ASSOCIATING GENETIC RESISTANCE TO *Plasmodium falciparum* MALARIA INFECTION WITH ETHNIC GROUPS RESIDENTS OF MALARIA ENDEMIC AND NON-ENDEMIC REGIONS OF KENYA

GEORGE AYODO (BSc, MSc Human Genetics)
184/0031/03

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR DEGREE OF DOCTOR OF PHILOSOPHY IN GENETICS OF INFECTIOUS DISEASES IN THE SCHOOL OF HEALTH SCIENCE OF KENYATTA UNIVERSITY

April, 2009

DECLARATION

s thesis is my original work and has not been submitted for a degree to any other university or for any other award.

Signature Ayodo. Date 4/5/09

George Ayodo
184/0031/03

We confirm that the report in this thesis was carried out by the candidate under our supervision.

Signature Date 05/05/2009

Dr. Micheal F Otieno
Department of Pre-clinical Sciences, Kenyatta University and

Signature Date May 5, 2009.

Prof.Alloys SS Orago
Department of Public Health, Kenyatta University and
National AIDS Control Council, Kenya

Signature David Reich Date 28/04/2009

Prof. David Reich
Harvard Medical School, Harvard University USA.

DEDICATION

This research document is dedicated to my wife; Penelope Ayodo, son; David Ayodo, father; Alfred Ayodo and mother; Alice Ayodo. Their concern with my progress kept me alert both day and night.

ACKNOWLEDGEMENTS

I wish to thank all my supervisors for their support during the period of this study. I am particularly grateful to Prof. David Reich (Harvard Medical School, USA) for the mentorship in all aspects of medical research work. I am also thankful to Prof. Alloys SS Orago and Dr. Micheal F.Otieno for healthy discussions and comments throughout this work. It would be indeed a serious mistake or typo-error if I failed to thank Drs. Alkes Price, Alon Kenoin, and Nick Patternson for statistical discussions and for putting extra effort to train me on different components of computational data management. Lastly, I am deeply indebted to Dr.Alicja, Gavin and Christine, who trained me in many aspects of bench work including genotyping process, data curations and setting up quality control measures. I can not exit without thanking Misters Shem Ochieng and David Samoea in the Pre-clinical Department, Kenyatta University, and the entire community members of Broad Institute of Harvard and MIT, USA.

TABLE OF CONTENTS

LIST OF APPENDICES

APPENDIX A: DATA SETS FOR CASE-CONTROL ASSOCIATION STIDIES AND DETAILS OF PREVIOUSLY PUBLISHED ASSOCIATIONS OF 10 GENES SELECTED FOR THIS STUDY.

APPENDIX B: DATA SETS FOR POPULATION DIFERENTIATION TESTS.

APPENDIX C: REVIEW OF PREVIOUS ASSOCIATIONS OF OTHER GENETIC VARIANTS INCLUDING THE 10 TESTED IN THIS STUDY

APPENDIX D: THE LIFE CYCLE OF MALARIA PARASITE

APPENDIX E: RESEARCH INSTRUMENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS AND ABBREVIATIONS

CI	Confidence Interval
CS	Circumsporozoite Protein
DARC	Duffy Antigen / Chemokine Receptor
DNA	Deoxyribonucleic Acid
EHH	Extended Haplotype Homozygozity
FCγR	Fc gamma receptor
G6PD	Glucose 6 Phosphate Dehydrogenase
GTP	Guanosine Triphosphate
HbAS	Sickle cell trait
HLA	Human Laucocyte Antigen
ICAM	Intercellular Adhesion Molecule
IFNAR1	Interferon –Alpha Receptor-1
IFN-α	Interferon alpha
IFN-γ	Interferon Beta
Ig	Immunoglobin
IL-1	Interleukin - 1
INOS	Inducible Nitric Oxide Synthase
LD	Linkage disequilibrium
MAEBL	Membrane Apical Erthyrocyte Binding Surface Protein

MAL/TRAP	
MB	Mega-Base
MHC	Major Histocompatibility Complex
NO	Nitric Oxide
OR	Odd /ratio
PF	*Plasmodium falciparum*
PfEMP-1	*Plamodium falciparum* Membrane Protein - 1
PRBCs	Parasitized Red Blood Cells
PVM	Parasitophorus Vacuole
RBC	Red Blood Cells
RSP-1	Ring Surface Protein -1
SMA	Severe Malarial Anaemia
SNPs	Single Nucleotide Polymorphisms
SP	Sulphadoxine and Pyrimethamine
TRD	Thrombospondin Related Anonymous Protein
VCAM-1	Vascular Adhesion Molecule - 1
VSA	Vascular Surface Antigen

DEFINITION OF TERMS

Haplotype: A set of closely linked alleles (genes or DNA polymorphisms) that are inherited as a unit.

Polymorphism: A variation in the sequence of genetic information on a segment of DNA

Single Nucleotide Polymorphism (SNP): A variation in the genetic code at a specific point on the DNA sequence.

Locus (pl. loci): The position of a gene on a chromosome or other chromosome marker; also, the DNA at that position.

Allele: Alternative forms of a gene; a single allele from each locus is inherited independently from each parent.

ABSTRACT

Malaria causes death of millions of people in sub-Saharan Africa and about 80% are children and women under 5 years of age. Infection has therefore exerted pressure on human genome and as a consequence clinical manifestations appear variable in endemic and non-endemic populations. Part of the reason for this epidemiological difference is hypothesized that over the last few thousand years, endemic populations have built up genetic resistance to severe malaria infection. To test this hypothesis, the study searched for evidence of natural selection in malaria exposed and unexposed populations by (a) carrying out a large-scale collection in Kenya of severe malaria cases and controls from the Luo ethnic group and also of population controls from the Masai and Kikuyu ethnic groups, (b) carrying out an association study at 10 genetic variants previously associated with malaria resistance, (b) studying frequency differences across populations to determine which of these variants have been subject to selection for malaria resistance in the past few thousand years, and (c) also studying haplotype and linkage disequilibrium patterns around malaria resistance genes to search for evidence of natural selection. In the Luo case-control samples, the previously described associations at CD36-GT (P value < 0.004) and HbAS (P value = 0.015) were replicated. Strikingly, there was unusually high frequency differentiation of the HbAS and CD36-GT variants in the exposed (Luo and Yoruba) vs. relatively unexposed (Kikuyu and Masai) populations compared to a panel of 1,454 randomly chosen markers that were studied in the same samples (P < 0.00036 and 0.00096 respectively). By statistically combining the case-control association and frequency differentiation statistics, the power of the association analysis was increased by several orders of magnitude (HbAS with P value < 0.000018 and CD36-GT with P value < 0.00043), which provides a potential tool for researchers to find risk factors for infectious disease in future. Further assessment of haplotype blocks flanking HbAS-T, CD36-G and ICAM-T suggested that exposed and un-exposed populations exhibit different haplotype block patterns, supporting the evidence of natural selection. CD 36-GT appears to be under selection in both the Luo and Yoruba ethnic group, whereas HbAS is under selection in the Yoruba ethnic only but not in Luo ethnic groups. These results suggest Yoruba and Luo—perhaps because they are on different sides of the African continent—evolved different genetic response to malaria because they had been exposed to the disease for thousand years. This study has not only developed a novel method to identify malaria variants but has also provided an insight on the possibility of exploiting haplotype block patterns to map causal genes.

CHAPTER ONE

INTRODUCTION

1.1 Background Information

Malaria parasite infection is an ancient disease that leads to enormous human mortality and morbidity (Miller *et al.*, 2002). Today more than half the world population is at risk of infection, about 3.2 billion people and there are between 350 million and 500 million clinical cases reported each year resulting in 1 million deaths (Guerra et al., 2006; Anonymous Authors., 2007). About 80% who die are children under the age of 5 years and pregnant women (Anonymous Authors 2007)

Over the last 10,000 years, malaria has exerted strong selection pressure on the human genome (Bamshad *et al.*, 2003). Because of this pressure, the clinical manifestations appear variable in the endemic and non endemic populations suggesting the existence of a distinct acquired and innate immunity (Clarke *et al.*, 2004). Indeed, it has been established that despite high infection rates in endemic areas, only 1-2% of patients develop life threatening complications and profound anaemia, much less than the rate in non-endemic areas (Kwiatkowski., 2005). These observations suggest that there are genetic variants that modulate risk to *P. falciparum* malaria, similar to what has been observed with the Duffy antigen that protects against *P. vivax* (Miller et al., 1976). The genetics of the host therefore play an important role in the protection and susceptibility to malaria (Mackinnon *et al.*, 2005). Several host genetic variants have been identified that influence the clinical outcome of malaria and these include Apha globin, Beta globin, , Duffy Antigen Chemokine receptor (Hill *et al.*,2006) and G6PD (Moorman *et al.*, 2003)

The mechanism of protection of these variants is still unknown and there is evidence that they act in synergy (Hill, 2006; Goldstein, 2008). Most of the variants identified so far exhibit moderate effect ranging from odds ratios of 0.75 to 1.5, and genes associated with acquired immunity are often but not always involved (Petretto *et al.,* 2007). Despite the progress in malaria research, there is still lack of detailed understanding of the molecular and biochemical properties involved in the aetiology of malaria.

Of the genetic variations that exist in our species, 85-90% have been ascribed to differences among individuals in any given group and only 10-15% are due to differences between members of different groups (Goldstein & Hirschhorn., 2004). It has also been noted that functional variants can differ remarkably in their allele frequencies across populations (Goldstein & Hirschhorn., 2004). It is this observation is behind the idea to study populations that have been differentially exposed to malaria infection. The study hypothesized that differential exposure to malaria hundreds or even thousands of years could have resulted in an increased proportion of malaria resistance variants in endemic populations. Several such variants have indeed been identified in endemic populations by case-control association analysis. Part of the present study is a replication case-control association analysis, in which the association of 10 genetic variants to malaria in the endemically Luo ethnic group was tested. Comparisons are made between allele frequencies in children with severe malaria and uninfected adults (cases and controls) from the Luo ethnic group. Furthermore, the frequency differentiation of these variants in three ethnic groups: Yoruba, Maasai and Kikuyu were compared. The significance is to replicate previous associations and also to explore a novel method that improves the power to identify risk alleles to malaria.

The study further looked at the pattern of Linkage Disequilibrium (LD) around HbAS, ICAM and CD36, which are all malaria candidate genes. The hypothesis behind the investigation is that differential exposure to malaria infections has influenced the haplotype pattern around genes that are associated with resistance to malaria infection. Consequently, understanding disturbances to the haplotype pattern in exposed populations (which presumably reflect the influence of natural selection) can provide insight into the causal genes. Previously, it has been shown that parasitic infections such as malaria and schistosomaisis influence the LD pattern in the 5q31 region of the human genome but no study has looked at the haplotype block patterns around malaria candidate regions (Goldstein & Weale., 2001; Luoni *et al*., 2005; Rihet *et al*., 1998; Marquet *et al*., 1996). This component of the study is important because it provides basis for consider changing understand haplotype-related associations to infections and the demographic history associated infections. This study gives an insight to variants that are either biologically causal or just statistically correlated with malaria infection.

1.2 STATEMENT OF THE PROBLEM

Genetic variants are important for the strategic development of drug and vaccine because they govern parasite invasion, remodeling, growth, reinvasion of erythrocyte and also influence the efficacy of potential drugs and vaccines. Case-control association study approach has been used to identify variants but the challenges are: a) Inconsistent and false positive results and limited replications have been carried out for validation purposes (Goldstein & Hirschhorn., 2004). Different SNPs from a single gene have been associated with both protection and susceptibility (Kwiatkowski., 2005). In addition, lack of clear phenotype definitions of malaria reduces the case control tests because cases and controls are poorly defined in case-control study populations (Kwiatkowski., 2005). There has been a discussion on improving power of case control tests by statistical weights of whole genome scan but there has been no empirical demonstration (Roeder et al., 2006).

Several studies have been undertaken on African populations but stratification and admixture of populations has not been tested. Other studies however show the significance of testing for stratification in order to avoid false positives and inconsistent associations (Hirschhorn et al., 2002). There is therefore a need to understand population structure of any study population.

With the limited power of case control tests, there is a need to improve or develop a strategy to identify more malaria variants and possibly the causal genes. Limited work has been done to search for search other genes by defining haplotype blocks around malaria variants of endemic and non-endemic populations. LD has been used in non-malaria studies to identify flanking haplotype block with variants associated with diseases or infections. No study has attempted to compare haplotype blocks in differentially malaria exposed populations and also assessed and compared recombination break down of haplotype that carry malaria variants.

. 1.3 RESEARCH QUESTIONS

1.3.1 Is it possible to replicate the associations of previously association test in Luo ethnic group?

1.3.2 Can combination of case control and allele frequency statistics increase power of association tests?

1.3.3 Is there population stratification among the four 4 ethnic groups and do the linguistic and genetic cluster match?

1.3.4 What is the haplotype pattern around malaria variants? Are malaria variants under selection? .

1.4 NULL HYPOTHESES

1.4.1 The genetic effect is not the same in all the populations and the direction of the association is not similar in all the populations.

1.4.2 Combining case-control associations and frequency differentiation statistics does not increase the power of association test.

1.4.3 There is stratification or admixture of population among ethnic groups in Kenya

1.4.4. LD pattern of malaria associated genes is the same in both the exposed and unexposed populations, and variants associated with malaria are under selection with the same demographic information..

1.5 OBJECTIVES OF THE STUDY

1. 5.1 General objectives

The study replicate previous malaria associations, explores method for increasing power of association test, assesses the LD pattern around malaria candidates' genes and also , attempt to discern haplotype block pattern, and lastly compare the extent and effect of natural selection in malaria exposed populations.

1.5.2 SPECIFIC OBJECTIVES

a) Test the association of 10 malaria variants in Luo ethnic group that had been previously associated with malaria susceptibility or resistance.

b) Increase power of case control association tests by combining case-control and allele differentiation chi-square statistics.

c) Determine the genetic structure of ethnic groups in Kenya and find out if it can alter or confound association and allele frequency differentiation tests.

d) Define haplotype blocks of selected malaria variants and assess signal of natural selection

1.6 JUSTIFICATION

Identification of malaria variants in human that interact with the parasite is critical for development of drug and vaccine. (Kwiatkowski., 2005). Malaria variant with absolute protection against *P. falciparum* can be used to develop a vaccine, just like Null Duffy antigen that provides absolute protection against *vivax* malaria is presently being used to develop a vaccine against *vivax* malaria. (Saravia *et al*., 2008). The chellenge to this strategy is that the variants that have so far been associated with *P. falciparum* show moderate genetic effects of protection (Ayodo et al., 2007) with several inconsistent association results. This study therefore replicates previous association by using a different study design and also improves the power of case control association tests by combining statistical weights of case control association and allele frequency differentiation tests. The approach empirically demonstrates recent discussion by Roeder et al., 2006 that proposed increasing power of association tests by statistical weights of whole genome scan and opens an avenue for identification of more malaria variants..

In addition, it has been observed that stratification and admixed population leads to false positive in case control association studies. The study not only determines the population structure but also tests for genomic allele frequency differences between the closely related populations due to malaria exposure. Two approaches therefore validate case control and allele frequency differentiation tests.

Furthermore, allele frequency differentiation in endemic and non-endemic population provides a platform to use haplotypes block instead of single SNPs because using haplotype improves the chance of identifying causal genes, and other malaria variants and it is also cost-effective. The studies of natural selection are not only interesting in their own right. It demonstrates that by combining this frequency differentiation information statistically with association tests, elusive variants can be identified despite overlapping malaria phenotypes.

1.7 CONCEPTUAL FRAMEWORK

As outlined in Table 1.1, the design is to replicate previous association studies and explore how to identify other unknown genes associated with malaria infection. The concept is based on the fact that populations that have been exposed to malaria exhibit more malaria resistance genes than unexposed populations. Similarly, young cases of severe malaria are expected to exhibit under-representation of resistance genes. Conceptually, there is some similarity of this work with some previous workers, who used information from linkage genome scans to improve the power of the association (Roeder *et al.*, 2006). In the same way, information from selection studies is used to improve the power of association.

Table 1.1: The Conceptual framework for this study

Parameters	**Cases (exposed)**	**Controls (exposed)**
Parasitaemia	High	Low or absent
Haemoglobin concentration	Low or high	Normal
Severity of malaria infection	Symptomatic	Asymptomatic / normal
Rate of transmission	High	High

1.8 OPERATIONAL FRAMEWORK

The study involved three phases namely; Sample collection of DNA, genotyping and data management. The DNA samples were collected from schools and Bondo District hospital over a period of one and half years. After conducting quality control in Kenya, all DNA samples were shipped to Harvard Medical School, USA for the second genotyping . This second phase took 7 months at the Laboratory of David Reich, Harvard Medical School and Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), USA. The result for replication of case control association and combining case control and allele frequency differentiation statistical weights have already been published in American Journal of Human Genetics. Further work involved the assessment of haplotype blocks and detection of natural selection. The Figure 1.1 illustrates all the details of the study.

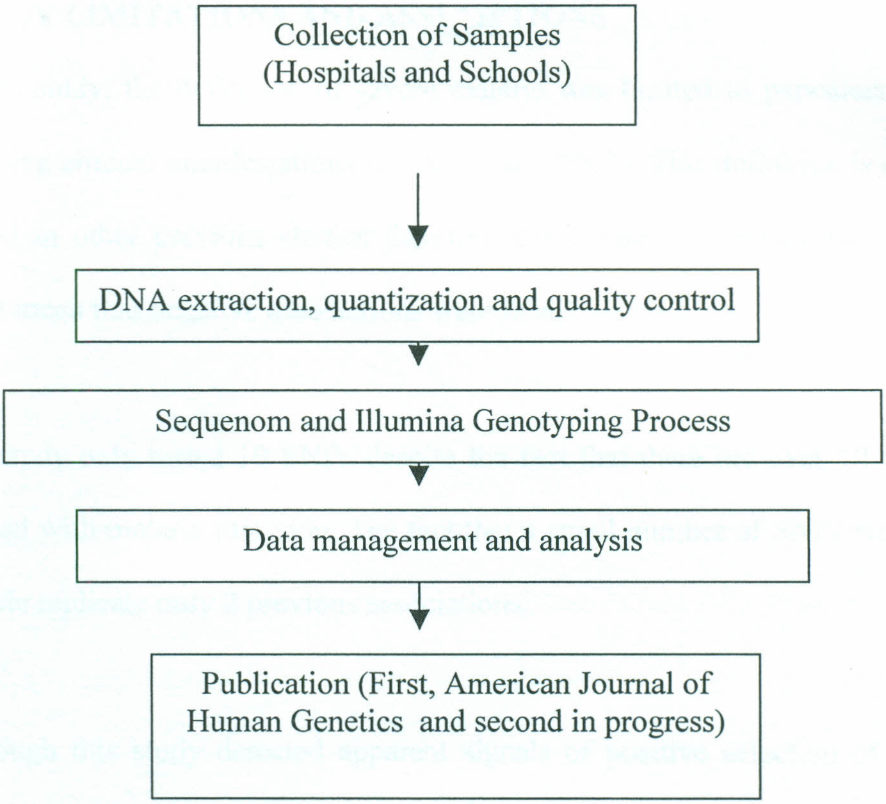


Figure 1.1: The illustration of the operational framework

1.9 STUDY LIMITATIONS AND ASSUMPTIONS

a) In this study, the definition of severe malaria was limited to parasitaemia and other overlapping clinical manifestations (Anstay *et al.*, 2007). This definition is different from that used in other previous studies therefore the failure to replicate these associations does not mean that negative associations were false.

b) The study only tested 10 SNPs despite the fact that there are over 20 SNPs that are associated with malaria infection. The fact that a small number of SNPs were used made it possible replicate only 2 previous associations.

c) Although this study detected apparent signals of positive selection of HbAS-T and CD36-G, this study lacked genomic control to assess whether signals were unique. Thus, the conclusion that these variants indicate strong signals of selection is tentative pending further confirmation.

d) The study failed to detect a signal of selection based on LD patterns around HbAS-T in the Luo ethnic group. This could be because the mutation was very old and was exposed to long period of LD decay due to recombination. The failure to detect this selection also suggests that either the density of SNPs used in this study is too low or there was an under-sampling of genetic variation in the study populations.

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview

The Malaria parasites *Plasmodium falciparum, P. vivax, P. malariae* and *P. ovale*, are responsible for about 500 million new malaria infections and 1.2 million deaths annually. (Greenwood *et al*., 1999). Malaria in humans caused by *P. falciparum* is the focus of the current study. With the emergence of epidemics of multi-drug resistant malaria parasites, understanding how human biology influences the pathogenesis of malaria infection can contribute to the design of new therapeutic strategies to treat the disease, and identify potential new targets for drug therapy.

The clinical manifestation of malaria infection is unclear. Most infected people in endemic areas remain asymptomatic and very few people who are infected progresses to severity status of the disease. Infected persons present different overlapping clinical symptoms and there are no clear phenotypes associated with severe or mild malaria. The clinical manifestations such as hypoglycemia, respiratory distress, circulatory collapse, fluid and electrolyte imbalance, renal failure, aneamia etc often appear as different combination and in most cases they overlap without showing specific patterns (Hill., 2006) and Mackinnon *et al*., 2006 suggested that genetic variants contribute to 25% of host resistance. Despite this findings, several studies have concluded that resistance to malaria is highly polygenic and there exist variable resistance genes in populations, for example the restriction of HbC in West African populations but not East Africa (Hill., 2006)

The importance of genetic variability to malaria susceptibility is underscored by the fact that there are many variations among humans in their inherent susceptibility to the infection (Tishkoff *et al.,* 2001; Sabeti *et al.* 2002a). Indeed, natural selection for resistance to malaria resistance may explain why only a small fraction of malaria infections in Africa lead to life threatening complications such as cerebral malaria or profound anemia (Kwiatkowski *et al.,* 2000; Sagel and Hill, 2003).

Inorder to dentify genes that have been implicated so far in resistance to severe malaria, scientists have focused on genes that *a priori* affect biological pathways relevant to malaria (Hill et al., 1997). These are polymorphisms in blood cell related genes, or in immune response and receptor associated genes such as Tumor Necrosis Factor (*TNF*), Nitric Oxide Synthase (*NOS*), Intercellular Adhesion Molecule (*ICAM-1*), CD40 ligand, and CD36 (McGuire *et al.,* 1994, 1998; Deitsch and Wellems, 1996; Burger *et al.,* 1998: Knight *et al.,* 1999; Wattavidanage *et al.,* 1999; Kun *et al.,* 1999, 2001). However, the full complement of variants governing susceptibility has not yet been identified (Kidson *et al.,* 1981; Hill *et al.,* 1991, Ruwende *et al.,* 1995). The resistance genes identified have fallen into 5 categories: blocking sporozoite and merozoite invasion into liver cells and erythrocytes, b) inhibiting parasite development in the hepatocytes, c) activating macrophages to phagocytose intra-erythrocytic parasites and free merozoites, d) neutralizing parasite glycosylphosphatidylinositol (and other key molecules) and e) Inhibiting the inflammatory cytokine cascade and mediating complement-dependent lysis of extracellular parasite gametes and development of zygotes.

Figure 2.1: Global distribution of malaria infection (Adapted from website (WHO., 2000)

🟪Areas where malaria infections are common

🟨Areas with limited risk

🟩No malaria

2.2 Geographic distribution of malaria. Malaria is endemic to the tropics and subtropics. In 1990, about 80% of the infections occurred in Africa, while the remaining ones were observed in Brazil, India, Afghanistan, Sri-Lanka, Thailand, Indonesia, Vietnam, Cambodia and China [Figure 2.1] (WHO, 2000). Over the past three decades, the incidence in Africa in particular has increased 2-3 fold. This upsurge—reversing the trend that would be expected due to modern treatment as a result of drug-resistant parasites, insecticide-resistant mosquitoes, increasing poverty and political instability

(Hartl, 2004). In the past 13 years, malaria epidemics have also spread to previously non-endemic areas, raising the prevalence in Africa from 20% to 60% (Hartl, 2004).

The present study focuses on the genetics of malaria in East Africa and specifically in Kenya. Although annually there are an estimated 8.2 million cases of malaria in Kenya (Githeko and Ndegwa, 2001), minimal research has been done in East Africa than in West Africa to identify genes that are relevant to malaria resistance, making it important to study the genetics of malaria in this region. Kenya presents a favorable environment in which to study the genetics of malaria infection because of its great geographic and ethnic diversity..

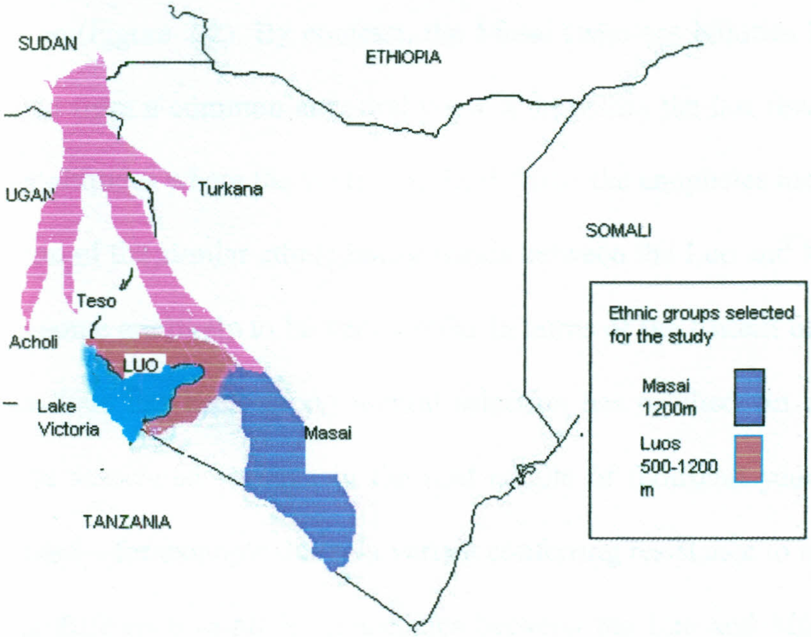


Figure 2.2: Migration history of the Luo and Masai ethnic groups in Kenya. [Information curated from Ochieng., 1985 and illustrated in this thesis]

As illustrated in Figure 2.2 the pink colour shows the likely path of migration of the Nilotic ancestors of both populations from Sudan and southern Ethiopia, beginning several thousand years ago and continuing until the 17th century. At present, the Masai tend to live at higher altitudes than the Luo (>1,000 meters) where malaria is much less common

A striking feature of the Kenyan population—which can in principle be used to increase the power of this study—is that communities of common ethnic ancestry live in places where the prevalence of malaria has different. Specifically, The Luo continuously has lived for thousands of years in low-altitude areas near Lake Victoria where malaria is endemic (Figure 2.2). By contrast, the Masai (who are Niloties like the Luo and likely descend from a common ancestral population within the last few thousand years) live at higher altitudes where the vector for the disease, the anopheles mosquito, is rare.

Because of the similar ethnic backgrounds between the Luo and Maasai, most sections of the genome are likely to be very similar in terms of the pattern of genetic variation. Only where there has been strong natural selection has resulted in changes in frequency of malaria associated variants in the past couple of thousand years since the populations separated—for example due to a variant conferring resistance to malaria—would expect a strong difference in allele frequencies between the Luo and Masai. It may therefore be possible to find new susceptibility genes to malaria by searching for regions of high genetic differentiation between Luo and Masai.

