.31	0.10	0.03	163
	Rufigi	0.36	0.01	0.12	0.11	0.38	0.02	455	0.74	0.15	0.07	0.03	413
	Ulanga	0.51	0.02	0.07	0.11	0.28	0.01	376	0.68	0.18	0.12	0.02	363
	Average	0.43	0.01	0.12	0.10	0.32	0.01		0.66	0.22	0.10	0.03	

Table 3-2: Allele frequencies of the most common *dhfr* and *dhps* haplotypes. *Rare alleles were defined as having a frequency of less than 5% in the population and were pooled

As observed in the northern Tanzania study we found a limited number of haplotypes at the selected loci, namely a total of 5 common (>5% frequency) haplotypes at *dhfr* and 3 at *dhps*. The frequencies of each haplotype at each site in every country are given in Table 3-2. We detected a number of rare alleles (<5% frequency in population), and these are summed together in the table. We observed most rare alleles in the country with the largest sample size, Tanzania where a further 2 haplotypes of 1 or 2 point mutations at *dhfr* (C59R; N51I+C59R) and a further 6 haplotypes of 1 or 2 point mutations at *dhps* were seen (S436F; S436C; A437G; K540E; A581G; S436F+K540E). In both Mozambique and South Africa there were no additional haplotypes at *dhfr* and only a further 2 at *dhps* (A437G; S436A+A437G in Mozambique S346F; S436A + A437G + K540E in South Africa). We did not detect a point mutation at codon 164 of *dhfr* and thus the most highly pyrimethamine resistant mutant, comprised of four substitutions, was not observed in any of the three countries.

b Between country differentiation through allele frequencies at dhfr and dhps

Of the common *dhfr* and *dhps* haplotypes, the frequencies were broadly similar across the 3 countries (Table 3-2). Surprisingly, the frequency of the allele conferring the greatest pyrimethamine resistance the *dhfr* triple mutant (N51I+C59R+S108N) allele was highest in Mozambique (0.60) compared to 0.52 in South Africa and 0.32 in Tanzania. Similarly, the most highly resistant *dhps* allele, the double mutant (A437G+K540E) was highest in Mozambique (0.21), a frequency double that observed in Tanzania and South Africa (both 0.10).

		dhps		
		Mozambique	Mpumalanga	Tanzania
dhfr	Mozambique	-	0.030	0.041
	Mpumalanga	0.022	-	0.040
	Tanzania	0.120	0.048	-

Table 3-3: Between population F_{ST} analysis at *dhps* (above the diagonal), and *dhfr* (below the diagonal). Cells shaded in light grey are significant at the p = 0.05 level

In Tanzania, the generally lower frequencies of these particular resistance alleles, in addition to the greater number of alleles at *dhfr* and *dhps*, was more in line with our expectations, given the short SP selection history. Summarizing the allele frequencies, we observed that the resistance levels of the three countries could be ranked thus: Tanzania <Mpumalanga <Mozambique. To determine whether there was any significance to the differences between countries, the frequencies of *dhfr* and *dhps* alleles were compared.

Table 3-3 show the Wright's fixation index and significance values for these pairwise comparisons. The F_{ST} values for comparisons with Tanzania were highest, particularly so at *dhfr* (Tanzania versus Mozambique F_{ST}=0.12, p<0.0001), probably due to the three times higher frequency of the sensitive *dhfr* alleles in Tanzania.

c Sub -population differentiation through allele frequencies at dhfr and dhps

Sites within countries were compared through pairwise F_{ST} and the results are shown in Table 3-4. Within Mozambique the *dhps* allele frequencies at all sites were not significantly different, while at *dhfr* the Salamanga subpopulation differed

significantly from three others (with the exception of Catuane). Inspection of Table 3-2 reveals that an unusually high frequency of the C59R + S108N double mutant allele is common to Salamanga and Catuane.

Comparing South African sites,the subpopulation of St eenbok was significantly different from Mangweni and Komatipoort at *dhps*. The frequency of the double mutant *dhps* allele in the subpopulation of Steenbok was 0.22 compared to 0.07, and 0.02 in Mangweni and Komatipoort respectively. At *dhfr* Steenbok differed significantly from Mangweni, the former having a higher frequency of the triple mutant allele and greatly reduced frequency of sensitive. Inter-site differences at *dhfr* and *dhps* are consistent with Steenbok having the most resistant alleles in both genes.

Finally in Tanzania, the subpopulation of Morogoro was significantly different having the single mutant S436A allele at a frequency of twice that found in Rufigi or Kilombero/Ulanga. The high frequency of this allele is curious, since the alanine substitution at codon 436 is thought to not have a great deal of influence in sulphadoxine resistance as it lacks the phenyl side chain that can cause conformational chains in the enzyme active site (D. Warhurst, Personal Communication) The three sites did not differ at *dhfr*.

Comparing all sub-populations pairwise some interesting observations emerge. Inspection of table 3-4 indicated that broadly there were fewer significant differences among pairwise intra country comparisons versus inter country pairwise comparisons. In particular the Mozambican sample sites were identified as having more significant and moderately high F_{ST} values in pairwise comparisons with South African or Tanzanian locations. Interestingly, one exception to this was the population of

Steenbok, which was similar to the Mozambican subpopulations at both *dhfr* and *dhps*. All pairwise comparisons involving the Tanzania population of Morogoro resulted in high and significant F_{ST} values at *dhps*. This was due to the particularly high frequency of the *dhps* single mutant allele S436A, exceptional in this dataset.

d The dhfr-dhps two locus genotypes

The frequency of the two locus genotypes is given in Appendix 1 and illustrated in Figure 3-1. Tanzania had the highest number of two locus combinations (29) reflecting the high number of alleles present at *dhfr* and *dhps*,but over half of these genotypes had frequencies of less than 5%. By contrast there was an average of eight two locus combinations in the other two countries. The frequency of the most highly resistant combination, triple mutant *dhfr* – double mutant *dhps* was highest in Mozambique, followed by South Africa then Tanzania. The general trend of the F_{ST} analysis was that only the differences in allele frequencies between the Tanzanian subpopulations and those from South Africa or Mozambique were significant for the majority of pairwise comparisons (Table 3-5). Within country pairwise comparisons were broadly non significant.

To look for evidence of recent selection through use of SP we tested for linkage disequilibrium. Steenbok was the only population with significant linkage disequilibrium between the *dhfr* triple mutant allele and the *dhps* double mutant allele (D'=0.8026 p=0.011). This might have arisen because of a higher degree of relatedness between the parasites sampled from this epidemic setting.

		dhps										
		Komatipoort	Steenbok	Mangwe ni	MatolaRio	Beluluane	BelaVista	Salamanga	Catuane	Morogoro	Rufigi	Ulanga
dhfr	Komatipoort	-	0.11	0.00	0.14	0.15	0.10	0.09	0.04	0.22	0.00	0.12
	Steenbok	0.01	-	0.05	-0.01	-0.01	-0.01	-0.01	0.00	0.10	0.04	0.00
	Mangweni	0.00	0.05	-	0.07	0.07	0.04	0.03	0.00	0.13	-0.01	0.06
	Matola Rio	0.04	0.00	0.09	-	-0.01	-0.01	-0.01	0.01	0.08	0.06	0.00
	Beluluane	0.05	0.01	0.11	-0.02	-	0.00	-0.01	0.02	0.07	0.07	-0.01
	BelaVista	0.03	-0.01	0.08	-0.02	-0.02	-	-0.02	-0.01	0.13	0.03	0.02
	Salamanga	0.06	0.05	0.05	0.06	0.06	0.06	-	-0.01	0.08	0.02	0.00
	Catuane	0.01	0.00	0.03	0.01	0.01	0.00	0.00	-	0.10	0.00	0.01
	Morogoro	0.01	0.07	0.00	0.12	0.13	0.11	0.11	0.05	-	0.12	0.02
	Rufigi	-0.01	0.01	0.01	0.03	0.04	0.02	0.04	0.00	0.00	-	0.05
	Ulanga	0.00	0.04	-0.01	0.09	0.10	0.07	0.07	0.03	-0.02	0.00	-

Table 3-4: Within and between population pairwise F_{ST} values for comparisons between the individual sampling locations. Above the highlighted diagonal *dhps*, below the diagonal *dhfr*. Boxed groups of pairwise comparisons indicates within population comparisons indicates within population

To test this, a subset of 60 unrelated individuals (as established by typing of 8 unlinked microsatellite loci – see below) were also tested for LD between dhfr and dhps resistance alleles. We found significant LD between the triple mutant *dhfr* allele and the *dhps* double mutant allele (D'=0.8413 p=0.0011) in this sample of unrelated individuals.

In Mozambique the 3 out of 5 populations had moderately positive D' values between these two alleles (D'= 0.455 to 0.657), but not significantly so. In Tanzania, Rufigi and Ulanga have weakly positive D' values (D'= 0.161 to 0.255) whereas in Morogoro the triple double combination was positively disassociated (D' = -0.533); a reflection of the high frequencies of the S436A substitution in the population. That in only one population were the *dhfr* triple and *dhps* double mutant in linkage disequilibrium is interesting.

The presence of the all five mutations comprising the *dhfr* triple and *dhps* double mutant is statistically associated with treatment failure, and intense use of SP will maintain an association between these sites.

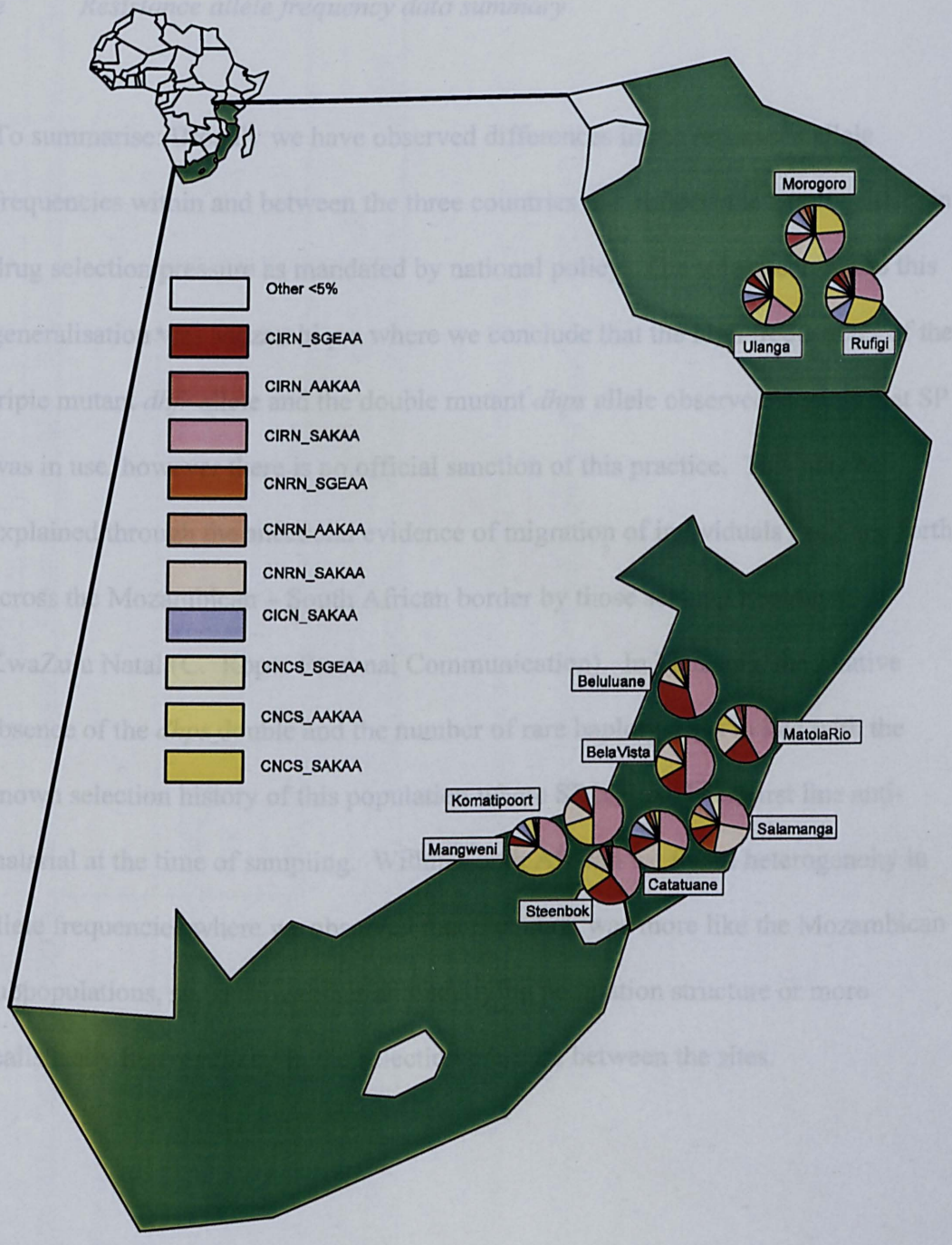


Figure 3-1: Map showing the frequencies of the *dhfr*/*dhps* two locus genotypes of alleles in the approximate location of the population sampled

e Resistance allele frequency data summary

To summarise: Broadly we have observed differences in the resistance allele frequencies within and between the three countries that reflects the heterogeneity in drug selection pressure as mandated by national policy. The sole exception to this generalisation was Mozambique where we conclude that the high frequencies of the triple mutant *dhfr* allele and the double mutant *dhps* allele observed suggest that SP was in use, however there is no official sanction of this practice. This may be explained through the anecdotal evidence of migration of individuals back and forth across the Mozambican – South African border by those seeking treatments in KwaZulu Natal (C. Roper Personal Communication). In Tanzania, the relative absence of the *dhps* double and the number of rare haplotypes is in line with the known selection history of this population where SP had not been first line anti-malarial at the time of sampling. Within South African we found heterogeneity in allele frequencies where we observed that Steenbok was more like the Mozambican subpopulations, suggesting either an underlying population structure or more realistically heterogeneity in the selection pressure between the sites.

		dhps										
		Komatipoort	Steenbok	Mangweni	Catatuane	Beluluane	Salamanga	MatolaRio	BelaVista	Morogoro	Rufigi	Ulanga
dhfr	Komatipoort	-										
	Steenbok	0.01	-									
	Mangweni	0.01	0.02	-								
	Catatuane	0.01	0.01	-0.01	-							
	Beluluane	0.04	0.00	0.08	0.04	-						
	Salamanga	0.03	0.03	0.02	-0.01	0.06	-					
	MatolaRio	0.00	-0.01	0.03	0.00	0.00	0.01	-				
	BelaVista	-0.02	-0.01	0.02	0.01	0.00	0.02	-0.01	-			
	Morogoro	0.06	0.06	0.01	0.01	0.11	0.03	0.06	0.07	-		
	Rufigi	0.04	0.04	0.00	0.00	0.09	0.03	0.05	0.04	0.00	-	
	Ulanga	0.08	0.07	0.02	0.03	0.13	0.07	0.09	0.09	0.01	0.01	-

Table 3-5: Within and between population pairwise comparisons F_{ST} 2 locus combination. Boxed groups of pairwise comparisons indicates within population comparisons. Cells shaded in light grey are significant at the p = 0.05 level

f Population differentiation at neutral polymorphic markers

We postulated that the differences in resistance allele frequencies exist whilst the parasite populations are continuous and overlapping. To test this we selected a subset of samples from a number of sites within Mozambique (n= 96), South Africa (n = 188) and Tanzania (n=475) and genotyped them at eight microsatellite loci (Anderson et al. 1999; Anderson et al. 2000b). The samples were taken regardless of the *dhfr* or *dhps* allele present, thus avoiding any biasing in allele frequencies at the neutral alleles as a result of changes in population structure due to drug selection. Due to sample number limitations the sites within Mozambique were grouped into East (Boane and Catuane; n=48) and West Mozambique (Bela Vista and Salamanga; n=48).

The eight microsatellite loci used are from sites located throughout the genome and unlikely to be biased by a loss of diversity in any one region of the genome. The genomic locations of the microsatellites can be found in table 3-6. Three loci are found on chromosome 12 within a region of over 70cM, sufficient distance apart that they can be assumed to be in linkage equilibrium (Su et al. 1999). This also applies to the two loci present on chromosome 6. One locus, Poly-a, is found on chromosome four, but is sufficiently distant from *dhfr* (>215kb) that any reduction in gene diversity at this locus due to drug selection at *dhfr* is unlikely. The region of reduced diversity described on selected chromosomes in southeast Asian populations only extended as far as ~100kb around *dhfr* (Nair et al. 2003).

Marker	Chromosome	Mpumalanga n=188	Mozambique n=96	Tanzania n=475
ARAII	11	0.875	0.879	0.832
G377	12	0.623	0.573	0.503
PfPk2	12	0.911	0.911	0.896
PolyA	4	0.883	0.845	0.860
TA109	6	0.838	0.818	0.814
TA42	5	0.498	0.507	0.380
TA87	6	0.858	0.845	0.853
TA102	12	0.849	0.809	0.839
Mean	-	0.792	0.773	0.747

Table 3-6: Microsatellite locations and diversity values per population.

The gene diversity of each locus was calculated for each country and is given in table 3-6. In each country the average gene diversity across all markers was high (0.747 - 0.792) and comparable to previous estimates for other African populations as expected for neutral loci (0.76-0.80) (Anderson et al. 2000a). Loci TA42 and G377 had lower gene diversity than the other loci, averaging across the 3 countries at 0.462 and 0.566 respectively. The locus on chromosome 4, Poly-a, has gene diversity values comparable to those for the majority of the loci sampled, namely 0.883 in South Africa, 0.845 in Mozambique and 0.860 in Tanzania. There was no dramatic reduction in the gene diversity at each of the microsatellite markers when the country samples were subdivided into the individual *dhfr* or *dhps* resistance allele classes.

g Population differentiation at the international level at neutral loci

The three populations of Mozambique, South Africa and Tanzania were compared using Wrights fixation index (Table 3-7). In all three pairwise comparisons the F_{ST} values were low, ranging from 0.0087 to 0.0152, indicating little difference in the frequencies of the alleles present in the three populations suggesting that the parasite

population in southeast Africa is panmictic. The pairwise comparisons involving Tanzania generated moderately higher F_{ST} values than that between the two southern countries. However, the 95% confidence intervals on the F_{ST} in the comparison excluding Tanzania overlapped, with those including Tanzania and we rejected the hypothesis of increased differentiation over distance.

The F_{ST} values for the comparison of the three populations are comparable to a prior estimate of F_{ST} between the African countries of Uganda, Congo and Zimbabwe (0.003 to 0.012) as shown by Anderson *et al.* A conclusion of panmixia was reached to explain the relationship between parasites from these countries (Anderson et al. 2000a). We can not directly compare the alleles present in the three previously reported countries with those reported here as the two studies were analysed on different instrumentation and there is no standardised allele size between the two instruments. However, it is likely that if the six nations were to be analysed together we would not find any significant differences, as the locations of the two studies interweave.

	Mpumalanga	Mozambique
M pumalanga	-	
Mozambique	0.0087 (±0. 0072)	-
Tanzania	0.0104 (±0.0 05 5)	0.0152 (±0.0 12)

Table 3-7: Pairwise population differentiation (F_{ST}) using microsatellite allele frequencies. 95% CI given in parenthesis.

In addition to calculating F_{ST} we also determined the degree of population differentiation using Slakin's R_{ST} (Slatkin 1995). This is analogous to F_{ST} but is adapted to use with microsatellites by assuming a high-rate stepwise mutation model (SMM) rather than an infinite-allele mutation model. This assumption of a stepwise mutation model, may not be applicable to a majority of *P. falciparum* microsatellites, that due to complex mutations such as indels and duplications do not follow simple rules of length variation (Anderson et al. 2000b). Of the eight microsatellites used here, SMM is the appropriate model for four, namely TA87, ARAII, G377 and PfPK2 (Anderson et al. 2000b).

Using the software Microsat (Minch et al. 1997), we calculated R_{ST} values between the three populations based on this subset of microsatellites (Table 3-8). The values generated were not significantly different from zero as indicated by the 95%CI calculated from bootstrapping the data exhaustively, supporting a conclusion of panmixia.

	Mpumalanga	Mozambique
Mpumalanga	-	
Mozambique	0.047 (±0.106)	-
Tanzania	0.024 (±0.045)	0.011 (±0.031)

Table 3-8: Population differentiation using Slatkin's R_{ST} on only loci where the stepwise mutation model could be applied. 95% CI given in parenthesis

h Population differentiation at the international level by population assignment

As gene diversity was high at each of the microsatellite markers, the F_{ST} values maybe an underestimation of the actual population differentiation (Hedrick 1999). Therefore, we ran the dataset through a Bayesian probabilistic clustering algorithm implemented by the software Structure (Pritchard, Stephens, and Donnelly 2000). The algorithm addresses whether the current population is structured due to an ancestral admixture event between user-defined 'k' numbers of ancestral populations. The algorithm uses a Markov chain Monte Carlo (MCMC) scheme to allocate individuals to one of the 'k' number of clusters on the basis of similarities in the allele distribution at a combination of loci, generating a likelihood value such that the most probable 'k' can be identified. It also estimates the proportion, Q, of membership of each individual to each cluster, estimating the proportion of an individuals' genome inherited from some ancestral population. For example if k=2 was assumed and the proportion of an individuals genome coming from ancestral population 1 was Q = 0.5 ($=^1/_k$), its' genome will come from each 'k' population with equal proportions. In this particular example such a result repeated over an entire dataset would infer that there was only one population present in the data set rather than the two tested for.

The algorithm performs this clustering individual by individual and ignores all information regarding the sampling populations. Thus we used Structure to determine whether the individuals comprising the datasets of ancestral, double mutant and triple mutant chromosomes would be clustered into populations matching those defined by the allele present at the selected site, *dhfr*.

The algorithm can implement three models to test for subpopulations and determine the population ancestry of individuals, namely the 'no admixture', 'admixture' and 'linkage between sites' models. As the majority of these markers are physically unlinked, we used the admixture model, and ignored the less appropriate no admixture model that assumes absolute population subdivision. The admixture proportions for each individual are independently modelled from a symmetrical Dirichlet distribution with hyper-variable a. If a is high the distribution is random, whereas if a is below 1, the distribution approximates a negative binomial distribution and it models each individual as having originated mostly from a single population, with each population being equally likely.

We tested a number of 'k' clusters from k=1 to 8 to test for between country population differentiation. We found that k=1 was the most likely number of clusters, supporting the F_{ST} estimates and the conclusion of panmixia. At all other k the likelihood was low and the Dirichlet parameter a was above one indicating that each individual has allele copies originating from all countries in equal proportions (Pritchard, Stephens, and Donnelly 2000).

i Population differentiation within countries at neutral loci

We calculated the Wright fixation index between the sites within each country (Table 3-9) and found that there was no significant population differentiation between the various sites, with the one exception of Morogoro versus Rufigi in Tanzania where we observed a significant but low F_{ST} value. R_{ST} estimates of population differentiation, using only the four microsatellites that the SMM could be applied to,

found no significant population differentiation values for within country pairwise comparisons (Table 3-10).

Comparing all sites pairwise across countries significant differences become apparent (Table 3-9). The population of Steenbok stands out as being significantly different from all non South African populations (F_{ST} in the range 0.017- 0.023), although this is not supported by the R_{ST} analysis which finds no significant population differentiation. The South African population of Mangweni is also significantly different from 3 of the five non South African populations; the difference between Mangweni and Morogoro and Rufigi supported by significant, but low R_{ST} values.

The highest significant F_{ST} values observed are those between East Mozambique and the Tanzania populations of Morogoro and Rufigi at 0.032 and 0.033 respectively, yet somewhat lower than the upper range of significant population differentiation observed in the *dhfr* and *dhps* allele frequencies (0.13 *dhfr*; 0.22 *dhps* from Table 3-4).

	Komatipoort	Mangweni	Steenbok	W. Mozambique	E. Mozambique	Ulanga	Morogoro
Komatipoort	-						
Mangweni	-0.005	-					
Steenbok	0.003	0.004	-				
W. Mozambique	-0.001	0.007	0.018	-			
E. Mozambique	0.014	0.018	0.023	0.021	-		
Kilombero/Ulanga	0.006	0.011	0.017	0.020	0.022	-	
Morogoro	0.006	0.016	0.022	0.024	0.032	0.009	-
Rufigi	0.006	0.008	0.023	0.020	0.033	0.006	0.009

Table 3-9: Within and between population pairwise comparisons F_{ST} of microsatellite alleles. Boxed groups of pairwise comparisons indicate within population comparisons. Cells shaded in light grey are significant at the p = 0.05 level.

	Komatipoort	Mangweni	Steenbok	W. Mozambique	E. Mozambique	Ulanga	Morogoro
Komatipoort	-						
Mangweni	-0.004 (±0.014)	-					
Steenbok	0.019 (±0.069)	0.041 (±0.051)	-				
W. Mozambique	0.017 (±0.071)	0.039 (±0.057)	0.086 (±0.21)	-			
E. Mozambique	-0.005 (±0.024)	0.02 (±0.027)	0.069 (±0.159)	-0.003 (±0.020)	-		
Ulanga	-0.004 (±0.008)	0.011 (±0.031)	0.018 (±0.047)	0.058 (±0.125)	0.023 (±0.078)	-	
Morogoro	0.045(±0.084)	0.074 (±0.067)	0.105(±0.202)	-0.004(±0.012)	0.006 (±0.022)	0.051 (±0.114)	-
Rufigi	-0.011 (±0.004)	0.03 (±0.029)	0.027 (±0.076)	0.045 (±0.125)	0.015 (±0.059)	0.004 (±0.006)	0.044 (± 0.082)

Table 3-10: Within and between population pairwise comparisons using Slatkin's R_{ST} of microsatellite alleles for only those that a stepwise mutation model could be applied. 95% CI given in parenthesis.

j Population differentiation at the intra national level by population assignment

Intra-population analyses for all three countries using the software 'Structure' (Pritchard, Stephens, and Donnelly 2000) found that in Tanzania k=1 was the most likely number of ancestral populations, denoting no population differentiation. In Mozambique, k=2 was the most likely, but the Dirichlet parameter a was high, and therefore k=1 was the most robust conclusion. In South Africa the admixture model testing k=3 began to identify individuals that clustered together, however k=1 was the most likely number of clusters. Closer inspection of the individuals clustering in the k=3 run identified the majority as from Steenbok. Testing this population separately, k=3 identified a number of individuals as clustering together,diff erent from the majority of the Steenbokian individuals. However, the variance on the log likelihood for this run was large, casting doubt on the robustness of the k=3 conclusion,and k=1 was determined as the most robust.

Thus in all three countries analysis of population differentiation using structure reached a conclusion of panmixia between all sites. However, the tendency of some individuals in Steenbok to cluster together in k>2 tests, in addition to the relatively high F_{ST} values for pairwise comparisons involving Steenbok and non South Africa populations, suggested that this population warranted further investigation.

k Epidemics in South Africa

One explanation for the observed high F_{ST} values in the comparisons between non South African sites and Steenbok, and to a lesser extent Mangweni (Table 3-9), is that

whilst microsatellite diversity was high in each population, an over representation of one or more alleles in one of the two compared populations,would be sufficient to generate the magnitude of F_{ST} values we have observed.

An over representation of an allele or set of alleles in a population can occur due to the epidemic expansion of a single lineage. This may be detected by looking for evidence of linkage disequilibrium. *Plasmodium falciparum* has a very high recombination rate at 17kb/cM (Su et al. 1999) but this is tempered by the degree of inbreeding in a population, determining the effective recombination rate (Dye and Williams 1997; Conway et al. 1999). In the majority of African populations, malaria transmission is high and a large proportion of infections contain multiple genetically distinct parasite lineages. Linkage disequilibrium distances tend to be small between loci on the same chromosome, and are even less likely to occur between physically unlinked sites, such as the majority of these loci are. However, epidemic expansions of a parasite lineage increase the inbreeding coefficient, reducing the effective recombination rate and can generate linkage disequilibrium between sites.

The Entomological Inoculation Rate (EIR) is a measure of transmission intensity that can be used as an approximate guide to expectations of effective recombination rates in a population, namely that when EIR is high, the inbreeding coefficient is low and the effect recombination rate is high. In Tanzania an EIR of 584 infective bites per adult per annum (Charlwood et al. 1998) has been estimated for a population within the study area. An EIR of this magnitude implies a high effective recombination rate as multiple infections are prevalent. The EIR for Mozambique is substantially lower at 12 infective bites per adult per annum (Hay et al. 2000). Contrast this with the EIR of the South African population, estimated as <1 infective bite per adult per annum

(K. Barnes Personal Communication). Populations such as South Africa are prone to epidemic expansions of parasite lineages as transmission is so low that clinical immunity, gained over successive resolved cases of malaria, develops later in life (in the teens as opposed to in childhood) and as such the prevalence of naïve hosts favours epidemic transmission (Kleinschmidt and Sharp 2001).

The linkage disequilibrium measure called the "index of association (I^S_A)" (Haubold and Hudson 2000) tests for a departure from random associations between the microsatellite loci. The expectation was that where the EIR is high there would be no association between the eight markers. Where the EIR is low the effective recombination rate will be lower and one may expect there may be associations between loci, particularly those on chromosome 12, that would not persist in a higher recombination rate.

In line with this expectation we found that in both the South African populations of Steenbok and Mangweni there was significant departure from a null hypothesis of linkage equilibrium (Table 3-11), suggesting that there is a high degree of inbreeding occurring in this population to maintain linkage between the eight physically unlinked loci.

	Standardized IA	Parametric P
Komatipoort	0.00 8	0.3332
Mangweni	0.02 7	0.0001
Steenbok	0.033	0.0000
W. Mozambique	-0.0 30	1.000
E. Mozambique	-0.005	1.000
KiloUlanga	0.005	0.34 0
Morogoro	0.006	0.28 4
Rufigi	0.01 1	0.088

Table 3-11: Departures from an assumption of linkage equilibrium between the eight microsatellite loci as calculated by LIAN (Haubold and Hudson 2000).

1 Identification of epidemically expanding lineages in South Africa

To identify the relationships of the isolates causing the LD in Steenbok and Mangweni, a neighbour joining tree was constructed using the distance measure of 1-*Ps* (Bowcock et al. 1994) where *Ps* is the proportion of alleles shared in a pairwise comparison of two isolates, calculated using the software "Microsat" (Minch et al. 1997)(Figure 3-2). *Ps* is a measure of similarity between two isolates calculated as the proportion of alleles shared summed over loci/(2 x number of loci compared) (Bowcock et al. 1994). Pairs of samples that have the highest *Ps* values are those that are most alike. The distance file of 100 bootstraps was made and the trees were constructed using "neighbor" and a consensus tree was compiled using the "consense" components of the software collection PHYLIP (Felsenstein 2004).

A node had to be present in greater than 75% of the bootstrapped trees for inclusion into the consensus tree. Inspection of the consensus tree clearly shows that the 3 major countries are not structured as the majority of the samples are equidistant from each other (Figure 3-2). There is no clustering of parasites into groups as defined by the population that they were sampled from, unsurprising given the earlier conclusion of panmixia.

What is notable is that there were a number of independent expanding lineages containing between 2 and 5 isolates. The isolates within these expanding lineages were all from South African sites explaining the LD values in Steenbok and Mangweni. The details of those isolates involved can be found in table 3-12.

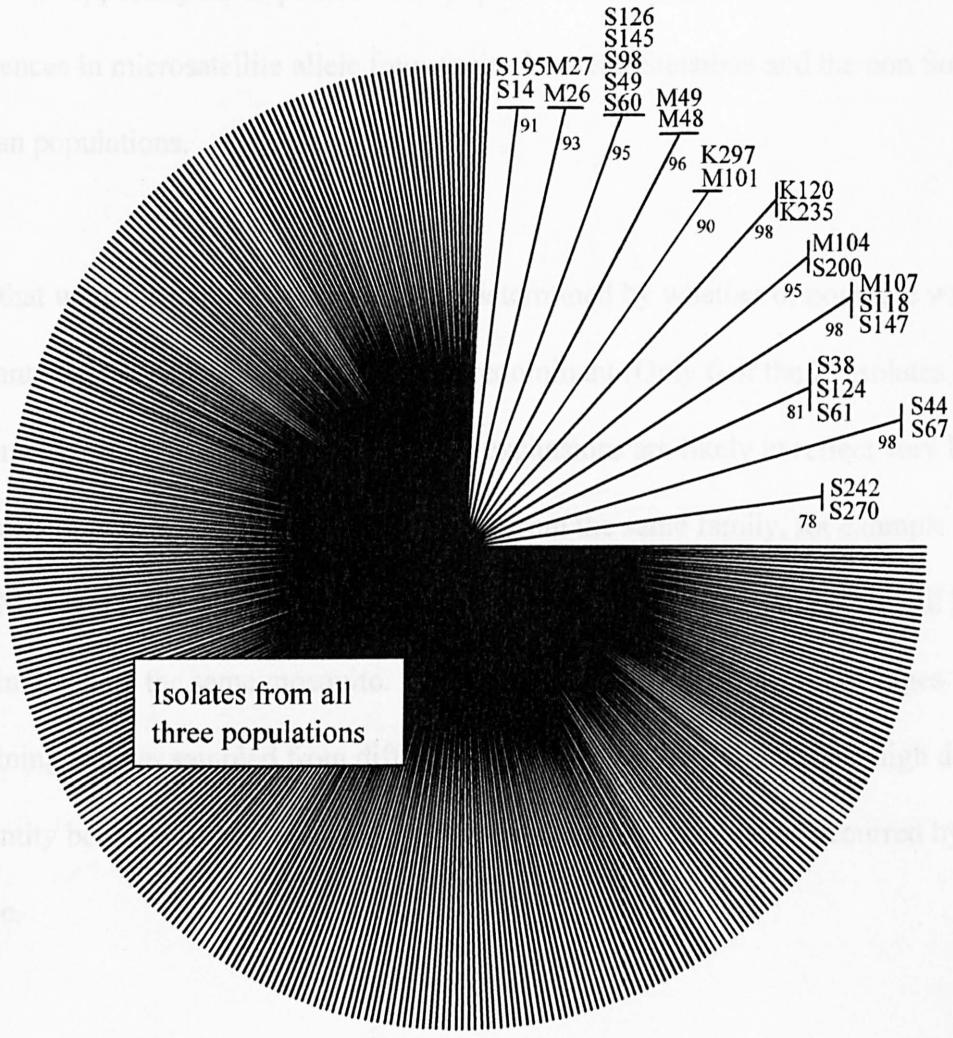


Figure 3-2: Neighbour joining tree constructed using the distance measure 1-*Ps*. Bootstrap values are given for each of the expanding lineages. Sample names are found at branch tips.

Close inspection identifies that the majority of the epidemic expansions occurred in Steenbok supporting the hypothesis that population structure events are behind the differences in microsatellite allele frequencies between Steenbok and the non South African populations.

Note that whilst the expansion was not predetermined by whether or not there was a resistant allele at *dhfr* or *dhps*, they were predominant (Only 6 of the 27 isolates have the sensitive allele at *dhfr*). A few of these expansions are likely to reflect very local transmission where samples have been taken from the same family, for example M26/M27 and M48/M49. It is likely that in these situations the patients will all have been infected by the same mosquito. However, there were also several lineages containing isolates sampled from different sites within South Africa. The high degree of identity between these isolates across the 8 loci is unlikely to have occurred by chance.

The epidemiology inferred by the presence of the clonal lineages of parasites in South Africa supports the hypothesis that selection pressure in Steenbok was high, as implied by the higher frequencies of the resistance alleles compared to the other South African populations. Epidemically expanded lineages imply a population of hosts with a minimal prevalence of clinical immunity to *P. falciparum*, which in turn implies an increased reliance on antimalarial resolution of infections.

Sample	Population	*dhfr*	*dhps*	ARAII	G377	PolyA	PfPk2	TA109	TA42	TA87	TA102
S14	Steenbok	CIRN	SGEAA	69	100	166	192	176	186	102	129
S195	Steenbok	CIRN	SGEAA	69	100	166	192	176	186	102	129
M26	Mangweni	CNCS	SAKAA	66	100	159	189	173	186	102	141
M27	Mangweni	CNCS	SAKAA	66	100	159	189	173	186	102	141
S145	Steenbok	CIRN	SAKAA	105	97	172	177	173	189	108	132
S49	Steenbok	CIRN	SAKAA	105	97	172	177	173	189	108	132
S98	Steenbok	CIRN	SAKAA	105	97	172	177	173	189	108	132
S60	Steenbok	CIRN	SAKAA	105	97	172	177	173	189	108	132
S126	Steenbok	CIRN	SAKAA	105	97	172	177	173	189	108	132
M48	Mangweni	CIRN	SAKAA	69	97	178	189	173	186	102	117
M49	Mangweni	CIRN	SAKAA	69	97	178	189	173	186	102	117
K297	Komatipoort	CIRN	SGEAA	69	100	175	181	173	186	96	129
M101	Mangweni	CIRN	SGEA A	69	100	175	181	173	186	96	129
K120	Komatipoort	CNRN	SAKAA	78	100	143	162	176	242	102	129
K235	Komatipoort	CNRN	SAKAA	78	100	143	162	176	242	102	129
M104	Mangweni	CNCS	SAKAA	72	97	153	175	158	186	100	117
S200	Steenbok	CNCS	SAKAA	72	97	153	175	158	186	100	117
M107	Mangweni	CIRN	SAKAA	75	100	159	169	161	189	100	126
S118	Steenbok	CIRN	SAKAA	75	100	159	169	161	189	100	126
S147	Steenbok	CIRN	SAKAA	75	100	159	169	161	189	100	126
S38	Steenbok	CNCS	SAKAA	72	97	153	166	161	242	108	117
S124	Steenbok	CNCS	SAKAA	72	97	153	192	161	242	108	117
S61	Steenbok	CNCN	SAKAA	72	97	153	192	161	242	108	117
S44	Steenbok	CNRN	SAKAA	72	100	175	172	170	242	87	123
S67	Steenbok	CNRN	SAKAA	72	100	175	172	170	242	87	123
S242	Steenbok	CIRN	SAKAA	66	100	156	186	173	186	96	126
S270	Steenbok	CIRN	SAKAA	66	100	156	186	173	186	96	126

Table 3-12: Sample details of the 'epidemic' expansions.

m Observation regarding the rate of de novo point mutations at dhfr

The mutation rate for substitutions at codon 108 of *dhfr* has been determined in the laboratory as being less than $2.5x10^{-9}$ mutations/DHFR gene/replication (Paget-McNicol and Saul 2001) and it can been argued numerically that, given that a patent infection contains $10^{10} - 10^{12}$ parasites, there is an expectation for single mutant resistance parasites to arise regularly if not within the course of a single infection. This observation was confirmed in description of numerous flanking sequence haplotypes of *dhfr* S108N single mutants in KwaZulu Natal (Roper et al. 2003).

It is with this in mind that we note that in one isolate within an expanding lineage of *dhfr* sensitive parasites there has been a substitution of asparagine for serine at codon 108. The single mutant isolate S61 from Steenbok has an identical genotype to isolate S124 and only one allele different to isolate S38, which both carry the sensitive allele at codon 108. This suggests that the mutation has appeared within the recent lifetime of the lineage expansion, although this was not confirmed by direct sequencing.

V Discussion

We have typed the point mutations present at the SP resistance loci *dhfr* and *dhps* in a population sample of *Plasmodium falciparum* infections taken from the field in three southeast African countries, Mozambique, Tanzania and South Africa. We found that throughout the region there were only a limited number of common alleles, namely 5 at *dhfr* and 3 at *dhps*, echoing a finding of Chapter 2. We found that whilst there are

other alleles present outside of the common eight, they are rare with less than 5% frequency in the population.

We have shown that significant heterogeneities in the frequencies of the 5 *dhfr* and 3 *dhps* resistance alleles exist both between and within the three countries. For example the frequencies of the most highly resistant *dhfr* triple mutant (N51I+C59R+S108N) and the *dhps* double mutant (A437G+K540E) alleles were found to be highest in Mozambique followed by South Africa and then Tanzania. In the case of the latter two countries, this pattern is almost certainly due to heterogeneity in selection pressure as mandated by differences in national treatment policy, where as the exception to this generalisation, Mozambique, has no official drug selection history.

We observed some heterogeneity in *dhfr* and *dhps* allele frequencies within countries most notably in South Africa where the frequencies of the *dhfr* triple mutant and *dhps* double mutant were considerably higher in Steenbok than in the neighbouring South African sites, such that the allele frequencies in Steenbok were not significantly different from those in Mozambique. However, when we considered the frequencies of the two locus genotypes, the within country differentiation between sites was not as pronounced as the greater differences existing between countries.

The hypothesis that these differences in allele frequencies exist between countries due to heterogeneity in selection pressure can be tested through description of population differentiation using selectively neutral markers. Using 8 microsatellites we found that in both inter and intra country comparisons there was no evidence of population subdivision and we make the conclusion of panmixia of parasite populations within

and between Tanzania, Mozambique and South Africa. Therefore that the heterogeneity exists between these countries in the frequencies of the resistance alleles is evidence of the heterogeneity in drug selection intensity as the gene flow in the region that has homogenised neutral allele frequencies has failed to do so at the selected loci.

a Measures of Selection

The high frequencies of the resistance alleles in Mozambique suggest that selection is occurring in this population, but this conclusion does not square with the known history of selection and national policy. Therefore is selection actually occurring in this population or are the high frequencies of the resistance alleles a result of gene flow from neighbouring South Africa, in particular KwaZulu Natal where the drug had been abandoned due to the high numbers of treatment failures (Roper et al, 2003). Linkage disequilibrium between *dhfr* and *dhps* resistance alleles is an indirect measurement of the level of selection occurring in a population. The combination of the *dhfr* triple mutant and *dhps* double mutant is the highest form of SP resistance present in southeast Africa and in populations where selection is intense these two alleles although physically unlinked will become associated through assortative mating. Therefore the presence of LD between these two alleles in a population would suggest that an association is being maintained through intense selection. In Mozambique there was no significant linkage disequilibrium between the *dhfr* triple mutant and *dhps* double mutant at any of the sites suggesting that selection was not occurring at the time of sampling or at least it was not intense or of sufficient coverage in the population to maintain an association between the two loci. Clearly linkage disequilibrium is only an indirect indicator of selection, and only a direct

observation, such as plasma levels of drug in patients would give the best estimate of the level of selection in the population. However, assuming that the indication is correct and there is no selection occurring, the high level of gene flow occurring in the region points to another cause of the widespread resistance allele, one that has significant policy implications for the region.

b The effect of gene flow

The evidence from the microsatellite markers strongly concluded that the parasite population in the region was panmictic across the three countries, and that a great deal of mixing was occurring in the parasite population such that resistance alleles could diffuse from a site where the drug is in use to sites where drug pressure is officially absent. This has been observed historically by Clyde (Clyde 1967) who in Tanzania in the 1950s,trea ted semi-immune individuals in the three villages of Mkuzi, Enzi and Kilulu found to the north of Muheza with a monthly dose of pyrimethamine. Resistance was quick to develop and spread out of the area to sites where selection was absent. At sites up to 100miles away and over a period of several years, individuals treated with 25mg or 75mg of pyrimethamine showed the presence of resistant parasites. The initial focus of intense selection was chosen by Clyde because it was thought that its isolation from other parasite populations due to isolation from other human populations would reduce the opportunity of gene flow of resistance genes. This was conclusively found not to be the case.

This extent of gene flow between neighbouring countries will have significant implications for national policy towards antimalarial usage and for the future use of that antimalarial in a much wider area than described by national borders particularly if high level resistance alleles have already entered the region.

c Epidemics in South Africa

The only site where significant LD was observed between the *dhfr* triple mutant and the *dhps* double mutant was in the South African population Steenbok. In conjunction with the high frequencies of the SP resistance alleles in Steenbok, the significant and strong linkage disequilibrium values suggests that selection was more intense at this site than elsewhere in the country.

However, the microsatellite data identified that in Mangweni and Steenbok epidemic expansions had occurred, predominantly amongst parasites carrying the *dhfr* and *dhps* resistance alleles. It is not clear whether these expansions are as a result of intense SP selection pressure or the cause of it in the population.

d Microsatellites: the pros and cons of their use

One concern about the applicability of microsatellites in description of population differentiation is that the high levels of polymorphism can lead expected heterozygosity values within and between populations to reach unity (Hedrick 1999) and thus underestimate the population differentiation. To account for this we included the probabilistic individual assignment test implement by the software Structure, in addition to allele frequency based tests such as Wrights fixation index F_{ST}. We found that both types of test concluded that the populations were panmictic at the microsatellite loci.

However, there is an additional concern about the use of microsatellites, which would not be resolved by the use of different tests, namely the assumption of selective neutrality of the microsatellite size. Microsatellites are useful because they can be considered as selectively neutral. However, it has been postulated through a modelling approach there is likely to be constraint on the size of a microsatellite (Nauta and Weissing 1996) through the tendency of mutation pressure to shift different populations towards the same distribution of alleles and empirical evidence identified an equilibrium between slippage rate enforcing a maximum size and point mutation rate removing repeat sequences from a microsatellite (Goldstein and Pollock 1997; Kruglyak et al. 1998). This constraint could generate a distribution of allele sizes centred on a mean stable microsatellite repeat size (Kruglyak et al. 1998). In such cases neither fixation indices nor population assignment tests would be able to provide resolution. The similar distributions of alleles in the two hypothetical divided populations would generate a low F_{ST} value, and the overlapping allele identity would confound population assignment. For example such a constraint on microsatellite size and the allele frequency distributions has been blamed for poor differentiation of populations of closely related bear species from more divergent populations in North America (Paetkau et al. 1997).

Applying this to the population under discussion here it can be argued that populations of malaria parasites in southeast Africa if not continuous, could appear so because the levels of genetic diversity at the microsatellites are very high and plausibly there has been sufficient time for the microsatellite allele distributions to drift to a similar mean. However, the conclusion of large scale gene flow in southeast Africa is the most rational, particularly as this is supported by the rapid dispersal of

the triple mutant allele after its migration from Southeast Asia (Roper et al. 2003; Roper et al. 2004).

It is questionable whether such an approach as microsatellite typing should be applied to ancient populations with high genetic diversity, if there is any reason to expect population subdivision.

VI Concluding remarks

The intensity of selection for drug resistance alleles in east Africa is heterogeneous between sites generating heterogeneity in allele frequencies. This is unequivocally occurring in the face of a strongly homogenising gene flow across a large region. In the subsequent three chapters I shall explore the effects of different intensities of selection on chromosomes carrying the selected sites, *dhfr* and *dhps*. Having established that differences in allele frequencies between Mpumalanga and Tanzania are maintained despite gene flow, I shall use these populations to describe the effects of the different selection pressure intensities on the resistance alleles present on chromosomes four and eight.

Chapter 4 Reduced Variation Around Drug Resistant *dhfr* Alleles in African *Plasmodium falciparum.*

I Abstract

We have measured microsatellite diversity at 26 markers around the *dhfr* gene in pyrimethamine sensitive and resistant parasites collected in southeast Africa. Through direct comparison with diversity on ancestral chromosomes we have found significant loss of diversity across a region of 70kb around the most highly resistant allele which is evidence of a selective sweep attributable to selection through widespread use of pyrimethamine (in combination with Sulphadoxine) as treatment for malaria. Retrospective analysis through four years of direct and continuous selection from use of sulphadoxine-pyrimethamine as first-line malaria treatment on a *Plasmodium falciparum* population in KwaZulu Natal, South Africa has revealed how recombination significantly narrowed the margins of the selective sweep over time. A deterministic model incorporating selection coefficients measured during the same interval indicates that the transition was towards a state of recombination-selection equilibrium. We compared loss of diversity around the same resistance allele in two populations at either extreme of the range of entomological inoculation rates (EIR), namely under one infective bite per year in Mpumalanga, South Africa, and more than one per day in southern Tanzania. Entomological inoculation rates determine effective recombination rates and are expected to profoundly influence the dimensions of the selective sweep. Surprisingly the dimensions were broadly consistent across both populations. We conclude that despite different recombination rates and contrasting drug selection histories in neighbouring countries, the region-wide movement of resistant parasites has played a key role in the establishment of

resistance in these populations and the dimensions of the selective sweep are dominated by the influence of high initial starting frequencies.

This work has been accepted for publication and is currently in press with Molecular Biology and Evolution

II Introduction

Drug treatment remains the primary means of clearing potentially lethal *Plasmodium falciparum* malaria infections and drug use has applied strong positive directional selection for resistance mutations. Theory predicts that selection will have a significant impact on genomic diversity (Smith and Haigh 1974). Initially there is a complete association between the new favoured mutation and the genome in which it arose. As the selected allele increases in frequency more distant associations are quickly broken down by recombination until only associations between the selected allele and sequences immediately flanking the gene remain – this type of association is often referred to as hitchhiking (Smith and Haigh 1974; Kaplan, Hudson, and Langley 1989). Eventually the signature of selection is a pattern of reduced gene diversity (expected heterozygosity (H_e)) at the region of sequence flanking the selected locus (Kaplan, Hudson, and Langley 1989). This loss of diversity is known as a selective sweep. Selective sweeps have been described in maize, humans and drosophila (Quesada et al. 2003; Bersaglieri et al. 2004; Palaisa et al. 2004), often long after the initial selection events which created them. Over time the eroding effects of new mutations and recombination obscure the signature of the original selection event, but in the case of *P. falciparum* drug selection is recent and measurable. We have looked at selection of resistance alleles at *dhfr* through use of pyrimethamine for treatment of *P. falciparum* in the present day, and made direct measurement of selection coefficients over a 4 year period of drug use (Roper et al. 2003). Here we examine the impact of selection, recombination and migration on chromosome diversity around the highly pyrimethamine resistant triple-mutant allele which is prevalent throughout southeast Africa.

The antimalarial drug pyrimethamine is a competitive inhibitor of the folate biosynthesis pathway that targets the active site of the enzyme dihydrofolate reductase (DHFR) and resistance to pyrimethamine is associated in vitro with substitutions within the active site of DHFR (Cowman et al. 1988; Peterson, Walliker, and Wellems 1988; Snewin et al. 1989). In Africa resistant forms of the enzyme contain substitutions at three different sites and permutations of these substitutions confer a range of sensitivities to the drug, the higher the number of substitutions, the greater the insensitivity to pyrimethamine (Wu, Kirkman, and Wellems 1996). In southeast Asia more resistant parasites are found containing a mutation at a fourth position, codon 164.

Microsatellite variation has been described around *dhfr* carrying 2-4 mutations sampled from 11 southeast Asian populations and diversity was significantly reduced in sites within a 100kb region around *dhfr*; a region of 12kb with strongly reduced gene diversity within a valley extending from 58kb upstream to over 40kb downstream (Nair et al. 2003). Selection was not concurrent with the study as use of the drug was discontinued some 10 years prior to the study (Nosten et al. 2000). Importantly all alleles had identical or very similar flanking microsatellites indicating a single ancestral origin. In Africa the situation is different. Although microsatellite flanking markers are tightly associated with *dhfr* resistance alleles containing 2 or more mutations, three separate origins of parasites carrying two mutations, but just a single origin of parasites carrying the triple mutant (N51I+C59R+S108N) *dhfr* alleles was identified in two populations in southeast Africa (Roper et al.2003) . In both Africa and Asia alleles containing a single mutation all have different flanking microsatellites suggesting multiple origins for single mutants. Comparing resistance alleles from Africa and southeast Asia and using allele sharing at 6 markers over a

30kb region around *dhfr*, we showed that the Africa triple mutant shared ancestry with the southeast Asian resistance expansion (Roper et al. 2004), indicating that it was introduced into Africa and introgressed into the population over considerable distances (>4000km).

To further explore the dynamics underlying the introgression of this allele in Africa and to describe the extent of the selective sweeps around it, we have measured the loss of diversity at microsatellite loci through comparison to a baseline of high diversity described on chromosomes carrying the drug sensitive allele. Loci are described at increasing distances from the *dhfr* gene in three different population contexts each with differing recombination rates and selection histories. It is these parameters that determine the size of a selective sweep. Selection acts to maintain high frequencies of the favoured allele and thereby the associations with hitchhiking loci (Barton 2000; Kim and Stephan 2002). A high recombination rate will act to dissolve the associations between hitchhiker and selected site, increasing variation at the flanking loci. The intensity of malaria transmission, by multiplying the number of genotypes infecting the same individual and increasing opportunities for outcrossing has a profound effect on effective rates of recombination.

Analysis of a genetic cross has estimated the recombination rate in *Plasmodium falciparum* to be 17kb/cM (Su et al. 1999), which is considerably higher that found in other eukaryotes. However this rate is moderated in the field according to the epidemiology of the parasite. Blood-stage *Plasmodium* is haploid and recombination only occurs during meiosis in the mosquito vector. The rate of outcrossing and therefore detectable recombination is determined by transmission intensity (Dye and Williams 1997). In regions of higher transmission intensity, as indicated by high

entomological inoculation rates (EIR), there are more multiple infections, increasing the number of different clones ingested by the mosquito and increasing the probability of out crossing during meiosis. The *effective* recombination rate is a more accurate measure of the recombination rate of the parasite as it also incorporates the degree of inbreeding (F) in the population (Babiker et al. 1994; Dye and Williams 1997; Conway et al. 1999).

Using a panel of 26 microsatellite loci we have mapped the gene diversity along drug sensitive chromosomes, reflecting ancestral state of diversity on chromosome 4 and triple mutant chromosomes with *dhfr* as a central point. We have examined a single population where the selection coefficient is known during a 4 year period of drug selection (Figure 4-1). Having a standardized effective recombination rate across two time points, we can describe the reduction in gene diversity at flanking loci over time due to a known strength of selection. To quantify the effect of recombination on the dimensions of the selective sweep around *dhfr* resistant chromosomes we compared loss of diversity around the same resistance allele from two populations on either extreme of a spectrum of transmission intensity in Southeast Africa. In this comparison we test the hypothesis that we should see significant differences in the dimensions of selective sweeps for parasite populations with differing effective recombination rates.

