POLYMORPHISMS IN THE *PLASMODIUM FALCIPARUM ERYTHROCYTE MEMBRANE PROTEIN1 (Pfemp1)* GENE AT TWO MALARIA ENDEMIC SITES IN KENYA

FRANCIS W. MAKOKHA

DOCTOR OF PHILOSOPHY
(Molecular Medicine)

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

2018

Polymorphisms in the *Plasmodium falciparum Erythrocyte Membrane Protein1 (Pfemp1)* Gene at Two Malaria Endemic Sites in Kenya

Francis W. Makokha

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy in Molecular Medicine in the Jomo Kenyatta University of Agriculture and Technology

2018

DECLARATION

This thesis is my original work and has not been presented to any other university by any other person.

Singed…….. Date……..

Francis W. Makokha

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

Signed …….. Date …….……..

Dr. Sabah Omar, PhD

KEMRI, Kenya

Signed…….Date……

Prof. Gabriel Magoma, PhD.

JKUAT, Kenya

DEDICATION

To my son and friend Philip Wekesa, my greatest inspiration and the best gift God ever brought into my life.

ACKNOWLEDGEMENT

I would like to sincerely thank Dr. Sabah Omar who admitted me into malaria laboratory and guided me into this project. I also wish to thank Prof. Gabriel Magoma of JKUAT, for his insightful and professional guidance as my supervisor and also the many words of encouragement.

To the members of staff and students in the Malaria laboratory, I say a big thank you for all the assistance you offered me as I worked on the bench. Your many kind words of encouragement really kept me going and I will forever remain grateful. In a special way, I would like to thank Mr. Francis Kimani, for his assistance in PCR trouble-shooting and many insightful suggestions and discussions concerning my project. I also recognize the efforts of Charity Hungu, for her assistance during bench work. To Rahma Udu, Stephen Kaniaru and Edwin Too for helping with sample collection and the many words of encouragement. I also want to appreciate the kindness of the Director and staff members of the Centre for Biotechnology and Research Development (CBRD), who created a conducive environment for my research work.

I would like to acknowledge Dr. Peter Bull for his guidance during the initial design of the project and for allowing me a short stay at the KEMRI-Welcome Trust, Kilifi laboratory to learn bench techniques and bioinformatics. His many suggestions on almost every step of the project, personal help with analysis of 454-next generation sequence data and suggestions on how to handle the data was of great help to me. To Dr. Etienne de Villiers, I say thank you for assisting me understand the 454 –sequencing data. Indeed without allowing me access to your software, and personally helping out in the analysis, I wouldn't have managed to handle the huge data. I further thank him for allowing me to attend the bioinformatics training on tools for analyzing

next generation sequence data at Ebiokit. I thank Dr. George Githinji of KEMRI Kilifi and James Mbora (formerly of ILIRI), fortheir assistance with bioinformatics analysis of my data. I also thank Wellcome Trust for giving me opportunity to attend training on working with pathogen genomes.

Finally, thanks to my entire family and all those who in one way or another assisted, encouraged, prayed with me or cared to inquire how far I was with my PhD work. Their support and prompting became my pillar of strength in this journey.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The variant *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) found on infected erythrocyte surface in malaria patients is encoded by *Pfemp*1 gene. This protein has been implicated in mediating different forms of malaria pathology such as rosseting of infected erythrocytes and cytoadherence of infected erythrocytes on endothelial cells of blood capillaries. This results in severe forms of the disease such as cerebral malaria and severe anemia. This PfEMP1 protein also mediates antigenic variation by *P. falciparum* thus rendering immune responses ineffective. PfEMP1 is coded for by highly variable *var* genes, with each parasite haploid genome containing over sixty copies of the gene. Only one gene is expressed at a given time and the expression pattern is mutually exclusive. The broad aim of this project was to profile the sequence tags of the DBLα domain of *Pfemp*1 genes in field isolates from the two malaria endemic sites. Blood samples from malaria positive patients were collected on Whatmann filter paper during a clinical field study at Mbita (Western Kenya) and Tiwi (Coastal Region, Kwale). Parasite DNA was extracted from the samples followed by PCR analysis using primers that target the DBLα domain of *Pfemp*1 genes. Some of the PCR product was sequenced by 454-next generation sequencing, Roche. The sequence reads were then translated into protein sequences of DBLα sequence tags and classified into various groups based on the number of cysteine residues in the sequence and positions of limited variations (PoLVs). This analysis revealed that group1/group1-like DBLα sequence tags were more prevalent in isolates from Tiwi than those from Mbita. Group 1 sequences are associated with expression of group A *var* genes that have been associated to severe malaria symptoms. Their presence in these isolates indicated that the patients from both sites were prone to developing severe symptoms like cerebral malaria and severe malarial aneamia. It was also found that group 4 sequence tags were the most frequent tags in field isolates from both study sites. These sequence tags have been associated to the expression of group B and C *var* genes. Expression of these genes is associated to mild symptoms of malaria. The sequence data can predicatively indicate the level of disease severity the circulating parasite population can course. Since the binding capacity of particular sequence types does not depend on expression level, these results suggested that the patients could have progressed to develop severe forms of malaria had they not be taken for early intervention. The high frequency of group-4 DBLα sequence tags indicated that a majority of the patients had started developing semi-immunity to malaria since they predict mild forms of malaria symptoms. A further analysis of sequence tags revealed sequences which were absent in the database after a Blast search at NCBI. This study therefore reports sequences unique to field isolates from the study sites. Further, it was observed that one sequence tag from Tiwi isolates possessed both MFK and REY motifs at PoLV1 and PoLV2 respectively. These two motifs have been found to be mutually exclusive hence this observation suggests that there is a possibility for the two motifs to co-exist in the same sequence tag although the chances of co-existence remain rare. A network was constructed to assess the genetic relationship between sequence tags based on position specific polymorphic blocks (PSPBs) in DBLα sequence tags. Sequences from Mbita study site and those from Tiwi largely clustered into separate giant networks with only a limited number of sequences from the two sites linking to each other. This observation suggested that parasite populations from the two endemic sites could be genetically distinct and that PfEMP1 sequencing could be a useful tool of understanding the genetics of parasite populations. This observation could also inform future efforts in the development of malaria vaccine. Thus the network approach of studying relationships between DBLα sequences is a useful tool of uncovering the genetic structure of parasite populations circulating in different malaria endemic region

