POPULATION GENETIC ANALYSIS OF *PLASMODIUM FALCIPARUM* FROM LAKE VICTORIA ISLANDS USING A NOVEL SINGLE FLUORESCENT LABELED MICROSATELLITE ASSAY

FELIX MUTHAMA MULENGE

MASTERS OF SCIENCE

(BIOINFORMATICS AND MOLECULAR BIOLOGY)

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

2017

Population Genetic Analysis of *Plasmodium falciparum* from Lake Victoria Islands Using a Novel Single Fluorescent Labeled Microsatellite Assay

Felix Muthama Mulenge

A thesis submitted in partial fulfillment of the requirements for the Degree of Master of Science in Bioinformatics and Molecular Biology of Jomo Kenyatta University of Agriculture and Technology

2017

DECLARATION

This research project is my original work and has not been submitted for any other degree or professional qualification.

Signature :……. Date ……..

Felix Muthama Mulenge

This thesis has been submitted for examination with our approval as supervisors.

Signature:…….. Date:…….

Prof Esther Magiri

JKUAT, Kenya

Signature:..... Date:.....

Dr Carol Hunja

CREATES, Kenya

Signature: **……** Date:.....

Dr Rashid Aman

ACCT, Kenya

DEDICATION

This thesis is dedicated to my entire family for their endless support. Their prayers and moral support have been the driving momentum that has enabled me to achieve the targeted goal in my studies.

ACKNOWLEDGEMENT

First, I take this opportunity to thank the almighty God for his love and blessings throughout the period of this project. I would like to express my special gratitude to my supervisors Prof Esther Magiri, Dr Hunja and Dr Rashid. Prof Magiri my university supervisor, thank you for your guidance, counsel, professional comments and encouragement that enabled me to go all the way to writing this work. I will always cherish each and every advice you gave. Secondly, I wish to express my most profound gratitude to my external supervisor Dr Hunja (CREATES) who allowed me into her project and supported me with technical support. You ably gave your guidance to my project and all related aspects timely and promptly. There other attribute that I have learnt besides the project work, which I consider to have added value to me as an aspiring scientist. Thank you for your valuable time. To my other external supervisor Dr Rashid (ACCT), am grateful for your enthusiasm and hard work. I owe a lot to your laboratory.

I also feel heavily indebted to the department of Biochemistry (JKUAT) for support and advice whenever things were tough for me. I would like also to thank each and every colleague that I worked with from laboratory. Geoffrey Masankwa, thank you for your technical training and support. For those I may have forgotten to mention yet they extended a helping hand in one way or another, thank you so much and may lord the almighty bless you all abundantly.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

WHO	World Health Organization
MoH	Ministry of Health
HEX	Hexachloro-6-carboxy-fluorescine
ROX	6-carboxy-X-rhodamine
TET	Tetrachloro-6-carboxy-fluorescine
FAM	6-carboxy-fluorescine
PCR	Polymerase Chain Reaction
LD	Linkage Disequilibrium
MOI	Multiplicity of Infection
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
PAGE	Polyacrylamide Gel Electrophoresis
bp	base pair
µL	Microliter
CREATES	Center for Research in Therapeutic Science
RBC	Red Blood Cells
SNP	Single Nucleotide Polymorphism
DBS	Dried Blood Spot
rfu	Relative fluorescent unit

ABSTRACT

Despite unprecedented efforts to control *Plasmodium falciparum* over decades, it remains entrenched in Africa, accounting for 90% of global malaria deaths. Control efforts have been difficult due in part to high levels of *P. falciparum* genetic diversity, drug resistance, poor vector management, poor drug access drug due to poverty. Attempts to understand the genetic structure and transmission dynamics of this parasite are underway in different regions prior to implementation of interventions. However, to date, few studies have examined the genetic diversity of malaria parasites in Lake Victoria islands where malaria transmission is stable. In order to examine population genetics, the current genotyping platforms based on laser induced fluorescence detection are relied on. The cost, however, of producing a new fluorescently labeled primer is prohibitive to many laboratories. To overcome such challenges a novel genotyping assay was adopted. The sensitive, specificity and reproducibility of the assay in typing *P. falciparum* cultures was determined. The assay proved to be highly reproducible with have detection sensitivity of up to 50 parasites/µL. The validated assay was used to genotype 188 *P. falciparum* samples from Lake Victoria basin in order to measure the extent of genetic diversity and population structure. High levels of genetic diversity were measured throughout the region (Mean *He* = 0.84) and low levels of population structure. Overall F_{ST} value was 0.044 indicating that approximately 5% of the overall allelic variation is due to differences between the populations. Based on these results, the study concludes that parasite population structure in the studied islands is shaped by human migration patterns that maintain extensive parasite gene flow between the sites. Consequently, any malaria elimination and interventions strategies in the study area will have to be carried out broadly on all four islands and adjoining mainland region.

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

Malaria arising from *Plasmodium falciparum* remains a leading global health threat, accounting for 438,000 deaths in 2015 (WHO 2015). Attempts to control and eliminate malaria are underway in different countries with variable transmission intensities, but the diverse parasite genetic structure that allows rapid evolution and spread of advantageous traits, threatens to undermine these efforts. For instance, the emergence and spread of parasite resistance to affordable antimalarial drugs such as chloroquine and sulfadoxine-pyrimethamine, have greatly contributed to the public health burden of malaria (White 1998). More worryingly, reports are appearing of evolving resistance to artemisinin, the current frontline treatment throughout the world (Dondorp 2009).

Therefore, there is urgent need to define the diversity, distribution and dynamics of malaria parasite populations, as parasite populations tends to differ geographically, due to confounding factors such as prevalence, vector species, host genetics and a variety of environmental influences. Using population genetic measurements such as expected heterozygosity, multiplicity of infection, linkage disequilibrium and F-statistics, factors affecting parasite population structure and genetic diversity such as inbreeding, epidemic expansion, geographic isolation and gene flow or introgression of foreign parasites can be characterized (Anderson *et al*. 2000). Thus, this study aimed to perform population genetic analysis of *P. falciparum* in Lake Victoria islands to determine the extent of parasite diversity, population structure and gene flow. Such findings can provide invaluable information, which may inform intervention strategies to reach elimination targets in the study area.

Numerous molecular genotyping techniques are in existence to analyze genetic diversity, transmission dynamics and population structure of *P. falciparum* field isolates. Early epidemiological studies focused on analysis of polymorphic genes encoding antigens notably merozoite surface protein (Joshi 2003), circumsporozoite surface protein (Chenet *et al*. 2008), glutamine rich protein (Greenhouse *et al*. 2006) and apical membrane antigen (Cortés *et al*. 2003). These antigenic motifs are often under strong immune selection pressure (Hughes and Hughes 1995) and therefore, genotyping results provided by these loci can potentially lead to a masked and distorted view of the parasite population structure and genetic diversity. Currently, microsatellites, which are presumed to be selectively neutral, are potential markers for population genetic studies (Anderson *et al*. 2000; Pumpaibool *et al*. 2009; Mobegi *et al*. 2012).

Although several methods such as agarose gel electrophoresis and polyacrylamide gel electrophoresis have been used for resolving microsatellite alleles, the current genotyping protocol is based on automated laser induce detection systems using capillary electrophoresis (Schuelke 2000). This technique requires the coupling of PCR fragments with fluorescent dyes and therefore, one of the primers in the PCR must be labeled at 5' end with fluorescent dyes. Labeling PCR fragments with fluorescent dyes has several advantages over other techniques like radioactive labeling or silver staining methods, including precise allele calling and potential for high-throughput operation (Li *et al*. 2009). Primer labeling coupled with automated fragment analysis allows rapid screening of sequence variation at multiple genomic loci. Intrinsic to this analysis is co-amplifying multiple loci with primer end labeled with different fluorescent dyes. The expense, however, of synthesizing a new fluorescently labeled primer for such large set of potentially informative loci is prohibitive and may prevent laboratories with limited budgets such as those in African settings from typing large numbers of microsatellite markers.

1.2 Statement of problem

Kenya launched its National Malaria strategy with a notably ambitious vision for a "malaria free Kenya" (Kenya Ministry of Health 2009). However, malaria arising from *P. falciparum* remains a major health problem among communities in Lake Victoria, Western Kenya. To achieve malaria elimination goals, it has been argued that understanding the genetic structure and transmission dynamics of *P. falciparum* parasites is crucial prior to the implementation of malaria interventions. To date, few studies have examined the genetic diversity, population structure and gene flow patterns of malaria parasites in Lake Victoria islands where malaria transmission is stable. Being an island surrounded by malaria endemic mainland region, the role of human movement and its aftermath effects on malaria interventions in these islands remains unknown. Thus, the aim of this study was to determine levels of parasite genetic diversity, population structure and gene flow, as this will allow tracking of parasite migratory routes in the study sites prior to the implementation of interventions.

In order to perform parasite genetic analysis as well as gene flow patterns, the current genotyping standard based on automated systems using fluorescently labeled PCR fragments are relied to offer precise allele calling. In this method, one of the primers used in the PCR is 5' end labeled with multiple fluorescent dyes. However, the cost of synthesizing fluorescently labeled primer is a limiting factor for many labs, as labeled primer in 50nmole range can cost US $100–130, depending on the dye which is five to ten times more than unlabeled ones. The expense of synthesizing this fluorescent-labeled primer is prohibitive to many African laboratories due to financial constrains and where malaria is still a public health problem. This study aimed at adopting a novel microsatellite assay that would use single fluorescent dye instead of multiple dyes to genotype *P. falciparum* and help overcome this financial burden.

1.3 Justification

Plasmodium falciparum remains the single most important threat to public health at global scale, accounting for more than 90% of the world's malaria mortality (WHO 2015). Despite unprecedented efforts to control this parasite, it remains entrenched in Africa. Thus, understanding the epidemiology, diversity, distributions and dynamics/exchange of gene flow of *P. falciparum* in different epidemiological regions is crucial in developing specific control tools against this parasite. With malaria elimination on the global agenda, mapping parasite population structure is essential prior to establishing goals for elimination and the rolling out of interventions. The Malaria Eradication Consultative Group recently recommended that countries to assess feasibility of undertaking malaria elimination campaign through a preliminary study to accumulate and analyze information required for realistic planning (malERA 2011). This is highly pertinent before embarking on a costly and potentially ineffective campaign.

Located in a high transmission region of Western Kenya where *P. falciparum* is the predominant species, the Lake Victoria islands present great opportunity for feasibility of malaria elimination and intervention studies. Islands are goods targets for such studies as they are assumed to be isolated with negligible possibility of reintroduction of malaria parasites by migration. An example is Vanuatu, an endemic archipelago in the South-West Pacific region has extensively been studied to determine the feasibility of malaria elimination in low transmission settings as well as how gene flows within and between the islands affects intervention plans (Lum *et al*. 2004). Indeed, successful elimination of malaria has been registered in this region and others such as Caribbean, Cyprus, Mauritius, Maldives, Reunion, Taiwan, and Singapore islands (WHO 2009).

Therefore, performing population genetic analysis on Lake Victoria islands would help identity routes of parasite transmission and gene flow patterns. Such information provides a valuable insight on how control and elimination strategies need to be implemented to disrupt the underlying dynamics and evolution of parasites in targeted foci. Population genetic indices such as the extent of diversity and genetic differentiation would be robust predictive markers of whether control strategies have succeeded in reducing the parasite population. Therefore, results from this study would provide a pertinent model for guiding future malaria-elimination program in Kenya.

In order to overcome financial difficulty associated with synthesizing multiple fluorescent-labeled primers, a novel procedure described by Li *et al* (2007) was adopted. This technique incorporates the concept of three primers as described by Schuelke (2000) with major modification on the use of single fluorescent dye. This technique has successfully been used for typing rodent parasite *P. yoelii* (Li *et al*. 2007) and could be a valuable tool for genotyping *P. falciparum*. It proved to be robust, reproducible, cheap and high throughput method for typing large set of microsatellite markers. Further, it presents a practical advantage in that the fluorescence-labeled primer can be used for amplification of all loci included in genotyping assay, so labeling of multiple primers with different fluorescent dyes is no longer necessary. Additionally, this method is designed to allow multiplex of first PCR reaction, in which several markers can simultaneously be co-amplified. This increases the amount of information generated per assay, and substantively reduce consumables and labor costs. Since malaria is for the most part, endemic in developing countries with limited budgets and techniques for use in malaria research, this methodology would ideally be cheap and of great benefit to research groups who perform low-throughput genetic analyses with a high number of microsatellite markers.

1.4 Hypothesis

H_0: Lake Victoria islands do not have a repertoire of genetically diverse *P. falciparum* parasites with no distinct population structure.

1.5 Objectives

1.5.1 General objective

To determine population genetic structure of *P. falciparum* in Lake Victoria islands using a novel single fluorescent-labeled microsatellite assay.

1.5.2 Specific objectives

1. To determine specificity, sensitivity and reproducibility of a novel microsatellite assay in genotyping *P. falciparum.*
2. To determine genetic diversity and multiplicity of infection of *P. falciparum* isolates from Lake Victoria islands.
3. To determine population structure and gene flow patterns of *P. falciparum* in Lake Victoria islands.
4. To evaluate linkage disequilibrium in the populations of *P. falciparum* from Lake Victoria islands.

CHAPTER TWO

LITERATURE REVIEW

2.1 Etiology of malaria

Malaria is a life-threatening, vector-borne infectious disease caused by intraerythrocytic protozoan parasites of the genus *Plasmodium*. So far, five species: *P. falciparum, P. vivax, P. knowlesi, P. ovale* and *P. malariae* have been identified as human malaria causative agents (Antinori *et al*. 2012). It is known that *P. falciparum* is the most virulent of the human malaria parasite species. It is transmitted by the female *Anopheles* mosquito and causes mild to lethal complications to the vertebrate host if early diagnosis and treatment is not adopted (Sinka *et al*. 2012).

2.2 Global distribution of malaria burden

Malaria is one of the world's most common infectious diseases and one of greatest global public health problems. The disease is broadly distributed in both the subtropics and tropical regions. According to the World Malaria Report (2015), 85 countries are classified as malaria endemic with 3.2 billion people being at risk of contracting and developing the disease. Globally, there were 214 million clinical episodes reported in the year 2015, of which 88% were from Africa, 10% from S. E Asia and 2% from Eastern Mediterranean region. Similarly, 438,000 deaths arising from the disease were reported in the same year, with 90% of the cases experienced in Africa south of the Sahara, 7% in S. E Asia and 2% in E. Mediterranean. Of these global estimates, 306,000 deaths occurred in children under the age of 5 years. Many epidemiological studies suggest that vast majority of malarial infections in Sub-Saharan Africa are mainly due to *P. falciparum* (WHO 2015). Despite more than a century of efforts to eradicate

or control this devastating disease, the existence and wide spread nature of the *P. falciparum* vector – the mosquito *Anopheles gambiae* has made malaria control in Africa more difficult.

2.3 Malaria situation in Kenya

Malaria remains a major public health problem in Kenya, accounting for 18% of all outpatient consultations and 6% of in-patient admissions (MoH 2016). Although, *P. falciparum* is the species frequently associated with severe malaria and accounts for 80-90% of clinical cases in Kenya, *P. malariae, P. ovale,* and *P. vivax* also exist in the country (Okara *et al*. 2010). Malaria transmission and risk of infection in the country are largely influenced by altitude, temperature and rainfall patterns. Therefore, prevalence of malaria varies considerably by season and across geographic regions. The variations in altitude and terrain create contrasts in the country's climate, which ranges from tropical along the coast to temperate in the interior to very dry in the north and northeast. The country is divided into four eco-epidemiological strata of malaria:

Endemic areas: These areas experience stable malaria have altitudes ranging from zero in the coastal region to 1,300 meters around the Lake Victoria basin in western Kenya (Figure 2.1). Malaria transmission is intense throughout the year, with *P. falciparum* prevalence between 20-40% and high entomological inoculation rates of 29.2 per year (MoH 2016). The coastal region has malaria prevalence ranging from 5–20%. Of the total Kenyan population, 26% lives in a malaria-endemic zone.

Highland epidemic prone areas: Epidemics are experienced in western highlands of Kenya where malaria transmission is seasonal with considerable year-to-year variation. The entire population is vulnerable and case-fatality rates during an epidemic can be greater than in

endemic regions. Approximately 39% of Kenyans live in these areas. The malaria prevalence in these areas ranges from 1–10% but some foci experience prevalence between 10% and 20%.

Seasonal malaria transmission areas: This epidemiological zone comprises arid and semiarid areas of northern and southeastern parts of the country, which experience short periods of intense malaria transmission during the rainy seasons. Although this is the largest zone in terms of geographic size, only 14% of the population lives in areas where the malaria prevalence is less than 5%.

Low malaria risk areas: This zone covers the central highlands of Kenya including Nairobi. Approximately 21% of the population lives in this area where there is little to no malaria transmission.

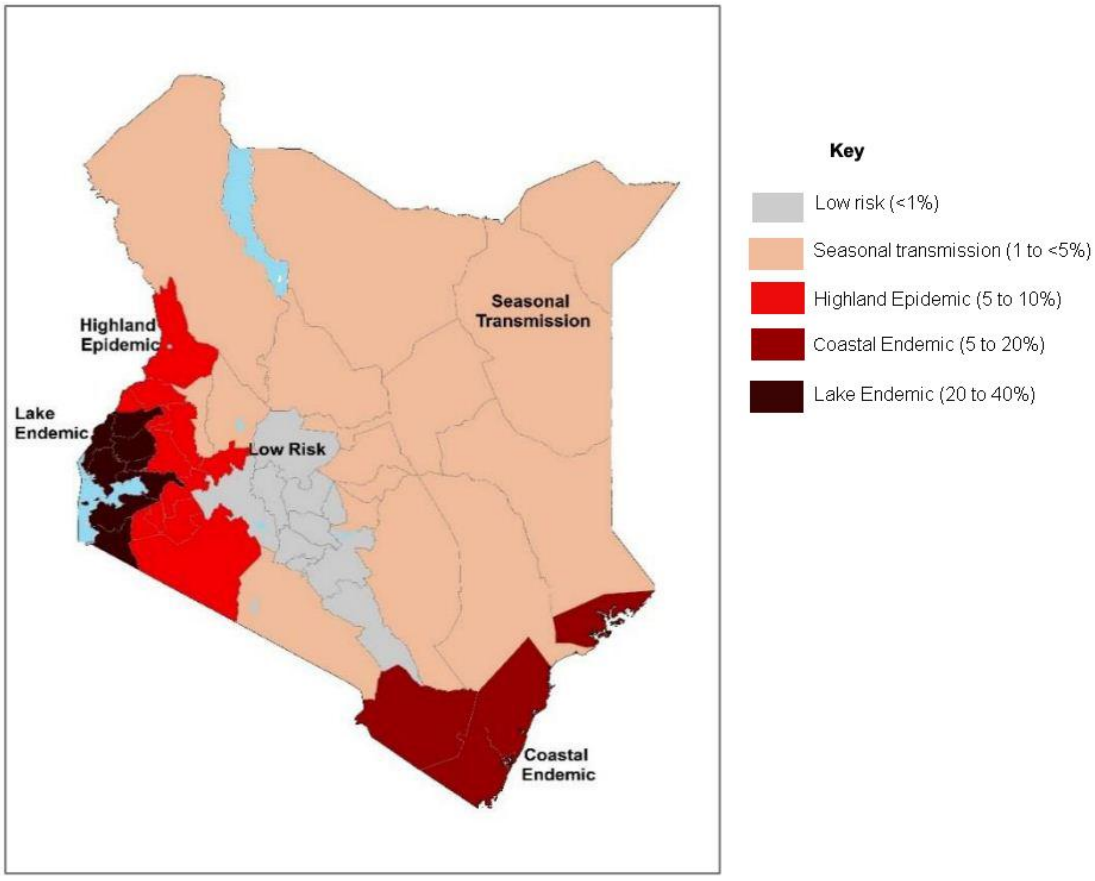

Figure 2.1 A country malaria endemicity map showing transmission zones (MoH 2016).