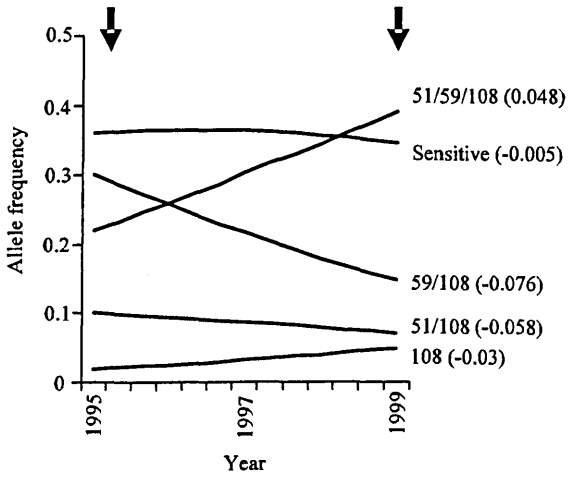


Figure 4-1: The change in the frequency of the five most common *dhfr* allelic haplotypes. The haplotype names are derived from the mutated positions in *dhfr* in the codon order 50, 51, 59 and 108. Selection coefficients (in brackets) were calculated from the change in allele frequencies over 12 generations (1 year being equivalent to 3 generations). The arrows indicate the sample points. Data taken from (Roper et al. 2003)

Population and sample date	History of SP[$] usage (first line unless stated otherwise)	Frequency of the triple mutant *dhfr* allele (51/59/108)	EIR? (Infectious bites per annum)
KwaZulu Natal 1996-1999	1988 - 2000	22% - 38%[*] (42% - 62%)[*§]	<0.8[†]
Mpumalanga 2001	1997 - present day	52% [?§]	<0.14[†]
Tanzania 2001	2001 - present day (after 18 years second line)	40% [?]	584[‡]

Table 4-1: Study population details * (Roper et al. 2003); $ Sulphadoxine-Pyrimethamine; § clinical samples; ? unpublished;† personal communication K. Barnes; ‡ (Charlwood et al. 1998)

III Methods and Materials:

a Study Samples:

To explore the effects of direct selection through time, parasite populations were sampled over two time points (1996 and 1999) in KwaZulu Natal. Samples were collected from patients presenting to a healthcare facility. The sites used for contrasting effective recombination rates were in the south of Tanzania and in Mpumalanga, a province in the northeast of South Africa. The Tanzanian samples were collected during household surveys of the three districts of Kilombero, Ulanga and Rufiji as a part of the Interdisciplinary Monitoring Project for Antimalarial Combination Therapy in Tanzania (IMPACT-Tz) artesunate combination therapy trial. The samples from Mpumalanga were from symptomatic malaria patients prior to treatment, as a component of the Southeast African Combination Antimalarial Therapy (SEACAT) evaluation. Table 4-1 summarises the drug use and

epidemiological context of the populations sampled. All three study sites are endemic for *P. falciparum.*

Typing of point mutations at codons 50/51, 59, 108, 164 was performed according to the method of (Pearce et al. 2003), using hybridization of sequence specific oligonucleotide probes (SSOP) to PCR amplified products as described in chapter 2. As blood stage parasites are haploid, the point mutation haplotypes are immediately evident except where multiple genotypes are co-infecting one person. For the purposes of looking at flanking sequence polymorphism on chromosomes carrying specific allelic forms of *dhfr* we deliberately selected a subset of samples that were unmixed at any polymorphic loci at *dhfr*.

b PCR amplification and analysis of microsatellite sequences:

The microsatellites primer sequences can be found in Appendix 3. The microsatellites were amplified in a semi-nested manner. The primary reaction comprised: 1µl template, 3.0mmol/l Mg2+, 0.75pmol/L primer and 1 unit of *Taq* polymerase. The reaction was cycled as follows: 2 min at 94°C and then 25 repeated cycles of 30s at 94°C, 30s at 42°C 30s 40°C and 40s at 65°C followed by 2 min at 65°C.

A third fluorescently labelled primer (Applied Biosystems, Warrington, Cheshire, United Kingdom) incorporated into a second round PCR of total volume 11µl containing 2.5mmol/l mg2+, 2pmol/L primer, 1 unit of Taq polymerase and 1µl of outer nest template. Cycling conditions were: 2min 94°C then 25 cycles of 20s at 94°C 20s at 45°C 30s at 65°C final step of 2 min at 65°C.

Samples were diluted 1 in 100 and run with LIZ-500 size standard on an ABi 3730 genetic analyzer (Applied Biosystems, Warrington, Cheshire, United Kingdom). Fragments were sized using the GeneMapper software (Applied Biosystems, Warrington, Cheshire, United Kingdom). As the samples were pre selected and consequently multiple alleles in the same isolate was a rare occurrence. In the event of a two or more alleles being detected, the majority allele was used if the minority peaks were less than 50% of the height of the majority. If peaks were of equivalent height, the data was recorded as missing for that locus in that isolate.

c Statistics and Software

Gene diversity values were calculated as $He = [n/(n-1)][1-\sum p_i^2]$ where n is the number of samples and p_i is the frequency of the *i*th allele. The variance of the gene diversity was calculated using Nei and Roychoudhury's formula (Nei and Roychoudhury 1974):

$$Var = \frac{2}{n(n-1)}\{(n-1)[\sum p_i^3 - (\sum p_i^2)^2] + \sum p_i^2 - (\sum p_i^2)^2\}$$

The software PowerMarker (Liu and Muse 2004) was used to calculate population differentiation theta values. The software 'bottleneck' (Cornuet and Luikart 1996) was used to determine H_e levels given the observed number of alleles using a coalescent simulation. The output was used to test the significance of allele distribution shown in the data. Whilst the stepwise mutation model (SMM) is a suitable model for evolution for a majority of microsatellites, the frequency of

complex mutations in *P. falciparum* microsatellites suggests that observations based on purely SMM could produce some spurious results (Anderson et al. 2000b). Therefore, both SMM and infinite allele model (IAM) were used, representing the upper and lower limits of the analysis. Differences between observed and 'bottleneck' predicted gene diversity values were tested for significance using Wilcoxon's tests.

We measured the significance in differences between gene diversity between pairs of chromosomal populations by permutation (Nash et al. 2005). The ratio of gene diversities between a pair of populations was calculated at each locus in the observed data sets and in 10000 simulated datasets where the alleles at each locus were reshuffled amongst all parasites. To obtain the level of significance we counted the number of occasions that the simulated ratio of diversities exceeded that for the observed data. The statistical package R (Team 2005) was used to generate the permutations.

Wiehe (Wiehe 1998) developed a model to describe the pattern of reduced variation in microsatellites flanking a selected site. This framework was subsequently modified by (Nair et al. 2003) who adapted it for *P. falciparum* specifically to describe changes in expected heterozygosity (or gene diversity) rather than variance in repeat number. This was necessary because of the often complex nature of *P. falciparum* microsatellite repetitive sequences. The model simulates the combined effects of recombination (incorporating the inbreeding coefficient to determine an effective rate of recombination), mutation, selection, and effective population size to predict the dimension of a selective sweep at equilibrium. The assumption of the model is that the allele under selection has reached fixation in the population. In the

case of southeast African *P. falciparum* populations (unlike those in Thailand as described by Nair et al 2003) the alleles associated with SP resistance are not fixed. We selected chromosomes according to whether they carried a resistance allele, and we contrasted diversity on these and on chromosomes carrying the sensitive allele. This artificially creates a "selected population" consisting only of resistance alleles, but importantly these chromosomes are still in a dynamic phase in the wider population and not in equilibrium or fixation. The important difference between the sampled chromosomes and those in the model is that they are still able to recombine with sensitive chromosomes carrying high diversity and or with resistance chromosomes surrounded by selective sweeps of different origin. The purpose of using the model is to guide interpretation of changes and differences in selective sweep dimensions measured under a number of different scenarios, and to examine what the predicted endpoint would be given the selection coefficients, recombination and migration rates in them.

1. *Effective recombination rate, r'*: The recombination rate r has been estimated by (Su et al. 1999) from observations of the genetic cross between parasite lines Hb3 and Dd2 to be 5.88×10^{-4} Morgan/kb/generation. The transmission intensity in a region determines the proportion of *P. falciparum* infections that result in self fertilization (Paul et al. 1995) and this inbreeding affects the apparent rate of recombination. Dye and Williams (Dye and Williams 1997) established a relationship between the inbreeding coefficient, F, and the recombination rate, r, such that $r' = r(1-F)$. The value of F used here was estimated from the number of mixed infections detected at *dhfr* in the original survey. As we are most interested in the amount of recombination that occurs between chromosomes carrying the different *dhps* alleles these estimates are

not unduly affected by being unable to detect multiple genotypes hidden by a shared *dhps* allele. It should however be noted that the inbreeding coefficients calculated this way maybe an underestimate.

2.*Mutation rate,* μ: The mutation rate of microsatellite loci in *P. falciparum* has been calculated as $\mu=1.59\times10^{-4}$ mutations/locus/generation (Anderson et al. 2000b).

3.*Selec tion coefficient, s*: The calculation of s from directly observed frequency changes in these populations is described in (Roper et al. 2003)

4.*Effective population size, N_e*: N_e for each population were estimated under infinite alleles model using the formula $N_e\mu = H/4(1-H)$, where H = gene diversity of neutral markers on sensitive *dhfr* chromosomes (i.e. not under the influence of selection) and μ, the mutation rate. We used the estimate of μ of $1.59\text{x}10^{-4}$ mutations/locus/generation made by (Anderson et al. 2000b).

5.

These parameters are incorporated into the formula:

$$H_e = 1 - e^{(8\mu+2r')/s}$$

where *e* is the initial frequency of the favourable allele in the population. (Nair et al. 2003) estimated *e* as $1/N_e$ because the resistance allele mutations had occurred de novo in southeast Asia. For this study of African resistance mutations we considered *e* to be the initial frequency of the mutant allele to be more variable because spread of resistance by gene flow is more likely how these mutants were introduced into African populations, rather than de novo mutation. Therefore *e* is equal to m/N_e where *m* is the number of migrants from one population to another. To adjust the

diversity predicted by the model to the baseline level of diversity per locus, the predicted values at a locus was multiplied by the observed gene diversity at that locus on the sensitive chromosomes.

IV Results and Interpretation:

Microsatellites occur in high frequency in the *P. falciparum* genome on average one per kilobase (Su et al. 1999). *dhfr* occurs at a central point on chromosome 4 using the published sequence of the 3D7 parasite line (Gardner et al. 2002). We were able to identify microsatellite markers at or close to 10, 20, 30, 40, 50, 60, 70, 75, 80, 90, 100, 125, 150, 250 and 350kb both upstream and downstream from codon 108 of *dhfr* (Primer sequences, chromosome location, repeat unit and allele size range for each locus are summarised in Appendix 3). The markers spanned 700kb, 58% of chromosome 4. Of the 30 markers, 5 were discarded either for having an inconsistent repeat size or greater than 35% missing data suggesting the occurrence of null alleles. The KwaZulu Natal sample consisted of 80 unmixed infections collected from patients in 1996 and 1999 which were selected as representative of the triple mutant *dhfr* allele (1996 n=14; 1999 n=28) and the sensitive allele (1996 n=14; 1999 n=14). Ten were removed because they were of a chromosomal haplotype present more than once in the sample set and were therefore presumed to be siblings. A further 11 samples were removed as they had greater than 35% missing data.

a The pattern of diversity around the sensitive dhfr allele

The pattern of diversity at all microsatellites flanking the sensitive alleles supports the interpretation that the sensitive form of the *dhfr* allele is considered to be the ancestral

state of *dhfr*. The mean level of gene diversity in microsatellite loci on sensitive chromosomes was similar in the two time points being 0.678 ± 0.077 in 1996 (n= 9) and 0.699 ± 0.077 in 1999 (n=12). The sample sets of the two years were merged as there was no significant differentiation between them (Wrights' fixation index F_{ST} = 0.025 *average* of all loci). The mean gene diversity of the merged ancestral population was 0.784, consistent with previous estimates in African populations of between 0.76 and 0.80 based on polymorphism at microsatellite markers dispersed throughout the genome (Anderson et al. 2000a). Table 4-2 summarises the average expected heterozygosities of markers when grouped by the repeat unit size. The gene diversity value for the trinucleotide repeats is slightly lower than previously reported, but not significantly so. The gene diversity values of the individual markers together with 95% confidence intervals (Nei and Roychoudhury 1974) plotted against distance from codon 108 of *dhfr* are shown in Figure 4-2a.

Gene diversity at marker loci around sensitive alleles should reflect baseline gene diversity in the ancestral state. To test for population events which may potentially confound this assumption we tested for excess of heterozygosity. The software '*bottleneck*' (Cornuet and Luikart 1996) carries out a Wilcoxon signed rank test comparing observed heterozygosity at each locus across the 700kb region and expected values generated under IAM and SMM. It was found that under IAM there was significant excess of expected heterozygosity (Wilcoxon test one tail p=0.008), but not so under SMM indicating that the population had not recently undergone a reduction in size. The high level of diversity at each marker, and the absence of a difference in alleles present between the two time points, together with the lack of evidence for a population structure event, means that the sensitive population can be used as representation of the ancestral state of each locus.

Repeat type	N (loci)	Expected Gene diversity (from sensitive population)	Expected Gene Diversity (from (Anderson et al. 2000b))
Di-	15	0.839 ± 0.076	0.781
Tri-	7	0.593 ± 0.098	0.688
Other	3	0.920 ± 0.114	0.636

Table 4-2: Gene diversity at microsatellite loci divided by repeat type

b The pattern of diversity around the triple mutant dhfr alleles sampled in KwaZulu Natal, South Africa

The diversity at loci on triple mutant chromosomes was compared with that on sensitive chromosomes. Figures 4-2b and c show plots of gene diversity on the sensitive chromosomes compared with that on triple mutant chromosomes collected in 1996 (n=13) and in 1999 (n=17) respectively. The most striking observation in each case is the valley of significantly reduced gene diversity on the selected chromosomes. In addition to this there are differences between 1996 and 1999. In 1996 when the frequency of triple mutants in the population was 22% (Figure 4-1) the core region of significantly reduced diversity extended across 70kb from locus U60 upstream to D10 downstream. In 1999 when the frequency of the triple mutant had increased to 38% (Figure 4-1) the core region was constricted to 50kb, extending from U40 upstream to D10 downstream (Figure 4-2c).

The core region of reduced gene diversity was taken as the range of markers between which there was an unbroken significant reduction in gene diversity. In some cases, where single locus spikes in diversity defined the outer edge of the core region there was significant diversity loss beyond the limits of the core region, for example the dip in diversity at locus U70 particular in 1996. In that case this may be considered to

have occurred as a result of selection on the favourable allele and an extended region of interrupted reduced diversity could be described. However this locus was not included in defining the core region of the selective sweep that would be compared between populations.

Interestingly, at both time points a significant asymmetry in the shape of the selective sweep is apparent with the region of reduced gene diversity extending 4-6 times further upstream than downstream. An additional microsatellite locus was identified between U75 to U70 to ascertain whether the dip in diversity relative to the sensitive chromosome population at U75 was an anomaly or suggestive of an additional site under selection. Diversity at this intermittent marker was midway between the two flanking markers in 1996 and had returned to baseline levels in 1999. It was concluded that if selection was occurring at a site proximal to locus U75 it was only very weak and more likely that the dip in diversity was an anomaly.

The size and direction of difference in gene diversity at microsatellite loci on the triple mutant chromosome from 1996, the triple mutant chromosomes from 1999 and the sensitive chromosomes are summarised in Figure 4-3. The significance levels were determined by the proportion of permutated diversity ratios that exceeded the observed diversity ratio for a given locus. Paired comparison of sensitive with triple mutant chromosomes from 1996 found diversity values were by permutation of the dataset significantly lower than on sensitive chromosomes at 8 loci spanning the 70kb core region. On the downstream side of *dhfr* in that comparison there several loci at which diversity was lower on the sensitive chromosomes indicating that diversity had fully returned on the resistant chromosomes.

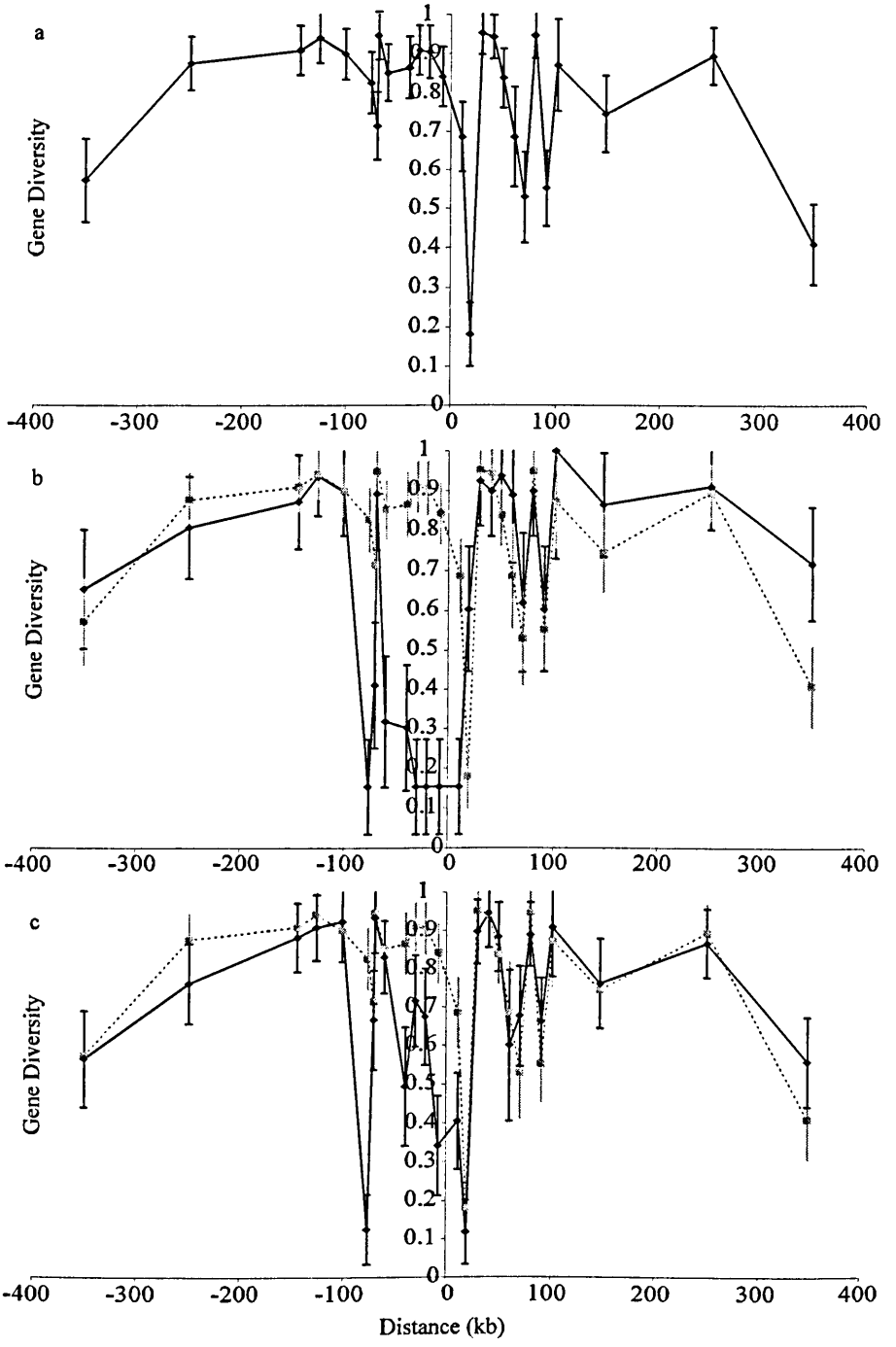


Figure 4-2: The gene diversity at each locus for populations of a) merged sensitive *dhfr* alleles (dashed line in all subsequent charts); b) triple mutant *dhfr* allele from KwaZulu Natal in 1996 (solid line), c) triple mutant *dhfr* allele from KwaZulu Natal in 1999 (solid line). Distances shown are actual physical distances from codon 108 of *dhfr* in bp. 95% CI were calculated from the unbiased variance of gene diversity as given by Nei and Roychoudbury (1974)

On the triple mutant chromosomes from 1999, diversity was significantly lower at each of the 7 core loci (spanning 50kb) and the magnitude of the difference in diversity was reduced from 0.53-0.76 in 1996 to 0.19-0.50 in 1999.

c Time and its effects on the pattern of diversity around dhfr: longitudinal data from KwaZulu Natal

The observed recovery in diversity over time could be caused by recombination through outcrossing with sensitive chromosomes or by the generation of diversity through de-novo mutation at microsatellite loci. To further examine the contributory factors influencing the dimensions of the selective sweep we have used the Wiehe model (Wiehe 1998), modified by Nair et al (Nair et al. 2003) which simulates the combined effects of recombination, mutation, selection, effective population size and inbreeding to predict the dimension of a selective sweep at equilibrium. To estimate the size of the effective population of *P. falciparum* in KwaZulu Natal we used gene diversity averaged across the loci on the sensitive (ancestral) chromosomes which gives a value of 4904 under an assumption of IAM. The selection coefficient is calculated as 0.048 based upon the measured changes in frequency of the triple mutant allele during the interval between 1996 and 1999 (described in figure 4-1). An inbreeding coefficient (F) of 0.7 was assumed. This is the rate at which selfing occurs and was estimated from the number of infections in which a single *dhfr* haplotype was detected in the original survey material. As it has been shown that the triple mutant allele did not arise in Africa but was imported (Roper et al. 2003), rather than assume an initial frequency reflecting the single mutation event, we used an estimate that reflects the number of migrants (m) between populations within the region and therefore the spread by gene flow.

Pop 1	Pop 2	U350	U250	U150	U125	U100	U75	U70d	U70	U60	U40	U30	U20	U10	MA1	MA2	dhfr	D10	D20	D30	D40	D50	D60	D70	D80	D90	D100	D150	D250	D350
Sensitive	Triple '96						0.67	0.30		0.56	0.50	0.76	0.75	0.69	0.90	0.44		0.53	-0.42			-0.10	-0.20				-0.13	-0.12		-0.31
Sensitive	Triple '99		0.12				0.70				0.37	0.19	0.23	0.50	0.90	0.83		0.28		0.06										
Triple '96	Triple '99							-0.26		-0.51		-0.56	-0.52		Fixed				0.48				0.29				0.09	0.10		0.16

Figure 4-3: The significance and direction of differences in gene diversity between sensitive and triple mutant chromosomes at each locus. Light grey = $p<0.05$; dark grey = $p<0.001$; black = $p<0.0001$. Significance was calculated as the proportion of occasions that H_e values greater than that observed were seen across 10000 permutations of the dataset.

We estimated the baseline number of migrants into a southeast African population from the F_{ST} estimate of population differentiation of parasites randomly sampled from Mpumalanga province of South Africa and Tanzania and typed at eight microsatellites dispersed throughout the parasite genome (Chapter 3). The F_{ST} of 0.011 estimates 24 migrants per generation from one population to the other. This is comparable to a F_{ST} estimate between two East African populations separated by approximately 2000km, Zimbabwe and Uganda, where m was determined as 20 (Anderson et al. 2000a). Estimating the initial frequency based on the number of migrants in this manner is a coarse approach as it may be an over estimate, as this simple approach assumes that all migrants are carrying the triple mutant allele. However, it provides a relative scale of the differences between populations in the initial starting frequencies in the explanation of the observed selective sweep data. For this analysis we used an estimate of 20 migrants per generation.

The predicted values from the model were adjusted at each locus for the baseline level of gene diversity on the sensitive chromosomes. The dimensions of the selective sweep in 1999 very closely resemble those of one predicted on the basis of these parameters (Figure 4-4). The goodness of fit was measured by the modelled data being not significantly different, by 95% confidence intervals, from the observed data (loci marked by an asterisk in Figure 4-4). This was true at 4 out of the 5 loci within the core region U40 to D10 already defined as having reduced diversity significantly below that of the sensitive population, and at 17 out of the total 26 loci, with the majority of the similarity in the portions of the selective sweep closest to *dhfr*. There is an apparent transition between 1996 and 1999 as the competing effects of selection and recombination reach equilibrium.

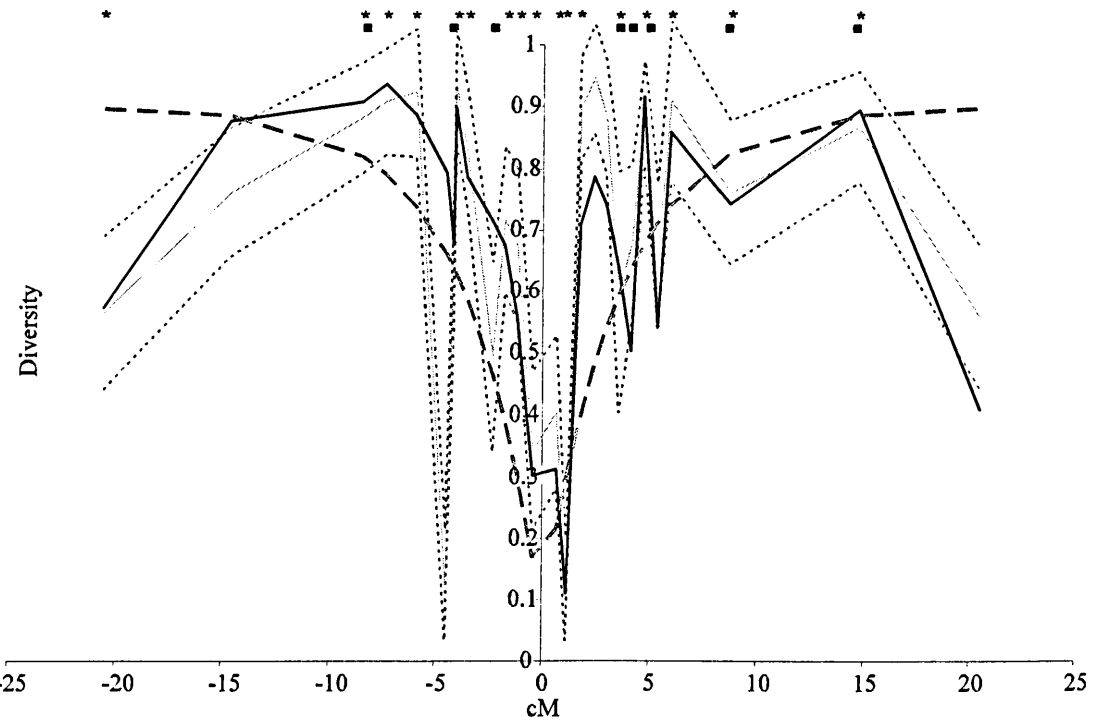


Figure 4-4: Prediction of a line of best fit by a deterministic model of hitchhiking on the chromosome. The grey solid line is the plot of the actual data in KwaZulu Natal in 1999. Dashed lines above this line represent the 95% confidence intervals around the actual data. The solid black line and dashed grey line are the model predictions for KwaZulu Natal and Thailand respectively. Parameters used for KwaZulu Natal were $N_e = 4904$, $\mu = 1.59 \times 10^{-4}$, $r = 5.88 \times 10^{-4}$, $s = 0.048$, $m = 20$, $F = 0.7$. The asterisks above the chart indicate when the modelled data using these parameters falls within the 95% CI of the actual data. Parameters used for the Thai prediction were as reported by Nair et al (2003). The squares indicate when the modelled data for the Thai population falls within the 95% CI of the KwaZulu Natal actual data

A selective sweep around *dhfr* has been described in a *P. falciparum* population on the Thai-Myanmar border. In this population an additional mutation at codon 164 is present and resistance alleles contain between 2-4 mutations. Drug selection, which took place between 1976 and 1989, completely displaced the sensitive alleles (Nair et al. 2003). Approximately 90 generations after drug selection ceased the selective sweep conformed to a shape predicted on the basis of a selection coefficient of 0.1, inbreeding co-efficient of 0.8 and a population size of 1000. The underlying microsatellite mutation rate and basic recombination rate not adjusted for the effects of inbreeding are assumed to be the same between KwaZulu Natal and Thailand. The assumption of the same underlying recombination rate makes the assumes that the genetic cross between parasite lines Hb3 and Dd2 is representative of the recombination rate throughout the total parasite population (Su et al. 1999).

The selective sweeps predicted under these two site-specific sets of circumstances are shown in Figure 4-4. Whilst the model data from parameters estimated for southeast Asia has not been adjusted for the baseline level of gene diversity at each locus for that population, the predicted line also appears fit the observed data set from 1999, yet only 1 out of 5 loci is not significantly different from the observed data within the region of significantly reduced gene diversity and only 8 out of 26 loci across the whole region (loci marked by a square in figure 4-4). We argue that the model using the parameters estimated for KwaZulu Natal is a better fit to the observed data. In KwaZulu Natal, the selective sweep is half as wide and just two thirds of the depth of that found in southeast Asia (Nair et al.2003; Anderson 2004) and this difference is explained by the lower selection coefficient (Figure 4-4) and to a lesser degree the 5 times smaller population size in southeast Asia. The effect of reduced selection is more opportunities for recombination to occur between selected and unselected

chromosomes. A smaller population size results in a shorter time until fixation of a favourable mutation, thereby also reducing the opportunities for recombination events (Kim and Stephan 2002). The difference in the magnitude of the selection coefficient between southeast Asia and KwaZulu Natal is probably a true reflection of the coverage of treatment and may also reflect the presence of even more highly resistant alleles containing substitutions at codon 164 in southeast Asia. There is a spectrum of selection coefficients acting on drug resistance loci in natural populations of *P. falciparum* and these are reviewed elsewhere (Anderson 2004). Effective recombination rates in Africa are much higher in general but in the case of KwaZulu Natal they are very similar to Thailand.

d The effect of different effective recombination rates: from low in Mpumalanga, South Africa to high in Tanzania

Although inbreeding coefficients of KwaZulu Natal and Thai-Myanmar populations were similar, transmission intensity in the African region spans an enormous range and populations in general have a much higher effective recombination rate there (Babiker et al. 1994; Anderson 2004). To examine the role of effective recombination rates we compared the selective sweep on triple mutant chromosomes in KwaZulu Natal with Mpumalanga, where the inbreeding coefficient (F) is high and southern Tanzania where F is low (Table 4-1).

The valleys of reduced gene diversity found surrounding the triple mutant *dhfr* allele present in populations in Tanzania (n=48), and Mpumalanga (n=56) are shown in Figure 4-5a and 4-5b. Consistent with previous findings, hitchhiking alleles immediately flanking the triple mutant alleles in all three populations were identical

indicating that they are descendants of one ancestral triple mutant (Roper et al. 2003; Roper et al. 2004). In the samples taken from the Tanzanian population, a 50kb core region extending from loci U40 to D10 was found to have significantly lower gene diversity than the sensitive baseline. In Mpumalanga the core region of significantly reduced gene diversity was 70kb extending from loci U60 to D10.

When the population pair wise comparisons of the gene diversity values for each matched pair of loci were plotted, we found a strong correlation for comparisons between the triple mutant allele populations of Mpumalanga with KwaZulu Natal 1999 ($r^2 = 0.934$ $p<<0.0001$), Tanzania with KwaZulu Natal 1999 ($r^2 = 0.889$ $p<<0.0001$) and Mpumalanga with Tanzania ($r^2 = 0.772$ $p=0.001$). The strong correlate is a good indication of the similarities in the dimensions of the selective sweeps in Tanzania, Mpumalanga and KwaZulu Natal in 1999. However, correlates between Mpumalanga, Tanzania and the triple mutant population from KwaZulu Natal in 1996 were much weaker (Mpumalanga $r^2 = 0.664$ $p=0.010$; Tanzania $r^2 = 0.620$ $p=0.018$).

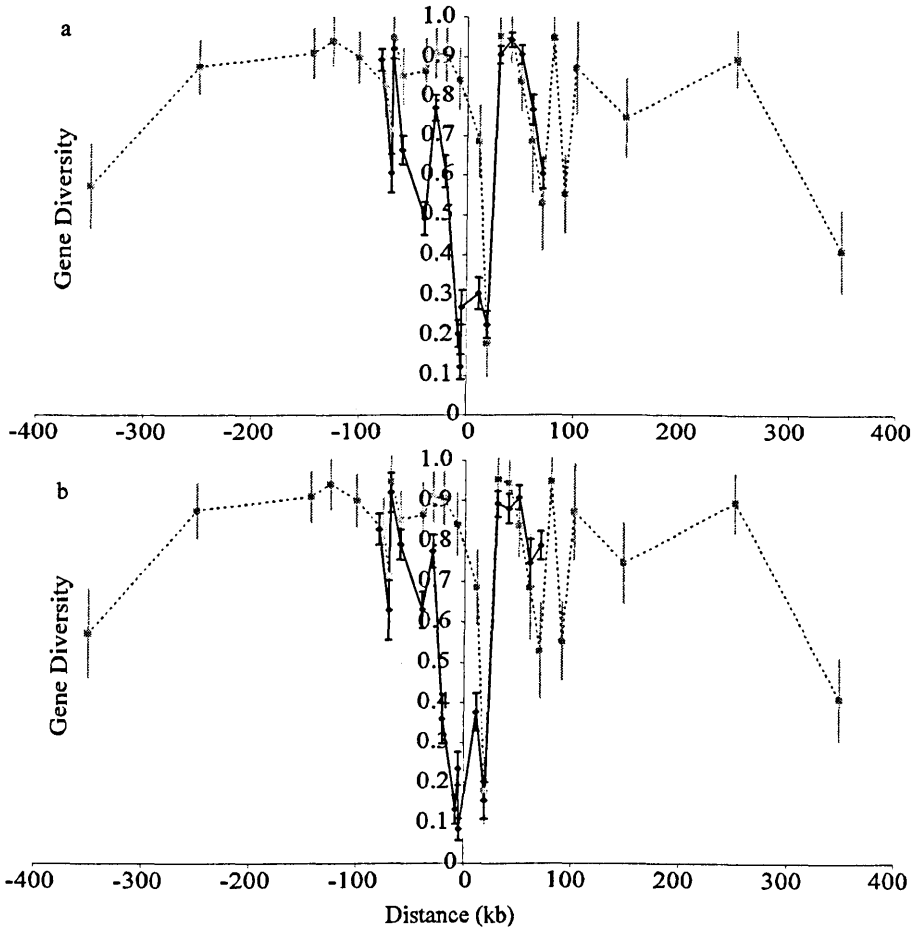


Figure 4-5: The gene diversity at each locus for populations of a) triple mutant *dhfr* chromosomes from Mpumalanga and b) triple mutant *dhfr* chromosomes from Tanzania. Distances shown are actual physical distances from codon 108 of *dhfr* in bp. 95% CI were calculated from the unbiased variance of gene diversity as given by Nei and Roychoudbury (1974).

Pop 1	Pop 2	U70d	U70	U60	U40	U30	U20	U10	MA1	MA2		D10	D20	D30	D40	D50
Sensitive	Triple Mpu			0.19	0.37	0.14	0.30	0.64	0.78	0.65	dhfr	0.38		0.05		-0.07
Sensitive	Triple Tanz				0.23	0.13	0.55	0.71	0.67	0.83		0.31		0.06	0.06	
Sensitive	Triple '99				0.37	0.19	0.23	0.50	0.90	0.83		0.28		0.06		
Triple '99	Triple Mpu			0.17												
Triple '99	Triple Tanz						0.32	0.21		0.003						
Triple Mpu	Triple Tanz			-0.13	-0.14	-0.004	0.25			0.19					0.06	

Figure 4-6: The significance and direction of differences in gene diversity between sensitive and triple mutant chromosomes from Mpumalanga, Tanzania and KwaZulu Natal '99 at each locus. Light grey = $p<0.05$; dark grey = $p<0.001$; black = $p<0.0001$.

The significant differences between loci and their direction and magnitude are summarised in Figure 4-6. The differences between triple mutant chromosomes from the three sites are limited. Two significant observations were that the selective sweep from Mpumalanga is larger, and secondly, the KwaZulu Natal 1999 selective sweep has reduced depth at the markers closest to *dhfr* relative to that in Tanzania. The lack of major differences in the dimensions of the three selective sweeps from Mpumalanga and Tanzania in 2001 as compared with that from KwaZulu Natal in 1999 is in contrast to the reduction in size that occurred over three years from 1996 to 1999 in KwaZulu Natal. We have suggested that a relatively weak selection coefficient versus a high rate of recombination explains the differences in the size of the selective sweeps in KwaZulu Natal between the two time points. Extrapolating the rates of breakdown of the selective sweep between 1996 and 1999 one may have expected on the Mpumalangan and Tanzanian derived chromosomes in 2001 to be more than 10kb smaller than that found in KwaZulu Natal in 1999. By contrast we find that the Mpumalangan selective sweep is actually significantly wider, and that the Tanzanian selective sweep whilst being the same width, is significantly deeper than in 1999. Differences in both width and depth imply the recombination has not occurred to the same extent in these populations as it has in KwaZulu Natal. The differences between the populations expected from the differences in EIR and effective recombination rate, were not apparent. In addition to EIR, assortative mating through drug pressure killing sensitive parasites in mixed infections could also cause an increase in the inbreeding coefficient.

e The importance of gene flow and initial starting frequencies on patterns of diversity

Using the model of expected diversity along a chromosome at equilibrium, we determined a line of best fit to the valleys of reduced gene diversity in Tanzania and Mpumalanga, using population-specific parameters calculated as follows. Inbreeding coefficients of 0.7 and 0.4 were estimated for Mpumalanga and Tanzania respectively from the number of single infections detected at *dhfr* across all samples in the initial surveys. The effective population sizes (N_e) were estimated from the average gene diversity of 8 microsatellites (Anderson et al. 1999) dispersed throughout the parasite genome (Chapter 3). The N_e was determined to be 5987 in Tanzania and $N_e = 4642$ in Mpumalanga. These estimates are comparable to KwaZulu Natal and to other southeast African populations (Anderson et al. 2000a) and larger than estimated for southeast Asia. The selection coefficient of 0.048, directly measured from frequency changes in KwaZulu Natal (Figure 4-1) was used and is the only direct measurement available for Africa at present.

In southeast Asia the initial frequency of the favoured allele will reflect the underlying mutation rate of *dhfr*, whereas in southeast Africa it is a reflection of the amount of gene flow between populations. Using the model, we estimated the number of migrants (m) required for best fit to the actual data was 1000 in Tanzania and 200 in Mpumalanga, 50- and 10-fold larger than the estimate based on an F_{ST} in the region. These may be overestimates because, unlike KwaZulu Natal, the selective sweep in populations of Tanzania and Mpumalanga are not yet likely to be at equilibrium. Alternatively the selection coefficient used in our simulation may be underestimated. In the case of Tanzania, an increase in the selection coefficient to

approximately 0.2 is equivalent to increase of m from 20 to 1000. However, it is unlikely that the coverage of pyrimethamine usage in Tanzania was higher than in KwaZulu Natal, Mpumalanga or on the Thailand-Myanmar border, since in Tanzania the first-line treatment at the time of sampling was chloroquine, the EIR is high resulting in acquired malaria immunity and more asymptomatic infections (Snow et al. 1997; Kleinschmidt and Sharp 2001).

In attempting to identify key parameters that explain the dimensions of the selective sweeps in the different sites using the model it becomes clear that it is not possible to distinguish a clear effect of recombination. However, over and above selection and recombination is an unquantifiable but important influence, namely the homogenizing effect of gene flow across the entire southeast African region. Drug treatment use in the various nation states has historically varied widely and local frequencies of resistance alleles will reflect this. The number of migrant triple mutant alleles entering a country will be determined by movement of people from neighbouring states and the frequency of that allele in the region or country they come from.

Although we have quantified recombination rates, the strength of pyrimethamine selection and its duration cannot be inferred from national treatment policy histories. In Tanzania, SP was the recommended second-line treatment at the time of sampling and had been for 18 years. In addition it is available through private suppliers and used in self treatment (Goodman et al. 2004). In northern Tanzania the frequency of the triple mutant allele can be as high as 84% (Pearce et al. 2003) reflecting a higher local level of use and a history of resistance (Ronn et al. 1996; Trigg et al. 1997). Gene flow between northern and southern Tanzania is likely to be high (Clyde 1967). A possible explanation for the larger than expected dimensions of the selective sweep

is that its initial frequency in the population was generated by high levels of gene flow from northern Tanzania, effectively reducing the population size and thus opportunities for out-crossing.

The parameters used to fit the line of equilibrium to the selective sweep on the triple mutant chromosomes in Mpumalanga are feasibly close to those of the actual equilibrium values. The estimate of *s* equivalent to the m of 200 is only slightly higher at 0.06 than the measured *s* in KwaZulu Natal (0.048). The drug selection history in Mpumalanga is shorter than in KwaZulu Natal as SP only became first line anti-malarial in 1997, nine years after its implementation in KwaZulu Natal.

f *Asymmetric pattern of diversity around dhfr*

A major characteristic of the selective sweeps around the triple mutant *dhfr* allele in KwaZulu Natal, Tanzania, Mpumalanga and southeast Asia is a pronounced asymmetry, and interestingly in every case the region of reduced gene diversity extends further on the upstream side of *dhfr*. There are a number of competing explanations for this phenomenon, and the most obvious is that rates of recombination differ on either side of the gene. Considerable variation in the distribution of crossover events, differing greatly from their genome average can reflect recombination hotspots or cold spots as well as other genome features such as proximity to centromeres and telomeres. (Barnes et al. 1995; Lichten and Goldman 1995). However the *dhfr* gene is in the centre of chromosome 4 and the distribution of meiotic crossover events on *P. falciparum* chromosomes has been shown to be relatively uniform in the progeny of a laboratory cross (Su et al. 1999).

Alternatively, the asymmetry may be due to stochastic nature of recombination events during short phases of intense selection. In a model of the pattern of genetic variation along a recombining chromosome, Kim and Stephan (Kim and Stephan 2002) showed that in a population where the time to fixation was short, such as when effective population size (N_e) was small, the selective sweep would be asymmetrical around the selected site. The short lineage of the rapidly expanding selected allele decreases the amount of time for recombination events to occur and the pattern is more of stochastic noise. In southeast Asia, pyrimethamine resistance swept to fixation in only six years (White 1992). In Africa the triple mutant is yet to reach fixation, so it is likely that the lineage is long and therefore the asymmetry can not be easily explained. Intense selection pressure can result in rapid epidemic expansions of resistant parasites in local populations in Africa and as such this would facilitate the formation of the asymmetry. The caveat to this hypothesis is that it assumes that the asymmetry is random, and therefore by chance it has fallen to same side of *dhfr* in East Africa as it has in southeast Asia.

Recent analysis of the genetic cross between parasite lines Hb3 and Dd2 has identified a 48.6kb region of chromosome 4 as in complete linkage with the folate salvage phenotype, thought to abrogate the killing effect of the sulphadoxine component of SP (Wang et al. 2004a). On this fragment were 7 open reading frames including *dhfr*. If there were a resistance enhancing or a resistance compensating adaptation in the region upstream of the *dhfr* triple mutant allele, such as one favourable for folate salvage then there would be reason to expect consistent asymmetry in the same direction. However the changes we observed over time do not support this as the area of lowest diversity became more condensed and mapped onto the *dhfr* gene itself. We observed a gradient in the depth of the selective sweep,

which becomes is progressively deeper as you get nearer to the gene. We observed no further dips in the levels of diversity along the chromosome that would indicate a putative second site under selection.

V Concluding remarks

In African populations there was a consistently narrower selective sweep than found on the Thailand – Myanmar border and this was expected because effective recombination rates are in general much higher in Africa due to the higher transmission intensity. We observed that over time in KwaZulu Natal the selective sweep reduced in size towards equilibrium between selection and recombination.

When we compared the three African populations of KwaZulu Natal, Mpumalanga and Tanzania we did not observe a clear relationship between the width of the area of reduced diversity and transmission intensity/effective recombination rates. The key factors appear to be the strength of selection, length of time over which drug selection has acted and the starting frequency of resistance mutants. Although the extent of gene flow is high between populations in this region, resistance levels are heterogeneous because of selection imposed in neighbouring countries implementing different treatment policies. Importantly, the founder events in Asia were determined by mutation rates whereas in Africa, these were dictated by resistance levels in neighbouring states and the contribution of diffusion of resistance alleles in determining starting frequencies has been significant.

The size of the observed selective sweep in Tanzania is contrary to expectation given the population genetic parameters of the region, but may be explained by large

numbers of migrants into that population. Our data underlines the importance of gene flow in the spread of resistance between African countries.

Chapter 5 Characteristics of a Selective Sweep Flanking the Double Mutant *dhfr* Allele in Two Southeast African *Plasmodium falciparum* Populations.

I Abstract

The *dhfr* double mutant allele has previously been shown not to confer the ability to survive the initial concentrations of pyrimethamine in the host plasma and was apparently displaced by the triple mutant allele in KwaZulu Natal under SP selection. Yet lineages possessing the double mutant allele are widespread throughout the southeast African region. To examine the effects of selection on this allele we looked for evidence of selective sweeps. Here we describe the genetic diversity at thirteen microsatellite loci covering an approximately 120kb region flanking the pyrimethamine resistant *dhfr* C59R+S108N double mutant allele. By comparing diversity present on sensitive chromosome 4 with double mutant chromosomes,we identified a region of significantly reduced diversity. Samples from Mpumalanga province of South Africa and in southern Tanzania were compared. In each case the region of reduced gene diversity was smaller than that previously described on triple mutant chromosomes sampled in the same populations. However, it was not as small as predicted by a deterministic model given a directly observed selection coefficient and quantified degree of gene flow. We find evidence of recombination between double and triple mutant chromosomes that had obscured the size of the original selective sweep on the double mutant chromosome. We discuss the implications of this with reference to the size and stability of the selective sweep on double mutant chromosomes.

II Introduction

The two *dhfr* double mutant alleles common to southeast Africa are comprised of amino acid substitutions at codons C59R+S108N, or N51I+S108N (Pearce et al. 2003; Roper et al. 2003). Descriptions of flanking sequence polymorphism on South African and Tanzanian double mutant chromosomes within 5.5kb of *dhfr*, identified that these alleles had arisen infrequently; once in the case of the 59/108 allele and twice in the case of the 51/108 allele (Roper et al. 2003). All double mutant lineages had separate ancestries from the triple mutant N51I+C59R+S108N allele (Roper et al. 2003). As the two alleles identified in KwaZulu Natal have separate origins, the identification of hitch hiking alleles would become complicated if the two 51/108 double mutant chromosomes had recombined, and thus we have chosen to describe the selective sweep on double mutant chromosomes carrying the single ancestry C59R+S108N *dhfr* allele.

The double mutant *dhfr* allele confers both a selective advantage in the presence of drug and a fitness cost in the absence of drug selection (Watkins et al.1997 ; Fohl and Roos 2003). It has been shown in vitro that the double mutant allele confers the ability to survive plasma concentrations of pyrimethamine inhibitory to sensitive parasites, but not the higher full treatment dose concentrations that the triple mutant allele can survive (Watkins et al. 1997). There is for the double mutant allele effectively a window of survival beginning 7 to 15 days after initial dosing until at least 52 days subsequently, in which the concentrations of SP in the plasma select for double mutant and higher forms of pyrimethamine resistance but are inhibitory to the sensitive parasites. Arguably selection for the double mutant allele can only occur if a new infection occurs or if the parasite emerges from the liver stage within the

selective window. The chances of new infection by a parasite carrying the *dhfr* double mutant allele within this 7-52 day window are increased in populations where the intensity of transmission is high.

It is likely that the double mutant allele has existed in the region for longer than the triple mutant allele. In KwaZulu Natal we determined the frequency change of the *dhfr* alleles between 1995 and 1999, whilst SP selection was operating (Roper et al. 2003). We found that whilst the frequency of the triple mutant allele was increasing (22% to 38%), the frequency of the double mutant allele decreased (30%-15%). It is clear from the ancestry of the two resistance allele lineages that the frequency changes observed are not due to transitional steps of double mutant to triple mutant in the evolution of resistance as observed in southeast Asia (Nair et al. 2003), but reflect the fitness differential between the two.

We have looked for evidence of a selective sweep around the double mutant allele in the form of regions of reduced diversity on the 59/108 double mutant chromosomes through comparison to diversity on ancestral chromosomes. These regions of reduced diversity around the double mutant allele are likely to be more transient due to its' longer time in the population and weaker SP selection operating on it around double mutant than triple mutant alleles and thus we compare resistant chromosomes sampled from the same population. As with the work on the triple mutant allele populations (see chapter 4), we address the effect of transmission intensity in the breakdown of the selective sweep by sampling double mutant chromosomes from Tanzania, where the entomological inoculation rate is high, and Mpumalanga where it is low (See Table 5-1).

Population and sample date	History of SP[$] usage (first line unless stated otherwise)	Frequency of the double mutant *dhfr* allele (59/108)	EIR? (Infectious bites per annum)
Mpumalanga 2001	1997 - present day	14% [?§]	<0.14[†]
Tanzania 2001	2001 - present day (after 18 years second line)	10.5% [?]	584[‡]

Table 5-1: [*] (Roper et al. 2003); [$] Sulphadoxine-Pyrimethamine; [§] clinical samples; [?] unpublished;[†] personal communication K. Barnes; [‡] (Charlwood et al. 1998)

III Methods and Materials

a Study samples:

The study sites used in this investigation are regions of southern Tanzanian and the north eastern province of South Africa, Mpumalanga. The Tanzanian samples were collected in 2001 during household surveys of the three districts of Kilombero, Ulanga and Rufiji,par t of the Interdisciplinary Monitoring Project for Antimalarial Combination Therapy in Tanzania (IMPACT-TZ) artesunate combination therapy trial. The samples from Mpumalanga were collected in 2001 from symptomatic malaria patients prior to treatment, as a component of the Southeast African Combination Antimalarial Therapy (SEACAT) evaluation. Table 5-1 summarises the drug use and epidemiological context of the populations sampled. Sensitive chromosomes were sampled in KwaZulu Natal in South Africa where the frequency of multiple infections is low. A full description of these samples can be found in the methods of Chapter 4. The typing of the point mutations present in *dhfr* and *dhps* were performed as described in chapter 2.

b PCR amplification and analysis of microsatellite sequences:

The microsatellites primer sequences can be found in Appendix 3. The PCRs were performed as described in Chapter 4.

c Statistics and Software

Gene diversity values were calculated as $He = [n/(n-1)][1-\sum p_i^2]$ where n is the number of samples and p_i is the frequency of the *i*th allele. The variance of the gene diversity was calculated using Nei and Roychoudhury's formula (Nei and Roychoudhury 1974):

$$Var = \frac{2}{n(n-1)}\{(n-1)[\sum p_i^3 - (\sum p_i^2)^2] + \sum p_i^2 - (\sum p_i^2)^2\}$$

The software PowerMarker (Liu and Muse 2004) was used to calculate population differentiation theta values. We measured the significance in differences between gene diversity between pairs of chromosomal populations by permutation (Nash et al. 2005). The ratio of gene diversities between a pair of populations was calculated at each locus in the observed data sets and in 10000 simulated datasets where the alleles at each locus were reshuffled amongst all parasites. To obtain the level of significance we counted the number of occasions that the simulated ratio of diversities exceeded that for the observed data. The statistical package R (The R Core Development Team 2005) was used to generate the permutations.

Wiehe (Wiehe 1998) developed a model to describe the pattern of reduced variation in microsatellites flanking a selected site. The model is described in full in Chapter 4. The selection coefficient 's' directly observed from the frequency changes in

Tanzania was calculated using $P_n=P_o \times e^{sn}$ where P_n = frequency at n generations, P_o = frequency at start, and n= number of generations (Hartl and Clark 1997).

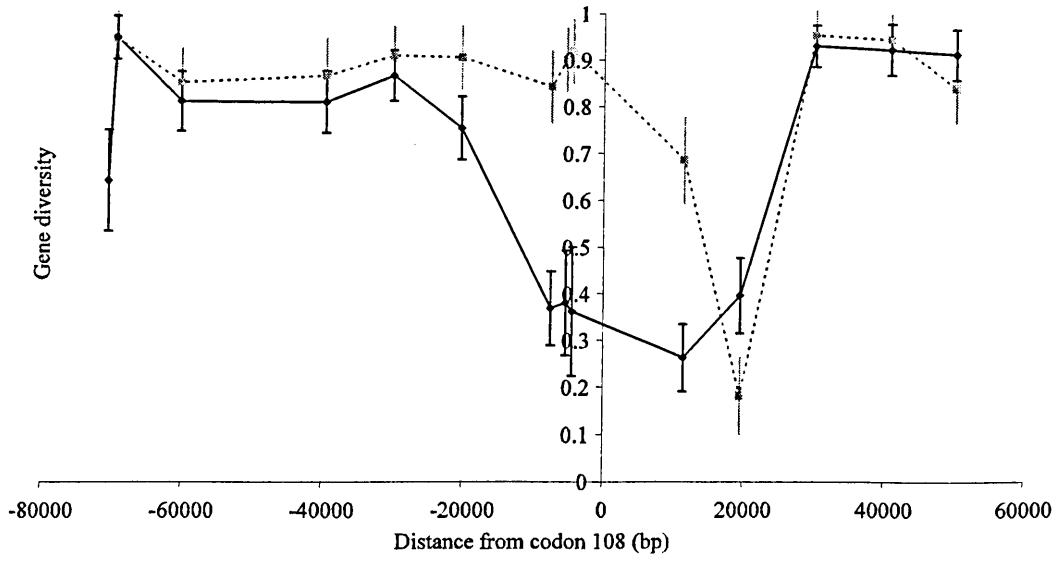
Clustering of individuals into 'k' populations was carried out using the software Structure (Pritchard, Stephens, and Donnelly 2000; Falush, Stephens, and Pritchard 2003). Burnin (minimising the effects of the starting parameter configuration) and run lengths were 10^5 and 10^6 Markov chain Monte Carlo (MCMC) iterations respectively. All other parameters were as default for the software.

IV Results and Interpretation

Previous work (Chapter 4 and Pearce et al. MBE in press) has shown reduced diversity around the triple mutant allele that had, at its greatest width, reduced diversity from 60kb upstream to 10kb downstream. Accordingly, a narrow subset of microsatellites was used in this study of double mutant chromosomes. These were 70, 60, 40, 30, 20 and 10kb upstream and 10, 20, 30, 40 and 50kb downstream of codon 108 of *dhfr*. In addition, two microsatellites located at 5.3kb and 4.4kb upstream used in a previous study were included (Roper et al. 2003). Levels of diversity at each microsatellite locus were compared with that in the sample of sensitive chromosomes from KwaZulu Natal described in Chapter 4, which we used as representative of the region. The high gene diversity and low population differentiation described between southeast African populations in Chapter 3 indicates that there is no reasonable expectation that patterns of high gene diversity observed on sensitive chromosomes identified in KwaZulu Natal would differ on non selected chromosome 4 populations in either of the other populations.

a Patterns of reduced gene diversity on double mutant chromosomes

At the time of sampling the frequencies of the double mutant alleles in the field were 14% in Mpumalanga and 10.5% in Tanzania. To examine the effect of selection we compared the Mpumalanga (n = 27) and Tanzanian (n = 21) double mutant chromosomes with the baseline of diversity on the sensitive chromosomes. We plotted gene diversity with 95% confidence intervals for each microsatellite against distance from codon 108 of *dhfr* (Figure 5-1). Significantly reduced gene diversity existed around the selected site covering 30kb (U20 to D10) in Mpumalanga and 50kb (U40 to D10) in Tanzania. The direction and extent of change in gene diversity at loci with significant differences between ancestral and double mutant chromosomes are given in Figure 5-2 and were determined as the proportion of permutated diversity ratios that exceeded the observe diversity ratio for a given locus. This process identified a core region of U40 to D20 in Mpumalanga and U40 to D20 in Tanzania as being significantly different, but in both cases locus D20 had considerably lower diversity in the sensitive chromosome population. As in Chapter 4 the core region of the selective sweep is defined as the range of markers between which there was an unbroken significant reduction in gene diversity. Both of the selective sweeps are asymmetrical with greater diversity on the downstream side of *dhfr*. Interestingly the same asymmetry was observed in the selective sweep described around the *dhfr* triple mutant chromosome.