2.3 Published genetic associations to severe malaria.

It is clear that replication in genetic studies is important for the validation of associations However, numerous published associations to common, complex genetic phenotypes have not been replicated in follow-up studies, which may be explained by genetic heterogeneity among populations (Hirschhorn *et al.* 2002). It is increasingly being thought that false-positives explain a substantial proportion of the associations published in the literature (Ioannidis *et al.* 2001; Hirschhorn *et al.* 2002; Lohmueller *et al.* 2003). For instance, a recent large replication study involving 16 variants previously published as conferring susceptibility to Type II Diabetes, only found replication at one variant PPARγ (Altshuler *et al.* 2000). Replication is thus essential in demonstrating biological and medical relevance. Below is a description of 9 genes containing variants that have been previously associated with malaria that will be tested in the current study (See Table 2.1 and Table 1 in appendix D for the samples sizes and details of allele frequency details)

Table 2.1 Previous association studies for susceptibility to severe *P. falciparum* malaria

Genotype	Population	P-value	Odds-ratio (OR)	95% CI	References
HbCC	Burkina Faso	0.0011	0.07	0 – 0.48	Agrawal et al.,2000
HbAC	Burkina Faso	0.0008	0.71	0.58 – 0.87	Agrawal et al.,2000
HbAS	Burkina Faso	0.01	0.27	0.17 – 0.42	*Modiano et al., 2001*
HbAS	Kenya	0.0004	0.04	0.02 - 0.37	*Aidoo et al., 2002*
G6PD-A females	Gambia	0.02	0.45	0.22 – 0.42	*Ruwende et al., 1995*
G6PD-A females	Kenya	0.11	0.6	0.33-1.11	*Ruwende et al., 1995*
G6PD-A hemizygous males	Gambia	0.02	0.23	0.06 – 0.77	*Ruwende et al., 1995*
G6PD-A hemizygous males	Kenya	0.12	0.54	0.25 – 1.15	*Ruwende et al., 1995*
NOS2A-117 C→T homozygote (TT)	Tanzania	0.0006	0.12	0.03 – 0.048	*Hobbs et al., 2002*
NOS2A-117 C→T homozygote (TT)	Kenya	0.012	0.31	0.12-0.84	*Hobbs et al., 2002*
NOS2A-117 C→T heterozygote (CT)	Kenya	0.005	0.25	0.09 – 0.66	*Kun et al., 1999*
NOS2A 1659T heterozygote	Gambia	0.04	1.31	1.0 – 1.72	*Burgner et al., 2003*
NOS2A -969(G→C)	Gabon	0.04	0.67	0.46-0.96	*Kun et al., 2001*
ICAM1 kilifi heterozygote	Gabon	0.012	0.52	0.30 – 0.92	*Fernandez et al.,1998*
ICAM1 kilifi homozygote	Gabon	0.027	0.18	0.04 – 0.86	*Kun et al., 1999*
CD32 Arg/Arg 131	Kenya	0.0021	0.28	0.12 – 0.63	*Shi et al.,, 1999*
CD32 Arg/Arg 131	Gambia	0.03	0.71	0.52 – 0.98	*Cooke et al., 2003*
CD36 1264T →G Homozygotes	Thailand	0.0069	0.59	0.40 – 0.87	*Aitman et al., 2000*
CD36 T188G heterozygote	Kenya	0.036	0.74	0.55 – 0.99	*Omi et al., 2003*
CD40L-726C female carriers	Gambia	0.002	0.74	0.48-1.15	*Sabeti et al., 2002*
CD40L-726C Male hemizygotes	Gambia	P<0.05	0.52	0.34-0.81	*Sabeti et al., 2002*
*TNF*308A Homozygote	Gambia	0.039	3.6	1.1-12.4	*McGuire et al., 1994*
*TNF*308A Homozygote	Kenya	0.16	3.3	0.62-17.6	*Knight et al., 1999*
TNF-238 Heterozygote	Gambia	0.04	1.7	1.02-2.8	*Knight et al., 1999*
TNF-376A heterozygote	Gambia	0.008	4.3	1.5-12.8	*McGuire et al., 1994*
*TNF*α*,*2 Heterozygote	Sri-Lanka	0.013	2.76	*	*Wattavidanage et al 1999*
IFNAR1-17470G/G homozygotes	Gambia	0.018	0.74	0.57-0.95	*Aucan et al., 2003*
IFNAR1-LI68V Homozygote	Gambia	0.031	0.76	0.59-0.97	*Aucan et al., 2003*

* Data could not be found in the published article

i) *HbCC* (Thalassemia) and *HbAS* (Sickle-cell trait): Both malaria-resistance mutations occur in the 6th position in the β-globin amino acid sequence, where glutamic acid is substituted by either lysine or valine respectively. The heterozygotes and homozygotes for the β-globin thalassemia allele (*HbAC* and *HbCC*) are estimated to have 29% and 93% reduced risk to severe and cerebral malaria respectively (Agarwal *et al.,* 2000;

Modiano *et al.*, 2001). The reduction in risk due to the very common *HbAS* genotype is about 73% (there is no data for homozygotes as they usually die at a young age due to sickle cell anaemia) (Bloland *et al.,*1999; Modiano *et al.,* 2001a, 2001b; Aidoo *et al.,* 2002: Hobbs *et al.,* 2002).

ii) *G6PD:* This enzyme has been detected in virtually all cells and plays a crucial role of catalyzing the first and rate-limiting step of the hexose monophosphate pathway, pathway for pentose phosphate synthesis (Vulliamy *et al.,* 1991). Mutations in the X-linked *G6PD* gene cause low levels of expression of the gene, resulting in anemia in response to certain food and drug exposures in over 400 million people worldwide (Beutler, 1990). The *G6PD-A* variant, which is responsible for the deficiency, has also been shown to confer a survival advantage to carriers, providing protection against severe malaria in about 46% in heterozygous females and 58% hemizygous males (Ruwende *et al.,* 1995) .

iii) *ICAM-1*: Is a host receptor, which has been shown in histopathologic studies to bind to parasite ligands, in particular *P. falciparum* erythrocyte membrane protein-1 (PfEMP1). The *ICAM-1 kilifi* variant shows protective effects in Gabon and Kenya as shown in Table 2.1 (Fernandez-Rayes *et al.,* 1997; Kun *et al.,* 1999).

iv) *CD32*: Is in the family of Fc receptors (Fcy RIIa) that are expressed on the surface of lymphocytes and monocytes / macrophages. The molecule provides a link between the humoral and cellular immune systems. It has been shown that the recognition of CD32 is influenced by polymorphism within the gene FcYRIIa (Shi *et al.,* 2001). Associations have been described in Gambia and Kenya (Shi *et al.,* 2001; Cooke *et al.,* 2003) (Table 2.1).

v) *CD36*: The human CD36 ligand is expressed on *P. falciparum*-infected red blood cells (Ockenhouse *et al.,* 1991; Baruch *et al.,* 1999). It is thought that CD36 interacts with malaria pathology by sequestering infected red blood cells and inhibiting the immune response to the parasite (Urban *et al.,* 1999). Interestingly, the CD36 variants shown in Table 2.1 are strongly associated with a reduction in susceptibility to severe cerebral malaria in Thailand and Kenya (Aitman *et a*l., 2000); Pain *et al.,* 2001; Omi *et al.,* 2003).

vi) *CD40L*: This is a type II membrane glycoprotein expressed by activated CD4+ T cells. CD40L is involved in B cell proliferation, antigen presenting cell activation, and Ig class switching (Durie *et al.,* 1994a; 1994b). The CD40L variant has been associated with a significant reduction in risk for severe malaria with odds ratio: 0.52, $p<0.002$ (Sabeti *et al.,* 2002b) (Table 2.1). Strong evidence of natural selection around the CD40 ligand gene has been demonstrated by this group in west African (Nigerian) as well as in Bantu (Zimbabwean) populations (Sabeti *et al.*, 2002b), presumably reflecting a history of conferring resistance against severe malaria.

vii) *TNF*: This plays a critical role in immune response to infectious and inflammatory diseases. Homozygotes for the TNF–308A variants have been associated with susceptibility to severe malaria in Kenya and Gabon, whereas the heterozygotes have been shown to have increased susceptibility (McGuire *et al.,* 1994, 1998; Knight *et al.*1999). Comaparison in of the populations, show that heterozygotes in Sri Lankan population were less susceptibility to malaria than the African populations (Wattavidanage *et al.,* 1999) [See Table 2.1]. Attempts to understand *TNF* haplotype structure in Gambian and Malawian populations have shown that the SNPs most strongly associated with susceptibility differ between these two African populations (Ackerman *et al.,* 2003), emphasizing the importance of defining the haplotype structure in each study population separately. The strategy for this study is to create a haplotype map for *ICAM*, Bata globin and *CD36* genes specifically for the Luo, Masai, Kikuyu and Yoruba.

viii) *NOS2A*: This gene is a mediator of immunity. The effects of each variant genotype associated with malaria infection are shown in Table 2.1 (Kun *et al.,* 1999; Hobbs *et al.,* 2002; Burger *et al.,* 2003). The heterozygous variant G-954C in Gabon may be as protective as sickle cell trait (Kun *et al.,* 2001).

ix) *IFN–γ*: This gene is a key element for innate and adaptive immunity and has been implicated in a wide range of infectious and inflammatory disease processes, especially during a clinically important episode of malaria (Kwiatkowski, 2000). A variant *IFNGR-*56 at the promoter region appears to often protect against malaria infection among Gambian children (Koch at el, 2002). The other variants found to be associated with malaria are the Interferon–alpha receptor–1 (IFNARI), 17470 and LI68V variants (Aucan *et al.,* 2003' Hill, 2003). The effects of the variants are reported in Table 2.1

The Duffy antigen / Chemokine receptor (DARC) has also been associated with resistance to malaria. An allele that disrupts a GATA-1 binding site in the DARC promoter region specifically inhibits the expression of the Duffy antigen on erythrocytes but not other cells (Miller *et al.,* 1976; Tournamille *et al.*, 1995). The absence of an expressed Duffy antigen on erythrocytes seems to confer complete resistance to *P. vivax* malaria (the frequency of this mutation is nearly 100% in West Africa, but essentially 0% outside Africa, presumably because of selection for this protective variant). Because the focus of this study is on *P. falciparum* malaria, the one responsible for highest level of morbidity and mortality in African populations, the replication of Duffy mutation was excluded in this study.

2.4 Parasite's life cycle

The parasite exists in three to four different forms and each form is specialized to survive in or for survival within in a certain environments. The gametocyte is the form that infects the mosquito and reproduces itself, as if it a hemophodite. When the mosquito has sucked blood containing gametocytes, these pass into the salivary glands of the mosquito, where they develop into a new form, the sporozoite. The sporozoite can be passed on to man when the mosquito bites, injecting its saliva into blood vessels. The sporozoite travels with the blood to the liver and enters the liver cells. In the liver some of the sporozoites divide (tachysporozoites) and become thousands of merozoites. The merozoites are released from the liver to the blood where they are taken up by red blood cells. Some of these turn into ring-formed trophozoites, which split again to form schizonts? Schizonts burst the red blood cells at a certain stage, releasing the merozoites. This release coincides with the rapid increase in body temperature following Malaria infection.. The trophozoites that are left over during division can, in the course of the next day, develop into the sexual form, the gametocyte, which can be taken up by any blood-sucking mosquito to start another cycle. The incubation period is about 10 to 15 days. This period can be much longer depending on medication. *Plasmodium ovale* and *Plasmodium vivax* can also produce a dormant form, a hypnozoite, which can cause relapses of the disease for several months or years after the original disease (relapsing malaria), because it remains dormant in the liver cells (See Appendix D, Figure 1).

2.5 Genetic understanding of malaria parasite.

Plasmodium falciparum has a complex life cycle because it invades both mosquitoes and vertebrate host cells. Target cells include mosquito salivary glands, red cells and liver cells. The invasion process in each cell is still unclear; nonetheless, other proteins such as cirumsporozoite (CS), thrombospondin–related anonymous protein (TRAP) and membrane apical erythrocyte binding like surface protein (MAEBL) have been associated with the invasion. These are therefore target proteins for biomedical research to developing immunoprophylactic and theurapeutic methods. The broad genetic variation of the *Plasmodium falciparum* populations however is a challenge to understanding how the parasite overcomes chemotherapeutic agents, vaccines and vector control strategies. The parasite (*P. falciparum*) demonstrate rich diversity (η =1.16 x 10^{-}3) (Volkman *et al.,* 2006) and comparative genomics of *P. falciparum* and *P. reichenowi* shows that adaptive evolution has occurred in the *P falciparum* at the level of transcript, gene function, protein expression and cellular localization. (Jeffares *et al.,* 2007). It has also been observed that non-paralog genes contain immunogene and other known drug targets consistent with the idea that human immune systems and drug use contributed to the parasite's evolution (Kidgell *et al.,* 2006). Others have also pointed at the alteration of target enzyme by point mutations resulting in an unusual complement of mismatch of repair genes contributing to the drug resistance (Kidgell *et al.,* 2006). Similarly, gene amplification of GTP cyclohydrolase enzyme in folate biosynthesis pathway has been associated with the resistance to anti-foliate drugs (; Bethke *et al.,* 2006)

2.6 Pathogenesis of malaria.

Malaria stands out among the systematic infectious diseases of humans. The fact that *P falciparum* is easy to locate within the blood vessels or is fixed onto the blood vessels means that there is a plausible explanation that malaria pathogenesis is due to poor oxygen delivery because of obstructed blood vessels. Nonetheless, focusing on malaria in isolation has caused a great deal of controversy as far as its pathogenesis is concerned. Evidence suggests that cell to cell interactions are responsible for the pathogenesis but clinical manifestations are protean. Adherence and invasion appear to generally correlate with severity (Pasvol *et al.,* 2001). The interactions occur mostly at the asexual reproductive stage of parasite. Some of the well defined interactions include areas as follows:

2.6.1 The invasion of red cells

The severity of malarial relates is thought to be correlated with parasite density which may determine the ability of the parasite to invade and multiply within the red blood cells (Chotivanich *et al.,* 2000, Mackinon.,2005). *P. falciparum* from severe malaria patients possess high *in vitro* multiplication rate, which are approximately three times higher than those from uncomplicated cases. They could invade the cells without any restriction in their invasion.

2.6.2 What is the mechanism?

Invasion is a highly specific, ordered and sequential process that lasts about 30 seconds. Merozoites attach to a susceptible red blood cell, re-orient itself so that the apical end is apposed to red cell membrane and then slowly moves into a localized invagination of the red cell which subsequently envelops it as the Parasitophorous Vacuolar Membrane (PVM). The red cell membrane however poses a formidable barrier to invasion. Currently, some intensive research is on the identification of the molecules in the red cell surface that merozoites bind to.

.

2.6.2.1 The binding of parasitized red cells (PRBC) to uninfected red cells (rosetting)

Rosetting has been associated with severe malaria but the mechanism under which it leads to disease is still obscure. A study in The Gambia showed that all isolates of *P. falciparum* obtained from children with cerebral malaria were capable of rosetting, whereas many isolates from patients with mild disease were not (Carlson *et al.*,1990). One other studies have implicated variants of PfEMP-1 and CD36 with the rosetting (Handunnetti *et al.*,1992). This kind of interaction is thought to enhance invasion and parasite multiplication especially under conditions of blood flow but this has never been proved through experiment (Pasvol.,2001).

2.6.2.2 The binding of PRBC to endothelial cells in critical organs (cytoadherence)

Cytoadherence is the process in which the mature PRBCs bind to endothelial cells in post-capillary venules, this explains why PRBCs are rarely seen in peripheral samples. Upon scanning using Electron Microscope (EM), several regular and symmetrically arranged “knobs” appear on the surface of the infected cell as *P. falciparum* matures. These knobs are thought to be the sites at which the parasitized red cell attaches to endothelial cells in deep tissues; they are also capable of cyto-adherence *in vitro*. Several molecules such as *PfEMP-1*, ring surface proteins (*RSP 1* and *2*) have been identified in the immature PRBCs. *PfEMP-1* variants are of particular interest because the amino terminal head structure of the molecule – which includes a Duffy-like binding domain and a cysteine-rich inter-domain region – appear to mediate adherence to a diverse set of host receptors (Barry *et al.,* 2007) Host - Parasite interactions led to immune evasion due to differential expression of var genes of *PfEMP-1* (Rottmann *et al.,* 2006; Barry *et al.,* 2007). Nonetheless, analyses of cellular interactions have revealed marked heterogeneity in molecular specificity that suggests complexity of pathogenesis. First, by immobilizing the parasites in various organs, cytoadherence prevents their passage through the spleen, a major site of parasite destruction. Localizing maturing parasites at the sites of reduced oxygen tension favours parasite growth and may facilitate invasion of uninfected red blood cells. Secondly, though is still unclear cytoadtherant parasites presumably lead to microvasculature obstruction. Lastly, cytoadherence may lead to local endothelial cell activation, release of cytokines and subsequent damage to adjacent tissues.

2.6.2.3 Induction of pro-inflammatory cytokines.

There is considerable evidence supporting the notion that malaria infection is primarily an inflammatory cytokine driven disease (Clark *et al*., 2006; 2007). The broad clinical manifestations of malaria infection have been associated with the balance between pro-inflammatory and anti-inflammatory cytokines (Clark *et al*., 2007). Furthermore, the clinical effect of low or high level of cytokine production is similar in many bacteria, rickettsia and viral infections (Clark *et al*., 2006). Recent studies involving different populations in West Africa (The Gambia), East Africa (Kenya) and the Far East (Vietnam) associated a SNP polymorphism (IFNARI C-G) with susceptibility to severe malaria. The combined effect was an odd ratio of 1.38, 95% confidence Interval of 1.17-1.64 and a P value of 0.00017 (Khor *et al*., 2007). Attempts are underway to use anti-inflammatory synthetic drugs to block the production of TNF, which has been associated with severe malaria (Specht *et al*., 2008)

2.7 Natural selection

Human adaptation occurs at two levels: species wide genetics adaptation and population adaptation (Tishkoff *et al*., 2002). Genetic adaptation involves physiological, biochemical and behavioral adjustments. To acquire evidence substantiating evidence of natural selection, one needs to understand demographic events in human history such as population expansion, bottlenecks, and population structure, all of which have left a unique signature across patterns of genetic variation in the genome. So far some of the strongest evidence of selection and genetic adaptation in Africans involves the polymorphism at G6PD (Sabeti *et al*., 2002a; Tishkoff *et al*., 2002)), Beta globin (Kwiatkowski,. 2000) and at the duffy blood group (Miller., 1976) most of which are maintained by selective pressure from *P. falciparum* and *P.vivax* malaria.. Furthermore, advances in comparative genomics, population genetics, and molecular evolutionary theory have provided data and methodology that can be used to test hypotheses of natural selection and adaptation. The greater difficulty in testing for natural selection is isolation of signal of selection due to malaria infection and demographic history.. Nonetheless, the availability of large datasets has made it possible to make inferences about human demographic history.

The DNA sequences that differ among individuals are known as polymorphisms and the nucleotide that differ among species are often referred to as fixed differences; differences among individuals are known as SNPs. Tests at the SNP level are powerful because there are large-scale datasets that can be used to empirically and robustly test departures from neutrality and infer population parameters. The neutral model states that evolutionary changes at molecular level are a consequence of genetic drift and not adaptive evolution (Kimura *et al* 1983). Selection under neutral model therefore contributes little to the observed variation within or between species (Kimura *et al* 1983). A current area of high interest is examining the selection pattern of candidate's genes across populations in the entire genome. The putative genes under selection can then be further studied from functional perspectives.

The study of selection is based on null hypothesis of neutrality. Here are some of the theoretical predictions under neutral model about the relationship between the rate of mutation and evolutionary parameters (Kimura and Ahta., 1978). $H = 4 Nu / (1+4 Nu)$ where H stands for expected heteozygosity, N is the effective population size, u is the permutation rate, (Kimura and Ahta., 1978). The set of assumptions for this test are genetic drift explains evolutionary change and the population is panmictic devoid of sub-populations, and that the population size has remained constant for a long time. The neutrality test therefore tests in effect the null-hypothesis of neutrality-equilibrium.

Natural selection is expected to after the frequencies in the genetic variant and therein alter individual survival probabilities. Favored mutations arise with the linked markers; thus selection extends beyond the immediate selected regions.

2.8 Evidence for natural selection at genes associated with malaria resistance.

Selection for resistance to malaria is one of the few areas where the impact of selection on genetic variation has been well-documented. The high gene frequency of haemoglobin S (sickle-cell trait) in Africa, has been shown to be protective against severe malaria in heterozygous individuals. However, the frequency of the variant is moderated by the fact that homozygotes have fatal sickle-cell anaemia (Kwiatkowski., 2000). The geographic distribution of Hemoglobin S variant shows a clear correlation with the prevalence of malaria (Flint *et al.*, 1986), confirming further the importance of natural selection on variant.

Evidence for natural selection for malaria resistance in Africa has also been documented at two other genes. Tishkoff *et al.*, 2002 and Verreli *et al.,* 2002 showed evidence for selection at the *G6PD-A* allele. Sabeti *et al.*, 2000b also showed that malaria infection has left a powerful signature of selection on the core haplotypes containing the *G6PD-A* and *TNFS5* alleles. Each gene appeared to be just a few thousand years old based on the flanking pattern of linkage disequilibrium, but is much more frequent than expected in the absence of selection.

CHAPTER THREE

MATERIALS AND METHODS

3.1: Overview

This chapter explains the study design by describing study areas, populations and the selection criteria, approach used to conduct case-control associations test, genotyping methods and data management tools.

3.2: Study design

The designed of this study is based on the fact that malaria infection has exerted pressure on the human genome and therefore there is differential selection in the endemic and non-endemic populations. The study has therefore considered two exposed and other two with limited exposed malaria. In addition, the design includes unmatched age between cases and controls. In usual approach case-control association studies, the age of cases and controls are matched. The average age of the cases was 2.6 years (Table 3.1), reflecting a focus on individuals with no previous immunological protection against malaria. For the controls, the study focused on older controls because had survived to an older age; thus, they are likely to have a higher frequency of variants protecting against severe malaria, making it easier to detect associations.

3.3: Study areas

3.3.1: Areas for the collection for case-control association tests samples.

All the malaria samples were collected from Bondo District Hospital in the Western part of Kenya. The district is at the shore of Lake Victoria at an altitude of ~1240 m above sea level (Figure 3.1) .All severe malaria cases were collected from the Bondo District hospital's children's emergency ward or from its out-patient clinic between May 2004 and August 2005. The control samples however were collected from secondary schools namely: Barkowino, Barchando, Chianda, Mituri, Nyagoma and Nyamira within Bondo District.

3.3.2: The study areas for collection of selection study samples.

The samples were collected in two regions in Kenya namely:-

3.3.2.1: Nyeri District

The samples were collected from Nyeri District in Central Kenya (Figure 3.1). The district is found at an altitude of ~ 1950 m above sea level and is non-malaria endemic area. All the samples were collected from the secondary schools such Nyeri Mixed Complex, St' Mary Boys, Kangubiri girls in the in Central Kenya

3.3.2.2: Narok District

The samples were collected from Narok District in the Rift Valley province of Kenya (Figure 3.1). The district is found at an altitude of ~ 1880m above sea level. The samples were also collected from the secondary schools such as Narok Boys, Masai girls and St Mary's girls.

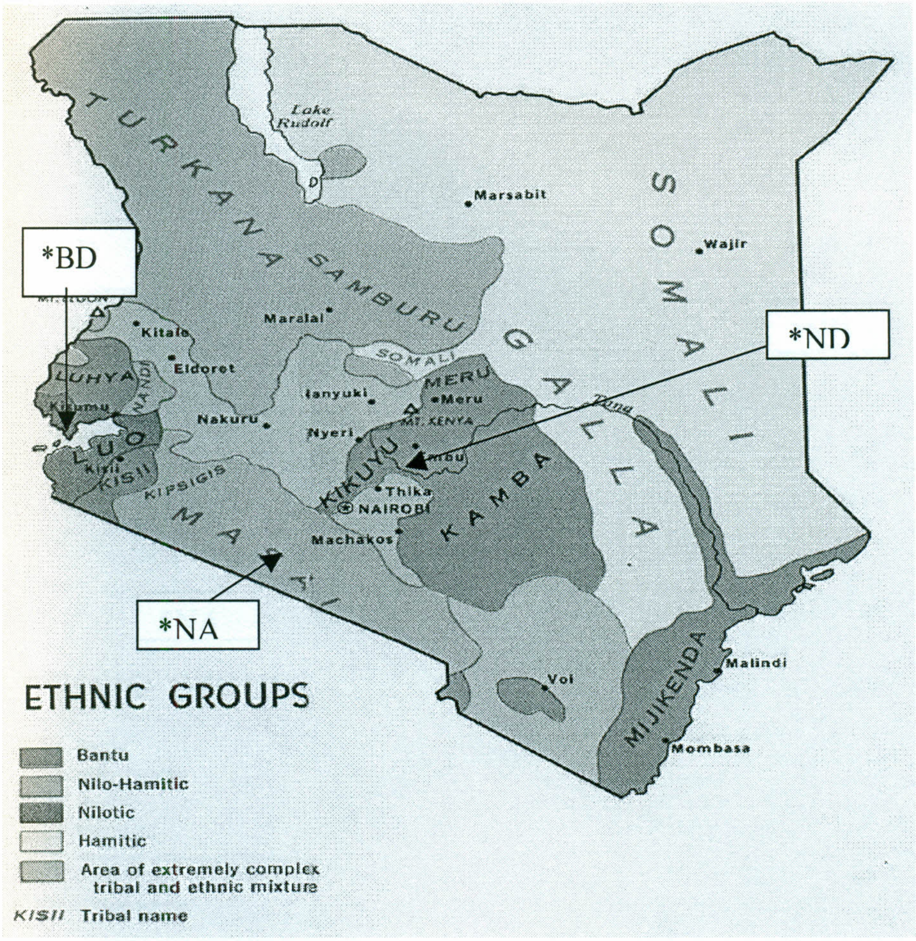


Figure 3.1: Shows the map of Kenya describing distribution and ethnic profiles of 3 ethnic used in this study [*Note: NAD ~ Narok District; ND ~ Nyeri District; BD ~ Bondo District]* [Adapted and revised from www.cia/gov/library., 1974]

3.3.2.3: International Haplotype Map project

The samples for Yoruba ethnic group were obtained from the International Haplotype Map (HapMap) project at their USA centre at Broad Institute of MIT and Harvard.

3.4: Study populations

The study involved four ethnic groups namely Luo, Masai, Kikuyu and Yoruba. The samples were collected as shown in Table 3.1. The Luo ethnic group speaks a Nilotic language and lives in a malaria-endemic region in western Kenya. Controls for case-

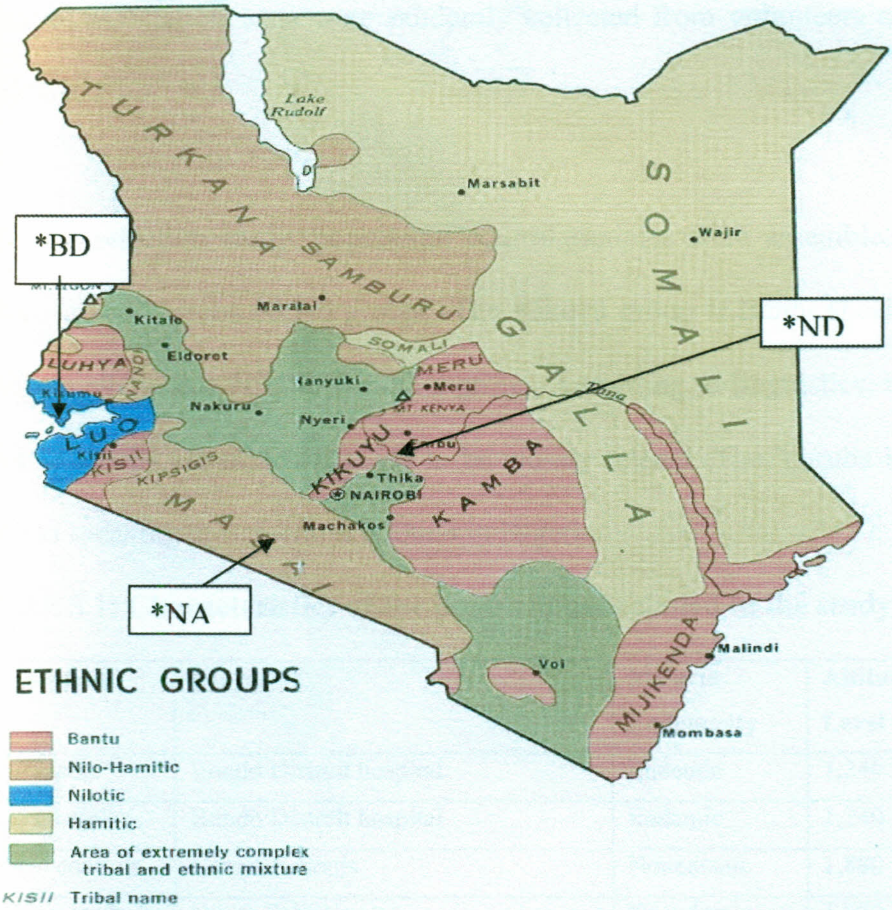


Figure 3.1: Shows the map of Kenya describing distribution and ethnic profiles of 3 ethnic used in this study [*Note: NAD ~ Narok District; ND ~ Nyeri District; BD ~ Bondo District]* [Adapted and revised from www.cia/gov/library., 1974]

3.3.2.3: International Haplotype Map project

The samples for Yoruba ethnic group were obtained from the International Haplotype Map (HapMap) project at their USA centre at Broad Institute of MIT and Harvard.