2.4 The life cycle of *Plasmodium* parasite

The *P. falciparum* malaria parasite has a highly complex and extraordinary multistage life cycle occurring in both the mosquito and human host (Tuteja 2007). Human infection begins when an infectious mosquito inoculates sporozoites into the human host during a blood meal (Figure 2.2). The sporozoites then migrate to the liver where they invade hepatocytes and undergoes non-pathogenic asexual replication phase producing thousands of merozoites (Cowman *et al*. 2012). The mature merozoites are released and invade red blood cells (RBC).Within the RBC, each merozoite replicate their DNA 4 to 5 times within 48 hours and release 16 to 32 daughter merozoites back to the blood stream to infect other RBC. Clinical manifestations such as chills, fever, splenomegaly, anemia, celebral malaria and sweats characterize this phase. While the erythrocytic phase produces millions of haploid asexual parasite, a small proportion of parasites do not undergo this phase but differentiate into male and female sexual forms (gametocytes) which aid the parasite to infect female *Anopheles* mosquito a fresh (Kebaier *et al*. 2009; Lew *et al*. 2003).

When a feeding mosquito ingests the gametocytes during a blood meal, they develop into male and female gametes, fuse and form a diploid zygote that develops into an ookinete. Genetic recombination and meiosis occurs at this phase. The motile ookinete tunnel into the midgut of the host and develops into oocyst. Each oocyst produces thousands of haploid forms (sporozoites) after rounds of mitotic divisions. The sporozoites are released and invade salivary glands. When the mosquito takes another blood meal, the sporozoites are released into the blood stream of the human host initiating another infections (Barillas-Mury and Kumar 2005).

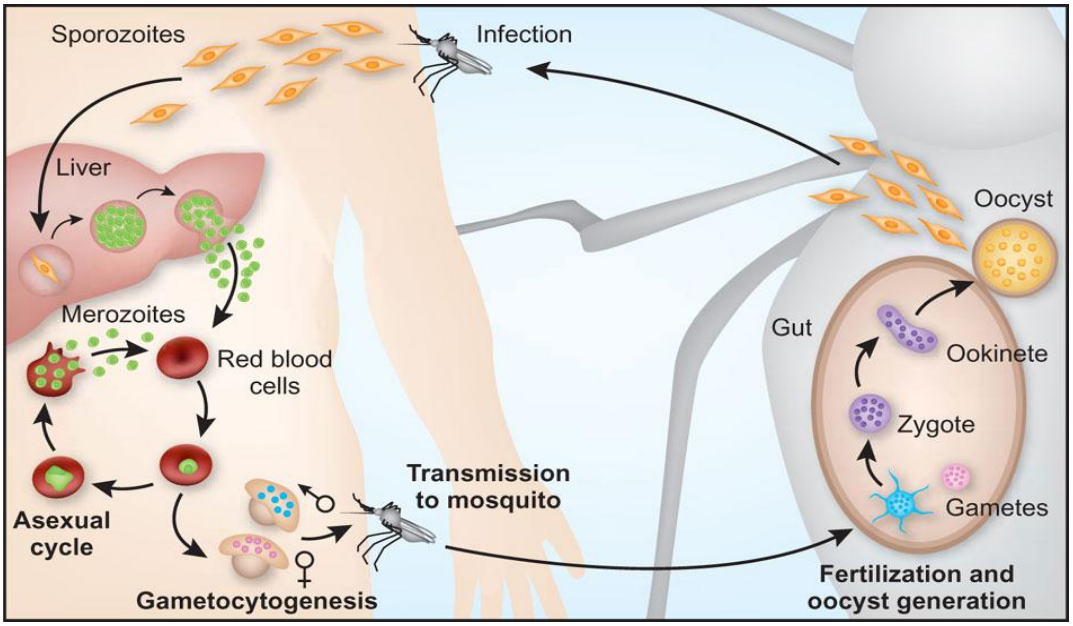

Figure 2.2 The life cycle of malaria parasite (Pasvol 2010)

2.5 Genetics of *Plasmodium falciparum*

The parasite *P. falciparum* has threes genomes: apicoplast genome, mitochondrial genome and a nuclear genome. The apicoplast is a 35kb circular genome that resembles the plastid DNA of non-photosynthetic plants, encoding translation and transcription-related protein genes, rRNAs and tRNAs. The genome is universally conversed in *Plasmodium* species. The function of apicoplast remains unclear, though it appears to be specified by gene transferred to the nucleus (Wilson *et al.* 1996).

The mitochondrial genome is the smallest genome of *Plasmodium*; about 6kb. It contains three protein-coding genes; cytochrome c oxidase I, cytochrome c oxidase III and cytochrome b (Feagin 1992). All these three genes are essential for a range for cellular processes. Analysis of mitochondrial genome has proven to be highly relevant in inferring inter and intra-specific evolutionary history of *Plasmodium* species (Joy *et al*. 2003).

The nuclear genome comprises 14 linear chromosomes that are haploid in nature with approximately 23.3Mb genomic sequence and more than 5600 genes of whom 1,817 have known functions (Gardner *et al*. 2002). The nuclear genome of *P. falciparum* is the most AT rich genome sequenced to date, with an overall AT composition of approximately 81%, which rises to 90% in intergenic regions and introns. Homologous chromosomes of *Plasmodium* vary considerably in size between clones, arising from continuous deletions, cross-over and chromosomal aberrations occurring preferably at telomeric regions (Jeffares *et al*. 2007). The telomeric regions of the chromosome carry genes encoding surface proteins motif recognized by human immune response and so they are under constant selection pressure. This gives the parasite a considerable capacity for changes in antigen expression and thereby developing immune evasion tactics.

2.6 Genetic diversity of *Plasmodium falciparum*

Genetic diversity is the variation in the genetic composition within or among individuals' population or species. Genetic recombination that is classically defined as the process, by which organisms with novel combinations of genes are produced in crosses between two parent organisms, is arguably considered as a major factor in creating genetic diversity (Conway and McBride 1991). In *Plasmodium*, this event occurs primarily at meiosis phase and allows genomic sites of the parasite to assort independently, thus it may act as a diversifying force, generating new genetic variants that can spread through the population driven by positive selection (Wootton *et al.* 2002). The genetic divergence vary considerably among the parasite populations and is often associated with local endemicity, transmission rates, geographic isolation levels and migration patterns of human and vector hosts (Anderson *et al*. 2000; Bogreau *et al*. 2006; Pumpaibool *et al*. 2009).

In regions of high endemicity of *P. falciparum* such as Sub-Saharan Africa, human hosts often harbor two or more genetically distinct clones of the same parasite (Fraser-Hurt *et al*. 1999) and mosquitoes have high chances of taking up mixture of genetically distinct gametocytes during a blood meal. Fusion of these two genetically distinct parasites (crossover fertilization) results into generation of a large repertoire of genetically diverse parasites. Although self-fertilization between identical gametes can occur, it exclusively occurs when the blood meal contain one parasite genotype. Such out-crossing variability have direct implication in the emergence of parasites' advantageous phenotypes (Wootton *et al*. 2002).

The intensity of *P. falciparum* transmission by *Anopheles* mosquito plays a crucial role in determining genetic diversity of the parasite. Malarial transmission intensity is highly dependent on the vectorial capacity, which is defined by the density, longevity and bionomics of *Anopheles* vector prevalent in a given area and the climatic suitability for the *Plasmodium* species. Although this confounding factors vary geographically, high levels of transmission rates are dominant in African populations (Anderson *et al*. 2000). In Kenya, high malaria transmission is common in low-elevation regions (0-1300m above sea level) around Lake Victoria and along the coastal regions with habitats that provide suitable breeding ground for *Anopheles* mosquitoes (Chen *et al*. 2004). High proportions of mixed parasite genotypes arising from frequent meitotic recombination events in mosquito are observed in individuals residing in these regions. Contrary, in low malaria transmission regions, parasite clones as low as one clone per infection are common and this enhances chances of inbreeding and reduces the effective recombination rates (Zhong *et al*. 2007).

Epidemiological studies on *P. falciparum* genetic diversity show that high levels are generally predominant in African populations (Anderson *et al*. 2000), low in S. America

populations (Machado *et al*. 2004) and intermediate in S. E Asia populations (Iwagami *et al*. 2009). In Africa, the diverse parasite populations are not strongly isolated from each other due to dynamic migratory patterns of the human host and different levels of geographic isolation (Lynch and Roper 2011). Exceptions are observed in remote islands such as Comoros (Rebaudet *et al*. 2010) and on the very edge of malaria endemic zones such as Djibouti (Bogreau *et al*. 2006), that show fragmented parasite population structures due to restricted parasite gene flow. This diverse parasite population structure observed globally is directly implicated in the emergence of advantageous phenotypes such as drug resistance and antigenic variants (Pumpaibool *et al*. 2009; Talisuna *et al*. 2007), with persistent migration of vectors and humans playing an important role in the spread of these traits (Roper *et al*. 2004; Lum *et al*. 2004). These factors continue to curtail any progress made towards mitigating this disease.

2.7 Global malaria control and elimination campaign

During the global malaria eradication program in the early 1950s, 37 endemic countries succeeded in eliminating malaria transmission while others aimed for control, not elimination (Yekutiel, 1960). In sub-Saharan Africa, Kenya, Nigeria, Ghana and Tanzania were among the countries where elimination was considered unfeasible due to their high malaria endemicity. Over the last decade, there has been a substantial increase in international funding for malaria control through major international financing agencies such as Bill and Melinda Gates foundation and the US President's Malaria Initiative. This, together with renewed political commitment in endemic countries, has resulted in increased coverage of malaria interventions and a reduction in malarial disease and death in several countries, including several in sub-Saharan Africa where malaria burden is greatest. Inspired by these achievements and by the momentum created by global advocacy, several countries are now revising their strategic plans

and are considering elimination as an alternative to maintaining control measures indefinitely (Moonen *et al*. 2010).

With elimination on the global agenda, different countries have launched or are considering launching their National Malaria Strategic plans with a notable vision of eliminating malaria in defined geographical areas. However, it has been argued that premature commitment to elimination may be counterproductive as it could divert limited resources and negatively impact efforts to reduce the high burden of malaria (Snow and Marsh 2010; Najera *et al*. 2011). Moreover, the existence of diverse parasite population with advantageous phenotypes such as drug resistance and antigenic variability has been cited to have practical implication on malaria control and elimination strategies (Rebaudet *et al.* 2010). For instance, the emergence of reduced susceptibility to artemisinin in Thailand (Dondorp 2009), West Cambodia (Amaratunga *et al*. 2012) and Myanmar (Shah *et al*. 2011) threatens to curtail any progress made towards mitigating this scourge. Thus, there is an urgent need for clear, evidence-based guidance to assess whether malaria elimination represents a realistic goal in a given area.

In the context of achieving malaria elimination targets, islands have been used as study models to determine feasibility of malaria elimination in different transmission settings. Islands are considered ideal targets for such studies due to their isolated nature. An example is Vanuatu, an endemic archipelago in the South-West Pacific region that has extensively been studied to determine the feasibility of malaria elimination in low transmission settings as well as how human movement within and between the islands affects malaria intervention plans (Lum *et al*. 2004). The study revealed that *P. falciparum* in this region was largely isolated on individual islands and human movements between islands were responsible for malaria parasite gene flow. This implied that malaria interventions needed to be carried out on an island-by-island basis to

curtail possible gene flow and subsequent parasite importation (Reid *et al*. 2010). Indeed, successful elimination of malaria has been registered in other islands such as Caribbean, Cyprus, Mauritius, Maldives, Reunion, Taiwan, and Singapore islands where transmission is low and unstable (WHO 2009).

However, despite extensive malaria elimination feasibility studies on islands with low malaria transmission, not much has been done on islands in high transmission regions. Past failures of malaria elimination experienced in malaria endemic islands such as Zanzibar off the coast of Tanzania (LeMenach *et al*. 2011) and Comoros archipelago off the eastern coast of Africa (WHO 2008) raised questions on the extent of human, and vector migratory trends and parasite population structure in these intense transmission islands. Therefore, *P. falciparum* population genetic studies are crucial in evaluating the extent of parasite genetic diversity, which is an indicator of the parasite populations' resilience to control measures. Population genetic structure is pertinent, to help map spatial distribution of genetic diversity over geographical space and thus inferring parasite migratory patterns, prior to implementation of malaria interventions (Reid *et al*. 2010).

2.8 Genetic markers for genotyping malaria parasite

A genetic marker is any visible trait (phenotype) or easily identifiable piece of genetic material (DNA) which serves as indices for presence or absence of variations of individuals within the population. Numerous studies have been done to reveal substantial parasite diversity within and among vertebrate and insect hosts, although most studies center on parasite clonal diversity within the vertebrate (Anderson *et al*. 2000; Machado *et al*. 2004; Campino *et al*. 2011). *Plasmodium* population structure and epidemiological studies have extensively relied on surface protein loci notably Merozoite surface proteins 1 and 2 (MSP1 and MSP2) (Akter *et al*.

2012) and less frequently the circumsporozoite surface protein (CSP) (Chenet *et al*. 2008) and glutamine rich protein (GLURP) (Greenhouse *et al*. 2006) to reveal repeat lengths and sequence variation.

These loci are antigenic motifs that are continuously exposed to host immune responses, which suggest that the loci are under selection pressure. Due to strong immune selection there is likelihood of mutation on the surface protein loci that substantially affects population dynamics of the parasite (Jiang *et al*. 2011). This is evidenced by work of Gilbert *et al*. (1998) which demonstrated a non-randomized CSP locus sequence pattern in *Plasmodium*, a strong indicator of immune selection. Clearly, to confer parasite population structure using data derived from these loci is problematic, since it is not clear whether the patterns observed reflect population history or natural selection. Therefore, selecting a neutral stable marker that accurately figures out genomic variation across the *Plasmodium* genome would alleviate this problem. Fortunately, many other polymorphic markers are available to study genetics of *Plasmodium* at single gene level (Sunnucks 2000) as they provide sufficient information on allelic variation at a given locus within the genome of the organism. They include isoenzymes, Single Nucleotide Polymorphisms (SNPs), Amplified Fragment Length Polymorphism (AFLP) and microsatellites.

2.8.1 Isoenzymes

Among the different genetic markers, the first to be established were the isoenzymes in the 1960s (Lewontin and Hubby 1966). Isoenzymes markers are allelic forms of enzymes that occur due to genetic variations. They are simple, cheap, codominant although not highly polymorphic hence they have extensively been used for evolutionary studies, estimating genetic variation and to less extent monitoring various forms of gene conservation (Gaensslen *et al*. 1987). However, the use of these molecular markers is limited since there are only small number of identified

markers within a limited number of genes hence the genetic information obtained is not a representative of the whole genome. These were initially used as there were no alternative ways to determine genetic variations.

In *Plasmodium* genetic studies, glucose phosphate isomerase, glutamate dehydrogenase, adenosine deaminase and lactate dehydrogenase enzymes were extensively relied to delineate genetic variations (Freese and Markus 1990; Joshi *et al*. 1989; Myint-Oo 1986; Thaithong *et al*. 1981). A study by Freese and Markus (1990) using the listed isoenzymes reported limited genetic variation among the *P. falciparum* isolates in African population despite high transmission intensities. Such low genetic variations may partly be due to the nature of the markers. Isoenyzmes are protein in nature and reflect alteration on DNA sequence through changes in amino acid composition. Changes in amino acid composition may less often change the conformation of the enzyme due to redundancy of the genetic code thereby producing no change in electrophoretic mobility of the band (that is, visualization of different alleles). This renders them less informative markers. Nevertheless, with technological breakthrough, other informative markers were adopted.

2.8.2 Single Nucleotide Polymorphism (SNPs)

SNPs are single base change in a DNA sequence, with a usual alternative of two possible nucleotides at a given position. For such a base position with sequence alternatives in genome sequence to be considered as an SNP, it is considered that the least frequent allele should have a frequency of 1% or greater in the population. SNPs are relatively easy to assay, co-dominant and bi-allelic markers since variation at each locus is defined by two alleles. In *Plasmodium*, SNPs occurs abundantly in the genome including promoters, introns and exons, although their distributions vary greatly among various chromosomal regions or different genes (Mu *et al*.

2007; Volkman *et al*. 2007). SNPs within a coding region may not necessarily change amino acid sequence of the protein due to degeneracy nature of genetic code (Feng *et al*. 2003). A change in a SNP in which both forms lead to the same protein sequence is termed as synonymous (silent mutations). SNPs occurring in non-coding sequence are subject to gene splicing or transcription factor binding.

Generally, SNPs are deemed to have relatively low frequency of single nucleotide substitutions, estimated to being between 1×10^{-9} and 5×10^{-9} per nucleotide per generation (Li *et al*. 1981). Therefore, the probability of two independent base changes occurring at a single position is very low. To carry out genetic studies using these markers calls for screening of a large number of SNPs for adequate information which is expensive and requires a lot of extensive in silico analysis to delineate the obtained sequences (Morin *et al*. 2004). Despite their shortcoming, SNPs have received considerable attention recently as potential marker for estimating genetic diversity (Feng *et al*. 2003), population dynamics and inferring the evolutionary history of species populations (Mu *et al*. 2002).

2.8.3 Microsatellite

Despite the radical emergence of novel molecular genotyping methods and high-throughput platforms in recent years, microsatellites remain among the most popular and informative marker in population genetics (Anderson *et al.* 2000). Microsatellites often referred to as short tandem repeats (STRs) or simple sequence repeats (SSRs) consist of repetitive units of approximately 2-6 base pairs that are arrayed at particular chromosomal locations (Schlötterer 2000). They are generally co-dominant and locus specific with high mutation rates (between 10^{-2} and 10^{-6} per locus per generation, and on average 5×10^{-4}). As consequence of their elevated mutation rates, microsatellites are typically highly polymorphic and therefore, individual

variations are manifested as repeat number differences. They are considered important neutral markers, as they are not subjected to selection pressure from host immunity. The neutral source of microsatellite polymorphism is replication slippage, which is a commonly observed replication error in repetitive sequences that occurs when the new strand mis-pairs with the template strand (Russell *et al*. 2006).