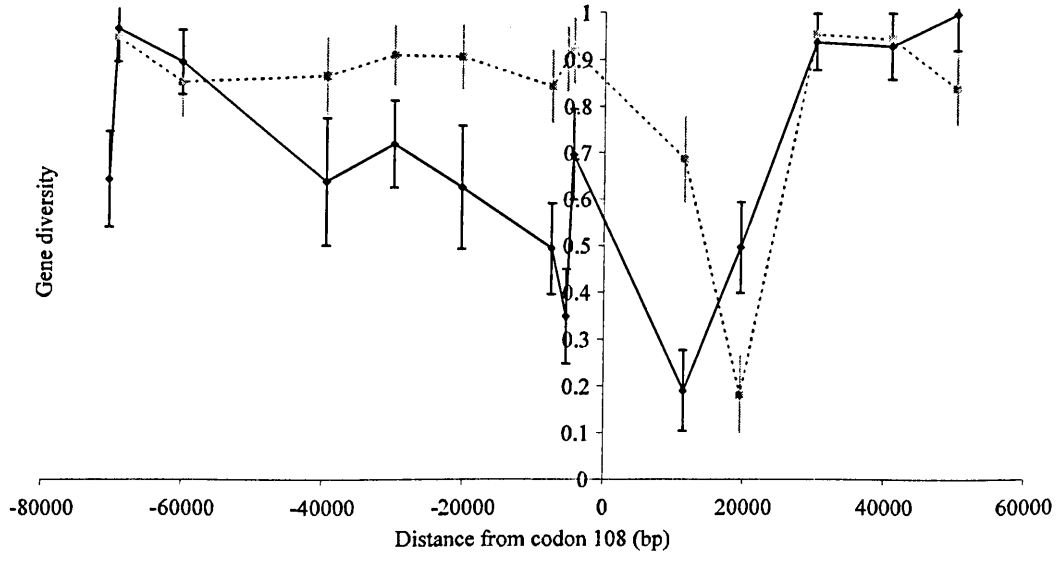


Figure 5-1: The gene diversity at each locus for populations of a) the double mutant *dhfr* chromosomes sampled from Mpumalanga, South Africa (solid line) and b) double mutant *dhfr* chromosomes sampled from Tanzania (solid line). On both plots diversity at loci flanking sensitive *dhfr* alleles sampled from KwaZulu Natal, represented by the dashed line. Distances shown are actual physical distances from codon 108 of *dhfr* in bp. 95% CI were calculated from the unbiased variance of gene diversity as given by Nei and Roychoudbury (1974).

Pop 1	Pop 2	U40	U30	U20	U10	U5.3	U4.4	dhfr	D10	D20
Sensitive	Double Mpu	0.055	0.044	0.15	0.47	0.52	0.56		0.42	
Sensitive	Double Tanz	0.23	0.19	0.28	0.35	0.55	0.22		0.50	-0.31

Figure 5-2: The significance and direction of differences in gene diversity between ancestral and double mutant chromosomes at each locus from each population. Light grey = $p<0.05$; dark grey = $p<0.001$; black = $p<0.0001$.

The upstream markers U5.3 and U4.4 are the microsatellite loci closest to the selected site. Alleles of 199bp and 186bp at these loci respectively are the same as previously reported for KwaZulu Natal and northern Tanzania (Roper et al. 2003), confirmation that the monophyletic *dhfr* double mutant has spread widely.

b The selective sweep present on the double mutant chromosomes is smaller than that on triple mutant chromosomes

Comparisons were made of the size of the selective sweeps on double mutant and triple mutant chromosomes sampled from the same population (Figure 5-3). In Mpumalanga the region of reduced diversity on the triple mutant chromosome extended from marker U60 upstream to D10 downstream. The double mutant chromosome population has significantly higher gene diversity at each locus within this range, with the exception of loci U4.4 and D10. In Tanzania, the triple mutant selective sweep extended from U40 to D10; however only at U20-U4.4 was the diversity significantly lower than that on the double mutant chromosome; there was no significant difference at the remaining sites (U40-U30). Across the two double mutant chromosomes there were only two loci at which there was a significant difference in the depth of the selective sweep namely U4.4 and U60, reflecting a lower gene diversity on the Tanzanian double mutant chromosomes.

c The archetypal selective sweeps in Tanzania and South Africa

As the two derived chromosomal lineages coexist in Tanzania and Mpumalanga, it is likely that recombination has occurred between them. Recombination between the double mutant chromosomes and an expanding lineage of triple mutant chromosomes could cause an apparent extension in the size of the selective sweep on the double

mutant chromosomes. Conversely, recombination with sensitive chromosomes will reduce the extent of chromosome covered by the selective sweep. To investigate the extent of triple and double mutant recombination, we determined the 'archetype' triple mutant and double mutant chromosomes for each population. These were inferred by selecting the allele with the highest frequency at each locus for resistance allele in the two populations. Archetypal alleles are shown in Figure 5-4 for the regions of the chromosome where diversity is significantly reduced.

Inspection of figure 5-4 shows clear differentiation at close loci U4.4 and U5.3 but further out on the double mutant chromosomes there were alleles common to both the double mutant and triple mutant lineages, namely loci U10, D10 and D20. Locus U10 on the sensitive chromosomes had a distribution of alleles more typical for a site where diversity is high suggesting that at this locus recombination had occurred between the two resistant chromosomes. At locus D20 convergence is explained by the fact that it is also most common on the sensitive chromosome. At locus D10 the shared allele had moderately high frequency among the sensitive chromosomes. The pattern of allele sharing between the triple and double mutant chromosomes at D10 and D20 can be explained by the lack of variability in the ancestral state. These loci are not informative in determining how much recombination has occurred.

Pop 1	Pop 2	U60	U40	U30	U20	U10	U5.3	U4.4	dhfr	D10	D20
Triple Mpu	Double Mpu	-0.15	-0.32	-0.10	-0.14	-0.16	-0.26				
Triple Tanz	Double Tanz	-0.10			-0.27	-0.36	-0.11	-0.61			-0.34
Double Tanz	Double Mpu	-0.08						-0.33			

Figure 5-3: The significance and direction of differences in gene diversity between double mutant and triple mutant chromosomes from Mpumalanga and Tanzania at each locus. Light grey = $p<0.05$; dark grey = $p<0.001$; black = $p<0.0001$.

Chromosome	60U	40U	30U	20U	10U	U5.3	U4.4	10D	20D
CNRN Mpu 1		210	151	203	257	199	186	136	283
CNRN Mpu 2		220	165	211	257	199	186	136	283
CNRN Tz		210	151	203	257	199	186	136	283
CIRN Mpu	230	210	163	201	257	203	177	136	283
CIRN Tz		210	163	201	257	203	177	136	283

Figure 5-4: The archetypal haplotypes of flanking sequence loci for sites within the region of reduced diversity defined on the triple mutant chromosome. See text for method of definition of archetype alleles

Excluding the loci described above, it was observed that there was a notable frequency of contaminating archetype triple mutant alleles at a number of loci on the double mutant chromosome sampled in both Mpumalanga and Tanzania. In Mpumalanga the loci on the double mutant chromosomes with the triple mutant archetypal hitchhikers were U60 (31%), U40 (32%), U20 (19%), U5.3 (5%) and U4.4 (13%). In Tanzania archetype triple mutant alleles were found at loci U60 (15%), U40 (45%), U20 (5%), U5.3 (10%) and U4.4 (35%). In neither double mutant chromosome population was the archetypal triple mutant allele found at locus U30.

The contamination of loci on the double mutant chromosomes by archetypal triple mutant alleles obscures the extent of reduced gene diversity occurring purely due to selection on the double mutant allele. To estimate the extent to which recombination with triple mutants had reduced diversity around the double mutant, we recalculated the gene diversity at each site using only the frequency in the data set of the archetype double mutant alleles, ignoring the frequency of the triple mutant chromosome associated alleles and of the rare alleles (which contribute least to the calculation). In Mpumalanga, the recalculation of gene diversity increased the range of loci where gene diversity was significantly lower than on the sensitive chromosomes to U30 – D10 as compared to U20 – D10 when all alleles present were used in the calculation. In Tanzania this recalculation decreased the number of loci with significantly lower diversity than on ancestral chromosomes to U30 – D10 (excluding U4.4), smaller than the region of U40 – D10 described when all alleles were included in the calculation.

d Recombination and epidemics in Mpumalanga

In Mpumalanga we found evidence of an epidemic expansion of a recombinant double mutant chromosome. In the process of defining the archetypal double mutant chromosomal haplotype in Mpumalanga, at each of the loci U30 and U20 were two alleles that could be defined as double mutant archetypal hitchhiking alleles (Figure 5-4). The uniquely Mpumalangan haplotype 165/ 211 on the double mutant chromosome is found in the population at a frequency of 29.6%. The U30/U20 haplotype of alleles 151/203 is common to both study populations of double mutant chromosomes, specifically found at a frequency of 33.3% in Mpumalanga. When testing for the presence of the 165/211 haplotype within the Mpumalangan triple mutant chromosome population it was not found. However,the 165 a llele at U30 was found at 5% frequency, whereas the common 151 allele at U30 was found at a frequency of 15%. The U20 alleles 211 or 203 were found in much lower frequencies although 151/203 was found as a complete haplotype in the Mpumalangan triple mutant chromosome population on two occasions.

From the core haplotype of U30 and U20 alleles it was possible to extend the haplotype in the 165/211 type double mutant chromosome to loci U40 and U60 constructing the haplotype 232/220/165/211, which had a frequency of 14.8% in the population. On double mutant chromosomes that had the common east African 151/203 haplotype there was no associations with specific alleles further upstream past U40.

The presence of two double mutant subpopulations is consistent with the rapid epidemic expansions of parasite lineages which we have described in Chapter 3. It

appears that the novel double mutant lineages derive from epidemic expansion of a recombinant double mutant chromosome at loci U60 through to U20.

e 'Structure' analysis – allocation of populations

We tested the robustness of the definition of archetypes by running the dataset through a probabilistic clustering algorithm implemented by the software Structure (Pritchard, Stephens, and Donnelly 2000; Falush, Stephens, and Pritchard 2003). The algorithm infers whether the current population is structured due to an ancestral admixture event between an user defined 'k' number of ancestral populations using a Markov chain Monte Carlo (MCMC) scheme to allocate individuals to one of the 'k' number of clusters on the basis of similarities in the allele distribution at a combination of loci, generating a likelihood value for the allocations. It also estimates the proportion, Q, of membership of each individual to a cluster, essentially estimating the proportion of an individuals' genome inherited from an ancestral population. For example, if k=2 was assumed and the proportion of an individuals' genome coming from ancestral population 1 was Q = 0.5 ($=^1/_k$), it will have equal proportions of its' genome from both assumed k populations; such a result repeated over an entire dataset would infer that there was only one population rather than the two tested for. For each individual,t he admixture proportion 'q' is independently modelled from a symmetrical Dirichlet distribution with hyper-variable a. If a is high the distribution is random, where as if a is below 1, the distribution approximates a negative binomial distribution and it models each individual as having originated mostly from a single population, with each population being equally likely.

The algorithm performs the clustering individual by individual and ignores all information regarding the sampled populations. Thus we used Structure to determine whether the individuals comprising the datasets of ancestral, double mutant and triple mutant chromosomes would be clustered into populations matching those defined by the allele present at the selected site, *dhfr*.

The algorithm can implement three models to test for subpopulations and determine the population ancestry of individuals, namely the 'no admixture', 'admixture' and 'linkage between sites' models (Pritchard, Stephens, and Donnelly 2000; Falush, Stephens, and Pritchard 2003). The latter model was the most applicable to the dataset of linked markers, whereas the other models assume linkage equilibrium between the markers. The 'linkage between sites' model infers the ancestral admixture event by estimation of the population origin of chromosomal chunks of linked markers passed between admixing individuals during meiotic crossover (Falush, Stephens, and Pritchard 2003). The chromosomal chunks are defined by correlations of allele frequencies between markers in individuals within a cluster.

In addition to producing a value for the proportion of occasions that an individual was grouped into one specific cluster, the linkage between sites also produces a locus by locus estimate of the proportion q. From this it is possible to identify chunks of chromosome that have remained in linkage during the assumed admixture event and identify the ancestral population they may have originated in.

Structure was run on the dataset of 103 and 90 individuals from the ancestral, double mutant and triple mutant chromosomes from Mpumalanga and Tanzania respectively, using loci U70dd to D50 and *dhfr* itself. We included the allele at the selected site,

dhfr, as we found that its inclusion did not bias the clustering. Using the linkage model, we found that whilst the k=3 assumption was the most likely; the model was unable to cluster individuals into a specific population, deviating only slightly from proportions close to 1/k. This may be because over all loci in the dataset there were insufficient differences between the populations to distinguish them. Furthermore, the Dirichlet hyper parameter a was above 1, indicating that each individual had allele copies originating from all populations in equal proportions (Pritchard, Stephens, and Donnelly 2000).

The locus by locus analysis of k=3 populations is shown in Figure 5-5. In both Mpumalanga and Tanzania large chunks of chromosome were identified as belonging to the same cluster. In all cases loci closest to *dhfr* were most confidently placed in one specific cluster. Clusters broadly matched the subdivision of populations into subpopulations defined by the allele at *dhfr* as expected. Within the two derived populations the chunks of chromosome identified by the algorithm as having come from the same ancestral population extended across a similar number of markers as when the region of the chromosome was described by reduced diversity.

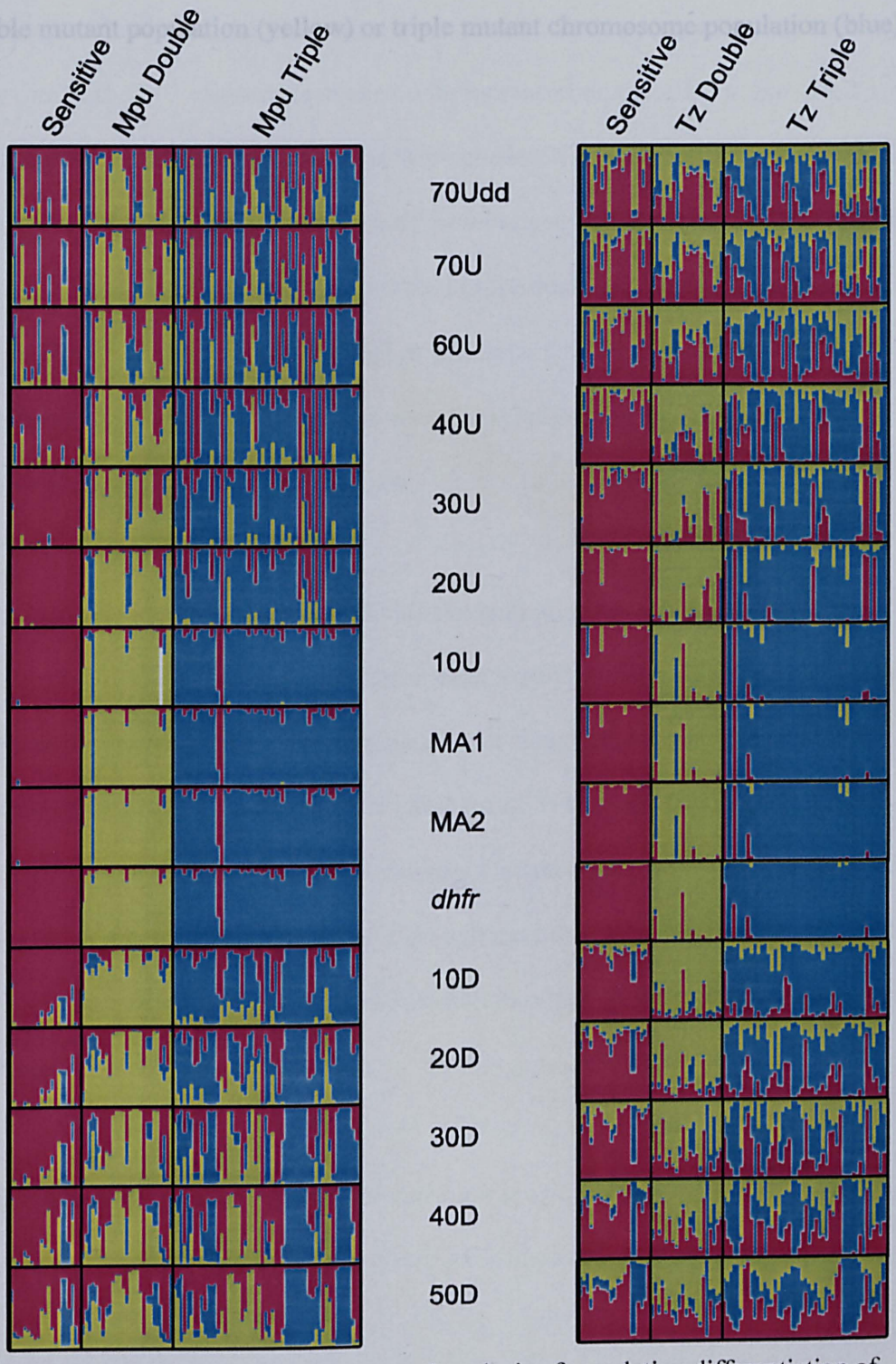


Figure 5-5: The locus by locus 'structure' analysis of population differentiation of sensitive, double mutant and triple mutant chromosomes in a) Mpumalanga and b) Tanzania. Each separate column of colour represents at each locus (cont'd over) the

proportion of an individual coming from either the sensitive population (red), the double mutant population (yellow) or triple mutant chromosome population (blue)

Outside the central core of loci where almost all individuals were confidently placed in one specific cluster, the proportion q of an individuals genome finding ancestry in only one of the k=3 clusters decreased with increased distance from *dhfr*, such that at the loci furthest from the selected site were clustered in all three k clusters with near equal proportions. This was particularly pronounced in Tanzania. In Mpumalanga, a number of double mutant and triple mutant individual chromosomes remained in one cluster for the entire extent of the region genotyped. This reflected the number of chromosomes within the dataset that were the relatives, but not direct siblings, of recent epidemically expanded lineages.

By calculating the proportion of individuals that had greater than q = 0.9 proportion of ancestry in one of the k populations at each locus, we determined the dimensions of the region of reduced diversity using a lower threshold of 20% as the minimum number of individuals exceeding this value for q. It was observed that on both the Mpumalangan and Tanzanian double mutant chromosomes this process described a region from marker U30 to the gene *dhfr*, agreeing the description of the selective sweeps dimensions as determined by calculation of the gene diversity in the absence of triple associated and rare alleles.

f A modelled description of the selection coefficient and amount of gene flow required to generate observed patterns of gene diversity

Wiehe (1998) developed a model later modified by Nair et al. (2003) that describes the pattern of diversity at microsatellite loci on chromosome flanking a selected site when there is equilibrium between the recombination, selection and mutation rates acting on the chromosome, when the selected allele is at fixation. Using this model

we can determine the strength of selection required to generate a line of best fit to the observed gene diversity on the chromosome.

The model has been described in full in chapter 4. The values for recombination rate ($r = 5.88x10^{-4}$) and microsatellite mutation rate ($\mu = 1.59x10^{-4}$), in addition to population size (N_e = 4624 Mpumalanga; 5987 Tanzania) and inbreeding factor (F = 0.7 Mpumalanga; 0.4 Tanzania) were as used in that prior section. A selection coefficient was calculated from the frequency change between 2000 and 2001 of the double mutant allele in Tanzania (s=0.025) based on the change from 9.8% to 10.5%, and weaker than the selection coefficient s=0.056, based on the change in triple mutant allele frequency (33% to 39%) (Malisa et al unpublished data) over the same period. In South Africa from 1995 to 1999 the selection coefficients for the double and triple mutant were -0.076 and 0.048 respectively (Roper et al. 2003).

g Selection versus gene flow

We input the values calculated for the parameters detailed above, and used the model to estimate the upper number of migrants (m) required to generate the gene diversity seen on the double mutant chromosomes. In Mpumalanga this was m ~ 50 and in Tanzania it was m = 500 (Figure 5-6).

The m values predicted for the double mutant chromosomes of m ~ 50 and m = 500 are four times lower than the m = 200 and m =2000 estimated by the model for the triple mutant populations of Mpumalanga and Tanzania respectively. In chapter 3 we described the patterns of gene flow between Tanzania, Mpumalanga and Mozambique, concluding that the sites were panmictic.

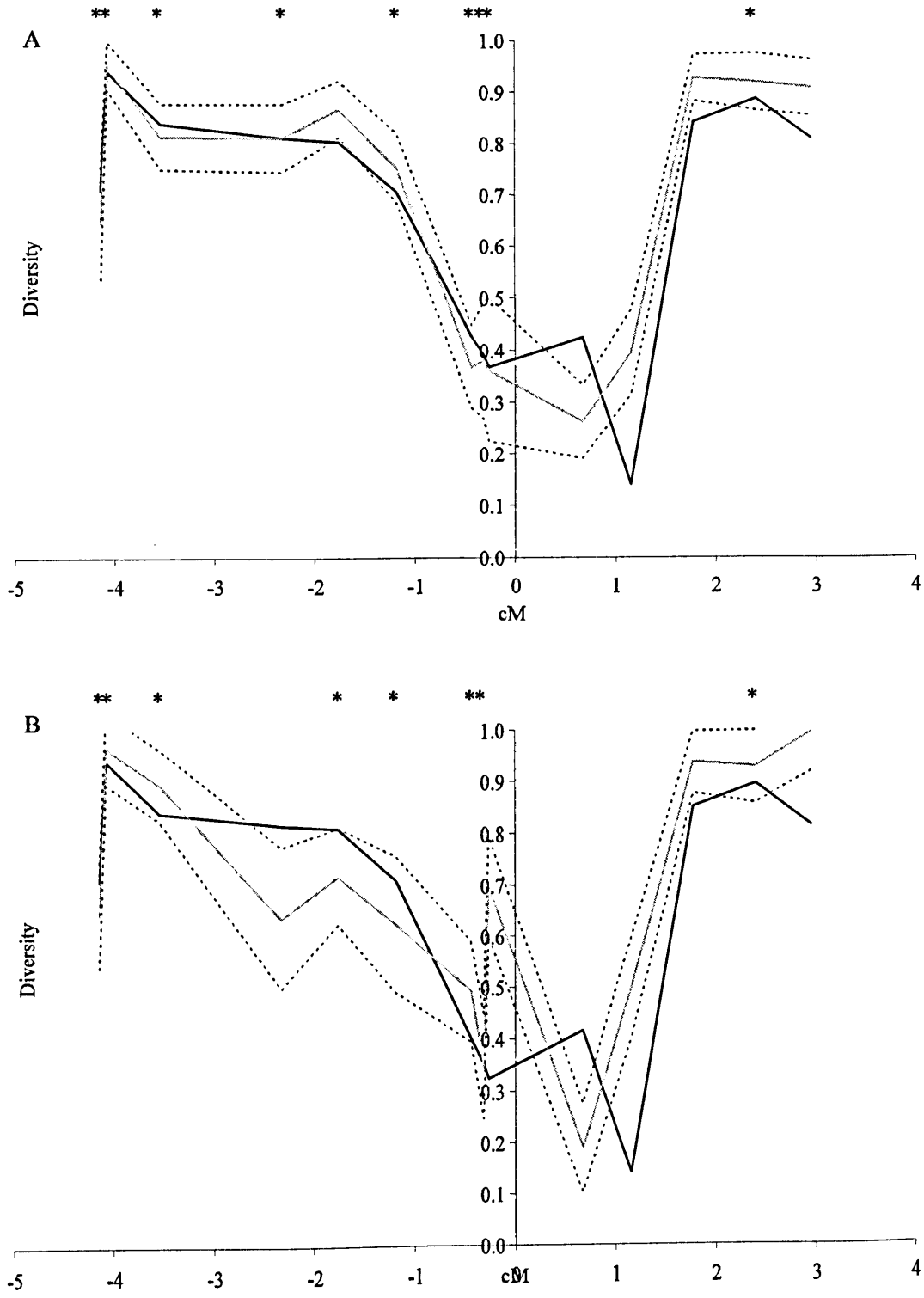


Figure 5-6: Prediction of a line of best fit by a deterministic model of hitchhiking on the chromosome. The grey solid line is the plot of the actual data in a) Mpumalanga and b) Tanzania. Dashed lines flanking this line represent the 95% confidence intervals around the actual data. The solid black line is the model predictions for each population using the parameters defined in the text (cont'd over)

and s= 0.025 and m=50 (Mpumalanga) or m=500 (Tanzania). The asterisks above the chart indicate when the modelled data using these parameters falls within the 95% CI of the actual data.

The numbers of migrants per generation in any single direction assuming symmetrical migration calculated from Wrights' fixation index were in the range of 16 to 28, considerably lower than the values of m predicted by the model.

If the initial starting frequency was fixed at the measured level of gene flow, the predicted sweep would be very much smaller. The high initial starting frequencies of the resistance allele in the population are required because the measured selection coefficient is too weak to counter the effects of recombination.

It is probable that selection coefficients were higher in the recent past. The KwaZulu Natal study shows the C59R+S108N frequency dropping, displaced by the triple mutant allele (Roper et al. 2003). By fixing the starting frequency of the double mutant allele to that when selection began, the selection coefficient required to generate the best fit line to the observed selective sweep on the double mutant chromosome can be estimated. In Mpumalanga s was only moderately higher than the estimated selection coefficient from the frequency changes of the Tanzanian double mutant allele when we assumed that the starting frequency was 20 migrants. However, using this assumption in Tanzania, the selection coefficient would have needed to be double its actual value to generate the same goodness of fit to the observed data.

V Discussion

We have described the presence of a selective sweep on the C59R+S108N double mutant chromosome sampled from populations in South Africa and Tanzania. The selective sweeps in both populations cover a significantly smaller region of the

chromosome than the selective sweep present on the triple mutant chromosome, which is consistent with their weaker resistance predicted by estimates based on measured selection coefficients for this allele in South Africa and Tanzania. In fact the extent of the sweep was considerably larger than predicted by the deterministic Wiehe model (Wiehe 1998). Gene flow may be an important factor in maintaining the size of the selective sweep in the population when selection was absent or very low. Also recombination with triple mutant chromosomes was identified as a factor important in maintaining the width of the selective sweep present on the chromosomes sampled from both populations.

The double mutant allele is a peculiarity. The very presence of a detectable selective sweep is surprising when one considers the factors affecting the survival of the allele in a population. The treatment dose of SP is lethal to *dhfr* double mutant alleles. There is an approximately 45 day window between 7 and 52 days post treatment with SP when a double mutant allele has fitness advantage over the sensitive allele (Watkins et al. 1997). Selection for the double mutant allele can only occur when chance re-infection or emergence from the liver occurs during this window. The probability of a double mutant parasite infecting a host during this selective window will depend on the intensity of transmission in the region and the frequency of the double mutant which regionally is only 10-15% (Chapter 3).

Furthermore, in the absence of drug, resistance mutation at *dhfr* may actually have a fitness cost. The C59R+S108N double mutant allele was shown to have a reproductive impairment relative to the drug sensitive allele (Fohl and Roos 2003). This was identified through transfection work in *Toxoplasma gondii* where the *P. falciparum* double mutant and sensitive alleles of *dhfr* replaced the native gene and

the rates of growth were compared in vitro and in vivo in absence of pyrimethamine. It was found that in vivo the double mutant allele conferred a 1.8% fitness cost on the numbers of parasites produced per generation (Fohl and Roos 2003). In the absence of selection, this fitness cost is unlikely to be sufficient to remove the double mutant allele from the parasite population in short time; a decline in the population frequency of the double mutant allele from 20% to 10% would take approximately 13 years and over 4 times that to drop below 1%.

The fitness of the double mutant allele relative to the triple mutant allele in the absence of drug is not known. It has been suggested that the triple mutant has a greater fitness defect as the alterations in the active site of *dhfr* are greater (Warhurst 2002), however, this is not resolved empirically.

a Selection conditions at the emergence of the double mutant allele

One explanation for the selective sweep around the double mutant in 2001 is that the selection conditions at the onset of emergence of this allele may have been different. In the 1950s pyrimethamine monotherapy was used widely (Peters 1970) and these alleles may date from then. If that is the case, then it becomes an open question what the effect of present day use of pyrimethamine has on the form of drug selection to the selective sweep. There is evidence for present day selection for the double mutant allele, such as the survival of the selective sweep in areas such as Tanzania where the high rate of effective recombination would otherwise be expected to rapidly break up the selective sweep.

b The fitness cost of the double mutant and assortative mating

It can be argued that the selective sweep present around the double mutant allele is maintained by assortative mating brought about by selective killing of sensitive parasites by drug or poor competitive fitness in mixed infections (Taylor, Walliker, and Read 1997; Taylor and Read 1998). If, in the absence of drug the double mutant allele has a fitness cost relative to the sensitive allele then during co-infections the double mutant parasites will be out competed. The result of this scenario would be that the proportion of mutants outcrossing with the sensitive chromosome would be reduced. Such inbreeding can reduce the effects of recombination on the selective sweep around the double mutant allele. In the presence of drug the same would occur, except that the relative fitness of the two alleles is reversed.

c Recombination, asymmetry and epidemics

We observed that the selective sweeps present on the double mutant chromosomes were asymmetric. The same asymmetry was observed on triple mutant chromosomes and we concluded this was due to a stochastic pattern of recombination breakpoints (Kim and Stephan 2002), that had by chance occurred more often on the downstream arm of the chromosome. However, the convergence of these observations might suggest recombination hotspots downstream of *dhfr*, but the occurrence of two invariant loci downstream of *dhfr* at the edge of the double mutant sweep has clouded the description.

In Mpumalanga we identified an indirect method for maintenance of the selective sweep on the double mutant chromosome without recourse to positive selection.

During epidemic expansions the rapid increase in the frequency of hitchhiking alleles maintains associations with the selected site and outcrossing becomes rare. Recombination between siblings within a lineage has no effect on diversity and the size of the selective sweep is maintained.

VI Concluding Remarks

If the double mutant allele can only increase in frequency and maintain a selective sweep by chance, it is not clear why the selective sweep on the double mutant chromosome present in a population like Tanzania has not disappeared completely. Are the chance infections of hosts with sufficiently low drug plasma concentrations sufficient to maintain its' frequency in the population and the sweep around it? Or are there low levels of drug usage in the population from poor drug quality and sub therapeutic dosing? We address these questions in chapter 7.

Chapter 6 Reduced Variation Around Drug Resistant *dhps* Alleles in African *Plasmodium falciparum*

I Abstract

Using 25 microsatellites covering an approximately 675kb region of chromosome 8 we have described the patterns of diversity of sequence flanking the drug sensitive and drug resistant double mutant alleles at *dhps*, the locus encoding dihydropteroate synthetase. This folate biosynthesis pathway enzyme is the target of the sulphonamide component of the antimalarial sulphadoxine-pyrimethamine (SP). A key difference to mutations at *dhfr* is that emergence of *dhps* resistance alleles has been considerably more recent. In KwaZulu Natal, a province of South Africa, the double mutant resistance allele was absent in 1995 but by 1999 it had reached a frequency of 15%. On double mutant chromosomes sampled from this population we observed an approximately 275kb region of significantly reduced diversity that extended from 125kb upstream to 150kb downstream; a region considerably larger than that described on pyrimethamine resistant *dhfr* triple mutant chromosomes from the same population and time point. We then described the patterns of diversity on resistant and sensitive chromosomes in neighbouring province Mpumalanga in 2001 and found that that the region of reduced gene diversity had narrowed to 80kb in width extending from 50kb upstream to 30kb downstream. We argue that the difference reflects the instability of selective sweeps around recently emerged resistance alleles under strong directional selection and their movement towards equilibrium over time.

II Introduction

Resistance to the sulphonamide component of sulphadoxine pyrimethamine (SP) is associated in vitro with a series of substitutions within the active site of the target enzyme of the folate biosynthesis pathway, dihydropteroate synthase (DHPS) (Brooks et al. 1994; Triglia and Cowman 1994). This has been demonstrated through laboratory based in vitro sensitivity tests and transfection experiments (Triglia et al. 1997; Wang et al. 1997b; Triglia et al. 1998)

In the enzyme DHPS 14 substitutions at five codon sites have been characterised world wide of which six have so far been recorded in Africa in a variety of haplotypic conformations (Figure 6-1) (Wang et al. 1997a; Wang et al. 1997b; Eberl et al. 2001). In southeast Africa the most prevalent of the resistance alleles is that comprised of substitutions A437G and K540E.

The role of these in conferring resistance to sulphadoxine in vivo has been inferred from studies showing the over-representation of mutations in recrudescent infections after treatment (Khan et al. 1997; Basco, Tahar, and Ringwald 1998; Curtis, Duraisingh, and Warhurst 1998; Jelinek et al. 1999b; Basco et al. 2000; Nzila et al. 2000a). This important *dhps* allele is referred to here as the double mutant allele.

Despite the above evidence, the role of mutations at *dhps* in clinical SP treatment failure has been argued to be minimal (Watkins et al 1997) in comparison to *dhfr*. It was argued that the ability of *P. falciparum* to use exogenous folate in media or in host plasma effectively bypasses the need for functional *dhps* in the folate biosynthesis pathway. However, it has recently been shown through transfection

studies that parasites require *dhps* activity to remain viable (Wang et al. 2004b). This indicates that the action of sulphadoxine in disabling *dhps* is lethal and consequently the mutations confer resistance.

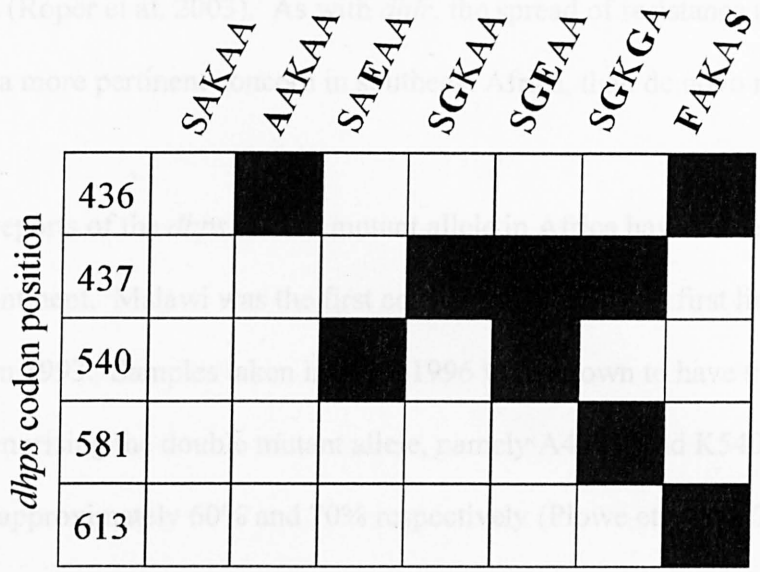


Figure 6-1: The amino acid substitutions at *dhps* reported in Africa. The allele name is taken from the amino acid present at each codon, with letters in bold identifying the locus at which the mutation occurred. Data taken from: (Wang et al. 1997a; Wang et al. 1997b; Eberl et al. 2001)

There is growing evidence that *dhps* mutations are predictive of treatment failure. (Omar, Adagu, and Warhurst 2001; Kublin et al. 2002). A recent study in Uganda found that the substitution K540E at *dhps* was the strongest predictor of treatment failure (Dorsey et al. 2004), where as others have found that the presence of three mutations (N51I+C59R+S108N) at *dhfr* together with (A436G+K540E) at *dhps* were associated with treatment failure (Kublin et al. 2002).

Further evidence that *dhps* has a role in SP resistance is the observation that there is a population expansion of the *dhps* double mutant. In KwaZulu Natal the double mutant allele increased in frequency from 0% to 15% over four years whilst SP

treatment failures were the subject of a rapid increase. Analysis of closely linked microsatellite markers (with 8kb) has shown that one particular lineage of this double mutant alleles has spread over a region of at least 4000km spanning Tanzania and South Africa (Roper et al. 2003). As with *dhfr*, the spread of resistance alleles by gene flow is a more pertinent concern in southeast Africa, than de novo mutation.

The earliest reports of the *dhps* double mutant allele in Africa have come out of the east of the continent. Malawi was the first country to adopt SP as first line antimalarial in 1993. Samples taken in 1995-1996 were shown to have the point mutations comprising the double mutant allele, namely A437G and K540E, at a frequency of approximately 60% and 70% respectively (Plowe et al. 1997). While in Tanzania samples from Magoda village in the Tanga region collected in 1995 and 1996 were found to have high frequencies of both the A437G and K540E substitutions (Jelinek et al. 1998). In west Africa the *dhps* double mutant is rare. In a recent study in Ghana, the double mutant allele (A437G+K540E) was only found at a frequency of 2.6%, whereas the A437G alone was 93.4% and the *dhfr* triple mutant was found at a frequency of 60% (Marks et al. 2005). The authors had not typed codons 436, 581, 613. Two populations sampled in southern Mauritania were more extensively screened and the K540E allele was entirely absent, yet there were mutations at other loci in particular S436A S436F A437G and A613S, but it is not possible to identify in what haplotypic conformations these mutations were found.

To address the issue of selection on mutant *dhps* we have examined the extent of the region of reduced gene diversity around double mutant alleles during first line use of SP sampled from two South African populations, KwaZulu Natal in 1999 and Mpumalanga in 2001 (Table 6-1).

III Methods and Materials:

a Study Samples:

Samples from KwaZulu Natal were taken at two time points (1995 and 1999), collected from malaria patients presenting to a healthcare facility, either Mozvold Hospital or its' satellite clinic at Ndumu. Samples were also taken in Mpumalanga, from symptomatic malaria patients prior to treatment, as a component of the Southeast African Combination Antimalarial Therapy (SEACAT) evaluation. Table 6-1 summarises the drug use and epidemiological context of the populations sampled.

Population and sample date	History of SP[$] usage (first line unless stated otherwise)	Frequency of the double mutant *dhps* allele (437/540)	EIR? (Infectious bites per annum)
KwaZulu Natal 1995-1999	1988 – 2000	0% - 15%[*] (0% - 47%)[*§]	<0.8[†]
Mpumalanga 2001	1997 - present day	10%[?§] (2%-22%)[‡]	<0.14[†]

Table 6-1: [*] (Roper et al. 2003); [$] Sulphadoxine-Pyrimethamine; [§] clinical samples; [?] unpublished;[†] personal communication K. Barnes; [‡]Range of allele frequencies among the three sampling sites; evidently the majority of samples were taken from Steenbok

P. falciparum malaria at both sites is historically endemic, but features characteristics of epidemic malaria (Kleinschmidt and Sharp 2001).

The typing of the point mutations present in *dhfr* and *dhps* were performed as described in chapter 2.

b PCR amplification and analysis of microsatellite sequences:

The microsatellites primer sequences can be found in Appendix 6. The PCRs were performed as described in Chapter 4.

c Statistics and Software

Gene diversity values were calculated as $H_e = [n/(n-1)][1-\sum p_i^2]$ where n is the number of samples and p_i is the frequency of the *i*th allele in the sample set (Nei 1987). The variance of the gene diversity was calculated using Nei and Roychoudhury's formula (Nei and Roychoudhury 1974):

$$Var = \frac{2}{n(n-1)}\{(n-1)[\sum p_i^3 - (\sum p_i^2)^2] + \sum p_i^2 - (\sum p_i^2)^2\}$$

The software PowerMarker (Liu and Muse 2004) was used to calculate population differentiation theta values. The use of the software 'bottleneck' (Cornuet and Luikart 1996) is described in Chapter 4. We measured the significance in differences between gene diversity between pairs of chromosomal populations by permutation (Nash et al. 2005). The ratio of gene diversities between a pair of populations was calculated at each locus in the observed data sets and in 10000 simulated datasets where the alleles at each locus were reshuffled amongst all parasites. To obtain the level of significance we counted the number of occasions that the simulated ratio of diversities exceeded that for the observed data. The statistical package R (The R Core Development Team 2005) was used to generate the permutations.

The distance measure 1-*ps* (Bowcock et al. 1994) was calculated using the software Microsat (Minch et al. 1997), bootstrapping the dataset 100 times. Neighbour joining

trees were calculated using the PHYLIP v3.6 (Felsenstein 2004) package "neighbor". Consensus trees were constructed using "consense" (Felsenstein 2004) and visualised using the software TreeView (Page 1996).

The Wiehe model used in this chapter is as described in chapter 4. N_e was calculated for these populations from the gene diversity of neutral markers on sensitive *dhps* chromosomes.

IV Results and Interpretation

a Gene diversity levels on sensitive chromosome populations from two time points sampled from KwaZulu Natal

To describe the extent of reduced gene diversity on chromosome 8 in parasites carrying the *dhps* double mutant allele, it was first necessary to describe the pattern of diversity along chromosomal sequence flanking the sensitive allele. Samples of sensitive chromosomes were taken from two time points, namely in 1995 when the double mutant resistance allele was not detected at all, and in 1999 where it was found at a frequency of 15% (Roper et al. 2003). Over this time, the frequency of the sensitive allele decreased apparently displaced by the double mutant allele.

Forty microsatellite markers spanning a 963kb region of chromosome 8 were identified from the 3D7 genome sequencing project (Gardner et al. 2002). Of the forty loci, ten loci were discarded as null alleles. We compared gene diversity on sensitive chromosomes sampled from KwaZulu Natal in both 1995 and 1999 to test that there was no change over time or that the emergence of the double mutant chromosome had no effect on diversity of the sensitive chromosomes. The gene

diversity at each locus on the sensitive chromosomes from 1995 (n=24) and 1999 (n=20) was calculated. These values are summarised in Table 6-2. Diversity at the majority of the microsatellite loci was high (0.423 – 0.945 in 1995; 0.511 - 0.961 in 1999), however in both years three loci were found to be monomorphic (D30, D40 and D450) and a further two loci had notably low diversity (U40.2 and D20). Comparison of 95% confidence intervals between the two time points showed that there were few significant differences in diversity at a loci apart from U40.2 (p=0.014) and U10 (p=0.025) where diversity was higher in 1995.

As the 1995 sample of sensitive chromosomes was taken prior to the emergence of the double mutant resistance allele we can exclude the possibility that gene diversity at this time point was affected by recombination with resistant chromosomes. Loci that were monomorphic or had low gene diversity (<0.200) were excluded from analyses,as they were uninformative.

To determine whether there was a significant excess or deficiency in the observed gene diversity at each locus when compared to the expected distributions simulated under stepwise (SMM) or infinite allele (IAM) mutation models for the observed number of alleles, a Wilcoxons' rank test implemented by the software 'bottleneck' (Cornuet and Luikart 1996) was used. As no specific mutation model has been fitted to the microsatellite loci used here, the IAM and SMM represent the lower and upper limits of expected gene diversity respectively. The sensitive chromosomes sampled from KwaZulu Natal in 1995 had a significant excess of gene diversity under IAM (p = 0.0063) but not under SMM. Locus U40 had a significant excess of gene diversity relative to the expected generated under both SMM and IAM (H_{obs} = 0.944 vs. SMM H_{exp} =0.916 p=0.009; IAM H_{exp} =0.897 p=0.001).

		KZN 1995			KZN 1999			
Marker	Distance	n=	Gene diversity	95% CI	n=	Gene Diversity	95% CI	P value
U125	-124300	24	0.772	0.076	20	0.853	0.077	0.153
U100	-100502	23	0.929	0.054	19	0.906	0.074	0.625
U80	-77874	24	0.562	0.081	20	0.574	0.102	0.854
U75	-74856	21	0.871	0.07	19	0.901	0.073	0.574
U60	-59750	19	0.912	0.07	19	0.889	0.074	0.656
U50	-50542	22	0.801	0.076	20	0.842	0.08	0.468
U50.2	-49173	16	0.917	0.083	20	0.911	0.066	0.909
U40	-40411	22	0.944	0.052	19	0.942	0.063	0.958
U40.2	-37676	24	0.359	0.087	20	0.195	0.089	0.014
U30	-30534	22	0.797	0.076	20	0.711	0.096	0.173
U30.2	-26958	24	0.754	0.073	20	0.726	0.092	0.647
U20	-19600	23	0.858	0.066	18	0.895	0.078	0.472
U20.2	-18014	23	0.423	0.091	20	0.558	0.106	0.064
U10	-10749	12	0.723	0.152	20	0.511	0.102	0.025
D10	10721	22	0.892	0.064	20	0.9	0.069	0.864
D20	18810	24	0.083	0.052	19	0.105	0.072	0.624
D20.2	21592	10	0.911	0.146	19	0.912	0.07	0.987
D30.2	29011	9	0.944	0.151	19	0.889	0.074	0.473
D30	31957	1	Monomorphic	n/a	20	0.195	0.089	-
D40	39849	10	0.2	0.167	1	Monomorphic	n/a	-
D50.2	50736	10	0.644	0.2	19	0.76	0.097	0.259
D60	61522	15	0.81	0.11	19	0.766	0.091	0.552
D70	71248	8	0.857	0.202	14	0.879	0.106	0.837
D75	73587	21	0.719	0.089	16	0.65	0.125	0.371
D100	100542	23	0.866	0.065	20	0.932	0.061	0.158
D125	126314	23	0.925	0.054	19	0.871	0.078	0.264
D150	147878	23	0.945	0.05	18	0.961	0.062	0.690
D350	352476	21	0.871	0.069	16	0.783	0.106	0.165
D450	451061	24	0.083	0.052	1	Monomorphic	n/a	-
D550	530114	17	0.721	0.115	17	0.809	0.101	0.267
Average	-	19	0.727	-	18	0.744	-	

Table 6-2: The observed gene diversity at microsatellite loci flanking the sulphadoxine sensitive *dhps* allele (Grey shaded boxes indicate where $p<0.05$)

Fourteen loci were identified as having a deficiency in observed gene diversity (H_{obs}), but only two were significantly different from expected diversity (H_{exp}) under SMM, namely locus U125 (H_{obs} = 0.772 vs. H_{exp} =0.876 p=0.001) and locus D550 (H_{obs} = 0. 721 vs. H_{exp} =0.850 p=0.0110) and none under IAM.

The sample taken in 1999 had a significant excess of gene diversity under IAM (p=0.0044) but under SMM there was a mildly significant deficiency (p= 0.0263). Seventeen of the 25 loci tested had lower gene diversity under the SMM than the observed, but only in two was the difference significant, namely U50 (H obs. = 0.842 vs. H exp. =0.870 p=0.0320) and U100 (H obs. = 0.906 vs. H exp. 0.953 p= 0.0040). No single locus had a significant excess of observed gene diversity.

The 'bottleneck' analysis indicated that the sensitive chromosomes from both time points had not undergone any population event that would cause a reduction in gene diversity and equally none that would cause an excess of gene diversity such as a bottleneck where the number of alleles is reduced faster than the gene diversity, such that the observed gene diversity is higher than the expected from the observed allele number.

b Population differentiation of sensitive chromosomes between 1995 and 1999 populations in KwaZulu Natal

Having determined that there was no significant difference in diversity levels between the two time points, the allele distributions at each locus for the two populations of sensitive chromosomes were compared using Wrights' fixation index (F_{ST}), with a

view to merging them (Table 6-3). Significant differences were identified at two loci (U50 and D550) due to the presence of private alleles. However, the overall F_{ST} between the two populations was low ($F_{ST} = 0.0089$). In subsequent descriptions, the diversity at each locus on the sensitive chromosome is described from the merged dataset. The sensitive chromosomes are a baseline against which to compare the effect of selection on resistant alleles and maybe regarded as the ancestral state of chromosome 8.

The microsatellites were classified according to repeat type and compared with previous estimates of gene diversity calculated for different microsatellite repeat types in *P. falciparum* (Anderson et al. 2000b) (Table 6-4). They follow a similar trend where the simpler repeats have the highest levels of diversity and overall diversity was slightly higher at these markers than those published previously.

The gene diversity at each locus for the merged ancestral dataset is plotted against distance from the midpoint of *dhps* in Figure 6-2. The pattern across the ancestral chromosome was of high diversity with occasional dips to more moderate diversity values. These dips in diversity were not restricted to any one microsatellite repeat type, or to any single region on the sensitive chromosomes. Loci D350 and D550 are excluded from the figure for the convenience of scale. The diversity at these loci was comparable to elsewhere on the chromosome shown in table 6-2.

Locus	F_{ST}	p-Value
U125	-0.002	0.434
U100	-0.013	0.700
U80	-0.009	0.404
U75	-0.004	0.297
U60	0.009	0.234
U50	0.078	0.018
U50.2	0.002	0.181
U40	0.004	0.319
U30	0.066	0.167
U30.2	-0.033	0.696
U20	0.000	0.269
U20.2	-0.021	0.468
U10	0.023	0.133
D10	-0.003	0.607
D20.2	-0.033	0.841
D30.2	-0.005	0.286
D50.2	-0.024	0.650
D60	-0.052	0.848
D70	0.064	0.095
D75	-0.047	0.977
D100	0.001	0.446
D125	0.032	0.093
D150	-0.012	0.752
D350	-0.024	0.654
D550	0.181	0.003
Overall	0.0089	0.000

Table 6-3: Population differentiation values and significance for comparisons at each microsatellite locus between *dhps* sensitive chromosomes sampled from KwaZulu Natal in 1995 and 1999

c *Diversity of sequence flanking the double mutant allele*

A sample of thirteen double mutant chromosomes from KwaZulu Natal in 1999 was typed at the 25 informative microsatellite loci. The gene diversity at each locus is compared using 95% confidence intervals with diversity on sensitive chromosomes (Figure 6-3). Loci D350 and D550 were excluded from the figure, but mutant and sensitive chromosomes did not differ significantly. The major trend is for lower diversity at all loci on the resistant chromosomes. The core region of statistically significantly lower gene diversity covers 125kb of the double mutant chromosomes extending from U50 upstream to D75 downstream. Within this region two loci were completely invariant.

Repeat Type	Gene Diversity	95% CI	Expected Gene Diversity (from Anderson et al 2000)	95% CI
di-	0.818 (n=16)	0.041	0.781 (n=12)	0.050
tri-	0.773 (n=5)	0.043	0.688 (n=12)	0.044
Other	0.883 (n=1)	0.034	0.636 (n=12)	0.169

Table 6-4: Gene diversity by repeat type in the merged KwaZulu Natal sensitive chromosome population

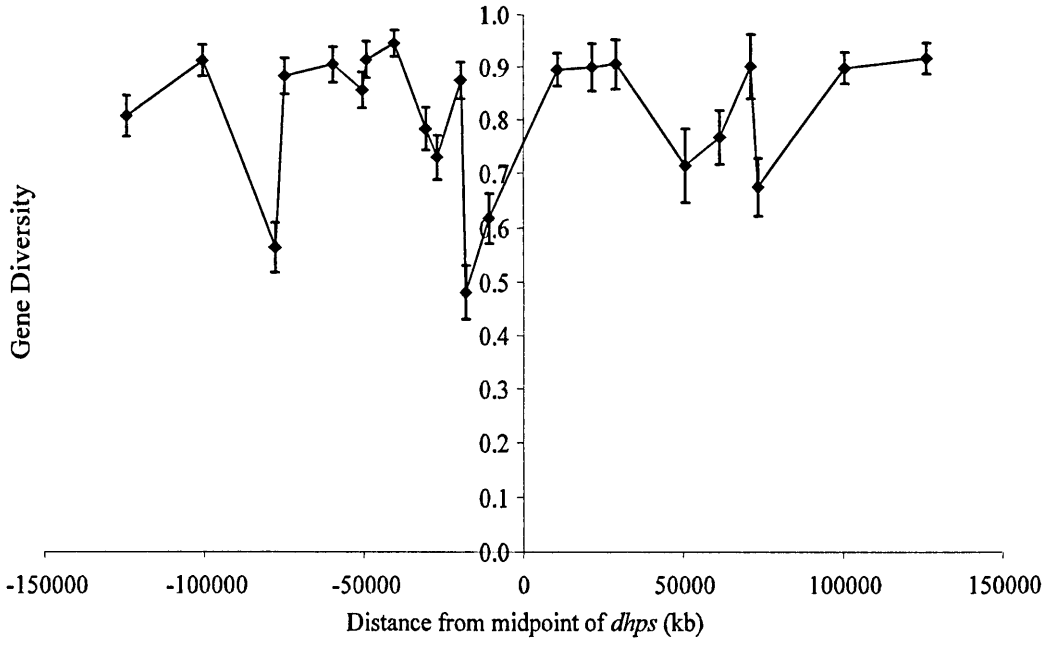


Figure 6-2: Gene diversity at microsatellite loci at increasing distances upstream and downstream from the midpoint of *dhps* on drug sensitive/ancestral chromosomes sampled from KwaZulu Natal. 95% confidence limits calculated from variance in gene diversity

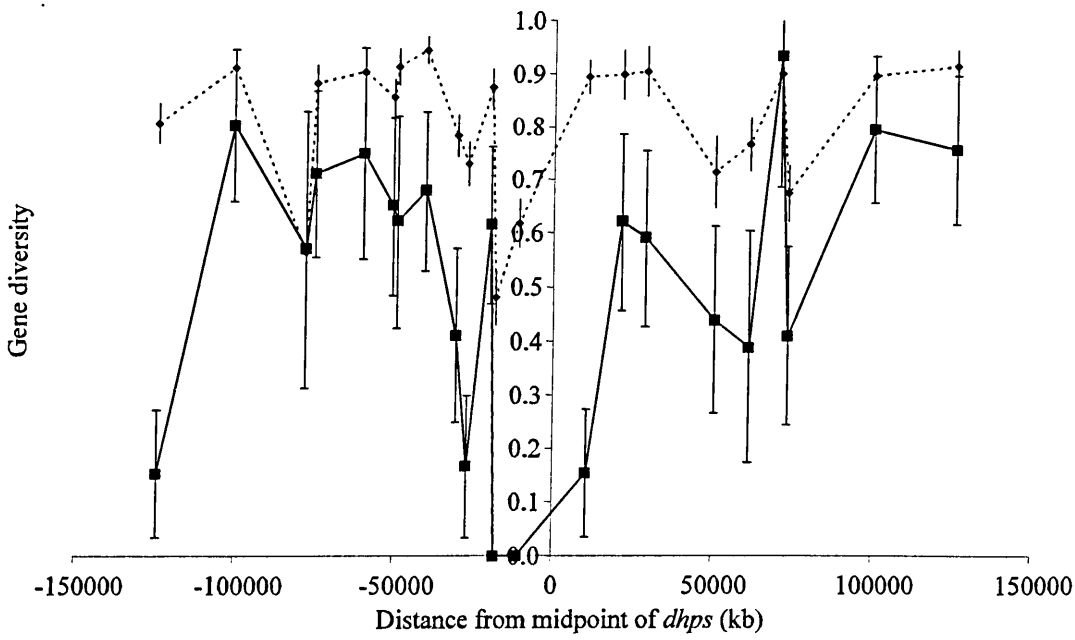


Figure 6-3: Gene diversity at microsatellite loci at increasing distances upstream and downstream from the midpoint of *dhps* on drug sensitive chromosomes (dashed line) and *dhps* double mutant chromosomes (solid line) sampled from KwaZulu Natal. 95% confidence limits calculated from variance in gene diversity.

The gene diversity at each locus on the double mutant and sensitive chromosomes was tested for significant difference using the method described in chapter 4. The extent and direction of the difference is given in Figure 6-4. With the exception of the loci U80,D50.2 and D70 the gene diversity on the double mutant chromosome was significantly lower at all loci from U125 to D150, within which there was a central core of U75 to D30.2 of consecutive markers where diversity was significantly lower than on the sensitive chromosomes. However, diversity does not return gradually with increasing distance from the selected site, rather a substantial jump in gene diversity occurs only a relatively short distance from *dhps*.

Contrasting this selective sweep with that around the *dhfr* triple mutant allele in the same population in 1999 there are two notable differences; firstly, the selective sweep on the *dhps* double mutant chromosomes is considerably larger with the statistically significant drop in diversity covering approximately 275kb as compared to approximately 50kb on the *dhfr* triple mutant chromosome. The larger size of the *dhps* double mutant chromosome selective sweep is because of the shorter history of this allele in the population. The frequency of the *dhfr* triple mutant allele in 1995 was 22% while the *dhps* double mutant was absent.

Secondly, there is not such pronounced asymmetry to the selective sweep present on the *dhps* double mutant chromosome,althoug h there are more significant differences upstream than downstream as was found with *dhfr*. We explained the asymmetry of the *dhfr* triple mutant sweep as a reflection of the stochastic pattern of recombination over time. The direction of asymmetry is the same in both cases and in the absence of a better explanation we assume this is due to chance.