3.4: Study populations

The study involved four ethnic groups namely Luo, Masai, Kikuyu and Yoruba. The samples were collected as shown in Table 3.1. The Luo ethnic group speaks a Nilotic language and lives in a malaria-endemic region in western Kenya. Controls for case-control association tests were randomly collected from volunteers at nearby secondary schools

For the selection study, population control samples were assembled from the Maasai, Kikuyu and Yoruba ethnic groups. The Kikuyu group is Bantu speaking group whereas Maasai ethnic is a Nilotic speaking group. According to linguistics, the Maasai and Luo ethnic groups are often regarded to be closely related. The Yoruba is often regarded as Bantu speaking group who are found in Nigeria.

Table 3.1: Characteristics of the populations included in the study

Population	**Source**	**Malaria Endemicity**	**Altitude above sea Level (m)**
Luo cases	Bondo Distrcit hospital	Endemic	1,240
Luo controls	Bondo Distrcit hospital	Endemic	1,240
Masai controls	Narok, Schools	Nonedemic	1,880
Kikuyu controls	Nyeri, Schools	Nonedemic	1,950
Yoruba controls	International Haplotype map project	Nonedemic	1,950

3.5: Inclusion criteria

3.5.1: Severe malaria

All children under 10 years of age with severe malaria infection were included in the study. Severe malaria was defined according to the World Health Organization (WHO) criteria and also as used in the previous association studies. (WHO., 2000; Mockenhaupt *et al*., 2006). The therefore considered or included all the subjects with the overlapping clinical manifestations at the time of hospitalization. The clinical manifestations included: respiratory distress, convulsions, prostration and hyperthermia (> 38^0C). Severe malaria status was not only determined by clinical manifestations but was also confirmed through blood smears and Giemsa staining. A subject was therefore declared to be suffering from severe malaria when he or she presented malaria associated clinical manifestation with 12 or more parasites per 200 red blood cells.

3.5.2: Selection study

The controls for severe malaria cases were expected to be students from the Luo community in Bondo District who have been exposed to malaria for about 13 years. Other populations for the selection study were also expected to be above 14 years of age. Matching the ages of the control subject was critical because they had survived through diseases that target infants. For the Yoruba, the study focused on unrelated men and women, the parents in HapMap mother-father-child trios.

3.6: Exclusion criteria

The selection of severe malaria patients excluded all children that were tested HIV positive, malnourished with the sign of under- weight. All children who were hospitalized but were negative for the parasite or had parasite counts less than 12 per 200 red blood cells were excluded in the study..

3.7: Ethical considerations

All the participants gave informed consent while for children an informed consent was obtained from either their parents and/or guardians (See Appendix E part 1). The study was reviewed and approved by Harvard Medical School and Kenyatta University ethical reviews boards, and both ministry of health and education (See Appendix E part 2, 3 and 4).

3.8: Sample collection and DNA extractions procedure

About 2 ml of blood was obtained by venipuncture for all the samples that I collected from study areas. The DNA was extracted within 10 hours following blood collection using a Qiagen DNA Blood mini kit, and then stored at –20 ^{0}C before being shipped to Kenyatta University for long-term storage. For details as regarding DNA extraction procedures look at this website (http://www1.qiagen.com/SelectLocation.aspx).

3.9: Genotyping at 10 candidate SNPs

All human subjects were genotyped for 13 candidate malaria Single Nucleotide Polymorphisms (SNPs) using mass spectrometry (Sequenom). SNPs with minor allele frequencies averaging less than 5% across the four ethnic groups were discarded, leaving 10 SNPs for subsequent analysis (Table 4.3). Although the X-linked *G6PD* and *CD40* genes are important candidates for malaria resistance, these were excluded from the study because the focus was on autosomal SNPs that could be compared to an empirical panel of autosomal variants in the genome. Details for the sequenom genotyping protocol is available in this website

(http://www.broad.mit.edu/gen_analysis/genotyping/sequenom_ncrr.html)

To assess genotyping quality for 85 genotypes obtained in duplicate, had 2 discrepancies and a discordance rate of 2.4%. After removing samples with < 80% genotyping completeness, the average completeness of genotypes was 97.8%. Genotyping results of 9 SNPs in Yoruba samples were compared to data from HapMap **Error! Bookmark not defined.** of 30 genotypes obtained in duplicate, there was 1 discrepancy (3.33%). All SNPs were in Hardy-Weinberg equilibrium in all the four ethnic groups studied (P>0.05).

3.10: Genotyping at 1454 random SNPs

For the assessment of allele frequency differentiation at random SNPs, the Illumina Bead Lab System was used to genotype 1,536 random SNPs from the Illumina linkage panel (covering chromosomes 1, 2, 3 and 22) in 45 of the Luo controls, 47 Masai controls and 37 Kikuyu controls. Genotypes for these SNPs in 55 Yoruba samples were obtained from the HapMap databbase.. Of these SNPs 1,454 passed standard quality control checks and had been genotyped in all the four populations. For the details protocol for illumina genotyping are available in below website)

(http://www.broad.mit.edu/gen_analysis/genotyping/illumina_infinium_ncrr.html)

3.11: Genotyping at 500,000 random SNPs

For the detection of signal of positive selection, Affymetrix Micro-Array was used to genotype 500,000 random SNPs covering all chromosomes in 90 Luo controls, 70 Masai controls and 80 Kikuyu controls (Tang *et al*., 1999). Genotypeswere also obtained for these SNPs in 55 Yoruba samples from the HapMap database (The International HapMap., 2005). Of these SNPs 80% passed standard quality checks and had been genotyped in all four populations.

For the details of Affymetrix Micro-Array genotyping are available in below website.

(http://www.affymetrix.com/products/arrays/specific/500k.affx)

3.12: Case-control association analysis

Statistical significance of allele frequency differences was assessed between Luo cases and Luo controls using a χ^2 test with one degree of freedom. A 1-tailed P-values was used to test for statistical significance since our interest was in testing whether a genotype or allele previously associated with malaria is more common in cases than in controls. Odds ratios (ORs) were computed as $A = (P_{case}/1\text{-}P_{case}) / (P_{control}/\ 1\text{-}P_{controls})$, where P_{case} is the frequency in cases and $P_{control}$ is the frequency in controls. A 95% confidence interval was also computed to give the range of ORs that produced a likelihood ratio consistent with the data (P>0.05). Specifically, the standard error of the log-odds ratio was estimated as $(1/n_{case\text{-}ref}+1/n_{case\text{-}var}+1/n_{control\text{-}ref}+1/n_{control\text{-}var})^{0.5}$, where $n_{case\text{-}ref}$ and $n_{case\text{-}var}$ are the counts of the reference and tested genotypes in cases, and $n_{control\text{-}ref}$ and $n_{control\text{-}var}$ are the analagous quantities in controls. The 95% confidence interval is quoted as the range $(e^{\ln(A)\text{-}1.65B}, e^{\ln(A)+1.65B})$.

3.13: Statistical test for natural selection

The model of allele frequency differentiation between two populations that were used to test for selection at a given marker is that the difference in population frequencies is normally distributed with mean 0 and variance $cp(1\text{-}p)$, where p is the ancestral frequency of the marker. This model is similar to that of Nicholson *et al.,2002* who showed that for populations with modest genetic divergence times, it is a good approximation of allele frequency differentiation (Nicholson *et al.,* 2002). Under certain assumptions, the c parameter is expected to equal two times F_{ST}. From a population genetic perspective, c can be viewed as measuring genetic drift between populations.

To estimate c empirically, data from the 1,454 randomly chosen markers was used. For a given pair of populations, c as the empirical variance of the difference in population frequencies was estimated after normalizing by $p(1\text{-}p)$ and accounting for sampling noise, which has variance $p(1\text{-}p)(1/N_1 + 1/N_2)$ where N_1 and N_2 are total allele counts for the two populations at a given marker. The normalization term $p(1\text{-}p)$ was approximated by setting p equal to the average of observed frequencies of the two populations, and approximated binomial sampling noise as normally distributed. The same approximations were applied both to our estimation of c and our subsequent analysis of individual markers. SNPs with average minor allele frequency less than 5% for the two populations being analyzed were omitted from all computations, since the normal approximation becomes less reliable.

To test whether an individual marker was more differentiated than expected between two populations, I compared the observed difference in frequency to the expected distribution $N(0, p(1\text{-}p)(c + 1/N_1 + 1/N_2))$, using the value of c estimated above, and computed a χ^2 statistic with one degree of freedom. A feature of this test is that the χ^2 statistic has mean value of 1 across the set of markers used to infer c. The test appropriately handles different sample sizes for candidate markers vs. random markers used to infer c.

3.14: Combining case-control association and the test for differentiating selection

The combined test formally evaluates whether the observed data is consistent with the null-null model of no case-control association and no selection. The test is carried out by summing the association χ^2 statistic and the differentiation χ^2 statistic, forming a χ^2 statistic with two degrees of freedom. The association χ^2 statistic used in this test is by definition a 2-tailed statistic. Sum for each pair of populations was computed by using the same association statistic in each case. When one of the two populations being compared is the Luo population, the summed counts of Luo cases and Luo controls in the combined statistics reported in appendix B Table 5 was used. This generally leads to less significant *P*-values than using Luo controls only (and so is conservative). Using summed counts of Luo cases and Luo controls is appropriate under the null assumption of no association, and ensures that the association statistic and differentiation statistic are independent. However, for the selection-only statistics reported in Appendix B Table 5. Luo controls only was used to evaluate the evidence of selection in the control population without regard to evidence of case-control association.

3.15: Assessing the haplotype pattern around or in ICAM, HbAS and CD36 genes.

14-20 SNPs (these were tagged SNPs with $r^2 > 8$ in Hapmap database) flanking ICAM, HbAS and CD36 variants were selected in order to compare the LD pattern in both exposed and un-exposed populations (Weale *et al.*, 2003). Haploview program was used to analyze this data, it is designed to simplify and expedite the process of haplotype analysis by providing a common interface to several tasks relating to haplotype block analyses. See (http://www.broad.mit.edu/mpg/haploview/)

3.16: Detection of the signals of selection from LD data

80 and 158 SNPs within 1 Mb of the CD36 and HbAS genes were assessed respectively. The goals were to use the Long Range Haplotype test (LRH) to test for alleles of high frequency with long range linkage disequilibrium suggesting that the haplotype a rose to high frequency before recombination could break the association of the nearby SNPs. the

below formula was applied to test for Extended Haplotype Homozygosity (EHH), which is a signal of selection.

$$EHH_t = \frac{\sum_{i=1}^{s} \binom{e_{ti}}{2}}{\binom{c_t}{2}}$$

Where c is the number of samples of a particular core haplotype, e is the number of samples of a particular extended haplotype and s is the number of unique extended haplotypes. (http://www.broad.mit.edu/mpg/sweep)

CHAPTER FOUR:

RESULTS

4.1 Overview

This chapter shows the results of replication of case and control associations tests of 10 malaria variants, improved case and control association test by combining statistical weights of case control and allele frequency differentiation tests, population structure of the four study populations by principal component analysis approach, LD patterns of SNPs around three selected genes and lastly preliminary result signal of selections.

4.2 Case –control associations

Ten variants were tested for association to malaria by comparing Luo cases to controls. Variants HbAS and CD36-GT showed nominally statistically significant associations by 2-sided and 1-sided tests. All the 2-sided and I-sided tests on either allele or genotypes were statistically significant (See Table 4.1 and appendix A Table 1, 2, 3 and 4). The highest and lowest genotype and allele frequency were 25% and 3% respectively

Table 4.1: Replication analysis for 10 previous genetic associations

Genotype / allele previously associated to malaria susceptibility	Direction of previous association	Genotype / allele frequency in controls	Odds ratio (95% confidence interval)	*P*-value of test of association to the previously associated genotype or allele
HbAS((Aidoo *et al.*,2002)	Protection [a,c]	24%	0.57 (0.41-0.79)	0.0004
CD36-GT (Aitman *et al.*,2000)	Risk [b]	12%	1.5 (1.03-2.18)	0.015
ICAM-TT (Kun et al.,1999)	Protection [a]	7%	0.71 (0.42-1.21)	0.10
NOS2A-1659-AA (Burgner et al.,2003)	Risk [b]	6%	0.42 (0.21-0.83)	0.99
TNF-238 A allele (Knight et al.,2002; McGuire *et al.*,1994)	Risk [b]	10%	1.0 (0.73-1.39)	0.49
CD32-AA (Shi et al.,2001;Cooke *et al.*,2003)	Protection [d,e]	25%	0.95 (0.71-1.29)	0.38
IFNARI-LI168V –CC (Aucan et al.,2003)	Protection [b]	3%	1.18 (0.54-2.07)	0.76
TNF-308-A allele (Knight et al.,1999; Flori *et al.*,2005)	Risk [b]	9%	1.13 (0.82-1.56)	0.21
IFNARI-17470-CC (Aucan et al.,2003)	Protection [b]	10%	0.85 (0.53-1.36)	0.34
TLR-4-AG (Mockenheupt et al.,2006)	Risk [a]	10%	1.36 (0.85-2.17)	0.10

Published association was to: [a]Severe malaria, [b]Cerebral malaria, [c]Mild malaria, [d]Severe malarial anemia, or [e]Parasitaemia.

A well-known association in which heterozygotes for the sickle cell trait HbAS (HbAS-T) was replicated with a P < 0.0004, Odds Ratio (OR) of 0.57 and 95% Confidence Interval (CI) of 0.41-0.79 (Materials and Methods, Section 3.12 and Table 4.1). Association in which heterozygotes for CD36-GT are at increased risk for severe malaria was also observed with the P<0.015, OR of 1.50 and 95% CI = 1.03-2.18) *NOS* (rs8078340) gives a nominally significant 2-sided P-value, but when tested against susceptibility as observed in previous study, the P < 0.99. (Table 4.2; Appendix A Table 2 and 4). (Burgner *et al*., 2003).

Table 4.2: Shows the distribution of allele frequency in 4 ethnic groups.

SNP Information		Allele frequency (# alleles used in assessment)			
Allele name	Ref. SNP ID	Luo	Yoruba	Masai	Kikuyu
HbAS-T	rs334	13% (908)	11% (102)	0% (194)	0% (200)
CD36-G	rs3211938	6 % (914)	22% (100)	1% (186)	0% (202)
ICAM-T	rs5491	25% (910)	24% (100)	16% (186)	18% (206)
NOS2A-1659-A	rs8078340	21% (910)	19% (98)	25% (188)	21% (204)
TNF-238-A	rs361525	9% (914)	1% (100)	21% (192)	16% (202)
CD-32-A	rs1801274	50% (900)	50% (98)	50% (190)	44% (210)
IFNARI-L168V-C	rs2257167	16% (914)	16% (98)	25%(192)	19% (208)
TNF-308-A	rs1800629	9% (866)	6% (96)	6% (188)	7% (204)
IFNARI-17470-C	rs1012335	32% (904)	22% (108)	33% (190)	35% (202)

4.2.1 Allele differentiation in the four study populations

The sickle cell allele HbAS-T is present at appreciable frequency in the Luo (13%) and Yoruba (11%), but is absent in the non-endemic Masai and Kikuyu. The *CD36*-G allele is present at 22% in the Yoruba and 6% in the Luo, but occurs at only ~1% frequency in the non-endemic populations (See Table 4.2). This result is illustrated in Figure 4.1 and the result shows that frequency of HbS-T in Yoruba and Luo ethnic groups not statistically significant whereas that of CD36-G is statistically significant (error bars are not overlapping).

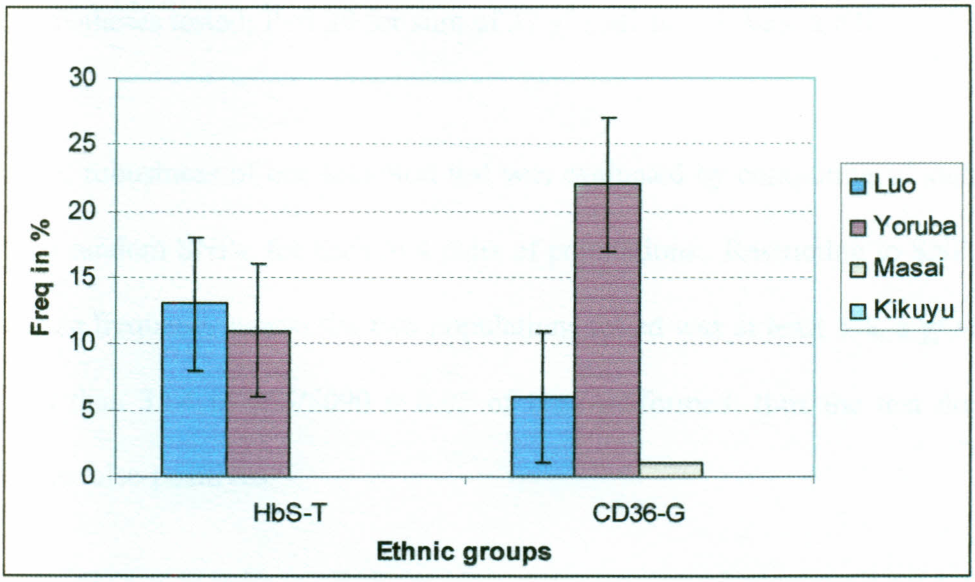


Figure 4.1: Shows Allele frequency distribution of HbS-T and CD36-G alleles that associated with protection and risk to severe malaria.

4.3: Combined case-control and allele selection.

By contrast, the 8 SNPs that do not give positive case-control association show no evidence for differentiating selection (P>0.25 for each SNP and pair of populations after correcting for multiple hypotheses tested; P=0.39 for sum of 32 χ^2 statistics at these 8 SNPs) (Table 4.3).

Furthermore, robustness of our selection test was evaluated by computing χ^2 statistics for each of the 1,454 random SNPs, for each of 4 pairs of populations. Restricting to SNPs in which the average allele frequency across the two populations tested was at least 5%, a χ^2 value observed was greater than 3.84 in 255/5090 = 0.05 of tests performed; thus the test does not have a propensity to false-positives.

Table 4.3: Tests for differentiating selection between endemic and non-endemic populations

SNP Information		Comparison of endemic and non-endemic population			
Allele name	Ref. SNP ID	Luo vs Masai	Luo vs Kikuyu	Yoruba vs Masai	Yoruba vs Kikuyu
HbAS-T	rs334	0.00149	0.00036	0.044	0.025
CD36-G	rs3211938	n/a	n/a	0.00590	0.00096
ICAM-T	rs5491	0.19	0.25	0.41	0.48
NOS2A-1659-A	rs8078340	0.62	0.86	0.57	0.89
TNF-238-A	rs361525	0.04	0.13	0.00741	0.010
CD-32-A	rs1801274	1.0	0.37	1.0	0.56
IFNARI-L168V-C	rs2257167	0.20	0.61	0.37	0.72
TNF-308-A	rs1800629	0.60	0.74	1.0	0.84
IFNARI-17470-C	rs1012335	0.92	0.64	0.32	0.18
TLR-4-G	rs4986790	0.60	1.0	0.6	0.84

P-values for selection are based on allele frequency differentiation tests between malaria endemic (Luo and Yoruba) and non-endemic populations (Kikuyu and Masai). Statistics for SNPs with average minor allele frequency less than 5% for the two populations analyzed are reported as n/a.

The evidence from association with the evidence from the selection tests was formally combined as noted in the methods. Combined test evaluateed whether the observed data is consistent with the model of no case-control association and no selection. While the evidence for association at HbAS-T and *CD36*-G is only moderate by the association analysis alone (See 2-sided P-values in Appendix A, Table 2), significance is boosted dramatically when the association and selection evidence are combined: P=0.000018-0.00029 for HbAS-T and P=0.00043-0.017 for *CD36*-G, depending on which populations are compared (Table 4.4). These results remain statistically significant after correcting for 40 comparisons tested (P=0.00072 for HbAS-T and P=0.017 for *CD36*-G)

Table 4.4: Combining case-control analysis and tests of natural selection

SNP Information		P-value from combining case-control association with tests for differentiating selection			
Allele name	Ref. SNP ID	Luo vs Masai	Luo vs Kikuyu	Yoruba vs Masai	Yoruba vs Kikuyu
HbAS-T	rs334	0.000056	0.000018	0.00048	0.00029
CD36-G	rs3211938	n/a	n/a	0.0023	0.00043
ICAM-T	rs5491	0.19	0.23	0.32	0.36
NOS2A-1659-A	rs8078340	0.033	0.038	0.032	0.038
TNF-238-A	rs361525	0.12	0.30	0.028	0.038
CD-32-A	rs1801274	0.96	0.65	0.96	0.81
IFNARI-L168V-C	rs2257167	0.34	0.68	0.52	0.73
TNF-308-A	rs1800629	0.63	0.69	0.76	0.75
IFNARI-17470-C	rs1012335	0.91	0.79	0.56	0.38
TLR-4-G	rs4986790	0.42	0.41	0.38	0.42

4.4: Test for population structure

The genetic differentiation among populations ranges from 0.0012 between Masai and Kikuyu (lowest differentiation), to 0.021 between Yoruba and Masai (highest differentiation). . The F_{ST} values estimated between pairs of populations are Yoruba-Luo (0.008), Yoruba-Masai (0.021), Yoruba-Kikuyu (0.015), Luo-Kikuyu (0.008), Luo-Masai (0.011) and Kikuyu-Masai (0.0012) (Figure 4.2).The top eigenvector is highly statistically significant by principal components analysis ($P<<10^{-12}$). The second eigenvector is not significant ($P = 0.09$). Evidence for population substructure in the Luo which could in principle confound our case-control tests of association was not observed.

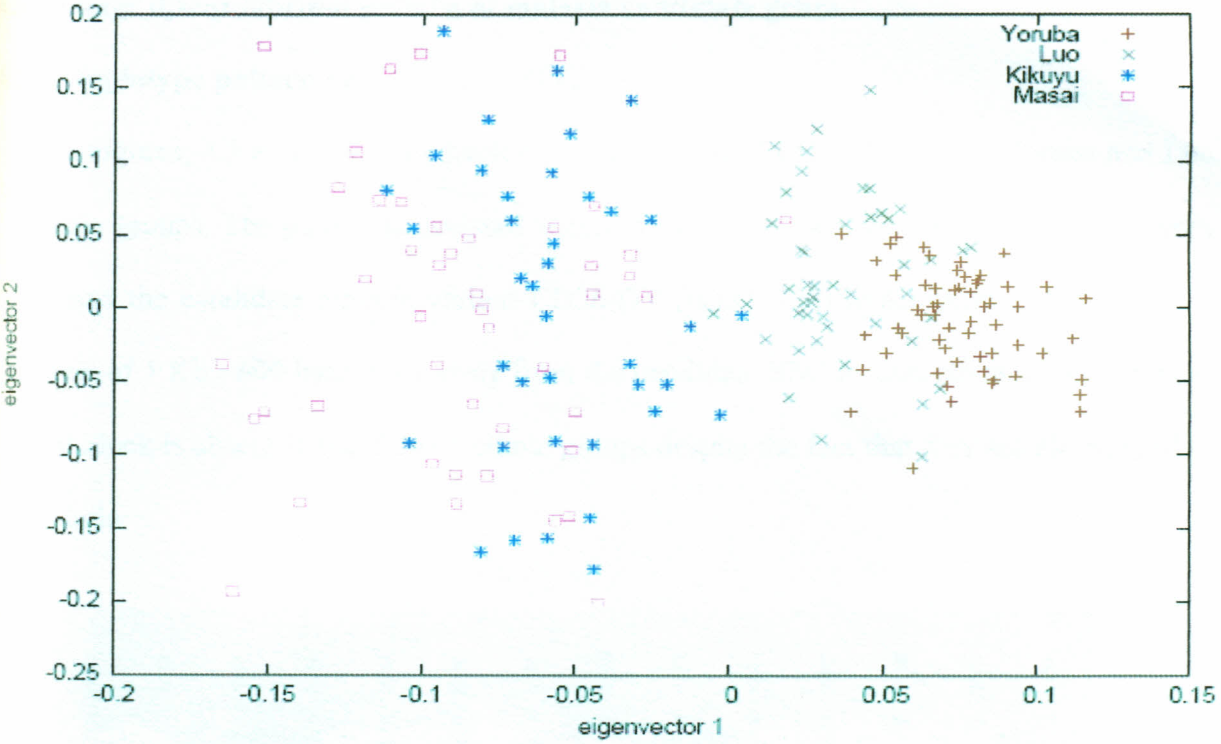


Figure 4.2 Principal components analysis of samples from four different populations genotyped at 1,454 SNPs.

4.5 Linkage disequilibrium pattern of malaria candidate genes

4.5.1: Haplotype pattern around CD36-GT.

The Figures; 4.3 and 4.4, show haplotype blocks around CD36-GT in the Yoruba and Luo ethnic groups. The pattern is assessed within 55 Kb with 14 SNPs. The haplotype pattern around the candidate malaria variant CD36-GT (rs3211938) looks quite different with a block of 1 Kb ~600 base pairs away from the candidate SNP in Luo ethnic group whereas that block is absent in the Yoruba ethnic groups despite the fact that they are all exposed to malaria.