CHAPTER ONE
INTRODUCTION

1.1. Background

Malaria remains a major public health challenge globally. The World Health Organization (WHO) estimates that there were 212 million malaria cases in 2015 (WHO, 2016). 90% of these cases occurred in Africa, 7% of the cases were reported in South East Asia while 2% was reported in Eastern Mediterranean Region (WHO; 2016). Over 3.4 billion people are still at risk of getting malaria (WHO, 2016). The report (WHO, 2016) also records an estimate of 429000 deaths attributable to malaria. 92% of the deaths were reported in Africa, 6% in South East Asia and 2% in Eastern Mediterranean region.

Those at highest risk are children younger than 5 years and pregnant women, particularly primigravida. Of the more than one million deaths that occur in Africa due to malaria, most are children under 5 years old. Apart from acute disease episodes and deaths in Africa, malaria also contributes significantly to anemia in children and pregnant women, adverse birth outcomes such as spontaneous abortion, stillbirth, premature delivery and low birth weight, and overall child mortality (Barnes and White, 2005).

In malaria endemic regions such as sub-Saharan Africa, children develop significant immunity against *P. falciparum* (Bull *et al*, 2008; Marsh, 1992). This naturally acquired immunity is poorly understood at a molecular level. Furthermore, it has been hard to explain why some children suffer severe malaria and die while others fail to succumb to life-threatening disease despite having the same exposure (Bull *et al*, 2005).

Plasmodium falciparum erythrocyte membrane protein1 (*PfEMP1*) are parasite proteins found on the surface on infected erythrocytes (IE) (Baruch, 1996). These proteins have been shown to be responsible for cytoadherence of IEs to the endothelial cells of blood capillaries. Cytoadherence of IEs in the microvasculature results in occlusion and contributes to the pathology of severe forms of disease such as cerebral malaria. PfEMP proteins act as ligands on erythrocyte cell surface that interact with receptors which mediate the cytoadherence process to endothelial cells. They include: CD36, thrombospondin

(TSP), and intercellular adhesion molecule 1 (ICAM-1). Vascular cell adhesion molecule 1 (VCAM-1) and endothelial leukocyte adhesion molecule 1 (ELAM-1) can also act as endothelial cell receptors for a minority of *P. falciparum* parasites. Thus PfEMP1 play a central role in host parasite interactions (Bull *et al*, 2005)

The PfEMP proteins are highly variable between parasites and are reported to show antigenic switching. This results in distinct cytoadeherence and agglutination phenotypes and variation during chronic infections in the human host (Baratin *et al*, 2007). This would imply that different parasite populations show different virulence patterns and levels of diseases severity. Additionally, variations in parasite derived antigens on IE surface would induce different forms of immune responses that may either aid in parasite clearance or aggravate pathology of disease (Baratin *et al*, 2007)

About 60 different *var* genes encode PfEMP1 proteins, but deletions, recombinations and gene conversions create a virtually unlimited repertoire. *Var* gene expression is mutually exclusive such that only a single copy is transcribed in one parasite at any one time (Baratin *et al*, 2007; Freitas-Junior *et al*, 2000; Chen *et al*, 1998). These PFEMP1 proteins are made up of a combination of different domains, each mediating a specific range of interactions with host molecules on endothelial cells, platelets, uninfected erythrocytes and dendrite cells (Bull *et al*, 2005; Urban *et al*, 1999). These interactions are thought to be responsible for severe pathology associated with *P. falciparum* malaria (Warimwe *et al*, 2009).