The *P. falciparum* genome contains over 900 microsatellites markers, predominantly having dinucleotide, trinucleotide or tetranucleotide repeats, with an average of one microsatellite locus every 2–3 kb (Anderson *et al*. 1999). Such high abundance seems to correlate positively with genome adenine-thymine (AT) content, which is extremely high (average AT content reaches 95% in repetitive units). As a consequence, microsatellites have increasingly been used for genetic mapping, genetic structure analyses and genetic diversity studies (Ellegren 2004; Mittal 2009; Ferdig and Su 2000). In malaria studies, the highly polymorphic nature of this marker permits detection of polyclonal infections (Anderson *et al*. 2000), which is important in describing the history of endemicity and the stability of transmission with a specified foci (Branch *et al*. 2011; Sutton *et al*. 2011). Moreover, they have the potential to provide contemporary estimates of parasite migrations (Rebaudet *et al*. 2010), distinguish relatively high rates of migration from panmixia (Razakandrainibe *et al*. 2005) and estimates of genetic relatedness (Selkoe and Toonen 2006). Due to high mutation rates, microsatellites are highly associated with sites of recombination and therefore, they are useful for detecting recent population dynamics because of the accumulated mutations, which is the hallmark of population expansion (Morin *et al*. 2004).

Despite many advantages, microsatellite markers also have several challenges and pitfalls that at best complicate the data analysis, and at worst greatly limit their utility and confound their

analysis. However, all marker types have some downsides, and the versatility of microsatellites to address many types of ecological questions outweighs their drawbacks for many epidemiological applications. Fortunately, many of the drawbacks common to microsatellites can be mitigated by adopting novel genotyping alternatives or carefully optimizing genotyping parameters of the marker.

2.9 Detection methods for fragments generated using microsatellites

Microsatellite markers were first genotyped by Litt and Luty (1989). The initial use of this marker extensively relied on manual methods of fragment separations such as agarose gel electrophoresis and estimation of alleles sizes with naked eye under UV-light after staining with ethidium bromide. However, differentiation of alleles with as little as 2bp size difference was difficult on such a matrix. Thus, polyacrylamide gel electrophoresis (PAGE) was adopted to offer greater fragment resolution and accurate allele sizing and agarose gel electrophoresis was rapidly replaced by sophisticated and semi-automated PAGE platforms such as ABI373 and ABI377 (Applied Biosystems). However, mobility of the fragments on PAGE platforms was confounded by sequence composition of the repeats units. For instance, CA strands tends to move faster on PAGE than GT strands resulting into stutter bands which adversely affected fragment analysis. In population analysis using microsatellite, precise allele calling is an important aspect to avoid genotyping errors. Fortunately, the emergence of fluorescent dye chemistry (FAM, HEX, TET, ROX, JOE) amenable to laser-based detection systems in mid 90s revolutionized this field (Schuelke 2000). This lead to introduction of capillary electrophoresis fluorescence detection systems by various manufactures, such as Applied Biosystems, Amersham Biosciences and Beckam Coulter to improve fragment resolution and precision (Butler *et al*. 2004).

In capillary electrophoresis, forward or reverse primer of one of locus is directly labeled with a fluorescent tag. Then this fluorescence is incorporated into PCR products, which can be detected in the capillary sequencer upon laser excitation. The relative bp sizes of the fragments are estimated in relation to migration time of an internal fluorescent size standard using specific software such as GeneMapper. This technique has high resolution of one bp and results are reproducible. The use of this method, however, can be costly due to high price of the fluorescent label, which must be carried by one of the primers in the primer pair (Wenz *et al*. 1998). Meanwhile, a novel procedure in which three primers are used for the amplification of a defined microsatellite locus was proposed and improved by Schuelke (2000). This system is based on a forward primer with M13 tail at its 5' end, a reverse primer and the universal fluorescent primer. The M13 tail is sequence derived from bacterial vector (Steffens *et al*. 1993). Although the M13 sequence is the most widely used tail, any sequence that is not homologous to the target genome could be used. In multilocus analysis, the universal primer are labeled with different fluorescent dyes and combined to co-amplify multiple loci through multiplex PCR (Schuelke 2000). Such approach increases the amount of information generated per assay and reduces labor costs.

Unfortunately, the use of these multiple fluorescent dyes considerably increases the cost technique especially when a large number of loci are involved. Moreover, the use of these multiple fluorescent dyes with large differences in fluorescence intensities may cause genotyping inaccuracies due to the inability of genotyping software to fully correct for background fluorescence, which results from the overlap of the emission spectra of the dyes (Liljander *et al.* 2009). This phenomenon is called "pull-up" and results in the detection of allele where no dye-labeled amplicon actually exists. Such challenges may hinder laboratories with limited budgets and technological knowhow from adopting this technique. In order to overcome these

difficulties, Li *et al.* (2007) introduced an economical procedure for typing large number of microsatellite loci. This technique incorporates the concept of Schulke (2000), but rather than using multiple fluorescent dyes under multiplex PCR, only one fluorescent dye is needed for all microsatellites to be typed. Effectiveness and robustness of this assay had been validated on rodent malaria parasite *P. yoelii*, however, it remains unutilized approach in other malaria parasites especially human malaria parasites. Based on the well-know concept of Li *et al*. (2007), our aim was to adopt this methodology and optimize it for genotyping *P. falciparum* microsatellite markers.

2.10 Principle of single fluorescent labeled microsatellite assay

A widely relied method for typing microsatellites of *P. yoelii* (Li *et al.* 2007) parasites was adopted to genotype *P. falciparum*. This assay combines the principles of the tailed primer method as described by Schuelke (2000) and a single fluorescent dye concept described by Li *et al.* (2007) in two-step PCR amplification. The principle of novel genotyping system is illustrated in this study (Figure 2.3). Briefly, in the first PCR reaction, outer locus-specific primers are multiplexed to co-amplify sets of target marker loci from genomic DNA. The locus-specific primers become fully incorporated into PCR products, which serve as primer binding sites for second stage PCR. Multiplexing helps reduce amplification bias between amplicons by normalizing primer hybridization kinetics (Vos *et al*. 1995) and provides more uniform amplification of the target sequences during the second stage amplification.

In the second stage amplification, three primers namely; Universal M13 forward labeled with HEX dye, M13 tagged forward and a reverse primer amplifies the first PCR products in uniplex as follows. The M13 tagged forward primer anneals to first PCR target sequence but the tagged region does not and selectively amplifies downstream region of the DNA (step 1). At the

next step, the extension product of the sequence-specific forward primer acts as a template for extension by a sequence-specific reverse primer (step 2). The reverse extension product carries a complementary tag sequence at its 3' end, so the universal M13 forward primer labeled with HEX dye anneals to this complementary tag sequence. Since the M13 fluorescent labeled primer does not contain any homology to *Plasmodium* genomic sequence, it anneals only to the extension product of the complementary sequence-specific primer (step 3). During the reaction, the tagged locus-specific primer is depleted owing to a low starting concentration and the fluorescently labeled universal primer is then subsequently incorporated resulting into fluorescently labeled PCR products. At least half of the complementary strands of the tagged sequence served as templates for tagged primer and within a few cycles sufficient dye labeled amplicons are generated for detection with DNA genetic analyzer.

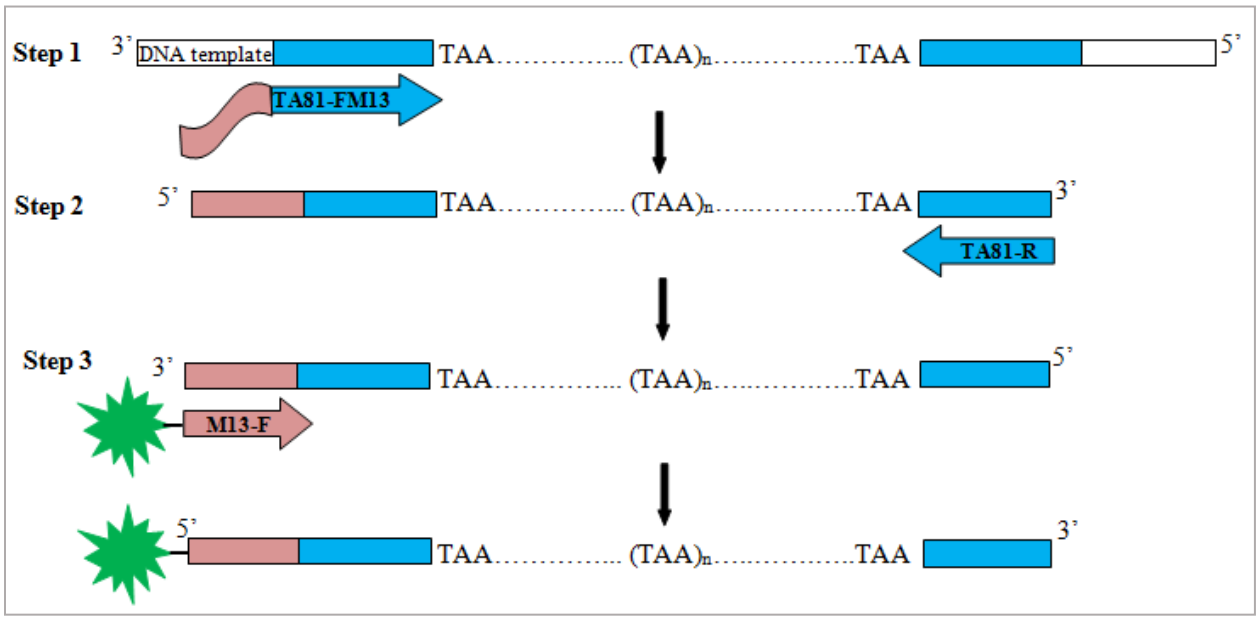

Figure 2.3 Principle of the assay using single fluorescent-labeled primer. The forward primer (TA81-FM13) is designed to have a 19bp tag at it 5' end. The complementary strands generated by reverse primer (TA81-R) with have 19bp tag at its 3' end. The fluorescent-labeled primer (M13-F) can anneal to the 3' end sequence tag and fluorescent dye becomes fully incorporated into the strand.

2.11 Advantages of single fluorescent labeled assay

The single fluorescent-labeled assay provides several technological and practical advantages. First, the ability to perform multiplexing under standardized reaction condition enables any combination of markers to be deployed for multiplexed amplification and can improve assay throughput, reduce cost of reagents, and consumables and template DNA used. Although the initial optimization can be time-consuming, this process is essentially the same as that typically required for any new primer set for conventional PCR.

Secondly, it facilitates accurate detection and allele scoring, since problem associated with "spectral bleed" are reduced. Spectral bleed are genotyping inaccuracies due to the inability of genotyping software to fully correct for background fluorescence artifacts resulting from emission overlaps especially on multi-dye assays (Blacket *et al*. 2012) which results into detection of an allele where no dye-amplicon actually exists.

A third advantage is the substantial saving of cost for fluorescent primer labeling, since the synthesis of specific fluorescently labeled primer for each microsatellite marker is not required. Rather single fluorescent-labeled assay requires only one dye labeled primer to fully genotype sets of markers. Further, the ability to dye label primer with a fluorophore of choice provides flexibility to choose different dyes either HEX, FAM, ROX or TET depending on their cost for genotyping. These two advantages dramatically reduces project costs.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study area

The study samples were collected from permanent residents of four Lake Victoria isalnds: Mfangano, Ngodhe, Kibuogi and Takawiri, and a mainland shoreline region (Ungoye) of Homa Bay County, Western Kenya (Figure 3.1). Mfangano Island is the largest (66km^2) and the most densely populated with a population of approximately 25,000 people followed by Ungoye with a population of about 2,000 (Kenya National Bureau of Statistics 2010). Ngodhe, Kibuogi and Takawiri are small islands with estimated human populations of between 700-1000 depending on the fishing season. The molecular analysis of the collected samples was conducted at CREATES laboratory, Strathmore University.

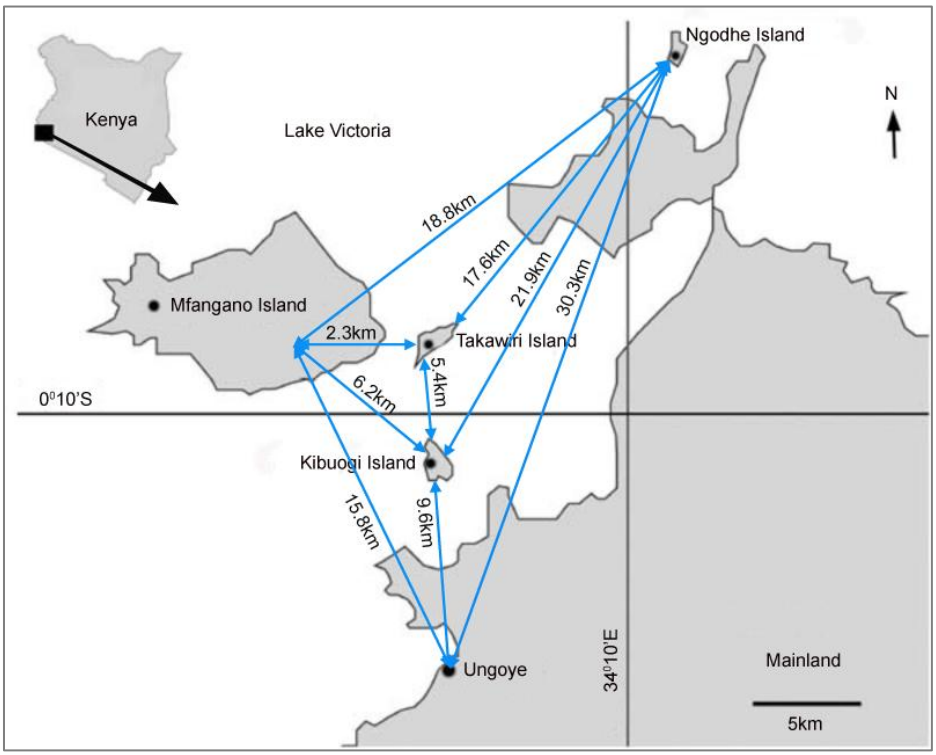

Figure 3.1 A map of Lake Victoria region showing the study sites. Blue lines indicate the geographic distances between the sites.

3.2 Study design

This study was derived from a cross-sectional epidemiological study of *P. falciparum* infection prevalence in resident populations of islands in Lake Victoria, Kenya. To test the hypothesis, experimental procedures conducted in this study were categorized into two phases. Phase one involved developing single fluorescent-labeled microsatellite assay to genotype *P. falciparum*. Here, genotyping principle described by Li *et al*. (2007) was adopted and which has successfully been used in typing rodent malaria parasite *P. yoelii* (Li *et al*. 2009). Selection of microsatellite markers for this study was done based on existing literature, in which eight polymorphic microsatellites markers described and validated by Anderson *et al*, (1999) were used in typing *P. falciparum.*

Four laboratory grown *P. falciparum* lines were used to optimize and validate the methodology. In order to evaluate the sensitivity of the assay, amplification was done on DNAs extracted from red blood cells cultures with varying parasite densities for each *P. falciparum* line. The specificity of the developed assay was evaluated by examining its ability to accurately identify the correct alleles in the four laboratory lines. All PCR amplifications were reamplified and rescored at least thrice to assess the reproducibility detection and size determination of the assay.

Phase two involved genotyping *P. falciparum* field isolates using the validated assay. Field samples identified to be *P. falciparum* positive by rapid diagnostic test and microscopy were selected for genetic analysis. To estimate genetic complexity of the parasite, each locus in each individual isolate, the predominant allele (where multiple alleles were detected) or the only allele (where only a single allele was scored) was counted for population genetic analyses. Summary indices including expected heterozygosity, numbers of alleles, and allelic frequencies per locus

within each population was generated and used to determine genetic diversity, population structure, gene-flow and linkage disequilibrium based on the shared frequencies of DNA genotypes defined by sequence variations.

3.3 Ethical clearance

Ethical approval to conduct the study was granted by the joint Kenyatta National Hospital and University of Nairobi Ethical Review Board (No P7/01/2012). Written informed consent was obtained from all study participants or their parents/guardians.

3.4 Sampling technique

This research was based on purposive sampling technique to obtain experimental subjects. Inclusion criteria; asymptomatic individuals positive for *P. falciparum* by light microscopy and rapid diagnostic test were considered as potential candidates for the study after giving informed consent. Exclusion criteria; Patients under malaria treatment and those who did not give informed consent.

3.5 Experimental techniques

3.5.1 Sample collection

Study was part of an ongoing bigger population-wide epidemiological survey project where finger-prick blood samples were obtained from study participants in accordance with the inclusion criteria. Briefly, 393 blood samples were collected and screened in the field for *P. falciparum* by microscopy and rapid diagnostic test. Parasitemia levels of *P. falciparum* positive samples were determined and recorded for archiving. In total, 188 dried blood spots (DBS) from *P. falciparum* positive samples were prepared by absorbing approximately 50 µl of the collected

blood onto Whatman filter paper (Whatman, Maidstone, UK) and air-drying. DBS were stored at ambient temperature at in sealed zip-lock bags with desiccant awaiting DNA extraction.

3.5.2 Extraction of parasite DNA

From dried blood spots, genomic DNA was extracted using Qiagen QIAamp DNA Mini Kit (Qiagen, Crawley, UK). In brief, three punches (6mm in diameter) were taken from each blood spot and extracted according to manufacturer's instructions. DNA was eluted in a final volume of 200µl, aliquoted and kept under frozen conditions (-20^0C) until use.

3.5.3 Laboratory cultured parasite lines

Laboratory cultured *P. falciparum* lines MAD20, FCR3, K1 and THAI 838 were donated by Osaka University in Japan, through assistance of Dr. Richard Culleton. At this facility, the parasite lines were cultured using standard methods. Parasitemia levels were determined using flow cytometry and red blood cell concentration was measured with a hemocytometer. Red blood cells from culture were mixed with uninfected whole blood to yield varying parasite densities (1 – 5,000 parasites/µl) for each line. Parasite DNA was extracted from each cell line culture using E.N.Z.A blood DNA kit (Omega Bio-Tek, Inc). These extracted DNA samples were used for microsatellite assay optimization and validation procedures.

3.5.4 Determination of DNA yield and quality

The concentration and purity of the extracted DNA samples were determined by measuring the absorbance of 2µl of DNA using NanoVue plus spectrophotometer (Thermo scientific, USA). The ratio of optical density at wavelengths of 260nm and 280nm was used to assess the purity of DNA. A ratio between 1.8 and 2.0 denotes that the absorption in the ultraviolet range is due to nucleic acid. DNA yield was measured at a concentration factor of 50ng/µl.

3.5.5 Testing for *Plasmodium falciparum* in extracted DNA samples

Identification of *P. falciparum* in the extracted DNA samples was done through PCR-based procedure. PCR amplification targeting cytochrome c oxidase III (cox3) gene in mitochondrial genome of *Plasmodium* was performed using MtNst_falF (5′-GAACACAATTGTCTATTCGT-ACAATTATTC-3′) and MtNst_falR (5′-CTTCTACCGAATGGTTTATAAATTCTTTC-3′) primers under PCR condition previously described by Isozumi *et al.* (2015).

3.5.6 Microsatellite marker loci

Trinucleotide repeat microsatellite loci previously described and published for use in genotyping filter-paper blood samples (Anderson *et al*. 1999) were chosen. These polymorphic loci are distributed throughout the *P. falciparum* genome. They included (chromosome assignments are given in parentheses): TA1 (Chr6), TA40 (Chr10), TA42 (Chr5), ARA2 (Chr11), Pfg377 (Chr12), TA87 (Chr6), TA109 (Chr6) and TA81 (Chr5).