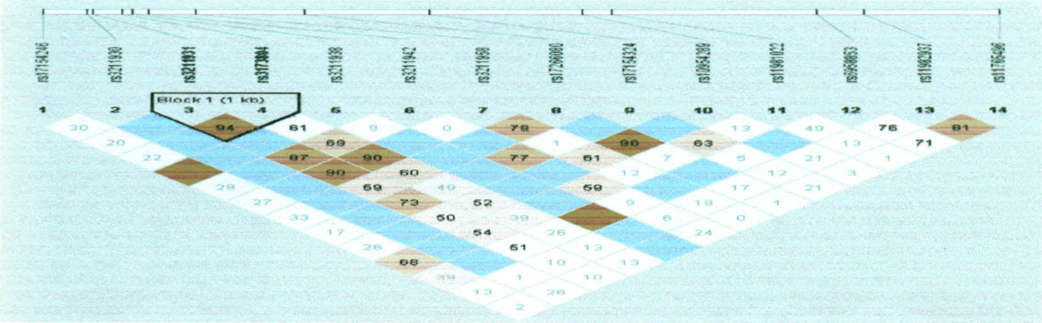


Figure 4.3: Haplotype block pattern of CD36 in Luo ethnic group

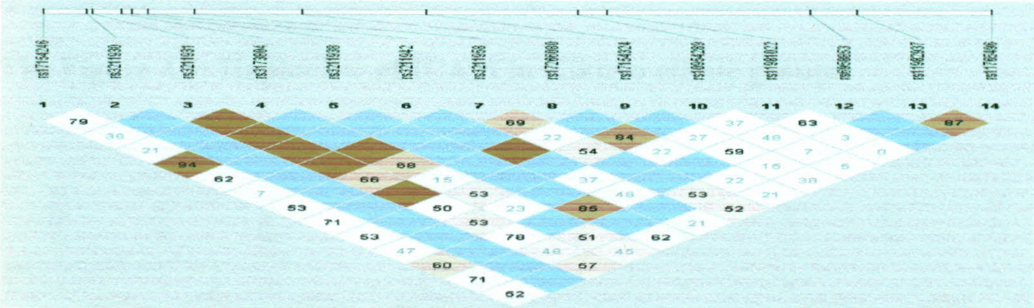
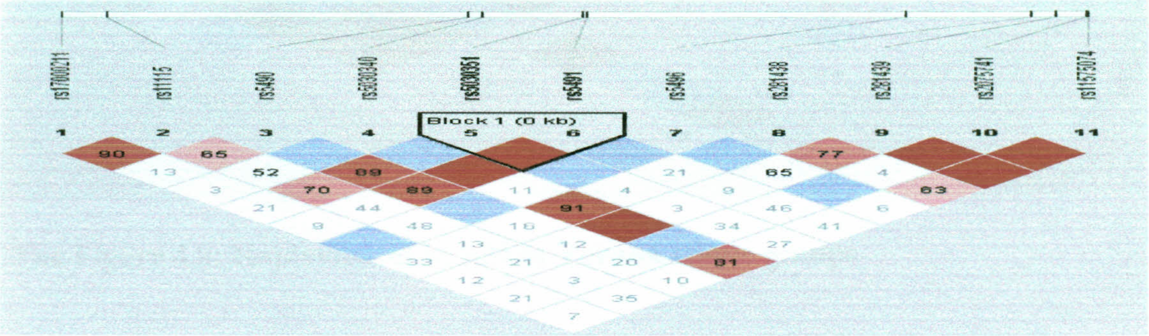


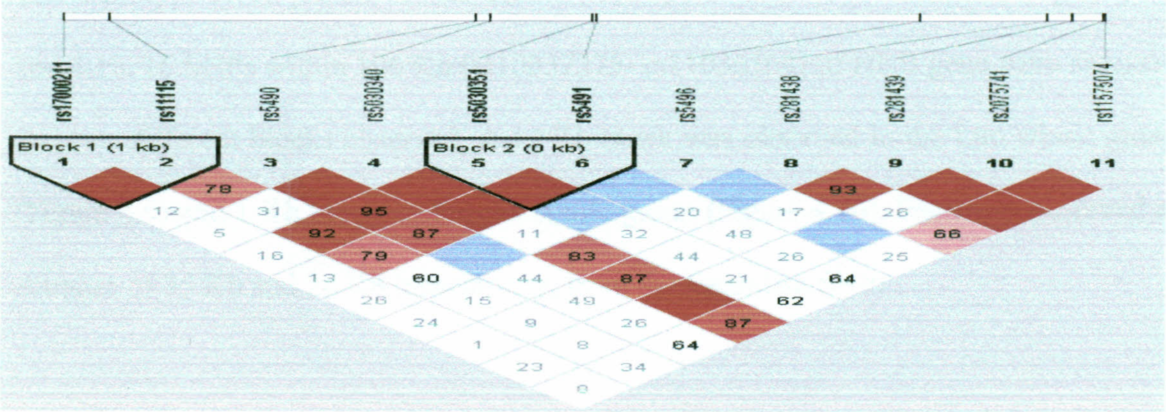
Figure 4.4: CD36 in Yoruba ethnic group

4.5.2: Haplotype pattern around ICAM-T

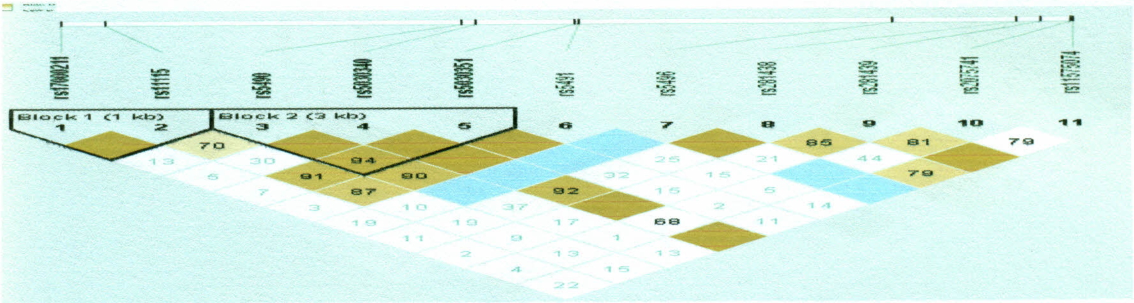
ICAM-T (rs5491) was present in both the exposed and unexposed populations and haplotype patterns in the 4 populations are as illustrated in figures; 4.5, 4.6, 4.7 and 4.8. The assessment was done within 32 Kb region consisting of 11 SNPs. Two blocks were observed. The block (~ 600 base pairs) that carried the candidate SNP was present in Kikuyu, Masai and Yoruba but absent in the Luo ethnic group. However, one unique block (~ 3 Kb) that carried the candidate SNP was observed in the Luo ethnic group. Another single block 16 Kb away from the rs5491 was present only in the exposed populations (Luo and Yoruba ethnic groups).



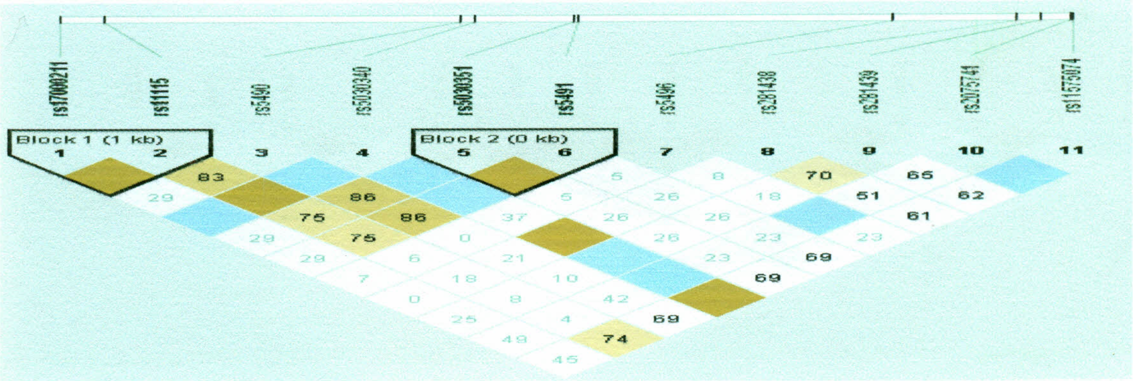
The Figure 4.5: Haploview of ICAM in Kikuyu ethnic groups



The Figure 4.6: Haploview of ICAM in Masai ethnic groups



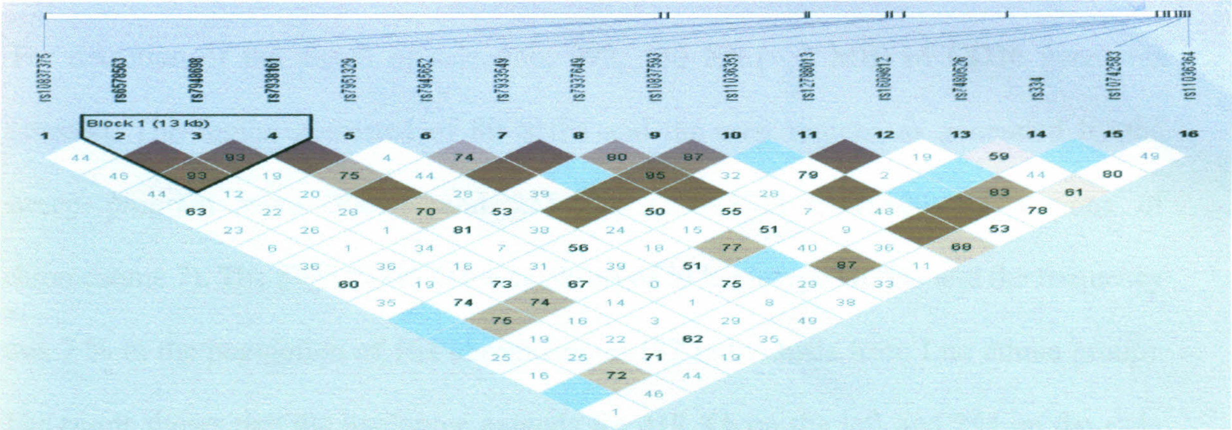
The Figure 4.7: Haploview of ICAM in Luo ethnic groups



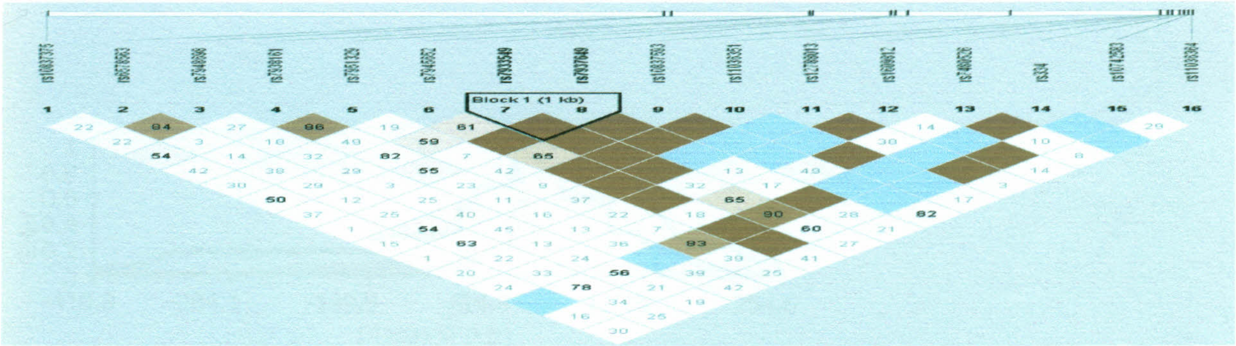
The Figure 4.8: Haploview of ICAM in Yoruba ethnic groups

4.5.3: Haplotype pattern around HbAS

The haplotype block pattern around HbAS (rs334) was observed as shown in figures; 4.9 and 4.10. 16 SNPs within 106 Kb (rs10837375- rs11036364) of HBB gene were assessed and two different blocks observed. A 13 Kb block was observed in the Luo ethnic group (Figure 4.9) and 1 Kb block in Yoruba ethnic group (Figure 4.10). Each was observed at distance of 35 Kb and 26 Kb respectively, from the candidate SNP.



The Figure 4.9: Haploview of HbAS in Luo ethnic groups



The Figure 4.10: Shows haploview of HbAS in Yoruba ethnic groups

4.6: Preliminary results of signal for selection (HbAS and CD36).

For detection of the signal of selection, 992, 130 Kb (~ 1 MB) of CD36 gene was assessed. The region consisted of 89 SNPs and the core regions as generated by the sweep program were 306Kb and 255 Kb (79942371-79945100 Kb regions of chromosome 7). The SNP CD36-GT was observed in haplotype GCTG and the frequency was 7 % in the population of 164 chromosomes (82 individuals from Luo ethnic group). The result shows that the haplotype extends for 418 Kb on the left and 254 on the right suggesting that it recently arisen in the Luo ethnic group (See figure 4.11)

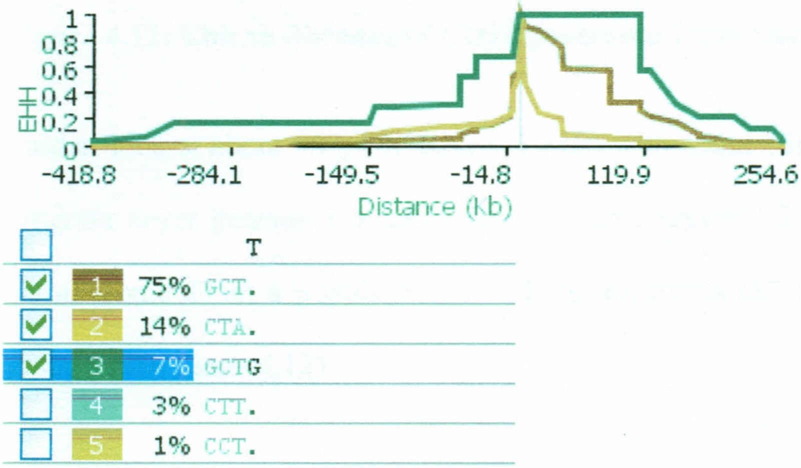
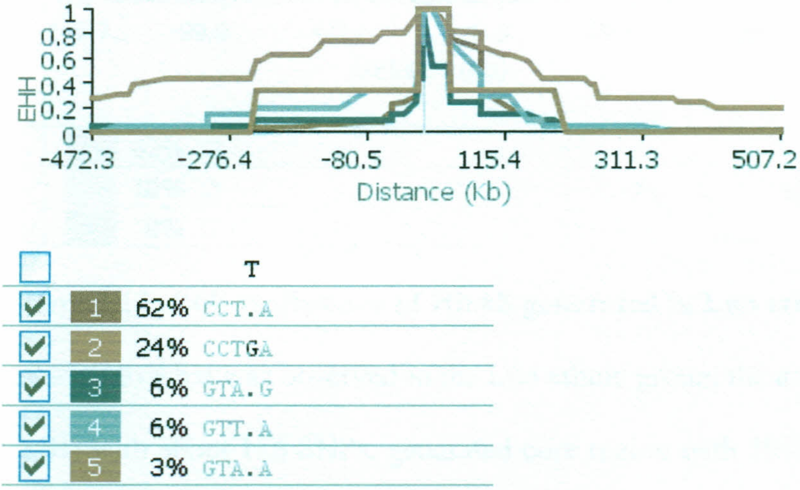


Figure 4.11: Ehh vs distance of CD36 generated from Luo ethnic group

In Yoruba ethnic group, the region was spanning 985,887 KB (~ 1 MB) with 84 SNPs. The core region as generated by the sweep program consisted of 306Kb and 306 Kb that were overlapping within 79942371-79948723 regions of chromosome 7. The SNP CD36-GT is in haplotype GCTG and its frequency is 24 % in a population of 120 chromosomes.. The result shows that the haplotype extends for a long distance 472 Kb on the left and 507 on the right similarly suggesting that SNP CD36-GT must have recently arisen. (See figure 4.12)



☐		T
☑	1	62% CCT.A
☑	2	24% CCTGA
☑	3	6% GTA.G
☑	4	6% GTT.A
☑	5	3% GTA.A

Figure 4.12: Ehh vs distance of CD36 generated from Yoruba ethnic group

Further look at about 967, 736 Kb (~ 1 MB) with about 156 SNPs revealed that sweep program never generated blocks within the core region (5204808-5205217) regions of chromosome 11 in a population of 164 chromosomes (82 individuals from Luo ethnic group). (See figure 4.12)

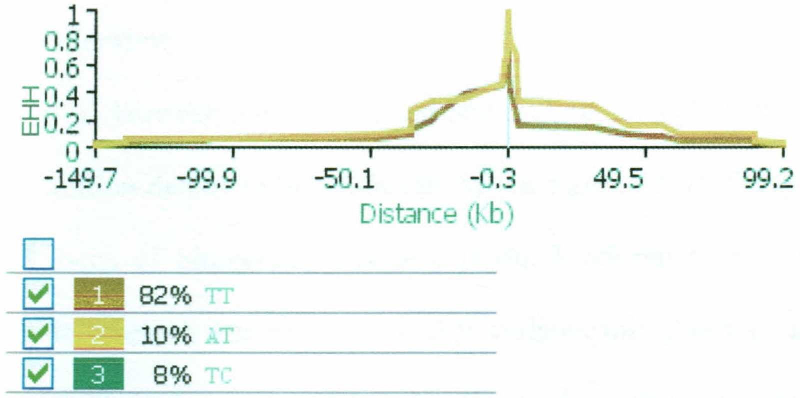


Figure 4.13: Ehh vs distance of HbAS generated in Luo ethnic group

Contrary to what was observed in the Luo ethnic group, the assessment of 967, 736 Kb (~ 1 MB) with about 156 SNPs, generated core region with 299Kb and 300 Kb blocks. The blocks were overlapping within 79942371-79945100 regions of chromosome 11. The SNP HbAS is in haplotype AA and it frequency is 12 % in the population of 240 chromosomes (120 individuals from Yoruba ethnic group). The result shows that the haplotype extends beyond 467 Kb on the left and 500Kb on the right suggesting that this SNP HbAS must have recently arisen in the Yoruba ethnic group (See figure 4.14)

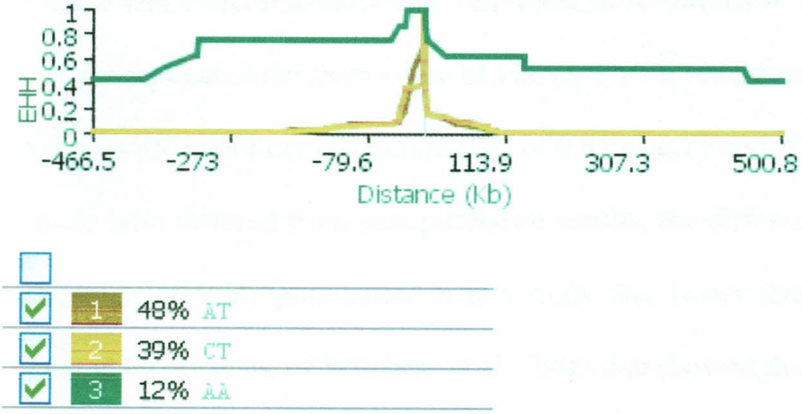


Figure 4.14: Ehh vs distance of HbAS generated in Yoruba ethnic group

CHAPTER FIVE:

DISCUSSION

5.1: Overview

There are between 350 million and 500 million clinical cases reported each year resulting in 1 million deaths in Sub-Saharan Africa especially *P. falciparum* (Guerra et al., 2006) The focus of biomedical research is on development of vaccine and new drugs but genetic research attempts to identify variants that provides absolute protection against *P.falciparum* that are candidate for drug and vaccine development. As a strategy to take genetic approach to another level, this study attempts to replicate previous associations studies in order to provide insurance against errors, investigators biases and over-elaborate data exploration (Cardell and Clayton., 2005). The study also attempts to improve power of association tests, assesses the pattern of haplotype blocks around malaria candidate genes and also tests for signals of selection for variants associated with malaria in the study in order to facilitate identification variants which may have absolute protection against *P.falciparum* malaria.

5.2: Case and control association and allele differentiation tests.

The study replicated the association of HbAS-T with severe malaria. The odds ratio (OR) was 0.57 with a confidence interval (CI) of 0.41 to 0.79 and P value of 0.0004. Although the odds ratio differed from past published results, the difference was due to the fact that the average of study population in this study was lower than previous studies. This is supported by the work of Williams et al., 2005 that showed that the protection of HbAS is strongly associated with the age. The study also replicated the association of CD36 with severe malaria and the result was OR = 1.50, 95%CI =1.03 -2.18 and P value = 0.015.

The odds ratio and confidence interval overlapped with those previous association study of CD36 (Aitman *et al.*,2000). Further association of NOS revealed OR = 0.42, 95%CI = 0.21-0.83 and P value = 0.99 but the variant was associated with protection contrary to what others observed (Burgner *et al.*, 1998; 1999; 2003). Other 7 variants were not associated with severe malaria in this study. This could be as result of use parasitemia as a baseline characteristic that was contrary to previous phenotype used in previous associations (Appendix A, Table 4).

Malaria infection selection pressure on the human genome has been demonstrated in several studies (Tishkoff et al., 2007). The most outstanding example in Duffy antigen in African populations (Miller.,1976). Other studies have demonstrated frequency differentiation of sickle cell trait and G6PD-A in African populations suggesting the effect of differential exposure (Moormann *et al.,* 2003). This study confirms differential distribution of the sickle cell allele HbAS-T in endemic and non-endemic populations. The Luo ethnic group have HbAS-T frequency of 13% and 11% was observed among the Yoruba group.but HbAS-T is absent in the non-endemic Masai and Kikuyu ethnic groups.

This study however reports for the first time, the differential distribution of CD36-G allele in the endemic and non-endemic populations .The *CD36*-G allele is present at 22% in the Yoruba and 6% in the Luo, but occurs at only ~1% frequency in the non-endemic populations Kikuyu and Maasai. The striking observation is that CD36 variant is seen to be associated with increasing susceptibility in Luo ethnic group but has a higher frequency in the endemic populations. The explanation could be that historically the selection on this variant has changed over time because of host-parasite genetic interactions (Ayodo *et al*., 2007).

On comparing allele frequency differentiation between malaria-endemic and non-endemic populations, Luo and Kikuyu comparison showed a significant differentiation at HbS-T with P of 0.00036 whereas Yoruba and Kikuyu showed P value of 0.00096 at CD36-g. Other 8 SNPs showed no sign of differentiating selection. The observation leads to the rejection of null hypothesis.

5.3: Improved power of case control tests.

This study validated a long standing idea to increase power of association test by combining the evidence of association with that from test of natural selection (Roeder et al., 2006). Previous studies have priotized SNPs by natural selection on the basis of a combination of the allele being frequent and surrounded by a long-range haplotype (Sabeti et al., 2002: Voight et al., 2006). The study however demonstrates that combining association analysis with evidence of natural selection can increase power to detect risk variants by order of magnitude.- up to P = 0.000018 for HbS – T and P = 0.00043 for CD36-G (Table 4.4).

5.4: Population structure of the 4 ethnic groups

Population sub-structure results in false negative or false positive and in this study no sub-structure of the Luo ethnic groups was observed and therefore no cause for false-positives or false negatives. .(Freedman *et al*., 2004: Patterson *et al*., 2006; Tishkoff *et* *al*., 2007). . However, we observed unusual cluster pattern between Luo and Yoruba, and Masai and Kikuyu (See figure 4.2). This observation was inconsistent with the linguistic cluster pattern where Masai and Luo who are Nilotic speakers were expected to cluster together.

5.5: LD pattern in the exposed and un-exposed populations.

The study considered the LD pattern around 3 malaria variants in CD36, HBB and ICAM genes because CD36-G and HbAS-T were associated with severe malaria in this study but ICAM-1 provides an example of adaptive evolution. There is evidence that ICAM-1 interacts with the different strains of *P. falciparum* and also influences the interferon response (Howell *et al.,* 2008; Bertonati and Tramontano., 2007; Jenkins *et al*., 2007; 2005; Baratin *et al*., 2007; Logan *et al*., 2005). The observation of different haplotype block patterns around CD36 and HBB (Figure 3.3, 3.4, 3.9 and 3.10) may demonstrate an interesting history of malaria infections in the Luo and Yoruba ethnic groups. Further work can in principle discern the haplotype block pattern and can be useful in mapping causal genes (Reich *et al*.,2001; Patil *et al*.,2001; Johnson *et al.,* 2003; Gabriel *et al* 2002;Carlson *et al*.,2003; Khor *et al*.,2007). On other hand, haplotype block patterns around ICAM-1 in both exposed and unexposed population (See Figure 3.5 and 3.8) suggest that ICAM-1 can be used to empirically demonstrate adaptive evolution as the gene interacts with the malaria parasite (Bertonati and Tramontano., 2007)

5.6: Detection of signal of positive selection.

The study of genetic imprints has profound implications in studying the population history and medicine (Sabeti *et al*., 2002b). Here, a Long range haplotype test in two populations (Luo and Yoruba) that have been exposed to malaria for thousands of years were carried out. The genes of interest have been implicated in malaria infection in several studies (Ayodo *et al.,* 2007) and this study confirms their associations with malaria. The study reports preliminary result that the surrogate haplotype that carry the locus CD36-G is under selection in both the malaria endemic or exposed Luo and Yoruba ethnic groups. This is supported by both high frequency and EHH (See figure 4.11 and 4.12) suggesting that the mutation on the haplotype is recent and arose faster than expected through neutral evolution (Sabeti *et al.*,2002b)). Interestingly, the preliminary results show evidence of selection for HbAS in the Yoruba but not the Luo ethnic group (See Figure 3.14 and 3.14). These observations tentatively support the view as observed in the case of CD36 gene that the Luo and Yoruba may have different history of exposure or transmission intensity to the malaria infection. Although this study suggests different length of time exposure to malaria infections between Luo and Yoruba ethnic groups, there is a need to determine the period of selection. There is also a need to assess the evidence of selection of HbAS using dense SNPs as the selection at HbAS looks quite old in the Luo ethnic group (Stephens *et al.*, 1998). This means the haplotype that carries the HbAS-T allele has undergone recombination and thus the LD has decayed over time.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Overview

The study replicated two associations, observed unusual frequency differentiation of the two variants in endemic and non-endemic populations, combined the statistical weights of case-control association and allele frequency differentiation to increase power of association test, observed no sub- structure of population Luo ethnic group, observed different haplotype block patterns in exposed and un exposed population and also observed signal selection of malaria associated variants., The work that stands out in this study is that it has come up with a novel ways to improve the power of genetic studies to identify host factors modulating response to infectious disease.

6.2 Summary of conclusions

- This study has not only replicated 2 previous associations (HbAS and CD36-GT) but has also developed a novel method that increases power to identify malaria variants using limited sample-size (<60 samples). By statistically combining the case-control association and frequency differentiation statistics, the power of the association tests increased by several orders of magnitude (HbAS with P value = 0.000018 and CD36-GT with P value =0.00043) (See Table 4.5)
- This study has also ascertained that genetic clusters are not consistent with linguistic clusters. In linguistic cluster, Luo ethnic group are expected to be closer to Masai than Kikuyus. This was however contrary to the observation as the Masais were closer to Kikuyus than Luos despite the fact that they speak Nilotic language (See cluster pattern in figure 4.2).

- By observing different haplotype patterns in exposed and un-exposed populations, the haplotype patterns look appear different suggesting that linkage disequilibrium can be sued to identify causal polymorphisms or genes related to specific infections. The outstanding example is of ICAM- 1 (Figures 4.5, 4.6, 4.7 and 4.8)
- The investigation of positive signal of selection in the exposed populations, particularly on the replicated SNPs suggest that each ethnic group may have had either a unique demographic history or an unstable fluctuating malaria infection selection pressures (See the pattern of CD36 and HbAS in Figures 4.11, 4.12, 4.13 and 4.14).

6.3 Recommendation for further research

- This study has developed a novel method that combines case-control and allele frequency differentiation statistics that may be useful in the context of whole genome scan association mining scans and can be used for future identification of malaria variants.
- The study recommends that any association tests should clearly define phenotypes as in most cases the phenotypes overlap and may hinder the success of any replication.
- There is need to conduct thorough population structure assessment as this study suggests that linguistic cluster pattern does not represent genetic cluster. Any attempt to ignore this in association study may result into false positives or lack of replication.

- There is a need to improve the sweep program so as to correct for variable sizes of overlapping core haplotype.

REFERENCES

Ackerman, HC., Ribas, G., Jallow, M., Mott, R., Neville, M., Sisay-Joof, F., Pinder, M., Campbell, RD and Kwiatkowski DP (2003). Complex haplotypic structure of the central MHC region flanking TNF in a West African population. *Genes Immun.* **4**: 476-486.

Agarwal, A., Guindo, A., Cissoko, Y., Taylor, JG., Coulibaly, D., Kone, A., Kayentao, K., Djimde, A., Plowe, CV., Doumbo, O., Wellems, TE and Diallo, D.(2000). Hemoglobin C associated with protection from severe malaria in the Dogon of Mali, a West African population with a low prevalence of hemoglobin S. *Blood*, **96** : 2358-2363.

Aidoo, M., Terlouw, DJ., Kolczak, MS., McElroy, PD., Kuile, FO., Kariuki, S., Nahlen, BL., Lal, AA and Udhayakumar, V (2002). Protective effects of the sickle cell gene against malaria morbidity and mortality. *Lancet.* **359** :1311-1312.

Aitman, TJ., Cooper, LD., Norsworthy, PJ., Wahid, FN., Gray, JK., Curtis, BR., McKeigue, PM., Kwiatkowski, D., Greenwood, BM., Snow, RW., Hill, AV and Scott J (2000). Malaria susceptibility and CD36 mutation. *Nature*. **405**:1015-1016.

Altshuler, D., Hirschhorn, JN., Klannemark, M., Lindgren, CM., Vohl, MC., Nemesh, J., Lane, CR., Schaffner, SF., Bolk, S., Brewer, C., Tuomi, T., Gaudet, D., Hudson, TJ., Daly, M., Groop, L and Lander, ES (2000). The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat.Genet*. **26**: 76-80.