Additionally, PFEMP1 proteins are part of the variable surface antigens (VSA). They form some of the best VSAs that act as immune targets for naturally acquired immunity (Bull *et al*, 2005). They undergo clonal antigenic variation meaning that a single genotype can evade host antibodies by switching between *var* genes (Smith *et al*, 1995). After repeated exposure to infection, a repertoire of variant-specific antibodies that can recognize the variant surface antigens expressed by most parasite isolates builds up. The anti-VSA antibodies can exert a selection pressure on the repertoire of VSA expressed during

infection. Thus piecemeal acquisition of such antibodies leads to the development of naturally acquired immunity to malaria (Warimwe *et al*, 2009).

Var genes have a two-exon structure (Su *et al*, 1995). The first exon is large (~3.5 to 9.0 kb) and encodes multiple adhesion domains called the Duffy binding-like (DBL) and cysteine-rich inter-domain region (CIDR). The second exon is smaller (~1.0 to 1.5 kb) and codes for a more conserved cytoplasmic tail. Although PfEMP1 sequences are highly diverse, the adhesion domains can be grouped by sequence similarity (Smith *et al*, 2000) into seven types of DBL domains (α, α1, β, γ, δ, ε, and x) and four types of CIDR domains (α, α1, β, and γ) that have been used as criteria for dissecting PfEMP1 protein domain structures and binding functions (Kraemer *et al*, 2007).

The DBL α domain of the *var* gene has been a target for investigations of the gene repertoire to study diversity in laboratory strains and wild *P. falciparum* isolates. The flanking region of the DBLα domain is highly conserved while the middle region shows extreme diversity. Thus PCR typing has been used to investigate the nature of genetic diversity with respect to the *var* gene repertoire of laboratory-cultured parasites of different geographical origin and clinical field isolates. The results show certain PCR amplified products of parasites which clearly distinguish parasites from different locations and *in vitro* cultured parasites with clinical field isolates from malaria patients. This technique can be used to predict diseases severity (Ozarker *et al*, 2007).

Some studies have indicated that expression of certain cys2 *var* sequences could be associated with severe malarial anemia or impaired consciousness. There is evidence that the two severe forms of malaria show a differential association with expression of sequences containing an MFK motif (Warimwe *et al*, 2009). While expression of MFK motifs could be associated with impaired consciousness, expression of sequences containing REY motifs was associated with severe malarial aneamia. Further expression of these two motifs is mutually exclusive (Warimwe *et al*, 2009). Thus their presence in expressed sequences could be used to categorize malaria severity.

1.2 Statement of the problem

Malaria still remains a major public health problem despite the fact that various anti-malarial drugs are available. According to WHO's World Malaria Report (WHO, 2016), Kenya accounted for 14% of parasite prevalence in the Eastern and Southern African region in 2015. Additionally, there has been a report confirming treatment failure of the currently used artemisinin based combination therapy (ACTs) among individuals returning to United Kingdom after visiting Africa (Sutherland *et al*, 2017). Two of these individuals had visited Uganda, indicating that *P. falciparum* parasites circulating in the region could have developed resistance. Development of an effective vaccine has been hampered by variations in the parasite surface antigens. Although it is known that *var* genes code for parasite erythrocyte surface proteins, the polymorphism in the *pfemp*1 genes have not been associated to treatment outcomes among malaria patients in Kenya. This study will try to identify *Pfemp*1 one sequences that can be used as markers for disease severity.

1.3 Justification of the study

Considerable attention has been given to naturally acquired immunity to malaria. This immunity still remains poorly understood at molecular level. It has also been difficult to explain why some children get severe disease and die while others don't succumb to life threatening disease, despite having similar exposure (Bull et al, 2005). *Pfemp*1 gene products play an important role in host immune response to plasmodium infection. These surface proteins interact with various host receptors leading to different levels of disease severity. Since the proteins are coded for by variant forms of genes and only one subtype is expressed per parasite clone, they induce different levels of immune responses and disease severity. Several molecular tools are available and can be used to map and characterize *Pfemp*1 sequences. There is therefore need to use some of these tools to characterize *Pfemp1* sequences in field isolates. This data should then be compared to clinical data of each patient. This can provide a basis of developing molecular and epidemiological tools that can be used to predict disease severity. Further, the data generated can form a molecular basis for malaria vaccine development.

1.4 Research Questions

1. Which polymorphisms are found in the DBLα domain sequence tags of *Pfemp1*?
2. What are the differences in the DBLα sequence tags isolated from samples collected from the two sites?
3. Which dominant *var* sequences are expressed in field isolates?
4. What is the relationship between polymorphisms in *Pfemp*1 gene and clinical presentation malaria?

1.5.1 General Objective

To determine sequence profiles in the DBLα domains of *Pfemp*1 gene, in field isolates collected from two malaria endemic sites in Kenya.

1.5.2 Specific Objectives

1. To type polymorphisms in DBLα sequence tags of PfEMP1in field isolate samples collected at two malaria endemic sites.
2. To determine frequency and distribution of PfEPM1 DBLα sequence tags in genomic DNA of field isolates from Mbita and Tiwi.
3. To compare the frequency of PfEMP1 DBLα sequence tags in sequences isolated from Mbita and Tiwi field isolates.