Pop 1	Pop2	U125	U100	U80	U75	U60	U50	U50.2	U40	U30	U30.2	U20	U20.2	U10	*dhps*	D10	D20.2	D30.2	D50.2	D60	D70	D75	D100	D125	D150	D350	D550
KZN SAKAA	KZN SGEAA	0.65	0.11		0.17	0.15	0.20	0.29	0.27	0.37	0.56	0.26	0.48	0.62		0.74	0.28	0.31		0.38		0.27	0.10	0.16	0.13	0.11	

Figure 6-4: The direction and extent of significant differences in gene diversity between sensitive and double mutant chromosomes sampled from KwaZulu Natal. Shading indicates the following levels of significance Light grey = $p<0.05$; dark grey = $p<0.001$; black = $p<0.0001$

d Deterministic modelling of the selective sweep on KwaZulu Natal double mutant chromosomes

To examine differences in the contribution of various factors that influence sweep dimensions, we used the Wiehe model (Wiehe 1998), modified by Nair et al. (2003) to generate the expected dimension of a selective sweep, using the parameters defined in the methods section.

Using the Wiehe model it is possible to estimate either the selection coefficient acting on the selected site or the initial frequency of the allele in the population, through describing a line of best fit to the observed data. In the case of the KwaZulu Natal population, a selection coefficient for the *dhps* double mutant has been estimated from direct observation of the change in allele frequencies between 1995 and 1999 where the frequency in the samples taken from cross-sectional surveys rose from 0% to 15% (Roper et al. 2003). Roper et al (2003) calculated an upper estimate of the initial frequency of the *dhps* double mutant of 0.03 based on binomial statistical probability. Based on this a conservative estimate of the selection coefficient,'s' was calculated as 0.13 assuming 3 parasite generations per annum.

The initial starting frequency ($e = m/N_e$) of the double mutant allele in the KwaZulu Natal parasite population is unknown. We can infer that it was introduced by migration rather than de novo mutation since there are reports of the *dhps* double mutant at other sites in southeast Africa which precede its first appearance in KwaZulu Natal (Plowe et al. 1997; Jelinek et al. 1998). The number of migrants moving into KwaZulu Natal can be estimated from an F_{ST} calculated from 8 unlinked microsatellites (Chapter 3). The number of migrants between Tanzania and South

Africa per generation in one direction is approximately (m=) 20. This is comparable to an estimate of the number of migrants between the east African populations of Zimbabwe and Uganda (Anderson et al. 2000a).

Using s = 0.13 and m = 20 we find that the shape and size of the modelled selective sweep follows closely the line of the observed data (Figure 6-5). However, even if these parameters were varied, it was not possible to generate a theoretical sweep that had achieved fixation at the two sites closest to the selected site and a return to diversity to the more distant sites at the same rate as that in the observed dataset.

e The extent of the chromosome covered by the selective sweep may be obscured by inbreeding

It could be argued that the large size of the region of reduced diversity on the resistant chromosomes was because we had sampled related individuals in a resistant malaria epidemic. Such a scenario would artificially lower diversity among all resistant chromosomes. The presence of "siblings" in the data set is a risk of sampling within populations of low transmission intensity where epidemic expansions of parasite lineages are not uncommon. We tested the entire KwaZulu Natal resistant and sensitive chromosome dataset for the presence of siblings. Neighbour joining trees were constructed using the distance measure 1- *Ps* where *Ps* is a measure of similarity between two isolates calculated as 'the proportion of alleles shared summed over loci/2 x number of loci compared' (Bowcock et al. 1994). As the majority of the loci on the resistant chromosomes share alleles the, differentiation between samples was effectively only tested on a minimum of 5 loci i.e. those which were not

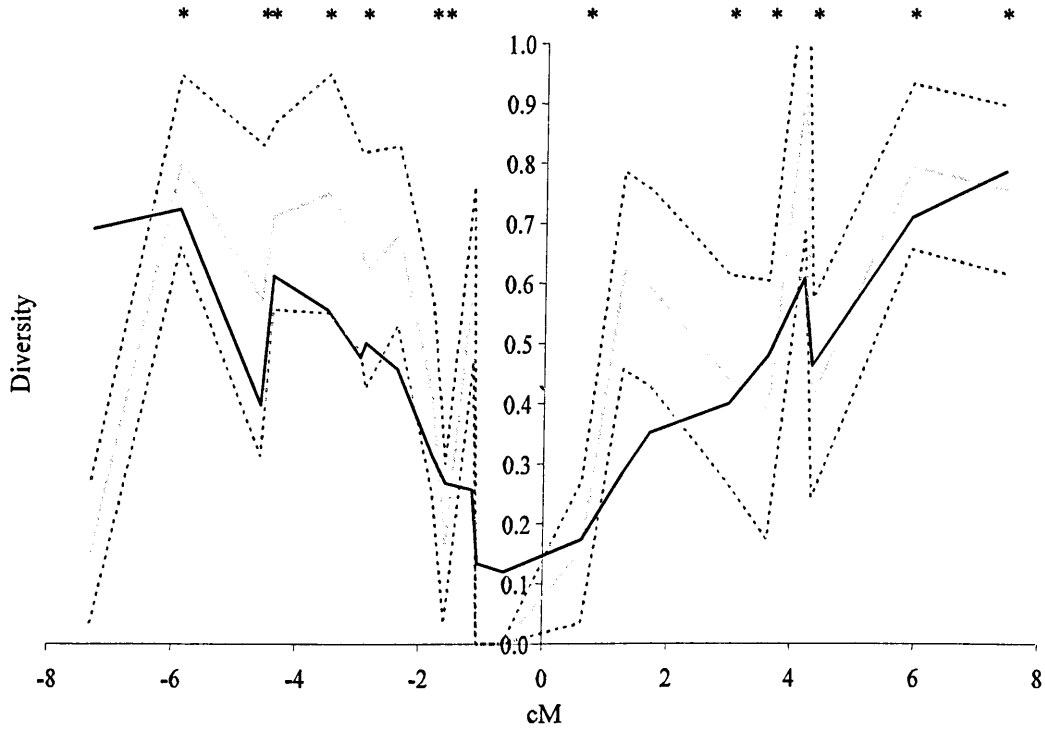


Figure 6-5: Deterministic prediction of the size and shape of the selective sweep around the *dhps* double mutant allele at equilibrium between recombination, mutation and selection. The parameters used are given in the text, however: selection coefficient s = 0.13, m=20. The grey line represents the observed diversity with the dashed grey lines representing the 95% CI. The black line represents the model predication. Asterisks above indicate loci at which the modelled line fell within the 95%CI

significantly different from the diversity on the sensitive chromosomes. The dataset was bootstrapped 100 times in Microsat (Minch et al. 1997) and a consensus neighbour joining tree drawn using the PHYLIP packages Neighbor and Consense (Felsenstein 2004) (Fig 6-6).

The majority of the pooled sensitive and resistant dataset were unrelated and equally dissimilar. Four lineages of near identical chromosomes were identified, three consisting of two isolates and one of 6 isolates. Only one lineage contained sensitive alleles. When the lineage of six isolates were examined at *dhfr* and microsatellites on chromosome 4, only three matched indicating that these three are siblings; the remaining three carried different alleles. The other two resistant lineages which consisted of pairs of isolates were found to be unrelated at chromosome four.

f A selective sweep present on double mutant chromosomes sampled in Mpumalanga in 2001

A key observation of the selective sweep on the triple mutant chromosome observed in KwaZulu Natal was the reduction in size from 70kb in width to 50kb between 1995 and 1999. By describing the region of reduced diversity on a second South African population at a later time point we hoped to examine whether the same thing had happened. KwaZulu Natal stopped using SP in 2000 so we used a sample from Mpumalanga which is a neighbouring province in South Africa. It is likely that there is considerable gene flow between KwaZulu Natal and Mpumalanga (Chapter 3). Use of SP began in only 1997. In 2001 the frequency of the *dhps* double mutant allele in Mpumalanga was 10% ranging between 2 and 22% in the populations

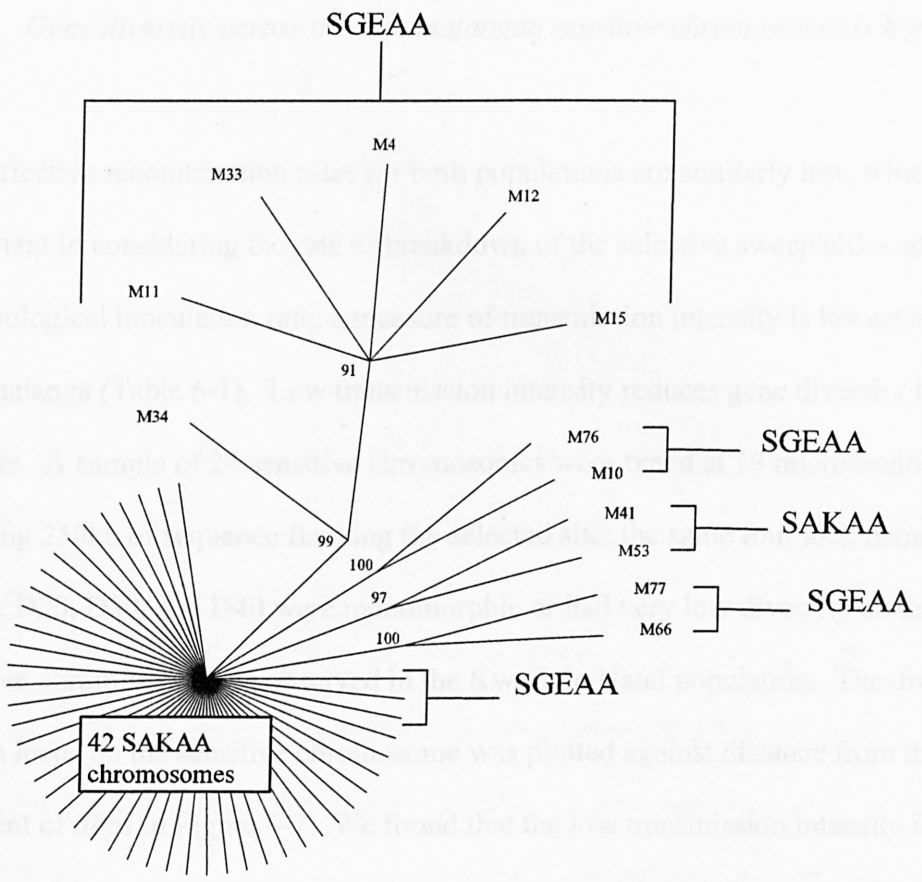


Figure 6-6: 'Neighbour joining' tree of relatedness using the measure of 1-*Ps* of sensitive and double mutant chromosomes sampled in KwaZulu Natal. Bootstrap values are given on the nodes of the expanding lineages

sampled. This frequency was broadly equivalent to 15% in KZN in 1999 (Roper et al. 2003).

g Gene diversity across the Mpumalangan sensitive chromosomes is high

The effective recombination rates for both populations are similarly low, which is important in considering the rate of breakdown of the selective sweep although the entomological inoculation rate, a measure of transmission intensity is lowest in Mpumalanga (Table 6-1). Low transmission intensity reduces gene diversity in a parasite. A sample of 24 sensitive chromosomes were typed at 19 microsatellites covering 250kb of sequence flanking the selected site; the same four loci, namely U40.2, D20, D30, and D40 were monomorphic or had very low diversity on the sensitive chromosomes, as observed in the KwaZulu Natal population. The diversity at each locus on the sensitive chromosome was plotted against distance from the midpoint of *dhps* in Figure 6-7. We found that the low transmission intensity in Mpumalanga had not affected the levels of diversity on the chromosome.

With the exception of the U10 and D20.2 there were high levels of diversity across the entire region of the chromosome. The observed gene diversity was tested for either a significant excess or deficiency using Wilcoxons' rank test through comparison with expected diversity generated under IAM or SMM mutation models (Cornuet and Luikart 1996). It was found that under IAM there was a significant excess of gene diversity but under SMM there was neither a significant deficiency nor excess of gene diversity.

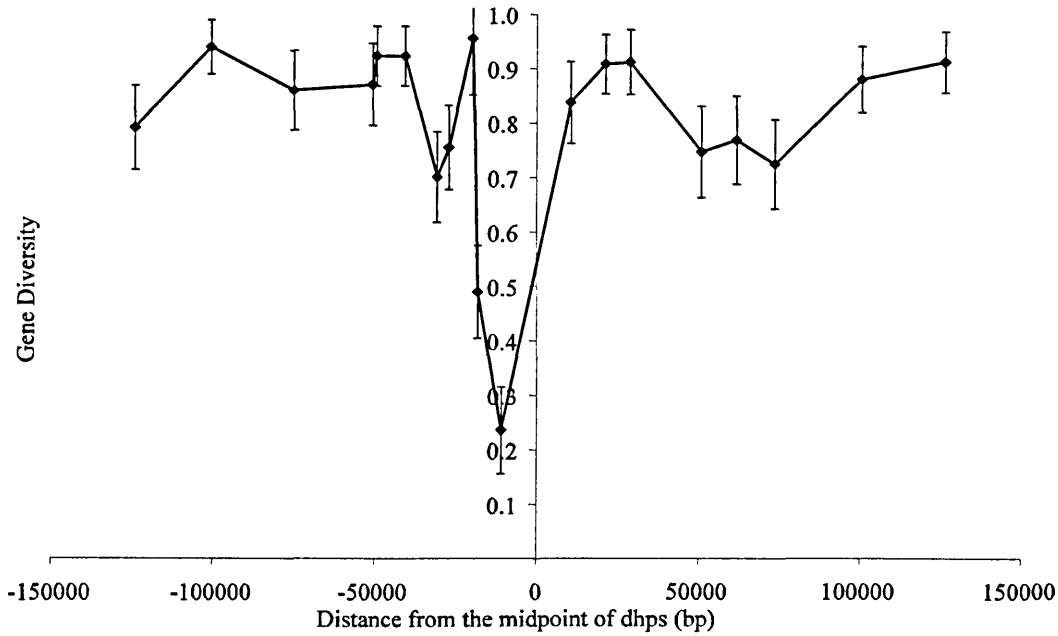


Figure 6-7: Gene diversity at microsatellite loci at increasing distances upstream and downstream from the midpoint of *dhps* on drug sensitive chromosomes sampled from Mpumalanga. 95% confidence limits calculated from variance in gene diversity

h The selective sweep present on double mutant chromosome in Mpumalanga is smaller than that in KwaZulu Natal

Gene diversity at each of the 19 loci in a sample of 24 double mutant resistant chromosomes was plotted against distance from the midpoint of *dhps* in Figure 6-8. The selective sweep present on the double mutant chromosomes sampled from Mpumalanga was smaller than that in KwaZulu Natal. Comparing 95% confidence intervals between sensitive and resistant chromosomes it was determined that the core of the selective sweep covered approximately 80kb from loci U50.2 to D30.2. Whilst this is smaller than the 275kb *dhps* in KwaZulu Natal, it was larger than the 70kb selective sweep present on chromosome 4 in Mpumalanga in 2001 and KwaZulu Natal in 1999. It is asymmetrical with diversity having returned quicker on the downstream side. Paired comparison of the sensitive and resistant chromosomes in Mpumalanga are summarised in Figure 6-9.

Curiously, the deepest part of the region of reduced diversity is not centred on the selected site, but upstream of the gene itself. The overall shape of the sweep is also asymmetrical extending much further upstream. This was also the case when we had observed asymmetry in the selective sweep present on the *dhfr* triple mutant chromosome. In the case of the *dhfr* we concluded that the stochastic pattern of recombination breakpoints was the cause (Kim and Stephan 2002) and this may well be the case for *dhps* too. However with *dhfr* the deepest part of the selective sweep was most proximal to *dhfr*,wher eas with *dhps* the deepest part of the selective sweep is upstream of *dhps*. This suggests a second site linked to *dhps* is also subject to selection.

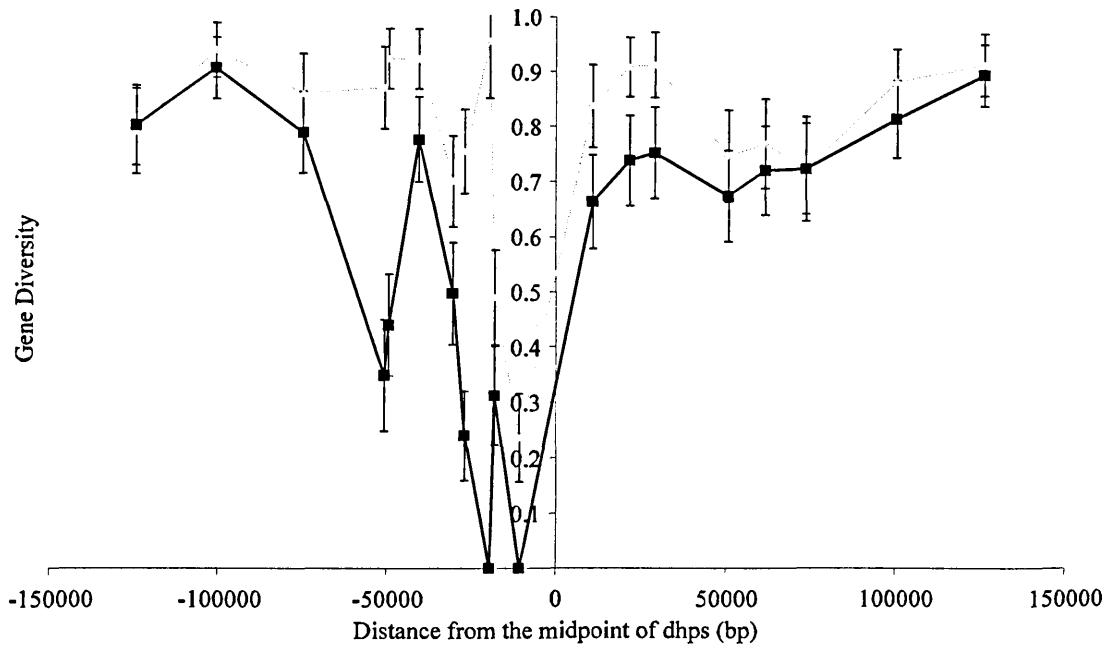


Figure 6-8: Gene diversity at microsatellite loci at increasing distances upstream and downstream from the midpoint of *dhps* on drug sensitive/ancestral chromosomes (dashed line) and *dhps* double mutant chromosomes (solid line) sampled from Mpumalanga. 95% confidence limits calculated from variance in gene diversity

Pop 1	Pop2	U125	U100	U80	U75	U60	U50	U50.2	U40	U30	U30.2	U20	U20.2	U10	*dhps*	D10	D20.2	D30.2	D50.2	D60	D70	D75	D100	D125	D150	D350	D550
KZN SAKAA	KZN SGEAA	0.65	0.11		0.17	0.15	0.20	0.29	0.27	0.37	0.56	0.26	0.48	0.62		0.74	0.28	0.31		0.38		0.27	0.10	0.16	0.13	0.11	
Mpu SAKAA	Mpu SGEAA		-0.01	n/a		n/a	0.52	0.48	0.15	0.20	0.51	0.95		Fixed		0.17	0.17	0.16			n/a		0.07		n/a	n/a	n/a

Figure 6-9: The direction and extent of significant differences in gene diversity between ancestral and double mutant chromosomes sampled from KwaZulu Natal and Mpumalanga. Shading indicates the following levels of significance Light grey = $p<0.05$; dark grey = $p<0.001$; black = $p<0.0001$.

i Comparing the double mutant lineages of Mpumalanga and KwaZulu Natal

Before committing to any further comparison of the selective sweeps present on double mutant chromosomes from KwaZulu Natal and Mpumalanga it was first necessary to address the question of shared ancestry. Archetypal resistant chromosomes were constructed using all loci where significantly reduced diversity was found. The allele with the highest frequency at every locus was considered the most likely archetype (Figure 6-10). Comparing KwaZulu Natal and Mpumalanga allelic archetypes it is clear that they are identical within the U50-D10 region; only U40, D20.2 and D30.2 were different. As the KwaZulu Natal sample was taken at an earlier time point, we considered that the alleles at the three loci are the original hitchhikers associated with the mutation event at *dhps* and that the variants are Mpumalanga specific. At a few loci there were two alleles that could be classified as archetypal alleles and these were not solely restricted to the outer portions of the selective sweep (U60, D125) but were also found at loci proximal to the selected site (U20). It is likely that these differences are products of recombination between other allelic chromosomes that have subsequently increased in frequency. In chapter 3 of this thesis we showed that epidemic like expansions of lineages had occurred in Mpumalanga and it is reasonable to assume that the differences between the archetypal double mutant chromosomes from both populations were formed in this way.

Population	U125	U100	U80	U75	U60	U50	U50.2	U40	U30	U30.2	U20	U20.2	U10	D10	D20.2	D30.2	D50.2	D60	D70	D75	D100	D125	D150
KZN SGEAA	348	289		316	317/308	121	217	181	181	226	402/406	90	109	152	409	203	169	264		197	123	386/376	221
Mpu SGEAA						121	217	187/175	181	226	402	90	109	152	415	215							

Figure 6-10: The archetypal haplotypes of flanking sequence loci for sites within the region of reduced diversity defined on the *dhps* double mutant chromosome. See text for method of definition of archetype alleles.

j Deterministic identification of the cause of the difference in size of the region of reduced diversity between the two populations

The Wiehe (Wiehe 1998) model of a selective sweep at equilibrium between recombination, selection and mutation was used to determine the differences in selection and required migration between the selective sweep present in the two populations. Unlike in KwaZulu Natal, no directly observed selection coefficient for the double mutant allele was available. We input parameters identical to those used for KwaZulu Natal; inbreeding coefficient (F=0.7) based on the number single genotypes present at *dhps* and effective population size (N_e =5100), and altered selection coefficient to find the line of best fit to the observed data. Assuming an initial frequency of the allele in the population of 20, the same as used in KwaZulu Natal, it was determined that the line of best fit occurs when s=0.08 (Figure 6-11). This is a reduction in the magnitude of the selection coefficient from that directly observed in KwaZulu Natal where s = 0.13. When s = 0.13, the number of migrants establishing the initial starting frequency of the double mutant allele was one, much less than the m=20 used for KwaZulu Natal.

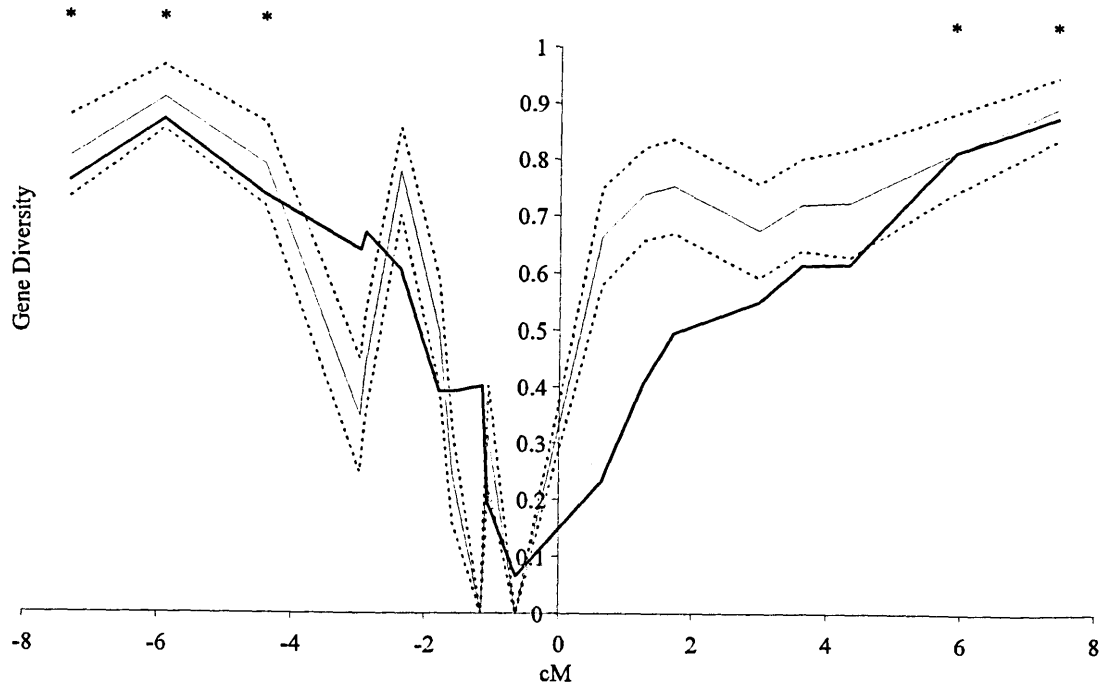


Figure 6-11: Deterministic prediction of the size and shape of the selective sweep around the *dhps* double mutant allele at equilibrium between recombination, mutation and selection. The parameters used are given in the text however: selection coefficient s = 0.08, m=20. The grey line represents the observed diversity with the dashed grey lines representing the 95% CI. The black line represents the model predication. Asterisks above indicate loci at which the modelled line fell within the 95%CI

V Discussion

In this thesis we have described selective sweeps around the key *dhfr* resistance allele, the triple mutant, and now around the *dhps* double mutant allele. Selection operating on these two loci must be equivalent at any given point in time as the two drugs which target these enzymes are co-administered. Therefore, it is interesting to compare the selective sweep around the *dhps* double mutant allele with that around the *dhfr* triple mutant allele. We have described the selective sweeps around both the *dhfr* and *dhps* alleles in two populations, KwaZulu Natal and Mpumalanga. By comparing sweeps around *dhfr* and *dhps* within a population, the effective recombination rate and selection pressures are effectively standardised, as are the other factors contributing to the dimensions of selective sweeps.

The selective sweep present on resistant chromosomes from KwaZulu Natal in 1999 extended from approximately 125kb upstream to 150kb downstream with a central core of 125kb of unbroken significantly low gene diversity. This 275kb region far exceeded the 50kb region of reduced gene diversity present on *dhfr* triple mutant chromosomes also sampled in KwaZulu Natal in 1999 (Chapter 4). In Mpumalanga in 2001,a smaller se lective sweep was found around the resistant *dhps* allele. In this case the region of significantly reduced diversity covered a 80kb region similar, but slightly larger than the 70kb around *dhfr* triple mutant alleles sampled contemporaneously.

a Why is the selective sweep around the dhps double mutant wider than that around dhfr: Time?

The difference in the size of the selective sweeps around the two genes is notable, but consistent with our observation that selective sweeps reduce in size over time. The appearance of the resistant *dhps* allele was much more recent than *dhfr*. In 1995 the *dhfr* triple mutant allele had a frequency of 22%, whilst the *dhps* double mutant allele was absent. The longer that an allele has existed in a population the greater the length of time that recombination will have had to act to reduce the size of a selective sweep, and as neither of the alleles discussed have reached fixation in the population, there are still opportunities for recombination, with sensitive chromosomes.

b Why is the selective sweep around the dhps double mutant wider than that around dhfr: Migration and starting frequencies?

One of the most striking observations of the work describing the selective sweeps present on the resistant *dhfr* chromosomes was the role of migration in determining the size and shape of the region of reduced gene diversity (Chapter 4). Using a deterministic model of a selective sweep (Wiehe 1998; Nair et al. 2003), we found that rather than an initial starting frequency of $1/N_e$ in the population, the model predicted that multiple migrants carrying the resistance allele were involved in establishment of the allele in these populations.

Using constant parameters for mutation and recombination in addition to a directly observed selection coefficient, we observed in KwaZulu Natal that an estimation of an initial frequency of the *dhfr* triple mutant allele in the population generated by 20

migrants generated a good fit to the observed data. Neutral markers indicate extensive gene flow between sites in the southeast Africa region (Chapter 3), so this value seems reasonable. Applying the selection coefficient observed in KwaZulu Natal to Mpumalanga the model indicated that 200 migrant triple mutant chromosomes founded the initial starting frequency. Since SP use began in Mpumalanga nine years later than in KwaZulu Natal, the regional frequencies of the *dhfr* triple mutant allele were very probably a lot higher. Chapter 3 showed that the frequency of the resistance alleles in Mozambique in 2001 was high. This country had very little SP use so the high frequencies suggests passive diffusion gene flow into the population. Of course the selection coefficient directly observed in KwaZulu Natal may be an underestimate of that in Mpumalanga but both populations have a similar malaria epidemiology and treatment policy and treatment seeking behaviour are similar in the two provinces.

In the case of the *dhfr* resistance allele, the shape of the selective sweep that we observed is a product of migration and a low selection coefficient. How migration performs this role will be discussed in the following chapter, however, to briefly summarise, it is evident that the moderately high starting frequencies of an allele in a new population places the allele in the new population at frequency closer to fixation than a low starting frequency. If the allele had an initial frequency in the population equivalent to $1/N_e$ the directly observed selection coefficient would take longer to reach the current allele frequency in the population and thus greater opportunity for recombination reducing the size of the selective sweep.

Using the selection coefficient estimated from the *dhps* double mutant allele frequency change, we found that in KwaZulu Natal the initial starting frequency will

have required 20 migrants into the population carrying the double mutant allele. This estimate of 20 is, however, an upper boundary as the selection coefficient calculated and used in the model makes the assumption that the double mutant allele appeared in 1995 (and therefore that 12 generations have passed), whereas in reality it was undetected at this time point and may have appeared subsequently. If so, the selection coefficient would have been higher. Applying the same selection coefficient to Mpumalanga identified that the initial starting frequency of the allele in the population was approximately $1/N_e$.

Thus we have observed that the initial starting frequencies of the *dhps* double mutant allele in KwaZulu Natal and Mpumalanga were lower than those estimated for the *dhfr* triple mutant allele, and this was most pronounced in Mpumalanga. What underlies this difference in initial starting frequencies? In Chapters 2 and 3 we described heterogeneity in the *dhfr* and *dhps* allele frequencies between populations across southeast Africa and in all these populations the frequencies of the *dhps* resistance alleles were lower than that of the *dhfr* resistance alleles, which given the recent emergence of *dhps* in the population is unsurprising. We would expect that this difference in the resistance allele frequencies between loci would extend to the proportions of the total number of migrants occurring through natural levels of gene flow that carry either of these resistance alleles. The high starting frequencies estimated for the triple mutant alleles in Mpumalanga reflects the high frequencies of the allele in neighbouring populations when use of SP selection began in the South African province, whereas the lower estimate of migrants with the *dhps* double mutant allele reflects its lower frequency in neighbouring populations.

c Why is the selective sweep around the dhps double mutant wider than that around dhfr: Selection coefficients?

The final major difference between the resistant *dhps* and *dhfr* alleles is the size of the selection coefficient operating on them. Roper et al estimated the selection coefficients for both the *dhfr* triple mutant and the *dhps* double mutant alleles from frequency changes in KwaZulu Natal between 1995 and 1999 (Roper et al. 2003). The magnitude of the selection coefficient for the *dhps* allele, estimated over 12 generations, was notable for being almost three times larger than that estimated over the time period at *dhfr* (0.13 vs. 0.048). As we have stated earlier the selection pressure applied by SP use is simultaneously applied on both loci so at first sight the difference in selection coefficients is curious.

However, we know that the *dhps* allele arrived into the population when the resistant *dhfr* allele was already established. The combination of the *dhfr* triple mutant and the *dhps* double mutant together is predictive of survival of treatment doses of SP (Omar, Adagu, and Warhurst 2001; Kublin et al. 2002; Dorsey et al. 2004), whereas *dhfr* alone is not (Alifrangis et al. 2003).

Clearly the two locus resistant genotype has higher fitness than other genotypes as evidenced by observation in chapters 2 and 3 that there is strong linkage disequilibrium between the *dhfr* triple mutant and the *dhps* double mutant (Omar, Adagu, and Warhurst 2001; Pearce et al. 2003). It can be postulated that the measurable association of the *dhfr* triple mutant and the *dhps* double mutant in the field and post treatment is what is observed in the high selection coefficient. To test this hypothesis would require a second population where a resistant *dhps* allele had

appeared in the population first. Such populations can be found in west Africa, where alleles at *dhps* other than the A437G+K540E double mutant can be found in high frequencies (Eberl et al. 2001; Marks et al. 2005) although the association of the west African *dhps* alleles with sulphadoxine resistance is not as clear cut as the east African double mutant described here.

d Effects on the dhps selective sweep of the presence of the dhfr triple mutant allele

In addition to speculating whether the observed association between the *dhfr* and *dhps* resistant alleles has had an effect on the selection coefficient of the *dhps* double mutant allele, it is interesting to speculate what effect the association of the two alleles would have on the selective sweep around the *dhps* double mutant allele over time.

The *dhfr* triple mutant can survive treatment with SP, but addition of the *dhps* double mutant allele increases the survival of the parasite. However, as the frequency of the *dhfr* triple mutant allele increases in the population, it would have the effect of reducing the fitness differential between the *dhps* double mutant and *dhps* sensitive allele, as arguably both *dhps* alleles would survive SP treatment if present in a parasite with the *dhfr* triple mutant. The reduction in fitness differential would increase the time to fixation for the *dhps* double mutant allele, increasing the time that recombination would have to reduce the selective sweep flanking the allele.

We observed in Mpumalanga that the selective sweeps present on the *dhfr* and *dhps* resistant chromosomes were 70kb and 80kb respectively. The *dhps* double mutant

allele was at a lower frequency in the population than the *dhfr* triple mutant allele (10% vs. 52%) and thus more prone to outcrossing decaying the selective sweep. Yet it had existed in the population for a shorter period of time and thus less time for recombination to occur.

It is interesting to speculate whether the rate in reduction of the size of the selective sweep on the *dhps* resistant chromosomes would be quicker than observed on the *dhfr* triple mutant chromosomes despite both being subject to the same underlying effective recombination rate. It may be expected that both selective sweeps would reach a unity when there is equilibrium in the population genetic parameters. Yet, the effect of a reduced fitness differential would lead to the selective sweep on the *dhps* resistant chromosome being smaller at fixation than that present on the *dhfr* resistant chromosomes.

However a more probable explanation for the similarity in size between the selective sweeps observed in Mpumalanga around the *dhfr* triple mutant allele and *dhps* double mutant allele, despite the difference in age of the alleles, is the instability of large selective sweeps over smaller ones. The larger a region of reduced diversity is the greater the length of sequence within which stochastic recombination breakpoints can occur. In Mpumalanga these would occur rarely and larger sequences have a greater probability of crossover. It would be interesting to describe the sweep on resistant chromosomes sampled in Mpumalanga at a later time point to observe the rate of decay of the sweep from this starting point similar sized sweeps.

VI Concluding remarks

Watkins et al have argued in the past that sulphadoxine plays no role in the killing effect of SP and that the variety of point mutations at *dhps* is argument against its role in resistance (Watkins et al. 1997; Watkins et al. 1999). This is disputed by a number of lines of evidence such as the proven role of *dhps* through allelic replacement and gene knock out transfection studies (Triglia et al. 1997; Triglia et al. 1998; Wang et al. 2004b) and the predictive association of *dhps* mutations with treatment failure (Omar, Adagu, and Warhurst 2001; Kublin et al. 2002; Dorsey et al. 2004).

We have shown the presence of a large selective sweep around the *dhps* double mutant allele in two South African populations. In conjunction with the fact that these resistant alleles increased in frequency under SP drug pressure, the presence of a selective sweep is clear indication that the *dhps* double mutant allele is under positive directional selection. The role of this double mutant allele and the importantance of sulphadoxine resistance to parasite survival is very clear.

Chapter 7 Metadiscussion

This final section of this thesis is divided into two parts. The first part will bring together the conclusions of the five chapters of experimental work. The second part discusses the conclusions in the wider context.

I Part 1:

a Local and regional heterogeneity in haplotype frequencies and the relationship with selection pressure

We began by developing a novel technique for the high throughput detection of the point mutations associated with sulphadoxine/pyrimethamine (SP) resistance at the loci encoding the drug targeted enzymes dihydrofolate reductase, *dhfr*, and dihydropteroate synthetase *dhps* (Pearce et al. 2003). Based on the sequence specific oligonucleotide probing dot blot (SSOP-dot blot) hybridisation method (Conway et al. 1999), we developed an approach that enables the determination of "allelic haplotypes" of the different SNPs coding for amino acid substitutions. Each allelic haplotype confers known differences in drug tolerance, as determined by in vitro studies, for example the triple mutant allele comprised of substitutions N51I+C59R+S108N is 1600 times less sensitive to pyrimethamine than the sensitive allele (Wu, Kirkman, and Wellems 1996). Using the SSOP-dot blot approach we described the frequencies of the different SP resistance alleles present in field samples collected as dried bloodspots from infected patients in three sites 150 miles apart in northern Tanzania. We found that there were a small number of alleles (5 at *dhfr* and 3 at *dhps* – see figure 7-1) and these were common to each population.

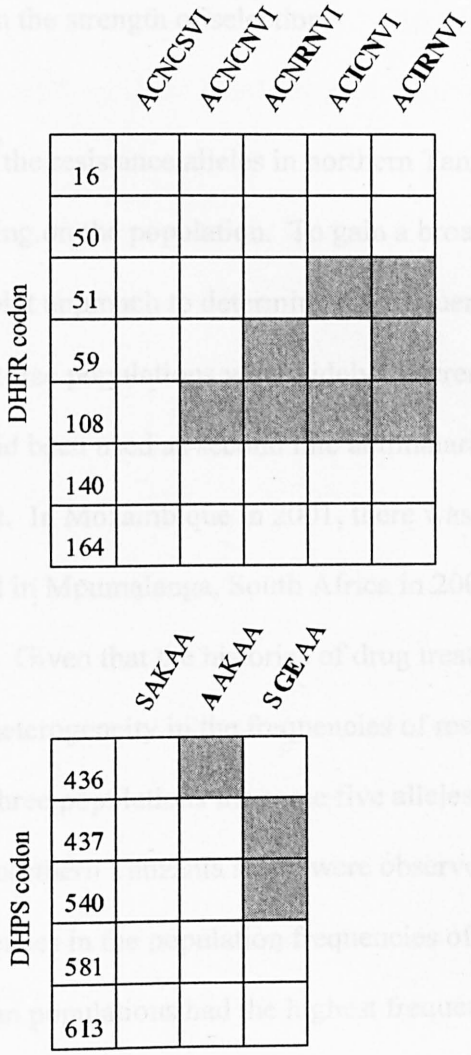


Figure 7-1: Scheme of *dhfr* and *dhps* alleles found within in this study. Shaded blocks indicate the site at which the constituent SNPs of each haplotype arise. Names of the alleles are composed of the amino acids present at each of the sites, in consecutive order of the codon number, described as having a role if SP resistance.

The *dhfr* triple N51I+C59R+S108N mutant and the *dhps* double A437G+K540E mutant were the most highly resistant alleles identified. The frequencies of these were high and differed significantly between the three sites strongly suggesting regional heterogeneity in the strength of selection.

The high frequencies of the resistance alleles in northern Tanzania suggested a history of drug selection operating on the population. To gain a broader picture of the region we used the SSOP-dot blot approach to determine the frequency of *dhfr* and *dhps* alleles in samples from three populations with widely different selection histories. In Tanzania in 2000, SP had been used as second line antimalarial treatment after chloroquine for 18 years. In Mozambique in 2001, there was officially no use of SP in malaria treatment and in Mpumalanga, South Africa in 2001, SP had been firstline antimalarial since 1997. Given that the histories of drug treatment policy were different, we expected heterogeneity in the frequencies of resistance alleles between the populations. In all three populations the same five alleles at *dhfr* and three at *dhps* described in the earlier northern Tanzania study were observed. As expected we found significant differences in the population frequencies of each allele, but surprisingly Mozambican populations had the highest frequencies of the highly resistant alleles. It was clear that past national policy was not a good indicator of the resistance status of the parasite populations. We speculated that selection had occurred through unofficial use; there is anecdotal evidence of Mozambicans crossing the border seeking treatment in nearby border clinics in South Africa, where SP is used for treatment of their infections.

b The shared ancestry of resistance alleles sampled across a wide region suggests significant gene flow

Microsatellite loci in the immediate flanking region of the *dhfr* and *dhps* genes have been used to infer the ancestry of these point mutation haplotypes (Roper et al. 2003). Comparison of South African and northern Tanzanian samples showed that alleles comprised of multiple mutants were monophyletic in origin and had dispersed throughout the southeast African region. The exception was the *dhfr* double mutant N51I-S108N allele, for which two ancestral lineages were identified; a common widely dispersed lineage and a rare independent lineage in South Africa.

To quantify the extent of regional gene flow we examined allele frequencies at 8 unlinked neutral microsatellite loci in Tanzania, Mozambique and South Africa. Nesting tests of population differentiation by scales of distance we found high amounts of gene flow both between local populations and across the region, despite the regional and local differences in the frequencies of resistance alleles. The evidence of strongly homogenising gene flow across a region, but strong differentiation between populations in the frequencies of resistance alleles is evidence of the heterogeneity in selection intensity across the region. The observation of a high level of gene flow explains the common ancestry of the resistance alleles region wide and is consistent with early reports of highly mobile pyrimethamine resistance (Clyde 1967). Pyrimethamine resistance was selected by intense prophylactic use of pyrimethamine in the Muheza district of Tanzania in the 1950s. Diffusion of resistance up to 100miles from the original focus was described. Such a pattern of diffusion outwards from the initial site of intense selection to populations where selection was absent as observed in the 1950s and the low population differentiation

between sites described here, is strongly suggestive of a diffusion of resistance genes through a parasite population that is continuous over a great geographical distance in southeast Africa.

c Transcontinental gene flow

As a source of gene flow, the movement of people particularly across large distances through aviation has also contributed to the spread of SP resistance. Through comparison of the flanking markers around triple mutant *dhfr* alleles sampled from southeast Africa and numerous *dhfr* resistance alleles from southeast Africa, we identified the southeast African triple mutant allele as a transcontinental migrant from Asia (Roper et al. 2004). This reinforces the view that de novo development of SP resistance mutants in southeast Africa has occurred rarely. The lineages of the double mutant C59R+S108N and the two N51I+S108N alleles present in southeast Africa are highly likely to be "home grown" as their flanking haplotypes were different.

d Selective sweeps around the dhfr triple mutant allele in Africa

dhfr is found on chromosome four; *dhps* is found on chromosome 8. By comparing diversity at microsatellites the length of chromosomes carrying sensitive and resistance alleles at both these loci, we showed a loss of diversity specifically around mutant alleles at both loci.

We compared selective sweeps around the *dhfr* triple mutant allele on chromosomes sampled from Tanzania, and the KwaZulu Natal and Mpumalanga provinces of South Africa. The inner part of the sweep shared identity with the selective sweep present

on southeast Asian *dhfr* resistance chromosomes confirming previous data (Roper et al. 2004). The width of the selective sweep in African populations was reduced relative to that described in Asian populations (Nair et al. 2003). The sweep was widest in KwaZulu Natal in 1995 where the region of significantly reduced gene diversity covered a smaller region of 70kb, smaller than that recorded in southeast Asia (>100kb) (Nair et al. 2003). This difference in size is even more striking when one considers that at the time of sampling in southeast Asia, SP selection pressure had been absent for at least ten years (Nosten et al. 2000).

e Interaction between recombination and selection and the additional role of gene flow

Between the three southeast African populations we observed significant differences in the sizes of the selective sweeps around the triple mutant allele. The key factors in explaining these differences were the intensity of drug selection operating in the region and the effective recombination rate, which varies according to transmission intensity (Paul et al. 1995; Dye and Williams 1997). Recombination acts to limit the distance from the favourable allele that hitchhiking flanking alleles persist, whereas selection by increasing the frequencies of both favoured alleles and hitchhikers acts to cement these associations into the population by fixation (Smith and Haigh 1974; Kaplan, Hudson, and Langley 1989). The opposing effects of these parameters mean that neither recombination nor selection can be treated independently.

We observed that the effect of recombination over time was to narrow considerably the selective sweep around the *dhfr* triple mutant allele. In KwaZulu Natal the region of reduced gene diversity was narrowed from 70kb in 1995 to a region of 50kb in

1999, despite ongoing selection through the use of SP as first line treatment. Sampling from the same population standardises the effective recombination rate between the two time points, and highlighted how selection for the triple mutant allele in the population was insufficient to counter the high recombination rate. Recombination reduces the size of a selective sweep when derived chromosomes recombine with ancestral chromosomes. This occurs when drug coverage in the population is not absolute. In African populations absolute drug coverage is rare as a significant proportion of infections are asymptomatic, disease recognition is poor (von Seidlein et al. 2002) or other forms of treatment are used; Knowledge-Attitude-Practice studies show that in the marketplace a number of alternatives to SP including non antimalarials are widely available and are used in malaria treatment (Djimde et al. 1998).

Description of selective sweeps present on triple mutant chromosomes sampled from Tanzania and Mpumalanga, suggested that there is a limit to the variability in the size of selective sweeps despite differences in selection histories and transmission intensities. This may imply stability to the selective sweep even when resistance alleles are not at fixation. The region of reduced gene diversity on the triple mutant chromosomes sampled from Tanzania was expected to be very small as the population has a high effective recombination rate and selection in the population was at the time of sampling was by second line only. However, it was 50kb wide, the same as that found in KwaZulu Natal in 1999 after 11 years as firstline. This suggested that another factor was involved that compensated for the weak selection pressure and maintained the region of reduced diversity in the face of the high effective recombination rate. This factor was gene flow as determined by its incorporation into a deterministic model (Wiehe 1998; Nair et al. 2003) within the

parameter of the initial starting frequency of the selected allele. Using estimates of selection based on measured frequency change and recombination rate adjusted for intrapopulation inbreeding rates, we showed that the initial starting frequency was a key variable reinforcing the importance of gene flow in the evolution of SP resistance in this region.

The deterministic model used (Wiehe 1998; Nair et al. 2003) predicts a selective sweep at equilibrium between the parameters of selection, mutation and recombination. If this is not the case then the initial frequency required to compensate for the low selection coefficient may be an over estimate. However it is doubtful that the values for the initial frequency in the population would ever have been equal to $^{1}/_{Ne}$,bec ause we know that de novo mutations giving rise to the resistance alleles occurred in Asia (Roper et al. 2004) and subsequently starting frequencies will always have been determined by gene flow. It makes sense that the selective sweep often requires an initial starting frequency greater than $^{1}/_{Ne}$.

f The selective sweep present on C59R+S018N dhfr double mutant chromosomes

Having described the selective sweep around the most highly pyrimethamine resistant allele in the region, we carried out the same analysis of selective sweeps around the *dhfr* C59R+S108N double mutant allele in two populations, South Africa and southern Tanzania. The observation that the level of resistance to SP conferred by this allele is insufficient to survive treatment doses of pyrimethamine in the hosts' plasma and selection for this allele over the sensitive allele is reliant on infection of a host, or emergence from the liver stages into the bloodstream within a specific

window of sub therapeutic drug levels that occurs 7-15 days after treatment (Watkins et al. 1997), led to the expectation that the size of the selective sweep on the double mutant chromosomes would be smaller than on the triple mutant chromosome. We used the deterministic model (Wiehe 1998; Nair et al. 2003) to determine the size based on a directly observed selection coefficient from Tanzania and found that the real sweep was much larger than predicted. It is probable that this allele was present in the southeast African parasite population prior to the triple mutant allele and perhaps dates from the earliest reports of Pyr resistance in the 1950's (Figure 7-2) (Peters 1970). If that is the case the recombination rates and the irregularity of selection would reduced the extent of the region of reduced gene diversity. Thus it was surprising to find that in the two study populations a substantial selective sweep was still detectable.

g Comparing the selective sweeps around the dhfr double mutant and triple mutant alleles

In Tanzania the width of the selective sweeps present on the double mutant (U30-D10) and triple mutant chromosomes (U40-D10) sampled from the same time point were similar. In Mpumalanga we found that the difference was more pronounced, with a wider selective sweep around triple mutant alleles (U60-D10) than double mutant alleles (U30-D10) probably reflecting the greater relative fitness of the former over the latter.

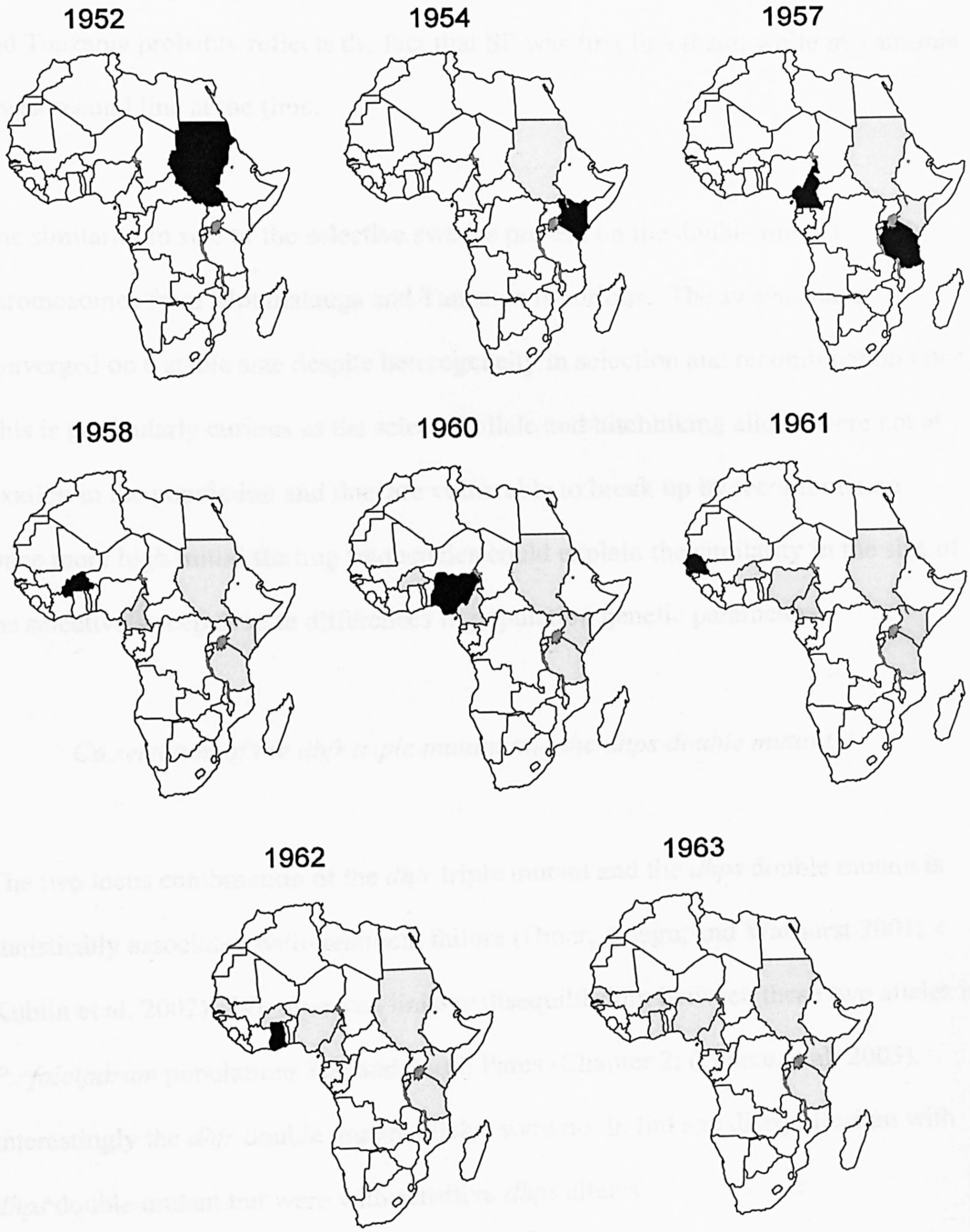


Figure 7-2: The emergence of pyrimethamine resistance on continental Africa between 1952 - 1963 adapted from (Peters 1970)

The differences in size of sweep around the triple mutant allele between Mpumalanga and Tanzania probably reflects the fact that SP was first line there, while in Tanzania it was second line at the time.

The similarity in size of the selective sweeps present on the double mutant chromosomes from Mpumalanga and Tanzania is curious. The sweeps have converged on a stable size despite heterogeneity in selection and recombination rates. This is particularly curious as the selected allele and hitchhiking alleles were not at fixation in the population and thus are vulnerable to break up by recombination. Once more high initial starting frequencies could explain the similarity in the size of the selective sweep despite differences in population genetic parameters.

h Co selection of the dhfr triple mutant and the dhps double mutant

The two locus combination of the *dhfr* triple mutant and the *dhps* double mutant is statistically associated with treatment failure (Omar, Adagu, and Warhurst 2001; Kublin et al. 2002). We observed linkage disequilibrium between these two alleles in *P. falciparum* populations Hai and South Pares (Chapter 2; (Pearce et al. 2003). Interestingly the *dhfr* double mutant alleles were not in linkage disequilibrium with *dhps* double mutant but were with sensitive *dhps* alleles.

The linkage disequilibrium between the *dhfr* triple mutant and *dhps* double mutant alleles is indicative of the co selection by the drug components of SP. It shows that combined the two alleles confer greater fitness than each separately supporting the view that they both have a role in the survival of SP treatment in the field.

The appearance of the *dhfr* triple mutant allele preceded emergence of the *dhps* double mutant allele. In KwaZulu Natal in 1995 the *dhps* double mutant allele was absent from the population, and by 1999 it had achieved a frequency of 15%, an estimated selection coefficient of 0.13, assuming 3 generations per annum (Roper et al. 2003). This was greater than the selection coefficient estimated for the *dhfr* triple mutant over the same time period. It is worth considering that the larger selection coefficient for the *dhps* allele is a reflection of this advantage of the temporary association. Any advantage conferred on the *dhfr* triple mutant allele through association with the *dhps* double mutant allele is not clear from its selection coefficient estimated over the same period as the *dhps* allele was absent at the earlier time point.

One of the original questions we were to attempt to address in this thesis when first designing the study, was whether we could use the width of a selective sweep present on a selected chromosome as a proxy for the strength of selection occurring in that population. Clearly for such a measure to be made the effective recombination rate between the two populations would need to be standardised, hence use of a longitudinal dataset from KwaZulu Natal. However, it soon became apparent with the observation of the important role of gene flow in the formation of the selective sweeps, that it would be impossible to distinguish between selection occurring in the study site and selection occurring in neighbouring populations.

Therefore we propose that a better proxy for the intensity of selection occurring in the population would be the linkage disequilibrium between the two locus *dhfr*/*dhps* combination of the *dhfr* triple mutant and *dhps* double mutant as we would expect that if there was no selection these two alleles would be rapidly uncoupled.

i The selective sweep present on the dhps double mutant chromosomes

In KwaZulu Natal in 1999, it was notable how much larger the selective sweep around *dhps* was than around the *dhfr* triple mutant allele population in 1999 (>275kb vs. 50kb). In neighbouring province of Mpumalanga where the effective recombination rate and selection coverage is similar the selective sweep on double mutant chromosomes sampled in 2001 was narrowed to 80kb and *dhfr* was 70kb.

The reduction in size of the selective sweep is due to disequilibrium between recombination and selection pressure. Deterministic modelling (Wiehe 1998; Nair et al. 2003) of the predicted size of the selective sweep at equilibrium between known effective recombination and microsatellite mutation rates and directly observed selection coefficients, found that as with sweeps described on chromosome 4, there was a requirement of an initial frequency of the allele in the population greater than $^{1}/_{Ne}$. The value for the number of migrants estimated by the model was 20, but as the selection coefficient used was the minimum estimate this number of migrants could be lower. In Mpumalanga the initial starting frequency of the allele was closer to $^{1}/_{Ne}$. As the frequency of the double mutant allele in the region is low the low initial starting frequencies of the allele predicted by the model are unsurprising, particularly as selection was operating at the time, whereas arguably *dhfr* resistance alleles accumulated passively in some populations prior to selection.