Anonymous author/s (2007). Is malaria eradication possible? *Lancet*. **27**., 370

Anstay, NM and Price, RN (2007). Improving case definition for severe malaria. *PLOS Med.* **4**., e267.

Aubouy, A., Migot-Nabias, F and Deloron P (2007). Correlations between treatment outcome and both anti-MSP119 antibody response and erythrocyte-related genetic factors in Plasmodium falciparum malaria. *Infect Genet Evol*. **7**:147-154.

Aucan, C., Walley, AJ., Hennig, BJ., Fitness, J., Frodsham, A., Zhang, L., Kwiatkowski, D and Hill, AV (2003). Interferon-alpha receptor-1 (IFNAR1) variants are associated with protection against cerebral malaria in the Gambia. *Genes Immun.* **4**., 275-82.

Ayodo, G., Price, AL., Keinon, A., Ajwang, A., Otieno, MF., Orago, ASS., Patterson, N and Reich D (2007). Combing Evidence of Natural Selection with association Analysis Increases Power to Detect Malaria – Resistance Variants. *Am. J. Hum Genet*. **<u>81</u>**: 234-242.

Bamshad, M and Wooding SP (2003). Signatures of natural selection in the human genome. *Nat Rev Genet* **4**:99-111.

Baratin, M., Roetynck, S., Pouvelle, B., Lemmers, C., Viebig, NK., Johansson, S., Bierling, P., Scherf, A., Gysin, J., Vivier, E and Ugolini S (2007). Dissection of the role of PfEMP1 and ICAM-1 in the sensing of plasmodium falciparum-infected erythrocytes by natural killer cells. *PLoS ONE*. **2**:e228.

Barry,AE., Leliwa-Sytek, A., Tavul, L., Imrie, H., Migot-Nabias, F., Brown, SM., McVean, GA and Day KP (2007). Population genomics of the immune evasion (var) genes of Plasmodium falciparum. *PLoS Pathog*. **3**:e34.

Baruch DI. (1999). Adhesive receptors on malaria-parasitized red cells. *Baillieres Best Pract Res Clin Haematol*. **12**:747-761.

Bertonati, C and Tramontano A.(2007). A model of the complex between the PfEMP1 malaria protein and the human ICAM-1 receptor. *Proteins*. **69**:215-222.

Bethke, LL., Zilversmit, M., Nielsen, K., Daily, J., Volkman, SK., Ndiaye, D., Lozovsky, ER., Hartl, DL and Wirth DF (2006). Duplication, gene conversion, and genetic diversity in the species-specific acyl-CoA synthetase gene family of Plasmodium falciparum. *Mol Biochem Parasitol.* **150**:10-24.

Beutler, E. (1990). The genetics of glucose-6-phosphate dehydrogenase deficiency. *Semin Hematol.* **27**:137-164.

Bloland, PB., Ruebush, TK., McCormick, JB., Ayisi, J., Boriga, DA., Oloo, AJ., Beach, R., Hawley, W., Lal, A., Nahlen, B., Udhayakumar, V and Campbell CC (1999). Longitudinal cohort study of the epidemiology of malaria infections in an area of intense malaria transmission .,Description of study site, general methodology, and study population. *Am J Trop Med Hyg.* **60**:635-640.

Burgner, D., Rockett, K and Kwiatkowski D (1999). Nitric oxide and infectious diseases. *Arch Dis Child.* **81**:185-188.

Burgner, D., Usen, S., Rockett, K., Jallow, M., Ackerman, H., Cervino, A., Pinder, M and Kwiatkowski DP (2003). Nucleotide and haplotypic diversity of the NOS2A promoter region and its relationship to cerebral malaria. *Hum Genet* **112**:379-386

Burgner, D., Xu, W., Rockett, K., Gravenor, M., Charles, IG., Hill, AV and Kwiatkowski D (1998). Inducible nitric oxide synthase polymorphism and fatal cerebral malaria. *Lancet.* **352**:1193-1194.

Cardell, HJ and Clayton DG (2005). Genetic association studies. *Lancet* **366:**1121-1131

Carlson, CS., Eberle, MA., Rieder, MJ., Smith, JD., Kruglyak, L and Nickerson, DA (2003). Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nat Genet*. **33**:518-521.

Carlson, J., Helmby, H., Hill, AVS and Brewster, D, *et al.*(1990). Human cerebral malaria: association with erythrocyte rosetting and lack of anti-rosetting antibodies. *Lancet* **336**:1457-1460.

Carpenter, D., Abushama, H., Bereczky, S., Färnert, A., Rooth, I., Troye-Blomberg, M., Quinnell, RJ and Shaw,MA (2007). Immunogenetic control of antibody responsiveness in a malaria endemic area. *Hum Immunol*. **68**:165-169.

Chotivanich, K., Udomsangpetch, R., Simpson, JA., Newton, P., Pukrittayakamee, S., Looareesuwan, S and White, NJ (2000). Parasite multiplication potential and the severity of Falciparum malaria. *J Infect Dis*. **181**:1206-1209.

Clark IA (2007). The advent of the cytokine storm. *Immunol Cell Biol*. **85**: 271-273.

Clark AG (2001). Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science*. **293**: 455-62

Clark, IA., Alleva, LM., Mills, AC and Cowden WB (2004). Pathogenesis of malaria and clinically similar conditions.*Clin Microbiol Rev*. **17**:509-539.

Clark, IA., Budd, AC., Alleva, LM and Cowden WB (2006). Human malarial disease: a consequence of inflammatory cytokine release. *Malar J*. **5**:85.

Clarke, SE., Brooker, S., Njagi, JK., Njau, E., Estambale, B., Muchiri, E and Magnussen, P (2004). Malaria morbidity among school children living in two areas of contrasting transmission in western Kenya. *Am J Trop Med Hyg* **71**:732-738.

Carlson, CS., Eberle, MA., Rieder, MJ., Smith, JD., Kruglyak, L and Nickerson, DA (2003). Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nat Genet*. **33**:518-521.

Carlson, J., Helmby, H., Hill, AVS and Brewster, D, *et al.*(1990). Human cerebral malaria: association with erythrocyte rosetting and lack of anti-rosetting antibodies. *Lancet* **336**:1457-1460.

Carpenter, D., Abushama, H., Bereczky, S., Färnert, A., Rooth, I., Troye-Blomberg, M., Quinnell, RJ and Shaw,MA (2007). Immunogenetic control of antibody responsiveness in a malaria endemic area. *Hum Immunol*. **68**:165-169.

Chotivanich, K., Udomsangpetch, R., Simpson, JA., Newton, P., Pukrittayakamee, S., Looareesuwan, S and White, NJ (2000). Parasite multiplication potential and the severity of Falciparum malaria. *J Infect Dis*. **181**:1206-1209.

Clark IA (2007). The advent of the cytokine storm. *Immunol Cell Biol*. **85**: 271-273.

Clark AG (2001). Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science*. **293**: 455-62

Clark, IA., Alleva, LM., Mills, AC and Cowden WB (2004). Pathogenesis of malaria and clinically similar conditions.*Clin Microbiol Rev*. **17**:509-539.

Clark, IA., Budd, AC., Alleva, LM and Cowden WB (2006). Human malarial disease: a consequence of inflammatory cytokine release. *Malar J*. **5**:85.

Clarke, SE., Brooker, S., Njagi, JK., Njau, E., Estambale, B., Muchiri, E and Magnussen, P (2004). Malaria morbidity among school children living in two areas of contrasting transmission in western Kenya. *Am J Trop Med Hyg* **71**:732-738.

Cooke, GS., Aucan, C., Walley, AJ., Segal, S., Greenwood, BM., Kwiatkowski, DP and Hill, AV (2003). Association of Fcgamma receptor IIa (CD32) polymorphism with severe malaria in West Africa. *Am J Trop Med Hyg* **69**:565-568.

Daly, MJ., Rioux, JD., Schaffner, SF., Hudson, TJ and Lander, ES (2001). High-resolution haplotype structure in the human genome. *Nat Genet*.**29**:229-232.

Deitsch, KW and Wellems TE (1996). Membrane modifications in erythrocytes parasitized by *Plasmodium falciparum*. *Mol Biochem Parasitol.* 76:1-10

Dobaño, C., Rogerson, SJ., Mackinnon, MJ., Cavanagh, DR., Taylor, TE., Molyneux, ME and McBride, JS (2008). Differential Antibody Responses to Plasmodium falciparum Merozoite Proteins in Malawian Children with Severe *Malaria. J Infect*. (Epub ahead of print)

Durie, FH., Foy, TM and Noelle, RJ (1994b). The role of CD40 and its ligand (gp39) in peripheral and central tolerance and its contribution to autoimmune disease. *Res Immunol*. **145**:200-205.

Durie, FH., Foy, TM., Masters, SR., Laman, JD and Noelle, RJ (1994a). The role of CD40 in the regulation of humoral and cell-mediated immunity. *Immunol Today*. **15**: 406-411.

Enevold, A., Nkya, WM., Theisen, M., Vestergaard, LS., Jensen, AT., Staalsoe, T., Theander, TG., Bygbjerg, IC and Alifrangis M (2007). Potential impact of host immunity on malaria treatment outcome in Tanzanian children infected with Plasmodium falciparum. *Malar J*. **6**:153.

Fernandez-Reyes, D., Craig, AG., Kyes, SA., Peshu, N., Snow, RW., Berendt, AR., Marsh, K and Newbold, CI (1997). A high frequency African coding polymorphism in the N-terminal domain of ICAM-1 predisposing to cerebral malaria in Kenya. *Hum Mol Genet.* **6**:1357-1360.

Flint, J., Hill, AV., Bowden, DK., Oppenheimer, SJ., Sill, PR., Serjeantson, SW., Bana-Koiri, J., Bhatia, K., Alpers, MP and Boyce, AJ, *et al* (1986). High frequencies of alpha-thalassaemia are the result of natural selection by malaria. *Nature*. **321**:744-750.

Flori, L., Delahaye, NF., Iraqi, FA., Hernandez-Valladares, M., Fumoux, F and Rihet P (2005). TNF as a malaria candidate gene: polymorphism-screening and family-based association analysis of mild malaria attack and parasitemia in Burkina Faso.*Genes Immun* **6**:472-480.

Freedman, ML., Reich, D., Penney, KL., McDonald, GJ., Mignault, AA., Patterson, N., Gabriel, SB., Topol, EJ., Smoller, JW., Pato, CN., Pato, MT., Petryshen, TL., Kolonel, LN., Lander, ES., Sklar, P., Henderson, B., Hirschhorn, JN and Altshuler D (2004). Assessing the impact of population stratification on genetic association studies. *Nat Genet*. **36**:388-393.

Gabriel, SB., Schaffner, SF., Nguyen, H., Moore, JM., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, SN., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, ES., Daly, MJ and Altshuler D (2002). The structure of haplotype blocks in the human genome.*Science*.**296**:2225-2229.

Githeko A and Ndegwa (2001). Predicting epidemics in the Kenyan highlands using climate data., a tool for decision makers. *Global Change & Human Health*. **2**. 54-63.

Goldstein DB (2008). Genomics and Biology come together to fight HIV. *Plos Biology* **6**:e76

Goldstein, DB and Hirschhorn, JN (2004) In genetic control of disease, does 'race' matter? *Nat Genet*.**36**:1243-1244.

Goldstein, DB and Weale, ME (2001). Population genomics: linkage disequilibrium holds the key. *Curr Biol*. **11**:R576-579.

Graves P, Gelband H.(2007) WITHDRAWN: Vaccines for preventing malaria. *Cochrane Database Syst Rev*. **18**:CD000129.

Graves, P and Gelband H (2006) Vaccines for preventing malaria (SPf66) *Cochrane Database Syst Rev*.**2**:CD005966.

Graves, P., Gelband, H and Garner, P (1998). The SPf66 Malaria Vaccine: What is the Evidence for Efficacy? *Parasitol Today*. **14**:218-220.

Greenwood B (1999). Malaria mortality and morbidity in Africa. *Bull World Health Organ.***77**: 617-618.

Guerra CA, Snow RW, Hay SI (2006). Defining the global spatial limits of malaria transmission in 2005. *Adv Parasitol* **62**:157-79.

Hanchard, NA., Rockett, KA., Spencer, C., Coop, G., Pinder, M., Jallow, M., Kimber, M., McVean, G., Mott, R and Kwiatkowski DP (2006). Screening for recently selected alleles by analysis of human haplotype similarity. *Am J Hum Genet* **78**:153-159.

Handunnetti, SM., Schravendijk, MR., Hasler, T and Barnwell JW, *et al* (1992) Involvement of CD36 on erythrocytes as a rosetting receptor for Plasmodium falciparum-infected erythrocytes. *Blood.***80**: 2097-3104.

Hartl, DL (2004). The origin of malaria: mixed messages from genetic diversity. *Nat Rev Microbial*. **2**:15-22.

Hill AVS (2006) Aspect of genetic susceptibility to human infectious diseases. *Annu. Rev Genet*. **40**: 469-486.

Hill, AV (2003). Interferon-alpha receptor-1 (IFNAR1) variants are associated with protection against cerebral malaria in the Gambia. *Genes Immun.***4**:275-282.

Hill, AV., Allsopp, CE., Kwiatkowski, D., Anstey, NM., Twumasi, P., Rowe, PA., Bennett, S., Brewster, D., McMichael, AJ and Greenwood BM (1991). Common west African HLA antigens are associated with protection from severe malaria. *Nature*.**352**: 595-600.

Hill, AV., Jepson, A., Plebanski, M and Gilbert, SC (1997). Genetic analysis of host-parasite coevolution in human malaria. *Phil. Trans. R. Soc. Lond. B*. **352**: 1317-1325.

Hirschhorn, JN., Lohmueller, K., Byrne, E and Hirschhorn K (2002). A comprehensive review of genetic association studies. *Genet Med*. **4**:45-61.

Hobbs, MR., Udhayakumar, V., Levesque, MC., Booth, J., Roberts, JM., Tkachuk, AN., Pole, A., Coon, H., Kariuki, S., Nahlen, BL., Mwaikambo, ED., Lal, AL., Granger, DL., Anstey, NM and,Weinberg, JB(2002). A new NOS2 promoter polymorphism associated with increased nitric oxide production and protection from severe malaria in Tanzanian and Kenyan children.*Lancet*.**360**:1468-1475.

Howell, DP., Levin, EA., Springer, AL., Kraemer, SM., Phippard, DJ., Schief, WR and Smith JD (2008). Mapping a common interaction site used by Plasmodium falciparum Duffy binding-like domains to bind diverse host receptors. *Mol Microbiol*. **67**:78-87.

Ioannidis, JP., Ntzani, EE., Trikalinos, TA and Contopoulos-Ioannidis, DG (2001) Replication validity of genetic association studies. *Nat Genet*. **29**:306-309.

Jeffares, DC., Pain, A., Berry, A., Cox, AV., Stalker, J., Ingle, CE., Thomas, A., Quail, MA., Siebenthall, K., Uhlemann, AC., Kyes, S., Krishna, S., Newbold, C., Dermitzakis, ET and Berriman, M (2007). Genome variation and evolution of the malaria parasite Plasmodium falciparum. *Nat Genet* 39:120-125.

Jenkins, N., Wu, Y., Chakravorty, S., Kai, O., Marsh, K and Craig, A (2007) *Plasmodium falciparum* intercellular adhesion molecule-1-based cytoadherence-related signaling in human endothelial cells. *J Infect Dis*. **196**:321-327.

Jenkins, NE., Mwangi, TW., Kortok, M., Marsh, K., Craig, AG and Williams TN (2005) Polymorphism of intercellular adhesion molecule-1 is associated with a reduced incidence of nonmalarial febrile illness in Kenyan children. *Clin Infect Dis*. **41**:1817-1819.

Johnson, GC., Esposito, L., Barratt, BJ., Smith, AN., Heward, J., Di Genova, G., Ueda, H., Cordell, HJ., Eaves, IA., Dudbridge, F., Twells, RC., Payne, F., Hughes, W., Nutland, S., Stevens, H., Carr, P., Tuomilehto-Wolf, E., Tuomilehto, J., Gough, SC., Clayton, DG and Todd JA (2001). Haplotype tagging for the identification of common disease genes. *Nat Genet*.**29**:233-237.

Khor, CC., Vannberg, FO., Chapman, SJ., Walley, A., Aucan, C., Loke, H., White, NJ., Peto, T., Khor, LK., Kwiatkowski, D., Day, N., Scott, A., Berkley, JA., Marsh, K., Peshu, N., Maitland, K., Williams, TN and Hill AV(2007). Positive replication and linkage disequilibrium mapping of the chromosome 21q22.1 malaria susceptibility locus.*Genes Immun*. **8**:570-576.

Kidgell, C., Volkman, SK., Daily, J., Borevitz, JO., Plouffe, D., Zhou, Y., Johnson, JR., Le Roch, K., Sarr, O., Ndir, O., Mboup, S., Batalov, S., Wirth, DF and Winzeler, EA (2006). A systematic map of genetic variation in Plasmodium falciparum. PLoS Pathog. 2:e57.

Kidson, C., Lamont, G., Saul, A and Nurse GT (1981). Ovalocytic erythrocytes from Melanesians are resistant to invasion by malaria parasites in culture. *Proc Natl Acad Sci U S A*. **78**:5829-5832.

Kimura, M and Ahta, T (1978). Stepwise mutation model and distribution of allele frequencies in a finite population. *Proc. Natl. Acad. Sci. USA*. **75**: 2868-2872.

Kimura, M,. (1983) Rare variant alleles in the light of the neutral theory. *Mol Biol Evol. 1*:84-93.

Knight, JC., Udalova, I., Hill, AV., Greenwood, BM., Peshu, N., Marsh, K and Kwiatkowski D (1999a). A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. *Nat Genet*. **22**:145-150.

Knight, JC., Udalova, I., Hill, AV., Greenwood, BM., Peshu, N., Marsh, K, and Kwiatkowski D. (1999b). A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. *Nat Genet*. **22.**,145-150.

Koch, O., Awomoyi, A., Usen, S., Jallow, M., Richardson, A., Hull, J., Pinder, M., Newport, M and Kwiatkowski D (2002) IFNGR1 gene promoter polymorphisms and susceptibility to cerebral malaria. *J Infect* Dis.**185**:1684-1687.

Kun, JF., Klabunde, J., Lell, B., Luckner, D., Alpers, M., May, J., Meyer, C and Kremsner, PG (1999). Association of the ICAM-1Kilifi mutation with protection against severe malaria in Lambarene, Gabon. *Am J Trop Med Hyg*. **61**:776-779.

Kun, JF., Mordmuller, B., Perkins, DJ., May, J., Mercereau-Puijalon, O., Alpers, M., Weinberg, JB and Kremsner, PG (2001). Nitric oxide synthase 2 (Lambarene) (G-954C), increased nitric oxide production, and protection against malaria. *J Infect Dis*.**184**:330-336.

Kwiatkowski D. (2000). Genetic susceptibility to malaria getting complex. *Curr Opin Genet Dev*. **10**:320-324.

Kwiatkowski DP (2005). How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet*. **77**:171-192.

Logan, KE., Chambers, MA., Hewinson, RG and Hogarth, PJ.(2005). Frequency of IFN-gamma producing cells correlates with adjuvant enhancement of bacille Calmette-Guèrin induced protection against Mycobacterium bovis. *Vaccine*. **23**:5526-5532.

Lohmueller, KE., Pearce, CL., Pike, M., Lander, ES and Hirschhorn, JN (2003). Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet*. **33**:177-182.

Luoni, G., Forton, J., Jallow, M., Sadighi, AE., Sisay-Joof, F., Pinder, M and Hanchard, N., Herbert, M., Kimber, M., Mott, R., Hull, J., Rockett, K and Kwiatkowski D (2005). Population-specific patterns of linkage disequilibrium in the human 5q31 region.*Genes Immun*_**6:**723-727.

Mackinnon, MJ., Mwangi, TW., Snow, RW., Marsh, K and Williams TN (2005) Heritability of malaria in Africa. *PLoS Med* **2**:e340.

Marquet, S., Abel, L., Hillaire, D., Dessein, H., Kalil, J., Feingold, J., Weissenbach, J and Dessein AJ (1996). Genetic localization of a locus controlling the intensity of infection by Schistosoma mansoni on chromosome 5q31-q33. *Nat Genet*. **14:**181-184.

McGuire, W., D'Alessandro, U., Stephens, S., Olaleye, BO., Langerock, P., Greenwood, BM and Kwiatkowski, D (1998). Levels of tumour necrosis factor and soluble TNF receptors during malaria fever episodes in the community. *Trans R Soc Trop Med Hyg*. **92**:50-53.

McGuire, W., Hill, AV., Allsopp, CE., Greenwood BM and Kwiatkowski D (1994) Variation in the TNF-alpha promoter region associated with susceptibility to cerebral malaria. *Nature* **371**:508-510.

Miller, KH., Mason, SJ., Clyde, DF and McGinnis, MH (1976). The resistance factor to *Plasmodium vivax* in Black the Duffy-blood group genotype Fζ Fζ. *N. Engl. J. Med.* **295**: 302-304.

Mockenhaupt, FP., Cramer, JP., Hamann, L., Stegemann, MS., Eckert, J., Oh, NR., Otchwemah, RN., Dietz, E., Ehrhardt, S., Schroder, NW., Bienzle, U and Schumann, RR (2006). Toll-like receptor (TLR) polymorphisms in African children: Common TLR-4 variants predispose to severe malaria. *Proc Natl Acad Sci. USA*. **103**:177-182.

Modiano, D., Luoni, G., Sirima, BS., Lanfrancotti, A., Petrarca, V., Cruciani, F., Simpore, J., Ciminelli, BM., Foglietta, E., Grisanti, P., Bianco, I., Modiano, G and Coluzzi, M (2001). The lower susceptibility to Plasmodium falciparum malaria of Fulani of Burkina Faso (west Africa) is associated with low frequencies of classic malaria-resistance genes. *Trans R Soc Trop Med Hyg*. **95**:149-152.

Modiano, D., Luoni, G., Sirima, BS., Simpore, J., Verra, F., Konate, A., Rastrelli, E., Olivieri, A., Calissano, C., Paganotti, GM., D'Urbano, L., Sanou, I., Sawadogo, A., Modiano, G and Coluzzi, M (2001). Haemoglobin C protects against clinical *Plasmodium falciparum* malaria. *Nature* **414**:305-308.

Moormann, AM., Embury, PE., Opondo, J., Sumba, OP., Ouma, JH., Kazura, JW and John, CC (2003). Frequencies of sickle cell trait and glucose-6-phosphate dehydrogenase deficiency differ in highland and nearby lowland malaria-endemic areas of Kenya. *Trans R Soc Trop Med Hyg*. **97:**513-514.

Murray, SS., Oliphant, A., Shen, R., McBride, C., Steeke, RJ., Shannon, SG., Rubano, T., Kermani, BG., Fan, JB., Chee, MS and Hansen MS (2004). A highly informative SNP linkage panel for human genetic studies. *Nat Methods*. **1**:113-137.

Nasr, A., Elghazali, G., Giha, H., Troye-Blomberg, M and Berzins, K.(2008) Interethnic differences in carriage of haemoglobin AS and Fcgamma receptor IIa (CD32) genotypes in children living in eastern Sudan. *Acta Trop.***105**:191-195.

Nasr, A., Iriemenam, NC., Troye-Blomberg, M., Giha, HA., Balogun, HA., Osman, OF., Montgomery, SM., ElGhazali, G and Berzins, K (2007). Fc gamma receptor IIa (CD32) polymorphism and antibody responses to asexual blood-stage antigens of Plasmodium falciparum malaria in Sudanese patients. *Scand J Immunol*. **66**:87-96.

Nicholson, G., Smith, AV., Jonsson, F., Gustafsson, O., Stefansson, K and Donnelly P (2002). Assessing population differentiation and isolation from single nucleotide polymorphism data. J. R. Statist. Soc. 64, 695-715.

Ochieng, WR (1985) History of Kenya. Macmillan Education, pg 34.

Ockenhouse, CF., Ho, M., Tandon, NN., Van Seventer, GA., Shaw, S., White, NJ., Jamieson, GA., Chulay, JD and Webster HK (1991). Molecular basis of sequestration in severe and uncomplicated Plasmodium falciparum malaria: differential adhesion of infected erythrocytes to CD36 and ICAM-1.J *Infect Dis*.**164**:163-169.

Omi, K., Ohashi, J., Patarapotikul, J., Hananantachai, H., Naka, I., Looareesuwan, S and Tokunaga K (2003). CD36 polymorphism is associated with protection from cerebral malaria. *Am. J. Hum. Genet.*72: 364-374.

Pain, A., Urban, BC., Kai, O., Casals-Pascual, C., Shafi, J., Marsh, K and Roberts DJ (2001) A non-sense mutation in Cd36 gene is associated with protection from severe malaria. *Lancet*.**357**:1502-1503.

Pasvol G (2001). Cell-cell interaction in the pathogenesis of severe falciparum malaria *Royal College of Physicians*. 1: 495-500.

Patil, N., Berno, AJ., Hinds, DA., Barrett, WA., Doshi, JM., Hacker, CR., Kautzer, CR., Lee, DH., Marjoribanks, C., McDonough, DP., Nguyen, BT., Norris, MC., Sheehan, JB., Shen, N., Stern, D., Stokowski, RP., Thomas, DJ., Trulson, MO., Vyas, KR., Frazer, KA., Fodor, SP and Cox, DR (2001). Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science*. **294**:1719-1723.

Patterson, N., Price, AL and Reich, D (2006) Population Structure and Eigenanalysis. *PLoS Genet*. **2**:e190.

Pearce, EJ and Schofield, L (2007) Host-pathogen interactions: pas de deux. *Curr Opin Immunol.* 19:373-375.

Petretto, E., Liu, ET and Aitman TJ (2007) A gene harvest revealing the archeology and complexity of human disease.*Nat Genet*. **39**:1299-1301.

Reich, DE., Cargill, M., Bolk, S., Ireland, J., Sabeti, PC., Richter, DJ., Lavery, T., Kouyoumjian, R., Farhadian, SF., Ward, R and Lander ES. (2001a). Linkage disequilibrium in the human genome. *Nature.* **411**:199-204.

Reich, DE., Goldstein and DB (2001b) Detecting association in a case-control study while correcting for population stratification. *Genet Epidemiol*. **20**:4-16.

Rihet, P., Traoré, Y., Abel, L., Aucan, C.,Traoré-Leroux, T and Fumoux, F (1998) Malaria in humans: Plasmodium falciparum blood infection levels are linked to chromosome 5q31-q33.*Am J Hum Genet*. **63**::498-505.

Riley, EM., Olerup, O., Bennett, S., Rowe, P., Allen, SJ., Blackman, MJ., Troye-Blomberg, M., Holder, AA and Greenwood, BM.(1992). MHC and malaria: the relationship between HLA class II alleles and immune responses to Plasmodium falciparum. *Int Immunol.* **4**:1055-1063.

Roeder, K., Bacanu, SA., Wasserman, L and Devlin B (2006). Using linkage genome scans to improve power of association in genome scans. *Am. J. Hum. Genet.* **78**:243-252.

Rottmann, M., Lavstsen, T., Mugasa, JP., Kaestli, M., Jensen, AT., Muller, D., Theander, T and Beck, HP (2006). Differential expression of var gene groups is associated with morbidity caused by Plasmodium falciparum infection in Tanzanian children. *Infect Immun* . **74**:3904-3911.

Ruwende, C., Khoo, SC., Snow, RW., Yates, SN., Kwiatkowski, D., Gupta, S., Warn, P., Allsopp, CE., Gilbert, SC and Peschu N, *et al.*(1995). Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. *Nature*.**376**: 246-249.

Sabeti, P., Usen, S., Farhadian, S., Jallow, M., Doherty, T., Newport, M., Pinder, M., Ward, R and Kwiatkowski, D. (2002a) CD40L association with protection from severe malaria. *Genes Immun*.**3**:286-291.