CHAPTER TWO
LITERATURE REVIEW

2.1 Global malaria burden

Malaria continues to be one of the major parasitic diseases with at least 300 to 500 million new cases being reported annually. About 60% of the clinical cases and 80% of deaths occur in Africa, south of Sahara (D'Alesandro and Butteins, 2001). As at 2004, 107 countries had reported areas at risk of malaria transmission (WHO, 2005). Although this number is less than that in the 1950's, when 140 countries and territories were endemic, 3.2 billion people are still at risk.

Recent World Health Organization report (WHO, 2011)indicates that there is a reduction in malaria episodes in some countries. More than 50% reduction has been recorded between 2000 and 2010 in 43 of the 99 countries with ongoing transmission, while downward trends of 25%–50% were seen in 8 other countries. There were an estimated 216 million episodes of malaria in 2010, of which approximately 81%, or 174 million cases, were in the African Region. There were an estimated 655 000 malaria deaths in 2010, of which 91% were in Africa. Approximately 86% of malaria deaths globally were of children under 5 years of age. The estimated incidence of malaria globally has reduced by 17% since 2000 and malaria-specific mortality rates by 26%. These rates of decline are lower than internationally agreed targets for 2010 (reductions of 50%) but nonetheless, they represent a major achievement (WHO, 2011). This downward trend could however be threatened by reported resistance to artemisinin combination therapy used in the treatment of *P. falciparum*, the course of severe malaria cases, in Southeast Asia (WHO, 2011). It could further be compounded by developed insecticide resistance by the mosquito vector.

Those at highest risk are children younger than 5 years and pregnant women, particularly primigravida. Of the more than one million deaths that occur in Africa due to malaria, most are children under 5 years old. Apart from acute disease episodes and deaths in Africa, malaria also contributes significantly to anemia in children and pregnant women, adverse birth outcomes such as spontaneous abortion, stillbirth, premature delivery and low birth weight, and overall child mortality (Barnes and White, 2005).

In addition to being a public health problem, malaria flourishes in regions where societies are least prosperous. According to WHO, the global distribution of per-capita gross domestic product shows a striking correlation between malaria and poverty. The disease is estimated to be responsible for an estimated average annual reduction of 1.3% in economic growth for those countries with the highest burden. An annual economic loss of US$ 12 billion is estimated to be incurred globally (WHO, 2000).

Malaria is endemic in countries with slow economic growth. This could be due to various factors; which include effects of disease on fertility, population growth, saving and investment, worker productivity, absenteeism, premature mortality and medical costs. The malaria burden is further increased in countries in which the disease is endemic by parasite resistance to available anti-malarial drugs. Chloroquine and sulfadoxine-pyrimethamine drugs, which have been largely in use, have been rendered almost ineffective by resistance of *P. falciparum* (Shiff, 2002). Thus there has been a high rate of both clinical and parasitological failure of these drugs. Some studies indicate that after a steady decline in malaria deaths, the mortality rate due to malaria has increased in the last two decades in eastern and southern Africa, despite the drop in all-cause deaths among children, due to anti-malarial drug resistance. Despite advances in understanding malaria ecology, and development of interventions, more than 50% of the world's populations are still exposed to malaria. This is an increase of close to 10% over the past decade (Breman *et al*, 2004). This has thus increased the public health burden because of prolonged or recurrent illness and progression to severe malaria associated with increased hospitalization and death (Barnes and White, 2005). For instance, 69% of malaria deaths in Siaya, western Kenya, are attributable to drug resistance (Zucker *et al*, 2003).

According to the ministry of health, 20 million Kenyans are still at risk of getting infected with malaria. It accounts for 30 to 50% of out-patient visits and 29% of hospital admissions. 6000 women suffer from anemia in their first pregnancy while 4000 babies are born with low birth weight due to malaria. The disease is also thought to be a major contributor to rising poverty and significant loss of working hours and school attendance.

2.2 Plasmodium parasite

Malaria is caused by protozoan parasites of the genus ***Plasmodium*.** Four species of *Plasmodium affect man and cause* the disease in its various forms (Figure 1.1). They are:

a) Quartan species: *P. malariae,*

b) Benign tertian species: *P. vivax*

c) Malignant tertian species: *P. falciparum*

d) Ovule species: *P. (P.) ovale*.

*P. falciparum*is the most widespread and cause the mostsevere malaria of these four species and if untreated can lead to fatal cerebral malaria.