The de novo origin of the *dhps* SGEAA double mutant allele is unknown, however it's probably of African origin, being found in high frequencies throughout east Africa, but very low prevalence in southeast Asia and South America, where the

SGKGA and the SGEGA resistance alleles are the common types. Thus, the initial frequency of the allele in perhaps Tanzania or Malawi will have reflected the de novo mutation event ($^1/_{Ne}$). Since then gene flow has clearly played the important role in establishing the mutation in the KwaZulu Natal and Mpumalanga populations.

j Features common to the three selective sweeps described

None of the three resistance alleles,*dhfr* triple, *dhfr* double or *dhps* double have achieved fixation in the populations studied. The rate that the frequency of the allele increases has implications for the size of the sweep size because it determines the opportunities for outcrossing. In African populations the expectation is that selective sweeps present around the drug selected alleles will always tend towards a smaller size reflecting the high rate of recombination and the weaker selection. It might also be expected that selective sweeps in African parasite populations will be less permanent than in malaria populations elsewhere.

One recurring observation was the asymmetrical nature of the regions of reduced gene diversity around all three alleles. In each case, diversity was consistently returned more quickly on the downstream side of the selected site. This was particularly pronounced on the *dhfr* triple mutant chromosomes. On the *dhfr* double mutant chromosomes the asymmetry identified could be secondary, occurring through recombination with the expanding population of triple mutant chromosomes. On the *dhps* double mutant chromosomes the asymmetry was clear, with the deepest part of the region of reduced gene diversity being maintained on the upstream side consistently over two time points.

In the absence of a better explanation we argue that the stochasticity of recombination breakpoints was to blame (Kim and Stephan 2002), with more recombination events occurring on the downstream side of the selected site. Other causes of the asymmetry could be recombination hot/cold spots or an additional site such as a compensatory mutation under selection (discussed below), but we observed that in the case of the *dhfr* triple mutant chromosome where the asymmetry is most pronounced, over time the deepest part of the selective sweep condenses onto *dhfr*.

Curiously, the same asymmetry was found in the southeast Asian selective sweep around *dhfr*. The triple mutant allele was introduced from Asia to Africa (Roper et al. 2003). The common asymmetry to sweeps in the two continents supports the view that the introduction involved multiple migrants. If a single migrant had entered the population, the asymmetry would happen by chance to be on the same side of *dhfr* as in southeast Africa. By contrast if a larger number of migrants entered the African parasite population, and undergone a rapid expansion, the same asymmetry as southeast Asian chromosomes would be transferred.

II Part 2:

a Why do dhfr and dhps multiple mutants emerge so rarely?

Prior to recent discoveries, the predominant hypothesis explaining the evolution of resistance to SP suggested that resistance conferring mutations had occurred de novo frequently (reviewed in (Hastings 2004)). The argument for multiple de novo origins is a numerical one. The mutation rate of *dhfr*, for example, is sufficiently high

($2.5x10^{-9}$ mutations/*dhfr*/generation) that at least one mutation would arise at *dhfr* per patent infection (10^{10}-10^{12} parasites).

One explanation for the difference between empirical evidence and theoretical prediction is that the probability of survival of a new mutation in the population is low (Hastings 2004). If a beneficial allele arises during the course of a drug treated infection, for it to survive in the population it must be transmitted to other hosts. Only a small proportion of blood stage parasites develop into gametocytes (<1%) (Taylor and Read 1997; Anderson and Roper 2005), and it is possible that the new mutation can be lost. Furthermore, the transmission of the de novo mutant is dependant on the parasites survival against the hosts' immune defence. *Plasmodium falciparum* can perform antigenic switching, conferred by expressing different var genes and thus avoid immune destruction. (Gatton, Martin, and Cheng 2004) showed that parasite lineages that had only recently switched the var gene being expressed survived and went on to be transmitted. So while mutations may occur frequently, they would need to coincide with var gene switching to evade host immunity and be ingested in a mosquito blood meal. It is perhaps then not surprising that the majority of high level drug resistance has occurred in places where the majority of individuals are semi immune or non-immune.

It is argued that high transmission intensities would greatly assist in the establishment of the allele in the population (Mackinnon 1997), as in low transmission regions there is a lower chance of transmission to other hosts. However, high transmission increases the challenges to the hosts' immune systems, speeding up the development of clinical immunity in the host (Snow et al. 1997).

A second explanation for low rate of emergence of de novo mutations observed is that additional adaptations are required to alleviate or compensate for a fitness cost of the resistance mutation itself. There is strong evidence of fitness costs associated with *dhfr* mutations as identified in growth rate transfection experiments comparing *dhfr* sensitive and double mutant alleles in the presence and absence of pyrimethamine selection (Fohl and Roos 2003). In the case of chloroquine resistance the observed drop in the frequency of the *Pfcrt* chloroquine resistance allele in Malawi dropped after SP replaced chloroquine as first line antimalarial suggesting that fitness costs of resistance *Pfcrt* alleles still remain (Kublin et al. 2003). However, in bacteria compensatory adaptations to alleviate the cost of the antibiotic resistance have been shown to exist (Schrag, Perrot, and Levin 1997).

b Evidence for compensating adaptations

A compensatory mutation is one which alleviates the fitness cost of the mutations occurring at the drug targeted site. This may occur through a variety of mechanisms such as intragenic mutations in RNA or protein molecules, intergenic mutations altering the interaction between cofactors, bypass mechanisms or emergence of a new function and, finally, increased expression of the mutated protein (Maisnier-Patin and Andersson 2004).

Of the first of these compensatory mechanisms, there has been no descriptions of intragenic mutations other than those already described as associated with SP resistance, although substitutions at codons 51 and 59 could be argued to be compensatory as they enhance the substrate binding of enzymes with substitutions at codon 108 (Warhurst 2002).

In terms of compensatory overproduction of the mutated protein, it is interesting to note that DHFR binds its own mRNA as post transcriptional control of protein production (Goldberg 2002; Zhang and Rathod 2002). The mRNA is bound within the joining region between DHFR and thymidylate synthase, preventing production of excess protein. If a mutation blocking mRNA binding occurred in the joining region, over expression of *dhfr* could have the effect of swamping out the drug through producing excess target. I undertook a brief project sequencing the joining region of a small number of pyrimethamine sensitive and resistant parasites sampled from the field, but did not identify any polymorphism. Subsequently, it has been shown that there is no evidence for over expression of *dhfr* in resistant parasites (Wang et al. 2004a).

There is potential for compensatory adaptations in form of folate salvage, bypassing the function of *dhps*. The human host is incapable of de novo synthesis of folate, and all its folate is derived from the diet. *P. falciparum* carries out folate synthesis but is also capable of scavenging folate from the host, although the concentration of folate in the blood of the host is particularly low in populations suffering from malnutrition (Wang et al. 1997b; Wang et al. 1999; Wang et al. 2004a). Resistance substitutions in *dhfr* and *dhps* interfere with the normal interactions with the substrate molecule and the production of folate and in this situation the ability to scavenge folate is advantageous to the parasite. If exogenous folate is limited (such as in the case of malnutrition of host) the ability to synthesise folate from precursor molecules would be more efficient. Therefore it is interesting to speculate whether *P. falciparum* has evolved a compensating mechanism for actively accumulating folate.

Recent work describing the activity of the drug Probenecid on *P.falciparum*, suggests that the parasite may indeed have such a mechanism. Probenecid, is an inhibitor of drug efflux mechanisms and folate salvage, and has been shown in vitro to chemosensitise drug resistant *P. falciparum* to SP (Nzila et al. 2003). It is possible that probenecid inhibits the uptake of exogenous folate into the parasite, either through blocking parasite specific folate carrier proteins, similar to its action in blocking folate transporters in mammalian cells, or through blocking the accumulation of PABA, an essential precursor of the folate biosynthesis pathway(Nzila et al. 2003). Field studies testing the efficacy of SP in combination with probenecid in Nigerian children found that the addition of probenecid increased the treatment efficacy of SP (Sowunmi et al. 2004). As yet a target of probenecid activity remains unidentified.

The asymmetry of the selective sweeps present around the *dhfr* mutant and *dhps* mutant alleles could indicate the presence of compensatory adaptations. At the latter time point in the description of the selective sweep on the *dhps* double mutant chromosome (Mpumalanga 2001), the deepest part of the selective sweep seemed not to be on *dhps* but slightly upstream of it. Diversity had returned quicker on the downstream side. Bioinformatic analysis of the region within the deepest part of the selective sweep identified a function unknown open reading frame (PF08_0093), which had some homology to a mitochondrial folate carrier proteins in other systems. However, the homology was only between conserved mitochondrial signalling motifs rather than to specific functional motifs and so this protein may not actually have a folate specific function.

Similarly the selective sweep around the *dhfr* triple mutant allele was asymmetrical favouring reduced diversity on the upstream side of the gene. Interestingly, (Wang et al. 2004a) mapped the ability to salvage folate and escape the lethal effects of sulphadoxine damaged *dhps*,to a 48kb re gion of chromosome 4 containing 7 open reading frames (ORFs) including *dhfr*. If there were a compensatory mutation within this 48kb region, we would expect to see reduced diversity around it and the asymmetry could be evidence of this. The position of *dhfr* in this 48kb region is with the majority of sequence downstream of the gene, and the asymmetry points to the shorter upstream arm of the region, where three additional ORFs are located, of which two are function unknown and the closest to *dhfr* has homology to an RNA binding protein. However, in KwaZulu Natal the selective sweep reduced in size to be more condensed on *dhfr* suggesting that stochasticity of recombination was the more likely a cause of asymmetry than one of the three ORFs identified on the upstream portion of the 48kb region.

c The emergence of dhfr double mutant alleles

The establishment of the 'home-grown' *dhfr* double mutant alleles and there widespread distribution in the southeast African population is a curiosity, mainly because of their inability to survive treatment concentrations of SP (Watkins et al. 1997). We found supporting evidence in the lack of LD between the *dhfr* and *dhps* double mutant alleles, suggesting that combination conferred insufficient fitness to maintain the association of the alleles through drug selected assortative mating. The *dhfr* double mutant allele has at most a 45 day window within treated people during which they enjoy a positive selective advantage over sensitive alleles (Watkins et al. 1997). Only chance infection post treatment would favour the *dhfr* double mutant

allele over the sensitive allele. In KwaZulu Natal the triple mutant displaced the double mutant allele over the 4 year period (Roper et al. 2003).

Our analysis of the selective sweep around the C59R+S108N double mutant allele clearly demonstrates the action of selection. Yet it is hard to see what advantage they have in the face of SP use. Intense use of SP treatment will increase the number of hosts in the population with the 45 day selective window, but equally the expansion in numbers of resistant parasites will be limited by chance infection of hosts with lethal SP concentrations.

It is possible that the double mutant allele became established in a population where pyrimethamine rather than SP selection was operating. The earliest reports of pyrimethamine resistance in Africa came from regions where pyrimethamine alone was used prophylactically in community wide treatment programmes in the 1950s – 1960s summarised by (Peters 1970) (Figure 7-2). It is possible that the double mutant alleles date from this era. If the C59R+S108N double mutant allele has existed since the 1950s this would have implications for the expected size of the selective sweep around the double mutant allele described in Chapter 5, because we would expect that 120 generations of recombination would remove the selective sweep, particularly as the allele has never achieved fixation.

d The emergence of the dhfr triple mutant

The *dhfr* triple mutant allele in southeast Africa has been shown to have shared ancestry with the clade of *dhfr* resistance alleles found in Southeast Asia. It is tempting to speculate that the triple mutant allele was introduced into African

populations at the same time as the *Pfcrt* resistance allele, also of Asian origin (Wootton et al. 2002). Certainly pyrimethamine resistance was well established in Thailand in the 1980s (White 1992), which is when chloroquine resistance first appeared in Africa (Payne 1989). It can be speculated that the triple mutant allele 'hitch-hiked' with chloroquine resistant *Pfcrt* for sufficient generations to increase the frequency of the *dhfr* triple mutant allele even in the absence of pyrimethamine selection (Roper et al. 2004). However, recombination would rapidly uncouple the association unless selection for the *dhfr* triple mutant resistance allele was occurring concomitantly with chloroquine use. This might have occurred once chloroquine treatment failures were treated with SP. SP became second line antimalarial in the majority of southeast African countries after 1982 (Mutabingwa et al. 2001).

It is probable that the *dhfr* triple mutant allele entered continental Africa through southeast Africa, as it is here that the earliest reports of high numbers of SP treatment failures were made (Ronn et al. 1996; Trigg et al. 1997; Wang et al. 1997a). It is not known how the southeast Asian triple mutant allele made its' way into southeast Africa; it can be speculated that having expanded out of southeast Asia into the Indian subcontinent and migration between southeast India and Africa had introduced the resistance allele into southeast African parasite populations. Alternatively it may have come directly from Thailand. In either case the amount of gene flow into Africa would be restricted by the proportion of travellers infected.

A recently published model (Santiago and Caballero 2005) describes the effects on genetic hitchhiking when a favourable mutation arises in only one population but enters the other through migration. The major prediction of the model is that the more restricted the gene flow between the two subpopulations, the narrower the

region of reduced gene diversity around the favourable allele in the second population. Where migration is low this causes an increase in diversity within the selective sweep and only at loci very close to the selected site does gene diversity remain at levels in the population of origin. It is likely that the source of higher diversity is recombination between the migrant sweep and the local population. As population differentiation approaches panmixia the very high number of migrants into the second population maintains the reduced diversity around selected alleles. Only at panmixia does the diversity flanking the beneficial allele in each 'population' become equivalent.

What does this model tell us about the effects of gene flow in the populations described in this thesis? In the first scenario, the *dhfr* triple mutant allele came directly from southeast Asia. Using unlinked microsatellites, a F_{ST} value of ~0.1 was estimated for comparisons between Southeast Asian and African parasite populations (Anderson et al. 2000a). This gives an estimate of 2.25 effective migrants per generation, or assuming symmetrical migration, 5-7 migrants per year, given 4-6 generations per year (Anderson and Roper 2005). Using the model described above to predict the effects on the selective sweep of migration between southeast Asia and southeast Africa, we find that there was likely to be an approximately 8% increase in diversity levels along the chromosome. Figure 7-3 shows the modelled diversity at increasing distance from *dhfr* along the chromosome in "Asia" where the favourable allele arose, then in "Africa" into which the allele migrated.

Note that even with these low rates of migration the width of the selective sweep is predicted to be only marginally narrowed. The level of diversity on the selected chromosomes in the "African" population returns to the pre selective sweep levels of

diversity approximately 10kb closer to the selected site than in the population of origin. This indicates that there is no theoretical basis to expect loss in the dimensions of the selective sweep due to migration to southeast Africa from southeast Asia. It is more likely that the difference in the size of the selective sweeps measured in southeast Asia and in Africa are due to events that have occurred within the African populations. Regrettably we do not have descriptions of the size of the selective sweep around the triple mutant allele in Africa immediately after its introduction from Southeast Asia and therefore can not directly verify the prediction from the model.

In the second scenario, namely that the triple mutant allele entered continental Africa via the Indian subcontinent, introduced into east Africa by numerous migrants travelling to and from Indian ex pat communities or as a result of flourishing trade links. As such the effect on the size of the selective sweep around the triple mutant allele in Africa would be even less than that resulting from migration from southeast Asia to Africa.

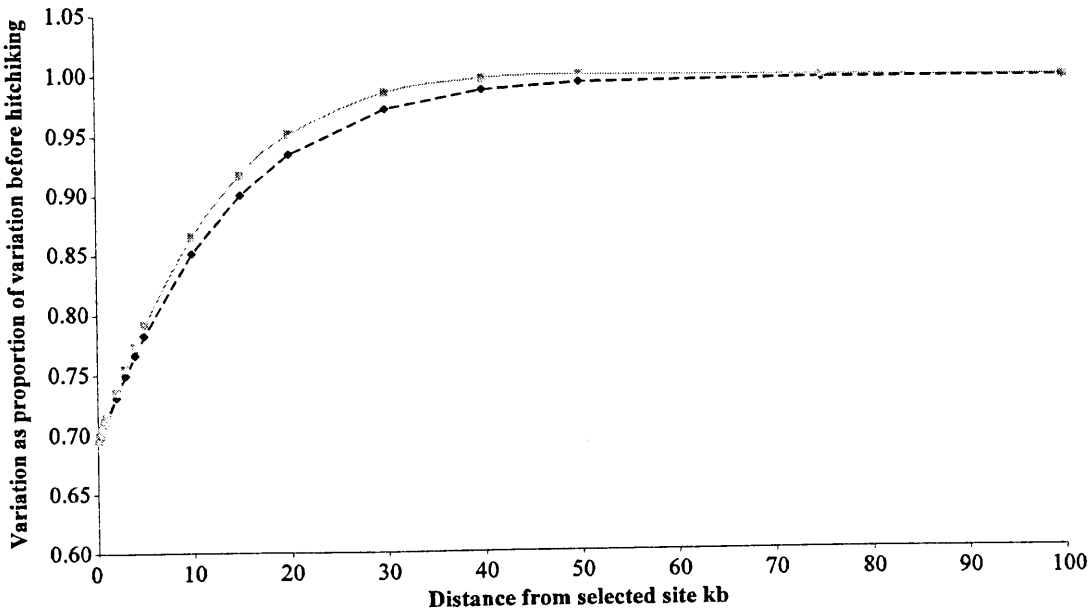


Figure 7-3: The predicted variation as a proportion of variation prior to hitchhiking on selected chromosomes present in subdivided populations. The parameters used in the model were as used by (Nair et al. 2003) in the (Wiehe 1998) deterministic model prediction of the shape of the selective sweep present in southeast Asia; population size N=10000, selection coefficient s=0.1, recombination rate r=0.000588 M/kb/generation F_{ST} = 0.1; Diamonds "Asia" where the favourable allele first arose, and Squares "Africa" to which the allele migrated. See text and (Santiago and Caballero 2005)

III Generalisable conclusions about the selective sweeps around SP resistance alleles in southeast Africa

a Movement of SP resistance alleles with in southeast Africa

A recurring theme in this thesis is the role of gene flow in the evolution of SP resistance in African populations. Firstly, microsatellite analysis of three parasite populations in Mozambique, South Africa and Tanzania found that there was little population differentiation between them, and thus heterogeneities in resistance allele frequencies were due to differences in the intensity of selection in these populations. Secondly gene flow was identified as passively disseminating and establishing the resistance alleles in populations throughout the region. We observed that the starting frequencies of an allele in the population as predicted by the model broadly reflected it frequency in neighbouring populations and the history of selection in the population. Passive accumulation of pyrimethamine resistance alleles in the absence of drug selection explaining the higher starting frequencies compared to the initial starting frequencies of the later emerging *dhps* double mutant allele.

The role of gene flow in the evolution of SP resistance has clear implications for the policy approach to maintaining the useful treatment lifetime (UTL) of antimalarial drugs as it is evident that unilateral national approaches to the use of antimalarials are detrimental. Whilst not making the drug unusable, large scale gene flow ensures that resistance alleles are present throughout the region when selection occurs, limiting the usefulness of the drug for neighbouring states. If the next generation of antimalarial treatments are to be protected it is key that the importation of parasites into Africa is limited to prevent migration of resistance alleles and secondly

multilateral antimalarial treatment policies need to be adopted to maximise the UTL of the drug in a region.

The gene flow between the southeast African populations described in this thesis is sufficiently high that the model of (Santiago and Caballero 2005) would predict no effect on the diversity levels of a selective sweep as it moved from one population to another. The numbers of migrants are sufficiently high that associations between hitchhiker and selected site are maintained. This is particularly pertinent considering our observation of the similarity in the sizes of the selective sweeps across southeast Africa despite very different effective recombination rates and selection histories. For example, a 50kb region of reduced gene diversity was described on the *dhfr* triple mutant chromosomes sampled from both Tanzania, where the effective recombination rates is high and KwaZulu Natal in 1999 where it is low. The similarity in the dimensions of the selective sweeps in the two populations despite different recombination and selection parameters, was deterministically modelled (Wiehe 1998; Nair et al. 2003) as due to the 100 times greater initial starting frequency of the allele in Tanzania. These high initial frequencies of the allele in Tanzania in the absence of selection slow the breakdown of associations between selected site and hitchhiker in the absence of selection by increasing the proportion of 'recombining' resistant chromosomes where there is minimal effect on diversity in the selective sweep.

b Detection of drug selected loci in African Populations

If there is an additional locus involved in SP resistance, describing patterns of diversity in the genome would be a reasonable method of detection (Anderson 2004),

although in the light of the findings in this thesis, African populations are a double edged sword. Whilst we identified selective sweeps around known sites under selection, they were smaller in the African populations described here than in populations in southeast Asia. Thus identifying unknown sites after an unknown period of selection in African populations would require a greater coverage of markers than on chromosomes sampled from southeast Asian populations. We would expect that to identify a novel resistance allele in African populations by description of patterns of diversity in the genome of parasites with a resistance phenotype, one would need a high coverage of genetic markers spaced approximately 10kb apart (which would have captured the 40kb sweep around the *dhfr* double mutant allele). If the resistance phenotype was only recently observed approximately 25-30kb apart would be sufficient (and would capture the *dhps* double mutant sweep).

The high recombination rate and weaker level of selection maybe a disadvantage of using African populations, but it also provides the advantage of condensing the region of reduced gene diversity to a smaller region of the chromosome than in other populations, for example South American populations where the entire genome can be in linkage disequilibrium (Cortese et al. 2002). Thus if additional novel mutations are involved they maybe more easily identified in Africa than in other populations.

Another advantage of using African populations is that resistance alleles and their phenotypes are more easily distinguished because their origin in the population is limited, either to a single de novo mutation event or a migration event. This is true for both SP resistance and chloroquine resistance and arguably will hold for future resistance to novel antimalarials. In other global locations, whilst also occurring rarely, there can be more variation in the number of resistance alleles that arise, that

whilst composed of different substitutions would all have a resistance phenotype. For example in southeast Asia the description of the selective sweep on the resistant *dhps* chromosomes is hampered by the presence of at least three origins of *dhps* resistance alleles, each with separate flanking haplotypes that show evidence of having recombined (Anderson and Roper 2005).

IV Final Remarks: Future work and the future of antimalarials

This thesis is not a comprehensive description of the evolution of SP resistance in southeast Africa and there are a number of avenues that could be explored with reward:

1. In West Africa, the frequency of SP resistance alleles are much lower than found in east Africa (Wang et al. 1997a; Eberl et al. 2001; Marks et al. 2005) C. Roper personal communication), as SP selection histories in west African countries are for the majority shorter than those in east Africa. Whereas at *dhfr* the substitutions found in west Africa are the same as in east Africa , at *dhps* the double mutant allele common to southeast Africa is very rare, and other substitutions are more common (Wang et al. 1997a; Eberl et al. 2001; Marks et al. 2005) C. Roper personal communication). Having now described the selective sweeps around the common southeast African SP resistance alleles, it is essential to do the same around the common west African alleles, particularly so at *dhps*. A mapping study identifying the distribution of the *dhfr* triple mutant allele across continental Africa, describing both allele frequencies and flanking haplotypes is currently in

progress (Pota and Roper). This work has yet to be extended to the alleles present at *dhps*.

2. Having made one longitudinal description of the selective sweep on *dhfr* triple mutant chromosomes in southeast Africa, it would be of interest to do an equivalent study in an additional population, particularly one where effective recombination rates are high. This would address the question of the apparent stability in size of the selective sweeps around drug-selected alleles regardless of recombination or selection history. In southern Tanzania the samples used in this thesis were taken prior to the national policy switch to SP in 2001. Samples taken at two further time points would address two questions. Firstly is there any effect on the region of diversity around the resistance alleles in this population when selection is eventually applied? Given that transmission intensity in this population is high and drug use generally lower, it is doubtful that the application of selection would result in an epidemic expansion of resistant parasites that would increase the size of the selective sweep. A second time point a few years after selection in the population had begun, when presumably there will have been an increase in the frequencies of the resistance allele, could be used to address the question of the interaction between recombination and selection as presumably the rate of breakdown of the sweep will have been greater prior to selection occurring, unless gene flow from neighbouring populations had sufficiently stabilised size of the selective sweep.

3. To attempt to address the role of gene flow in the size of the selective sweep on resistant chromosomes, it would be of interest to describe the region of

reduce diversity on chromosomes taken from specific geographical areas, with sufficient power that comparisons could be made between them. If we were able to longitudinally sample populations where selection was ongoing and neighbouring populations where selection was absent, between which a known amount of gene flow was occurring, we would comparatively test the theoretical role of gene flow in the size and shape of the selective sweeps in the populations where selection was absent. As with Clyde's observation of pyrimethamine resistance 100 miles from the initial foci of selection, would we observe that the selective sweep around an allele narrowed more in populations of greatest distance from the population where selection was initially applied? Clearly southeast African populations where SP resistance alleles and use of SP is widespread, would not be the most applicable such an investigation.

4.As a related point to that made above, finer resolution of the parasite population structure could be made if more microsatellites or other markers such as SNPs were used, which are less likely to overestimate population differentiation. Whilst there is unlikely to be much deviation from the conclusion of low population subdivision, it may illuminate more minor differences. With sufficient coverage it maybe possible to detect the introgression of non African drug resistant parasites into African populations shedding light on the route of spread. Clearly to do this on SP resistance would require good historical data. Furthermore, as gene flow is likely to be occurring in all directions all the time, homogeny across the region in neutral allele frequencies could make this difficult to determine.

Bibliography

Abdel-Muhsin, A. M., L. C. Ranford-Cartwright, A. R. Medani, S. Ahmed, S. Suleiman, B. Khan, P. Hunt, D. Walliker, and H. A. Babiker. 2002. Detection of mutations in the *Plasmodium falciparum* dihydrofolate reductase (*dhfr*) gene by dot-blot hybridization. American Journal of Tropical Medicine and Hygiene **67**:24-27.

Alifrangis, M., S. Enosse, I. F. Khalil, D. S. Tarimo, M. M. Lemnge, R. Thompson, I. C. Bygbjerg, and A. M. Ronn. 2003. Prediction of *Plasmodium falciparum* resistance to sulfadoxine/pyrimethamine in vivo by mutations in the dihydrofolate reductase and dihydropteroate synthetase genes: a comparative study between sites of differing endemicity. Am J Trop Med Hyg **69**:601-606.

Alifrangis, M., S. Enosse, R. Pearce, C. Drakeley, C. Roper, I. F. Khalil, W. M. Nkya, A. M. Ronn, T. G. Theander, and I. C. Bygbjerg. 2005. A simple, high-throughput method to detect *Plasmodium falciparum* single nucleotide polymorphisms in the dihydrofolate reductase, dihydropteroate synthase, and P. falciparum chloroquine resistance transporter genes using polymerase chain reaction- and enzyme-linked immunosorbent assay-based technology. Am J Trop Med Hyg **72**:155-162.

Anderson, T. J. 2004. Mapping drug resistance genes in *Plasmodium falciparum* by genome-wide association. Current Drug Targets Infectious Disorders **4**:65-78.

Anderson, T. J., B. Haubold, J. T. Williams, J. G. Estrada-Franco, L. Richardson, R. Mollinedo, M. Bockarie, J. Mokili, S. Mharakurwa, N. French, J. Whitworth, I. D. Velez, A. H. Brockman, F. Nosten, M. U. Ferreira, and K. P. Day. 2000a. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. Molecular Biology and Evolution **17**:1467-1482.

Anderson, T. J., and C. Roper. 2005. The origins and spread of antimalarial drug resistance: Lessons for policy makers. Acta Trop **94**:269-280.

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

Anderson, T. J., X. Z. Su, A. Roddam, and K. P. Day. 2000b. Complex mutations in a high proportion of microsatellite loci from the protozoan parasite *Plasmodium falciparum*. Molecular Ecology **9**:1599-1608.

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

Babiker, H. A., L. C. Ranford-Cartwright, and D. Walliker. 1999. Genetic structure and dynamics of *Plasmodium falciparum* infections in the Kilombero region of Tanzania. Trans R Soc Trop Med Hyg **93 Suppl 1**:11-14.

Barnes, T. M., Y. Kohara, A. Coulson, and S. Hekimi. 1995. Meiotic recombination, noncoding DNA and genomic organization in Caenorhabditis elegans. Genetics **141**:159-179.

Barton, N. H. 2000. Genetic hitchhiking. Philosophical Transactions of the Royal Society of London B Biological Sciences **355**:1553-1562.

Barton, N. H., and M. Slatkin. 1986. A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. Heredity **56 (Pt 3)**:409-415.

Basco, L. K., R. Tahar, A. Keundjian, and P. Ringwald. 2000. Sequence Variations in the Genes Encoding Dihydropteroate Synthase and Dihydrofolate Reductase and Clinical Response to Sulfadoxine-Pyrimethamine in Patients with Acute Uncomplicated Falciparum Malaria. The Journal of Infectious Diseases **182**:624-628.

Basco, L. K., R. Tahar, and P. Ringwald. 1998. Molecular basis of in vivo resistance to sulfadoxine-pyrimethamine in African adult patients infected with *Plasmodium falciparum* malaria parasites. Antimicrobial Agents and Chemotherapy **42**:1811-1814.

Bersaglieri, T., P. C. Sabeti, N. Patterson, T. Vanderploeg, S. F. Schaffner, J. A. Drake, M. Rhodes, D. E. Reich, and J. N. Hirschhorn. 2004. Genetic signatures of strong recent positive selection at the lactase gene. American Journal of Human Genetics **74**:1111-1120.

Bloland, P. B., P. N. Kazembe, A. J. Oloo, B. Himonga, L. M. Barat, and T. K. Ruebush. 1998. Chloroquine in Africa: critical assessment and recommendations for monitoring and evaluating chloroquine therapy efficacy in sub-Saharan Africa. Trop Med Int Health **3**:543-552.

Bloland, P. B., E. M. Lackritz, P. N. Kazembe, J. B. Were, R. Steketee, and C. C. Campbell. 1993. Beyond chloroquine: implications of drug resistance for evaluating malaria therapy efficacy and treatment policy in Africa. J Infect Dis **167**:932-937.

Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd, and L. L. Cavalli-Sforza. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature **368**:455-457.

Brooks, D. R., P. Wang, M. Read, W. M. Watkins, P. F. G. Sims, and J. E. Hyde. 1994. Sequence variation of the hydroxymethyldihydropterin pyrophosphokinase - dihydropteroate synthase gene in lines of the human parasite, *Plasmodium falciparum*, with differing resistances to sulfadoxine. European Journal of Biochemistry **224**:397-405.

Charlwood, J. D., T. Smith, E. Lyimo, A. Y. Kitua, H. Masanja, M. Booth, P. L. Alonso, and M. Tanner. 1998. Incidence of *Plasmodium falciparum* infection in infants in relation to exposure to sporozoite-infected anophelines. American Journal of Tropical Medicine and Hygiene **59**:243-251.

Clyde, D. F. 1967. Malaria in Tanzania. Oxford University Press, Oxford, UK.

Clyde, D. F. 1954. Observations on monthly pryimethamine ('Daraprim') prophylaxis in an East African village. East African Medical Journal **3**:41-46.

Cockerham, C. C., and B. S. Weir. 1984. Covariances of relatives stemming from a population undergoing mixed self and random mating. Biometrics **40**:157-164.

Conway, D. J., C. Roper, A. M. Oduola, D. E. Arnot, P. G. Kremsner, M. P. Grobusch, C. F. Curtis, and B. M. Greenwood. 1999. High recombination rate in natural populations of *Plasmodium falciparum*. Proceedings of the National Academy of Sciences USA **96**:4506-4511.

Cornuet, J. M., and G. Luikart. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics **144**:2001-2014.

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

Cortese, J. F., and C. V. Plowe. 1998. Antifolate resistance due to new and known *Plasmodium falciparum* dihydrofolate reductase mutations expressed in yeast. Molecular and Biochemical Parasitology **94**:205-214.

Cowman, A. F. 1997. The Mechanisms of Drug Action and Resistance in Malaria. Pp. 221-246 *in* J. D. Hayes, and C. Roland Wolf, eds. Molecular Genetics of Drug Resistance. Harwood Academic Publishers, Amsterdam.

Cowman, A. F., M. J. Morry, B. A. Biggs, G. A. M. Cross, and S. J. Foote. 1988. Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of *Plasmodium falciparum*. Proceedings of the National Academy of Sciences USA **85**:9109-9113.

Curtis, J., M. T. Duraisingh, and D. C. Warhurst. 1998. In Vivo Selection for a Specific Genotype of Dihydropteroate Synthetase of *Plasmodium falciparum* by Pyrimethamine-Sulfadoxine but Not Chlorproguanil-Dapsone Treatment. The Journal of Infectious Diseases **177**:1429-1433.

Djimde, A., C. V. Plowe, S. Diop, A. Dicko, T. E. Wellems, and O. Doumbo. 1998. Use of antimalarial drugs in Mali: policy versus reality. Am J Trop Med Hyg **59**:376-379.

Dorsey, G., C. Dokomajilar, M. Kiggundu, S. G. Staedke, M. R. Kamya, and P. J. Rosenthal. 2004. Principal role of dihydropteroate synthase mutations in mediating resistance to sulfadoxine-pyrimethamine in single-drug and combination therapy of uncomplicated malaria in Uganda. Am J Trop Med Hyg **71**:758-763.

Doumbo, O. K., K. Kayentao, A. Djimde, J. F. Cortese, Y. Diourte, A. Konaré, J. G. Kublin, and C. V. Plowe. 2000. Rapid Selection of *Plasmodium falciparum* Dihydrofolate Reductase Mutants by Pryimrthamine Prophylaxis. The Journal of Infectious Diseases **182**:993-996.

Drakeley, C. J., I. Carneiro, H. Reyburn, R. Malima, J. P. Lusingu, J. Cox, T. G. Theander, W. M. Nkya, M. M. Lemnge, and E. M. Riley. 2005. Altitude-Dependent and -Independent Variations in *Plasmodium falciparum* Prevalence in Northeastern Tanzania. J Infect Dis **191**:1589-1598.

Dye, C., and B. G. Williams. 1997. Multigenic drug resistance among inbred malaria parasites. Proceedings of the Royal Society of London B Biological Sciences **264**:61-67.

Eberl, K. J., T. Jelinek, A. O. Aida, G. Peyerl-Hoffmann, C. Heuschkel, A. O. el Valy, and E. M. Christophel. 2001. Prevalence of polymorphisms in the dihydrofolate reductase and dihydropteroate synthetase genes of *Plasmodium falciparum* isolates from southern Mauritania. Tropical Medicine and International Health **6**:756-760.

Edoh, D., H. Mshinda, J. Jenkins, and M. Burger. 1997. Pyrimethamine-resistant *Plasmodium falciparum* parasites among Tanzanian children: a facility-based study using the polymerase chain reaction. American Journal of Tropical Medicine and Hygiene **57**:342-347.

Ellman, R., C. Maxwell, R. Finch, and D. Shayo. 1998. Malaria and anaemia at different altitudes in the Muheza district of Tanzania: childhood morbidity in relation to level of exposure to infection. Ann Trop Med Parasitol **92**:741-753.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164**:1567-1587.

Felsenstein, J. 2004. PHYLIP (Phylogeny Inference Package). Distributed by the author., Department of Genome Sciences, University of Washington, Seattle.

Fohl, L. M., and D. S. Roos. 2003. Fitness effects of DHFR-TS mutations associated with pyrimethamine resistance in apicomplexan parasites. Mol Microbiol **50**:1319-1327.

Gardner, M. J., N. Hall, E. Fung, O. White, M. Berriman, R. W. Hyman, J. M. Carlton, A. Pain, K. E. Nelson, S. Bowman, I. T. Paulsen, K. James, J. A. Eisen, K. Rutherford, S. L. Salzberg, A. Craig, S. Kyes, M. S. Chan, V. Nene, S. J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft, M. W. Mather, A. B. Vaidya, D. M. Martin, A. H. Fairlamb, M. J. Fraunholz, D. S. Roos, S. A. Ralph, G. I. McFadden, L. M. Cummings, G. M. Subramanian, C. Mungall, J. C. Venter, D. J. Carucci, S. L. Hoffman, C. Newbold, R. W. Davis, C. M. Fraser, and B. Barrell. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature **419**:498-511.

Gatton, M. L., L. B. Martin, and Q. Cheng. 2004. Evolution of resistance to sulfadoxine-pyrimethamine in *Plasmodium falciparum*. Antimicrob Agents Chemother **48**:2116-2123.

Goldberg, D. E. 2002. Parasitology. When the host is smarter than the parasite. Science **296**:482-483.

Goldstein, D. B., and D. D. Pollock. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic interference. J Hered **88**:335-342.

Goodman, C., S. P. Kachur, S. Abdulla, E. Mwageni, J. Nyoni, J. A. Schellenberg, A. Mills, and P. Bloland. 2004. Retail supply of malaria-related drugs in rural Tanzania: risks and opportunities. Tropical Medicine and International Health **9**:655-663.

Greenwood, B., and T. Mutabingwa. 2002. Malaria in 2002. Nature **415**:670-672.

Hartl, D. L., and A. G. Clark. 1997. Principles of Population Genetics. Sinauer Associates, Sunderland, MA.

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

Hastings, I. M., W. M. Watkins, and N. J. White. 2002. The evolution of drug-resistant malaria: the role of drug elimination half-life. Philos Trans R Soc Lond B Biol Sci **357**:505-519.

Haubold, B., and R. R. Hudson. 2000. LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage Analysis. Bioinformatics **16**:847-848.

Hay, S. I., D. J. Rogers, J. F. Toomer, and R. W. Snow. 2000. Annual *Plasmodium falciparum* entomological inoculation rates (EIR) across Africa: literature survey, Internet access and review. Trans R Soc Trop Med Hyg **94**:113-127.

Hedrick, P. 1999. Perspective: Highly Variable Loci and their Interpretation in Evolution and Conservation. International Journal of Organic Evolution **53**:313-318.

Hyde, J. E. 1990. The dihydrofolate reductase-thymidylate synthetase gene in the drug resistance of malaria parasites. Pharmacol Ther **48**:45-59.

Jelinek, T., A. H. Kilian, J. Curtis, M. T. Duraisingh, G. Kabagambe, F. von Sonnenburg, and D. C. Warhurst. 1999a. *Plasmodium falciparum*: selection of serine 108 of dihydrofolate reductase during treatment of uncomplicated malaria with co-trimoxazole in Ugandan children. American Journal of Tropical Medicine and Hygiene **61**:125-130.

Jelinek, T., A. H. Kilian, G. Kabagambe, and F. von Sonnenburg. 1999b. *Plasmodium falciparum* resistance to sulfadoxine/pyrimethamine in Uganda: correlation with polymorphisms in the dihydrofolate reductase and

dihydropteroate synthetase genes. American Journal of Tropical Medicine and Hygiene **61**:463-466.

Jelinek, T., A. M. Ronn, J. Curtis, M. T. Duraisingh, M. M. Lemnge, J. Mhina, I. C. Bygbjerg, and D. C. Warhurst. 1997. High prevalence of mutations in the dihydrofolate reductase gene of *Plasmodium falciparum* in isolates from Tanzania without evidence of an association to clinical sulfadoxine/pyrimethamine resistance. Tropical Medicine and International Health **2**:1075-1079.

Jelinek, T., A. M. Ronn, M. M. Lemnge, J. Curtis, J. Mhina, M. T. Duraisingh, I. C. Bygbjerg, and D. C. Warhurst. 1998. Polymorphisms in the dihydrofolate reductase (DHFR) and dihydropteroate synthetase (DHPS) genes of *Plasmodium falciparum* and in vivo resistance to sulphadoxine/pyrimethamine in isolates from Tanzania. Tropical Medicine and International Health **3**:605-609.

Kaplan, N. L., R. R. Hudson, and C. H. Langley. 1989. The "hitchhiking effect" revisited. Genetics **123**:887-899.

Khan, B., S. Omar, J. N. Kanyara, M. Warren-Perry, J. Nyalwidhe, D. S. Peterson, T. Wellems, S. Kaniaru, J. Gitonga, F. J. Mulaa, and D. K. Koech. 1997. Antifolate drug resistance and point mutations in *Plasmodium falciparum* in Kenya. Transactions of the Royal Society of Tropical Medicine and Hygiene **91**:456-460.

Kim, Y., and W. Stephan. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics **160**:765-777.

Kleinschmidt, I., and B. Sharp. 2001. Patterns in age-specific malaria incidence in a population exposed to low levels of malaria transmission intensity. Tropical Medicine and International Health **6**:986-991.

Kruglyak, S., R. T. Durrett, M. D. Schug, and C. F. Aquadro. 1998. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. PNAS **95**:10774-10778.

Kublin, J. G., J. F. Cortese, E. M. Njunju, R. A. Mukadam, J. J. Wirima, P. N. Kazembe, A. A. Djimde, B. Kouriba, T. E. Taylor, and C. V. Plowe. 2003. Reemergence of chloroquine-sensitive *Plasmodium falciparum* malaria after cessation of chloroquine use in Malawi. J Infect Dis **187**:1870-1875.

Kublin, J. G., F. K. Dzinjalamala, D. D. Kamwendo, E. M. Malkin, J. F. Cortese, L. M. Martino, R. A. M. Mukadam, S. J. Rogerson, A. G. Lescano, M. E. Molyneux, P. A. Winstanley, P. Chimpeni, T. E. Taylor, and C. V. Plowe. 2002. Molecular Markers for Failure of Sulfadoxine-Pyrimethamine and Chlorproguanil-Dapsone Treatment of *Plasmodium falciparum* Malaria. The Journal of Infectious Diseases **185**:380-388.

Kun, J. F., L. G. Lehman, B. Lell, R. Schmidt-Ott, and P. G. Kremsner. 1999. Low-dose treatment with sulfadoxine-pyrimethamine combinations selects for drug-resistant *Plasmodium falciparum* strains. Antimicrobial Agents and Chemotherapy **43**:2205-2208.

Lichten, M., and A. S. Goldman. 1995. Meiotic recombination hotspots. Annual Review of Genetics **29**:423-444.

Liu, K., and S. Muse. 2004. PowerMarker: new genetic data analysis software.

Mackinnon, M. J. 1997. Survival probability of drug resistant mutants in malaria parasites. Proc Biol Sci **264**:53-59.

Maisnier-Patin, S., and D. I. Andersson. 2004. Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. Res Microbiol **155**:360-369.

Marks, F., J. Evans, C. G. Meyer, E. N. Browne, C. Flessner, V. von Kalckreuth, T. A. Eggelte, R. D. Horstmann, and J. May. 2005. High prevalence of markers for sulfadoxine and pyrimethamine resistance in *Plasmodium falciparum* in the absence of drug pressure in the Ashanti region of Ghana. Antimicrob Agents Chemother **49**:1101-1105.

Mberu, E. K., M. K. Mosobo, A. M. Nzila, G. O. Kokwaro, C. H. Sibley, and W. M. Watkins. 2000. The changing in vitro susceptibility pattern to pyrimethamine/sulfadoxine in *Plasmodium falciparum* field isolates from Kilifi, Kenya. American Journal of Tropical Medicine and Hygiene **62**:396-401.

McCombie, S. C. 1996. Treatment seeking for malaria: a review of recent research. Soc Sci Med **43**:933-945.

Miller, L. H., R. J. Howard, R. Carter, M. F. Good, V. Nussenzweig, and R. S. Nussenzweig. 1986. Research toward malaria vaccines. Science **234**:1349-1356.

Minch, E., A. Ruiz-Linares, D. Goldstein, M. Feldman, and L. L. Cavalli-Sforza. 1997. Microsat v.1.5d: A computer program for calculating various statistics on microsatellite allele data.

Mutabingwa, T., A. Nzila, E. Mberu, E. Nduati, P. Winstanley, E. Hills, and W. Watkins. 2001. Chlorproguanil-dapsone for treatment of drug-resistant falciparum malaria in Tanzania. The Lancet **358**:1218-1223.

Nair, S., J. T. Williams, A. Brockman, L. Paiphun, M. Mayxay, P. N. Newton, J. P. Guthmann, F. M. Smithuis, T. T. Hien, N. J. White, F. Nosten, and T. J. Anderson. 2003. A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. Molecular Biology and Evolution **20**:1526-1536.

Nash, D., S. Nair, M. Mayxay, P. N. Newton, J. P. Guthmann, F. Nosten, and T. J. Anderson. 2005. Selection strength and hitchhiking around two anti-malarial resistance genes. Proc Biol Sci **272**:1153-1161.

Nauta, M. J., and F. J. Weissing. 1996. Constraints on allele size at microsatellite loci: implications for genetic differentiation. Genetics **143**:1021-1032.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Nei, M., and A. K. Roychoudhury. 1974. Sampling variances of heterozygosity and genetic distance. Genetics **76**:379-390.

Nosten, F., M. van Vugt, R. Price, C. Luxemburger, K. L. Thway, A. Brockman, R. McGready, F. ter Kuile, S. Looareesuwan, and N. J. White. 2000. Effects of artesunate-mefloquine combination on incidence of *Plasmodium falciparum* malaria and mefloquine resistance in western Thailand: a prospective study. The Lancet **356**:297-302.

Nzila, A., E. Mberu, P. Bray, G. Kokwaro, P. Winstanley, K. Marsh, and S. Ward. 2003. Chemosensitization of *Plasmodium falciparum* by probenecid in vitro. Antimicrob Agents Chemother **47**:2108-2112.

Nzila, A. M., E. K. Mberu, J. Sulo, H. Dayo, P. A. Winstanley, C. H. Sibley, and W. M. Watkins. 2000a. Towards an Understanding of the Mechanism of Pyrimethamine -Sulfadoxine Resistance in *Plasmodium falciparum*: Genotyping of Dihydrofolate Reductase and Dihydropteroate Synthase of Kenyan Parasites. Antimicrobial Agents and Chemotherapy **44**:991-996.

Nzila, A. M., E. Nduati, E. K. Mberu, C. H. Sibley, S. A. Monks, P. A. Winstanley, and W. M. Watkins. 2000b. 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*. The Journal of Infectious Diseases **181**:2023-2028.

Omar, S. A., I. S. Adagu, and D. C. Warhurst. 2001. Can pretreatment screening for *dhps* and *dhfr* point mutations in *Plasmodium falciparum* infections be used to predict sulfadoxine- pyrimethamine treatment failure? Transactions of the Royal Society of Tropical Medicine and Hygiene **95**:315-319.

Paetkau, D., L. P. Waits, P. L. Clarkson, L. Craighead, and C. Strobeck. 1997. An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. Genetics **147**:1943-1957.

Page, R. D. 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci **12**:357-358.

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

Palaisa, K., M. Morgante, S. Tingey, and A. Rafalski. 2004. Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. Proceedings of the National Academy of Sciences USA **101**:9885-9890.

Paul, R. E., M. J. Packer, M. Walmsley, M. Lagog, L. C. Ranford-Cartwright, R. Paru, and K. P. Day. 1995. Mating patterns in malaria parasite populations of Papua New Guinea. Science **269**:1709-1711.

Payne, D. 1989. The History and Development of the WHO Standard *in vivo* and *in vitro* Test Systems for the Sensitivity of *Plasmodium falciparum* and other Human *Plasmodia* to Antimalarial Drugs. University of London (London School of Hygiene and Tropical Medicine), London.

Pearce, R. J., C. Drakeley, D. Chandramohan, F. Mosha, and C. Roper. 2003. Molecular determination of point mutation haplotypes in the dihydrofolate reductase and dihydropteroate synthase of *Plasmodium falciparum* in three districts of northern Tanzania. Antimicrobial Agents and Chemotherapy **47**:1347-1354.

Peters, W. 1970. Drug Resistance in the Human Malarias: 1. Drugs Influencing Folic Acid Metabolism *in* W. Peters, ed. Chemotherapy and Drug Resistance in Malaria. Academic Press, London and New York.

Peterson, D. S., D. Walliker, and T. E. Wellems. 1988. Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in falciparum malaria. Proceedings of the National Academy of Sciences USA **85**:9114-9128.

Plowe, C. V., J. F. Cortese, A. Djimde, O. C. Nwanyanwu, W. M. Watkins, P. A. Winstanley, J. G. Estrada-Franco, R. E. Mollinedo, J. C. Avila, J. L. Cespedes, D. Carter, and O. K. Doumbo. 1997. Mutations in *Plasmodium falciparum* dihydrofolate reductase and dihydropteroate synthase and epidemiologic patterns of pyrimethamine- sulfadoxine use and resistance. Journal of Infectious Diseases **176**:1590-1596.

Polley, S. D., and D. J. Conway. 2001. Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. Genetics **158**:1505-1512.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945-959.

Quesada, H., U. E. Ramirez, J. Rozas, and M. Aguade. 2003. Large-scale adaptive hitchhiking upon high recombination in Drosophila simulans. Genetics **165**:895-900.

Ranford-Cartwright, L. C., K. L. Johnston, A. M. Abdel-Muhsin, B. K. Khan, and H. A. Babiker. 2002. Critical comparison of molecular genotyping

methods for detection of drug-resistant *Plasmodium falciparum*. Transactions of the Royal Society of Tropical Medicine and Hygiene **96**:568-572.

Raymond, M., and F. Rousset. 1995. GENEPOP(version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity **86**:248-249.

Ronn, A. M., H. A. Msangeni, J. Mhina, W. H. Wernsdorfer, and I. C. Bygbjerg. 1996. High level of resistance of *Plasmodium falciparum* to sulfadoxine-pyrimethamine in children in Tanzania. Transactions of the Royal Society of Tropical Medicine and Hygiene **90**:179-181.

Roper, C., R. Pearce, B. Bredenkamp, J. Gumede, C. Drakeley, F. Mosha, D. Chandramohan, and B. Sharp. 2003. Antifolate antimalarial resistance in southeast Africa: a population-based analysis. The Lancet **361**:1174-1181.

Roper, C., R. Pearce, S. Nair, B. Sharp, F. Nosten, and T. Anderson. 2004. Intercontinental spread of pyrimethamine-resistant malaria. Science **305**:1124.

Sachs, J., and P. Malaney. 2002. The economic and social burden of malaria. Nature **415**:680-685.

Santiago, E., and A. Caballero. 2005. Variation after a selective sweep in a subdivided population. Genetics **169**:475-483.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin ver. 2.000: A software for population genetics data analysis., Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Schrag, S. J., V. Perrot, and B. R. Levin. 1997. Adaptation to the fitness costs of antibiotic resistance in Escherichia coli. Proc Biol Sci **264**:1287-1291.

Sibley, C. H., J. E. Hyde, P. F. G. Sims, C. V. Plowe, J. G. Kublin, E. K. Mberu, A. F. Cowman, P. A. Winstanley, W. A. Watkins, and A. M. Nzila. 2001. Pyrimethamine-sulfadoxine resistance in *Plasmodium falciparum*: what next? Trends in Parasitology **17**:582-588.

Sidhu, A. B. S., D. Verdier-Pinard, and D. A. Fidock. 2002. Chloroquine Resistance in *Plasmodium falciparum* Malaria Parasites Conferred by pfcrt Mutations. Science **298**:210-213.

Sirawaraporn, W., T. Sathitkul, R. Sirawaraporn, Y. Yuthavong, and D. V. Santi. 1997. Antifolate-resistant mutants of *Plasmodium falciparum* dihydrofolate reductase. Proceedings of the National Academy of Sciences USA **94**:1124-1129.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics **139**:457-462.

Smith, J. M., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genetic Research **23**:23-35.

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

Snow, R. W., C. A. Guerra, A. M. Noor, H. Y. Myint, and S. I. Hay. 2005. The global distribution of clinical episodes of *Plasmodium falciparum* malaria. Nature **434**:214-217.

Snow, R. W., J. A. Omumbo, B. Lowe, C. S. Molyneux, J. O. Obiero, A. Palmer, M. W. Weber, M. Pinder, B. Nahlen, C. Obonyo, C. Newbold, S. Gupta, and K. Marsh. 1997. Relation between severe malaria morbidity in children and level of *Plasmodium falciparum* transmission in Africa. The Lancet **349**:1650-1654.

Sowunmi, A., F. A. Fehintola, A. A. Adedeji, G. O. Gbotosho, C. O. Falade, E. Tambo, B. A. Fateye, T. C. Happi, and A. M. Oduola. 2004. Open

randomized study of pyrimethamine-sulphadoxine vs. pyrimethamine-sulphadoxine plus probenecid for the treatment of uncomplicated *Plasmodium falciparum* malaria in children. Trop Med Int Health **9**:606-614.

Su, X., M. T. Ferdig, Y. Huang, C. Q. Huynh, A. Liu, J. You, J. C. Wootton, and T. E. Wellems. 1999. A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. Science **286**:1351-1353.

Talisuna, A. O., P. Bloland, and U. D'Alessandro. 2004. History, dynamics, and public health importance of malaria parasite resistance. Clin Microbiol Rev **17**:235-254.

Taylor, L. H., and A. F. Read. 1997. Why so few transmission stages? Reproductive restraint by malaria parasites. Parasitol Today **13**:135-140.

Taylor, L. H., and A. F. Read. 1998. Determinants of transmission success of individual clones from mixed-clone infections of the rodent malaria parasite, Plasmodium chabaudi. Int J Parasitol **28**:719-725.

Taylor, L. H., D. Walliker, and A. F. Read. 1997. Mixed-genotype infections of malaria parasites: within-host dynamics and transmission success of competing clones. Proc Biol Sci **264**:927-935.

The R Core Development Team. 2005. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing.

Trigg, J. K., H. Mbwana, O. Chambo, E. Hills, W. Watkins, and C. F. Curtis. 1997. Resistance to pyrimethamine/sulfadoxine in *Plasmodium falciparum* in 12 villages in north east Tanzania and a test of chlorproguanil/dapsone. Acta Tropica **63**:185-189.

Triglia, T., and A. F. Cowman. 1994. Primary structure and expression of the dihydropteroate synthetase gene of *Plasmodium falciparum*. Proceedings of the National Academy of Sciences USA **91**:7149-7153.

Triglia, T., J. G. Menting, C. Wilson, and A. F. Cowman. 1997. Mutations in dihydropteroate synthase are responsible for sulfone and sulfonamide resistance in *Plasmodium falciparum*. Proceedings of the National Academy of Sciences USA **94**:13944-13949.

Triglia, T., P. Wang, P. F. G. Sims, J. E. Hyde, and A. F. Cowman. 1998. Allelic exchange at the endogenous genomic locus in *Plasmodium falciparum* proves the role of dihydropteroate synthase in sulfadoxine-resistant malaria. The EMBO Journal **17**:3807-3815.

von Seidlein, L., S. Clarke, N. Alexander, F. Manneh, T. Doherty, M. Pinder, G. Walraven, and B. Greenwood. 2002. Treatment uptake by individuals infected with *Plasmodium falciparum* in rural Gambia, West Africa. Bull World Health Organ **80**:790-796.

Wang, P., R. K. Brobey, T. Horii, P. F. Sims, and J. E. Hyde. 1999. Utilization of exogenous folate in the human malaria parasite *Plasmodium falciparum* and its critical role in antifolate drug synergy. Mol Microbiol **32**:1254-1262.

Wang, P., C.-S. Lee, R. Bayoumi, A. Djimde, O. M. Doumbo, G. Swedberg, L. D. Dao, H. Mshinda, M. Tanner, W. M. Watkins, P. F. G. Sims, and J. E. Hyde. 1997a. Resistance to antifolates in *Plasmodium falciparum* monitored by sequence analysis of dihydrofolate synthetase and dihydrofolate reductase alleles in a large number of field samples of diverse origins. Molecular and Biochemical Parasitology **89**:161-177.