Sabeti, PC., Reich, DE., Higgins, JM., Levine, HZ., Richter, DJ., Schaffner, SF., Gabriel, SB., Platko, JV., Patterson, NJ., McDonald, GJ., Ackerman, HC., Campbell, SJ., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R and Lander, ES (2002b). Detecting recent positive selection in the human genome from haplotype structure. *Nature.* **419**:832-837.

Saravia, C., Martinez, P., Granados, DS., Lopez, C., Reyes, C,. Patarroyo, MA., (2008). Identification and evaluation of universal epitopes in Plasmodium vivax Duffy binding protein. *Biochem Biophys Res Commun.*;**377**:1279-83

Schofield, L (2007). Rational approaches to developing an anti-disease vaccine against malaria. *Microbes Infect.* 9:784-791

Segal, S and Hill, AV (2003). Genetic susceptibility to infectious disease. *Trends Microbiol.* **11**:445-448.

Shi, YP., Nahlen, BL., Kariuki, S., Urdahl, KB., McElroy, PD., Roberts, JM and Lal, AA (2001). Fcgamma receptor IIa (CD32) polymorphism is associated with protection of infants against high-density Plasmodium falciparum infection. VII. Asembo Bay Cohort Project. *J Infect Dis.* **184:**107-111.

Specht, S., Sarite, SR., Hauber, I., Hauber, J., Görbig, UF., Meier, C., Bevec, D., Hoerauf, A and Kaiser A.(2008). The guanylhydrazone CNI-1493: an inhibitor with dual activity against malaria-inhibition of host cell pro-inflammatory cytokine release and parasitic deoxyhypusine synthase. *Parasitol Res.* (Epub ahead of print)

Stephens, JC., Reich, DE., Goldstein, DB., Shin, HD., Smith, MW., Carrington, M., Winkler, C., Huttley, GA., Allikmets, R., Schriml, L., Gerrard, B., Malasky, M., Ramos, MD., Morlot, S., Tzetis, M., Oddoux, C., di Giovine, FS., Nasioulas, G., Chandler, D., Aseev, M., Hanson, M., Kalaydjieva, L., Glavac, D., Gasparini, P and Dean M (1998). Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. *Am J Hum Genet.* **62**:1507-1515.

Stoute, JA and Ballou WR.(1998). The current status of malaria vaccines. *BioDrugs.* **10**:123-36.

Tang, K., Fu, DJ., Julien, D., Braun, A., Cantor, CR and Koster, H (1999). Chip-based genotyping by mass spectrometry. *Proc Natl Acad. Sci USA* **96**:10016-10020.

Terlouw, DJ., Aidoo, MA., Udhayakumar, V., Kolczak, MS., Oloo, AJ., Kager, PA., Lal, AA., Nahlen, BL and Kuile FO (2002) Increased efficacy of sulfadoxine-pyrimethamine in the treatment of uncomplicated falciparum malaria among children with sickle cell trait in Western Kenya. *J Infect Dis*. **186:**1661-16618.

The International HapMap Consortium (2005). A haplotype map of the human genome. *Nature.* **437**: 1299–1320.

The International HapMap Project (2004). A haplotype map of the human genome. *Nature*. **426**:789-796.

Tian, JH., Miller, LH., Kaslow, DC., Ahlers, J., Good, MF., Alling, DW., Berzofsky, JA and Kumar, S. (1996) Genetic regulation of protective immune response in congenic strains of mice vaccinated with a subunit malaria vaccine. *J Immunol*. **157**:1176-1183.

Tishkoff, SA and Williams, SM. (2002). Genetic analysis of African populations: human evolution and complex disease. *Nat Rev Genet*. **3**: 611-621.

Tishkoff, SA., Reed, FA., Ranciaro, A., Voight, BF., Babbitt, CC., Silverman, JS., Powell, K., Mortensen, HM., Hirbo, JB., Osman, M., Ibrahim, M., Omar, SA., Lema, G., Nyambo, TB., Ghori, J., Bumpstead, S., Pritchard, JK., Wray, GA and Deloukas, P (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* **39**:31-40.

Tishkoff, SA., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J., Piro, A., Stoneking, M., Tagarelli, A., Tagarelli, G., Touma, EH., Williams, SM and

.**Tournamille, C., Colin, Y., Cartron, JP and Le Van Kim, C (1995).** Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat Genet.*10:224-228.

Vallender, EJ and Lahn, BT (2004) Positive selection on the human genome. Hum. *Mol. Genet.* **13**: R245-254.

Verrelli, BC., McDonald, JH., Argyropoulos, G., Destro-Bisol, G., Froment, A., Drousiotou, A., Lefranc, G., Helal, AN., Loiselet, J and Tishkoff, SA (2002). Evidence for balancing selection from nucleotide sequence analyses of human G6PD.*Am J Hum Genet.***71**:1112-1128.

Volkman, SK., Sabeti, PC., DeCaprio, D., Neafsey, DE., Schaffner, SF., Milner, DA., Jr, Daily, JP., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Duraisingh, MT., Lukens, A., Derr, A., Stange-Thomann, N., Waggoner, S., Onofrio, R., Ziaugra, L., Mauceli, E., Gnerre, S., Jaffe, DB., Zainoun, J., Wiegand, RC., Birren, BW., Hartl, DL., Galagan, JE., Lander, ES and Wirth DF (2007). A genome-wide map of diversity in *Plasmodium falciparum*. *Nat Genet* **39**:113-139.

Vulliamy, TJ., Othman, A., Town, M., Nathwani, A., Falusi, AG., Mason, PJ and Luzzatto, L (1991). Polymorphic sites in the African population detected by sequence analysis of the Glucose-6-Phosphate Dehydrogenase gene outline the evolution of the variants A and A-.*Proc Natl Acad Sci USA*.**88**: 8568-8571.

Wattavidanage, J., Carter, R., Perera, KL., Munasingha, A., Bandara, S., McGuinness, D., Wickramasinghe, AR., Alles, HK., Mendis, KN and Premawansa, S. (1999) TNFalpha*2 marks high risk of severe disease during *Plasmodium falciparum* malaria and other infections in Sri Lankans. *Clin Exp Immunol*. **115**:350-355.

Weale, ME., Depondt, C., Macdonald, SJ., Smith, A., Lai, PS., Shorvon, SD., Wood, NW and Goldstein, DB (2003). Selection and evaluation of tagging SNPs in the neuronal-sodium-channel gene SCN1A: Implications for linkage disequilibrium gene mapping. *Am J Hum Genet* **73**:551-565.

Williams, TN., Mwangi, TW., Roberts, DJ., Alexander, ND., Weatherall, DJ., Wambua, S., Kortok, M., Snow, RW and Marsh, K (2005). An immune basis for malaria protection by the sickle cell trait. *PLoS Med* **2**:e128.

World Health Organization (2000). Severe falciparum malaria. *Trans R soc Trop Med Hyg*.**94**:1-90.

Zakeri, S., Lindergard, G., Davies, RM., Boudin, C., Louis, F and Hommel M (2006). Identification and typing of Cameroonian isolates of P. malariae using monoclonal anti-bodies against P. brasilianum. *Acta Trop*. **99**: 97-101

LIST OF APPENDICES

APPENDIX A: DATA SETS FOR CASE-CONTROL ASSOCIATIONS AND DETAILS OF PREVIOUSLY PUBLISHED ASSOCIATIONS OF 10 GENES SELECTED IN THIS STUDY.

Appendix A Table 1: Result of the replication test.

Genotype / allele previously associated to malaria susceptibility	Reference SNP ID	Genotype / allele frequency in controls	Cases / controls	Odds ratio (95% confidence interval)	*P*-value of test of association to the previously associated genotype or allele
HbAS	rs334	24%	447/454	0.57 (0.41-0.79)	0.0004
CD36-GT	rs3211938	12%	456/457	1.5 (1.03-2.18)	0.015
ICAM-TT	rs5491	7%	460/455	0.71 (0.42-1.21)	0.10
NOS2A-1659-AA	rs8078340	6%	450/455	0.42 (0.21-0.83)	0.99
TNF-238 A allele	rs361525	10%	459/457	1.0 (0.73-1.39)	0.49
CD32-AA	rs1801274	25%	454/450	0.95 (0.71-1.29)	0.38
IFNARI-LI168V -CC	rs2257167	3%	455/457	1.18 (0.54-2.07)	0.76
TNF-308-A allele	rs1800629	9%	450/433	1.13 (0.82-1.56)	0.21
IFNARI-17470-CC	rs1012335	10%	451/452	0.85 (0.53-1.36)	0.34
TLR-4-AG	rs4986790	10%	407/303	1.36 (0.85-2.17)	0.10

Appendix A Table 2: Full set of statistical tests for association (expansion of Table 4.2 in main text)

Genotype / allele name	**Reference SNP ID**	**Genotype/Allele frequency in Luo controls**	**Cases / controls**	**Odds ratio (95% Confidence Interval)**	**Tests of previously associated genotype: 2-sided (1-sided)**
HbAS-AT	rs334	24%	447/454	0.57 (0.41-0.79)	0.0008 (0.004)
CD36-GT	rs3211938	12%	456/457	1.5 (1.03-2.15)	0.03 (0.015)
ICAM-TT	rs5491	7%	460/455	0.71 (0.42-1.21)	0.20 (0.10)
NOS2A-1659-AA	rs8078340	6%	450/455	0.44 (0.23-0.83)	0.01 (0.99)
TNF-238 A allele	rs361525	10%	459/457	1.0 (0.73-1.39)	NA
CD32-AA	rs1801274	25%	454/450	0.95 (0.71-1.29)	0.76 (0.38)
IFNARI-LI168V-CC	rs2257167	3%	455/457	1.18 (0.54-2.07)	0.48 (0.76)
TNF-308-A allele	rs1800629	9%	450/433	1.13 (0.82-1.56)	NA
IFNARI-17470-CC	rs1012335	10%	451/452	0.85 (0.53-1.35)	0.68 (0.34)
TLR-4-AG	rs4986790	10%	407/303	1.36 (0.86-2.13)	0.20 (0.1)

Appendix A. Table 3: Data sets for case-control association analysis

Variants	Genotypes	Cases	Controls
HbAS-AT	TT	2	3
	AT	71	113
	AA	374	338
CD36-GT	GG	0	1
	GT	79	56
	TT	377	401
ICAM-TT	AA	253	262
	AT	182	159
	TT	25	34
NOS2A1659-AA	AA	12	28
	AT	163	140
	TT	275	287
TNF238-A	AA	2	5
	AG	77	71
	GG	380	381
CD32-AA	AA	111	114
	AT	231	226
	TT	112	110
IFNARI-LI168V-CC	CC	14	12
	CG	120	126
	GG	321	319
TNF308-A	AA	4	7
	AG	79	61
	GG	367	365
IFNARI-LI168V-CC	CC	14	12
	CG	120	126
	GG	321	319
TLR4-AG	AA	350	270
	AG	55	31
	GG	2	0

Appendix A. Table 4: Previous published association tests for 10 genes tested in this study.

Genotype / allele*	Direction of association	Severe cases	Mild Cases	Controls	Frequency in controls	Population	*P*-value	Odds/ risk**-ratio (95% Confidence Interval)
HbAS (AT) (Aidoo *et al.*,2002)	Protection[a]	1022 children cohort 2-16months			NA	Kenya	0.0001	0.45** (0.24-0.84)
HbAS (AT) (Modiano *et al.*,2001)	Protection[a]	359	476	3,513	10%	Burkina Faso	0.001	0.27(0.17-0.42)
CD36 (GT) (Aitman *et al.*,2000)	Risk[b]	97	0	430	NA	Gambia	0.01	1.5 (1.0-2.3)
			0	331	NA	Kenya, Kilifi		
ICAM1 kilifi (TT) (Kun *et al.*,1999)**Error! Bookmark not defined.**	Protection[a]	100	0	100	6%	Gabon	0.027	0.18 (0.04-0.86)
NOS2A 1659 (TT) (Burgner *et al.*,1999)	Risk[b]	246	0	259	10%	Gambia	0.03	1.8 (1.02-3.07)
*TNF-238A (Knight *et al.*,1999)**Error! Bookmark not defined.**	Risk[d]	193	349	371	6%	Gambia	0.02	1.9 (1.1-3.1)
**TNF-238A (McGuire* et al., *1994)*	Risk[b]	257	0	311	NA	Kenya, Kilifi	0.01	0.2 (0.08-0.71)
CD32 Arg/Arg 131 (Shi *et al.*,2001)	Protection[e]	320	132	121	34%	Kenya, Luo	0.0021	0.28 (0.12-0.63)
CD32 Arg/Arg 131(Cooke et al.,2003)	Protection[d]	353	524	558	50%	Gambia	0.03	0.71 (0.52-0.98)
IFNAR1-LI68V (GG)(Aucan *et al.*,2003)***Error! Bookmark not defined.***	Protection[b]	528	338	554	6%	Gambia	0.031	0.76 (0.59-0.97)
*TNF*308(AA) (Knight *et al.*,1999)	Risk[b]	384	0	371	NA	Gambia	0.039	3.6 (1.1-12.4)
*TNF308A (Flori *et al.*,2005)	Risk [c]	197 Individuals in 34 families			NA	Burkina Faso	0.03	NA
	Risk[e]						0.0005	NA
IFNAR1-17470(GG)(Aucan *et al.*,2003) (GG)	Protection[b]	528	338	554	45%	Gambia	0.018	0.74 (0.57-0.95)
TLR-4 (AG) [Mockenhaupt *et al.*,2006)	Risk[a]	290	0	290	22%	Ghana	0.049	1.53 (1.0-2.3)

* Published association was to the allele rather than to the genotype

Published association was to: [a]Severe malaria, [b]Cerebral malaria, [c]Mild malaria, [d]Severe malaria anemia or Parasitemia[e]

APPENDIX B. DATA SETS FOR POPULATION DIFERENTIATION TESTS.

ppendix B Table 5: Tests for differentiating selection between endemic and non-endemic populations\

	Luo controls		Luo cases		Masai		Kikuyu		Yoruba	
Alleles	ref	var	ref	var	ref	var	Ref	var	Ref	var
HbAS-T	789	119	819	75	194	0	200	0	Allele's name	11
CD36-G	856	58	883	79	184	2	202	0	78	22
ICAM-T	683	227	688	232	156	30	169	37	76	24
NOS2A-1659-A	714	196	713	187	141	47	162	42	79	19
TNF-238-G	833	81	837	81	152	40	169	33	99	1
CD32-A	454	446	453	455	95	95	92	118	49	49
IFNARI-LI168V-C	764	150	762	148	144	48	168	40	98	16
TNF-308-A	791	75	813	87	176	12	189	15	90	6
IFNARI-17470-C	616	288	622	280	128	62	131	71	84	24
TLR-4-G	571	31	755	59	165	13	161	9	109	5

Note:-ref. is the ancestral allele and var. is the derived allele

APPENDIX C: REVIEW OF PREVIOUS ASSOCIATIONS OF OTHER GENETIC VARIANTS INCLUDING THE 10 TESTED IN THIS STUDY.

Appendix C Table 6: Previous association studies for susceptibility to severe *P. falciparum* malaria

Genotype	Severe cases	Mild Cases	Controls	Genotype frequency in controls	Population	P-value	Odds-ratio (OR)	95% Confidence Interval for OR
HbCC	359	476	3,513	0.02	Burkina Faso	0.0011	0.07	0 – 0.48
HbAC	359	476	3,513	0.23	Burkina Faso	0.0008	0.71	0.58 – 0.87
HbAS	359	476	3,513	0.1	Burkina Faso	0.01	0.27	0.17 – 0.42
HbAS	69	45	42	0.17	Kenya	0.0004	0.04	0.02 - 0.37
G6PD-A females	255	0	182	0.14	Gambia	0.02	0.45	0.22 – 0.42
G6PD-A females	133	0	143	0.27	Kenya	0.11	0.6	0.33-1.11
G6PD-A hemizygous males	279	0	239	0.06	Gambia	0.02	0.23	0.06 – 0.77
G6PD-A hemizygous males	117	0	149	0.19	Kenya	0.12	0.54	0.25 – 1.15
NOS2A-117 C→T homozygote (TT)	134	0	45	0.2	Tanzania	0.0006	0.12	0.03 – 0.048
NOS2A-117 C→T homozygote (TT)	144	0	916	0.19	Kenya	0.012	0.31	0.12-0.84
NOS2A-117 C→T heterozygote (CT)	134	0	45	0.3	Kenya	0.005	0.25	0.09 – 0.66
NOS2A 1659T heterozygote	246	0	259	0.41	Gambia	0.04	1.31	1.0 – 1.72
NOS2A -969(G→C)	100	100	0	0.3	Gabon	0.04	0.67	0.46-0.96
ICAM1 kilifi heterozygote	100	0	100	0.42	Gabon	0.012	0.52	0.30 – 0.92
ICAM1 kilifi homozygote	100	0	100	0.06	Gabon	0.027	0.18	0.04 – 0.86

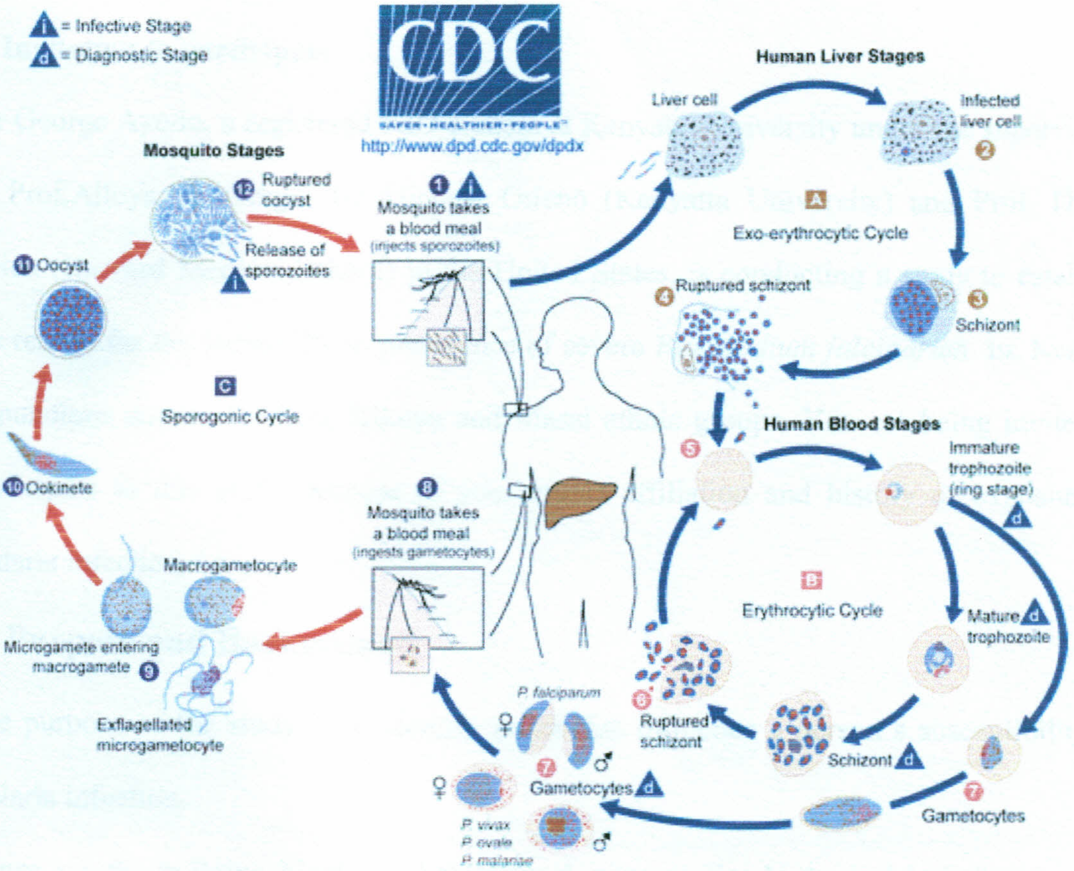
CD32 Arg/Arg 131	320	132	121	0.34	Kenya	0.0021	0.28	0.12 – 0.63
CD32 Arg/Arg 131	353	524	558	0.5	Gambia	0.03	0.71	0.52 – 0.98
CD36 1264T →G Homozygotes	272	0	203	0.32	Thailand	0.0069	0.59	0.40 – 0.87
CD36 T188G heterozygote	693	0	693	0.18	Kenya	0.036	0.74	0.55 – 0.99
IFNGR-56 Heterozygotes	562	0	569	0.51	Gambia	0.016	0.54	*
CD40L-726C female carriers	298	0	142	0.64	Gambia	0.002	0.74	0.48-1.15
CD40L-726C Male hemizygotes	332	0	129	0.5	Gambia	$P<0.05$	0.52	0.34-0.81
*TNF*308A Homozygote	384	0	371	*	Gambia	0.039	3.6	1.1-12.4
*TNF*308A Homozygote	257	0	311	*	Kenya	0.16	3.3	0.62-17.6
TNF-238 Heterozygote	193	349	371	0.12	Gambia	0.04	1.7	1.02-2.8
TNF-376A heterozygote	384	0	371	0.015	Gambia	0.008	4.3	1.5-12.8
*TNFα**,*2 Heterozygote	35	116	84	0.7	Sri-Lanka	0.013	2.76	*
IFNAR1-17470G/G homozygotes	528	338	554	0.45	Gambia	0.018	0.74	0.57-0.95
IFNAR1-LI68V Homozygote	528	338	554	0.06	Gambia	0.031	0.76	0.59-0.97

* Data could not be found in the published article

† References for these association studies are given in the text.

APPENDIX D: THE LIFE CYCLE OF MALARIA PARASITE

Appendix D Figure 1: The life cycle of malaria parasite.



The malaria parasite life cycle involves two hosts. During a blood meal, a malaria-infected female *Anopheles* mosquito inoculates sporozoites into the human host ①. Sporozoites infect liver cells ②and mature into schizonts ③, which rupture and release merozoites ④. (Of note, in *P. vivax* and *P. ovale* a dormant stage [hypnozoites] can persist in the liver and cause relapses by invading the bloodstream weeks, or even years later.) After this initial replication in the liver (exo-erythrocytic schizogony A), the parasites undergo asexual multiplication in the erythrocytes (erythrocytic schizogony B). Merozoites infect red blood cells ⑤. The ring stage trophozoites mature into schizonts, which rupture releasing merozoites ⑥. Some parasites differentiate into sexual erythrocytic stages (gametocytes) ⑦. Blood stage parasites are responsible for the clinical manifestations of the disease. The gametocytes, male (microgametocytes) and female (macrogametocytes), are ingested by an *Anopheles* mosquito during a blood meal ⑧. The parasites' multiplication in the mosquito is known as the sporogonic cycle C. While in the mosquito's stomach, the microgametes penetrate the macrogametes generating zygotes ⑨. The zygotes in turn become motile and elongated (ookinetes) ⑩ which invade the midgut wall of the mosquito where they develop into oocysts ⑪. The oocysts grow, rupture, and release sporozoites ⑫, which make their way to the mosquito's salivary glands. Inoculation of the sporozoites ①into a new human host perpetuates the malaria life cycle.

APPENDIX E: RESEARCH AUTHORIZATION DOCUMENTS FOR THIS RESEARCH

Appendix E Part 1 Consent form

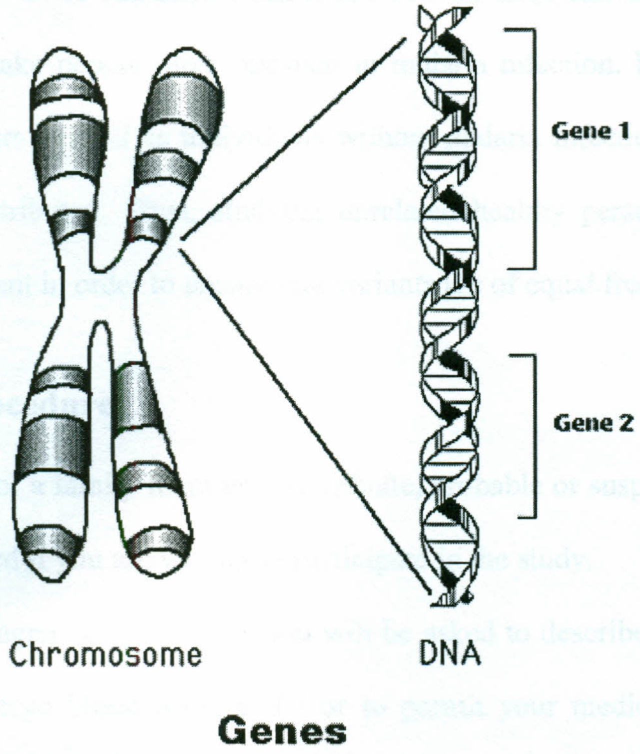
a) Invitation to participate

Mr George Ayodo, a registered PhD student at Kenyatta University under the supervision of Prof.Alloys SS Orago, Dr Micheal Otieno (Kenyatta University) and Prof. David Reich (Harvard Medical School) in the United States, is conducting a study to establish the reason for the variability in prevalence of severe *Plasmodium falciparum* in Kenyan populations among the Luo, Kikuyu and Masai ethnic groups. You are being invited to participate in this study because of your ethnic affiliation and history of exposure to malaria infection.

b) Purpose and Background

The purpose of the study is to identify genes that influence a person's susceptibility to malaria infection.

Genes are the building blocks of heredity. A gene carries biological information in a form that must be copied and transmitted from each cell to its offspring. Genes are made of DNA and are located on chromosomes. They are inherited in pairs, one from each parent. As you can see in the figure below, DNA is shaped like a twisted stepladder, the famous "double helix". The genetic information is carried on the rungs of the ladder.



It appears that resistance to malaria infection has a complex pattern of inheritance, as a single gene does not cause it., several genes or combinations of genes may make an individual more protected against the disease. At this time, the identities of only some of these genes are known. A vast body of previous research shows that there is a strong genetic influence on the development of malaria. In fact some families have more than one member with severe malaria, and persons that have an identical twin with severe malaria are at an increased risk of also developing the disease.

A vast body of previous research shows that there is a strong genetic influence on malaria infection. Several genes have been found to be associated with malaria susceptibility. In order to identify which genes may be involved., large numbers of people are being studied. DNA extracted from blood cells or from saliva is analyzed to find genes which may make people more resistant to malaria infection. It is important to screen affected members as well as individuals without malaria infection to determine how these genes are distributed. Thus, studying unrelated healthy persons without malaria infection is important in order to ensure that variants are of equal frequency in cases and controls.

c) Procedure

If you or a family member has definite, probable or suspected malaria infection, you may be asked if you are willing to participate in the study.

If you agree to participate, you will be asked to describe your symptoms in an interview, to undergo blood tests, and / or to permit your medical records to be reviewed by a member of the study team. An interview, lasting approximately fifteen minutes, consists of a series of questions related to your medical and family history as it pertains to malaria infection or to your physical signs. Parasite count and hemoglobin tests will be carried out to determine the severity of your infection. Your medical records will be used to verify your diagnosis for study purposes only, and will be reviewed exclusively by two people: the investigating scientist and the study coordinator. Although your DNA sample will be analyzed for this study in laboratories in Kenya and in the United States, no researchers except for Mr. George Ayodo will be able to connect your name to the randomized identification number that will go with the sample. The key connecting your name to that number will be kept in a locked safe.

It is possible, although highly unlikely, that review of your medical records will disclose a feature of your malaria infection that was previously undetected. If this occurs, I will speak to your primary care physician to obtain more information. Of course, any new information I find will be shared with your physician and thus contribute to your care.

Your will be asked to provide contact information (address and phone numbers) and demographic information such as my date of birth and ethnicity. You may also be contacted periodically (by mail or phone) by a member of the study team to inquire of any changes in clinical status or for other relevant items that may have occurred.

You will be asked to donate a small sample of blood, approximately five tablespoons (75 milliliters), through a vein in my arm. This will take place either at the hospital's premises or at the site of investigation. The actual blood draw will take about five minutes to perform.