3.5.7 Primer modification and preparation

These marker loci consist of conserved locus-specific sequences flanking the tandem repeats region in haploid *P. falciparum* genome. The nested primer sets sequences previously described by Anderson *et al*. (1999) were used to amplify the selected microsatellite marker loci. Here, the genotyping principle described by Li *et al*. (2007) for typing rodent parasite *P. yoelii* using single fluorescent-labeled primer was adopted. Each specific inner forward primer was modified to have a 19 bp M13 sequence (5′-CACGACGTTGTAAAACGAC-3′) (tags) added to the 5′ end matching the conserved region of the microsatellite repeat (Figure 3.2). A stand alone universal M13 forward primer with the same 19-basepair sequence and that binds tags was labeled with HEX fluorescent dye. This single fluorescent-labeled universal M13 forward primer was singly used across all the eight microsatellite marker loci.

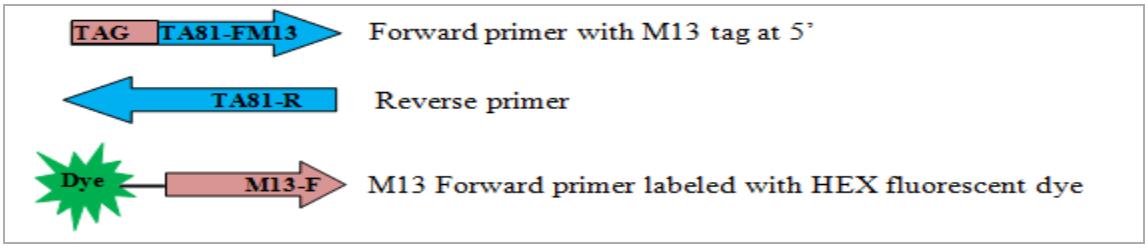

Figure 3.2 Diagrammatic representation of detection primer set for TA81 loci. The forward primer is modified to have a 19bp sequence tag at its 5' end, which bears no sequence homology to the *Plasmodium* genome.

The primers were ordered from FASMAC commercial company, Japan. They were dissolved in DNase-RNase free water to produce a 100µM stock solution for both fluorescent-labeled universal primer and locus- specific primers. All primer stocks were diluted ten-fold to form 10µM working solutions then stored in small aliquots at -20^{0}C. Details on the primer sequences, accession number and expected fragment size can be found in appendix I.

3.5.8 Genotyping of microsatellite loci using single fluorescent-labeled primer

A two-round hemi-nested PCR was used to amplify eight microsatellite loci from DNA isolates using a single universal M13 fluorescent-labeled primer as reporter primer. First-round PCR reaction was multiplexed, such that reactions for TA1, TA87, TA40, TA42 and ARA2, PFG377, TA109, TA81 marker loci were co-amplified in the same reaction tube using the outer species-specific primer sets.

Both the first and second PCR reactions were carried out in a total volume of 20µL per sample per locus and contained 1x Ex *Taq* buffer, 0.05 units of Ex *Taq* polymerase (Takara, Kyoto, Japan), 0.2mM dNTPs mix, 0.2µM of each locus-specific primer and 10ng of template DNA or 2µl of primary PCR products. PCR was performed in a GeneAmp 9700 instrument (Applied Biosystems, Foster City, CA) under the following conditions: An initial denaturation phase at 94^{0}C for 2 minutes, followed by 25 cycles of denaturation at 94^{0}C for 20 seconds,

annealing at 45^0C for 30 seconds, and extension at 65^0C for 40 seconds and a final extension at 65^0C for 2 minutes.

In the second PCR reaction, each marker loci was amplified separately using the inner reverse primer, M13 tagged forward primer and M13 universal forward primer coupled with HEX fluorescent dye under the same conditions except that the amplification cycles were increased to 35. Negative controls (no DNA template) and positive control *(P. falciparum* 3D7) were used in each run.

3.5.9 Fragment analysis by capillary electrophoresis

Fragment analysis of the amplified PCR products were done at the Macrogen Laboratory South Korea, using 3730 DNA Analyzer (Applied Biosystems, Foster City, CA). At the facility, analysis of data files was performed to achieve electropherogram peaks with 100 - 20,000 relative fluorescent unit (rfu) range. Samples that surpassed 20,000rfu mark were diluted 1:20 in DNase-free water to obtain peaks within the set range. The resulting raw data files were sent by "post" to CREATES, Kenya for subsequent analysis.

3.5.10 Microsatellite calling

PCR fragment sizes of the received raw data files were called using Genemapper software version 5.0 (Applied Biosystems, Foster City, CA). Briefly, all data files were imported into a new project within the software application. A panel, which contained all the eight microsatellite markers assayed, was created. Into this panel, a marker, which defines the name, dye color, repeat length and fragment size range (bp) of each microsatellite locus, was also prepared. Unlike the current marker calling procedure, that requires selection of multiple dye colors, green dye color was selected since the fragments were labeled with HEX dye (green in color). "Bins"

which corresponded to three base pair intervals (the size of microsatellite motif) were constructed. In the microsatellite analysis method of the application software, 100rfu was set as the minimal peak threshold for allele calling. Samples were evaluated after initial analysis for off-scale data, background fluorescence and sizing errors. Control (*P. falciparum* 3D7) which is run on each plate were inspected for the same artifacts to derive peak morphology. The alleles sizes of 3D7 were used to correct run to run variation among capillary electrophoresis runs. Microsatellite allele sizes were imported into excel sheet. Allele peaks that were spaced at intervals corresponding to the trinucleotide repeats in all eight loci genotyped were considered as true alleles.

3.5.11 Determination of sensitivity, specificity and reproducibility of the assay

The sensitivity of the assay was determined by testing its ability to amplify microsatellite loci in DNA samples of *P. falciparum* lines (KI, MAD20, FCR3 and THAI 838) with parasite densities corresponding to 5,000, 1,000, 500, 100, 50, 10, 5 and 1 parasites/µl. To simplify results interpretation and discussion in this study, these parasite densities were classified into three categories: high (1000- 5000), medium (50-500) and low (1-10). A parasite density with electropherogram peaks below the cut-off value of 100 rfu was deemed as the detection limit of the assay. Specificity was determined by examining ability of the assay to accurately identify the correct alleles in the four *P. falciparum* lines. Any observed alleles not expected based on the known genetic composition of the four laboratory-grown lines were designated false-positive alleles. The four *P. falciparum* lines at each parasite density were re-amplified and rescored at least three times to assess the reproducibility of detection and size determination of the system. The results were analyzed with reference to the initial data to ascertain concordance.

3.5.12 Genotyping of field isolates

The optimized and validated single fluorescent-labeled microsatellite assay was used to genotype DNA samples extracted from filter-paper parasitized blood samples collected from Lake Victoria islands. These samples were genotyped under same PCR procedure and conditions as described above.

3.6 Analysis of parasite population characteristics

3.6.1 Interpretation of alleles

For each individual isolate, only the predominant allele or the single allele at each locus was scored and counted for subsequent analyses. Since all eight markers are single-copy loci (haploid), presence of one or more additional alleles at a particular locus was interpreted as a multi-infection. Any additional allele(s) present due to multiple parasite infections was recorded if the peak was at least one-third the height of the predominant allele but the allele(s) did not count toward the analyzed sample. This because inclusion of these additional allele(s) from multi-infected isolates would result in a biased estimation of allelic frequencies in the populations (Anderson *et al*. 1999). Additionally, it is also impossible to match the different alleles obtained at each locus to construct a valid genotype.

3.6.2 Measure of genetic diversity

The overall genetic diversity in each of the geographic locations was assessed using Arlequin version 3.5.1.2 software (Excoffier *et al*. 2005) by determining the number of alleles per locus, allelic richness, and expected heterozygosity, calculated from allelic frequencies of the eight microsatellites. Number of alleles per locus, allelic frequencies, and expected heterozygosity were used as measures of level of polymorphism in the loci and to determine the

diversity of the populations. The expected heterozygosity index (*He*), was calculated as *He* = $\frac{n}{n-1}(1-\sum p^2)$, where p is the frequency of each different allele at each locus and n is the number of alleles in the sample. *He* is the probability that two alleles sampled from a population are different. Thus heterozygosity has a potential range from 0 meaning that all alleles are identical (no polymorphism) to 1 that all alleles within the sampled population are unique (highly polymorphic). Since the number of alleles per locus is highly dependent on sample size and the fact that different numbers of isolates were analyzed in the five sites, it was important to normalize data on the basis of the smallest sample size (in this case, Ngodhe with 17 isolates). We therefore computed allelic richness based on a rarefaction method to allow the comparison of genetic diversity across different samples sizes using FSTAT version 2.9.3.2 software (Goudet 1995).

3.6.3 Assessment of multiple infections

The multiplicity of infection (MOI) was defined as the mean number of genetically distinct parasites genotypes co-infecting an individual. MOI was estimated for each isolate from the locus with highest number of alleles. An infection was categorized as a multiple genotype infection if more than one peak was detected by any of the markers and where the additional peak was at least one-third the height of the primary peak. A single infection was categorized as one with only one peak detected at all the genotyped loci. Mean MOI was calculated based on the total number of parasites genotypes detected divided by the number of isolates analyzed.

3.6.4 Characterizing parasite population structure and gene flow patterns

Population genetic structure was investigated using Wright's pair-wise fixation index (F_{ST}) computed with Arlequin software (Larrañaga *et al*. 2013). F_{ST} index estimates genetic

differentiation by calculating weighted F statistics also known as theta (θ) for each locus and over all based on distinct number of alleles among the isolates (Michalakis and Excoffier 1996). A random permutation test (n = 10,000) was performed to test whether the F_{ST} indices observed significantly differed from zero. To validate the population genetic structure established by F statistics, Structure version 2.3.4 software was used to assign each genotype from the five populations into genetically related clusters. The software assigns individual genotype to a predetermined number of clusters (K) based on allelic frequencies at each locus (Pritchard *et al*. 2000; Pritchard 2010). K values between 2 to 8 within these populations was chosen based on Evanno and others (Evanno *et al*. 2005) calculation model that suitable K value ranges from K = 2 up to the true number of study populations plus three. 20 replicate runs were performed to choose the best *K* value, after a burn-in period of 10,000 steps followed by 10,000 iterations under admixture model and assumed correlated allele frequencies. Structure Harvester version 0.6.94 software was applied to evaluate Delta K values from Structure output files (Earl and VonHoldt 2011). The most likely *K* value was computed by the higher number in the change for *K* (ΔK) according to the method described by Evanno and others (Evanno *et al*. 2005). CLUMPP program v1.1.2 was used to aid the interpretation of cluster results (Jakobsson and Rosenberg 2007) and Distruct program v1 was used to facilitate graphical display of population clusters (Rosenberg 2004).

3.6.5 Measure of Linkage disequilibrium among parasite populations

To test for evidence of multi-locus linkage disequilibrium, the standardized index of association (I_A^S) among alleles from all eight loci in each population was computed using LIAN version 3.5 Web interface developed for multilocus haploid data (Haubold and Hudson 2000). Linkage disequilibrium (LD) is the non-random association of alleles across loci and can occur

as a result of a range of processes including population sub-structure and selection. LD analyses were performed to determine the independent assortment of alleles in this study population. LD analyses were performed in two ways: first, a curtailed dataset that excluded all mixed-clone infections so that only single clone isolates with predominant allele at each locus was constructed and analyzed. Second, a dataset that contained multilocus genotypes found once in each population set (unique genotypes) was reconstructed and analyzed, and the results from the two analyses were compared. This index was calculated as $I_A^S = \frac{1}{n-1}(\frac{V_D}{V_E} - 1)$ assuming a null hypothesis of complete random association among loci ($I_A^S = 0$) by Monte Carlo simulation at 10,000 permutations, where V_E is the expected mismatch variance under LD, V_D is the observed variance of the numbers of shared alleles in the population and *n* is the number of examined loci. Significant LD is detected if V_D value is greater than 95% confidence interval of the values derived from the reshuffled data sets.

3.7 Statistical data analyses

To test for evidence of statistically significant differences in the levels of genetic diversity as well as MOI between the five populations, pair wise comparisons were made using the Kruskal-Wallis test (Theodorsson-Norheim 1986).

CHAPTER FOUR

RESULTS

4.1 Concentration and purity of extracted DNA

DNA extracted from *P. falciparum* lines with varying parasite densities were quantified to obtain their concentrations and purity (Appendix II). All samples were of good quality as their optical density ratios reading were ranging from 1.8 to 2.0. Concentration readings indicated sufficient amount of DNA template was available for PCR amplification.

4.2 Identification of *Plasmodium falciparum* in extracted DNA

Analysis of the identity of *P. falciparum* in the extracted DNA samples was done using PCR procedure that targeted cox3 gene of mitochondrial genome. The amplified PCR fragments were resolved in 1% Agarose gel electrophoresis. Gel analysis revealed that all the 188 extracted DNA samples were positive for *P. falciparum* with expected band size of 400bp (Figure 4.1).

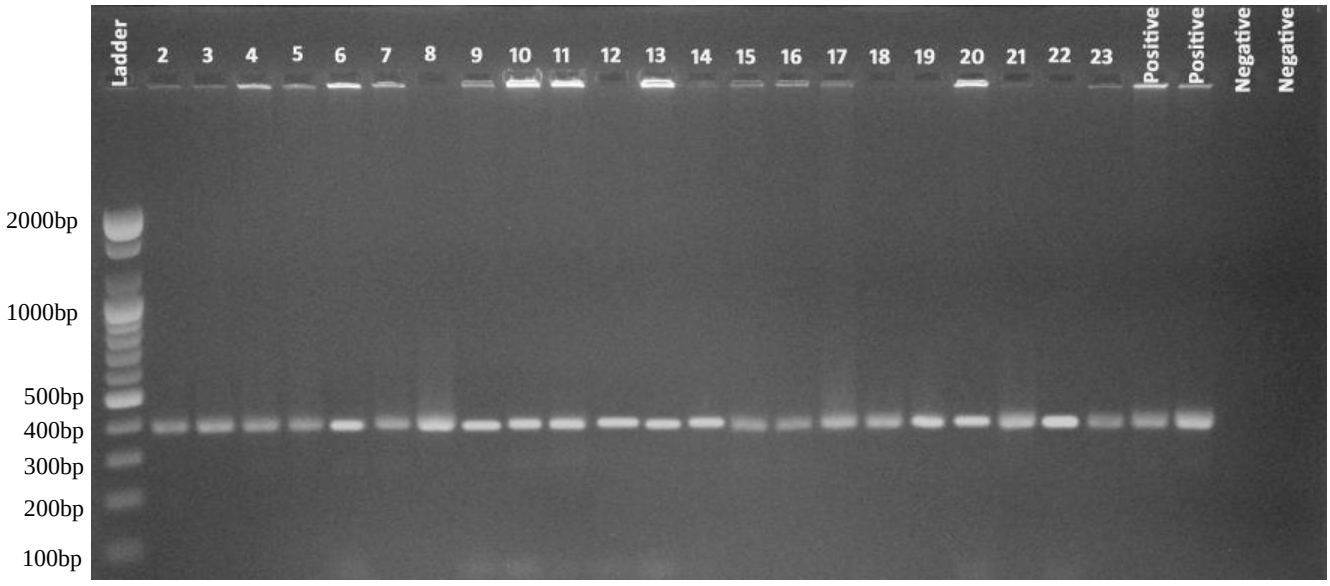

Figure 4.1 Gel image showing positive amplification of *P. falciparum*. Lane 1: 2kb DNA ladder, lane 2-23: DNA samples, lane 24-25: positive controls and lane 26-27: negative controls.

4.3 Microsatellite reaction artifacts

The single fluorescent-labeled primer method was initially set up under the same PCR conditions as described by Anderson *et al.* (1999). However, different artifacts were identified on the electropherograms during preliminary results analysis and interpretation. Some of the artifacts indentified include:

4.3.1 Fluorescent background artifacts

An inherent fluorescent background was observed in electropherograms of laboratory-grown lines analyzed (Figure 4.2a). The relative fluorescent unit (rfu) value for the fluorescent background was by large below the laid cut-off value (100rfu). Moreover, a slightly higher non-specific artifact peaks was observed in the background of the electropherograms of these analyzed samples. These fluorescent backgrounds artifacts were reduced by optimizing the amount of DNA template used in primary PCR reaction. A cut-off value of 100rfu was set in order to simplify identification of true alleles from non-specific artifact peaks in the optimized assay.

4.3.2 Stutter peaks

Besides the primary peak for each allele observed in the electropherogram, distinct patterns of “stutter” peaks were also observed. Amplification of laboratory lines with high parasite densities generated stutter peaks with a repetitive pattern. For instance, amplification of TA40 marker loci on FCR3 line generated stutter peaks with 1bp intervals from the true allele peak (Figure 4.2b).These stutter peaks were relatively lower than the true allele peak.

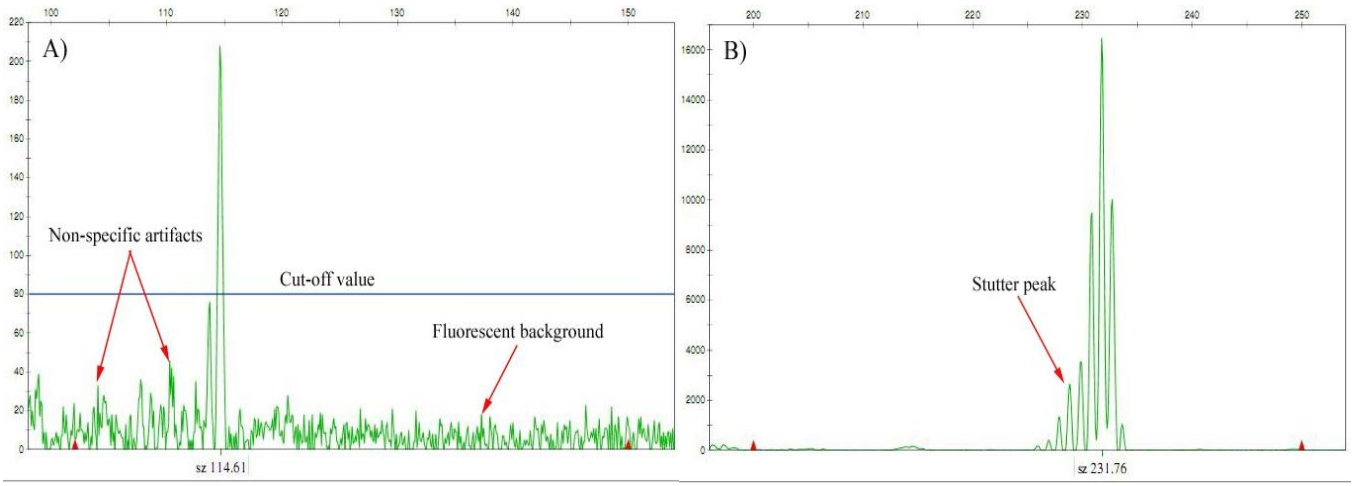

Figure 4.2 Microsatellite reaction artifacts. **A)** Fluorescent background and non-specific artifacts peaks in FCR3 lab-line sample amplified with TA87 marker. **B)** Stutter peaks artifacts in MAD20 lab-clone sample amplified with TA40 primers. The y-axis depicts the rfu while x-axis depicts fragment size in base pair.

4.4 Performance of optimized single fluorescently labeled assay on controls

Several assay parameters in nested PCR reaction and product preparation were optimized before capillary electrophoresis in order to reduce the above artifacts and ensure sustained sensitivity and specificity of the assay.