Wang, P., N. Nirmalan, Q. Wang, P. F. Sims, and J. E. Hyde. 2004a. Genetic and metabolic analysis of folate salvage in the human malaria parasite *Plasmodium falciparum*. Mol Biochem Parasitol **135**:77-87.

Wang, P., M. Read, P. F. G. Sims, and J. E. Hyde. 1997b. Sulfadoxine resistance in the human malaria parasite *Plasmodium falciparum* is determined by

mutations in dihydropteroate synthetase and an additional factor associated with folate utilization. Molecular Microbiology **23**:979--986.

Wang, P., Q. Wang, T. V. Aspinall, P. F. Sims, and J. E. Hyde. 2004b. Transfection studies to explore essential folate metabolism and antifolate drug synergy in the human malaria parasite *Plasmodium falciparum*. Mol Microbiol **51**:1425-1438.

Warhurst, D. C. 2002. Resistance to antifolates in *Plasmodium falciparum*, the causative agent of tropical malaria. Science Progress **85**:89-111.

Watkins, W. M., E. K. Mberu, P. A. Winstanley, and C. V. Plowe. 1997. The Efficacy of Antifolate Antimalarial Combinations in Africa: A Predictive Model Based on Pharmacodynamic and Pharmacokinetic Analysis. Parasitology Today **13**:459-464.

Watkins, W. M., E. K. Mberu, P. A. Winstanley, and C. V. Plowe. 1999. More on 'The Efficacy of Antifolate Antimalarial Combinations in Africa'. Parasitology Today **15**:131-132.

White, N. J. 1992. Antimalarial drug resistance: the pace quickens. Journal of Antimicrobial Chemotherapy **30**:571-585.

Wiehe, T. 1998. The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: hitchhiking of microsatellites. Theorectical Population Biology **53**:272-283.

Wootton, J. C., X. Feng, M. T. Ferdig, R. A. Cooper, J. Mu, D. I. Baruch, A. J. Magill, and X. Z. Su. 2002. Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. Nature **418**:320-323.

Wu, Y., L. A. Kirkman, and T. E. Wellems. 1996. Transformation of *Plasmodium falciparum* malaria parasites by homologous integration of plasmids that confer resistance to pyrimethamine. Proceedings of the National Academy of Sciences USA **93**:1130-1134.

Zhang, K., and P. K. Rathod. 2002. Divergent regulation of dihydrofolate reductase between malaria parasite and human host. Science **296**:545-547.

Zolg, J. W., J. R. Plitt, G. X. Chen, and S. Palmer. 1989. Point mutations in the dihydrofolate reductase-thymidylate synthase gene as the molecular basis for pyrimethamine resistance in *Plasmodium falciparum*. Molecular and Biochemical Parasitology **36**:253-262.

Appendix 1: Frequencies of the *dhfr* and *dhps* two locus genotypes (Chapter 3)

Locus	CNCS SAKAA	CNCS AAKAA	CNCS SGEAA	CICN SAKAA	CNRN SAKAA	CNRN AAKAA	CNRN SGEAA	CIRN SAKAA	CIRN AAKAA	CIRN SGEAA	Other <5%	N=
MatolaRio	0.09	0.00	0.00	0.00	0.16	0.00	0.02	0.42	0.04	0.20	0.07	45
Beluluane	0.06	0.00	0.00	0.00	0.09	0.00	0.03	0.46	0.03	0.34	0.00	35
BelaVista	0.13	0.00	0.03	0.00	0.10	0.00	0.07	0.50	0.00	0.17	0.00	30
Catatuane	0.19	0.03	0.00	0.08	0.17	0.03	0.03	0.31	0.00	0.11	0.06	36
Salamanga	0.09	0.00	0.06	0.06	0.25	0.06	0.09	0.28	0.00	0.09	0.00	32
Steenbok	0.18	0.00	0.00	0.00	0.10	0.00	0.02	0.40	0.03	0.25	0.02	60
Komatipoort	0.20	0.00	0.00	0.00	0.15	0.00	0.00	0.50	0.00	0.10	0.05	40
Mangweni	0.29	0.05	0.02	0.05	0.14	0.00	0.00	0.34	0.02	0.05	0.05	59
Morogoro	0.23	0.13	0.02	0.06	0.10	0.04	0.03	0.20	0.08	0.01	0.08	98
Rufigi	0.26	0.06	0.02	0.10	0.08	0.03	0.01	0.27	0.04	0.04	0.05	285
Kilo/Ulanga	0.36	0.10	0.04	0.06	0.05	0.01	0.02	0.18	0.08	0.05	0.06	190

Appendix 2: Allele frequencies (%) at the eight unlinked microsatellites (Chapter 3)

ARAII	Komatipoort	Mangweni	Steenbok	E. Mozambique	W. Mozambique	Ulanga	Morogoro	Rufigi
29							22.07	1.92
39	2.22	6.38	1.23		36.96			
42	2.22	2.13				1.96	4.14	0.64
50	17.78	8.51		34.04				
55	2.22			8.51	2.17			
57								0.64
60	2.22	2.13	2.47		2.17	1.96	2.76	7.69
63	2.22		9.88	2.13	4.35	10.78	2.76	8.33
66	8.89	12.77	14.81	10.64	13.04	35.29	26.90	28.21
69	20.00	31.91	19.75	12.77	4.35	17.65	17.24	30.13
72	11.11	8.51	16.05	12.77	8.70	12.75	7.59	7.05
75	11.11	14.89	16.05	8.51	17.39	9.80	7.59	4.49
78	11.11	10.64	3.70	4.26	4.35	3.92	4.83	7.05
81	6.67		1.23	2.13	4.35	4.90	2.07	1.92
84			4.94	4.26	2.17		0.69	1.28
87						0.98	1.38	0.64
93			1.23					
105	2.22	2.13	8.64					
	n=45	n=47	n=81	n=47	n=46	n=102	n=145	n=156

G377	Komatipoort	Mangweni	Steenbok	E. Mozambique	W. Mozambique	Ulanga	Morogoro	Rufigi
90								1.38
93	4.17	10.42	3.45	9.09	11.36	6.14	3.47	2.07
97	20.83	25.00	25.29	18.18	9.09	21.05	16.67	19.31
100	62.50	54.17	50.57	61.36	63.64	64.04	72.22	65.52
103	12.50	8.33	16.09	9.09	6.82	7.02	4.86	7.59
106		2.08	4.60	2.27	2.27	1.75	2.78	3.45
109					4.55			0.69
112					2.27			
	n=48	n=48	n=87	n=44	n=44	n=114	n=144	n=145

PolyA	Komatipoort	Mangweni	Steenbok	E. Mozambique	W. Mozambique	Ulanga	Morogoro	Rufigi
124						0.88		
130		4.55						
133			2.38		2.78	0.88		3.15
136		6.82	7.14	2.86	5.56		0.73	
139		2.27	5.95	11.43	5.56	1.77	0.73	2.36
143	4.65		3.57		5.56	5.31	7.30	2.36
146	4.65		2.38	2.86	2.78	8.85	2.92	5.51
150	11.63		2.38		8.33	11.50	10.95	7.09
153	18.60	18.18	17.86	17.14	13.89	20.35	16.79	27.56
156	11.63	15.91	9.52	22.86	5.56	14.16	16.79	8.66
160	6.98	15.91	9.52	14.29	2.78	5.31	10.95	9.45
163	4.65	4.55	3.57	2.86		4.42	4.38	8.66
166	11.63	4.55	9.52	8.57	11.11	9.73	8.03	9.45
169	4.65	2.27	3.57	5.71	2.78	3.54	4.38	3.94
172	6.98	2.27	9.52		8.33	5.31	3.65	2.36
175	6.98	4.55	9.52	5.71	13.89	4.42	5.84	3.15
178	4.65	11.36	3.57		8.33	1.77	3.65	2.36
181		2.27				0.88	0.73	2.36
184	2.33	4.55		2.86	2.78		0.73	1.57
187						0.88	0.73	
193				2.86			0.73	
	n=43	n=44	n=84	n=35	n=36	n=113	n=137	n=127

PfPk2	Komatipoort	Mangweni	Steenbok	E. Mozambique	W. Mozambique	Ulanga	Morogoro	Rufigi
140								0.62
144						1.03		1.86
157		2.13			2.33	1.03		0.62
160	14.89	4.26	7.78	4.65	9.30	6.19	2.82	4.97
163	27.66	19.15	22.22	32.56	30.23	23.71	30.99	28.57
166	8.51	12.77	8.89	13.95	11.63	7.22	7.04	7.45
169	8.51	14.89	18.89	16.28	9.30	8.25	12.68	12.42
172	4.26	2.13	5.56	9.30	4.65	18.56	13.38	15.53
175	4.26	10.64	8.89	4.65	11.63	11.34	9.86	6.21
178	4.26	2.13	11.11	2.33	2.33		3.52	3.11
181	10.64	6.38	3.33	2.33		5.15	4.23	5.59
184	6.38	4.26	1.11	4.65	2.33	1.03	2.11	1.86
187	2.13	6.38	7.78	2.33	2.33	3.09	7.04	4.97
190	4.26	10.64		4.65	6.98	8.25	2.11	4.35
193	4.26	2.13	4.44		6.98	4.12	2.11	
196		2.13		2.33			2.11	1.86
210						1.03		
	n=47	n=47	n=90	n=43	n=43	n=97	n=142	n=161

TA109	Komatipoort	Mangweni	Steenbok	E. Mozambique	W. Mozambique	Ulanga	Morogoro	Rufigi
152			1.10			0.81		1.35
155							0.67	
158		2.08	1.10				2.68	0.68
161	27.08	25.00	18.68	14.89	14.58	30.08	25.50	27.03
164	25.00	20.83	23.08	29.79	25.00	13.01	19.46	28.38
167				4.26			2.01	2.03
170		2.08	4.40			0.81	0.67	
173	10.42	12.50	20.88	10.64	6.25	11.38	12.75	9.46
176	22.92	10.42	14.29	17.02	22.92	17.07	30.20	18.92
179		4.17	1.10	2.13	2.08	2.44	0.67	2.03
182	2.08	2.08	3.30	2.13	4.17			
185	2.08		2.20			7.32	1.34	2.03
188		4.17	1.10	2.13	2.08	1.63	2.01	1.35
191			1.10			0.81	0.67	0.68
194						0.81		
197	6.25	12.50	5.49	14.89	22.92	13.82	1.34	4.73
200				2.13				1.35
203	2.08	2.08						
209	2.08	2.08	2.20					
	n=48	n=48	n=91	n=47	n=48	n=123	n=149	n=148

TA42	Komatipoort	Mangweni	Steenbok	E. Mozambique	W. Mozambique	Ulanga	Morogoro	Rufigi
180					2.94			
183					2.94			
186	73.91	78.05	59.77	73.81	61.76	69.23	77.87	83.65
189	10.87	9.76	21.84	7.14	11.76	4.81	7.38	3.77
192						0.96		
198				2.38				
201	2.17	4.88		4.76	8.82	12.50	5.74	5.66
204						0.96		
215	2.17		1.15				0.82	
236					2.94			
239		2.44						
243	10.87	4.88	17.24	11.90	8.82	8.65	6.56	5.66
246						0.96	0.82	1.26
249						1.92	0.82	
	n=46	n=41	n=87	n=42	n=34	n=104	n=122	n=159

TA87	Komatipoort	Mangweni	Steenbok	E. Mozambique	W. Mozambique	Ulanga	Morogoro	Rufigi
73				2.17	2.27	2.42		
80	2.22	2.08						
84								1.22
87	2.22	8.33	5.56	2.17	2.27	0.81	1.96	0.61
90	2.22	2.08		2.17		1.61	0.65	2.44
93	4.44	4.17	8.89	4.35	2.27	3.23	0.65	3.66
96	15.56	18.75	12.22	26.09	31.82	20.16	15.69	24.39
99	20.00	20.83	20.00	4.35	18.18	13.71	20.92	15.24
102	22.22	18.75	20.00	15.22	9.09	16.13	21.57	15.24
105	13.33	10.42	10.00	13.04	15.91	15.32	14.38	20.73
108	13.33	6.25	16.67	17.39	6.82	12.90	13.73	9.15
111	2.22	2.08	3.33	4.35	4.55	5.65	5.23	1.83
114		6.25	2.22			2.42	2.61	1.83
117			1.11	8.70	4.55	4.03	2.61	3.66
120	2.22					0.81		
123						0.81		
131					2.27			
	n=45	n=48	n=90	n=46	n=44	n=124	n=153	n=164

TA102	Komatipoort	Mangweni	Steenbok	E. Mozambique	W. Mozambique	Ulanga	Morogoro	Rufigi
111	2.22						0.69	
114								0.77
117	2.22	8.70	14.29	6.67	9.76	3.45	3.47	3.08
120	15.56	17.39	10.71	13.33	17.07	18.10	18.75	18.46
123	11.11	10.87	13.10	6.67	9.76	15.52	19.44	9.23
126	15.56	17.39	22.62	10.00	14.63	22.41	14.58	17.69
129	33.33	21.74	19.05	40.00	31.71	26.72	28.47	27.69
132	6.67	8.70	14.29	10.00	4.88	2.59	7.64	11.54
135	6.67	4.35	4.76	6.67		6.03	5.56	8.46
138	4.44	6.52	1.19	6.67	9.76	2.59	1.39	0.77
141	2.22	4.35			2.44	0.86		2.31
144						1.72		
	n=45	n=46	n=84	n=30	n=41	n=116	n=144	n=130

Appendix 3: All chromosome 4 mapping microsatellites described in this study (Chapters 4 & 5)

Primer Name	Sequence	Repeat Unit	Name	Distance from codon 108 of dhfr	Size Range
dhfr350kbU3F	CACATACTATGATCTACCC	AAT	U350	-349402	288-303
dhfr350kbUF-lab	CAGAAGATGTGCAACAAG				
dhfr350kbUR	CTTCGCAATTGAGTTGGT				
dhfr250kbU3F	GAGCAAATCCGTATGATCG	AAT	U250	-249083	261-287
dhfr250kbUF-lab	ATGGACGCATTTAACCCGA				
dhfr250kbUR	TCATAGGAGCGACTAGGTT				
dhfr150kbU3F	ATGCATACCTGTGAAGAA	AAAAT	U150	-143768	162-210
dhfr150kbUF-lab	TGCCCTTTTAATCAGAAC				
dhfr150kbUR	CTACAGTTTTGTGAGTTAT				
dhfr125kbU3F	TAAATTTCCTTCACATAGGCC	AT	U125	-125163	158-198
dhfr125kbUF-lab	CGTTGCAAATATACCATCTTG				
dhfr125kbUR	GGATGAATCCAAGTAACCAAT				
dhfr100kbU3F	AACTTGTATCACACACCCAA	AT	U100	-100618	309-338
dhfr100kbUF-lab	CCTGAGCATAAGATTGCAAAG				
dhfr100kbUR	AACTGAATCAGAAGTCATATTGG				
dhfr90kbUF	GTAATAGCAAATTAGAGGTGtt	AT	U90	-89100	n/a
dhfr90kbU3R	GAACACCAGATGGCCCTC				
dhfr90kbUR-lab	CAAATGGTATGAACGTGATTC				
dhfr80kbUF	AAGAGGTTGCACAAGATC	AT	U80	-80048	258-277
dhfr80kbU3R	CATATCGTATGTATGCCAC				
dhfr80kbUR-lab	GAGATGCTCTTACAATCG				
dhfr75kbU3F	CGAGGTGAATTATGGGAA	AT	U75	-76185	247-273
dhfr75kbUF-lab	CATACGGCAAAATGAGATa				
dhfr75kbUR	TATGTGCGAACGGTAAAA				
dhfr70DDkbUR	ACCTTTAAATACGACTAGCG	AT	U70DD	-70352	480-521
dhfr70DDkbU3F	GGCTGTTGTCGCTACTCG				
dhfr70DDkbUF-lab	GTATGATCACCTAAATCGAC				
dhfr70kbUF	TACACACATATGTGCACATATAA	AAAG	U70	-69219	301-395
dhfr70kbU3R	CCTTATTCTTATGTCCCGCG				
dhfr70kbUR-lab	CTTACGTTCTTACTTTCTTACG				
dhfr60kbUF	CGGTCTATATAGGAAATTTATC	AT	U60	-60113	193-242
dhfr60kbU3R	CGGTTCTGTGTAAACCCATG				
dhfr60kbUR-lab	CCCCAGCCTTAGACATATGC				

Primer Name	Sequence	Repeat Unit	Name	Distance from codon 108 of dhfr	Size Range
dhfr50kbU3F	CTTCCTTTGTATATCATATC	AT	U50	-51037	212-219
dhfr50kbUF-lab	TGATATTCTTATATGTTCATC				
dhfr50kbUR	GAACATTTCATATAGTTACAT				
dhfrii40kbU3F	GTTGCCTTGGCGTATAAG	AT	U40	-39464	200-242
dhfrii40kbUF-lab	aaGCAACATATGCAAGTTAAt				
dhfrii40kbUR	TTTATCAAAGGGACACTCC				
dhfr30kbU3F	GTGACAGTGAACAGGTAATA	AT	U30	-29895	133-173
dhfr30kbUF-lab	GAACGTACATAGGTTTATAC				
dhfr30kbUR	CTTCATAATTGTCGCACGTA				
dhfr20kbU3F	GCTTCATATCTTATATCATT	AT	U20	-20182	197-222
dhfr20kbUF-lab	CAAATTAACACTTCATAATC				
dhfr20kbUR	tTTCTAAAATGCTTATCCc				
dhfrii10kbU3F	CAGAAGGGTTATAATAAGAT	ATTT	U10	-7415	245-265
dhfrii10kbUF-lab	CGACATACATACATTTATCA				
dhfrii10kbUR	TACAAATGAAGGTCGATTTT				
dhfr					
dhfr10kbD3F	AGTAAAGAAACAGAAACAAGTACA	AT	D10	11450	130-153
dhfr10kbDF-lab	ATCAGGAGGAATAAGAAATGAAC				
dhfr10kbDR	GTTAACATACTATCCATATGAATG				
dhfr20kbD3F	GATGAAAATTTGCTTACTTAC	ATG	D20	19605	280-296
dhfr20kbDF-lab	TGTGCACATGAAATTTGTTC				
dhfr20kbDR	TGGTTCAATAAACGAGACC				
dhfr30kbD3F	TCATCATCTGGATCGATA	AT	D30	30199	126-166
dhfr30kbDF-lab	ACAAATTGTGTGTATCAC				
dhfr30kbDR	ATGCTGTACACAAATTGG				
dhfr40kbD3F	ACGATAAGGATGATACTC	AT	D40	40874	297-336
dhfr40kbDF-lab	TTGTGTTCATCTTCATCC				
dhfr40kbDR	TAACAAGCGAAAGAGGTA				
dhfr50kbD3F	AATGTACTACAGTAAAACAGA	AT	D50	50176	142-175
dhfr50kbDF-lab	AGCATATATGATGTAATAAGG				
dhfr50kbDR	AATACAAGTGGAAAGTTACC				
dhfr60kbDF	GTGCATGTTGCACAGAAATG	TAA	D60	60946	242-272
dhfr60kbD3R	GTCCTTCCCATTACAAGTTTC				
dhfr60kbDR-lab	GGTAAATAGGCCAACAACTC				

Primer Name	Sequence	Repeat Unit	Name	Distance from codon 108 of dhfr	Size Range
dhfr70kbDF	GTTATCACCAAAATGAGTTCG	TAT	D70	70672	215-232
dhfr70kbD3R	CAAATGCTCACTATGTGAAG				
dhfr70kbDR-lab	GTGCAGATAGGAATAACATG				
dhfr75kbD3F	ACAAATCACCACACCAAA	AT	D75	75424	177-199
dhfr75kbDF-lab	AACAAACACACACGTTCT				
dhfr75kbDR	TACATGTAAGAAAGAACGT				
dhfr80kbD3F	GGCTATTTGTTCATAATGTTC	AT	D80	80215	272-295
dhfr80kbDR	GTGGACAGGTAAAACATATTA				
dhfr80kbDF-lab	GCAAACATCCAATGGTATCA				
dhfr90kbDF	ACAAATATCATGTTAGTAATACA	AT	D90	91191	246-264
dhfr90kbD3R	AGGACGTTAAAATGACGGC				
dhfr90kbDR-lab	CTTTCCCATAGACACTTGAA				
dhfr100kbD3f	TGACAAATTAAATGTAGG	TAAA	D100	101466	107-139
dhfr100kbDF-lab	CCTATAAGAAAAGATGAG				
dhfr100kbDR	GCACATCATTCTAACAA				
dhfr125kbD3F	ATATATGCTGAAATGTGG	AT	D125	126196	n/a
dhfr125kbDF-lab	AGACAAAGGTACAAATAC				
dhfr125kbDR	TTCAATCCTAACAACAAC				
dhfr150kbD3F	TTTCATCTTCTCCTCTTGA	TTA	D150	148016	220-253
dhfr150kbDF-lab	CCTTTGATTCGTTCATGC				
dhfr150kbDR	CAACGAACATACTGAAGAA				
dhfrii250kbD3F	GGTGACAAGAAAAAGGCGAC	ATA	D250	250856	221-275
dhfrii250kbDF-lab	GAATGCAAATGATAATGTGAATG				
dhfrii250kbDR	CATCGTCTTCAGCATTCATCT				
dhfr350kbD3F	GAAATGAATTTGTCTACC	AAT	D350	348652	166-181
dhfr350kbDF-lab	CTGCAATAAACATACAAC				
dhfr350kbDR	TAACCACATTTACAGCTA				

Appendix 4: Allele frequencies (%) at each microsatellite locus on *dhfr* sensitive and triple mutant chromosomes

U350	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
288	65.00	53.85	64.71	42.86	
291	10.00		5.88		
294	10.00	7.69	17.65	42.86	
297	10.00	30.77	11.76	14.29	
303	5.00	7.69			
	n=20	n=13	n=17	n=7	n=0

U250	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
261	23.81	23.08	11.76	25.00	
264		30.77			
267	14.29	30.77	41.18	50.00	
270	9.52	7.69	29.41	12.50	
273	19.05		5.88		
276	4.76	7.69	5.88	12.50	
279	19.05		5.88		
285	4.76				
287	4.76				
	n=21	n=13	n=17	n=8	n=0

U150	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
162	4.76		5.88		
166	19.05	15.38	35.29	25.00	
168	14.29	7.69	5.88		
170	9.52	7.69	5.88		
172	9.52	30.77		12.50	
174		7.69	5.88		
176	19.05			12.50	
178	9.52				
180	4.76		5.88		
184			5.88		
192			11.76		
200		7.69			
202			5.88		
208	9.52	23.08	5.88	50.00	
210			5.88		
	n=21	n=13	n=17	n=8	n=0

U125	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
158	5.26				
164	10.53	7.69			
168	5.26	15.38		14.29	
170	21.05			14.29	
172	10.53	15.38	12.50	14.29	
174	10.53	15.38			
176	5.26		6.25	28.57	
178	10.53	15.38		28.57	
180	5.26		6.25		
182		15.38	25.00		
184	5.26	7.69	18.75		
190		7.69	6.25		
192			6.25		
194			12.50		
196	5.26				
198	5.26		6.25		
	n=19	n=13	n=16	n=7	n=0

U100	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
309	4.76		7.69		
311	4.76				
313				12.50	
315	19.05	23.08	7.69	62.50	
317	14.29	15.38	7.69		
319		7.69			
321	23.81		23.08		
323	4.76	23.08	15.38		
325		15.38	15.38	12.50	
327	9.52	7.69			
330			7.69		
332	9.52				
334	4.76		15.38	12.50	
338	4.76	7.69			
	n=21	n=13	n=13	n=8	n=0

U75	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
247	25.00				
255	20.00		6.25		
257	30.00	7.69			
259	10.00	92.31	93.75	100.00	
261	10.00				
265	5.00				
	n=20	n=13	n=16	n=8	n=0

U70DD	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
480	42.86	76.92	26.67	57.14	57.14
483	4.76		13.33		3.57
499	4.76				
501	33.33	15.38	53.33	21.43	17.86
503		7.69		2.38	3.57
521	14.29		6.67	19.05	17.86
	n=21	n=13	n=15	n=42	n=28

U70	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
301				7.55	
304		9.09		1.89	
310	5.26	9.09		1.89	15.38
312	5.26		7.14	1.89	3.85
314	5.26	9.09			7.69
320				1.89	3.85
322	5.26		14.29		
323				1.89	3.85
327				3.77	
328	5.26			3.77	
329					7.69
331				3.77	
335	21.05		14.29	9.43	11.54
337	5.26				
339	10.53				
341		9.09		3.77	
347	5.26		14.29	11.32	
349		9.09			3.85
351	10.53	36.36	14.29	22.64	23.08
356				5.66	3.85
360				5.66	
362					3.85
363		9.09	14.29		
364				3.77	
371	5.26			5.66	
373	10.53			1.89	
375			14.29		3.85
379		9.09			3.85
383	5.26				
387				1.89	
389			7.14		
395					3.85
	n=19	n=11	n=14	n=53	n=26

U60	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
193			17.65		
220				3.57	4.35
222	4.76		5.88	7.14	
224	4.76	8.33		7.14	
226	14.29		5.88	1.79	6.52
228	23.81		29.41	16.07	17.39
230	28.57	83.33	29.41	55.36	39.13
232	4.76	8.33	5.88		17.39
234	4.76				6.52
236					4.35
238	14.29		5.88	3.57	2.17
240					2.17
242				5.36	
	n=21	n=12	n=17	n=56	n=46

U40	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
202					2.08
204	5.26			3.64	
206					2.08
208	5.26				6.25
209				1.82	
210	21.05	83.33	71.43	70.91	60.42
212	5.26				4.17
214	5.26				
215				1.82	
216	31.58		7.14	5.45	10.42
218	10.53				4.17
220				1.82	4.17
222	5.26		14.29	7.27	2.08
224	10.53	16.67		7.27	4.17
242			7.14		
	n=19	n=12	n=14	n=55	n=48

U30	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
137				1.82	
140	4.76				
142	14.29				
145					4.55
147	4.76		18.75	1.82	2.27
149	19.05			9.09	13.64
151	9.52		18.75	14.55	15.91
153	4.76			3.64	2.27
155				1.82	4.55
157	4.76			1.82	4.55
159	19.05		6.25		
161	14.29		6.25		6.82
163	4.76	92.31	50.00	56.36	43.18
165				5.45	
167		7.69		3.64	
173					2.27
	n=21	n=13	n=16	n=55	n=44

U20	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
199	10.00				
201		92.31	56.25	60.38	78.72
203	5.00	7.69	12.50	16.98	6.38
205	20.00		12.50	1.89	
207	15.00		12.50	5.66	2.13
209	15.00			3.77	6.38
211	10.00			1.89	2.13
213				1.89	2.13
215			6.25	5.66	
217	10.00			1.89	
219	15.00				
222					2.13
	n=20	n=13	n=16	n=53	n=47

U10	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
245	25.00				
253	20.00		6.25		
255	25.00	7.69	12.50	1.82	
257	15.00	92.31	81.25	89.09	93.02
259	10.00				6.98
261				1.82	
263	5.00				
265				7.27	
	n=20	n=13	n=16	n=55	n=43

MA1	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
197	15.79				
199	10.53				
203	21.05	100.00	91.67		
205	21.05				
207	5.26		8.33		
209	15.79				
211	10.53				
	n=19	n=7	n=12	n=0	n=0

MA2	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
159	5.56				
170	16.67				
172	22.22				
174	11.11				
175	5.56	18.18	6.25		
177	16.67	81.82	87.50		
179	11.11				
183	5.56				
187			6.25		
189	5.56				
	n=18	n=11	n=16	n=0	n=0

DHFR	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
88	80.00		7.69		
99	5.00				
100	5.00				
108		100.00	92.31		
110	10.00				
	n=20	n=7	n=13	n=0	n=0

D10	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
130	4.76			1.85	
133	4.76		5.88	1.85	2.17
136	47.62	92.31	76.47	83.33	78.26
139	33.33	7.69	17.65	5.56	13.04
143	4.76			3.70	4.35
147					2.17
150				3.70	
153	4.76				
	n=21	n=13	n=17	n=54	n=46

D20	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
280		23.08	5.88	5.45	5.41
283	90.48	61.54	94.12	87.27	91.89
286	9.52	7.69		7.27	
290					2.70
296		7.69			
	n=21	n=13	n=17	n=55	n=37

D30	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
126	4.76				2.38
130			11.76	9.09	9.52
132	14.29	8.33	5.88	3.64	9.52
134		25.00	29.41	21.82	26.19
136	9.52	8.33	5.88	9.09	2.38
138	4.76	8.33		3.64	14.29
140	9.52		5.88	3.64	7.14
142	14.29		5.88	7.27	4.76
144	4.76	16.67	5.88		
145				9.09	2.38
147	4.76	16.67	17.65	12.73	4.76
149	4.76	8.33	5.88	9.09	7.14
151	4.76				2.38
154				3.64	
155	9.52				
156				5.45	
157	4.76	8.33	5.88		
159	9.52				
160				1.82	7.14
	n=21	n=12	n=17	n=55	n=42

D40	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
299	9.52	15.38	14.29	14.55	30.77
301			7.14		
303				1.82	2.56
307	9.52			7.27	5.13
309		7.69			2.56
311				7.27	2.56
313	9.52	7.69		5.45	7.69
315				5.45	12.82
317	19.05		7.14	5.45	5.13
319	4.76	30.77	7.14	5.45	2.56
321	9.52	7.69		9.09	2.56
323	9.52	15.38	14.29	7.27	10.26
325	4.76	7.69	14.29	3.64	5.13
327	4.76	7.69	14.29	5.45	5.13
329	4.76		7.14		
331	4.76			3.64	
332					2.56
333	9.52		14.29		
334				7.27	2.56
336				10.91	
	n=21	n=13	n=14	n=55	n=39

D50	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
142				1.89	2.38
144					2.38
147	4.76	15.38		3.77	7.14
149				7.55	2.38
151	4.76		18.75	5.66	9.52
153	33.33	15.38	25.00	5.66	21.43
155	14.29	7.69		15.09	7.14
157	4.76	15.38	12.50	5.66	11.90
159	4.76	15.38		3.77	2.38
161	4.76		18.75	7.55	2.38
163	23.81	15.38	12.50	22.64	14.29
165				5.66	4.76
167		7.69	6.25	1.89	9.52
169			6.25	9.43	
171	4.76			1.89	
173		7.69			
175				1.89	2.38
	n=21	n=13	n=16	n=53	n=42

D60	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
242	13.33			2.13	6.25
245	20.00			4.26	6.25
248		11.11		10.64	9.38
251	13.33	33.33	60.00	40.43	46.88
254	53.33		30.00	25.53	18.75
257		22.22	10.00	2.13	3.13
260		11.11		4.26	3.13
263		11.11		4.26	6.25
269				6.38	
272		11.11			
	n=15	n=9	n=10	n=47	n=32

D70	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
215	5.56	9.09	20.00	1.82	8.33
218	22.22	36.36	6.67	5.45	16.67
221	5.56		20.00	18.18	16.67
224	66.67	54.55	53.33	60.00	37.50
227				7.27	14.58
229				7.27	6.25
	n=18	n=11	n=15	n=55	n=48

D80	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
272			5.88		
273				14.29	
274	5.00		11.76		
275	5.00	7.69		14.29	
276	5.00				
277	5.00	7.69			
280	20.00	15.38	23.53	42.86	
281	15.00	7.69			
282	5.00	7.69	11.76	14.29	
283	5.00		5.88		
284	10.00	30.77	23.53		
285	5.00		5.88		
286	5.00	15.38			
287	5.00				
288	5.00		11.76		
289		7.69		14.29	
293	5.00				
	n=20	n=13	n=17	n=7	n=0

D90	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
246				12.50	
249		7.69			
252			5.88		
255	9.52	7.69	5.88	12.50	
258	61.90	61.54	52.94	50.00	
261	28.57	23.08	29.41	12.50	
264			5.88	12.50	
	n=21	n=13	n=17	n=8	n=0

D100	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
107		20.00			
115	7.69	20.00	9.09		
117	7.69				
121	7.69	20.00			
123	23.08		27.27	33.33	
125	30.77		9.09		
127	15.38		9.09		
129		20.00	9.09		
131	7.69		18.18	33.33	
133		20.00			
135			18.18	33.33	
	n=13	n=5	n=11	n=3	n=0

D150	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
220		8.33	13.33		
232	5.56			14.29	
235		8.33	20.00		
238	33.33	25.00	40.00	14.29	
241	38.89	8.33	26.67	71.43	
244	16.67	8.33			
247	5.56	33.33			
250		8.33			
	n=18	n=12	n=15	n=7	n=0

D250	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
221	21.05		11.76		
227				12.50	
230	5.26				
233	15.79	7.69	23.53	12.50	
236	5.26	23.08	5.88		
239			5.88		
242		7.69	29.41		
245	15.79	23.08	5.88	25.00	
248		7.69			
251	21.05	7.69		25.00	
254	5.26				
257	5.26	15.38		25.00	
260	5.26	7.69			
263			11.76		
266			5.88		
	n=19	n=13	n=17	n=8	n=0

D350	Sensitive	KZN CIRN 1995	KZN CIRN 1999	Mpu CIRN	Tanz CIRN
166		7.69		12.50	
169	76.19	30.77	58.82	50.00	
175	14.29	46.15	35.29	37.50	
178	9.52				
181		15.38	5.88		
	n=21	n=13	n=17	n=8	n=0

Appendix 5: Allele frequencies (%) at each microsatellite locus on *dhfr* double mutant chromosomes

U70DD	Mpu CNRN	Tanz CNRN
480	57.89	52.63
483		5.26
499	5.26	
501	15.79	31.58
503	5.26	
521	15.79	10.53
	n=19	n=19

U70	Mpu CNRN	Tanz CNRN
310	8.33	6.25
314		6.25
320	4.17	
323		6.25
335	16.67	18.75
337	4.17	
339	4.17	6.25
341		12.50
345		6.25
347	8.33	6.25
349	4.17	
351	8.33	6.25
356	8.33	
360	20.83	6.25
362	8.33	
364	4.17	
371		6.25
375		6.25
381		6.25
	n=24	n=16

U60	Mpu CNRN	Tanz CNRN
220	7.69	
222	7.69	20.00
224	3.85	10.00
226	3.85	15.00
228	11.54	10.00
230	30.77	15.00
232	30.77	20.00
234		5.00
236	3.85	5.00
	n=26	n=20

U40	Mpu CNRN	Tanz CNRN
200		6.67
208	8.00	
210	36.00	60.00
214	4.00	
216	16.00	6.67
218		13.33
220	32.00	
222		13.33
224	4.00	
	n=25	n=15

U30	Mpu CNRN	Tanz CNRN
133	3.70	
137	11.11	14.29
140		4.76
147		4.76
149	3.70	
151	25.93	52.38
153	7.41	4.76
155	3.70	9.52
161		4.76
165	37.04	4.76
167	7.41	
	n=27	n=21

U20	Mpu CNRN	Tanz CNRN
197		6.25
201	19.23	6.25
203	38.46	62.50
205		6.25
207		6.25
209		6.25
211	19.23	
213	23.08	
222		6.25
	n=26	n=16

U10	Mpu CNRN	Tanz CNRN
253		4.76
257	76.00	66.67
259	24.00	28.57
	n=25	n=21

MA1	Mpu CNRN	Tanz CNRN
195	10.53	4.76
199	78.95	80.95
201	5.26	
203	5.26	9.52
207		4.76
	n=19	n=21

MA2	Mpu CNRN	Tanz CNRN
170		5.26
177	13.33	36.84
185	6.67	
187	80.00	42.11
189		15.79
	n=15	n=19

D10	Mpu CNRN	Tanz CNRN
133	14.81	
136	85.19	90.00
143		10.00
	n=27	n=20

D20	Mpu CNRN	Tanz CNRN
280	14.81	28.57
283	77.78	66.67
286	3.70	4.76
296	3.70	
	n=27	n=21

D30	Mpu CNRN	Tanz CNRN
126	3.70	5.00
128	3.70	
130		5.00
132	3.70	15.00
134	7.41	10.00
136	3.70	10.00
138		15.00
140	11.11	
142		5.00
145	25.93	10.00
147	11.11	15.00
149		5.00
151	3.70	
154	3.70	5.00
156	14.81	
160	3.70	
166	3.70	
	n=27	n=20

D40	Mpu CNRN	Tanz CNRN
297	4.35	
299	4.35	5.56
303	4.35	16.67
311	4.35	5.56
313		16.67
317	17.39	5.56
319	4.35	
321	8.70	11.11
323	26.09	
325	4.35	5.56
327		16.67
329	4.35	5.56
334	17.39	11.11
	n=23	n=18

D50	Mpu CNRN	Tanz CNRN
149	8.33	11.76
151	16.67	5.88
153	16.67	17.65
155	16.67	11.76
157	8.33	17.65
159	8.33	5.88
163	12.50	5.88
165		5.88
167	8.33	17.65
169	4.17	
	n=24	n=17

Appendix 6: All chromosome 8 mapping microsatellites described in this study (Chapter 6)

Name	Sequence	Repeat Unit	Marker	Position relative to dhps	Size Range
dhps450kbU3F	TACATGTATAGTGCAAAGCCC	TA	U450	-432403	n/a
dhps450kbUF-lab	GCAAAGCCCTTCGGGTCC				
dhps450kbUR	GTGTAATAAGCGTGCACACG				
dhps350kbU3F	TGAGTCTTGCACACTACTA	CAT	U350	-327070	n/a
dhps350kbUF-lab	CCAGAATGGACAAGCTTC				
dhps350kbUR	GTGATTGCCATCGGTCG				
dhps250kbU3F	GCTAAAAGGATAAGGATAT	TAT	U250	-232209	n/a
dhps250kbUF-lab	GCGGTTTATAATGACATG				
dhps250kbUR	TGGAAAAGGTGTACATTC				
dhps150kbU3F	GAAATATTGTGAGCATTGG	TA	U150	-151317	n/a
dhps150kbUF-lab	AATAAACCCCGATCTGTC				
dhps150kbUR	GCATAACAGGGTTATCAC				
dhps125kbU3F	TGTAATATACACATGCTG	TA	U125	-124300	334-360
dhps125kbUF-lab	AACATTGTACAAGTAAGG				
dhps125kbUR	CTGAGGATTAAATTAAGAA				
dhps100kbU3F	GTGAAGACATACGTTCACT	TA	U100	-100502	227-319
dhps100kbUF-lab	CATAGAAAGAGTTAAGAATAC				
dhps100kbUR	GCTATTACCTGCTATTAATC				
dhps90kbUF	CTACACTAATAGGTAGTTAG	ATTTT	U90	-91246	n/a
dhps90kbUR-lab	CGTTTATGTATGTAGAATAAAT				
dhps90kbU3R	TCATGGATTTTATGAATGCTA				
dhps80kbUF	GACGACGCAATTTCACTG	ATT	U80	-77874	229-247
dhps80kbUR-lab	GGATTTTTACTTCAGTATATG				
dhps80kbU3R	CAGCCTTATGTAATGATGTA				
dhps75kbU3F	CGTATATACGTTTATCAT	CTTTTTTT	U75	-74856	258-316
dhps75kbUF-lab	TATTAAACCATTCAACTG				
dhps75kbUR	CCAATGCATATATGAAAA				
dhps70kbUF	GGAATATTCAATATTCATGC	ATT	U70	-67216	n/a
dhps70kbUR-lab	CGTTTAATAAGAACCAAGG				
dhps70kbU3R	GATTGTGACGAAGAATGG				
dhps60kbUF	CTGCATATAATTATGTAAATTG	ATT	U60	-59750	296-330
dhps60kbR-lab	TCAAGACGAAAACGTATTAG				
dhps60kb3R	CAAGAATGTGCAGAAGTG				

Name	Sequence	Repeat Unit	Marker	Position relative to dhps	Size Range
dhps50kbU3F	TAAGTCATATATCATTGG	TA	U50	-50542	109-137
dhps50kbUF-lab	GAAATTGAAAAATATGCC				
dhps50kbUR	TATTTAGGAGATATACAC				
dhps50.2kbUF	CCTGTAGTATTTCCTCCT	TA	U50.2	-49173	211-247
dhps50.2kbUR-lab	GGCAACTAACAGAGGAAT				
dhps50.2kbU3R	CCGAACCTTAGAATCCAA				
dhps40kbU3F	CGATATCGTTCGAATGTG	TA	U40	-40411	153-199
dhps40kbUF-lab	TGGACATATAACTGCACC				
dhps40kbUR	ACTCTCCCACATTCAACA				
dhps40.2kbUF	GGGATTTGTCATTTCTTCC	TTA	U40.2	-37676	n/a
dhps40.2kbUR-lab	GAAAGGGTCCTAGGAAAC				
dhps40.2kbU3R	GGGGTATAGACGATACTA				
dhps30kbU3F	CATCATGCTTTATTCGTA	TA	U30	-30534	177-197
dhps30kbUF-lab	ATTCTCATGATGTTTATGT				
dhps30kbUR	TACATGTGTGTTATTTCC				
dhps30.2kbUF	GAGTACGTTAGCATCATC	TAA	U30.2	-26958	218-239
dhps30.2kbUR-lab	CCACGTTTATATTGTTCG				
dhps30.2kbU3R	GCATAGTTCGTTCTCATC				
dhps20kbU3F	CTTTTGTTACATCAAATTCG	TA	U20	-19600	396-420
dhps20kbUF-lab	GGTATAATGTACATTTTCAC				
dhps20kbUR	ATAAAGACAGAGAAAATTATC				
dhps20.2kbU3F	CGTCCAAATATAACGAATTAA	TA	U20.2	-18014	85-105
dhps20.2kbUF-lab	CTCTTTATAAGGTTGTCCC				
dhps20.2kbUR	GACTTTTTATATTTGCCTGTG				
dhps10kbU3F	CTCCCTTTTGTAAAATGTA	TA	U10	-10749	107-115
dhps10kbUF-lab	CGTTTCACATATGGTCAA				
dhps10kbUR	GAAATAAAACTTAGCTATTTG			0	
dhps					
dhps10kbD3F	TTTCTCCTTTTCTTACGGA	TA	D10	10721	142-166
dhps10kbDF-lab	TACGGATTCATTGTGCAC				
dhps10kbDR	CTGTAGCTCAATCTTAAGA				
dhps20kbD3f	CCCTAAATTAGATGATCA	TAA	D20	18810	n/a
dhps20kbDF-lab	AGATATCACAACCATCAC				
dhps20kbDR	GTGACTATGATGACTATT				

Name	Sequence	Repeat Unit	Marker	Position relative to dhps	Size Range
dhps20.2kbD3F	GAAAGCATCGGGTGATAA	TAA	D20.2	21596	394-448
dhps20.2kbDF-lab	CGAACCTGATTATAAGGAA				
dhps20.2kbDR	TGTATTCATGTTCCCATGT				
dhps30.2kbD3F	TATCATATTGTTCTTGACATAAA	TA	D30.2	29011	187-234
dhps30.2kbDF-lab	CTGTAATCTCTTGTTAATAATC				
dhps30.2kbDR	GAGCTGAATATTAAAATGATAA				
dhps30kbD3F	AATAGTAACAAGTTTGTG	TAA	D30	31957	n/a
dhps30kbDF-lab	ATTTTGACTGCTTATTTG				
dhps30kbDR	CTGTTCATTCAAATTTAC				
dhps40.2kbD3F	TATCTCGTGGGAAGGAGC	TA	D40.2	39849	n/a
dhps40.2kbDF-lab	AATACCTGGAAGGATAAATTT				
dhps40.2kbDR	GAGGAAGAATCATTTTGAATA				
dhps40kbD3F	TGGTTTAGATGATAAGTC	TAA	D40	42075	n/a
dhps40kbDF-lab	TATAATAGGGCATACAAC				
dhps40kbDR	CATTTTGAATACCACCAT				
dhps50.2kbD3F	GGAACCTAGAGGATGATG	TA	D50.2	50028	159-183
dhps50.2kbDF-lab	CTCCCACTCAAAGTTCTC				
dhps50.2kbDR	TTACGGAATATCAGTGGTC				
dhps60kbDF	GGGGAATATATTTCCGATC	ATA	D60	61522	256-273
dhps60kbDR-lab	ATTATCAGAATAATGTATATCC				
dhps60kbD3R	GTAAAGGGCACAATATTGAT				
dhps70kbDF	GTCCTACACATTAAGTTGTTC	TA	D70	71248	121-160
dhps70kbDR-lab	GGTTATATATTGTTCACCATG				
dhps70kbD3R	TTGAAACATACTAAAGAAGCTAT				
dhps75kbD3F	CTTGTGTCATCACAATTC	TA	D75	73587	189-203
dhps75kbDF-lab	TCTCCTATCCATTTTTGC				
dhps75kbDR	CCTCAATTTTTTATGTTTAC				
dhps80kbDF	CAACTGTTTTATATGAATGGG	-	D80	80791	n/a
dhps80kbDR-lab	TTGTCTACATTTTGTAGAACG				
dhps80kbD3R	CATATTATTCATAATGTGGGG				
dhps90kbD3F	AAGGACAATTCACTGGGC	TAAA	D90	91767	n/a
dhps90kbDF-lab	GCGTATGTCAATATATGGAG				
dhps90kbDR	ATTAGACTTTGCATGGCCTT				

Name	Sequence	Repeat Unit	Marker	Position relative to dhps	Size Range
dhps100kbD3F	CGATACAGATTTATATAAC	TA	D100	100542	113-145
dhps100kbDF-lab	AATATGCCACATACATAAT				
dhps100kbDR	GTCACATTTTATGATATATG				
dhps125kbD3F	TTCACCGTCATGGACTT	TA	D125	126314	364-400
dhps125kbDF-lab	CTTGAAGTCTTCTATGCT				
dhps125kbDR	CTTGTCCTGTTATTCCG				
dhps150kbD3F	TCGTCTTCTACCTTTACG	CATA	D150	147878	108-221
dhps150kbDF-lab	ATTTTCACGCGCTTTTCC				
dhps150kbDR	TTATAAGGAATTGCCCTGA				
dhps250kbD3F	GTTCGTATGCTCTTTTGT	-	D250	250699	n/a
dhps250kbDF-lab	GAAAAAAGCTAGCCATGA				
dhps250kbDR	CCTATTACAACATATGTTG				
dhps350kbD3F	GACAGCGAGCTATTACGA	ATT	D350	352476	263-337
dhps350kbDF-lab	GAGGGTTGCTTTGGTTCA				
dhps350kbDR	GGGAATGAGGAAAAACGTT				
dhps450kbD3F	TATTCAGATAATGCAACG	ATT	D450	451061	n/a
dhps450kbDF-lab	TTTTTCACATTATCACCC				
dhps450kbDR	ATCTGAAGATCTTGTGGA				
dhps550kbD3F	ATTTTATTCGTGTTCTAC	AT	D550	530114	100-124
dhps550kbDF-lab	TAGTATACTCAGCATAAG				
dhps550kbDR	CACATACCCAATATGTTA				

Appendix 7: Allele frequencies (%) at each microsatellite locus on *dhps* double mutant chromosomes

U125	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
334	4.17				
336	8.33		4.35		
338		5.00	8.70		
340		5.00	4.35	7.69	
342	4.17	20.00	4.35		
344	4.17	5.00	17.39		20.83
346	16.67	20.00	8.70		25.00
348	45.83	30.00	43.48	92.31	33.33
350	4.17				
352	4.17				4.17
354	8.33	10.00	8.70		12.50
356		5.00			
360					4.17
	n=24	n=20	n=23	n=13	n=24

U100	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
227	21.74	31.58	4.17	16.67	
229	4.35	5.26	12.50	8.33	33.33
231	13.04	10.53	16.67	16.67	12.50
233	4.35				
237	4.35	5.26			
239	4.35	5.26	4.17		
241		5.26			
243		5.26	4.17		8.33
245	13.04		16.67		8.33
249					16.67
251	4.35				
259		5.26			
269					4.17
271		5.26	8.33		
273			4.17		
275		5.26	8.33		
277	8.70	5.26	4.17		12.50
283	4.35		4.17		
287		5.26	8.33	16.67	4.17
289	4.35			41.67	
292			4.17		
297	4.35				
301		5.26			
319	8.70				
	n=23	n=19	n=24	n=12	n=24

U80	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
229	45.83	30.00		57.14	
235		5.00			
241	50.00	60.00		42.86	
244		5.00			
247	4.17				
	n=24	n=20	n=0	n=7	n=0

U75	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
258	14.29	5.26	4.76		
260				16.67	4.35
262			9.52		
264	4.76				
266		10.53	4.76		
268	23.81	26.32	23.81	25.00	34.78
276		15.79			
284	14.29			8.33	
286		5.26	9.52		
288		5.26			
292	23.81	15.79	28.57		21.74
296	9.52	5.26	4.76		8.70
300	4.76	5.26	14.29		4.35
316	4.76	5.26		50.00	26.09
	n=21	n=19	n=21	n=12	n=23

U60	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
296	15.79	15.79			
299	5.26				
302	5.26	26.32			
305	5.26			11.11	
308	5.26	10.53		33.33	
311	21.05	15.79		11.11	
314	15.79	5.26			
317	15.79			44.44	
320	5.26	10.53			
324		10.53			
327		5.26			
330	5.26				
	n=19	n=19	n=0	n=9	n=0

U50	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
109			5.00		
111	4.55	5.00			
115					4.76
117	31.82		10.00		
119	4.55	35.00	10.00		
121	13.64	15.00	20.00	58.33	80.95
123	31.82	20.00	30.00	16.67	
125	4.55	5.00	10.00	8.33	4.76
127	9.09	5.00	5.00		9.52
129		5.00	10.00	16.67	
133		5.00			
137		5.00			
	n=22	n=20	n=20	n=12	n=21

U50.2	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
211	18.75	20.00	13.04		
213		15.00	8.70		
215		5.00			
217	12.50			60.00	75.00
219			8.70		
221	12.50	5.00	13.04		
223	12.50		4.35		4.17
225	18.75	20.00	8.70	20.00	8.33
227		10.00	21.74		4.17
229	12.50		8.70	20.00	
231	6.25		4.35		
233		10.00	8.70		8.33
237	6.25	5.00			
243		5.00			
247		5.00			
	n=16	n=20	n=23	n=10	n=24

U40	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
153		5.26			
155	9.09	10.53	4.35		4.35
157		10.53			
159	13.64	5.26	13.04		4.35
161		10.53	4.35		
163	4.55	21.05	4.35		
165	13.64	5.26			
167	9.09		4.35	7.69	4.35
169			13.04		
171	9.09	5.26	8.70		
173		5.26			4.35
175	9.09		17.39		30.43
177	9.09	5.26	17.39		
179	9.09	10.53	8.70		
181	9.09			53.85	4.35
183		5.26	4.35		
185					4.35
187	4.55			23.08	39.13
199				15.38	4.35
	n=22	n=19	n=23	n=13	n=23

U30	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
177			4.35		
181	13.64	50.00	47.83	76.92	70.83
183	31.82	15.00	17.39	7.69	8.33
185	31.82	20.00	26.09		8.33
187	9.09	5.00	4.35		4.17
189	9.09	10.00		15.38	8.33
197	4.55				
	n=22	n=20	n=23	n=13	n=24

U30.2	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
218	4.17		4.17		
221		5.00			
223	29.17	25.00	16.67		4.17
226	37.50	45.00	45.83	91.67	87.50
228			4.17		
230	20.83	20.00	16.67	8.33	4.17
233	8.33		4.17		
235			8.33		4.17
239		5.00			
	n=24	n=20	n=24	n=12	n=24

U20	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
396			8.33		
398	4.35	5.56			
400	8.70	5.56	8.33		
402	30.43	27.78	16.67	46.15	100.00
404	4.35	5.56	8.33	7.69	
406	21.74		8.33	46.15	
408		11.11			
410			16.67		
412	4.35	11.11	16.67		
414	13.04	16.67	8.33		
416	8.70				
418		11.11	8.33		
420	4.35	5.56			
	n=23	n=18	n=12	n=13	n=8

U20.2	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
85		5.00			4.35
90	73.91	65.00	66.67	100.00	82.61
95	21.74	15.00	29.17		13.04
100	4.35	15.00			
105			4.17		
	n=23	n=20	n=24	n=12	n=23

U10	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
107		5.00			
109	41.67	65.00	87.50	100.00	100.00
111	37.50	30.00	8.33		
113	8.33		4.17		
115	12.50				
	n=24	n=20	n=24	n=13	n=24

D10	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
142	4.55		28.57		8.70
144	13.64	20.00	4.76	7.69	
146	9.09	5.00	14.29		13.04
148	4.55	10.00	9.52		
150	27.27	10.00			26.09
152	13.64	25.00	28.57	92.31	52.17
154	4.55	5.00	4.76		
156	4.55	5.00			
158	13.64		4.76		
160		5.00	4.76		
162	4.55	10.00			
166		5.00			
	n=22	n=20	n=21	n=13	n=23

D20.2	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
394		5.26			4.35
397	10.00	5.26	4.17		
400	10.00	5.26			4.35
403		5.26	4.17		
406		5.26			
409	20.00	15.79	4.17	58.33	8.70
412	10.00	5.26	12.50		4.35
415	10.00	21.05	20.83	25.00	47.83
418	30.00	15.79	16.67		4.35
421		15.79	4.17		4.35
424			12.50		
427			4.17		
430			12.50	16.67	21.74
436	10.00				
448			4.17		
	n=10	n=19	n=24	n=12	n=23

D30.2	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
187			4.55		
189			4.55		
191			4.55		
193	22.22	21.05	18.18		
195					4.55
197	22.22	5.26			4.55
199	11.11			8.33	
201	11.11		13.64		18.18
203		21.05	4.55	58.33	18.18
205	11.11	21.05	9.09		
207		5.26	4.55		4.55
209	11.11	5.26	22.73		
211			4.55		
213	11.11				
215		10.53	4.55	33.33	45.45
217					4.55
219		5.26			
227		5.26			
234			4.55		
	n=9	n=19	n=22	n=12	n=22

D50.2	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
159	20.00	10.53			
165		10.53	4.55		
167	10.00		22.73		39.13
169	60.00	47.37	45.45	75.00	43.48
171		15.79	4.55	8.33	8.70
173	10.00	5.26	4.55	16.67	4.35
175		5.26	13.64		
179		5.26			
181					4.35
183			4.55		
	n=10	n=19	n=22	n=12	n=23

D60	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
256					4.55
258	6.67				
261		5.26	19.05		31.82
264	33.33	31.58	33.33	77.78	27.27
267	26.67	31.58	28.57	22.22	36.36
270	20.00	26.32	19.05		
273	13.33	5.26			
	n=15	n=19	n=21	n=9	n=22

D70	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
121				16.67	
124				16.67	
127	12.50				
138	12.50	14.29			
140				33.33	
143		7.14			
145	12.50	28.57			
147		21.43			
150				16.67	
152		14.29			
154	25.00	7.14			
156	37.50				
158				16.67	
160		7.14			
	n=8	n=14	n=0	n=6	n=0

D75	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
189					10.00
191	9.52	6.25			
193			4.55		
195	23.81	18.75	31.82	25.00	15.00
197	47.62	56.25	40.91	75.00	50.00
199	14.29	18.75	18.18		15.00
201	4.76				5.00
203			4.55		5.00
	n=21	n=16	n=22	n=12	n=20

D100	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
113		10.00			
119		5.00	21.74		
121	8.70		17.39		8.70
123	13.04	15.00	4.35	46.15	34.78
125	4.35		4.35	7.69	
127	17.39	10.00	17.39		
129	4.35	15.00	4.35	7.69	21.74
131	13.04	15.00	8.70	15.38	17.39
133	30.43	10.00	17.39		4.35
135				7.69	8.70
137	4.35	10.00	4.35		
139	4.35	5.00		7.69	
141				7.69	4.35
145		5.00			
	n=23	n=20	n=23	n=13	n=23

D125	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
364	4.35	5.26	9.09		4.17
366	4.35				
368			4.55	7.69	
370	13.04	26.32	13.64		
372		10.53	4.55		
374	13.04			15.38	8.33
376	17.39		18.18	23.08	20.83
378	4.35	5.26	13.64		4.17
380		10.53	4.55		
382	8.70	26.32	13.64		16.67
384					16.67
386	4.35	5.26	13.64	46.15	12.50
388	17.39	5.26		7.69	8.33
390		5.26	4.55		
396	4.35				
398	4.35				8.33
399					
400	4.35				
	n=23	n=19	n=22	n=13	n=24

D150	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
108	4.35	5.56			
112		5.56			
116	4.35				
120		5.56			
124	8.70	5.56		23.08	
128	17.39			15.38	
132	13.04	16.67		7.69	
136	4.35	11.11		7.69	
140	4.35				
145	13.04	5.56		7.69	
150	8.70	11.11			
154	4.35	11.11			
158	4.35	5.56			
170	4.35				
174		5.56			
178		5.56			
182	4.35				
186	4.35				
201		5.56			
221				38.46	
	n=23	n=18	n=0	n=13	n=0

D350	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
263	9.52			8.33	
281	14.29	6.25			
284	4.76				
300	19.05	31.25		25.00	
303		6.25		50.00	
319	23.81	31.25			
322	9.52			16.67	
337	19.05	25.00			
	n=21	n=16	n=0	n=12	n=0

D550	KZN Sensitive '95	KZN Sensitive '99	Mpu Sensitive	KZN SGEAA	Mpu SGEAA
100	11.76				
104	5.88	17.65			
106	11.76				
108	5.88	41.18		20.00	
110		11.76		20.00	
112	52.94	5.88			
116	5.88	5.88			
120		5.88		40.00	
122				10.00	
124	5.88	11.76		10.00	
	n=17	n=17	n=0	n=10	n=0

Appendix 8: Publications arising from, or forming an integral part of the thesis.