Biological materials obtained from blood cells (DNA, RNA, serum, plasma) may be shared in the future with other qualified researchers.

d) Risks/Discomforts/Confidentiality

Blood drawing could cause some temporary discomfort from the needle stick, bruising and/or rarely, infection at the puncture site

While participation in research also entails some loss of privacy, your records and other information that you share with the investigators will be handled in a confidential manner. Your blood donation will be coded with a number and any data obtained from the study will never have my name or any identifiable characteristics attached to it that can be traced back to you. Thus, except as part of the initial data gathering, you will not receive any results from the analysis that pertain to me as an individual.

e) Economic Consideration

You will not receive any monetary or other compensation for my participation in this study. There will be no charges for participation in this study.

f) Benefit

There is no direct benefit to you from my participation in this study. However, the knowledge that is gained from the study will help the investigators learn more about how malaria infection is inherited and may help to develop improved treatments for the disease in the future.

g) Alternatives

You may choose not to participate in this study or withdraw at any time. If I do so, this will not have any consequence on the care you receive.

h) INFORMED CONSENT FORM

Questions

This study has been explained to me by a member of the investigating team. If I have further questions, comments or concerns I can call George Ayodo on **Telephone, 0735 568114, P.O.Box 54, NYILIMA, WEST ASEMBO LOCATION, RARIEDA DIVISION, BONDO DISTRICT.**

If for some reason I do not wish to speak to the investigating team, I may contact the Kenyatta University ethical committee, which is concerned with the protection of volunteers in research projects. I may reach the committee office between 8:00 and 5:00, Monday through Friday, by calling (020) 810901, or by writing: Ethical Committee: P.O. Box 43844, Kenyatta University, Nairobi.

Consent

PARTICIPATION IN RESEARCH IS VOLUNTARY. I have the right to decline to participate and to withdraw at any point in this study without jeopardy to my continued medical care or my future interactions with the Kenyatta University.

I have been given a copy of this consent and a copy of the Experimental Subject's Bill of Rights to keep.

If I wish to participate I should sign below.

------------------- --

Date Signature of Study Participant

For participants below 18 years, parental or legal guardian permission is required

I wish to have my child participate in this study. Signing below gives my permission for my child to do so.

------------------- --

Date Signature of Parent or Legal Guardian

Person obtaining

Questionnaire

Note: George Ayodo will fill this questionnaire form while interviewing the child and the parent. Before the start of an interview, the content of the consent document will be discussed with the subject's parent. This will be followed by consenting.

a) Details of the subject

1) What is the child's name?---

2) How old is the child?---

3) What is the sex of the child?--

5) Where are the hospital records?--

6) Are all four grandparents from the same ethnic group as the child?-----------------

b) Physical address of the subject

1) Village, ---

2) Sub-location, ---

3) Location,--

4) District---

5) Nearest school?---

b) Retrieval of Genetic information. *The information obtained at this stage will be compared with the information obtained from the hospital records.*

1) How often do you bring your child to the hospital?------------------------------------

2) When did your child start experiencing this disease and how long has it lasted?------------------

----

3) Has your child previously ever had such a severe case of malaria------------------

4) How often have you taken your child to the hospital for this illness?-----------------------------

----

5) Have any of your other children had such severe malaria? ---------------------------

6) Are they twins? If yes has the other twin ever had a severe case of malaria?------

----

A) Clinical signs as described with the assistance of Hospital Physician:

- a) Unarousable coma (Blantyre score <3)--
- b) Impaired consciousness, ---
- c) Convulsions and respiratory distress---
- d) Clinical shock or severe prostration, --

e) History of convulsions,--

f) Extent of alertness--

B) <u>Laboratory tests results</u>:

1) Hypoglycemia glucose level ---g/dL

2) Parasite counts--- trophozoites / nL

3) Hemoglobin concentration---g/dL

Investigator's comments--

C) <u>Categorization</u>

1) Cerebral malaria ----------------------

3) Severe malarial anemia ----------------------

4) Mild or uncomplicated malaria ----------------------

5) Control ----------------------

Comments: by George Ayodo---

SIGNATURE--

Appendix E Part 2: Authorization letters

MINISTRY OF EDUCATION, SCIENCE AND TECHNOLOGY

legrams: "EDUCATION", Nairobi
ephone: Nairobi 334411
en replying please quote

JOGOO HOUSE "B"
HARAMBEE AVENUE
P.O. Box 30040-00100
NAIROBI

: No. MOEST 13/001/34C 159/2

24th January, 2005, 20......

George Ayodo
Kenyatta University
P.O. BOX 43844
NAIROBI

Dear Sir

RE: RESEARCH AUTHORISATION

Following your application for extension to conduct research on "Associating Genetic variation to Resistance to severe Malaria in Kenya", I am pleased to inform you that you have been authorised to conduct research in Bondo and Narok Districts for a period ending 31st June, 2005.

You are advised to report to the District Commissioner and the District Education Officers of the respective Districts of your research before embarking on your research project.

You are further expected to deposit two copies of your research findings to this Office upon completion of your research project.

Yours faithfully

B. O. ADEWA
FOR: PERMANENT SECRETARY

Cc
The District Commissioner
Bondo
Narok

The District Education Officer
Bondo
Narok

Recieved and permit to visit our schools:

DISTRICT EDUCATION OFFICER
BONDO DISTRICT
P.O. Box 580 BONDO
DATE 1/02/05

The Medical Officer of Health
Bondo
Narok Districts

APPENDIX G: PUBLISHED PAPERS AND MANUSCRIPTS

ARTICLE

Combining Evidence of Natural Selection with Association Analysis Increases Power to Detect Malaria-Resistance Variants

George Ayodo, Alkes L. Price, Alon Keinan, Arthur Ajwang, Michael F. Otieno, Alloys S. S. Orago, Nick Patterson, and David Reich

Statistical power to detect disease variants can be increased by weighting candidates by their evidence of natural selection. To demonstrate that this theoretical idea works in practice, we performed an association study of 10 putative resistance variants in 471 severe malaria cases and 474 controls from the Luo in Kenya. We replicated associations at *HBB* ($P = .0008$) and *CD36* ($P = .03$) but also showed that the same variants are unusually differentiated in frequency between the Luo and Yoruba (who historically have been exposed to malaria) and the Masai and Kikuyu (who have not been exposed). This empirically demonstrates that combining association analysis with evidence of natural selection can increase power to detect risk variants by orders of magnitude—up to $P = .000018$ for *HBB* and $P = .00043$ for *CD36*.

Malaria infection (MIM 248310) has exerted severe pressure on the human genome within the past 10,000 years,[1-3] and there are more cases today than ever before, with an estimated 300–660 million new episodes of clinical *Plasmodium falciparum* malaria every year.[4] Despite high infection rates, only 1%–2% of patients develop life-threatening complications, such as cerebral malaria and profound anemia,[5] so natural selection has likely operated, to a large extent, on severity. In the context of high infection rates, the genetics of host response are likely to play an important role.[6] In sub-Saharan Africa, the populations in which malaria is endemic generally have a lower proportion of cases with severe disease.[5,7] This suggests that there exist genetic variants that have risen to higher frequency in malaria-endemic populations because they modulate risk of *P. falciparum* malaria, similar to the case of the Duffy-null variant that protects against *P. vivax* malaria.[8]

A handful of genetic variants have already been associated with risk of or protection against severe malaria infection.[5] Our first objective in this study was to test variants of β-globin (*HbAS*[9,10]), intercellular adhesion molecule (*ICAM* TT[11]), *CD36* (*CD36* GT[12]), nitric oxide synthase (*NOS2A* 1659 AA[13]), tumor necrosis factor (*TNF* 238 A[14] and *TNF* 308 A[14-16]), Fc γ-receptor IIA (*CD32* AA[17,18]), interferon-α receptor-1 (*IFNARI* LI168V CC[19] and *IFNARI* 17470 CC[19]), and Toll-like receptor (*TLR4*[20]), which had previously been associated with malaria susceptibility. The particular phenotype we focused on was high levels of parasitemia in young children due to malaria infection.

Second, we compared the frequency differentiation in populations in which malaria is endemic and in closely related populations in which it is not endemic, searching for the differences that would be expected if natural selection had affected those alleles in one population but not in the other, because malaria began to affect only one group. Finally, we formally combined the evidence of association from case-control studies with evidence of natural selection in populations that have been exposed to malaria infection. We note that there has been discussion elsewhere of how one could formally combine case-control association studies with statistical weights obtained on the basis of evidence of natural selection.[21] Our goal in this study was to empirically demonstrate the power of this approach.

Material and Methods

Human Subjects

We collected 471 severe malaria cases and 474 controls from the Luo ethnic group, a population that speaks a Nilotic language and lives in a malaria-endemic region in western Kenya. All the severe malaria cases were collected from the Bondo District Hospital's children's emergency ward or from its outpatient clinic between May 2004 and August 2005. The average age of the cases was 2.6 years (table 1), reflecting our focus on individuals with no previous immunological protection against malaria. The controls were randomly collected from volunteers at nearby secondary schools, with an average age of 16.9 years (table 1). We focused on older controls, because we knew that they had survived to an older age. Thus, the control samples selected for this study may be slightly enriched for variants protecting against severe malaria, which should make it slightly easier to detect associations.

For the selection study, we assembled population control samples from the Masai, Kikuyu, and Yoruba ethnic groups. We collected samples from the Masai and Kikuyu from secondary schools in Narok and Nyeri, Kenya, respectively (table 1). The Yoruba samples were from the International Haplotype Map project[22]; we analyzed data from unrelated men and women, the parents in HapMap mother-father-child trios.

From the Department of Genetics, Harvard Medical School, and Broad Institute of Harvard and MIT (G.A; A.L.P; A.K; N.P; D.R), Boston; and Department of Pre-Clinical Sciences, Kenyatta University, Nairobi (G.A; A.A; M.F.O; A.S.S.O)

Received March 6, 2007; accepted for publication April 25, 2007; electronically published June 15, 2007.

Address for correspondence and reprints: Dr. David Reich, Harvard Medical School, Department of Genetics, 77 Avenue Louis Pasteur, New Research Building, Boston, MA 02115. E-mail: reich@genetics.med.harvard.edu

Am. J. Hum. Genet. 2007;81:234–242. 0002-9297/2007/8102-0005$15.00
DOI: 10.1086/519221

Table 1. Characteristics of the Populations Included in This Study

Population	Mean Age (Range)	No. in Sample (Male/Female)	Source	Malaria Endemicity	Altitude above Sea Level (m)
Luo cases	2.6 (1.5–10.0)	471 (232/239)	Bondo District Hospital, Kenya	Endemic	~1,240
Luo controls	16.9 (14–20)	474 (290/184)	Bondo schools, Kenya	Endemic	~1,240
Masai controls	16.9 (13–21)	97 (42/55)	Narok schools, Kenya	Nonendemic	~1,880
Kikuyu controls	17.1 (15–19)	110 (46/64)	Nyeri schools, Kenya	Nonendemic	~1,950
Yoruba controls	NA[a]	55 (27/28)	International Haplotype Map	Endemic	~700

[a] NA = not available.

About 2 ml of blood was obtained by venipuncture for all the samples we collected in Kenya. We extracted DNA within 10 h of blood collection, using a Qiagen DNA Blood mini kit, and then stored it at −20°C. All the participants provided informed consent, and, for children, informed consent was obtained from the parents and/or guardians. The study was reviewed and approved by the Harvard Medical School and Kenyatta University ethical review boards and by the Kenyan government.

Clinical Identification of Human Subjects with Severe Malaria

We identified human subjects who had severe malaria according to World Health Organization criteria. Blood smears and Giemsa staining were used to determine the asexual parasite count (parasitemia level). We identified cases as young children with >12 parasites per 200 red blood cells. All cases were also required to have overlapping clinical manifestations at the time of hospitalization, such as respiratory distress, convulsions, prostration, and hyperthermia (>39°C).

Genotyping at Candidate Genetic Variants

We genotyped all human subjects for 13 candidate malaria SNPs, using mass spectrometry (Sequenom).[23] We discarded SNPs with minor-allele frequency averaging <5% across the four ethnic groups, leaving 10 SNPs for subsequent analysis (table 2). Although the X-linked *G6PD* and *CD40* genes are important candidates for malaria-resistance genes,[3] we excluded them from this study because we wished to focus on autosomal SNPs that we could compare with an empirical panel of autosomal variants in the genome.

As an assessment of genotyping quality, we observed that, for 85 genotypes obtained in duplicate, there were 2 discrepancies, for a discordance rate of 2.4%. After removing samples with <80% genotyping completeness, we found that the average completeness of genotypes was 97.8%. We also compared genotyping of 10 SNPs in Yoruba samples with data from HapMap.[22] Of 30 genotypes obtained in duplicate, there was 1 discrepancy (3.33%). All SNPs were in Hardy-Weinberg equilibrium in all the ethnic groups we studied ($P > .05$).

Genotyping at 1,454 Random SNPs

For the assessment of allele-frequency differentiation at random SNPs, we used the Illumina Bead Lab System to genotype 1,536 random SNPs from the Illumina linkage panel (covering chromosomes 1, 2, 3, and 22) in 45 of the Luo controls, 47 Masai controls, and 37 Kikuyu controls. We also obtained genotypes for these SNPs in 55 Yoruba samples from the HapMap database.[22] Of these SNPs, 1,454 passed standard quality checks and had been genotyped in all four populations.

Case-Control Association Analysis

We assessed the statistical significance of allele-frequency differences between Luo cases and Luo controls, using a χ^2 test with 1 df. We used a one-tailed test of statistical significance, since our interest was in assessing whether a genotype or allele previously associated with malaria is more common in cases than in controls. We computed odds ratios (ORs) as $A = (f_{\text{case}}/1 - f_{\text{case}})/(f_{\text{control}}/1 - f_{\text{control}})$, where f_{case} is the frequency in cases and f_{control} is the frequency in controls. We also computed a 95% CI as the range of ORs that produced a likelihood ratio consistent with the data ($P > .05$). Specifically, we estimated the SE of the log OR as

$$B = \left(\frac{1}{n_{\text{case-ref}}} + \frac{1}{n_{\text{case-var}}} + \frac{1}{n_{\text{control-ref}}} + \frac{1}{n_{\text{control-var}}}\right)^{0.5} ,$$

where $n_{\text{case-ref}}$ and $n_{\text{case-var}}$ are the counts of the reference and variant genotypes in cases, and $n_{\text{control-ref}}$ and $n_{\text{control-var}}$ are the analagous quantities in controls. The 95% CI is quoted as the range ($e^{\ln(A)-1.65B}$ to $e^{\ln(A)+1.65B}$).

Epistasis Testing

To test for possible epistasis between any two SNPs, we used logistic regression. We compared the fit of three models with the data (case-control status for all the Luo samples): (1) genotype at the first SNP, (2) genotype at the second SNP, and (3) genotype at both SNPs.[24] We performed a one-tailed test for association with the genotypes previously associated with malaria. We calculated a Wald statistic and assessed significance for the epistatic interaction by a χ^2 test with 1 df.

Statistical Test for Natural Selection

The model of allele-frequency differentiation between two populations that we used to test for selection is that the difference in population frequencies at a given polymorphism is normally distributed with mean 0 and variance $cp(1-p)$, where p is the ancestral frequency. This model is similar to that of Nicholson et al.,[25] who showed that, for populations with modest genetic divergence times, it is a good approximation for allele-frequency differentiation. Under certain assumptions, the c parameter is expected to equal $2 \times F_{ST}$. From a population genetics perspective, c can be viewed as measuring genetic drift between populations.

To estimate c empirically, we used data from the 1,454 randomly chosen markers. For a given pair of populations, we estimated c as the empirical variance of the difference in population

Table 2. **Replication Analysis for 10 Genotypes or Alleles Previously Associated with Malaria Susceptibility**

Genotype or Allele	Reference SNP	Direction of Previous Association	Frequency in Controls (%)	No. of Cases/Controls Genotyped	OR (95% CI)	*P*
HbAS[10]	*rs334*	Protection[a,b]	25	447/454	.57 (.41–.79)	.0004
CD36 GT[12]	*rs3211938*	Risk[c]	12	456/457	1.50 (1.03–2.18)	.015
ICAM TT[11]	*rs5491*	Protection[a]	7	460/455	.71 (.42–1.21)	.10
NOS2A 1659 AA[13]	*rs8078340*	Risk[c,d]	6	450/455	.42 (.21–.83)	.99
TNF 238 A[14,15]	*rs361525*	Risk[c]	9	459/457	1.00 (.73–1.39)	.49
CD32 AA[17,18]	*rs1801274*	Protection[d,e]	25	455/447	.95 (.71–1.29)	.38
IFNARI LI168V CC[19]	*rs2257167*	Protection[c]	3	455/457	1.18 (.54–2.07)	.76
TNF 308 A[14,16]	*rs1800629*	Risk[c]	9	450/433	1.13 (.82–1.56)	.21
IFNARI 17470 CC[19]	*rs1012335*	Protection[c]	3	455/452	.85 (.53–1.36)	.34
TLR4 AG[20]	*rs4986790*	Risk[a]	10	407/303	1.36 (.85–2.17)	.10

[a] Previously published association with severe malaria.

[b] Previously published association with mild malaria.

[c] Previously published association with cerebral malaria.

[d] Previously published association with severe malarial anemia.

[e] Previously published association with parasitemia.

frequencies, after normalizing by $p(1-p)$ and accounting for sampling noise, which has variance $p(1-p)(1/N_1 + 1/N_2)$, where N_1 and N_2 are total allele counts for the two populations at a given marker. We approximated the normalization term $p(1-p)$ by setting p equal to the average of observed frequencies of the two populations, and we approximated binomial sampling noise as normally distributed. The same approximations were applied both to our estimation of c and to our subsequent analysis of individual markers. SNPs with average minor-allele frequency <5% for the two populations being compared were omitted from all computations, since the normal approximation becomes less reliable (table 3).

To test whether an individual marker was more differentiated than expected between two populations, we compared the observed difference in frequency with the expected distribution $N[0,p(1-p)(c + 1/N_1 + 1/N_2)]$, using the value of c estimated above, and computed a χ^2 statistic with 1 df. A feature of this test is that the χ^2 statistic has a mean value of 1 across the set of markers used to infer c. The test appropriately handles different sample sizes for candidate markers versus random markers used to infer c. A detailed statistical treatment will appear elsewhere (A.L.P, N.P., and D.R., unpublished data).

Combining Case-Control Association and the Test for Differentiating Selection

The combined test formally evaluates whether the observed data are consistent with the model of no case-control association and no selection. The test is performed by summing the association χ^2 statistic and the differentiation χ^2 statistic, forming a χ^2 statistic with 2 df. We note that the association χ^2 statistic used in this test is, by definition, a two-tailed statistic. We computed this sum for each pair of populations, using the same association statistic in each case. When one of the two populations being compared was the Luo population, we used the summed counts of Luo cases and Luo controls in the combined statistics reported in table 4. This generally leads to less significant P values than does using Luo controls only (and so is conservative). Using summed counts of Luo cases and Luo controls is appropriate under the null assumption of no association and ensures that the association statistic and differentiation statistic are independent. However, for the selection-only statistics reported in table 3, we used Luo controls only, since we wished to evaluate the evidence of selection in the control population, without regard to evidence of case-control association.

Results

Case-Control Association

We tested each of the 10 variants for association with malaria, comparing Luo cases with Luo controls. Two of the variants showed nominally statistically significant associations by one-tailed tests that searched for an association with the genotype or allele previously proposed to affect malaria resistance (table 3). We replicated the well-known association in which heterozygotes for the sickle-cell trait HbAS (*HbAS* T) are protected against severe malaria ($P = .0004$; OR 0.57 [95% CI 0.41–0.79]) (see the "Material and Methods" section). Although the OR of 0.57 is less strong than that observed in some previous studies,[9] it is in the same range as the OR of 0.45 (0.24–0.84), which was observed in another study of young children with a similar phenotype of severe malaria.[10] Different case-control studies focus on different phenotypes, and the protection of *HbAS* against severe malaria is known to vary with age,[26] so it is not surprising that the estimated ORs are heterogeneous across studies. We also replicated the association in which heterozygotes for *CD36* GT are at increased risk for severe malaria ($P < .015$; OR 1.50 [95% CI 1.03–2.18]).[12]

We note in passing that *NOSA* (*rs8078340*) gives a nominally significant P value (by a two-tailed test), but the association is in the opposite direction to previous reports ($P = .99$) (table 2). Our null findings at the other variants do not necessarily mean that they are unassociated; the CIs for the ORs are broad (table 2) and are often consistent with substantial association. We also note that our study included only individuals with parasitemia; we had no power to detect associations that were specific to cerebral malaria, a phenotype that was the focus of some previous studies.[13,27,28]

Table 3. Tests for Differentiating Selection between Malaria-Endemic and -Nonendemic Populations

Allele	Reference SNP	Allele Frequency (%) (No. of Alleles Used in Assessment)				P^a			
		Luo	Yoruba	Masai	Kikuyu	Luo vs. Masai	Luo vs. Kikuyu	Yoruba vs. Masai	Yoruba vs. Kikuyu
HbAS T	*rs334*	13 (908)	11 (102)	0 (194)	0 (200)	**.00149**	**.00036**	**.044**	**.025**
CD36 G	*rs3211938*	6 (914)	22 (100)	1 (186)	0 (202)	NA	NA	**.00590**	**.00096**
ICAM T	*rs5491*	25 (910)	24 (100)	16 (186)	18 (206)	.19	.25	.41	.48
NOS2A 1659 A	*rs8078340*	21 (910)	19 (98)	25 (188)	21 (204)	.62	.86	.57	.89
TNF 238 A	*rs361525*	9 (914)	1 (100)	21 (192)	16 (202)	**.04**	.13	**.00741**	**.010**
CD32 A	*rs1801274*	50 (900)	50 (98)	50 (190)	44 (210)	1.0	.37	1.0	.56
IFNARI L168V C	*rs2257167*	16 (914)	16 (98)	25 (192)	19 (208)	.20	.61	.37	.72
TNF 308 A	*rs1800629*	9 (866)	6 (96)	6 (188)	7 (204)	.60	.74	1.0	.84
IFNARI 17470 C	*rs1012335*	32 (904)	22 (108)	33 (190)	35 (202)	.92	.64	.32	.18
TLR4 G	*rs4986790*	5 (633)	4 (114)	7 (178)	5 (170)	.60	1.0	.6	.84

[a] *P* values for selection are based on allele-frequency differentiation tests between malaria-endemic (Luo and Yoruba) and -nonendemic populations (Kikuyu and Masai). Values in bold are significant. Statistics for SNPs with average minor-allele frequency <5% for the two populations analyzed are denoted as NA (not available).

Finally, we tested for epistatic interactions between each pair of variants,[29] but no pair showed a statistically significant interaction by a Wald test (not shown). We also tested for different strengths of association by sex but found no evidence of this (table A1).

Allele-Frequency Differentiation and Tests for Natural Selection

To test for differentiating natural selection, we compared the frequencies of the putative susceptibility variants between populations in which malaria is endemic and nonendemic (tables 3 and A2). We observed the most-significant frequency differentiation at the two SNPs that also showed the strongest associations (table 3). The sickle-cell allele *HbAS* T is present at appreciable frequency in the Luo (13%) and Yoruba (11%) but is absent in the malaria-nonendemic Masai and Kikuyu. The *CD36* G allele is present at 22% in the Yoruba and at 6% in the Luo but occurs at only ~1% frequency in the populations in which malaria is nonendemic.

To test whether these allele-frequency differences are greater than what could be explained in the absence of selection, we compared them with a panel of 1,454 random SNPs[30] for which we obtained genotypes in 45 Luo, 47 Masai, 37 Kikuyu, and 59 Yoruba. (We first assessed whether there was evidence of population substructure in the Luo,[31] which could, in principle, confound our case-control tests of association. No structure was detected, indicating that population stratification is not likely to cause false-positive or false-negative results in the association analysis.) We also used the data to assess the genetic relationships among the populations; understanding this is crucial to the tests for differentiating selection.

The genetic differentiation among populations ranges from $F_{ST} = 0.0012$ between Masai and Kikuyu (lowest differentiation) to $F_{ST} = 0.021$ between Yoruba and Masai (highest differentiation). We found that the Luo and Masai do not cluster genetically, despite the fact that they both speak Nilotic languages, whereas the Masai and Kikuyu are closely related (despite the fact that the Kikuyu speak a Bantu language) (fig. 1). These results show that the linguistic patterns in Kenya do not correlate with the genetic patterns, which is at odds with what has been suggested elsewhere.[32] Sampling of more populations should elucidate the relationships between genetic and linguistic groups in East Africa.[33]

To formally test for differentiating selection, we computed a χ^2 statistic for frequency differentiation at each tested SNP, assuming it was drawn from the empirical distribution defined by 1,454 random SNPs (see the "Material and Methods" section and table 3). Allele-frequency differentiation between malaria-endemic and -nonendemic populations is significant at *HbAS* T ($P = .00036$ for the most extreme Luo-Kikuyu comparison) and *CD36* G ($P = .00096$ for Yoruba-Kikuyu), with the results significant even after use of a Bonferroni correction for testing 40 comparisons of malaria-endemic and -nonendemic populations at 10 SNPs (this essentially involves multiplying the nominal *P* values by a factor of 40). By contrast, the eight SNPs that do not give positive case-control association show no evidence of differentiating selection ($P > .25$ for each SNP and pair of populations after correction for multiple hypotheses tested; $P = .39$ for the sum of 32 χ^2 statistics at these eight SNPs). We further evaluated the robustness of our selection test by computing χ^2 statistics for each of the 1,454 random SNPs for each of the four pairs of populations. If the test is robust, we would expect to achieve a χ^2 value >3.84, with probability 0.05. Restricting the analysis to SNPs in which the average allele frequency across the two populations tested was at least 5%, we observed a χ^2 value >3.84 in 255 (5%) of 5,090 of tests performed. Similarly, only 4 of 5,090 tests produced a *P* value <.001, and the lowest *P* value was not statistically significant after correction for 5,090 hypotheses tested ($P > .16$). These results show that our test for differenti-

Table 4. Formal Combination of Case-Control Association Analysis and Tests of Natural Selection

		P[a]			
Allele	Reference SNP	Luo vs. Masai	Luo vs. Kikuyu	Yoruba vs. Masai	Yoruba vs. Kikuyu
HbAS T	*rs334*	**.000056**	**.000018**	**.00048**	**.00029**
CD36 G	*rs3211938*	NA	NA	**.0023**	**.00043**
ICAM T	*rs5491*	.19	.23	.32	.36
NOS2A 1659 A	*rs8078340*	.033	.038	.032	.038
TNF 238 A	*rs361525*	.12	.30	.028	.038
CD32 A	*rs1801274*	.96	.65	.96	.81
IFNARI L168V C	*rs2257167*	.34	.68	.52	.73
TNF 308 A	*rs1800629*	.63	.69	.76	.75
IFNARI 17470 C	*rs1012335*	.91	.79	.56	.38
TLR4 G	*rs4986790*	.42	.41	.38	.42

[a] *P* values from combining case-control association studies with the test for differentiating selection between malaria-endemic (Luo and Yoruba) and -nonendemic (Masai and Kikuyu) populations. Values in bold are significant. NA = not available.

ating natural selection is not prone to false-positive results in a large selection of randomly chosen SNPs.

We note that both *HbAS* T and *CD36* G have been identified elsewhere as targets of recent positive natural selection.[22,34–36] However, the long-range haplotype test used to detect selection at these alleles detects evidence of selection from any cause and thus is not specific to a particular type of selection (e.g., for malaria resistance). The tests of allele-frequency differentiation we present here are much more specific to malaria. By comparing malaria-endemic and -nonendemic populations, we increase the probability that the loci detected as being affected by selection are specifically associated with malaria resistance. Of all the SNPs we tested for population differentiation—1,454 random SNPs and 10 candidates for malaria susceptibility—2 of those that achieve a nominal *P* value <.001 for at least one pair of populations were among the candidate malaria-resistance SNPs.