4.4.1 Detection sensitivity of the assay

The analytical sensitivity of the assay was determined by testing varying parasites densities of *P. falciparum* lab-grown lines containing from 1 to 5000 parasites/µl (Figure 4.3). Overall, the assay showed considerably sensitivity detection range across all loci and lab lines analyzed. Although sensitivity in terms of rfu values varied by loci, in high-density samples (>1000 parasites/µl) the rfu of the allele peaks for each lab line were within the recommended values (<30000 rfu). However, at low parasite densities (<10 parasite/µl), the rfu values were below the cut off value of 100 rfu (Figure 4.3d). The assay was being able to detect up to 50 parasites/µl in all four lines.

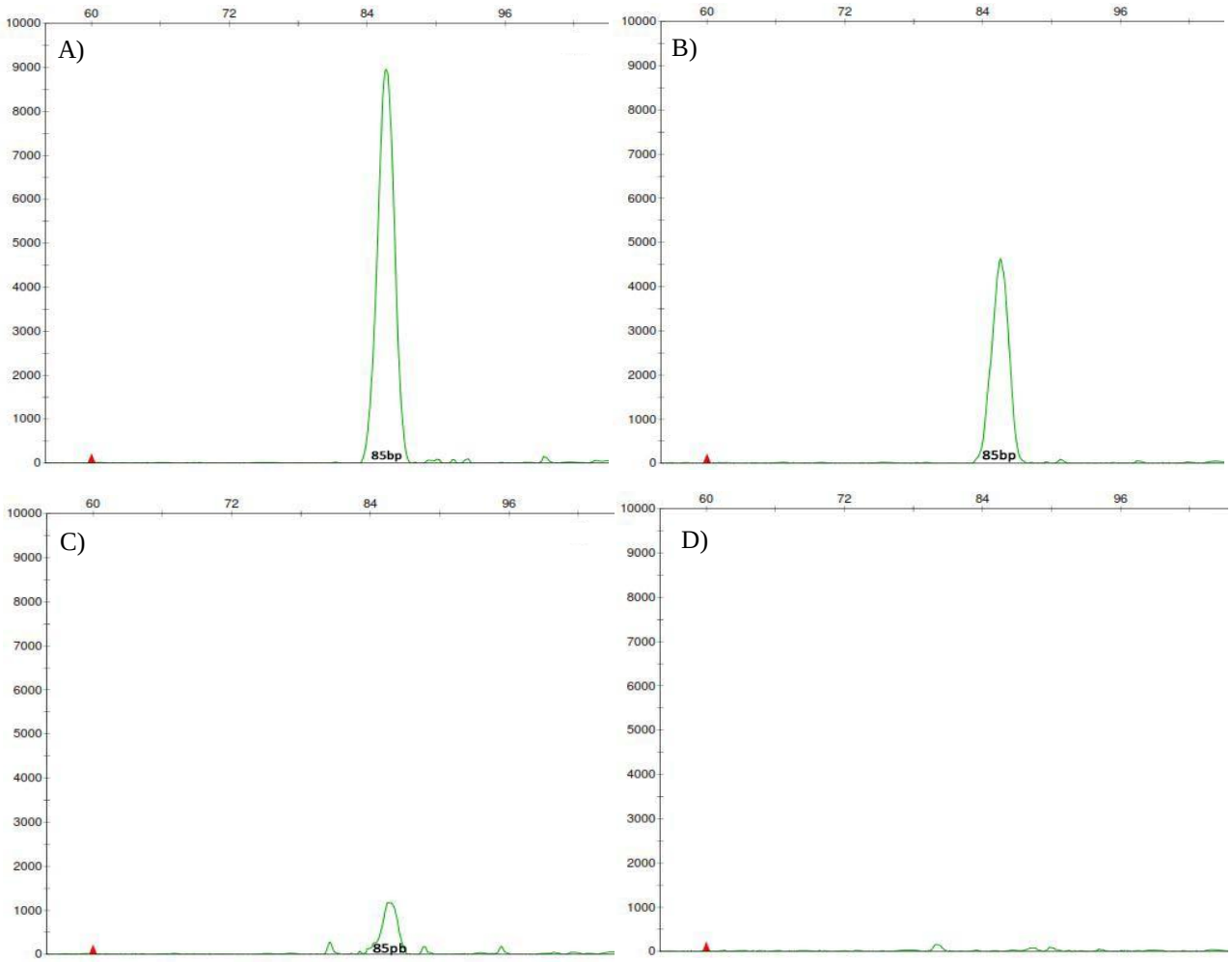

Figure 4.3 Peak profile of THAI 838 line amplified at varying parasite densities. Rfu values decreased proportionally with parasite densities. **A)** corresponds to parasite density of 500, **B)** parasite density of 100, **C)** parasite density of 50 and **D)** parasite density of 10. Below 50 parasites/µl no dectable peak was observed.

4.4.2 Detection of specificity

To determine the specificity of the assay, the ability of the assay to detect false-positives in the controls was investigated. An allele was designated as false-positive if the observed allele peak was not expected based on known genetic composition of lab-line and if it exceeded the laid cut-off value of 100rfu. Of all the four lab-grown lines used, none produced a false-positive allele for the eight microsatellites genotyped (Figure 4.4). The number and intensity of peaks were well characterized. Moreover, no visible peak was detected on the negative control sample, indicating high precision of the assay in mitigating genotyping errors.

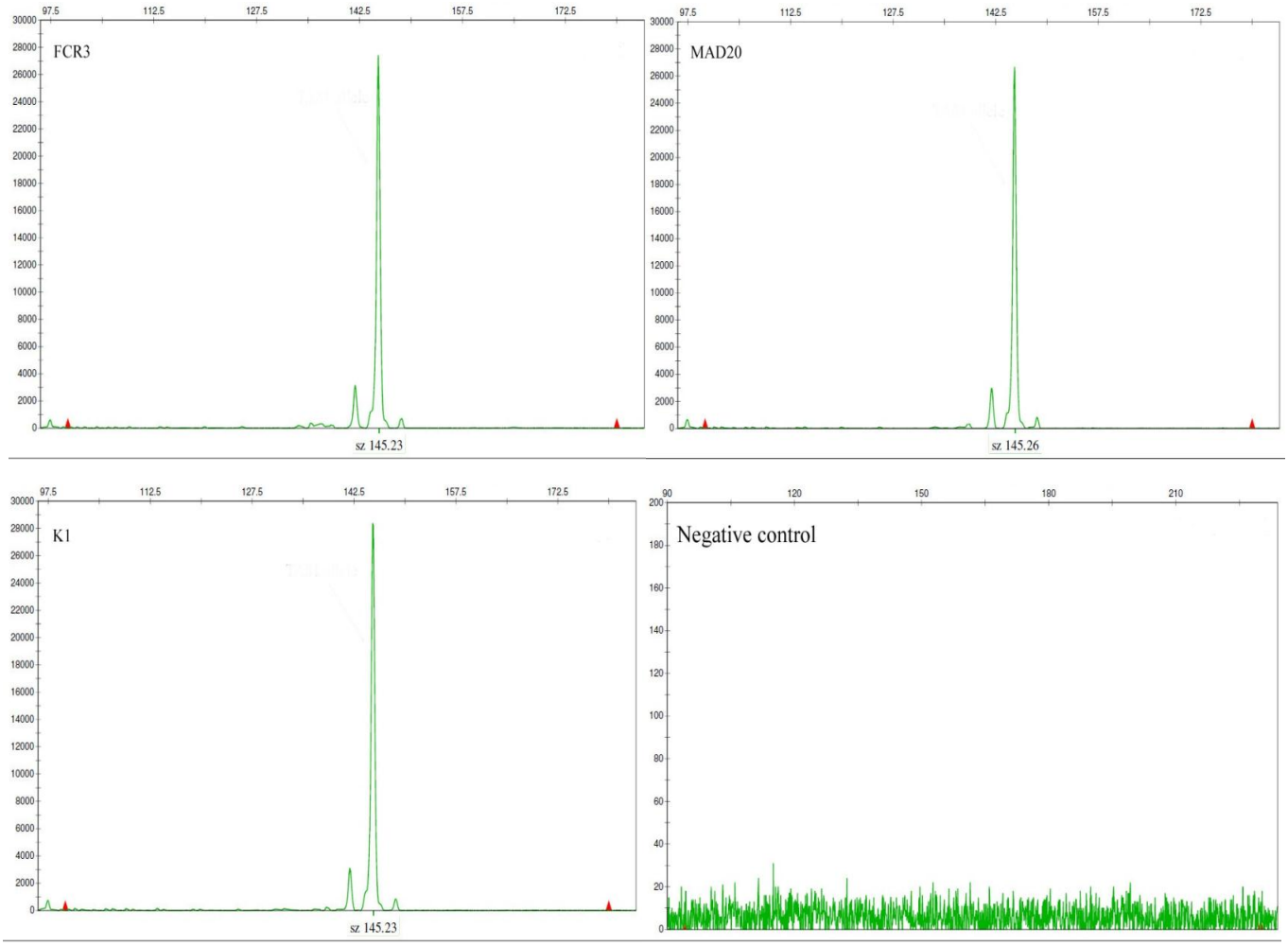

Figure 4.4 Profiles of FCR3, MAD20 and K1 lines amplified at TA81 loci. All controls show well-resolved alleles peaks sized at 145pb. Negative control sample showed fluorescent background signal with no allele detected on locus.

4.4.3 Reproducibility of the assay

The reproducibility of the assay was examined by re-amplifying lab lines of *P. falciparum* in replicates of three for each parasite densities. The samples were re-coded and reanalyzed without reference to the initial results. The number of alleles detected in the eight loci assay for the triplicates containing from 1 to 5000 parasites/µl of the four lab-lines are shown in table 4.1. The assay had reproducibility of three for three for parasite densities above 50, but scores decreased for low parasite densities.

Table 4.1 Reproducibility of the assay in detecting alleles in cultured parasite lines

	Number of alleles detected across the eight loci assayed											
	MAD20			K1			THAI 838			FCR3		
Parasite densities	1st replicate	2nd replicate	3rd replicate	1st replicate	2nd replicate	3rd replicate	1st replicate	2nd replicate	3rd replicate	1st replicate	2nd replicate	3rd replicate
>500	8	8	8	8	8	8	8	8	8	8	8	8
100	8	8	8	8	8	8	8	8	8	8	8	8
50	8	8	8	8	8	8	8	8	8	8	8	8
10	6	5	7	5	5	6	7	5	6	6	5	5
5	2	1	0	0	3	1	3	1	2	1	0	2
1	1	0	0	0	1	0	1	0	0	0	0	1

4.5 Population genetics of *P. falciparum* field isolates

4.6.1 Genetic diversity within populations

In total, 188 isolates: Kibuogi (n = 35), Mfangano (n = 50), Ngodhe (n = 17), Takawiri (n = 36), and Ungoye (n = 50) were successfully genotyped at all the eight loci (TA81, TA87, TA109, ARA2, pfg377, TA1, TA42, and TA40) and a full genotype profile generated (Appendix III). The eight markers examined were observed to be highly polymorphic, with overall number of distinct alleles per locus ranging from 11 (for locus ARA2) to 23 (for locus TA1; Table 4.2). There were no significant differences ($P = 0.16$) in the mean number of alleles per locus between the five parasite populations calculated using Kruskal-Wallis test. Allelic richness normalized as described in the methods, did not differ significantly between populations. This implied that the chosen marker loci were equally informative for both island and mainland parasite populations.

Table 4.2 Number of alleles, allelic richness and allelic diversity of eight loci from five sites. Allelic richness was normalized based on smallest sample size (n = 17).

	Kibuogi			Mfangano			Ngodhe			Takawiri			Ungoye			Total Number		
Locus	***A***	***Rs***	***He***	***A***	***Rs***	***He***	***A***	***Rs***	***He***	***A***	***Rs***	***He***	***A***	***Rs***	***He***	***A***	***Rs***	***He***
ARA2	6	5.58	0.52	7	6.19	0.79	7	7.00	0.71	7	6.31	0.69	11	9.23	0.85	11	8.08	0.80
PFG377	10	8.88	0.86	11	8.39	0.84	11	11.00	0.95	13	11.88	0.93	9	8.14	0.77	17	11.64	0.90
TA1	11	9.80	0.86	11	9.69	0.89	7	7.00	0.82	11	9.55	0.88	15	11.51	0.88	23	12.02	0.89
TA40	12	11.27	0.93	14	11.31	0.90	7	7.00	0.85	13	11.04	0.83	15	12.49	0.92	17	12.73	0.91
TA42	7	6.40	0.72	10	7.68	0.63	8	8.00	0.90	6	5.42	0.69	7	5.77	0.48	14	7.52	0.66
TA81	12	10.99	0.91	10	8.08	0.85	8	8.00	0.91	13	11.63	0.92	11	8.94	0.82	21	11.36	0.88
TA87	8	7.47	0.76	9	7.58	0.81	8	8.00	0.87	10	8.86	0.87	8	6.88	0.76	12	7.93	0.81
TA109	9	7.61	0.63	11	8.96	0.85	7	7.00	0.84	11	9.59	0.78	9	8.25	0.86	16	9.43	0.83
Mean	**9.38**	**8.50**	**0.78**	**10.38**	**8.48**	**0.82**	**7.88**	**7.88**	**0.86**	**10.50**	**9.29**	**0.82**	**10.63**	**8.90**	**0.79**	**16.38**	**10.09**	**0.84**

A = Number of alleles, *Rs* = Allelic richness and *He* = Allelic diversity

Allelic diversity per microsatellite loci was estimated by expected heterozygosity (*He*) based on frequency of alleles generated (Appendix IV) at each of the eight loci. On average, *He* index was 0.84, reflecting high genetic diversity among the isolates. The highest mean *He* values were observed in Ngodhe (*He* = 0.86) and the lowest in Kibuogi (*He* = 0.78; Table 4.2). This genetic variability though different by loci and by studied sites, did not meet the threshold for statistic significance. The mean number of alleles per locus (*A*) was highest in Ungoye (*A* = 10.63) and lowest in Ngodhe (*A* = 7.88).

4.6.2 Multiplicity of *P. falciparum* infection

The multiplicity of infection (MOI) was assessed based on the proportion of individual isolates with multiple alleles for each site. Since *Plasmodium* remains haploid in the human host, detection of multiple alleles at any of the genotyped loci in a sample indicates presence of multiple genotypes. Of the 188 samples analyzed, 151 isolates had at least two or more alleles detected by one of the eight loci. The proportion of isolates with multiple genotypes was highest in Ungoye (0.88) and lowest in Mfangano (0.72; Table 4.3). Further, the highest number of

parasite genotypes co-infecting an individual was six and was observed in Mfangano and Takawiri islands (Table 4.3). The highest, and lowest mean MOIs were recorded in Ngodhe and Kibuogi respectively although the differences across the regions were not significant by the Kruskal-Wallis test ($P = 0.81$).

Table 4.3 MOIs of *P. falciparum* infections in the five study areas

Site	Number of isolates genotyped	*Isolates with particular number of genotypes						#Mean MOI	Proportion of isolates with >1 genotype
		1	2	3	4	5	6		
Kibuogi	35	7	10	12	5	1	0	2.51	0.80
Mfangano	50	14	9	13	12	1	1	2.60	0.72
Ngodhe	17	3	2	3	5	4	0	3.29	0.82
Takawiri	36	7	10	14	4	0	1	2.53	0.81
Ungoye	50	6	6	13	22	3	0	3.20	0.88

*The isolates per study site with the minimum (1) and maximum (6) detected number of genotypes; #Mean MOI calculated as total number of parasite genotypes detected per number of isolates analyzed in each study site.

4.6.3 Genetic differentiation among the islands and population structure

Individual pair-wise differentiation (F_{ST}) values for all eight-marker loci across the five parasite populations ranged from 0.004 (for locus TA87) to 0.117 (for locus ARA2). Averaged across all marker loci, gene divergence among the different populations was 0.044 indicating that approximately 5% of the overall allelic variation is due to differences observed between the five populations. Figure 4.5 shows the F_{ST} indices between the five parasite populations. Overall, the F_{ST} values were low ranging from 0.014 to 0.081, and were significantly different from zero for all population comparisons ($P < 0.05$). The low F_{ST} indices observed virtually indicates absence of population sub-structuring among the studied populations.

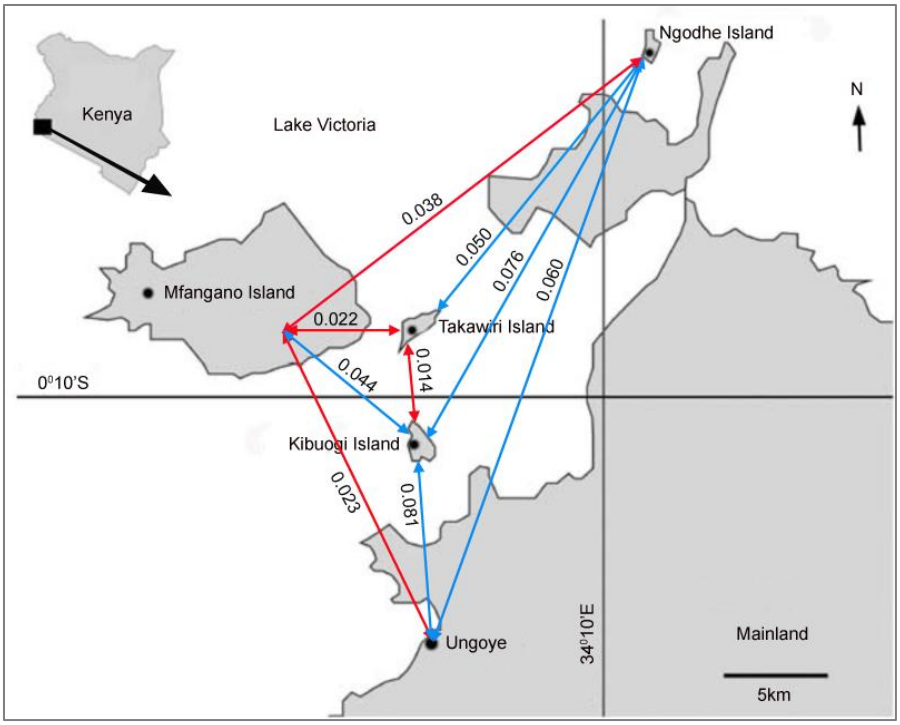

Figure 4.5 Levels of genetic differentiation between the five *P. falciparum* populations. Red lines indicate possible routes of parasite transmission as determined by low F_{ST} index between the sites.

Cluster analysis using structure software was then performed in order to determine the most accurate number of parasites with similar microsatellite genotypes circulating in these populations. Structure analysis identified three putative clusters *(*ΔK= 32.45) with much admixture in the five regions (Figure 4.6). The analysis assigned malaria parasites to particular genetic cluster based on membership coefficients, for each of the geographic regions representing parasite populations. Each cluster is represented by color codes; dominant cluster (orange), common cluster (blue), and rare cluster (yellow). The common cluster of genotypes included most of the parasites from Ngodhe (60.68%). The dominant cluster assigned 63.81% of isolates from Kibuogi and the rare cluster grouped 48.94% of isolates from Ungoye.