Contribution to published work by candidate

Pearce et al 2003, Antimicrobial Agents and Chemotherapy:

I participated in the design of the study. I was the lead investigator in the adaptation and refinement of the SSOP dot blot technique for use in typing point mutations at *dhfr* and *dhps*. I carried out the point mutation typing and the population genetic data analysis. I wrote the paper, in consultation with Dr C. Roper.

Roper et al 2003, The Lancet

I participated in the design and execution of the molecular genetic analysis of the *dhfr* and *dhps* SNPs and microsatellites flanking *dhps*. I was involved in the analysis of both the SNP and microsatellite data.

Roper et al 2004, Science

I participated in the design and execution of the SNP and microsatellite genetic analysis.

Endorsement by supervisor, Dr C. Roper:

ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Apr. 2003, p. 1347–1354
0066-4804/03/$08.00+0 DOI: 10.1128/AAC.47.4.1347–1354.2003

Vol. 47, No. 4

Molecular Determination of Point Mutation Haplotypes in the Dihydrofolate Reductase and Dihydropteroate Synthase of *Plasmodium falciparum* in Three Districts of Northern Tanzania

Richard J. Pearce,[1]* Chris Drakeley,[1,2] Daniel Chandramohan,[1] Frank Mosha,[2,3] and Cally Roper[1]

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, United Kingdom,[1] and Kilimanjaro Christian Medical Centre[3] and The Joint Malaria Programme (JMP)[2] Moshi, Tanzania

Received 24 September 2002/Returned for modification 12 November 2002/Accepted 28 January 2003

The antimalarial combination of sulfadoxine and pyrimethamine (SP) was introduced as first-line treatment for uncomplicated malaria in Tanzania during 2001 following 18 years of second-line use. The genetic determinants of in vitro resistance to the two drugs individually are shown to be point mutations at seven sites in the dihydrofolate reductase gene (*dhfr*) conferring resistance to pyrimethamine and five sites in the dihydropteroate synthase (*dhps*) gene conferring resistance to sulfadoxine. Different combinations of mutations within each gene confer differing degrees of insensitivity, but information about the frequency with which allelic haplotypes occur has been lacking because of the complicating effects of multiple infection. Here we used a novel high-throughput sequence-specific oligonucleotide probe-based approach to examine the present resistance status of three *Plasmodium falciparum* populations in northern Tanzania. By using surveys of asymptomatic infections and screening for the presence of all known point mutations in *dhfr* and *dhps* genes, we showed that just five *dhfr* and three *dhps* allelic haplotypes are present. High frequencies of both triple-mutant *dhfr* and double-mutant *dhps* mutant alleles were found in addition to significant interregional heterogeneity in allele frequency. In vivo studies have shown that the cooccurrence of three *dhfr* mutations and two *dhps* mutations in an infection prior to treatment is statistically predictive of treatment failure. We have combined data for both loci to determine the frequency of two-locus genotypes. The triple-*dhfr*/double-*dhps* genotype is present in all three regions with frequencies ranging between 30 and 63%, indicating that treatment failure rates are likely to be high.

Sulfadoxine-pyrimethamine (SP) has now replaced chloroquine as the first-line curative antimalarial in much of East Africa. The earliest reports of emerging SP resistance in Africa were from Muheza district in Tanzania during 1994 and 1995 (32, 37). SP remains in use throughout Tanzania and was recently officially adopted as the national first-line treatment for nonsevere *Plasmodium falciparum* malaria. It is now a priority to learn how widespread genetic determinants of SP resistance currently are in the larger regions of northern Tanzania.

To investigate this issue, we have carried out a population-based genetic analysis of *P. falciparum* in the North and South Pare Mountains and Hai district, which are distinct geographical areas with their own microclimates and tribes. There had been no malaria research in these districts since 1965 until recent work showed relatively low levels of transmission, with an estimated entomological inoculation rate of 24 infective bites/person/year in Hai district (C. Drakeley and D. Chandramohan, unpublished data) compared with an entomological inoculation rate in the range of 34 to 405 infective bites/person/year in Muheza district (14).

It is widely understood that people self-treat with antimalarial drugs, which can be freely purchased (24); as a consequence, a proportion of people attending health facilities with signs and symptoms of malaria may have had recent prior exposure to drugs. In this study we have analyzed material from community surveys of asymptomatic infections. These data, we believe, are less subject to bias due to prior drug selection and are therefore representative of the parasite population at large.

Resistance to SP is associated in vitro with a series of substitutions within the active site of target enzymes of the folate biosynthesis pathway, dihydropteroate synthase (DHPS) (4, 38) and dihydrofolate reductase (DHFR) (9, 29, 36), and this has been demonstrated through laboratory-based in vitro sensitivity tests and transfection experiments with DHFR with respect to pyrimethamine (43, 44) and DHPS with respect to sulfadoxine (39, 40, 42). The sequence changes coding for substitutions that are naturally occurring worldwide are summarized in Table 1. A Ser-to-Asn substitution at codon 108 of DHFR decreases sensitivity to pyrimethamine 100-fold (43) Additional substitutions at codons 51 (N51I), 59 (C59R), and 164 (I164L) progressively increase levels of resistance to pyrimethamine (15, 43). Isolates containing all four substitutions have been found in South America and southeast Asia but have yet to be reported in Africa (21, 25, 41). A total of 14 DHPS substitutions at five sites have been characterized worldwide, of which 6 have been recorded in Africa, with A437G and

* Corresponding author. Mailing address: Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel St., London WC1E 7HT, United Kingdom. Phone: (44) 207 927 2416. Fax: (44) 207 636 8739. E-mail: Richard.Pearce@lshtm.ac.uk.

TABLE 1. Summary of each of the sites at which an SNP occurs which is known to be associated with SP resistance and the oligonucleotide probe designed to detect it

Probe	Amino acid	Probe sequence[a]
DHFR		
Codon 16		
A*[b]	Ala	CC ATA TGT GCA TG T TGT A
S	Ser	CC ATA TGT TCA TG T TGT A
V	Val	CC ATA TGT GTA TG T TGT A
Codons 50 and 51		
CN*	Cys Asn	TGG AAA TGT AAT TCC CTA
CN2*	Cys Asn	TGG AAA TGT AAC TCC CTA
RN	Arg Asn	TGG AAA CGT AAT TCC CTA
RN2	Arg Asn	TGG AAA CGT AAC TCC CTA
RI	Arg Ile	TGG AAA CGT ATT TCC CTA
CI	Cys Ile	TGG AAA TGT ATT TCC CTA
Codon 59		
C*	Cys	AA TAT TTT TGT GCA GTT A
R	Arg	AA TAT TTT CGT GCA GTT A
Codon 108		
N	Asn	A AGA ACA AAC TGG GAA AG
S*	Ser	A AGA ACA AGC TGG GAA AG
T	Thr	A AGA ACA ACC TGG GAA AG
Codon 140		
V*	Val	AT GAA GAT GTT TAT ATC A
L	Leu	AT GAA GAT CTT TAT ATC A
Codon 164		
I*	Ile	GT TTT ATT ATA GGA GGT T
L	Leu	GT TTT ATT TTA GGA GGT T
DHPS		
Codons 436 and 437		
SA*	Ser Ala	GAA TCC TCT GCT CCT TTT
SG	Ser Gly	GAA TCC TCT GGT CCT TTT
FA	Phe Ala	GAA TCC TTT GCT CCT TTT
FG	Phe Gly	GAA TCC TTT GGT CCT TTT
AA	Ala Ala	GAA TCC GCT GCT CCT TTT
AG	Ala Gly	GAA TCC GCT GGT CCT TTT
CA	Cys Ala	GAA TCC TGT GCT CCT TTT
Codon 540		
K*	Lys	ACA ATG GAT AAA CTA ACA
E	Glu	ACA ATG GAT GAA CTA ACA
Codon 581		
A*	Ala	A GGA TTT GCG AAG AAA CA
G	Gly	A GGA TTT GGG AAG AAA CA
Codon 613		
A*	Ala	GA TTT ATT GCC CAT TGC
T	Thr	GA TTT ATT ACC CAT TGC
S	Ser	GA TTT ATT TCC CAT TGC

[a] Entries in column 3 indicate the amino acid changes that occur at the relevant codons following the point mutation. The sequences in bold represent the codon within which the point mutation occurs.

[b] *, wild-type sensitive codon.

K540E mutations being the most frequently reported (12, 41, 42).

The role of the point mutations at each locus in conferring resistance to SP in vivo has been inferred from studies showing predictive association of particular mutations with treatment failure (21, 28) and from overrepresentation of mutations in recrudescent infections after treatment (2, 3, 8, 10, 11, 13, 16–18, 20, 26). Such studies are complicated by mixed infections. As the blood-stage parasites are haploid, the cooccurrence of two or more genotypes in an infection means that variation at multiple sites cannot be assigned to an individual parasite line within the infection and, accordingly, that the predictive association with treatment outcome is rendered less straightforward.

To study the frequency of alleles at a population level, we have designed a PCR high-throughput approach which uses sequence-specific oligonucleotide probes (SSOP) for detection of known single nucleotide polymorphisms (SNP) to identify and construct haplotypes. Haplotypes are combinations of SNP that are in the same gene in the same parasite, as distinct from associations of point mutations that cooccur because there is a mixture of parasites of different genotypes within a single infection. Haplotypes are biologically meaningful, since they determine the resistance properties of parasites that are exposed to drugs at the time of treatment. For example, a triple-mutant *dhfr* haplotype of N51I+C59R+S108N has 1.5- to 3-fold-higher pyrimethamine resistance in vitro than either the N51I+S108N or C59R+S108N double-mutant haplotype (35). A mixed infection containing these two double-mutant alleles is less resistant to pyrimethamine than an infection containing the triple-mutant allele, despite all three mutations being present in either case. When comparing populations, it is important to measure the frequency of haplotypes rather than the prevalence of each point mutation separately, because haplotypes are the determinants of resistance levels.

The method we have employed involves the PCR amplification of sequences from the coding regions of *dhfr* and *dhps* genes, which are fixed onto membranes and probed with SSOP (7) designed to detect each of the single base pair substitutions at all positions summarized in Table 1. The SSOP method has advantages for high throughput, while retaining sensitivity and specificity equivalent to those of other methods used for detection of *dhfr* and *dhps* SNP (1, 31). We used tetramethylammonium chloride (TMAC) to standardize the melting temperature of digoxigenin-labeled oligonucleotide probes, thus enabling duplicate membranes to be probed and washed at a standard temperature and sequence variants at all SNP sites to be detected simultaneously. Chemifluorescent signals from the detection of probes for differing SNP at the same site were compared quantitatively using a Storm phosphorimager. This high-throughput method allowed us to command data from large sample sizes.

MATERIALS AND METHODS

The study area and the samples. *P. falciparum*-positive samples were collected from two separate studies within the same area of northeast Tanzania; one was a study of infants and young children in Hai district, and the second was a study of people up to 45 years of age in the North and South Pare sites. Both studies were cross-sectional malariometric surveys across an altitude band of 550 to 1,600 m, and most of the study subjects were asymptomatic.

For the villages in Hai district, samples were collected in May 2001. All children less than 5 years old from 16 randomly selected villages were invited to attend the survey clinic at a central clinic. A finger-prick blood sample for blood slides and a filter paper blood sample were collected from each participating child. The filter paper blood samples were air dried and stored at 4°C with desiccant. Bloodspots from blood film-positive children were selected retrospectively for genotyping.

Samples from the North and South Pare Mountains were collected in November 2001 during malariometric cross-sectional surveys. A random sample of 1,250 individuals (250 per village) under 45 years of age was recruited, and a finger blood sample was taken into EDTA Microtainer tubes. Filter paper bloodspots were made with 10 μl of packed cells from samples of individuals found to be parasite positive.

Ethics. Scientific and ethical clearance for both studies was granted from the National Institute of Medical Research in Tanzania and the London School of Hygiene and Tropical Medicine. Consent was obtained from all individuals or their guardians prior to enrollment in the surveys.

TABLE 2. Table of PCR primer sequences and reaction conditions for the nested amplification of *dhfr* and *dhps*

Gene and primer	Primer sequence	PCR conditions
dhfr		
Outer, M1	5′ TTTATGATGGAACAAGTCTGC 3′	94°C × 3 min
650 bp, M7	5′ CTAGTATATACATCGCTAACA 3′	94°C × 1 min, 52°C × 2 min, 72°C × 1 min, 40×; 72°C × 10 min
Inner, M3b	5′ TGATGGAACAAGTCTGCGACGTT 3′	94°C × 3 min
594 bp, M9	5′ CTGGAAAAAATACATCACATTCATATG 3′	94°C × 1 min, 44°C × 2 min, 72°C × 1 min, 4×; 94°C × 1 min, 44°C × 1 min, 72°C × 1 min, 34×; 72°C × 10 min
dhps		
Outer, N1	5′ GATTCTTTTTCAGATGGAGG 3′	94°C × 3 min
770 bp, N2	5′ TTCCTCATGTAATTCATCTGA 3′	94°C × 1 min, 51°C × 2 min, 72°C × 1 min, 40×; 72°C × 10 min
Inner, R2	5′ AACCTAAACGTGCTGTTCAA 3′	As described above for *dhps* outer primer sequence
711 bp, R	5′ AATTGTGTGATTTGTCCACAA 3′	

DNA extraction. DNA extraction from bloodspots on filter paper was carried out in a 96-well plate format. A segment of the bloodspot was first soaked in 0.5% saponin–1× phosphate-buffered saline overnight and was then washed twice in 1 ml of 1× phosphate-buffered saline. The segment was then boiled for 8 min in 100 μl of PCR-quality water–50 μl of 20% Chelex suspension in distilled water (pH 9.5).

PCR amplification of *dhfr* and *dhps*. A 711-bp fragment of *dhps* and a 594-bp fragment of *dhfr* containing the polymorphic codons were independently amplified by nested PCR in a 96-well plate format. PCR primer sequences and reaction conditions are indicated in Table 2. The 25-μl PCR mix contained primers at 0.25 μM final concentration, 2 mM $MgCl_2$, 250 μM each deoxynucleoside triphosphate, and 1× Bioline *Taq* polymerase. Template DNA (1 μl) was introduced to outer reaction mixtures. The *dhps* outer PCR product (1 μl) was introduced into a 25-μl inner amplification mixture. Aliquots of 1 μl of threefold-diluted *dhfr* outer PCR product were introduced into a 25-μl inner amplification reaction mixture.

Use of SSOP for molecular genotyping of point mutations. Final-round PCR products were heat denatured (95°C for 2 min), cooled, and then spotted onto nylon membranes in 1-μl volumes in a 12 by 8 grid. A panel of four PCR samples of known sequences representing all common sequence variants was spotted on every blot to act as positive-negative controls for probe specificity. Replicate blots were made of each array so that simultaneous probing with the oligonucleotide probes of the full panel for that gene could be conducted. After drying, cross-linking was performed with 1,200-J UV light. Sequence-specific 18-bp oligonucleotide probes 3′ end labeled with digoxigenin (DIG) (Roche Boehringer Mannheim, Mannheim, Germany) were each designed to complement the known sequence polymorphisms in *dhfr* and *dhps* listed in Table 1. SNP-specific hybridization was followed by high-stringency TMAC washes, and alkaline phosphatase-conjugated anti-DIG Fab fragments (Roche Boehringer Mannheim) were used to detect DIG-labeled probes as described by Conway et al. (7). Visualization was performed through the alkaline phosphatase-catalyzed breakdown of the fluorogenic substrate ECF (Amersham Pharmacia Biotech, Little Chalfont, Buckinghamshire, United Kingdom) and scanned on a Molecular Dynamics Storm 840 PhosphorImager (Amersham Pharmacia Biotech).

Scoring. We scored the presence, absence, or relative abundance of the variant sequence polymorphism at each site separately. Images of blots probed with variant sequences for a single locus were transferred as TIF files to ImageMaster Total Lab software (Amersham Pharmacia Biotech). In the array analysis subsection of the software, a standard area of each spot was defined and the intensity of chemifluorescence in that area was measured. Background data were adjusted for by subtraction of the volume of the negative controls from the volume data. Thus, the volume of chemifluorescence for each spot was calculated as Volume = (maximum intensity × spot area) − background. To determine the threshold of detection per se, the presence-flagging option was employed. By this method, the faintest spot considered present and not background was selected to set the flagging threshold value. To compare SNP-specific probes at a single site, Microsoft Excel bar charts were drawn comparing the volume data for each probe on every sample to the presence-flagging result in each case. The following rules were used to determine whether a SNP was present or absent at each site.

(i) An SNP was considered present in a PCR product when the volume value with a particular probe was higher than that of the background. When volume values were low, presence flagging provided an internal control for avoiding possible biasing between probes or blots.

(ii) An SNP was considered absent when all volume values below the first gridline on the chart were rejected. No set value can be given for this criterion, as volume value comparison is relative and differs depending on the strength of the probe labeling and binding.

Samples were categorized into the single, majority, or mixed category at each site as follows. Samples were considered to be of mixed haplotypes when the volume value of the minority SNP was more than half the volume value of the majority SNP. Samples were considered to be mixed but containing a majority SNP when the minority SNP value was less than half of the majority value but higher than that represented by the first gridline on the chart. Samples were considered to be single when only one SNP was present at a site according to the rules given above.

To combine data from all sites in a gene and construct haplotypes, it was necessary to discard samples in which a mixture was found but which did not contain a majority SNP. Thus, for the purpose of generating frequency data, one haplotype was scored from each sample; this was either a single or majority type, because haplotypes from mixed infections cannot be constructed.

Following the same principle, for measuring the frequency of two-locus genotypes, a subset of samples in which a single or majority allelic haplotype was found at both *dhfr* and *dhps* was used.

Statistical analysis. Statistical analysis of population differences in haplotype frequencies (Wright's F_{st} [fixation index]) (6) and linkage disequilibrium (LD), given as D′ coefficient, was carried out using Arlequin software (33). Statistical analysis of contingency tables of the association of haplotypes within two-locus combinations was performed using a χ^2 test.

RESULTS

Of the 165 bloodspots that yielded PCR products, 10.3% were mixed at *dhps* and 1.2% were mixed at *dhfr*; no majority haplotypes were found. The low number of mixed infections was a reflection of the low level of transmission in the area. On stratification of the populations of the North Pare and South Pare Mountains into the age ranges of 0 to 4 and 5 to 45 years, no significant difference was found in the frequencies of *dhfr* and *dhps* allelic haplotypes, allowing comparisons to be made between the samples from those sites and those from Hai district, where samples were taken exclusively from subjects <5 years of age.

The *dhfr* and *dhps* allelic haplotypes present in each region are shown in Fig. 1. Three point mutations were found in *dhfr*, and of the eight possible haplotypic conformations, five were found; this matches well with findings of point mutations present in single-genotype infections elsewhere in East Africa that have been described previously (26, 41). Three point mutations were likewise found in *dhps*, and of the eight possible haplotypes, only three were found in *dhps*.

Regional haplotype frequencies are summarized in Fig. 2. Not all alleles at *dhfr* and *dhps* were present in each of the

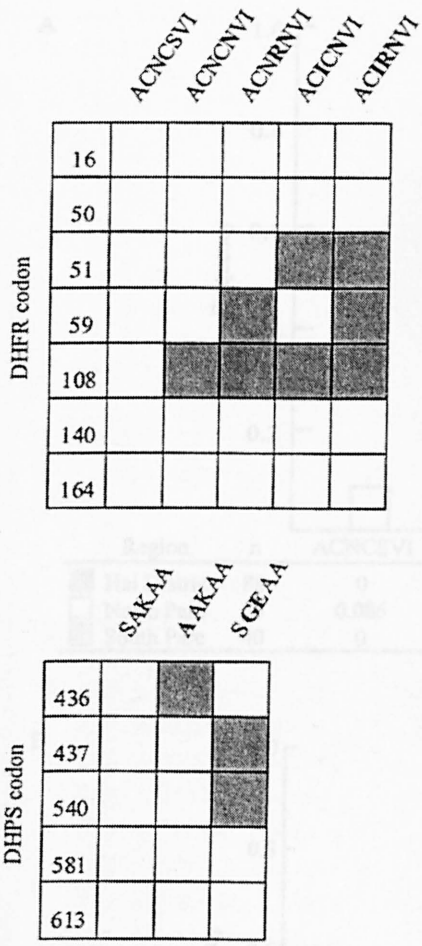


FIG. 1. Scheme of *dhfr* and *dhps* alleles found in this study. Shaded blocks indicate the sites at which the constituent SNP of each haplotype arise. Names of the alleles are composed of the amino acids present at each of the sites, in consecutive order by codon number, that are described as having a role in SP resistance. Substitutions are in boldface characters.

districts. The sensitive *dhfr* allele was not present in Hai district or the South Pare Mountains. The frequency of the *dhfr* triple-mutant allele was high in all districts and was highest in Hai (84.1%). The North Pare Mountains region had the lowest frequency of *dhfr* triple-mutant alleles; nevertheless, that frequency was five times greater than the frequency of sensitive *dhfr* alleles. We found no *dhfr* quadruple-mutant alleles (N51I+C59R+S108N+I164L).

The frequency of the *dhps* double-mutant allele was also greatest in Hai (64.2%) and lowest in the North Pare Mountains (43.3%); the frequency in the South Pare Mountains was intermediate (54.5%). Unlike that of *dhfr*, the sensitive allele of *dhps* was found in all districts. Furthermore, there was a low frequency (4 to 7.4%) of the single-mutant (S436A) allele in all districts.

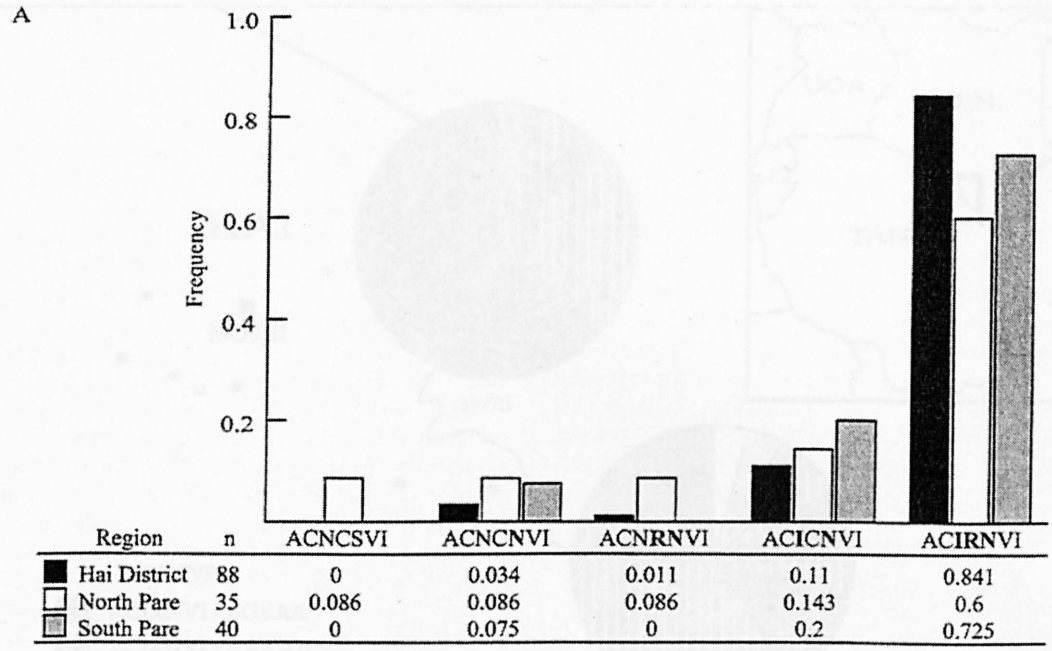
On calculation of the Wright's F_{st} (fixation index), the pairwise difference in haplotype frequencies at both loci between regions was shown to be statistically significant in comparisons of Hai district with the North Pare Mountains (*dhfr* F_{st} = 0.0733 [$P < 0.05$]; *dhps* F_{st} = 0.0748 [$P < 0.05$]). Interestingly, there was no significant difference in the remaining pairwise comparisons, despite the differences in haplotype frequencies between the North and South Pare Mountains.

Two-locus combinations were derived from single infections for which only one haplotype was recorded in both *dhfr* and *dhps*. The map in Fig. 3 shows the spatial distribution of frequencies of two-locus combinations. The reduction in sample size (n) reflects the loss due to mixed infections. The frequency of the most highly resistant genotype triple-mutant *dhfr*/double mutant *dhps* (ACIRNVI-SGEAA; substitutions are underlined) was found to be extremely high. The frequency was 63.2% in Hai district, 50% in the South Pare Mountains, and 22% in the North Pare Mountains. In the North Pare Mountains, there was more diversity at both loci and consequently a greater diversity of 2 locus genotypes. The Wright's F_{st} comparing the three populations at both loci further confirms the differences between the North and South Pare populations (F_{st} = 0.0583 [$P = 0.0054$]) and between the North Pare Mountains and Hai district (F_{st} = 0.149 [$P < 0.00001$]). Pairwise comparison of the South Pare and Hai district populations (F_{st} = 0.011) showed no significant difference and were merged for subsequent analysis. Statistical analysis of observed and expected two-locus combinations was performed on the population of the North Pare Mountains and the combined populations of Hai district and the South Pare Mountains. We found a significant departure from expected in the merged Hai-South Pare Mountains population (χ^2 test [$P = 0.0018$, 6 df]), whereas the distribution in North Pare was nonsignificant (χ^2 test [$P = 0.835$, 8 df]). LD analysis was performed on the combined Hai and South Pare Mountain data set, and we found three two-locus combinations to be in LD, namely, ACICNVI-SAKAA ($D' = 0.277$, $P = 0.01$), ACNRNVI-AAKAA ($D' = 1.0$, $P < 0.00001$), and ACIRNVI-SGEAA ($D' = 0.229$, $P = 0.031$) (substitutions are underlined). No other pair of alleles was found to be in LD.

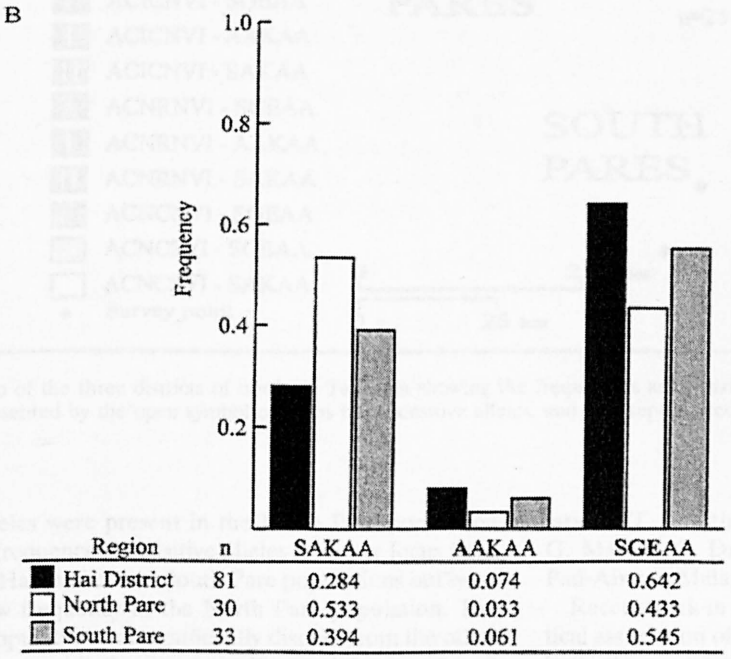
DISCUSSION

We have described a new approach by which blood survey material was used to determine the frequency of point mutation haplotypes in *P. falciparum* populations. This approach allows quantitation of resistance at the population level and enables direct comparison of population resistance levels even when they differ widely in the proportions of multiply infected individuals. The issue of multiple infections can be problematic when genotyping blood-stage parasites, because it causes haplotypic conformations of point mutations to be obscured and rare mutations to be oversampled. By recording one genotype per infection and discounting minority genotypes, we were able to avoid the oversampling of rare genotypes and to estimate the frequency of mutation haplotypes in the population in a manner which was standardized over all populations of different transmission intensities. With this consideration in mind, the SSOP method employed is designed for high-throughput screening of blood-stage infections to derive haplotype frequencies from survey material.

We have shown that only a subset of point mutations reported globally were present in North Tanzania in these populations, confirming previous reports of analyses of natural *P. falciparum* populations of East Africa (26, 41). Using single-



Region	n	ACNCSVI	ACNCNVI	ACNRNVI	ACICNVI	ACIRNVI
Hai District	88	0	0.034	0.011	0.11	0.841
North Pare	35	0.086	0.086	0.086	0.143	0.6
South Pare	40	0	0.075	0	0.2	0.725



Region	n	SAKAA	AAKAA	SGEAA
Hai District	81	0.284	0.074	0.642
North Pare	30	0.533	0.033	0.433
South Pare	33	0.394	0.061	0.545

FIG. 2. (A) Frequencies of *dhfr* alleles found in the district of Hai and the regions of the North and South Pare Mountains. (B) Frequencies of *dhps* alleles found in the district of Hai and the regions of the North and South Pare Mountains.

and majority-genotype infections, we were able to determine the point mutation haplotypes that were present in the three districts surveyed. We found five alleles at *dhfr*, and in reviewing the single-infection data from other studies in Africa, it is possible to see these same haplotypes in Kenya, Malawi, Tanzania, and Mauritania. We found three haplotypes in *dhps*, the sensitive-allele haplotype, the single-mutant allele S436A haplotype, and the double-mutant allele A437G K540E haplotype which has been widely recorded in East Africa (18, 19, 21, 25–27, 41) and not in West Africa (12, 30, 41).

Regional variation in population resistance. Allelic haplotype frequencies in the three regions differed significantly. Those of the Hai district and South Pare populations were highly resistant at *dhfr*, with triple-mutant allele (N51I+C59R+S108N) frequencies of >70% and *dhps* double-mutant allele (A437G+K540E) frequencies of >50%.

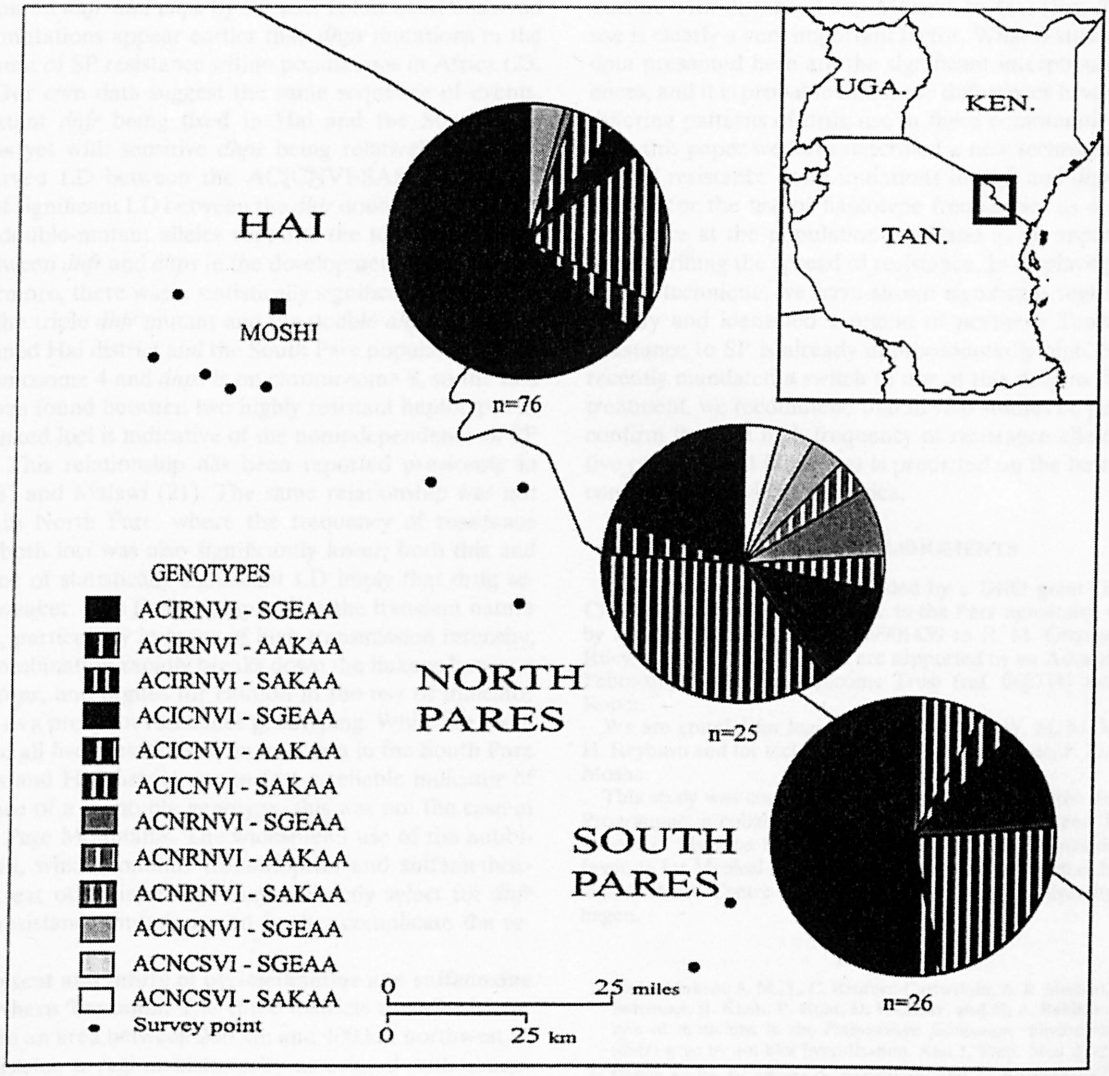


FIG. 3. Map of the three districts of northern Tanzania showing the frequencies and distribution of *dhfr/dhps* two-locus combinations. The genotype represented by the open symbol contains both sensitive alleles, and that represented by the black symbol contains the most resistant alleles.

The same alleles were present in the North Pare population but at lower frequencies. Sensitive alleles at either locus were absent in the Hai district and South Pare populations but were present at low frequency in the North Pare population. The North Pare population was significantly distinct from the other two populations, with the most significant difference found between those of the North Pare and Hai district populations. These findings point to a slower development of SP resistance in the North Pare Mountains. Possible causes of weaker selective pressure are greater use of alternative antimalarials such as amodiaquine and reduced use of antimalarials per se due to poor access to healthcare facilities or raised levels of acquired immunity. Such differences are maintained in the face of gene flow, which operates to make the parasite populations more homogeneous. Ongoing investigations into treatment-seeking behavior and overprescription by providers aim to further illuminate differences in drug use among the three study populations (T. Swarthout, D. Chandramohan, F. Mosha, A. Bell, G. Masuki, C. Drakeley, and H. Reyburn, Abstr. 3rd MIM Pan-African Malaria Conf., 2002).

Recent work in real populations in Malawi has found statistical association of the presence of all three *dhfr* mutations and both *dhps* mutations with failure to clear parasitemia after SP treatment (21). It is probable that in many cases this is due to the presence of the two most highly resistant alleles, triple *dhfr* and the double *dhps*—and this idea is supported by analysis of recrudescent infections following SP treatment (22). Here we directly measured the frequency of the highly resistant two-locus genotype consisting of the ACIRNVI-SGEAA (substitutions are underlined). We found that frequencies were twice as high in Hai district as in the South Pare Mountains and three times as high as in North Pare Mountains; hence, there is a clear and testable indication that SP treatment failure rates in these three regions are likely to differ.

Selection on *dhfr* and *dhps* by SP use. It has been observed that *dhfr* mutations appear earlier than *dhps* mutations in the development of SP resistance within populations in Africa (23, 26, 34). Our own data suggest the same sequence of events, with resistant *dhfr* being fixed in Hai and the South Pare Mountains yet with sensitive *dhps* being relatively common. The observed LD between the ACICNVI-SAKAA and the absence of significant LD between the *dhfr* double mutants and the *dhps* double-mutant alleles supports the idea of an interaction between *dhfr* and *dhps* in the development of resistance.

Furthermore, there was a statistically significant association between the triple *dhfr* mutant and the double *dhps* mutant in the combined Hai district and the South Pare populations. *dhfr* is on chromosome 4 and *dhps* is on chromosome 8, so the fact that LD was found between two highly resistant haplotypes at these unlinked loci is indicative of the nonindependence of SP selection. This relationship has been reported previously in Kenya (28) and Malawi (21). The same relationship was not apparent in North Pare, where the frequency of resistance alleles at both loci was also significantly lower; both this and the absence of statistically significant LD imply that drug selection is weaker. This finding emphasizes the transient nature of linkage, particularly in areas of high transmission intensity, where recombination rapidly breaks down the linkage between *dhfr* and *dhps*, and argues for caution in the use of indicator mutations as a proxy for resistance genotyping. While the cooccurrence of all five mutations in an infection in the South Pare Mountains and Hai district was in fact a reliable indicator of the presence of a quintuple genotype, this was not the case in the North Pare Mountains. The widespread use of the antibiotic Septrin, which contains trimethoprim and sulfamethoxazole, to treat other infections may indirectly select for *dhfr* and *dhps* resistance mutations and further complicate the relationship.

Past, present and future of pyrimethamine and sulfadoxine use in northern Tanzania. The three districts described here are found in an area between 200 km and 400 km northwest of Muheza district, a region historically associated with exceptionally high levels of antimalarial drug resistance. Resistance to pyrimethamine alone was reported in 1954 in Mngeza in Muheza district, following mass administration of prophylactic doses of pyrimethamine monotherapy over a five-month period during 1953 (5). Use of the SP combination began in Tanzania in 1982, when it was introduced as a second-line treatment for use in cases of chloroquine failure, and as early as 1984 it was the policy of the Muheza district hospital to use SP as a first-line antimalarial (25). Studies from the region report that SP was highly effective during the eighties, but resistance was recorded in Magoda village near Muheza in 1994 (32) and was subsequently reported in villages in the surrounding area (18, 19, 37). The emergence of resistance to SP in 1994 was attributed by the authors to be in part a result of the prophylactic intervention with weekly dapsone pyrimethamine treatment to all children less than 10 years old. Resistance to SP in Muheza district hospital is now reported to be as high as 45% (25).

Two explanations for the high frequency of resistance alleles in the three districts described here are the widespread use of SP or related drugs and the movement of resistance from Muheza. However, levels of resistance do not show a simple decline with distance from Muheza to Hai (Fig. 3), and drug use is clearly a very important factor. What is striking from the data presented here are the significant interpopulation differences, and it is probable that these differences have arisen from differing patterns of drug use in these communities.

In this paper we have described a new technique for detection of resistance point mutations in *dhfr* and *dhps*. We have argued for the use of haplotype frequencies as a measure of resistance at the population level and as an appropriate tool for describing the spread of resistance. In displaying the power of this technique, we have shown significant regional heterogeneity and identified a region of northern Tanzania where resistance to SP is already unprecedentedly high. Since policy recently mandated a switch to use of this drug as the first-line treatment, we recommend that in vivo studies be performed to confirm that the high frequency of resistance alleles is indicative of treatment failure, as is predicted on the basis of studies conducted elsewhere in Africa.

ACKNOWLEDGMENTS

Field studies in Hai were funded by a DfID grant (R7950 to D. Chandramohan et al.), and those in the Pare mountains were funded by a grant from MRC UK (G9901439 to B. M. Greenwood, E. M. Riley, et al.). C.R. and R.J.P. are supported by an Advanced Training Fellowship from The Wellcome Trust (ref. 060714) awarded to C. Roper.

We are grateful for logistical support from W. M. M. M. Nkya and H. Reyburn and for technical support from E. Nyale, F. Laizer, and M. Mosha.

This study was conducted under the auspices of the Joint Malaria Programme, a collaborative research initiative between The London School of Hygiene and Tropical Medicine, The Tanzanian National Institute for Medical Research, The Kilimanjaro Christian Medical College, and the Centre for Medical Parasitology, University of Copenhagen.

REFERENCES

1. **Abdel-Muhsin, A. M., L. C. Ranford-Cartwright, A. R. Medani, S. Ahmed, S. Suleiman, B. Khan, P. Hunt, D. Walliker, and H. A. Babiker.** 2002. Detection of mutations in the *Plasmodium falciparum* dihydrofolate reductase (dhfr) gene by dot-blot hybridization. Am. J. Trop. Med. Hyg. **67**:24–27.
2. **Basco, L. K., R. Tahar, A. Keundjian, and P. Ringwald.** 2000. Sequence variations in the genes encoding dihydropteroate synthase and dihydrofolate reductase and clinical response to sulfadoxine-pyrimethamine in patients with acute uncomplicated falciparum malaria. J. Infect. Dis. **182**:624–628.
3. **Basco, L. K., R. Tahar, and P. Ringwald.** 1998. Molecular basis of in vivo resistance to sulfadoxine-pyrimethamine in African adult patients infected with *Plasmodium falciparum* malaria parasites. Antimicrob. Agents Chemother. **42**:1811–1814.
4. **Brooks, D. R., P. Wang, M. Read, W. M. Watkins, P. F. G. Sims, and J. E. Hyde.** 1994. Sequence variation of the hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase gene in lines of the human parasite, *Plasmodium falciparum*, with differing resistances to sulfadoxine. Eur. J. Biochem. **224**:397–405.
5. **Clyde, D. F.** 1954. Observations on monthly pyrimethamine ("Daraprim") prophylaxis in an East African village. East Afr. Med. J. **3**:41–46.
6. **Cockerham, C. C., and B. S. Weir.** 1984. Covariances of relatives stemming from a population undergoing mixed self and random mating. Biometrics **40**:157–164.
7. **Conway, D. J., C. Roper, A. M. Oduola, D. E. Arnot, P. G. Kremsner, M. P. Grobusch, C. F. Curtis, and B. M. Greenwood.** 1999. High recombination rate in natural populations of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA **96**:4506–4511.
8. **Cortese, J. F., and C. V. Plowe.** 1998. Antifolate resistance due to new and known *Plasmodium falciparum* dihydrofolate reductase mutations expressed in yeast. Mol. Biochem. Parasitol. **94**:205–214.
9. **Cowman, A. F., M. J. Morry, B. A. Biggs, G. A. Cross, and S. J. Foote.** 1988. Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA **85**:9109–9113.
10. **Curtis, J., M. T. Duraisingh, and D. C. Warhurst.** 1998. In vivo selection for a specific genotype of dihydropteroate synthetase of *Plasmodium falciparum*

by pyrimethamine-sulfadoxine but not chlorproguanil-dapsone treatment. J. Infect. Dis. **177**:1429–1433.

11. **Doumbo, O. K., K. Kayentao, A. Djimde, J. F. Cortese, Y. Diourte, A. Konare, J. G. Kublin, and C. V. Plowe.** 2000. Rapid selection of *Plasmodium falciparum* dihydrofolate reductase mutants by pyrimethamine prophylaxis. J. Infect. Dis. **182**:993–996.
12. **Eberl, K. J., T. Jelinek, A. O. Aida, G. Peyerl-Hoffmann, C. Heuschkel, A. O. el Valy, and E. M. Christophel.** 2001. Prevalence of polymorphisms in the dihydrofolate reductase and dihydropteroate synthetase genes of *Plasmodium falciparum* isolates from southern Mauritania. Trop. Med. Int. Health **6**:756–760.
13. **Edoh, D., H. Mshinda, J. Jenkins, and M. Burger.** 1997. Pyrimethamine-resistant *Plasmodium falciparum* parasites among Tanzanian children: a facility-based study using the polymerase chain reaction. Am. J. Trop. Med. Hyg. **57**:342–347.
14. **Ellman, R., C. Maxwell, R. Finch, and D. Shayo.** 1998. Malaria and anaemia at different altitudes in the Muheza district of Tanzania: childhood morbidity in relation to level of exposure to infection. Ann. Trop. Med. Parasitol. **92**: 741–753.
15. **Hyde, J. E.** 1990. The dihydrofolate reductase-thymidylate synthetase gene in the drug resistance of malaria parasites. Pharmacol. Ther. **48**:45–59.
16. **Jelinek, T., A. H. Kilian, J. Curtis, M. T. Duraisingh, G. Kabagambe, F. von Sonnenburg, and D. C. Warhurst.** 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**:125–130.
17. **Jelinek, T., A. H. Kilian, G. Kabagambe, and F. von Sonnenburg.** 1999. *Plasmodium falciparum* resistance to sulfadoxine/pyrimethamine in Uganda: correlation with polymorphisms in the dihydrofolate reductase and dihydropteroate synthetase genes. Am. J. Trop. Med. Hyg. **61**:463–466.
18. **Jelinek, T., A. M. Ronn, J. Curtis, M. T. Duraisingh, M. M. Lemnge, J. Mhina, I. C. Bygbjerg, and D. C. Warhurst.** 1997. High prevalence of mutations in the dihydrofolate reductase gene of *Plasmodium falciparum* in isolates from Tanzania without evidence of an association to clinical sulfadoxine/pyrimethamine resistance. Trop. Med. Int. Health **2**:1075–1079.
19. **Jelinek, T., A. M. Ronn, M. M. Lemnge, J. Curtis, J. Mhina, M. T. Duraisingh, I. C. Bygbjerg, and D. C. Warhurst.** 1998. Polymorphisms in the dihydrofolate reductase (DHFR) and dihydropteroate synthetase (DHPS) genes of *Plasmodium falciparum* and in vivo resistance to sulphadoxine/pyrimethamine in isolates from Tanzania. Trop. Med. Int. Health **3**:605–609.
20. **Khan, B., S. Omar, J. N. Kanyara, M. Warren-Perry, J. Nyalwidhe, D. S. Peterson, T. Wellems, S. Kaniaru, J. Gitonga, F. J. Mulaa, and D. K. Koech.** 1997. Antifolate drug resistance and point mutations in *Plasmodium falciparum* in Kenya. Trans. R. Soc. Trop. Med. Hyg. **91**:456–460.
21. **Kublin, J. G., F. K. Dzinjalamala, D. D. Kamwendo, E. M. Malkin, J. F. Cortese, L. M. Martino, R. A. M. Mukadam, S. J. Rogerson, A. G. Lescano, M. E. Molyneux, P. A. Winstanley, P Chimpeni, T. E. Taylor, and C. V. Plowe.** 2002. Molecular markers for failure of sulfadoxine-pyrimethamine and chlorproguanil-dapsone treatment of *Plasmodium falciparum* malaria. J. Infect. Dis. **185**:380–388.
22. **Kun, J. F., L. G. Lehman, B. Lell, R. Schmidt-Ott, and P. G. Kremsner.** 1999. Low-dose treatment with sulfadoxine-pyrimethamine combinations selects for drug-resistant *Plasmodium falciparum* strains. Antimicrob. Agents Chemother. **43**:2205–2208.
23. **Mberu, E. K., M. K. Mosobo, A. M. Nzila, G. O. Kokwaro, C. H. Sibley, and W. M. Watkins.** 2000. 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.
24. **McCombie, S. C.** 1996. Treatment seeking for malaria: a review of recent research. Soc. Sci. Med. **43**:933–945.
25. **Mutabingwa, T., A. Nzila, E. Mberu, E. Nduati, P. Winstanley, E. Hills, and W. Watkins.** 2001. Chlorproguanil-dapsone for treatment of drug-resistant falciparum malaria in Tanzania. Lancet **358**:1218–1223.
26. **Nzila, A. M., E. K. Mberu, J. Sulo, H. Dayo, P. A. Winstanley, C. H. Sibley, and W. M. Watkins.** 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**:991–996.
27. **Nzila, A. M., E. Nduati, E. K. Mberu, C. Hopkins Sibley, S. A. Monks, P. A. Winstanley, and W. M. Watkins.** 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**:2023–2028.
28. **Omar, S. A., I. S. Adagu, and D. C. Warhurst.** 2001. Can pretreatment screening for dhps and dhfr point mutations in *Plasmodium falciparum* infections be used to predict sulfadoxine-pyrimethamine treatment failure? Trans. R. Soc. Trop. Med. Hyg. **95**:315–319.
29. **Peterson, D. S., D. Walliker, and T. E. Wellems.** 1988. Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in falciparum malaria. Proc. Natl. Acad. Sci. USA **85**: 9114–9128.
30. **Plowe, C. V., J. F. Cortese, A. Djimde, O. C. Nwanyanwu, W. M. Watkins, P. A. Winstanley, J. G. Estrada-Franco, R. E. Mollinedo, J. C. Avila, J. L. Cespedes, D. Carter, and O. K. Doumbo.** 1997. Mutations in *Plasmodium falciparum* dihydrofolate reductase and dihydropteroate synthase and epidemiologic patterns of pyrimethamine-sulfadoxine use and resistance. J. Infect. Dis. **176**:1590–1596.
31. **Ranford-Cartwright, L. C., K. L. Johnston, A. M. Abdel-Muhsin, B. K. Khan, and H. A. Babiker.** 2002. Critical comparison of molecular genotyping methods for detection of drug-resistant *Plasmodium falciparum*. Trans. R. Soc. Trop. Med. Hyg. **96**:568–572.
32. **Ronn, A. M., H. A. Msangeni, J. Mhina, W. H. Wernsdorfer, and I. C. Bygbjerg.** 1996. High level of resistance of *Plasmodium falciparum* to sulfadoxine-pyrimethamine in children in Tanzania. Trans. R. Soc. Trop. Med. Hyg. **90**:179–181.
33. **Schneider, S., D. Roessli, and L. Excoffier.** 2000. Arlequin ver. 2.000: a software for population genetics data analysis, ver. 2.000 ed. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.
34. **Sibley, C. H., J. E. Hyde, P. F. G. Sims, C. V. Plowe, J. G. Kublin, E. K. Mberu, A. F. Cowman, P. A. Winstanley, W. A. Watkins, and A. M. Nzila.** 2001. Pyrimethamine-sulfadoxine resistance in *Plasmodium falciparum*: what next? Trends Parasitol. **17**:582–588.
35. **Sirawaraporn, W., T. Sathitkul, R. Sirawaraporn, Y. Yuthavong, and D. V. Santi.** 1997. Antifolate-resistant mutants of *Plasmodium falciparum* dihydrofolate reductase. Proc. Natl. Acad. Sci. USA **94**:1124–1129.
36. **Snewin, V. A., S. M. England, P. F. Sims, and J. E. Hyde.** 1989. Characterisation of the dihydrofolate reductase-thymidylate synthetase gene from human malaria parasites highly resistant to pyrimethamine. Gene **76**:41–52.
37. **Trigg, J. K., H. Mbwana, O. Chambo, E. Hills, W. Watkins, and C. F. Curtis.** 1997. Resistance to pyrimethamine/sulfadoxine in *Plasmodium falciparum* in 12 villages in north east Tanzania and a test of chlorproguanil/dapsone. Acta Trop. **63**:185–189.
38. **Triglia, T., and A. F. Cowman.** 1994. Primary structure and expression of the dihydropteroate synthetase gene of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA **91**:7149–7153.
39. **Triglia, T., J. G. Menting, C. Wilson, and A. F. Cowman.** 1997. Mutations in dihydropteroate synthase are responsible for sulfone and sulfonamide resistance in *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA **94**:13944–13949.
40. **Triglia, T., P. Wang, P. F. Sims, J. E. Hyde, and A. F. Cowman.** 1998. Allelic exchange at the endogenous genomic locus in *Plasmodium falciparum* proves the role of dihydropteroate synthase in sulfadoxine-resistant malaria. EMBO J. **17**:3807–3815.
41. **Wang, P., C. S. Lee, R. Bayoumi, A. Djimde, O. Doumbo, G. Swedberg, L. D. Dao, H. Mshinda, M. Tanner, W. M. Watkins, P. F. Sims, and J. E. Hyde.** 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**:161–177.
42. **Wang, P., M. Read, P. F. Sims, and J. E. Hyde.** 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**:979–986.
43. **Wu, Y., L. A. Kirkman, and T. E. Wellems.** 1996. Transformation of *Plasmodium falciparum* malaria parasites by homologous integration of plasmids that confer resistance to pyrimethamine. Proc. Natl. Acad. Sci. USA **93**: 1130–1134.
44. **Zolg, J. W., J. R. Plitt, G. X. Chen, and S. Palmer.** 1989. Point mutations in the dihydrofolate reductase-thymidylate synthase gene as the molecular basis for pyrimethamine resistance in *Plasmodium falciparum*. Mol. Biochem. Parasitol. **36**:253–262.

Mechanisms of disease

Antifolate antimalarial resistance in southeast Africa: a population-based analysis

Cally Roper, Richard Pearce, Barry Bredenkamp, Jonathan Gumede, Chris Drakeley, Frank Mosha, Daniel Chandramohan, Brian Sharp

Summary

Background Sulfadoxine-pyrimethamine was first introduced for treatment of malaria in Africa during the early 1980s for cases when chloroquine treatment failed, and has since become the first-line treatment in many countries. Resistance to sulfadoxine-pyrimethamine is now increasing, especially in southeast Africa.

Methods We characterised genetic change in *dhfr* and *dhps* genes in the *Plasmodium falciparum* population of KwaZulu-Natal, South Africa, during 1995–99, a period of rapid deterioration of the effectiveness of sulfadoxine-pyrimethamine. We assessed the evolutionary origin of the resistance by analysing polymorphic microsatellite repeats in the flanking region of the *dhfr* and *dhps* genes, which show whether resistance alleles originated through shared or independent ancestral mutation events. We then assessed the current extent of dispersal of *dhfr* and *dhps* resistance alleles by doing the same analysis in *P falciparum* sampled from communities in the Kilimanjaro region of northern Tanzania in 2001.