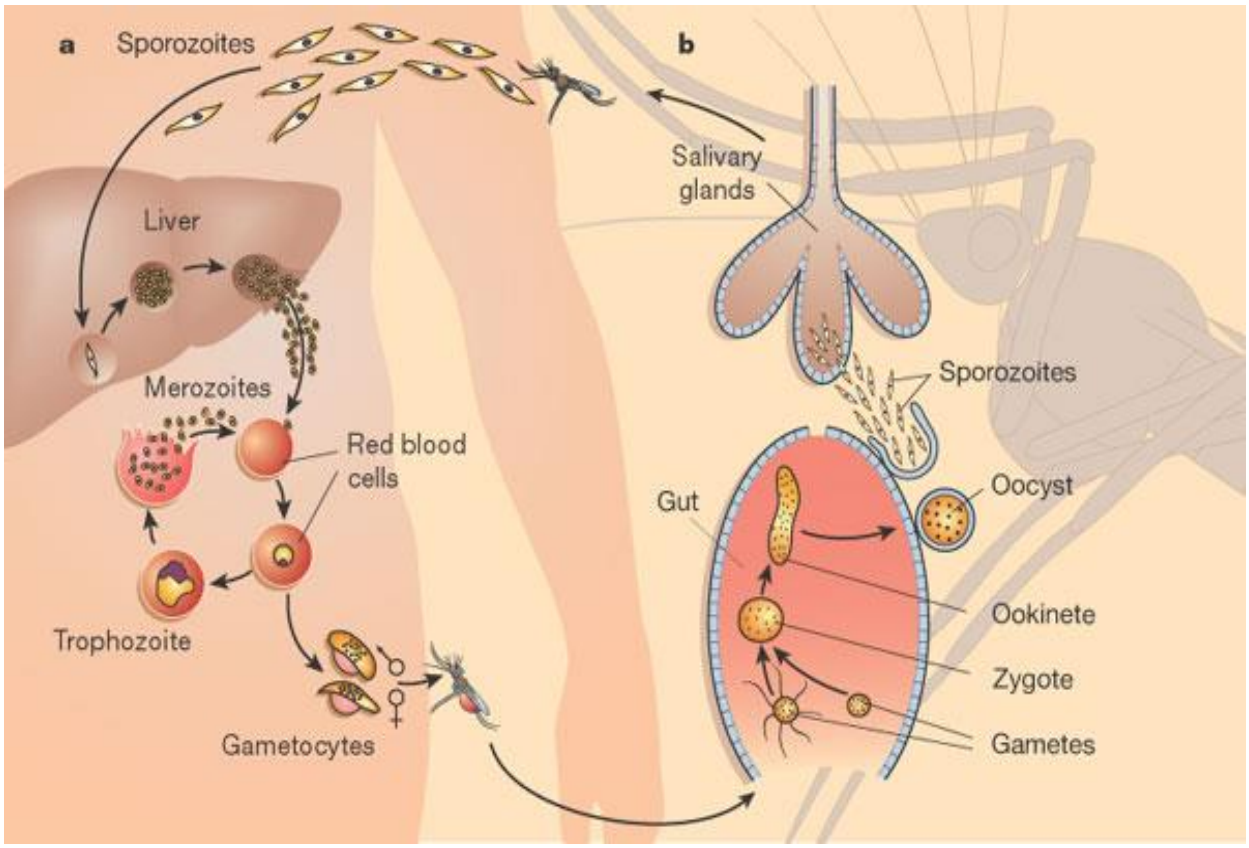


Figure 1.1The life cycle of the parasite Plasmodium falciparam

http://www.sanger.ac.uk/PostGenomics/plasmodium/presentations/plasmodium_lifecycle.shtml

2.3 Diagnosis of malaria

The World Health Organization (Bloland, 2001) recommends the methods below for malaria diagnosis.

2.3.1 Microscopy

Simple light microscopic examination of Giemsa-stained blood films is the most widely practiced and useful method for definitive malaria diagnosis. Advantages include differentiation between species, quantification of the parasite density, and ability to distinguish clinically important asexual parasite stages from gametocytes which may persist without causing symptoms. These advantages can be critical for proper case-management and evaluating parasitological response to treatment. Specific disadvantages are that slide collection, staining, and reading can be time-consuming and microscopists need to be trained and supervised to ensure consistent reliability (Bloland, 2001; Drakeley *et al* 2005).

2.3.2 Clinical (presumptive) diagnosis

Although reliable diagnosis cannot be made on the basis of signs and symptoms alone because of the non-specific nature of clinical malaria, clinical diagnosis of malaria is common in many malarious areas. In much of the malaria-endemic world, resources and trained health personnel are so scarce that presumptive clinical diagnosis is the only realistic option. Clinical diagnosis offers the advantages of ease, speed, and low cost. In areas where malaria is prevalent, clinical diagnosis usually results in all patients with fever and no apparent other cause being treated for malaria. This approach can identify most patients who truly need anti-malarial treatment, but it is also likely to misclassify many who do not (Bloland, 2001; Drakeley *et al*, 1999).

2.3.3 Antigen detection tests (rapid or "dipstick" tests)

This approach involves the rapid detection of parasite antigens using rapid immunochromatographic techniques. Commercially available kits are based on the detection of

the histidine-rich protein 2 (HRP-II) of *P. falciparum.* Compared with light microscopy, this test yields rapid and highly sensitive diagnosis of *P. falciparum* infection. Advantages to this technology are that no special equipment is required, minimal training is needed, the test and reagents are stable at ambient temperatures, and no electricity is needed. The principal disadvantages are a currently high per-test cost and an inability to quantify the density of infection (Bloland, 2005; Moody, 2002; WHO, 1999). However, circulating antibodies can hinder detection of HRP-2 protein by the RTD when parasetemia is low, leading to false negative (Ho *et al, 2014)*. Other studies indicate that deletions in hrp2 and hrp3 genes would also lead to false negatives (Beshir *et al,* 2017).