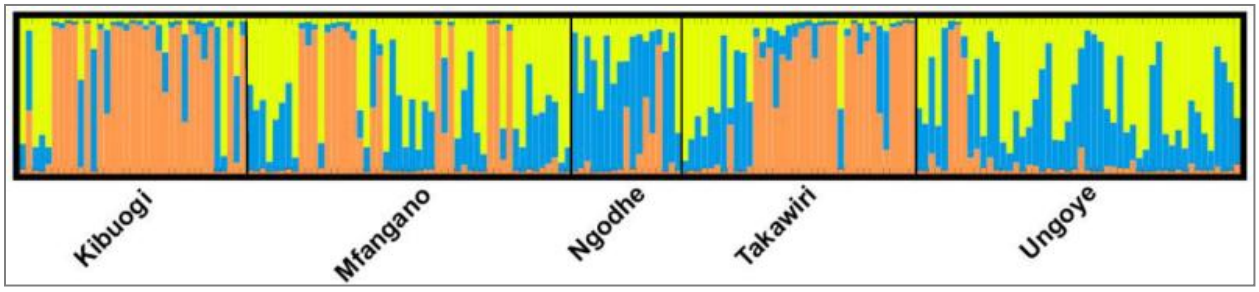

Figure 4.6 Structure analysis and assignment test for 188 *P. falciparum* genotypes. Each bar represents the proportion of each genotype in the defined clusters, each cluster being indicated by a different color; dominant cluster (Orange), common cluster (Blue), and rare cluster (Yellow).

4.6.4 Multilocus linkage disequilibrium in *P. falciparum* populations

A measure of non-random association among loci (multilocus LD) was calculated on mixed-clone infections and separately for single-clone infections using the index of association (I_A^S). For this analysis, mixed-clone infections included the whole dataset comprising 151 isolates. All these isolates were single representatives of each genotype in each population dataset (unique genotypes). Single-clone infections comprised a curtailed dataset of 37 isolates with the predominant allele at each locus. The latter analysis was used to confirm LD in the absence of genotypes detected from multiple infections, which can result in higher estimates of recombination and thus bias against the detection of LD. The degree of LD was highly variable in the five parasite populations. The overall I_A^S values ranged from 0.026 to 0.286 when single-clone infections were analyzed and -0.005 to 0.073 when mixed-clone infections were examined (Table 4.4). No evidence for multi-locus LD (P = 0.18) was found in all five populations when single-clone infections were analyzed. Mixed-clone infections only showed significant associations in Kibuogi and Takawiri populations.

Table 4.4 Multilocus linkage disequilibrium among *P. falciparum* populations

	Mixed-clone infections			Single-clone infections		
Population	**No.**	**I_A^S**	**P-value**	**No.**	**I_A^S**	**P-value**
Kibuogi	28	0.043*	0.01	7	0.095	0.08
Mfangano	36	0.009	0.10	14	0.026	0.16
Ngodhe	14	0.036	0.09	3	0.286	0.31
Takawiri	29	0.073*	0.01	7	0.027	0.31
Ungoye	44	-0.005	0.63	6	0.068	0.11
Total	**151**	**0.019***	**0.01**	**37**	**0.013**	**0.18**

No. indicates number of isolates for each measure. *Significant levels for a test of departure from 0 for IAS values ($P < 0.05$).

CHAPTER FIVE

DISCUSSION

5.1 Performance of single fluorescent-labeled microsatellite assay

A novel single fluorescent-labeled assay described in this study provides a number of attractive practical features as molecular tool investigating population genetics of *P. falciparum*. It is well known that sensitivity of PCR technique largely depends on various factors such as quality of DNA, parasite densities, amplification conditions, reagents and characteristics of the target locus. Jelinek *et al.* (1996), reported that sensitivity of PCR was much linked to parasite density. Considering the influence of such confounding factors, several experimental assays have demonstrated sensitivity detection limit of 50-100 parasites/µl on nested-PCR amplification of *P. falciparum* (Scopel *et al*. 2004; Coleman *et al*. 2006). The detection limit of the optimized single fluorescent-labeled assay was 50 parasites/µl blood, which was comparatively similar to those described in other experimental studies. However, Li *et al*. (2007), used this assay in typing *P. yoelii* microsatellites, but did not present any data on the assay's sensitivity.

There are several problems associated with microsatellite genotyping systems including the accuracy in estimating the sizes of PCR amplicons, which are not always consistent with the sizes predicted from the genome sequences. The M13 primer was expected to add 19 nucleotides to the PCR products. Comparison of alleles sizes from the predicted *P. falciparum* genome sequences and those estimated from our lab-grown lines showed that all of marker loci had alleles with added sizes of 19bp thus indicating successful incorporation of M13 tags. In this study, all laboratory-lines genotyped gave robust products of expected size on respective marker loci with no mispriming demonstrating specificity of the assay. Another pervasive problem is null alleles. Null alleles are caused by mutations in the primer binding regions especially the 3'

end and prevent amplification of the affected alleles. According to Chapuis and Estoup (2007), the presences of null alleles erroneously inflate genetic differentiation levels and affect population genetic analyses. However, this assay revealed absence of null alleles in all amplified loci, indicating high precision of the assay in mitigating genotyping errors.

5.2 Genetic characteristics of *P. falciparum* populations

This work presents the most extensive findings to date of the population and genetic structure of *P. falciparum* isolates from Lake Victoria islands using microsatellite markers analysis. Other studies had observed that the prevalence of malaria in this lake region is geographically and seasonally variable (Idris *et al*. 2014). This study was aimed at examining levels of parasite genetic diversity, population structure and gene flow, as this would allow tracking of parasite migratory routes in the study sites before the implementation of interventions.

5.3.1 Genetic diversity characterized by genotype profiles

This study revealed overall regional expected heterozygosity index value of 0.84 that were substantially higher than those observed in low transmission areas of South America: Brazil, (*He* = 0.14 - 0.62) (Hoffmann *et al*. 2003) and Southeast Asia: Thailand, (*He* = 0.65) (Pumpaibool *et al*. 2009). However, these heterozygosity index values were comparatively similar to those previously described in other countries with high malaria transmission intensities such as: Vietnam (*He* = 0.52- 0.91), Nigeria, (*He* = 0.79), Congo, (*He* = 0.80), and Uganda (*He* = 0.76) (Oyebola, *et al*. 2014; Durand *et al*. 2003; Anderson *et al*. 2000). Similarly, the results showed a high mean number of alleles (*A* = 16.38) for the five parasite populations. Generally, the number of alleles detected per locus is likely to be high in regions of high endemicity and low in regions of low endemicity (Anderson *et al*. 2000). Based on this, the number of alleles

recorded in this Lake region were comparable to levels observed in other high endemic regions of sub-Saharan Africa with allelic mean number ranging from 5.3 to 13.5 (Bogreau *et al*. 2006; Zhong *et al*. 2007).

5.3.2 Measurement of infection complexity

A large proportion of isolates reported here were multi-infected, with mean MOIs levels varying among the parasite populations. The high levels of multi-infections arguably resulted into high genetic diversity observed in this lake region. Overall, a mean MOI of 2.83 was recorded across populations and was comparatively similar to that described in other African areas; Guinea Bissau (MOI = 2.6), Gambia (MOI = 2.4), Senegal (MOI = 2.2) and Djibouti (MOI = 2.2), (Mobegi *et al.* 2012; Bogreau *et al.* 2006). Assuming each genotype observed is transmissible to the vector (mosquito) during a blood meal, the rate of cross-over fertilization and genetic recombination is likely to be high. The high rates of MOI reported in this study could be attributed in most part to high transmission rates in this malaria endemic part of Kenya.

5.3.3 Characterizing the population structure and gene flow

The genetic differentiation index (F_{ST}) revealed evidence of low sub-structuring among the studied populations. Most of the F_{ST} values were comparable to those revealed between *P. falciparum* populations from endemic regions in Africa (Anderson *et al*. 2000; Mobegi *et al*. 2012; Oyebola, *et al*. 2014) but substantially lower than those from less endemic regions such as Philippines (Iwagami *et al*. 2009), Papua New Guinea (Schultz *et al*. 2010), and Brazil (Machado *et al*. 2004). The level of genetic differentiation was high (F_{ST} = 0.081) between the mainland region (Ungoye) and Kibuogi Island despite their geographic proximity (9.6km apart) whereas the next closest site from the mainland region (Mfangano, 15.8km) recorded a relatively low level of genetic differentiation (F_{ST} = 0.023). Based on these observations, there might be

existence of gene flow barriers between these parasite populations. It has elsewhere been reported that parasite gene flow among populations could highly be influenced by human (Schultz *et al*. 2010; Lum *et al*. 2004) and mosquito (vector) dispersals (Costantini *et al*. 1996). Thus, these migrations could be playing significant role in shaping parasite population structure in the region. Besides migration, geographic barrier between the sites may be another cause of genetic differentiation by partial isolation by distance and breaking down of panmixis. With no air and road networks, migration of inhabitants across these geographically isolated islands is mainly through small fishing boats or ferry services.

Several epidemiological studies have shown that human movement and levels of genetic differentiation have inverse correlation (Lum *et al*. 2007; Lum *et al*. 2004; Schultz *et al*. 2010). For example, Lum *et al.* (2004) demonstrated that frequent human movements between geographically isolated islands were responsible for importation of new parasite genotypes into existing parasite population subsequently resulting in low F_{ST} indices. Thus, F_{ST} indices are pertinent in inferring gene flow through movement of inhabitants. Based on this, the results from this study strongly suggest frequent human traffic between the mainland (Ungoye) and Mfangano island than any other island as shown by F_{ST} value ($F_{ST} = 0.023$). Across the islands, human movement was high between Kibuogi and Takawiri ($F_{ST} = 0.014$). Lum *et al.* (2004) and Pumpaibool *et al.* (2009) have highlighted political and territorial conflicts, linguistic, and cultural diversity as major barriers impeding human interaction which is essential for malarial gene flow. However, with no such political and socio-cultural barriers within residents of this lake basin, routine socio-activities or seasonal migrations of the residents could be the major factors influencing parasite migratory routes. Despite possible direct migratory route between Ungoye (mainland) and Kibuogi Island, human movements are minimal as compared to Ungoye

and Mfangano Island, which are geographically far apart. This frequent human movement between Ungoye and Mfangano could be attributed to a ferry service available on this seaway. Further, high human traffic trends inferred from F_{ST} index between Kibuogi and Takawiri (F_{ST} = 0.014) could be due to immense seasonal fishing activities by migrant fishermen. Although Ngodhe is the furthest site from Ungoye (mainland), the extent of human movement into the region is difficult to confirm due to the small sample size used in this study.

5.3.4 Spatial distribution of parasite clusters

In this study, cluster analyses found a dominant cluster (orange) of genotypes circulating between the Kibuogi-Takawiri islands, which are consistent with the observed low levels of genetic differentiation (F_{ST} = 0.014) between the two sites. Furthermore, these dominant clusters appeared to be circulating between the Kibuogi-Takawiri-Mfangano islands. Based on observation this regular and informal human travel could accelerate the spread of these clusters of parasite genotypes and could have substantial influence on malaria epidemiology in the study areas. An earlier reports in Papua New Guinea (Schultz *et al*. 2010), suggests the importance of human movement in mapping malaria parasite migration routes on islands. The study showed that the "mapped" human migratory routes lead to the spread of malaria parasite and was responsible for seasonal malaria cases in the catchment areas. The results shows importance of mapping possible parasite transmission routes prior to implementation of malaria interventions.

5.3.5 Patterns of Linkage disequilibrium in the populations

Global studies have reported linkage disequilibrium (LD) to be inversely associated with high levels of malaria transmission (Anthony *et al*. 2005; Anderson *et al*. 2000). In areas with high transmission rates, LD is rapidly broken down due to increased proportion of mixed genotypes, leading to crossbreeding and meiotic recombination. On the other hand, low

transmission rates decrease the frequency of mixed genotypes and so inbreeding, which increases LD. Within each population, a large proportion of isolates were multi-infected and as expected, no significant deviation from random allelic association was observed on single-clone infections. The overall association index was 0.013 (P = 0.18) when single infections were analyzed, which was relatively weaker than that reported in low transmission areas (Pumpaibool *et al*. 2009; Larrañaga *et al*. 2013). This concurs with other epidemiological studies that have reported lack of LD in high transmission regions when single-clone parasite infections are analyzed (Mobegi *et al.* 2012; Oyebola *et al*. 2014).

As Kibuogi and Takawiri parasite populations had relatively low levels of MOI (2.51 and 2.53 respectively) compared to the other study sites, it was not surprising to find significant LD when mixed-clone infections were analyzed. Such significant LD could suggest high rates of ongoing inbreeding arising from selfing or as a consequent of high genetic relatedness among the genotypes from the isolates. Indeed, separate studies from Kenya (Razakandrainibe *et al*. 2005), Senegal (Leclerc *et al*. 2002), and the Democratic Republic of Congo (Durand *et al*. 2003) have revealed significant LD despite their high transmission rates. However, caution is needed when interpreting such kind of observations. It has elsewhere been shown that individuals residing in high endemic regions are prone to harbor a mixture of genetically distinct parasite genotypes which may be obtained from the single bite of a mosquito infected with more than one parasite genotype (Fraser-Hurt *et al*. 1999; Rosenberg *et al*. 1990). Since these parasites obtained from mixed-clone infections per isolate may be as a result of single recombination event in the mosquito, they are therefore expected to be closely related skewing the result to a higher and statistically significant LD.

CHAPTER SIX

CONCLUSION AND RECOMMENDATION

6.1 Conclusion

In conclusion, the study found that:

1. The single fluorescently labeled microsatellite assay for genotyping *P. falciparum* was highly sensitive, specific and reproducible. This PCR-based method offers a more practical and acceptable alternative for effective and accurate identification of *P. falciparum* genotypes. Although it extensively relies on capillary electrophoresis, which may not be available in many resource-poor settings, cheap genotyping services are available and so genotyping can be done "by post" and data files returned electronically to respective center. Therefore, there is no reason why this technique should remain exclusively for affluent countries.
2. The genetic diversity of Lake Victoria's *P. falciparum* populations are fairly diverse (*He* = 0.84) and falls within the expectation of a high malaria transmission zone. Similarly, the multiplicity of infections within the lake region was relatively high, a suggestive of high transmission and stable parasite population. Such characteristics may constitute an obstacle to malaria control strategies as genetic polymorphism may facilitate parasite population to adapt to their host and counteract any interventions, such as anti-malarial drugs or vaccines.
3. The population structure of *P. falciparum* in the lake basin was low (F_{ST} = 0.044) with no evidence of sub-structuring. The low level of population variations observed between the study sites is likely to be a consequence of immense gene flows into and out of the islands. Such pronounced parasite gene influx mediated by human movements is likely to represent another hindrance to malaria control. Indeed, high

parasite flux from one site to another could favor the transfer of advantageous parasite traits, such as anti-malarial drug resistant strains. Arguably, malaria control measures should be intensified in the region and that all malaria foci must be tackled simultaneously to reduce parasite reintroduction.

4. The LD among microsatellite loci was relatively low ($I_A^S = 0.013$), a clear indicator of high parasite transmission in the lake site. Significant LD observed in Kibuogi and Takawiri when mixed-clone infection were analyzed could suggest high rates of ongoing inbreeding arising from selfing or as a consequent of high genetic relatedness among the genotypes from the isolates.

6.2 Recommendation

This study represents a first attempt to analyze the genetic characteristic of *P. falciparum* population in Lake Victoria islands using a novel single fluorescent-labeled microsatellite assay. However, the study recommends that:

1. Since malaria is for the most part, endemic in developing countries with limited budgets and techniques for use in malaria research, this assay would ideally be of great benefit to research groups who perform low-throughput genetic analyses with a high number of microsatellite markers. It proved to be highly sensitive, specific and reproducible and it would be applicable in determining the genetic diversity, MOI, population structure and LD of *P. falciparum* populations.
2. On the lead up to malaria elimination goals in Lake Victoria basin, elimination strategies need to be implemented indiscriminately on the entire islands and adjoining mainland region to curtail possible parasite gene flow and subsequent malaria importation through human movement.

REFERENCES

Akter, J. *et al.* (2012). Genotyping of *Plasmodium falciparum* using antigenic polymorphic markers and to study anti-malarial drug resistance markers in malaria endemic areas of Bangladesh. *Malaria journal*, 11, p.386.

Amaratunga, C. *et al*. (2012). Artemisinin-resistant *Plasmodium falciparum* in Pursat province, western Cambodia: a parasite clearance rate study. *The Lancet infectious diseases*, 12(11), pp.851–8.

Anderson, T.J. *et al*. (2000). Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Molecular biology and evolution*, 17(10), pp.1467–82.

Anderson, T.J. *et al*. (1999). Twelve microsatellite markers for characterization of P*lasmodium falciparum* from finger-prick blood samples. *Parasitology*, 119 (2), pp.113–25.

Anthony, T.G. *et al*. (2005). Fragmented population structure of *Plasmodium falciparum* in a region of declining endemicity. *The Journal of infectious diseases*, 191(9), pp.1558–64.

Antinori, S. *et al*. (2012). Biology of human malaria Plasmodia including *Plasmodium knowlesi*. *Mediterranean Journal of Hematology and Infectious Diseases*, 4.

Barillas-Mury, C. and Kumar, S. (2005). *Plasmodium*-mosquito interactions: a tale of dangerous liaisons. *Cellular microbiology*, 7(11), pp.1539–45.

Blacket, M.J. *et al*. (2012). Universal primers for fluorescent labelling of PCR fragments-an efficient and cost-effective approach to genotyping by fluorescence. *Molecular Ecology Resources*, 12, pp.456–463.

Bogreau, H. *et al*. (2006). Genetic diversity and structure of African *Plasmodium falciparum* populations in urban and rural areas. *The American journal of tropical medicine and hygiene*, 74, pp.953–959.

Branch, O.H. *et al*. (2011). *Plasmodium falciparum* genetic diversity maintained and amplified over 5 years of a low transmission endemic in the peruvian amazon. *Molecular Biology and Evolution*, 28(7), pp.1973–1986.

Butler, J.M. *et al*. (2004). Forensic DNA typing by capillary electrophoresis using the ABI Prism

310 and 3100 genetic analyzers for STR analysis. *Electrophoresis*, 25(10–11), pp.1397–412.

Campino, S. *et al*. (2011). Population genetic analysis of *Plasmodium falciparum* parasites using a customized Illumina GoldenGate genotyping assay. *PloS one*, 6(6), p.e20251.

Chapuis, M.P. and Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, 24(3), pp.621–631.

Chen, H. *et al*. (2004). Population genetic structure of *Anopheles gambiae* mosquitoes on Lake Victoria islands, west Kenya. *Malaria journal*, 3, p.48.

Chenet, S.M. *et al*. (2008). Genetic diversity of vaccine candidate antigens in *Plasmodium falciparum* isolates from the Amazon basin of Peru. *Malaria journal*, 7, p.93.

Coleman, R.E. *et al*. (2006). Comparison of PCR and microscopy for the detection of asymptomatic malaria in a *Plasmodium falciparum/vivax* endemic area in Thailand. *Malaria Journal*, 5(1), pp.1–7.

Conway, D.J. and McBride, J.S. (1991). Population genetics of *Plasmodium falciparum* within a malaria hyperendemic area. *Parasitology*, 103(1), pp.7–16.

Cortés, A. *et al*. (2003). Geographical structure of diversity and differences between symptomatic and asymptomatic infections for *Plasmodium falciparum* vaccine candidate AMA1. *Infection and Immunity*, 71(3), pp.1416–1426.