Combined Analysis of Case-Control Association and Selection

Finally, we formally combined the evidence of association with the evidence from the selection test (see the "Material and Methods" section). The combined test evaluates whether the observed data are consistent with the model of no case-control association and no selection. Whereas the evidence of association at *HbAS* T and *CD36* G is only moderate by the association analysis alone (see *P* values in table A1), significance is greatly increased when the association and selection evidence is combined: $P =$ .000018–.00029 for *HbAS* T and $P =$ .00043–.017 for *CD36* G, depending on which populations are compared (table 4). These results remain statistically significant after correction for 40 hypotheses tested ($P =$ .00072 for *HbAS* T and $P =$ .017 for *CD36* G).

Discussion

We performed a case-control association study of malaria resistance in the Luo, an East African population, analyzing 10 previously implicated variants. We replicated associations at *HbAS* (OR 0.57 [95% CI 0.41–0.79]) and *CD36* (OR 1.50 [95% CI 1.05–2.18]). Our OR for *CD36* is in agreement with the results published elsewhere.[12] Similarly, the OR for *HbAS* is in agreement with the previously reported longitudinal study in the same population (OR 0.45 [95% CI 0.24–0.84]; $P =$ .0001).[10] For *HbAS*, the protective effect that we observed is smaller than in some previous reports, which is potentially due to the fact that the cases we studied were young (average age 2.6 years) and thus lacked an immune basis for *HbAS* protection. (Williams and colleagues showed that *HbAS* has a more protective effect for older individuals.[9,10,26]) A possible reason why we did not replicate all the previous associations is that, in our study, the phenotype was parasitemia, whereas previous studies sometimes focused on cerebral malaria (table 2). We also show in table 2 that the CIs for the ORs are broad; thus, many of the variants we tested are consistent with an effect on malaria susceptibility, even if we could not reject the hypothesis of no association.

A particularly striking observation is that, at *CD36*, where we observe significant case-control association and highly significant allele-frequency differentiation, the variant increasing susceptibility actually has higher frequency in malaria-endemic populations. A possible historical explanation is that the selection pressures on this variant may have changed over time because of host-parasite genetic interactions. For example, the variant may have historically reduced susceptibility to malaria, and then, as the parasite evolved to adapt to the human immune system, the allelic association might have reversed. This hypothesis would be consistent with the known temporal and geographical heterogeneity in *CD36* binding

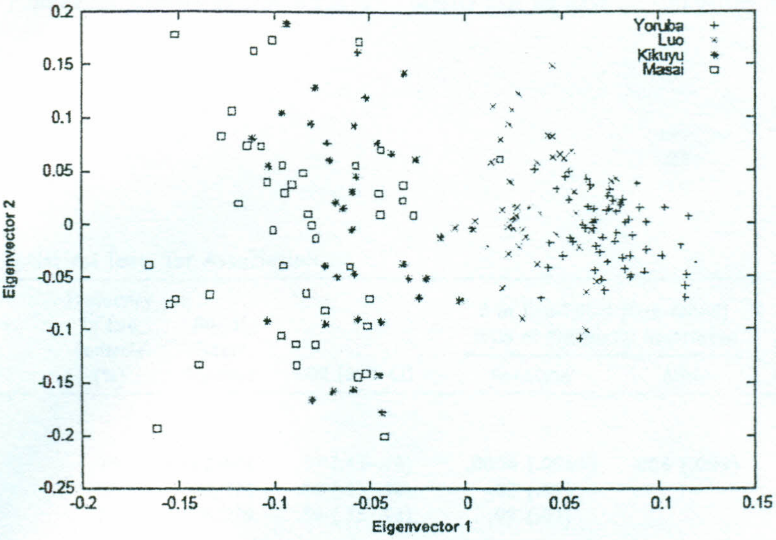


Figure 1. Principal-components analysis of samples from four different populations genotyped for 1,454 SNPs. The first eigenvector clusters the Yoruba, Luo, Kikuyu, and Masai. This is contrary to the expectation based on linguistics (the Luo and Masai both speak Nilotic languages) or geography. The F_{ST} values estimated between pairs of populations are as follows: Yoruba-Luo, 0.008; Yoruba-Masai, 0.021; Yoruba-Kikuyu, 0.015; Luo-Kikuyu, 0.008; Luo-Masai, 0.011; and Kikuyu-Masai, 0.0012. The top eigenvector is highly statistically significant by principal-components analysis[31] ($P \ll 10^{-12}$). The second eigenvector is not significant ($P = .09$).

and pathogenicity.[12] For example, genetic variation at the *PfEMP1* gene in the malaria parasite has been shown elsewhere to be associated with the pathogenicity,[37,38] and parasite *PfEMP1* and human *CD36* are known to interact.[39–41] In future studies, it will be interesting to explore whether human variants at *CD36* have different interactions with genetically different malaria parasite strains.[42]

These results finally provide empirical validation for a long-standing idea.[21] The idea is that, to increase power in case-control studies, one can combine the evidence of association with that from tests of natural selection. Previous studies have prioritized SNPs by natural selection on the basis of a combination of the alleles being frequent and being surrounded by a long-range haplotype[3,22,43]; the present study adds to this in several ways. First, we provide a formal χ^2 test of statistical significance, which can be combined with a case-control statistic to provide evidence that a SNP is a statistical outlier and, thus, a strong candidate for being associated with malaria. Second, our selection evidence is more specific to our phenotype of interest, since we are comparing frequency variants in populations differentiated by whether malaria has been historically endemic or nonendemic. Tishkoff et al.[33] recently applied a similar strategy to the phenotype of lactase persistence. They compared pastoral and nonpastoral populations in East Africa that have been differently exposed to diets including cow's milk. This analysis demonstrated high allele-frequency differences at variants near the lactase gene *LCT* and simultaneously showed that these highly differentiated variants also conferred the phenotype of lactase persistence.

We conclude that, in future whole-genome association scans, evidence from case-control comparisons can be combined with allele-frequency differentiation between differently exposed populations—and, potentially, other sources of evidence about recent selection[22,43]—to provide increased sensitivity and power in tests to detect disease-related genetic variants. It has been suggested that the identification of targets of selection may soon become a mainstream approach to finding genetic variants affecting human disease; our results provide empirical validation for this idea.[44] In our study, *P* values for *HbAS* and *CD36* were enhanced by several orders of magnitude with the use of <60 samples from each population analyzed, suggesting that this strategy may be cost effective relative to the number of additional samples needed to obtain a similar increase in power within the conventional case-control paradigm.

Acknowledgments

We are grateful to the human subjects and their families who participated in this study. We thank the medical officers and nursing staff in Bondo District Hospital, as well as the staff and students of the secondary schools where controls were collected. We are grateful to Pardis Sabeti, for thoughtful comments, and to Gavin McDonald, Alicja Waliszewska, Julie Neubauer, and Christine Schirmer, for technical advice. This work was supported by funds from Harvard Medical School, National Institutes of Health

(NIH) grant R21-AI64519, and a Burroughs Wellcome Career Development Award in the Biomedical Sciences to D.R. A.L.P. is supported by a Ruth Kirschstein K-08 award from the NIH. We declare that we have no competing financial interests.

Appendix A

Table A1. Statistical Tests for Association

Genotype/Allele, Reference SNP, and Sex	Frequency in Luo Controls (%)	No. of Cases/ Controls	OR (95% CI)	*P* in Two-Tailed (One-Tailed) Tests of Previously Associated		*P* in Tests Based on 3 × 2 Genotype Table
				Genotype	Allele	
HbAS AT:						
rs334:						
All	25	447/454	.57 (.41–.79)	**.0008 (.0004)**	**.006 (.003)**	**.0031**
Female	25	232/175	.55 (.33–.90)	**.02 (.01)**		
Male	25	215/279	.59 (.38–.93)	**.02 (.01)**		
CD36 GT:						
rs3211938:						
All	12	456/457	1.50 (1.03–2.18)	**.03 (.015)**	**.06 (.03)**	.061
Female	12	232/174	1.40 (.85–2.33)	**1.19 (.60)**		
Male	13	224/283	1.57 (.90–2.74)	.11 (0.06)		
ICAM TT:						
rs5491:						
All	7	460/455	.71 (.42–1.21)	.20 (.10)	.89 (.45)	.22
Female	8	235/176	.43 (.18–1.00)	**.04 (.02)**		
Male	7	225/279	1.05 (.53–2.09)	.90 (.45)		
NOS2A 1659 AA:						
rs8078340:						
All	6	450/455	.42 (.21–.83)	.01 (.99)	.69 (.5)	**.015**
Female	6	229/179	.55 (.22–1.40)	.21 (.11)		
Male	6	221/276	.28 (.09–.85)	.02 (.01)		
TNF 238 A:						
rs361525:						
All	9	459/457	1.00 (.73–1.39)	.28 (.14)	.97 (.49)	.47
Female	8	233/173	1.33 (.82–2.17)	.24 (.12)		
Male	9	226/284	.74 (.47–1.16)	.20 (.10)		
CD32 AA:						
rs1801274:						
All	25	455/447	.95 (.71–1.29)	.76 (.38)	.81 (.45)	.95
Female	24	229/176	1.06 (.67–1.67)	.37 (.19)		
Male	25	226/271	.45 (.30–.70)	.66 (.33)		
IFNARI LI168V CC:						
rs2257167:						
All	3	455/457	1.18 (.54–2.07)	.48 (.76)	.93 (.47)	.86
Female	2	231/180	1.37 (.45–4.17)	.37 (.19)		
Male	3	224/277	.37 (.11–1.24)	.66 (.33)		
TNF 308 A:						
rs1800629:						
All	9	450/433	1.13 (.82–1.56)	.21 (.11)	.42 (.21)	.25
Female	6	234/166	1.33 (.75–2.37)	.33 (.17)		
Male	11	216/267	1.17 (.78–1.74)	.45 (.23)		
IFNARI 17470 CC:						
rs1012335:						
All	3	455/452	.85 (.53–1.36)	.68 (.34)	.70 (.35)	.78
Female	2	234/180	1.73 (.52–5.70)	.34 (.17)		
Male	3	221/272	.46 (.15–1.43)	.66 (.33)		
TLR4 AG:						
rs4986790:						
All	10	407/299	1.36 (.86–2.17)	.20 (.10)	.067 (.02)	.20
Female	9	201/123	1.48 (.70–3.16)	.67 (.34)		
Male	11	206/176	1.33 (.74–2.40)	.17 (.09)		

Note.—This table is an expansion of table 2. Values in bold are significant.

Table A2. Allele Counts in Cases and Controls for the 10 Polymorphisms

Allele	Reference SNP	Luo Controls		Luo Cases		Masai		Kikuyu		Yoruba	
		Ref	Var	Ref	Var	Ref	Var	Ref	Var	Ref	Var
HbAS T	*rs334*	789	119	819	75	194	0	200	0	91	11
CD36 G	*rs3211938*	856	58	883	79	184	2	202	0	78	22
ICAM T	*rs5491*	683	227	688	232	156	30	169	37	76	24
NOS2A 1659 A	*rs8078340*	714	196	713	187	141	47	162	42	79	19
TNF 238 G	*rs361525*	833	81	837	81	152	40	169	33	99	1
CD32 A	*rs1801274*	454	446	453	455	95	95	92	118	49	49
IFNARI LI168V C	*rs2257167*	764	150	762	148	144	48	168	40	98	16
TNF 308 A	*rs1800629*	791	75	813	87	176	12	189	15	90	6
IFNARI 17470 C	*rs1012335*	616	288	622	280	128	62	131	71	84	24
TLR4 G	*rs4986790*	571	31	755	59	165	13	161	9	109	5

NOTE.—Ref = reference allele; Var = variant allele.

Web Resource

The URL for data presented herein is as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for malaria infection)

References

1. Bamshad M, Wooding SP (2003) Signatures of natural selection in the human genome. Nat Rev Genet 4:99–111
2. Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, Drousiotou A, Dangerfield B, Lefranc G, Loiselet J, et al (2001) Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science 293:455–462
3. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, et al (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837
4. Guerra CA, Snow RW, Hay SI (2006) Defining the global spatial limits of malaria transmission in 2005. Adv Parasitol 62:157–179
5. Kwiatkowski DP (2005) How malaria has affected the human genome and what human genetics can teach us about malaria. Am J Hum Genet 77:171–192
6. Mackinnon MJ, Mwangi TW, Snow RW, Marsh K, Williams TN (2005) Heritability of malaria in Africa. PLoS Med 2:e340
7. Clarke SE, Brooker S, Njagi JK, Njau E, Estambale B, Muchiri E, Magnussen P (2004) Malaria morbidity among school children living in two areas of contrasting transmission in western Kenya. Am J Trop Med Hyg 71:732–738
8. Miller LH, Mason SJ, Clyde DF, McGinniss MH (1976) The resistance factor to Plasmodium vivax in blacks: the Duffy-blood-group genotype, FyFy. N Engl J Med 295:302–304
9. Modiano D, Luoni G, Sirima BS, Simpore J, Verra F, Konate A, Rastrelli E, Olivieri A, Calissano C, Paganotti GM, et al (2001) Haemoglobin C protects against clinical Plasmodium falciparum malaria. Nature 414:305–308
10. Aidoo M, Terlouw DJ, Kolczak MS, McElroy PD, ter Kuile FO, Kariuki S, Nahlen BL, Lal AA, Udhayakumar V (2002) Protective effects of the sickle cell gene against malaria morbidity and mortality. Lancet 359:1311–1312
11. Kun JF, Klabunde J, Lell B, Luckner D, Alpers M, May J, Meyer C, Kremsner PG (1999) Association of the ICAM-1Kilifi mutation with protection against severe malaria in Lambarene, Gabon. Am J Trop Med Hyg 61:776–779
12. Aitman TJ, Cooper LD, Norsworthy PJ, Wahid FN, Gray JK, Curtis BR, McKeigue PM, Kwiatkowski D, Greenwood BM, Snow RW, et al (2000) Malaria susceptibility and CD36 mutation. Nature 405:1015–1016
13. Burgner D, Usen S, Rockett K, Jallow M, Ackerman H, Cervino A, Pinder M, Kwiatkowski DP (2003) Nucleotide and haplotypic diversity of the NOS2A promoter region and its relationship to cerebral malaria. Hum Genet 112:379–386
14. Knight JC, Udalova I, Hill AV, Greenwood BM, Peshu N, Marsh K, Kwiatkowski D (1999) A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. Nat Genet 22:145–150
15. McGuire W, Hill AV, Allsopp CE, Greenwood BM, Kwiatkowski D (1994) Variation in the TNF-alpha promoter region associated with susceptibility to cerebral malaria. Nature 371:508–510
16. Flori L, Delahaye NF, Iraqi FA, Hernandez-Valladares M, Fumoux F, Rihet P (2005) TNF as a malaria candidate gene: polymorphism-screening and family-based association analysis of mild malaria attack and parasitemia in Burkina Faso. Genes Immun 6:472–480
17. Shi YP, Nahlen BL, Kariuki S, Urdahl KB, McElroy PD, Roberts JM, Lal AA (2001) Fcγ receptor IIa (CD32) polymorphism is associated with protection of infants against high-density *Plasmodium falciparum* infection. VII. Asembo Bay Cohort Project. J Infect Dis 184:107–111
18. Cooke GS, Aucan C, Walley AJ, Segal S, Greenwood BM, Kwiatkowski DP, Hill AV (2003) Association of Fcgamma receptor IIa (CD32) polymorphism with severe malaria in West Africa. Am J Trop Med Hyg 69:565–568
19. Aucan C, Walley AJ, Hennig BJ, Fitness J, Frodsham A, Zhang L, Kwiatkowski D, Hill AV (2003) Interferon-alpha receptor-1 (IFNAR1) variants are associated with protection against cerebral malaria in the Gambia. Genes Immun 4:275–282
20. Mockenhaupt FP, Cramer JP, Hamann L, Stegemann MS, Eckert J, Oh NR, Otchwemah RN, Dietz E, Ehrhardt S, Schroder

NW, et al (2006) Toll-like receptor (TLR) polymorphisms in African children: common TLR-4 variants predispose to severe malaria. Proc Natl Acad Sci USA 103:177–182

21. Roeder K, Bacanu SA, Wasserman L, Devlin B (2006) Using linkage genome scans to improve power of association in genome scans. Am J Hum Genet 78:243–252
22. The International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437:1299–1320
23. Tang K, Fu DJ, Julien D, Braun A, Cantor CR, Koster H (1999) Chip-based genotyping by mass spectrometry. Proc Natl Acad Sci USA 96:10016–10020
24. Hosmer DW, Lemeshow S (1989) Applied logistic regression. Wiley, New York
25. Nicholson G, Smith AV, Jonsson F, Gustafsson O, Stefansson K, Donnelly P (2002) Assessing population differentiation and isolation from single nucleotide polymorphism data. J R Stat Soc 64:695–715
26. Williams TN, Mwangi TW, Roberts DJ, Alexander ND, Weatherall DJ, Wambua S, Kortok M, Snow RW, Marsh K (2005) An immune basis for malaria protection by the sickle cell trait. PLoS Med 2:e128
27. Burgner D, Rockett K, Kwiatkowski D (1999) Nitric oxide and infectious diseases. Arch Dis Child 81:185–188
28. Xu W, Humphries S, Tomita M, Okuyama T, Matsuki M, Burgner D, Kwiatkowski D, Liu L, Charles IG (2000) Survey of the allelic frequency of a NOS2A promoter microsatellite in human populations: assessment of the NOS2A gene and predisposition to infectious disease. Nitric Oxide 4:379–383
29. Williams TN, Mwangi TW, Wambua S, Peto TE, Weatherall DJ, Gupta S, Recker M, Penman BS, Uyoga S, Macharia A, et al (2005) Negative epistasis between the malaria-protective effects of α^+-thalassemia and the sickle cell trait. Nat Genet 37:1253–1257
30. Murray SS, Oliphant A, Shen R, McBride C, Steeke RJ, Shannon SG, Rubano T, Kermani BG, Fan JB, Chee MS, et al (2004) A highly informative SNP linkage panel for human genetic studies. Nat Methods 1:113–117
31. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2:e190
32. Cavalli-Sforza LL (2005) The human genome diversity project: past, present and future. Nat Rev Genet 6:333–340
33. Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, et al (2007) Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet 39:31–40
34. Hanchard NA, Rockett KA, Spencer C, Coop G, Pinder M, Jallow M, Kimber M, McVean G, Mott R, Kwiatkowski DP (2006) Screening for recently selected alleles by analysis of human haplotype similarity. Am J Hum Genet 78:153–159
35. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES (2006) Positive natural selection in the human lineage. Science 312:1614–1620
36. Moormann AM, Embury PE, Opondo J, Sumba OP, Ouma JH, Kazura JW, John CC (2003) Frequencies of sickle cell trait and glucose-6-phosphate dehydrogenase deficiency differ in highland and nearby lowland malaria-endemic areas of Kenya. Trans R Soc Trop Med Hyg 97:513–514
37. Kraemer SM, Smith JD (2003) Evidence for the importance of genetic structuring to the structural and functional specialization of the Plasmodium falciparum var gene family. Mol Microbiol 50:1527–1538
38. Rottmann M, Lavstsen T, Mugasa JP, Kaestli M, Jensen AT, Muller D, Theander T, Beck HP (2006) Differential expression of *var* gene groups is associated with morbidity caused by *Plasmodium falciparum* infection in Tanzanian children. Infect Immun 74:3904–3911
39. Ndungu FM, Sanni L, Urban B, Stephens R, Newbold CI, Marsh K, Langhorne J (2006) CD4 T cells from malaria-nonexposed individuals respond to the CD36-binding domain of Plasmodium falciparum erythrocyte membrane protein-1 via an MHC class II-TCR-independent pathway. J Immunol 176:5504–5512
40. Urban BC, Cordery D, Shafi MJ, Bull PC, Newbold CI, Williams TN, Marsh K (2006) The frequency of BDCA3-positive dendritic cells is increased in the peripheral circulation of Kenyan children with severe malaria. Infect Immun 74:6700–6706
41. Jeffares DC, Pain A, Berry A, Cox AV, Stalker J, Ingle CE, Thomas A, Quail MA, Siebenthall K, Uhlemann AC, et al (2007) Genome variation and evolution of the malaria parasite Plasmodium falciparum. Nat Genet 39:120–125
42. Volkman SK, Sabeti PC, DeCaprio D, Neafsey DE, Schaffner SF, Milner DA Jr, Daily JP, Sarr O, Ndiaye D, Ndir O, et al (2007) A genome-wide map of diversity in Plasmodium falciparum. Nat Genet 39:113–119
43. Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. PLoS Biol 4:e72
44. Vallender EJ, Lahn BT (2004) Positive selection on the human genome. Hum Mol Genet 13:R245–R254

APPENDIX F: ABSTRACTS FOR WORKSHOP AND CONFERENCES

ASSOCIATING GENETIC RESISTANCE WITH PLASMODIUM FALCIPARUM MALARIA INFECTION AMONG ETHNIC GROUPS RESIDENTS OF MALARIA ENDEMIC AND NON-ENDEMIC REGIONS OF KENYA

BY GEORGE AYODO

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR DEGREE OF DOCTOR OF PHILOSOPHY IN GENETICS OF INFECTIOUS DISEASES IN THE SCHOOL OF HEALTH SCIENCE OF KENYATTA UNIVERSITY

ABSTRACT -1

A striking difference between human populations where malaria is endemic and non-endemic is that infections in malaria-endemic areas include fewer life-threatening cases. We hypothesized that part of the reason for this epidemiological difference is that over the last few thousand years, endemic populations have built up genetic resistance to severe malaria infection. To test this hypothesis, we searched for evidence of natural selection in malaria exposed and unexposed populations by (a) carrying out a large-scale collection in Kenya of severe malaria cases and controls from the Luo ethnic group and also of population controls from the Masai and Kikuyu ethnic groups, (b) carrying out an association study at 10 genetic variants previously associated with malaria resistance, (c) studying frequency differences across populations to determine which of these variants have been subject to selection for malaria resistance in the past few thousand years, and (d) also studying haplotype and linkage disequilibrium patterns around malaria resistance genes to search for evidence of natural selection. In the Luo case-control samples, we replicated the previously described associations at CD36-GT (P value = 0.004) and HbAS (P value = 0.015). Strikingly, we also found that there was unusually high frequency differentiation of the HbAS and CD36-GT variants in the exposed (Luo and Yoruba) vs. relatively unexposed (Kikuyu and Masai) populations compared to a panel of 1,454 randomly chosen markers that we studied in the same samples (P = 0.00036 and 0.00096 respectively). We found that by statistically combining the case-control association and frequency differentiation statistics, we were able to increase the power of the association analysis by several orders of magnitude (HbAS with P value = 0.000018 and CD36-GT with P value =0.00043), which provides a potential tool for researchers to find risk factors for infectious disease in future. The further assessment of haplotype blocks flanking HbAS-T, CD36-G and ICAM-T suggested that exposed and un-exposed populations exhibit different haplotype block patterns, supporting the evidence of natural selection. CD 36-GT appears to be under selection in both the Luo and Yoruba ethnic group, whereas HbAS is under selection in the Yoruba ethnic only but not in Luo ethnic groups. These results suggests Yoruba and Luo—perhaps because they are on different sides of the African continent—evolved different genetic response to malaria since they began to be exposed to it several thousand years ago. This study has not only developed a novel method to identify malaria variants but has also provided an insight on the possibility of exploiting haplotype block patterns to map causal genes.

ABSTRACT-2

Combining evidence for natural selection with association studies increases power to detect malaria resistance variants

George Ayodo[1,2,3], Alkes L. Price[1,2], Alon Keinan[1,2], Arthur Ajwang[3], Michael F. Otieno[3], Alloys S.S. Orago[3], Nick Patterson[2] and David Reich[1,2]

[1]Department of Genetics, Harvard Medical School, Boston, MA USA; [2]Broad Institute of Harvard and MIT, Cambridge, MA USA; [3]Department of Pre-Clinical Sciences, Kenyatta University, Kenyatta, Kenya

It has been suggested that the power of case-control association studies to detect genetic variants can be increased by weighting polymorphisms according to their evidence for natural selection. As a proof-of-principle, we carried out an association study of 10 putative resistance variants in 464 severe malaria cases and 463 controls from the Luo ethnic group in Kenya. The study not only replicates associations at the *HBB* (P =0.0008) and *CD36* (P=0.03) genes, but also shows that the same two alleles are unusually differentiated in frequency between the Luo and Yoruba who have been historically exposed to malaria, and the Masai and Kikuyu who have not (compared with an empirical distribution based on 1,454 SNPs that we genotyped in the same populations). Combining evidence for association and/or selection, the evidence that these alleles are unusual becomes more significant by several orders of magnitude (as significant as P=0.000018 for *HBB* and P=0.00043 for *CD36*), demonstrating empirically that combining evidence from differentiating selection, with association studies, can boost power to detect variants. Interestingly, the frequency differentiation at *CD36* is in the opposite direction from what would be expected if it has always conferred resistance to malaria, raising the question of whether the pressure of selection on *CD36* has changed over time.

ABSTRACT-3

An Ethnic-Specific Polymorphism in the Catalytic Subunit of Glutamate-Cysteine Ligase Impairs the Production of Glutathione Intermediates that is implicated in several diseases.

Truc M. Le[1], Alecia S. Willis[2], Frederick E. Barr[1,3], Gary R. Cunningham[1,3], Jeff A. Canter[3], Rachel K. Price[3], George Ayodo[4], David Reich[4], and Marshall L. Summar[3,5]

[1]Division of Pediatric Critical Care Medicine, Vanderbilt University Medical Center, Nashville, TN [2] Department of Genetics, Baylor College of Medicine, Houston, TX [3] Center for Human Genetic Research, Vanderbilt University Medical Center, Nashville, TN [4]Department of Genetics, Harvard Medical School, Boston, MA, [5]Division of Medical Genetics, Vanderbilt University Medical Center, Nashville, TN

Glutathione plays a crucial role in free radical scavenging, oxidative injury, and cellular homeostasis. Glutamate cysteine ligase (GCL) is the rate-limiting enzyme in the glutathione biosynthesis pathway, and genetic variation in this gene can affect the ability to produce glutathione and respond to oxidative injury. Previously, we identified a non-synonymous polymorphism (462S) present only in individuals of African descent. Herein, we report that this ethnic-specific polymorphism (462S) encodes an enzyme with decreased or no *in vitro* activity when expressed in either a bacterial expression or a mammalian cell expression system. We also present evidence that apoptotic mammalian cells overexpressing the 462S enzyme have increased caspase activation and increased DNA laddering compared to cells overexpressing the wild-type enzyme. Finally, we also genotyped several African or African-descent populations and demonstrate that the allele frequency of the 462S polymorphism is increased in populations from malaria endemic areas and some African –descent in USA as compared to populations from Africa in non-malaria endemic areas. These findings describe a glutathione production pathway polymorphism common in individuals of African descent with decreased or no *in vitro* activity. Furthermore, demonstrating an increased frequency of the 462S GCLC allele among malaria endemic populations and African-descent in USA , we also show that this polymorphism has clinical significance. In addition, the complete lack of individuals who are homozygous for this genetic variant suggest that this genotype may exhibit a lethal phenotype in humans.

ABSTRACT-4

TITLE: - EMPIRICAL DEMONSTRATION OF THE CONSEQUENCE OF HUMAN AND P. FALCIPARUM CO-EVOLUTION.

The malaria parasites and the human hosts have co-evolved for thousands of years; consequently, there are individuals who remain asymptomatic despite infections. The co-evolution and its impacts on the host and parasite are still unclear. Using the observation from my previous work that demonstrated differential selection of ICAM and CD36 variant, I want characterize alleles and haplotypes of var genes of the parasite, Intercellular adhesion molecule 1 (ICAM-1) and CD36 of the human host. The var genes encode *Plasmodium falciparum* erythrocyte membrane proteins (pfEMP1) that interacts with the host receptors (ICAM-1 and CD36). The study will be a comparative genomic work where the subjects will consist of asymptomatic and severe malaria patients. I anticipate that this work will not only identify specific alleles or haplotype associated with malaria pathogenesis but will also identify a combination of host and parasite variants responsible for the parasite's fitness. This work in real–sense may demonstrate that virulence is consequence of parasite's fitness and it will further provide an insight as far as identifying drugs targets and vaccine candidate antigens