2.3.4 Molecular tests

Detection of parasite genetic material through polymerase-chain reaction (PCR) techniques is becoming a more frequently used tool in the diagnosis of malaria, as well as the diagnosis and surveillance of drug resistance in malaria (Bloland, 2005). Specific primers have been developed for each of the four species of human malaria. The major advantages of using a PCR-based technique are the ability to detect malaria parasites in patients with low levels of parasitemia and identify them to the species level. Infection with five parasites or less per µl can be detected with100% sensitivity and equal specificity. In addition, improved PCR techniques could prove useful for conducting molecular epidemiological investigations of malaria clusters or epidemics. Primary disadvantages to these methods are overall high cost, high degree of training required, need for special equipment, absolute requirement for electricity, and potential for cross-contamination between samples (Bloland 2005; Moody 2002).

2.3.5 Serology

Techniques also exist for detecting anti-malaria antibodies in serum specimens. Specific serological markers have been identified for each of the four species of human malaria. A positive test generally indicates a past infection. Serology is not useful for diagnosing acute infections because detectable levels of anti-malaria antibodies do not appear until weeks post infection and persist long after parasitaemia has resolved (Bloland, 2005; Howden *et al*, 2005).

2.4 Malaria Pathology

Pronounced changes during malaria occur in blood and blood forming systems, the spleen and the liver. Other changes may occur in other major organs depending on the type and severity of malaria. Pathological changes are more pronounced and severe in *P. falciparum* malaria. The pathology of malaria mainly involves erythrocytes in which the parasite destroys and feeds on hemoglobin. This causes anemia. In addition, the infected erythrocytes develop protuberances with histidine-rich parasite proteins that facilitate cytoadherence of the infected red blood cells onto receptors on venular and capillary endothelium. This sequestration of infected erythrocytes prevents their free circulation and promotes the parasite's survival by avoiding passage through the immunologically active spleen. Additionally, it may cause cerebral malaria if it occurs in the brain (Barnwell *et al*, 1989).

Clinical manifestations like vomiting and diarrhea occur when the gastrointestinal system is affected. When the central nervous system is affected, there could be seizures, prostration, coma, abnormal postures and deep tendon reflexes and brain stem abnormalities with evidence of raised intracranial pressure (Idro *et al*, 2005). Organs like the liver, spleen and lungs may also be affected during *P. falciparum* infection.

2.5 Malaria control

Until the nineteenth century, malaria transmission occurred in many parts of the world, including Europe and North America. But malaria reduced in these areas due to improved housing, ecological/land use changes (Arrow *et al*, 2004). This was augmented by the use of DDT, a highly effective insecticide introduced at the end of World War II. Thus through the use of residual insecticides and manipulation of environmental and ecological characteristics, malaria parasites were eliminated from Europe and N. America (Shiff, 2002)

By the 1950s' and 60s', residual spraying with DDT was the centre piece of global malaria eradication efforts. DDT's month long ability to kill/deter fed mosquitoes from resting on treated walls led to malaria decline in India, Sri Lanka and the former Soviet Union (Arrow *et al*, 2004). By 1966, DDT spraying and elimination of mosquito breeding sites freed more than 500 million people living in malarious zones. These efforts were not sustained because of the emergence of DDT resistant mosquito, community resistance to house spraying and high costs (Shiff, 2002; Arrow *et al*, 2004; Talisuna *et al*, 2004). In Sri Lanka, malaria eradication commenced in 1947, and in 1963, at the height of the eradication programs, only 17 cases were recorded throughout the country. Mass population movement into forested areas and the withdrawal of spraying with DDT due to the advanced level of control achieved contributed to faltering control, and in 1969, over 500,000cases were reported in the country. The eradication efforts through vector control was abandoned and the pendulum swung to over-reliance on the then effective drug, chloroquine.

In sub-Saharan Africa, no sustained malaria control efforts were ever mounted. The major vector in this region is *Anopheles gambiae*. It is wide spread, long-lived and more aggressive. Its entomological inoculation rates [EIR] (the frequency which a human is bitten by an infected mosquito) are over 1000. These EIRs rarely exceed 5/year in Asia and South America (Arrow *et*

al, 2004).Today, the global malaria burden is concentrated in sub-Saharan, where stable endemic disease is linked to poverty and a highly efficient vector.

The insecticide treated bed nets (ITNs) were first shown in the Gambia to reduce overall childhood mortality of children age between 1-4 years by 60% when combined with malaria chemoprophylaxis (Alonso *et al*, 1993). With respect to malaria's human reservoir, the big challenge facing Africa is the development of drug resistance by *P. falciparum* to readily available and affordable anti-malarial drugs such as sulfadoxine-pyrimethamine and chloroquine.