Costantini, C. *et al*. (1996). Density, survival and dispersal of *Anopheles gambiae* complex mosquitoes in a west African Sudan savanna village. *Medical and veterinary entomology*, 10(3), pp.203–219.

Cowman, A.F., Berry, D. and Baum, J. (2012). The cellular and molecular basis for malaria parasite invasion of the human red blood cell. *The Journal of cell biology*, 198, pp.961–71.

Dondorp, A.M. *et al*, (2009). Artemisinin Resistance in *Plasmodium falciparum* Malaria. *Drug Therapy*, 361, pp.455–467.

Durand, P. *et al*. (2003). Significant linkage disequilibrium and high genetic diversity in a population of. *Tropical Medicine*, 68(3), pp.345–349.

Earl, D.A. and VonHoldt, B.M. (2011). Structure harvester: a website and program for

visualizing structure output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), pp.359–361.

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature reviews. Genetics*, 5(6), pp.435–445.

Evanno, G., Regnaut, S. and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, 14(8), pp.2611–2620.

Excoffier, L., Laval, G. and Schneider, S. (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, 1, pp.47–50.

Feagin, J.E. (1992). The 6-kb element of *Plasmodium falciparum* encodes mitochondrial cytochrome genes. *Molecular and Biochemical Parasitology*, 52(1), pp.145–148.

Feng, X. *et al*. (2003). Single-nucleotide polymorphisms and genome diversity in *Plasmodium vivax*. *Proceedings of the National Academy of Sciences of the United States of America*, 100(14), pp.8502–7.

Ferdig, M.T. and Su, X.Z. (2000). Microsatellite markers and genetic mapping in *Plasmodium falciparum*. *Parasitology today (Personal ed.)*, 16(7), pp.307–12.

Fraser-Hurt, N. *et al*. (1999). Effect of insecticide-treated bed nets on haemoglobin values, prevalence and multiplicity of infection with *Plasmodium falciparum* in a randomized controlled trial in Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 93(1), pp.47–51.

Freese, J.A. and Markus, M.B. (1990). Enzyme typing of southern African isolates of *Plasmodium falciparum* by polyacrylamide gel electrophoresis. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 84(4), pp.519–520.

Gaensslen, R.E., Bell, S.C. and Lee, H.C. (1987). Distributions of genetic markers in United States populations: II. Isoenzyme systems. *Journal of forensic sciences*, 32, pp.1348–1381.

Gardner, M.J. *et al*. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, 419(6906), pp.498–511.

Gilbert, S.C. *et al*. (1998). Association of Malaria Parasite Population Structure, HLA, and Immunological Antagonism. *Science*, 279(5354), pp.1173-1177.

Goudet, J. (1995). FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. *Journal of Heredity*, 86, pp.485–486.

Greenhouse, B. *et al*. (2006). Validation of microsatellite markers for use in genotyping polyclonal *Plasmodium falciparum* infections. *The American journal of tropical medicine and hygiene*, 75(5), pp.836–42.

Haubold, B. and Hudson, R.R. (2000). LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage Analysis. *Bioinformatics (Oxford, England)*, 16(9), pp.847–848.

Hoffmann, E.H.E., Ribolla, P.E.M. and Ferreira, M.U. (2003). Genetic relatedness of *Plasmodium falciparum* isolates and the origin of allelic diversity at the merozoite surface protein-1 (MSP-1) locus in Brazil and Vietnam. *Malaria journal*, 2, p.24.

Hughes, M.K. and Hughes, A.L. (1995). Natural selection on *Plasmodium* surface proteins. *Molecular and Biochemical Parasitology*, 71(1), pp.99–113.

Isozumi, R. *et al*. (2015). Improved detection of malaria cases in island settings of Vanuatu and Kenya by PCR that targets the *Plasmodium* mitochondrial cytochrome c oxidase III (cox3) gene. *Parasitology International*, 64(3), pp.304–308.

Iwagami, M. *et al*. (2009). Genetic diversity and population structure of *Plasmodium falciparum* in the Philippines. *Malaria journal*, 8, p.96.

Jakobsson, M. and Rosenberg, N.A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), pp.1801–1806.

Jeffares, D.C. *et al*. (2007). Genome variation and evolution of the malaria parasite *Plasmodium falciparum*. *Nature genetics*, 39, pp.120–125.

Jelinek, T. *et al*. (1996). Geographic differences in the sensitivity of a polymerase chain reaction for the detection of *Plasmodium falciparum* infection. *The American journal of tropical medicine and hygiene*, 55(6), pp.647–651.

Jiang, H. *et al*. (2011). High recombination rates and hotspots in a *Plasmodium falciparum* genetic cross. *Genome Biology*, 12, p.33.

Joshi, H. (2003). Markers for population genetic analysis of human Plasmodia species, *P.*

falciparum and *P. vivax*. *Journal of Vector Borne Diseases*, 40(3–4), pp.78–83.

Joshi, H. *et al*. (1989). *Plasmodium vivax* enzyme polymorphism in isolates of Indian origin. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 83(2), pp.179–181.

Joy, D.A. *et al*. (2003). Early Origin and Recent Expansion of Plasmodium falciparum. *Science*, 300(5617), pp.318–321.

Kebaier, C., Voza, T. and Vanderberg, J. (2009). Kinetics of mosquito-injected *Plasmodium* sporozoites in mice: fewer sporozoites are injected into sporozoite-immunized mice. *PLoS pathogens*, 5(4), p.e1000399.

Kenya Ministry of Health, (2009). Revised Kenya Malaria Strategy 2009-2018. *National Malaria control*.

Kenya National Bureau of Statistics, (2010). The 2009 Kenya Population and Housing Census. , IC, pp.1–206.

Larrañaga, N. *et al*. (2013). Genetic structure of *Plasmodium falciparum* populations across the Honduras-Nicaragua border, 12, pp.1–10.

Leclerc, M.C. *et al*. (2002). Genetic diversity and population structure of *Plasmodium falciparum* isolates from Dakar, Senegal, investigated from microsatellite and antigen determinant loci. *Microbes and Infection*, 4(7), pp.685–692.

LeMenach, A. *et al*. (2011). Travel risk, malaria importation and malaria transmission in Zanzibar. , pp.1–7.

Lew, V.L., Tiffert, T. and Ginsburg, H. (2003). Excess hemoglobin digestion and the osmotic stability of *Plasmodium falciparum* – infected red blood cells. , 101(10), pp.4189–4194.

Lewontin, R.C. and Hubby, J.L. (1966). A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of Drosophila pseudoobscura. *Genetics*, 54(2), pp.595–609.

Li, J. *et al*. (2009). Hundreds of microsatellites for genotyping *Plasmodium yoeli*i parasites. *Molecular and Biochemical Parasitology*, 166, pp.153–158.

Li, J. *et al*. (2007). Typing *Plasmodium yoelii* microsatellites using a simple and affordable fluorescent labeling method. *Mol Biochem Parasitol*, 155(2), pp.94–102.

Li, W.H., Gojobori, T. and Nei, M. (1981). Pseudogenes as a paradigm of neutral evolution. *Nature*, 292(5820), pp.237–239.

Liljander, A. *et al*. (2009). Optimization and validation of multi-coloured capillary electrophoresis for genotyping of *Plasmodium falciparum* merozoite surface proteins (msp1 and 2). *Malaria journal*, 8, p.78.

Litt, M. and Luty, J.A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics*, 44(3), pp.397–401.

Lum, J.K. *et al*. (2007). Genetic diversity and gene flow of humans, *Plasmodium falciparum*, and *Anopheles farauti s.s*. of Vanuatu: Inferred malaria dispersal and implications for malaria control. *Acta Tropica*, 103(2), pp.102–107.

Lum, J.K. *et al*. (2004). Malaria dispersal among islands: human mediated *Plasmodium falciparum* gene flow in Vanuatu, Melanesia. *Acta tropica*, 90(2), pp.181–5.

Lynch, C. and Roper, C (2011). The Transit Phase of Migration: Circulation of Malaria and Its Multidrug-Resistant Forms in Africa. *PLoS Medicine*, 8(5).

Machado, R.L.D. *et al*. (2004). Genetic structure of *Plasmodium falciparum* populations in the Brazilian Amazon region. *The Journal of infectious diseases*, 190(9), pp.1547–55.

malERA, (2011). A research agenda for malaria eradication: basic science and enabling technologies. *PLoS medicine*, 8, p.e1000399.

Md Idris, Z. *et al*. (2014). Geographic and seasonal variation in malaria prevalence on islands in Lake Victoria (western Kenya): results from three cross sectional studies. *Malaria Journal*, 13(1), p.1.

Michalakis, Y. and Excoffier, L. (1996). A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics*, 142(3), pp.1061–1064.

Mittal, (2009). Microsatellite markers- A new practice of DNA based markers in molecular genetics. *Pharmacognosy Reviews*, 3(6), p.235.

Mobegi, V. A *et al*. (2012). Population genetic structure of *Plasmodium falciparum* across a

region of diverse endemicity in West Africa. *Malaria journal*, 11, p.223.

MoH, (2016). Malaria Indicator Survey. *National Malaria control*, p.2.

Moonen, B. *et al*. (2010). A framework for assessing the feasibility of malaria elimination. *Malaria journal*, 9(1), p.322.

Morin, P., Luikart, G. and Wayne, R. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, 19, pp.208–216.

Mu, J. *et al*. (2002). Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum. Nature*, 418(6895), pp.323–326.

Mu, J. *et al*. (2007). Genome-wide variation and identification of vaccine targets in the *Plasmodium falciparum* genome. *Nature genetics*, 39(1), pp.126–30.

Myint-Oo, (1986). Isoenzyme variation in schizonts of *Plasmodium vivax* from Burma. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 80(1), pp.1–4.

Najera, J.A., Gonzalez-Silva, M. and Alonso, P.L. (2011). Some lessons for the future from the global malaria eradication programme (1955-1969). *PLoS Medicine*, 8.

Okara, R.M. *et al*. (2010). Distribution of the main malaria vectors in Kenya. *Malaria journal*, 9, p.69.

Oyebola, lukosi, Y. a *et al*. (2014). Genetic diversity and complexity of *Plasmodium falciparum* infections in Lagos, Nigeria. *Asian Pacific journal of tropical biomedicine*, 4(Suppl 1), pp.S87-91.

Oyebola, M.K. *et al*. (2014). Microsatellite markers reveal low levels of population sub-structuring of *Plasmodium falciparum* in southwestern Nigeria. , pp.1–8.

Pasvol, G. (2010). Protective hemoglobinopathies and *Plasmodium falciparum* transmission. *Nature genetics*, 42, pp.284–285.

Pritchard, J.K. (2010). Documentation for structure software : Version 2 . 3. *In Practice*, 6(3), pp.321–326.

Pritchard, J.K., Stephens, M. and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), pp.945–959.

Pumpaibool, T. *et al*. (2009). Genetic diversity and population structure of *Plasmodium*

falciparum in Thailand, a low transmission country. *Malaria journal*, 8, p.155.

Razakandrainibe, F.G. *et al*. (2005). "Clonal" population structure of the malaria agent *Plasmodium falciparum* in high-infection regions. *Proceedings of the National Academy of Sciences*, 102(48), pp.17388–17393.

Rebaudet, S. *et al*. (2010). Genetic structure of *Plasmodium falciparum* and elimination of malaria, Comoros archipelago. *Emerging infectious diseases*, 16(11), pp.1686–94.

Reid, H. *et al*. (2010). Baseline spatial distribution of malaria prior to an elimination programme in Vanuatu. *Malaria journal*, 9, p.150.

Roper, C. *et al*. (2004). Intercontinental spread of pyrimethamine-resistant malaria. *Science (New York, N.Y.)*, 305(5687), p.1124.

Rosenberg, N.A. (2004). DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, 4(1), pp.137–138.

Rosenberg, R. *et al*. (1990). An estimation of the number of malaria sporozoites ejected by a feeding mosquito. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 84(2), pp.209–212.

Russell, B., Suwanarusk, R. and Lek-Uthai, U. (2006). *Plasmodium vivax* genetic diversity: microsatellite length matters. *Trends in Parasitology*, 22(9), pp.399–401.

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, 109, pp.365–371.

Schuelke, (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature biotechnology*, 18, pp.233–234.

Schultz, L. *et al*. (2010). Multilocus haplotypes reveal variable levels of diversity and population structure of *Plasmodium falciparum* in Papua New Guinea, a region of intense perennial transmission. *Malaria journal*, 9, p.336.

Scopel, K.K.G. *et al*. (2004). Low sensitivity of nested PCR using *Plasmodium* DNA extracted from stained thick blood smears: an epidemiological retrospective study among subjects with low parasitaemia in an endemic area of the Brazilian Amazon region. *Malaria Journal*, 3(1), pp.1–6.

Selkoe, K. A and Toonen, R.J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology letters*, 9(5), pp.615–29.

Shah, N.K. *et al*. (2011). Antimalarial drug resistance of *Plasmodium falciparum* in India: changes over time and space. *The Lancet infectious diseases*, 11(1), pp.57–64.

Sinka, M.E. *et al*. (2012). A global map of dominant malaria vectors. *Parasites & Vectors*, 5, p.69.

Snow, R.W. and Marsh, K. (2010). Malaria in Africa: Progress and prospects in the decade since the Abuja Declaration. *The Lancet*, 376(9735), pp.137–139.

Steffens, D.L., Sutter, S.L. and Roemer, S.C. (1993). An alternate universal forward primer for improved automated DNA sequencing of M13. *BioTechniques*, 15(4), p.580,582.

Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology and Evolution*, 15, pp.199–203.

Sutton, P.L., Torres, L.P. and Branch, O.H. (2011). Sexual recombination is a signature of a persisting malaria epidemic in Peru. *Malaria Journal*, 10, p.329.

Talisuna, A.O. *et al*. (2007). Intensity of malaria transmission and the spread of *Plasmodium falciparum*-resistant malaria: A review of epidemiologic field evidence. *American Journal of Tropical Medicine and Hygiene*, 77(6), pp.170–180.

Thaithong, S., Sueblinwong, T. and Beale, G.H. (1981). Enzyme typing of some isolates of *Plasmodium falciparum* from thailand. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 75(2), pp.268–270.

Theodorsson-Norheim, E. (1986). Kruskal-Wallis test: BASIC computer program to perform nonparametric one-way analysis of variance and multiple comparisons on ranks of several independent samples. *Computer Methods and Programs in Biomedicine*, 23(1), pp.57–62.

Tuteja, R. (2007). Malaria - an overview. *The FEBS journal*, 274, pp.4670–4679.

Volkman, S.K. *et al*. (2007). A genome-wide map of diversity in *Plasmodium falciparum*. *Nature genetics*, 39(1), pp.113–119.

Vos, P. *et al*. (1995). AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research*, 23(21), pp.4407–4414.

Wenz, H.M. *et al*. (1998). High-precision genotyping by denaturing capillary electrophoresis. *Genome Research*, 8(1), pp.69–80.

White, N.J. (1998). Drug resistance in malaria. *British medical bulletin*, 54(3), pp.703–715.

WHO, (2008). World Health Statistics 2008. *World Health Organisation*, p.112.

WHO, (2009). World Malaria Report 2009. *World Health Organisation*, 12, p.66.

WHO, (2015). World Malaria Report 2015. *World Health Organisation*, pp.1–280.

Wilson, (Iain) R.J.M. *et al*. (1996). Complete Gene Map of the Plastid-like DNA of the Malaria Parasite *Plasmodium falciparum*. *Journal of Molecular Biology*, 261(2), pp.155–172.

Wootton, J.C. *et al*. (2002). Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature*, 418, pp.320–323.

Yekutiel, P. (1960). Problems of epidemiology in malaria eradication. *Bulletin of the World Health Organization*, 22, pp.669–683.

Zhong, D. *et al*. (2007). *Plasmodium falciparum* genetic diversity in western Kenya highlands. *The American journal of tropical medicine and hygiene*, 77, pp.1043–1050.

APPENDICES

Appendix I Primer details for eight microsatellite markers

Loci	Primer name	Chromosomal number	Accession Number	Primer sequence (Red is M13 tagged Forward primer)	expected amplicon size
TA81	TA81-F	5	AF010510	GAAGAAATAAGGGAAGGT	141 bp
	TA81-R			TTTCACACAACACAGGATT	
	TA81-FM13			CACGACGTTGTAAAACGACTGGACAAATGGGAAAGGATA	
ARA2	ARA2-F	11	X17484	GTACATATGAATCACCAA	92 bp
	ARA2-R			GCTTTGAGTATTATTAATA	
	ARA2- FM13			CACGACGTTGTAAAACGACGAATAAACAAAGTATTGCT	
TA87	TA87- F	6	AF010571	ATGGGTTAAATGAGGTACA	119 bp
	TA87-R			ACATGTTCATATTACTCAC	
	TA87- FM13			CACGACGTTGTAAAACGACAATGGCAACACCATTCAAC	
TA40	TA40-F	10	AF010542	AAGGGATTGCTGCAAGGT	236 bp
	TA40-R			GAAATTGGCACCACCACA	
	TA40-FM13			CACGACGTTGTAAAACGACAAGGGATTGCTGCAAGGT	
	TA40-RI			CATCAATAAAATCACTACTA	
TA42	TA42-F	5	AF010543	ACAAAAGGGTGGTGATTCT	220 bp
	TA42-R			GTATTATTACTACTACTAAAG	
	TA42-FM13			CACGACGTTGTAAAACGACTAGAAACAGGAATGATACG	
TA1	TA1-F	6	AF010507	CTACATGCCTAATGAGCA	205 bp
	TA1-R			TTTTATCTTCATCCCCAC	
	TA1-FM13			CACGACGTTGTAAAACGACCCGTCATAAGTGCAGAGC	
TA109	TA109-F	6	AF010508	TAGGGAACATCATAAGGAT	193 bp
	TA109-R			CCTATACCAAACATGCTAAA	
	TA109-FM13			CACGACGTTGTAAAACGACGGTTAAATCAGGACAACAT	
PFG377	PFG377-F	12	L04161	GATCTCAACGGAAATTAT	117 bp
	PFG377-OR			TTATCCCTACGATTAACA	
	PFG377-FM13			CACGACGTTGTAAAACGACGATCTCAACGGAAATTAT	
	PFG377-R			TTATGTTGGTACCGTGTA	
	M13FHEX	Universal primer		CACGACGTTGTAAAACGAC	

Appendix II Spectrophotometer reading of DNA extracted from parasite culture lines

P. falciparum lines	Parasite density	DNA concentration (ng/µl)	Absorption 260/280
MAD20	5000	620.7	1.85
	1000	218.3	1.89
	500	171.8	1.82
	100	101.2	1.78
	50	67.9	1.83
	10	30.2	1.80
	5	18.6	1.91
	1	7.1	1.87
K1	5000	719.7	1.97
	1000	207.0	1.88
	500	142.9	1.83
	100	88.5	1.94
	50	49.3	1.87
	10	20.6	1.99
	5	10.4	1.82
	1	9.3	1.75
FCR3	5000	680.1	2.01
	1000	228.5	1.81
	500	182.7	1.77
	100	77.6	1.84
	50	55.0	1.83
	10	37.9	1.92
	5	17.2	1.89
	1	11.5	1.81
THAI 838	5000	590.8	1.86
	1000	202.4	2.05
	500	133.6	1.99
	100	94.1	1.82
	50	51.8	2.03
	10	30.2	1.84
	5	22.7	1.86
	1	8.2	1.94