Findings The large genetic change during 1995–99 in KwaZulu-Natal, South Africa, in both the health facility and the wider community surveys, was at the *dhps* locus, apparently because resistance at *dhfr* was established before 1995. The allelic determinants of resistance in this province share a common evolutionary origin with those found in Kilimanjaro, Tanzania, even though the two sites are 4000 km apart.

Interpretation Three resistant *dhfr* alleles, and one resistant *dhps* allele, each derived from independent ancestral lineages, have been driven through through southeast Africa. The movement by the *dhfr* alleles (pyrimethamine resistance) preceded that of the *dhps* allele (sulfadoxine resistance). Our findings emphasise that gene flow rather than new mutations has been the most common originator of resistance in African countries.

Lancet 2003; **361:** 1174–81
See Commentary page ????

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK (C Roper PhD, R Pearce MSc, C Drakeley PhD, D Chandramohan PhD); **Malaria Research Lead Programme, Medical Research Council, PO Box 70380, Overport 4067, Durban, South Africa** (B Bredenkamp MSc, B Sharp PhD); **Department of Health, Malaria Control Programme, KwaZulu-Natal, South Africa** (J Gumede EPH); **Joint Malaria Programme (JMP), Kilimanjaro Christian Medical Centre, Moshi, Tanzania** (C Drakeley, F Mosha PhD)

Correspondence to: C Roper, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, UK
(e-mail: Cally.Roper@lshtm.ac.uk)

Introduction

Resistance to sulfadoxine-pyrimethamine is increasing, especially in southeast Africa, and the immense human cost of the successive failure of these two drugs in Africa was described in 1998 as a malaria disaster.[1] The search is on for ways to protect and extend the useful life of the next generation of antimalarials, and a clear research priority is to understand the dynamics underlying the emergence of drug resistance and the extent and rate of its spread. The genetic basis of resistance to sulfadoxine and pyrimethamine is point mutations in the genes encoding their target enzymes—dihydropteroate synthase (DHPS) and dihydrofolate reductase (DHFR), respectively.[2,3] These common mutations are assumed to have arisen independently on numerous occasions.[4]

Pyrimethamine and sulfadoxine target two enzymes in the folate synthesis pathway of *Plasmodium falciparum*—DHPS and DHFR. When the two drugs are administered together as sulfadoxine-pyrimethamine, they act synergistically to disrupt folate synthesis and kill the parasite. Resistance to sulfadoxine-pyrimethamine has evolved worldwide, and is caused by point mutations that accumulate at several sites in the *dhfr* and *dhps* genes.[2,3] In both genes, every successive mutation incrementally increases the parasite's tolerance to the drug in vitro.[5–7] In Africa, each locus has just two or three point mutations that lead to resistance[8,9] and these mutations are a subset of those found worldwide.[2,3]

The rate at which point mutations occur in *P falciparum* is typical of eukaryotes, and a point mutation in *dhfr* that codes for a S108N substitution occurs at between $2{\cdot}5\times10^{-9}$ and 3×10^{-11} mutations per replication.[10] Since patients with malaria generally have between 10^{8} and 10^{12} haploid parasites circulating in their blood at the time of treatment, even at the most conservative estimate of mutation rate new resistance mutations could be expected to arise and be selected for within one patient.[11] For pyrimethamine and sulfadoxine, in which a few point mutations can directly confer resistance, resistance will probably arise rapidly and on many occasions.[4]

To assess the evolutionary origins of *dhfr* and *dhps* resistance determinants we investigated polymorphic MICROSATELLITE REPEATS within an 8 kb flanking region of both genes. Such markers are a robust method to distinguish between resistance alleles that share ancestry and those that have arisen independently. Results of a study[12] of microsatellite variation flanking the chloroquine resistance determining locus *pfcrt* have shown that resistance alleles found worldwide all derive from four ancestral mutants. We have therefore compared *dhfr* and *dhps* resistance alleles from South Africa and Tanzania to determine the extent to which alleles of common origin are dispersed. We also assessed the stability of these alleles over time in a longitudinal

GLOSSARY

HYPOENDEMIC
Regions are hypoendemic when the transmission intensity and incidence of a disease are sufficiently low that the population has little or no immunity

MICROSATELLITE REPEATS
Tandemly repeated sequence motifs of one, two, three, or four base pairs

POLYMORPHISM
Where more than one allele exists at a specific locus within a population

RELATIVE FITNESS
Fitness is the ability of a genotype to survive and reproduce—the change in frequency of a resistance allele is mathematically dependent on the fitness advantage it confers. Relative fitness is probability of survival and reproduction of a specific genotype relative to that of other genotypes in the same population

analysis of the *P falciparum* population of northern KwaZulu-Natal, South Africa.

The province of KwaZulu-Natal in South Africa replaced chloroquine with sulfadoxine-pyrimethamine as a first-line therapy in 1988. The first African country to change, Malawi, did not do so until 1993, so the *P falciparum* population in KwaZulu-Natal was exposed to first-line sulfadoxine-pyrimethamine selection early. In 1996, results of a study based at Mosvold hospital in the Ingwavuma district of KwaZulu-Natal showed that 20% of patients given sulfadoxine-pyrimethamine did not clear the parasites within 14 days of treatment. By 2000 this proportion had increased to 70%,[13] whereon the province switched to using coartemether.

Methods

Patients

In KwaZulu-Natal malaria is HYPOENDEMIC and transmission takes place during the second half of the rainy season (January to June, with peak transmission in April). We obtained clinical samples from outpatients seeking treatment at Mosvold Hospital and its satellite clinic at Ndumu in April, 1996, and April, 1999. In 1996, we recruited patients for an in-vivo clinical trial to test the effectiveness of chloroquine and sulfadoxine-pyrimethamine, and in 1999, we compared immunochromatographic diagnostic tests for malaria. The patients all had positive bloodfilms for *P falciparum* and we excluded those who reported receiving antimalarial drugs in the previous 14 days.

Procedures

During February, June, and September, 1995, we surveyed the blood of 1200 people living within a 12 km radius of Ndumu clinic. Blood spots were dried on filter paper (Whatman 3MM, Whatman International, Maidstone, UK) and stored at –20°C. The 1999 survey material was gathered by Department of Health personnel in routine active case detection in high risk areas of Ingwavuma district between January and March, 1999. ICT diagnostic test cards (Amrad ICT, Sydney, Australia) positive for *P falciparum* were retained and stored with desiccant. Parasite DNA was extracted from the paper blotter on which the fingerprick blood was introduced by the standard method.

P falciparum positive samples were gathered in villages spanning 250 km and at altitudes of 550–1600 m in Hai, North Pare, and South Pare districts. Most study participants were asymptomatic. In May, 2001, we obtained samples from children aged younger than 5 years from 16 villages in the Hai District. The filter paper blood samples were air dried and stored at 4°C with desiccant. Samples from North and South Pare districts were obtained in November, 2001 during malaria cross-sectional surveys. We recruited a consecutive sample of 1250 individuals (250 per village) younger than 45 years of age and obtained a finger blood sample, which was placed in an EDTA microtainer. Filter paper blood spots were made with 10 μL of packed cells from samples of individuals who were positive for the parasite and the samples were preserved as previously.

We screened the parasite DNA for all known sequence variants at sites where point mutations are known—namely codons 16, 50, 51, 59, 108, 140, and 164 of the *dhfr* gene, and codons 436, 437, 540, 581, and 613 of the *dhps* gene.[2] The DNA was extracted from bloodspots on filter paper or from positive ICT diagnostic test cards using a standard preparation method. First, we cut a segment of the dried blood spot or the excised paper blotter from the ICT diagnostic test card and (the site where the fingerprick blood is initially introduced to the test) using a sterile blade, and soaked the samples in 0·5% saponin-1× phosphate-buffered saline overnight. We then washed the sample in 1 mL 1× phosphate-buffered saline. The segment was then boiled for 8 min in 100 μL PCR quality water with 50 μL 20% chelex suspension in PCR quality water (pH 9·5).

We amplified a 594 bp fragment of *dhfr* and a 711bp fragment of *dhps*, each of which contained the coding sequence where known sites of sequence POLYMORPHISM are found, using nested PCR. These PCR reactions were done independently with primers complementary to conserved sequence in these genes.[14] Amplification of the *dhfr* or *dhps* fragment was therefore independent of the allele-specific sequence polymorphism. The resulting PCR products were then fixed to nylon membrane and probed for known sequence polymorphisms by hybridisation to specific oligonucleotide probes.[14]

Blood-stage *P falciparum* parasites are haploid, so when a patient was infected with a single parasite

Primer sequences for semi-nested PCR amplification of microsatellite loci

Primer	Sequence
DHFR0.3kb.3R	GGC ATA AAT ATC GAA AAC
DHFR0.3kb.F	ATT CCA ACA TTT TCA AGA
DHFR0.3kb.RHEX	TCC ATC ATA AAA AGG AGA
DHFR4.4kb.3F	GTT GTC AAT AAT TTC TGC ATC
DHFR4.4kb.R	CGA TAT ATC TGA TGG GTG A
DHFR4.4kb.FFAM	TAC CAT AGC AGT CTT TGC A
DHFR5.3kb.3F	TAC ATA ATT CAT ATG AAC TTG
DHFR5.3kb.R	CAC ATA TTA TAC AGG ACG A
DHFR5.3kb.FFAM	CCT GCA TTT GCA AGA AGT A
DHPS0.8kb.F	GAC CAA GTG TAA TTT AC
DHPS0.8kb.FFAM	GGA AAG TGC AAA CAT GT
DHPS0.8kb.3R	GAC ATA TAA TGA GCA TG
DHPS4.3kb.3F	AAC TTT TCG TGG GTA AAG
DHPS4.3kb.FHEX	GTA TGA ATA ATA TTA CCC TT
DHPS4.3kbR	GTT CGA TAT ATG CAC ACA
DHPS7.7kb3F	GTG TCC TAT AAG TAT TGA
DHPS7.7kbFFAM	GGT TAT CAA TAT GTA CAT
DHPS7.7kbR	TGA TAG TAC ATT ATG TAG

genotype, we could determine unequivocally the haplotype of resistance point mutations of that parasite, which is not possible when a patient is infected with several genotypes at once. Fortunately KwaZulu-Natal has low rates of malaria infection compared with other regions of Africa, so there are fewer mixed genotype infections than might be found elsewhere. Of 382 KwaZulu-Natal samples typed at all *dhfr* and *dhps* loci, 70% were infected by a single genotype. We included a further 8% of infections where a majority haplotype outnumbered the minority by greater than a two-to-one ratio, and excluded the remaining mixed infections. Likewise the cross-sectional surveys in Tanzania were across an altitude of 550–1600 m, where transmission rates are low and the proportion of single genotype infections was 80–90%.

To assess the evolutionary origins of *dhfr* and *dhps* resistance haplotypes, we investigated polymorphic microsatellite repeats within 8 kb flanking regions of both genes. We analysed microsatellite sequences located 0·3 kb, 4·4 kb, and 5·3 kb upstream of codon 108 of the *dhfr-ts* gene, which is on chromosome 4; and 0·8 kb, 4·3 kb, and 7·7 kb downstream from codon 437 of *pppk-dhps* gene, which is on chromosome 8. We amplified each locus by PCR then assessed the size of the fragments (reflecting the number of repeats) by gel electrophoresis. Alleles were PCR amplified and identified according to PCR fragment size, which was determined by electrophoresis on an ABI 377 and analysed with Genescan and Genotyper software (Applied Biosystems, Warrington, Cheshire, UK). We used a semi-nested PCR design, and primer sequences (panel) were designed from the 3D7 isolate genome PlasmoDb database[15] except DHFR.3R and DHFR.F which were described by Su and Wellems.[16] The primary reaction contained 1 μL template, 3·0 mmol/L Mg^{2+}, and 0·75 pmol/L each primer. Cycling conditions were: 2 min at 94°C, and then 25 repeated cycles of 30 s at 94°C; 30 s at 42°C; 30 s at 40°C; and 40 s at 65°C; followed by 2 min at 65°C. The third primer was fluorescently labelled and incorporated into a second round PCR reaction of total volume 11 μL, introducing 1 μL of product from the primary reaction, 2·5 mmol/L Mg^{2+}, and 2 pmol/L of each primer. Cycling conditions were: 2 min 94°C, then 25 cycles 20 s at 94°C; 20 s at 45°C; 30 s at 65°C, followed by a final step of 2 min at 65°C.

We used longitudinal analysis of the *P falciparum* population of KwaZulu-Natal to directly measure the RELATIVE FITNESS of resistant and sensitive alleles at *dhfr* and *dhps* while drug selection through use of sulfadoxine-pyrimethamine as firstline therapy was applied. We calculated relative fitness on the basis of frequency changes observed in the 1995 and 1999 community surveys by 1+s (the selection coefficient), where s was calculated from the formula:[17] $P_n=P_o\times e^{sn}$, where P_n=frequency after n generations, P_o=frequency at start, and n=number of generations (we assumed three generations a year). The curves between the points were drawn with predicted frequency changes based on these relative fitness values.

Statistical analysis

We compared the allele frequencies of *dhfr* and *dhps* in the KwaZulu-Natal time-series samples using Fisher's exact test in STATA version 7.0. To test the significance of the changes in frequency of one allele alone, we used χ^2 test or Fisher's exact test as appropriate.

Role of the funding source

The sponsors of this study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

In all *P falciparum* samples investigated, either from KwaZulu-Natal or Tanzania, we identified point mutations only at codons 108, 51, and 59 of the *dhfr* gene and at codons 436, 437, and 540 of the *dhps* gene. Allelic haplotypes—the conformations of point mutations—in *dhfr* and *dhps* were recorded only from single and majority genotype infections. Figure 1 shows the frequency of allelic haplotypes at *dhfr* and *dhps* in

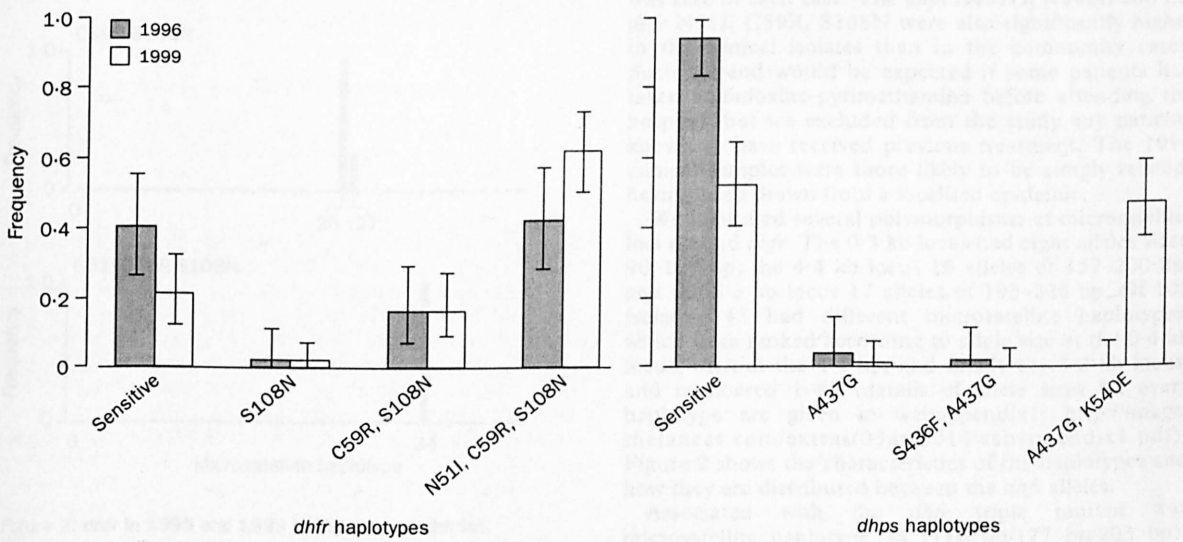


Figure 1: **Frequency of allelic haplotypes at *dhfr* and *dhps* in 1996 and 1999 in outpatients attending Mosvold Hospital and its satellite clinic at Ndumu, Ingwavuma District, ZwaZulu-Natal, South Africa**
Frequency values are shown with 95% CI (bars) calculated by binomial exact method.

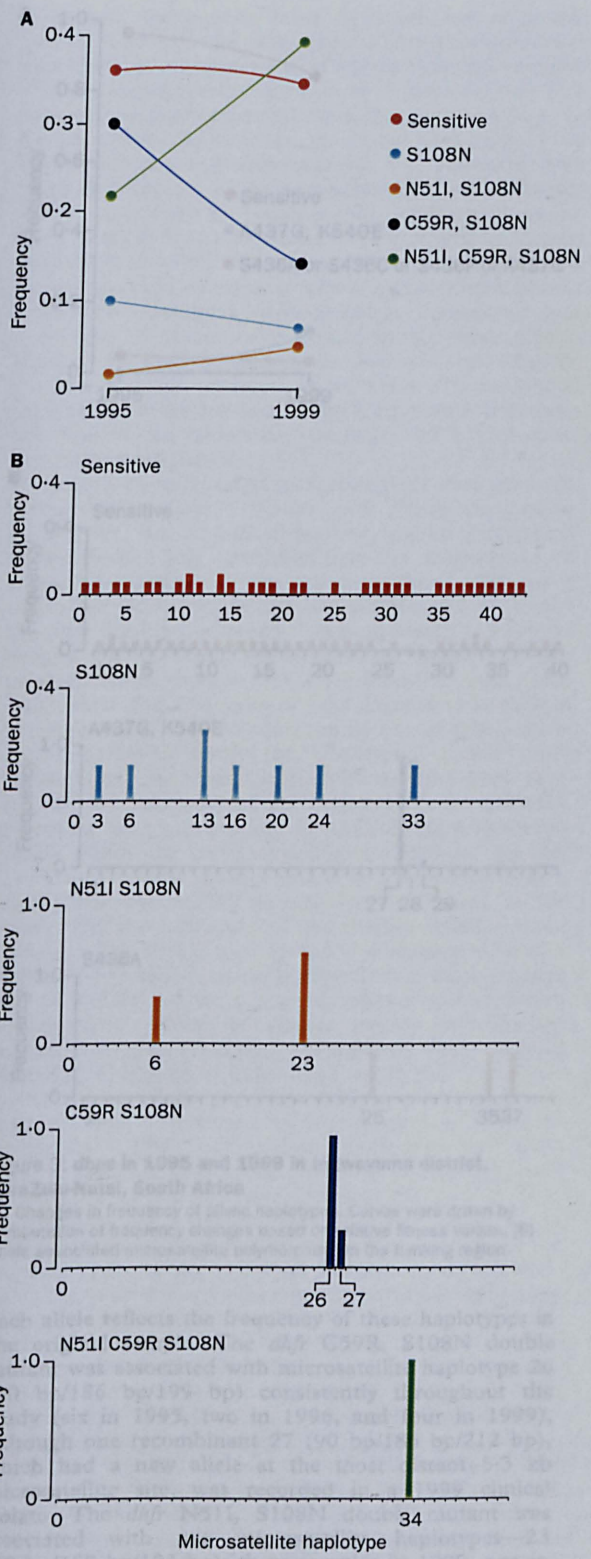


Figure 2: ***dhfr* in 1995 and 1999 in Ingwavuma district, KwaZulu-Natal, South Africa**

(A) Changes in frequency of allelic haplotypes. Curves were drawn by computation of frequency changes based on relative fitness values. (B) Allele associated microsatellite polymorphism in the flanking region.

parasites sampled from outpatients attending Mosvold hospital and its satellite clinic at Ndumu in 1996 (n=50) and in 1999 (n=73). Allele frequencies at the *dhfr* locus did not change significantly between 1996 and 1999 (p=0·07; figure 1), suggesting that pyrimethamine resistance had already been established. However, the frequency of the triple mutant *dhfr* N51I, C59R, S108N increased from 42% in 1996 to 62% in 1999 (p=0·04; figure 1). The frequency of the *dhps* double mutant A437G, K540E increased from 0% to 47% over this time (p<0·0001; figure 1).

To assess changes in *dhfr* and *dhps* throughout a wider population we retrospectively analysed blood samples collected by active surveillance surveys in the Ingwavuma district in 1995 and in 1999. The cases we detected were sampled over a wider region than in the previous part of the study and most people were asymptomatically infected. The frequency, measured from single and majority genotype infections in 1995 (n=50) and in 1999 (n=125) are shown in figure 2 for *dhfr* and figure 3 for *dhps*. Changes in the frequency of alleles at *dhfr* were non-significant (p=0·09; figure 2), the frequency of the *dhfr* sensitive haplotype remained roughly the same: 36% in 1995 and 34% in 1999, whereas the frequency of the *dhfr* triple mutant increased from 22% to 38%, apparently displacing the *dhfr* C59R, S108N double mutant, which fell from 30% to 15%.

Allele frequencies at *dhps* changed significantly between 1995 and 1999 (p=0·0006; figure 3) and the frequency of the double mutant rose from 0% in 1995 to 15% in 1999 (p=0·002), which displaced the sensitive haplotype, mirroring the changes seen in the hospital. We identified four single mutants of *dhps*—A437G, S436C, S436F, and S436A—but they were rare. S436A was identified in three isolates, whereas the others were only in one.

In 1999, the frequency of resistant alleles was significantly higher in the clinical isolates than in the community cases (*dhfr* p=0·002, *dhps* p<0·0001), but the frequency of the *dhfr* triple mutant did not differ between clinical and community cases in 1995–96 (p=0·052) and the frequency of the *dhps* double mutant was zero in each case. The *dhps* A437G, K540E and the *dhfr* N51I, C59R, S108N were also significantly higher in the clinical isolates than in the community cases. Such a trend would be expected if some patients had taken sulfadoxine-pyrimethamine before attending the hospital, but we excluded from the study any patients known to have received previous treatment. The 1999 clinical samples were more likely to be simply related, having been drawn from a localised epidemic.

We identified several polymorphisms at microsatellite loci around *dhfr*. The 0·3 kb locus had eight alleles sized 90–120 bp, the 4·4 kb locus 15 alleles of 157–200 bp, and the 5·3 kb locus 17 alleles of 193–226 bp. Of 102 isolates, 43 had different microsatellite haplotypes, which were ranked according to allele size at the 0·4 kb locus, then at the 4·4 kb, and finally the 5·2 kb locus, and numbered 1–43 (details of allele sizes for every haplotype are given in webappendix1; http://image.thelancet.com/extras/03art3014webappendix1.pdf). Figure 2 shows the characteristics of the haplotypes and how they are distributed between the *dhfr* alleles.

Associated with the *dhfr* triple mutant was microsatellite haplotype 34 (110 bp/177 bp/203 bp), which was identified in seven samples in 1995; one in 1996; and 35 in 1999, of which 22 were clinical and 13 actively detected. The number of replicate samples for

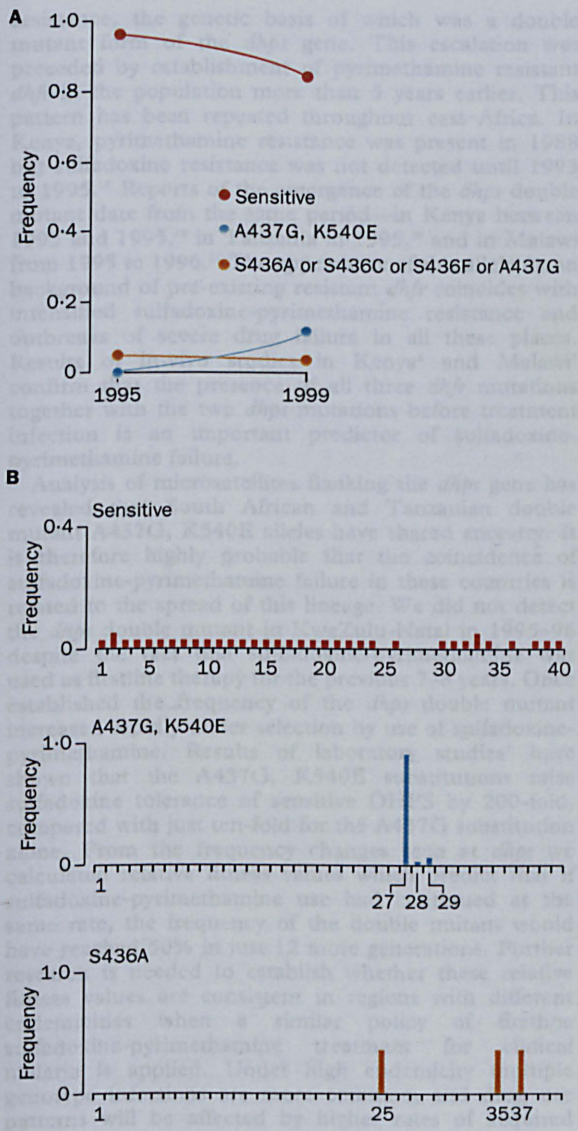


Figure 3: ***dhps* in 1995 and 1999 in Ingwavuma district, KwaZulu-Natal, South Africa**

(A) Changes in frequency of allelic haplotypes. Curves were drawn by computation of frequency changes based on relative fitness values. (B) Allele associated microsatellite polymorphism in the flanking region.

each allele reflects the frequency of these haplotypes in the original sample. The *dhfr* C59R, S108N double mutant was associated with microsatellite haplotype 26 (90 bp/186 bp/199 bp) consistently throughout the study (six in 1995, two in 1996, and four in 1999), although one recombinant 27 (90 bp/186 bp/212 bp), which had a new allele at the most distant 5·3 kb microsatellite site, was recorded in a 1999 clinical isolate. The *dhfr* N51I, S108N double mutant was associated with two microsatellite haplotypes—23 (90 bp/183 bp/193 bp): three samples in 1999, one in 1995) and 9 (90 bp/171 bp/217 bp): two samples in 1999. The single mutant S108N was associated with various haplotypes (four in 1995, three in 1999, and one clinical in 1999) as was the sensitive haplotype (eight in 1995, 11 in 1996, two in 1999, and 12 clinical in 1999).

Microsatellite polymorphisms flanking the *dhps* gene were characterised at sites 0·8 kb, 4·3 kb, and 7·7 kb from codon 437 in 70 isolates. There were 16 alleles of 104–142 bp at the 0·8 kb locus, seven alleles of 105–123 bp at the 4·3 kb locus, and 16 alleles of 110–142 bp at the 7·7 kb locus. Haplotypes were ranked according to allele size at 0·8 kb, then 4·3 kb, and finally 7·7 kb loci, and numbered 1–40 (full details of allele sizes for every haplotype are given in webappendix2; http://image.thelancet.com/extras/03art3014webappendix2.pdf. Figure 3 shows how these flanking haplotypes are distributed among the *dhps* alleles. As with *dhfr*, there was a high level of polymorphism around the sensitive alleles. The *dhps* double mutant was associated with microsatellite haplotype 27—consisting of 132 bp/105 bp/110 bp alleles at the 0·8 kb/4·3 kb/7·7 kb loci, suggesting a single origin. Of 29 double mutant A437G K540E isolates typed, three had recombined at the 7·7 kb locus, giving haplotypes of either 132 bp/105 bp/118 bp or 132 bp/105 bp/126 bp. By contrast with the *dhps* double mutant, the single mutant S436A alleles had unrelated flanking haplotypes 25 (128 bp/107 bp/132 bp), 35 (134 bp/105 bp/130 bp), and 37 (136 bp/105 bp/130 bp).

We compared the flanking sequences described above with that around alleles from malaria surveys in the Kilimanjaro region of northern Tanzania. Blood samples positive for *P falciparum* were obtained in 2001 in Hai, and in North and South Pare.[16] We characterised *dhfr* and *dhps* point mutations and found the same alleles as were present as in KwaZulu-Natal. We selected a subset of samples that were not mixed at any locus for microsatellite analysis. We could definitively characterise 86 samples that had the *dhfr* triple mutant haplotype and a single allele at each microsatellite locus—all of which had the identical 110 bp/177 bp/203 bp microsatellite haplotype as the triple mutants in KwaZulu-Natal. The *dhfr* N51I, S108N double mutant was assessed in ten single genotype infections. Each of them had the same 90 bp/183 bp/193 bp flanking microsatellite haplotype as the more common of the two variants characterised in the KwaZulu-Natal population. Likewise the *dhfr* C59R, S108N double mutant, characterised in just two Tanzanian isolates, had the same flanking haplotype as the equivalent KwaZulu-Natal mutant (90 bp/186 bp/199 bp). The pattern was repeated at *dhps*. Of 50 isolates with the A437G K540E double mutant allele, all had the characteristic 132 kb/105 kb/110 bp haplotype seen in the KwaZulu-Natal isolates.

At *dhfr* the relative fitness of the triple mutant N51I C59R S108N was standardised to 1·0 making the relative fitness of the sensitive allele 0·95 and the C59R, S108N double mutant 0·90. The frequencies of the N51I, S108N double mutant and the S108N single mutant were based on numbers fewer than ten at both time points and consequently measures of frequency change are prone to sampling error. We did not detect the *dhps* A437G K540E double mutant in 1995 or 1996. For fitness calculations an estimated initial frequency of 0·03 based on binomial statistical probability (with 95% confidence) was used. The fitness of the *dhps* sensitive allele was 0·866 relative to the double mutant A437G K540E.

Discussion

Our results suggest that the escalation of sulfadoxine-pyrimethamine resistance in KwaZulu-Natal during 1999/2000 coincided with the emergence of sulfadoxine

resistance, the genetic basis of which was a double mutant form of the *dhps* gene. This escalation was preceded by establishment of pyrimethamine resistant *dhfr* in the population more than 5 years earlier. This pattern has been repeated throughout east Africa. In Kenya, pyrimethamine resistance was present in 1988 but sulfadoxine resistance was not detected until 1993 to 1995.[18] Reports of the emergence of the *dhps* double mutant date from the same period—in Kenya between 1993 and 1995,[19] in Tanzania in 1995,[20] and in Malawi from 1995 to 1996.[21] The appearance of this allele on an background of pre-existing resistant *dhfr* coincides with intensified sulfadoxine-pyrimethamine resistance and outbreaks of severe drug failure in all these places. Results of in-vivo studies in Kenya[8] and Malawi[9] confirm that the presence of all three *dhfr* mutations together with the two *dhps* mutations before treatment infection is an important predictor of sulfadoxine-pyrimethamine failure.

Analysis of microsatellites flanking the *dhps* gene has revealed that South African and Tanzanian double mutant A437G, K540E alleles have shared ancestry. It is therefore highly probable that the coincidence of sulfadoxine-pyrimethamine failure in these countries is related to the spread of this lineage. We did not detect the *dhps* double mutant in KwaZulu-Natal in 1995–96 despite the fact that sulfadoxine-pyrimethamine was used as firstline therapy for the previous 7–8 years. Once established the frequency of the *dhps* double mutant increased rapidly under selection by use of sulfadoxine-pyrimethamine. Results of laboratory studies[7] have shown that the A437G, K540E substitutions raise sulfadoxine tolerance of sensitive DHPS by 200-fold, compared with just ten-fold for the A437G substitution alone. From the frequency changes seen at *dhps* we calculated relative fitness values which predict that if sulfadoxine-pyrimethamine use had continued at the same rate, the frequency of the double mutant would have reached 50% in just 12 more generations. Further research is needed to establish whether these relative fitness values are consistent in regions with different endemicities when a similar policy of firstline sulfadoxine-pyrimethamine treatment for clinical malaria is applied. Under high endemicity multiple genotype infections are more common, and drug use patterns will be affected by higher rates of acquired immunity. The potential confounders in such comparisons are factors not related to endemicity, such as drug availability and relative wealth.

Sensitive *dhfr* alleles had a high degree of polymorphism in their flanking sequences, reflecting the ancestral state of the wild-type population. By contrast, alleles with several resistance mutations showed flanking sequence variation that was highly conserved. The *dhfr* triple mutant was always associated with the same microsatellite haplotype, indicating that it has arisen only once, whereas the *dhfr* N51I, S108N double mutant was linked to two unrelated microsatellite haplotypes, suggesting that it has arisen twice, one of which was rare and found only in KwaZulu-Natal. The microsatellite haplotype flanking the *dhfr* C59R, S108N double mutant was different again. The distinctive flanking sequences of these double mutant alleles clearly show that none of them was ancestral to the triple mutant. Rather than arising many times, a few mutant alleles have been subject to positive drug selection, and been driven through the *P falciparum* population along with their flanking sequences. The fact that linked microsatellite alleles up to 5·3 kb away have been fixed in the mutant population testifies to the strength of that selection. Meanwhile the single mutant S108N was found on various flanking DNA backgrounds, suggesting either that the mutation has arisen on numerous occasions, or that the selection acting on it is too weak for the DNA sequences surrounding it to have become fixed, or both of these explanations combined.

The frequency of the *dhfr* triple mutant increased with time, apparently displacing a double mutant allele. This is consistent with in-vitro studies examining the relative levels of resistance conferred by these allelic haplotypes. The triple mutant is the most resistant to pyrimethamine, increasing the parasite's tolerance 1600-fold, whereas the S108N single mutation increases pyrimethamine tolerance 100-fold.[5] The double mutants C59R, S108N and N51I, S108N share broadly similar tolerance[22] between the single and triple mutants.[23]

Sequence analysis of *dhfr* and *dhps* in *P falciparum* sampled worldwide has shown that resistance alleles with multiple mutations predominate in regions of high drug use where sulfadoxine-pyrimethamine resistance problems are established.[2] Conversely they are rare, or absent, in places where the drug is used infrequently, showing that increasing resistance to sulfadoxine-pyrimethamine is caused by a progressive accumulation of mutations. As a consequence, it is widely expected that as sulfadoxine-pyrimethamine resistance intensifies the number of point mutations in resistance alleles will increase stepwise. We found no evidence of such a process in our analysis of *dhfr*. Although the C59R, S108N double mutant apparently preceded the triple mutant in KwaZulu-Natal and would theoretically be predicted to be replaced by it, it was not ancestral to it. Indeed none of the double mutant *dhfr* haplotypes gave rise to the triple mutant. This finding would suggest that new resistance mutations do not arise frequently and even constrain the rate at which the population can adapt to drug. Yet we did see rare single mutant resistance haplotypes in KwaZulu-Natal. And different versions of the *dhfr* S108N mutant or the *dhps* S436A mutant had unrelated flanking sequences, suggesting that they did arise independently, although it is possible that the selection acting on them is just too weak for the flanking sequences to have been conserved. Although single mutations seem to have arisen independently on multiple occasions, the fitness benefits they confer are perhaps too weak to overcome the associated fitness costs, and they do not seem to be strongly selected for. These alleles seem unlikely to spread throughout the population or to persist long enough to accumulate further mutations. On the rare occasions where a multiple mutant arises, it has high relative fitness and can spread rapidly throughout large geographic regions.

Chloroquine and sulfadoxine-pyrimethamine have been the cornerstone of antimalarial treatment in Africa for the past 40 years. With the efficacy of these drugs diminishing, experts warn that "a health calamity looms within the next few years".[11] For strategic planning of future resistance management the genetic analysis of flanking sequence around resistance genes is illuminating because it reveals the historical development of that resistance. A single allele of *pfcrt*, the chloroquine resistance determining locus, is found throughout Africa and flanking sequence has shown that it is a single lineage which first arose in southeast Asia.[24,12] We show here that the genetic determinants of sulfadoxine-pyrimethamine resistance in southeast Africa have also spread from a few ancestral mutants. Just one *dhps* and three *dhfr* alleles each of independent

and monophyletic origin are the basis of widespread resistance in the region. These findings are consistent with those of an analysis of the genetic determinants of sulfadoxine-pyrimethamine resistance in the South American Amazon.[25,26]

During the historical period when chloroquine, pyrimethamine, and sulfadoxine resistance was selected, routine use of these drugs for treatment has been surprisingly inefficient at establishing new resistance mutations. In the normal course of events in Africa, the proportion of parasites exposed to any particular drug is particularly low because a large proportion of infections are asymptomatic and consequently untreated. Pyrimethamine resistance emerged rapidly in both Tanzania and Kenya after mass administration of the drug on its own[27,28] the circumstantial evidence suggests that emergence of chloroquine resistance in southeast Asia and South America was associated with mass administration of that drug.[29] Therefore, drug coverage could be the most important driver of resistance evolution.

There are important lessons here for future management of resistance in Africa. Once established, drug resistance determinants to chloroquine, pyrimethamine, and sulfadoxine are highly mobile. Gene flow has played a key part in the emergence of resistance to these drugs in southeast African countries. A very important threat to new drugs or drug combinations introduced in these countries could be posed by importation of resistant or multidrug resistant genotypes that have been previously selected elsewhere. A priority for future planning for management of resistance is further research into the spatial extent of spread of the genetic determinants of resistance, and the potential of combination therapy for arresting or slowing this process.

Contributors

C Roper designed the study overall, wrote the report, had co-responsibility for the design and execution of the molecular genetic analysis, and analysed the data. R Pearce participated in the design and execution of the molecular genetic analysis, analysed the data, and reviewed the report. B Bredenkamp participated in the design of the longitudinal study in northern KwaZulu-Natal, the organisation of clinical studies at Mosvold Hospital during 1999, and the reviewing of the report. J Gumede participated in the community surveys done during 1999 in northern KwaZulu-Natal. C Drakeley participated in organisation of fieldwork, sample collection in the Pare mountains, and in reviewing the report. F Mosha organised the fieldwork and sample collection in Tanzania and reviewed the report. D Chandramohan organised the fieldwork and sample collection in the Hai district of northern Tanzania and reviewed the report. B Sharp participated in the design of the longitudinal study in northern KwaZulu-Natal, and as head of the South African MRC Malaria Lead Programme had co-responsibility for study design and co-ordination of field and laboratory research done in South Africa, and also reviewed the report.

Conflict of interest statement

None declared.

Acknowledgments

Cally Roper and Richard Pearce are supported by an Advanced Training Fellowship from The Wellcome Trust (ref 060714) awarded to C Roper. The Tanzania study was done under the auspices of the Joint Malaria Programme, a collaborative research initiative between The London School of Hygiene and Tropical Medicine, The Tanzanian National Institute for Medical Research, The Kilimanjaro Christian Medical College, and the Centre for Medical Parasitology, University of Copenhagen. Field studies in Hai were funded by a DfID grant (R7950) and those in the Pare mountains by a grant from MRC UK (G9901439).

We thank the staff of Mosvold hospital and Ndumu clinic, in particular Hervey Vaughan Williams, for their contribution; Chris Green of the Malaria programme of MRC Durban, who initiated the 1995 community study; David Conway and David Warhurst for advice and support; Paul Coleman and Ilona Carneiro for statistical advice; W Nkya and H Reyburn for logistical support; and E Nyale, F Laizer, and M Mosha for technical support.

References

1 Marsh K, Malaria disaster in Africa. *Lancet* 1998; **352**: 924.
2 Wang P, Lee C-S, Bayoumi R, 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* 1997; **89**: 161–77.
3 Sims P, Wang P, Hyde JE, On "The efficacy of antifolate antimalarial combinations in Africa". *Parasitology Today* 1998; **14**: 136–37.
4 Hastings IM, Watkins WM, White NJ. The evolution of drug-resistant malaria: the role of drug elimination half-life. *Philos Trans R Soc London Ser B* 2002; **357**: 505–19.
5 Wu Y, Kirkman LA, Wellems TE. Transformation of *Plasmodium falciparum* malaria parasites by homologous integration of plasmids that confer resistance to pyrimethamine. *Proc Natl Acad Sci USA* 1996; **93**: 1130–34.
6 Triglia T, Menting JGT, Wilson C, Cowman AF. Mutations in dihydropteroate synthase are responsible for sulfone and sulfonamide resistance in *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 1997; **94**: 13944–49.
7 Triglia T, Wang P, Sims PFG, Hyde JE, Cowman AF. Allelic exchange at the endogenous genomic locus in *Plasmodium falciparum* proves the role of dihydropteroate synthase in sulfadoxine-resistant malaria. *EMBO J* 1998; **17**: 3807–15.
8 Omar SA, Adagu IS, Warhurst DC, Can pretreatment screening for *dhps* and *dhfr* point mutations *in Plasmodium falciparum* infections be used to predict sulfadoxine-pyrimethamine treatment failure? *Trans R Soc Trop Med Hyg* 2001; **95**: 315–19.
9 Kublin JK, Dzinjalamala FK, Kamwendo DD, et al. Molecular markers for failure of sulfadoxine-pyrimethamine and chlorproguanil-dapsone treatment of *Plasmodium falciparum* malaria. *J Infect Dis* 2002; **185**: 380–88.
10 Paget-McNichol A, Saul A. Mutation rates in the dihydrofolate reductase gene of *Plasmodium falciparum*. *Parasitology* 2002; **1122**: 497–505.
11 White NJ, Nosten F, Looareesuwan S, et al.Averting a malaria disaster. *Lancet* 1999; **353**: 1965–67.
12 Wootton JC, Feng X, Ferdig MT, et al. Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* 2002; **417**: 320–23.
13 Bredenkamp BL, Sharp BL, Mthembu SD, Durrheim DN, Barnes KI. Failure of sulphadoxine-pyrimethamine in treating *Plasmodium falciparum* malaria in KwaZulu-Natal. *S Afr Med J* 2001; **91**: 970–72.
14 Pearce RJ, Drakeley C, Chandramohan D, Mosha F, Roper C. Molecular determination of point mutation haplotypes in the *dhfr* and *dhps* of *P falciparum* in three regions of northern Tanzania. *Antimicrob Agents Chemother* 2003; **47**: 1347–57.
15 Bahl A, Brunk B, Coppel RL, et al. PlasmoDB: the Plasmodium genome resource. An integrated database providing tools for accessing, analysing and mapping expression and sequence data (both finished and unfinished). *Nucleic Acids Res* 2002; **30**: 87–90.
16 Su X, Wellems TE. Toward a high resolution *Plasmodium falciparum* linkage map: polymorphic markers from hundreds of simple sequence repeats. *Genomics* 1996; **33**: 430–44.
17 Hartl DL, Clark AG. Principles of population genetics. 3rd edn. Sunderland, MA: Sinauer Associates, 1997: 430–44.
18 Mberu EK, Mosobo MK, Nzila AM, Kokwaro GO, Sibley CH, Watkins WM. The changing in vitro susceptibility pattern to pyrimethamine/sulfadoxine in *Plasmodium falciparum* field isolates from Kilifi Kenya. *Am J Trop Med Hyg* 2000; **62**: 396–401.
19 Nzila AM, Nduati E, Mberu EK, 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* 2000; **181**: 2023–28.
20 Curtis J, Duraisingh MT, Warhurst DC. In vivo selection for a specific genotype of dihydropteroate synthetase of *Plasmodium falciparum* by pyrimethamine-sufadoxine but not chlorproguanil-dapsone treatment. *J Infect Dis* 1998; **177**: 1429–33.
21 Plowe CV, Cortese JF, Djimde A, et al. Mutations in *Plasmodium falciparum* dihydrofolate reductase and dihydropteroate synthase and epidemiologic patterns of pyrimethamine sulfadoxine use and resistance. *J Infect Dis* 1997; **176**: 1590–96.
22 Cowman AF, Morry MJ, Biggs BA, Cross GA, Foote SJ. Amino acid changes linked to pyrimethamine resistance in the dihydrofolate

reductase thymidylate synthase gene of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 1998; **85:** 9109–13.

23 Peterson DS, Walliker D, Wellems TE. Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in *falciparum* malaria. *Proc Natl Acad Sci USA* 1988; **85:** 9114–18.

24 Su X, Kirkman LA, Fujioka H, Wellems TE. Complex polymorphisms in an approximately 330 kDa protein are linked to chloroquine resistant *P falciparum* in Southeast Asia and Africa. *Cell* 1997; **91:** 593–603.

25 Cortese JF, Caraballo A, Contreras CE, Plowe CV. Origin and dissemination of *Plasmodium falciparum* drug-resistance mutations in South America. *J Inf Dis* 2002; **186:** 999–1006.

26 Kublin JG, Witzig RS, Shankar AH. Molecular assays for surveillance of antifolate-resistant malaria. *Lancet* 1998; **351:** 1629–30.

27 Clyde D. *Malaria in Tanzania*. Oxford, UK: Oxford University Press, 1967.

28 Avery Jones S. Mass treatment with pyrimethamine: a study of resistance and cross resistance resulting from a field trial in the hyperendemic malarious area of Makueni, Kenya Sept, 1952–Sept, 1953. *Trans R Soc Trop Med Hyg* 1958; **52:** 547–61.

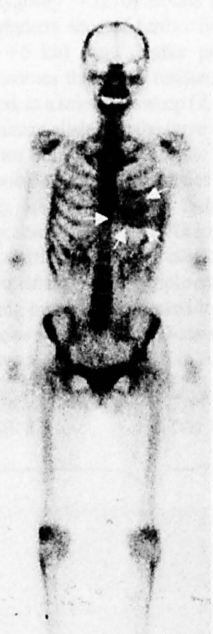
29 Payne D. Did medicated salt hasten the spread of chloroquine resistance in *Plasmodium falciparum*? *Parasitol Today* 1988; **4:** 112–15.

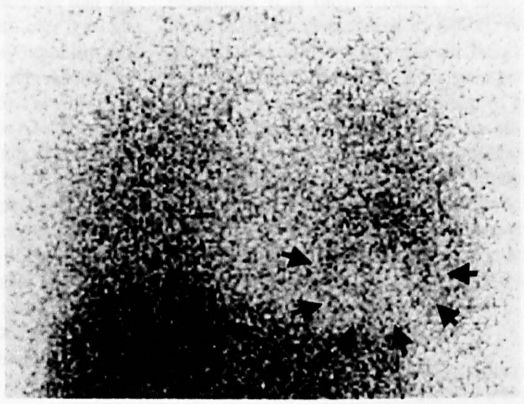
Clinical picture

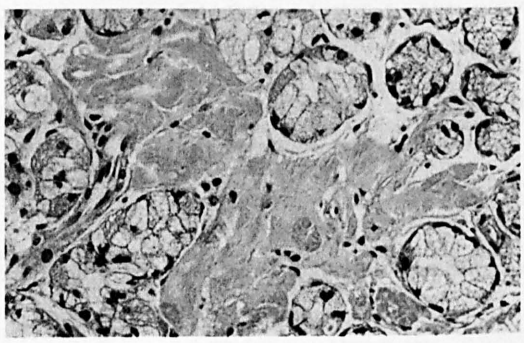
Cardiac denervation in amyloid polyneuropathy

Tomoaki Nakata, Kazuaki Shimamoto

A 43-year-old woman with abnormalities in bone scintigraphy performed several times following resection of right breast cancer presented with electrocardiographic abnormalities to our cardiology laboratory. Her history and results of physical examination disclosed neurological signs of peripheral sensory and motor disturbances with severe orthostatic hypotension and neurogenic bladder, which had started 6 months after the operation, but no symptoms or signs suggestive of myocarditis or an acute ischaemic event was documented. Massive ^{99m}Tc-MDP accumulation in the heart (figure, left) strongly suggested cardiac amyloidosis. An echocardiogram revealed cardiac hypertrophy but no wall motion abnormality; left ventricular wall thickness was 14 mm and left ventricular ejection fraction was 57%. There was no abnormality in myocardial perfusion imaging. Left ventricular diastolic function, however, was greatly depressed and cardiac sympathetic nerve imaging revealed a profound loss of cardiac uptake of ^{123}I-metaiodobenzylguanidine (figure, upper right); the heart-to-mediastinum ratio of 1·62. Her heart rate variation was also markedly reduced. Biopsy specimens from the salivary gland and intestine revealed depositions of amyloid proteins (figure, lower right). These findings indicated that systemic amyloid deposition is related not only to polyneuropathy but also to cardiac denervation. The patient currently has no serious complications or cardiac symptoms.







Graduate School of Medicine, Cardiovascular and Renal Medicine, Sapporo Medical University School of Medicine, Sapporo, Japan
(T Nakata MD, K Shimamoto MD)

Intercontinental Spread of Pyrimethamine-Resistant Malaria

Cally Roper,[1]* Richard Pearce,[1] Shalini Nair,[2] Brian Sharp,[3] François Nosten,[4] Tim Anderson[2]

Here we present molecular evidence demonstrating that malaria parasites bearing high-level pyrimethamine resistance originally arrived in Africa from southeast Asia. Chloroquine (CQ) is being replaced by sulfadoxine pyrimethamine (SP) for treatment of *Plasmodium falciparum* malaria in Africa. Mutations in the dihydrofolate reductase (*dhfr*) gene of *P. falciparum* underlie resistance to pyrimethamine. *Dhfr* alleles with one (108N) or two (108N plus 51I or 108N plus 59R) mutations result in increased parasite clearance times. Infections bearing triple-mutant *dhfr* (108N, 51I, and 59R) have high treatment failure rates, and quadruple-mutant *dhfr* alleles (108N, 51I, 59R, and 164L) render parasites untreatable (*1*). Parasites with one to three mutations occur in Africa, whereas all four mutations are common in southeast Asia (*2*) and South America (*3*). Triple-mutant alleles are replacing other alleles in Africa (*4*).

It is generally assumed that pyrimethamine resistance has evolved multiple times (*1*), because it is selectable in the laboratory, has a simple genetic basis, and appears rapidly after SP introduction. However, analysis of microsatellites that flank *dhfr* in African parasites sampled from sites 4000 km apart demonstrate just three independent origins of double mutants and a single origin of the triple mutant (*4*). Similarly, *dhfr* alleles with two to four mutations have a single evolutionary origin across five southeast Asian countries (*2*).

Genotyping of eight microsatellite markers flanking *dhfr* in both southeast Asian and African parasites shows that the triple-mutant *dhfr* allele in Africa shares a common origin with *dhfr* alleles bearing two to four mutations in southeast Asia (Fig. 1). The predominant five-locus microsatellite haplotype (–10 kb to +0.5 kb) associated with triple-mutant *dhfr* in Africa is identical to that associated with *dhfr* alleles carrying two to four mutations in southeast Asia. In contrast, these five loci show high levels of variation (mean expected heterozygosity = 0.76) around sensitive *dhfr* alleles. Markers situated further from *dhfr* (–20 kb and >+6 kb) show higher polymorphism on chromosomes that carry resistant *dhfr* alleles, as expected, in a selective sweep (*2*, *5*). However, the predominant alleles are the same on resistant chromosomes from both continents. In contrast, African double-mutant *dhfr* alleles have dissimilar flanking alleles, indicating independent origins (*4*). Because all southeast Asian *dhfr* alleles carrying >1 mutation have a single origin, the simplest explanation is that triple-mutant *dhfr* alleles spreading in Africa originated in southeast Asia.

Alleles at the major CQ-resistance locus *pfcrt* also have a common origin in African and Asian parasites (*5*, *6*). CQ-resistant *pfcrt* alleles and triple-mutant *dhfr* alleles may have arrived in Africa in the same parasite genome. Pyrimethamine resistance was widespread in Asia when CQ resistance was first recorded in Africa (*7*). Import of southeast Asian parasites has thus led to the demise of the two affordable drugs that have been the mainstay of malaria treatment in Africa.

Why did the triple-mutant allele not arise independently in Africa? Assuming a mutation rate of 10^{-9} per base per generation, we would expect 10 to 1000 independent origins of triple-mutant parasites in every infection (10^{10-12} parasites) containing double-mutant *dhfr* alleles. The implication is that complex compensatory mutations are required to restore parasite fitness.

Every year 30,000 malaria cases are imported into industrialized countries (*8*). The numbers of cases imported into Africa is unknown but likely to be substantial. Given that 67% of parasites sampled in Thailand, Cambodia, and Myanmar carry the 164L mutation in *dhfr* (*2*), as well as high levels of mefloquine and quinine resistance, it is only a matter of time before these invade and establish in Africa. We suggest that careful thought should be given to preventing further import of resistant parasites, perhaps by screening and treatment of passengers traveling from southeast Asia or South America to Africa. Widespread introduction of artemisinin-based combination therapy (*9*) could also help to minimize the foci from which resistant parasites can spread. Importantly, these data demonstrate that antimalarial drug resistance is an international problem requiring a coordinated international response.

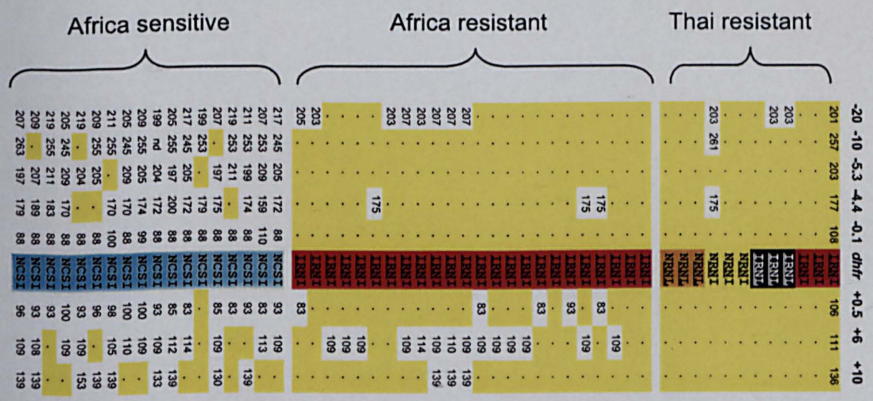


Fig. 1. *dhfr* alleles and flanking microsatellites of parasites from Africa and Thailand. The figure comprises data from 12 Thai parasites with two to four resistance mutations, 24 African parasites with triple-mutant alleles, and 18 African parasites with sensitive *dhfr* alleles. The four-letter codes describe amino acids present at positions 51, 59, 108, and 164 in the predicted *dhfr* protein (*10*). Amino acids conferring resistance are underlined, and *dhfr* alleles are shaded yellow, orange, red, and black in order of increasing resistance. Sensitive alleles are shaded turquoise. Allele lengths are shown for eight microsatellites positioned at –0.1, –4.4, –5.3, –10, and –20 kb upstream and +0.5, +6, and +10 kb downstream of *dhfr*. Dots and yellow shading indicate identical allele size to the predominant resistant haplotype (shown at right).

References and Notes

1. C. V. Plowe *et al.*, *J. Infect. Dis.* **176**, 1590 (1997).
2. S. Nair *et al.*, *Mol. Biol. Evol.* **20**, 1526 (2003).
3. J. F. Cortese, A. Caraballo, C. E. Contreras, C. V. Plowe, *J. Infect. Dis.* **186**, 999 (2002).
4. C. Roper *et al.*, *Lancet* **361**, 1174 (2003).
5. J. C. Wootton *et al.*, *Nature* **418**, 320 (2002).
6. D. A. Fidock *et al.*, *Mol. Cell* **6**, 861 (2000).
7. N. J. White, *J. Antimicrob. Chemother.* **30**, 571 (1992).
8. E. T. Ryan, K. C. Kain, *N. Engl. J. Med.* **342**, 1716 (2000).
9. N. White, *Philos. Trans. R. Soc. London Ser. B* **354**, 739 (1999).
10. Single-letter abbreviations for the amino acid residues are as follows: C, Cys; I, Ile; L, Leu; N, Asn; R, Arg; S, Ser.
11. Supported by NIH and the Wellcome Trust.

Supporting Online Material

www.sciencemag.org/cgi/content/full/305/5687/1124/DC1
Materials and Methods
References and Notes

7 April 2004; accepted 22 June 2004

[1]London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. [2]Southwest Foundation for Biomedical Research, San Antonio, TX, USA. [3]Medical Research Council, Durban, South Africa. [4]Shoklo Malaria Research Unit, Mae Sot, Tak, Thailand.

*To whom correspondence should be addressed. E-mail: cally.roper@lshtm.ac.uk