In general, basic principles of malaria control would focus on: measures to prevent human vector conduct; measures to prevent or reduce the dreading of mosquito; measures to destroy mosquito larvae; measures to kill/or reduce adult mosquito and measures to eliminate parasites from humans; measures to prevent and reduce malaria mortality (especially in high risk groups), reduce malaria morbidity and malaria transmission (Arrow *et al*, 2004).

Current control efforts rely on ITNs which interrupt human-vector contact and treatment of infected persons. ITNs also kill vectors and reduce local transmission. Indoor residual spraying can reduce vectors when a majority of houses in the targeted community are sprayed (Arrow *et al*, 2004).. This is preferred for vector control during malaria epidemics and in refugee camps since trained teams can rapidly cover likely areas *P. falciparum* transmission.

Case management by prompt access to healthcare, accurate diagnosis and effective treatment are the other cornerstones of malaria control. The current failure to control malaria with drugs often starts with a failure to deliver appropriate case management to malaria sufferers. In 1993, WHO outlined other control strategies that include early forecasting of malaria epidemics, development

of epidemiological information systems, capacity building in basic and applied research and ongoing assessment of ecological, social and economic determinants of disease within affected regions and countries (WHO 2006). The adoption of artemisinin based combination therapy as first line replacement to SP drugs in malaria treatment may further control transmission through their direct gametocytocidal effects.

The current malaria control strategies adopted by the Ministry of Health and the malaria control programme in Kenya focus on provision of prompt and effective treatment of at least 80% of children within 24 hours of fever onset; ensuring that 60% of children and pregnant women sleep under ITNs; containing epidemics through indoor residual spraying of houses and community mobilization to control the vector. Surveillance and early detection, identification of high risk areas, targeted indoor residual spraying and public education are other strategies of epidemic preparedness and response.

2.5 Malaria chemotherapy

Antimalarial drugs are medicines that prevent or treat malaria**.** There are various drugs used in the chemotherapy and chemoprophylaxis of both severe and uncomplicated malaria. Below are some antimalarials currently used for treatment and those that are still under development or are undergoing clinical trials as adopted from the Antimalarial Drugs Focus Group Report of the National institute of Health of the US government.

2.5.1 Aminoquinolines

2.5.1.1 Amodiaquine

Amodiaquine is a 4-aminoquinoline antimalarial drug similar in structure and activity to chloroquine. Like chloroquine, it also possesses antipyretic and anti-inflammatory properties.

Amodiaquine has the advantage over chloroquine of being more palatable and therefore easier to administer to children. Amodiaquine is no longer recommended for chemoprophylaxis because of the risk of severe adverse reactions (WHO, 2001; Mutabingwa *et al*, 2005).

2.5.1.2 Mefloquine:

Mefloquine is a 4-quinoline methanol chemically related to quinine. It is a potent long-acting blood schizonticide active against *P. falciparum* resistant to 4-aminoquinolines and sulfa drug-pyrimethamine combinations. It is also highly active against *P. vivax* and, *P. malariae* and most probably *P. ovale*. It is not gametocytocidal and is not active against the hepatic stages of malaria parasites. Owing to its long elimination half-life and consequent long-lived sub therapeutic concentrations in the blood, the development of resistance is to be expected especially in areas of high transmission. Since the late 1980s, resistance of *P. falciparum* to mefloquine has developed.

P. falciparum resistance to mefloquine is accompanied by cross-resistance to halofantrine and reduced sensitivity to quinine. In contrast, laboratory studies have shown some increase in the sensitivity of mefloquine-resistant isolates to chloroquine (WHO, 2001 Bavdekar *et al*, 1996).

2.5.1.3 Primaquine:

Primaquine is an 8-aminoquinoline highly active against the gametocytes of all malaria species found in humans and against hypnozoites of the relapsing malarial parasites, *P. vivax* and *P. ovale*. As a gametocytocide for *P. falciparum*, it is effective given in a single dose of 30-45 mg of base (0.5-0.75 mg of base per kg) (WHO, 2001; Bavdekar *et al*, 1996).

2.5.1.4 Chloroquine

Chloroquine is a 4-aminoquinoline derivative of quinine first synthesized in 1934. It had been the drug of choice for treatment of non-severe or uncomplicated malaria in most parts of malaria endemic areas for some time until, resistance was developed by the *P. falciparum* species. It is a very potent schizonticidal drug against erythrocytic stage of all the four *Plasmodium* species.

The use of chloroquine as a single first-line drug treatment is now increasingly limited following the evolution of chloroquine-resistant *P. falciparum*, but chloroquine remains the first-line drug of choice in most African countries south of the Sahara where acceptable clinical cure rates can be obtained. In areas where it is still used as a first-line drug, persistent parasitemia and lack of hematological recovery in children may be one of the early signs of chloroquine resistance. Even if the frequency of clinical failures is acceptable in the general population, a more effective first-line treatment may be required for vulnerable groups such as young children and pregnant women (WHO, 2002; WHO, 2001, Bavdekar *et al*, 1996).