Appendix III Allelic frequencies at 8 microsatellite loci

locus	Allele	Kibuogi	Mfangano	Ngodhe	Takawiri	Ungoye
ARA2	102	0.000	0.000	0.000	0.000	0.020
	72	0.029	0.000	0.059	0.028	0.040
	75	0.686	0.300	0.059	0.528	0.020
	78	0.057	0.040	0.059	0.056	0.100
	81	0.000	0.020	0.000	0.000	0.100
	84	0.086	0.280	0.176	0.167	0.300
	87	0.086	0.220	0.529	0.139	0.180
	90	0.000	0.040	0.059	0.000	0.120
	93	0.000	0.100	0.000	0.028	0.060
	96	0.000	0.000	0.059	0.056	0.040
	99	0.057	0.000	0.000	0.000	0.020
PFG377	102	0.029	0.000	0.000	0.000	0.000
	105	0.057	0.160	0.059	0.056	0.440
	111	0.000	0.160	0.000	0.083	0.000
	114	0.143	0.280	0.118	0.167	0.060
	117	0.029	0.160	0.059	0.083	0.060
	120	0.000	0.020	0.000	0.000	0.000
	129	0.000	0.000	0.176	0.028	0.020
	132	0.000	0.000	0.000	0.028	0.000
	141	0.000	0.000	0.000	0.056	0.000
	144	0.000	0.020	0.118	0.000	0.000
	147	0.000	0.020	0.118	0.000	0.000
	81	0.257	0.020	0.059	0.111	0.000
	84	0.029	0.020	0.059	0.056	0.140
	87	0.200	0.000	0.059	0.028	0.080
	90	0.143	0.020	0.118	0.139	0.080
	93	0.029	0.120	0.059	0.083	0.080
	96	0.086	0.000	0.000	0.083	0.040
TA1	140	0.029	0.000	0.000	0.000	0.000
	146	0.000	0.000	0.000	0.000	0.020
	149	0.000	0.000	0.000	0.000	0.020
	151	0.029	0.000	0.000	0.000	0.000
	152	0.000	0.020	0.000	0.000	0.000
	164	0.000	0.000	0.412	0.056	0.000
	173	0.000	0.040	0.000	0.000	0.000
	176	0.000	0.000	0.000	0.000	0.040
	179	0.314	0.200	0.118	0.222	0.220
	182	0.029	0.140	0.118	0.000	0.060
	185	0.057	0.140	0.000	0.167	0.200
	188	0.057	0.100	0.118	0.111	0.180
	191	0.171	0.140	0.059	0.028	0.040
	194	0.029	0.080	0.000	0.139	0.040
	197	0.114	0.080	0.000	0.167	0.060
	200	0.086	0.020	0.000	0.028	0.040
	203	0.000	0.040	0.000	0.028	0.020
	206	0.000	0.000	0.000	0.000	0.020
	209	0.000	0.000	0.059	0.028	0.000
	212	0.000	0.000	0.118	0.028	0.000
	215	0.000	0.000	0.000	0.000	0.020
	218	0.000	0.000	0.000	0.000	0.020

	227	0.086	0.000	0.000	0.000	0.000
TA40	200	0.000	0.020	0.000	0.000	0.060
	203	0.114	0.060	0.118	0.028	0.020
	206	0.086	0.020	0.059	0.028	0.060
	209	0.057	0.000	0.000	0.028	0.100
	212	0.000	0.060	0.176	0.056	0.100
	215	0.000	0.060	0.000	0.028	0.040
	218	0.057	0.020	0.059	0.389	0.060
	221	0.000	0.000	0.000	0.000	0.020
	224	0.086	0.040	0.000	0.000	0.100
	227	0.143	0.020	0.000	0.056	0.100
	230	0.143	0.160	0.294	0.139	0.180
	233	0.086	0.220	0.235	0.056	0.100
	236	0.029	0.180	0.059	0.056	0.000
	239	0.086	0.100	0.000	0.111	0.020
	242	0.029	0.020	0.000	0.000	0.000
	245	0.000	0.020	0.000	0.000	0.000
	248	0.086	0.000	0.000	0.028	0.040
TA42	162	0.000	0.020	0.000	0.000	0.000
	165	0.000	0.000	0.059	0.028	0.000
	180	0.029	0.000	0.059	0.000	0.000
	183	0.000	0.000	0.118	0.000	0.000
	198	0.000	0.040	0.000	0.000	0.060
	201	0.143	0.100	0.118	0.083	0.040
	204	0.486	0.560	0.235	0.472	0.680
	207	0.171	0.120	0.176	0.306	0.060
	210	0.057	0.060	0.000	0.083	0.040
	213	0.000	0.000	0.118	0.000	0.000
	219	0.086	0.040	0.118	0.028	0.100
	222	0.000	0.020	0.000	0.000	0.020
	225	0.000	0.020	0.000	0.000	0.000
	228	0.029	0.020	0.000	0.000	0.000
TA81	102	0.057	0.040	0.000	0.056	0.000
	105	0.029	0.000	0.000	0.000	0.000
	108	0.057	0.000	0.000	0.028	0.000
	111	0.086	0.000	0.000	0.000	0.000
	114	0.000	0.000	0.000	0.028	0.000
	126	0.000	0.020	0.118	0.000	0.020
	129	0.000	0.000	0.000	0.000	0.020
	132	0.057	0.000	0.176	0.028	0.080
	135	0.200	0.200	0.118	0.056	0.360
	138	0.114	0.140	0.176	0.083	0.120
	141	0.143	0.200	0.000	0.139	0.080
	144	0.114	0.220	0.059	0.194	0.180
	147	0.000	0.120	0.176	0.167	0.060
	150	0.029	0.020	0.000	0.083	0.000
	153	0.029	0.020	0.000	0.000	0.000
	159	0.000	0.000	0.000	0.000	0.020
	162	0.000	0.000	0.000	0.000	0.040
	165	0.086	0.020	0.118	0.083	0.020
	177	0.000	0.000	0.059	0.000	0.000
	99	0.000	0.000	0.000	0.056	0.000
	102	0.000	0.000	0.000	0.028	0.000

TA87	105	0.000	0.000	0.000	0.028	0.000
	108	0.029	0.000	0.059	0.000	0.020
	111	0.086	0.040	0.059	0.056	0.040
	114	0.400	0.280	0.294	0.250	0.420
	117	0.229	0.140	0.059	0.167	0.220
	120	0.057	0.140	0.118	0.194	0.080
	123	0.057	0.280	0.235	0.083	0.120
	126	0.057	0.040	0.118	0.083	0.080
	129	0.086	0.040	0.000	0.083	0.020
	132	0.000	0.020	0.000	0.028	0.000
	135	0.000	0.020	0.059	0.000	0.000
TA109	176	0.000	0.020	0.000	0.000	0.080
	179	0.086	0.240	0.118	0.056	0.260
	182	0.029	0.160	0.118	0.194	0.200
	185	0.029	0.060	0.059	0.028	0.100
	188	0.029	0.000	0.000	0.000	0.000
	191	0.086	0.060	0.294	0.028	0.060
	194	0.029	0.120	0.059	0.056	0.160
	197	0.000	0.020	0.000	0.028	0.000
	200	0.000	0.020	0.000	0.000	0.000
	203	0.000	0.020	0.059	0.056	0.040
	206	0.029	0.000	0.000	0.000	0.000
	209	0.000	0.000	0.000	0.028	0.000
	212	0.057	0.040	0.000	0.028	0.000
	215	0.600	0.240	0.294	0.444	0.100
	218	0.000	0.000	0.000	0.056	0.000
	227	0.029	0.000	0.000	0.000	0.000

Appendix IV Alleles scored in each of the 188 *P. falciparum* isolates genotyped at 8 loci

	ARA2				PFG						TA1				TA40						TA42					TA81				TA87					TA109			
Sample ID	Allele 1	Allele 2	Allele 3	Allele 4	Allele 1	Allele 2	Allele 3	Allele 4	Allele 5	Allele 6	Allele 1	Allele 2	Allele 3	Allele 4	Allele 1	Allele 2	Allele 3	Allele 4	Allele 5	Allele 6	Allele 1	Allele 2	Allele 3	Allele 4	Allele 5	Allele 1	Allele 2	Allele 3	Allele 4	Allele 1	Allele 2	Allele 3	Allele 4	Allele 5	Allele 1	Allele 2	Allele 3	Allele 4
K1	75				90	81	87	96			197				209	230					204					108				123					185	218		
K2	72	75			87						182				248						207					105				117					215			
K5	78	84			105	96	81	147			197	185			248						204					138				120	117				191			
K6	75				90	117					200				218						207					153				117					215			
K9	84	102	108		93						188				230						204					135				120	114	117	123		179	176	191	
K10	78	84			114	105					179	197			224						201					144				117	120				179	176	191	203
K13	75				81	84					197				248	227	236				207					150				117	132	120			215			
K15	99				96	114					191				233						204					138	147			123	114	126			179	182		
K17	75				81						179				203	224	239				207					144				114	111				212	182	215	
K18	87				87	114	108	126	144		152				233	230					204					144				108	120	123			179			
K19	75	72			81						191				206	227					219	201				138				129					215			
K20	75				81						191	143			227	233	239				201	219				138	165			114					215			
K28	75				87	96	105				179	224			236	227					204	201				141	165			111					215			
K36	75				81						227				224	203	239				201	207				141				117	111				215	197	218	224
K39	75	87			87	90	96				227				242						201					135				114	108				215			
K40	75				114						179				206						219					111				129	114	126			191			
K41	75				87						179				206						219					165	168			114	129				215			
K42	75				81						191				230						204					165	141	144		114					215			
K44	87				102	96					179				230						204					108				114					215			
K46	75				81						200				239	203					204	207				111				114					215			
K49	75	72			90	102					191				239						204					141				114					227			
K52	75				84	93	114				140				227						204					102				114					215			
K60	99	96			105	114					227				209						180	183	204			135				114					206	209		
K94	75	72			87						185				203	215					204					132	141			111	117				212			
K105	84				114						179				227						210					135				126	120				191			
K106	87	84			114						188				233						228	180				135				117					194			
K116	75				96	90					179				230	227					207					165				126	120	114			215	203		
K130	75				81	90	93				179				218						204	219				135				114	120	123			215			
MC1	93				114	117					179				230	227					204					144				120	117	114			191			
MC4	84	78	87	90	114	117					194	188	179	212	200						162					141	135	138	144	123	114	120	129		191	179		
MC11	87				93						179				230	227					210					138				123	117				194			
MC15	84	93			117	114					179				245						204	219				147	135			132					182	179		
MC16	87	93			120	117	114				182	179			233						207	204				138	135	144		123	117	120			191	194		
MC21	93				114	87					182	179	188	191	233						210					144	141	135		123	120				179	194		
MC34	90				105	84	96	147			188				239	224					204					165	177			123					182			

MC54	84				114						188	182	191		236	233					204					138				120					179			
MC56	84	87			105	147	96	81	90		188	182			242	236					204					144	135			123	120				185	194		
MC84	87				114						191				233						228					144				114					194			
MC93	87				111						197				230	227					219					135				114					182			
MC108	75	72			81						194				236	203	209	224	233	239	204	198				141	165			123	111				215			
MC123	75				111	117					191				230						207	204				141				117					182	185		
MC124	75				111						191				233	227					204	201				141				117					215			
MC125	75				111	81	93	117			191				212	203	224	239			204	207				102				114					212	179	182	218
MC127	75				93						191				233						204	207				138				129	114	120			215			
MC133	75				117						152				233						198					135	129			114	123	108			215			
MC134	75				114	84	117	144			179	188	224		215						201	204				141				114					215			
MC141	78	84			114						182	188			230						204					141	129	144		114					194			
MC145	84				93	84	90	120			185				233	230					225					147				135					182			
MC212	87	102			117	114	120				185				233						222					144				120					194	179		
MI23	84	90			84	96	105	117			185				233	230					204					147	144			129	114				182	206		
MI37	87	84			114						185				236	212					204					135				120					179			
MI38	87	90	93		117	105	114				203	179	188		215	233					219	204				144	135			117	123	108			185	179	172	194
MI41	78				105	81	96	147			179				230						204					135				111					197	182	194	
MI52	90				114						173				224	239					204	219				141	150	144	138	114	123				179			
MI54	87	78			117	114	120				200				227						204					144				117					179			
MI57	84				114						188	209			236	212					204	219				138	126			120					191	179		
MI74	84	96			114	117					182	179			215						204					138	141	144		114	117				182			
MI117	81	78			114						203	191			236						201					144				120	129				176			
MI198	75				93						197				239	209	227				204	207				150	138			114	120	126			215			
MI199	75				90	108					197				218	227					207	204				144				114	123	123			215			
MI200	93	84			147						194				212	239					198					135	138			123					179			
MI201	93	84			114						197				230						207	204				144	141			117	114	126			179	203		
MI202	75				117						179	185			203						204					147				117	126	129	114		215			
MI203	84				117						182				212	233	236				207					135				123					179			
N25	87				147	117	117	126	129		164				230						204					138				120					191			
N42	87				129	84	114	117	126		164				233	209					204					165				120					191	215		
N43	87	84			129						164				230						207					147				114					182	215		
N54	90				129	114	117	126			164				230	209					165					138				135					215	218		
N55	96				90	96	108				179				230						219	204				132				126	123				203			
N62	87	75			105	87	90	93			191	179			203	209	215				207	222				132				123					215	203		
N73	87				90						182				203	215	227	230			213	198				126				123					215			
N84	72	75	69		93						179				233	215	227				201					144				117	120				215			
N90	87				87						164				212	230					182					135				126	135				191	179		
N92	87	78			81	96	105	147			212				233						201	204				177				123	114	117			179	182		
N95	78				114	111					209				230						204					135				114					191			
N157	75				144	90	114				188	182			236						213					132	144	147		111	120				185	176		

N177	87				114	84	117	126	129		164				212						219					147				108					182			
N 213	84				147	87					212				212						180	183	186	204	228	147				114					179			
T9	75				81	86					179				218	212	230				207					144				114	111				212	180		
T26	75				93	84	87				194				218						207					165				129	111				215			
T30	72	75			81						194				230	218	224				207					117	141	147		126					215			
T41	87				132	114					185				212	233	236				165					144				120					179			
T64	75				111	96	114				185				218						204					144	111			114	126				215	218		
T68	96	75			111	87					185	182			218						204					102	105	114		123					218			
T72	96	75			111						209	206			239	203	224				204					141				123					218			
T73	75				93						197				218	227					204					150				120	117	132			215			
T77	75	72			93	90					194				218						204					147				114					215			
T81	75				90	93	99				197				218						204					150				114	111				215			
T101	78				114	84	87	96	105	114	188				215						219					138				120					191	194		
T114	87				105	114					188	182			227						204					141				126					194			
T115	84	87			117	120					179				233	230					204	207				147	144			129					182	180		
T119	75				87	84					179				218						210					141				129	120				179	180		
T122	84				129						188				236	233	245				207	195	192	189		147				120					185			
T126	75				90						185				230						207					138				117	126				179	180		
T128	84	93			105	84					191	194			233						204					138	141	144		117	123				182	203		
T137	75				96	87					188				218						207					114				117	126				215			
T142	87	75			96	87					212	185			218						204					144				111					215			
T166	75				90	84					194				227						201	204				108	101			120					212	180		
T169	75				90						185				239	227	248				207					102				117					182			
T172	75				141	90	126				200				206						207					99				117					215	183		
T176	75				96	141					197				203	224					207					165				102	114				197			
T185	75				141	80	126				197				236	218					204					150				114					215			
T197	87				114	87					164				212						204					132				105					203	191	188	200
T204	75				81	102					194				218	224	230				210					147				114					215			
T208	87				117						179	197			248						204					147				123					194			
T241	84				84	96	105	147			203				209						204					135				126					203	170		
T247	75				117	90					197				230						204					141	102	108	165	123					215			
U4	87				105	96	114	147			188				215						210					135	138	141	144	117	114	120	123		215			
U10	84	90			84	105					185				224	203	227				210					135	141	144		126	114				194			
U15	78	84			105	81	96	147			185	179			224						204					135				114	120				185	176	179	194
U16	84				105	96	114	147			179	191			200	212					207					144	135			120	123	126			179	191	194	
U17	90	78			105	96	114	147			188				224						204	201	207			144	141	135		114	117	123			176			
U19	102				105	96	147				185				233						204					138				108					185	182		
U21	84	81	96	99	105	96	147				179				212	227					204					138	141			114	120	126			176	185	194	206
U23	90	93			105	81	96	147			176				233						204	219				135	144			117					176	188	194	
U27	84				105	84	96	147			188				239	224					204					165	138	108		114	117	126			179	188		
U31	84				114	114					191				206	209					204					135				123					194			

U33	96	90			84	96	105				188				248						219	201				141				114					194			
U34	84	87	93		117	114					185	179	182		200	242					201					135	138	147	153	117	114	105	120	126	191	203		
U41	78	84			105	96	147				188	185			230	242					219	204				135	138			117					179			
U46	84				117						179	176	185		200						204					135	153			120					185	191	194	203
U49	96	78	84		105	96	147				182	197			209	224					204					144	147			111	117	123			179	191	194	
U50	87	84			105	96	147				149	212			248	248					204					159	138			117					182			
U53	84				105	96	147				179				233	206	215				204					147	144			114	108	117			179	194		
U55	93				105	81	96	147			185	179	197		230						204	219				144				120					185			
U57	93				105	81	96	147			179				233						219					141				123	117				182			
U61	90	84			114	96	105	147			182	191			230	206					219	204				138	135	144	147	126	114	117	123		191	182		
U64	87	84			87	84	105	114			185				236	224	230				204					144				114	126				179	182	194	
U65	84	84			87						176	179			227						204					144	147			120					179			
U70	78				96	81	105	147			188				212						204					135	138			123	120				194			
U74	87				105	84	96	147			179				230						204					138				114					179			
U79	81				84						185				206						204					132	141			114					194			
U86	81				105	84	96	147			215	194			209	200					204					135				114					194	179		
U94	81				105	96	147				197				215	206					204					162				114	117	120	123	126	194			
U95	81				105	81	96	147			179				218	212	233				204					135				123	114				194			
U101	81				105	81	96	147			179				212						204					144	147			123	117	129			182	187		
U106	87	90			117						200				230						204					132				114					182			
U107	78				105	84	96	147			194				233						204					138				129					179			
U117	72	75			90	93					179				203	215	227				198					141				117	120				215			
U118	75	72			90						179				218	230	239				204					135				117	120				215	203		
U123	90	87			93						203				218	230	239	245			204					135				117	120				179	182	218	
U125	87				93						179				227						204					132				117	120				203	185	212	
U126	87				93	117					188				230						204					135				117	120				182			
U129	72	75			90						188				230						204					129				114	117				188	179	182	212
U130	90	84	87		84						146				227						204					144				111	114				182			
U168	84	96			96	84	105	147			191				227	215					204					138	129	144		126	114	117	120	123	179	203		
U171	84	81			93						182				227						204					141	144	138		114	120	129			191	170	179	182
U218	78				105	81	96				197				230						204					135				126	114				182			
U237	84	87			84						194				212						204					135	144			114					215			
U244	84				129	114	126				206				209						204					162				114					179			
U246	84	87			114	81	96	105			218				209						198	204				135				114					185	203		

Highlighted light orange are predominant alleles while additional allele(s) are not highlighted.