2.5.2 Antifolates

2.5.2.1 Sulfadoxine-pyrimethamine

The only useful combinations of antifolate drugs for the treatment of malaria are synergistic mixtures that act against the parasite-specific enzymes, dihydropteroate synthetase and dihydrofolate reductase. Available combinations include the sulfa drug-pyrimethamine combinations, sulfadoxine-pyrimethamine and sulfalene-pyrimethamine, the former being more widely available. The long half-life of sulfa drug-pyrimethamine combinations provides a potent selective pressure for parasite resistance in areas of high transmission. In Africa since the late 1980s, *P. falciparum* sensitivity has decreased, particularly in East Africa where sulfadoxine-pyrimethamine has been used on a large scale, and resistance is demonstrable in parts of West Africa. Sulfa drug-pyrimethamine combinations have been successfully used in areas with highly developed *P. falciparum* resistance to chloroquine and during malaria epidemics. Compliance is high since they offer single-dose therapy (WHO, 2001). Intermittent treatment of pregnant women currently relies on sulfadoxine-pyrimethamine.

2.5.2.2 Proguanil

Proguanil is a synthetic biguanide derivative of pyrimidine with a marked effect on the primary tissue stages of *P. falciparum, P. vivax* and *P. ovale*. Proguanil is a dihydrofolate reductase inhibitor acting primarily through its major metabolite, cycloguanil. In addition, cross-resistance between cycloguanil and pyrimethamine is not absolute, resistance to the two drugs being controlled by different point mutations on the dihydrofolate reductase/thymidylate synthase (DHFR/TS) gene. It is known that proguanil has a second, non-antifolate mechanism of action and this may explain the effect of proguanil-atovaquone (WHO, 2001).

2.5.3 Quinine

Quinine is normally effective against falciparum infections that are resistant to chloroquine and sulfa drug-pyrimethamine combinations. Decreasing sensitivity to quinine has been detected in areas of South-East Asia where it has been extensively used for malaria therapy. This has occurred particularly when therapy was given in an unsupervised and ambulatory setting with regimens longer than 3 days. In these settings, patient adherence to therapy is low, leading to incomplete treatment; this may have led to the selection of resistant parasites. There is some cross-resistance between quinine and mefloquine, suggesting that the wide use of quinine in Thailand might have influenced the development of resistance to mefloquine in that country. Strains of *P. falciparum* from Africa are generally highly sensitive to quinine (WHO, 2001; Bavdekar *et al*, 1996).Quinine is still the drug of choice for severe falciparum malaria in most countries. It should only be used for uncomplicated malaria when alternatives are unavailable.

2.5.4 Halofantrine

Halofantrine, a phenanthrene methanol, is a blood schizonticide that is active against all malaria parasites. It is active against *P. falciparum* infections that are resistant to chloroquine and to sulfa

drug-pyrimethamine combinations. Early studies indicated that halofantrine was also active against some but not all isolates with reduced susceptibility to mefloquine. Halofantrine has no place in malaria control programmes because of its high cost, its variable bioavailability, its cross-resistance to mefloquine and the fact that fatal cardio toxicity has been reported in certain risk groups following standard therapy. It may be used on an individual basis in patients known to be free from heart disease in areas where multiple drug resistance is prevalent and no other effective antimalarial is available (Bavdekar *et al*, 1996; WHO, 2001).

2.5.5 Artemisinin and its derivatives:

Artemisinin *(qinghaosu)* is the anti-malarial principle isolated by Chinese scientists from *Artemisia annua* L. It is a sesquiterpene lactone with a peroxide bridge linkage. Artemisinin is poorly soluble in oils or water but the parent compound has yielded dihydroartemisinin, the oil-soluble derivatives artemether and arteether, and the more water-soluble derivatives sodium artesunate and artelinic acid. These derivatives have more potent blood schizonticidal activity than the parent compound and are the most rapidly effective antimalarial drugs known. They are used for the treatment of severe and uncomplicated malaria. They are not hypnozoiticidal but gametocytocidal activity has been observed (Barnes and White, 2005; WHO, 2001).

The anti-malarial activity of artemisinin and its derivatives is extremely rapid and most patients show clinical improvement within 1-3 days after treatment. However, the recrudescence rate is high when the drugs are used in monotherapy. Artemisinin derivatives are as listed below.

2.5.5.1 Artemisinin

Artemisinin is a sesquiterpene lactone with a peroxide bridge linkage that appears to be responsible for its anti-malarial activity. Artemisinin is a potent and rapidly acting blood

schizonticide, eliciting shorter parasite clearance times than chloroquine or quinine and rapid symptomatic responses (WHO, 2001).

2.5.5.2 Artemether

Artemether is an oil-soluble methyl ether derivative of dihydroartemisinin. As with artemisinin, it is effective against *P. falciparum* resistant to all other operationally used anti-malarial drugs (WHO, 2001). Artemether-lumefantrine combination is the current firstline drug for malaria treatment in Kenya.`

2.5.5.3 Artesunate

Artesunate, a water-soluble hemi-succinate derivative of dihydroartemisinin, is the most widely used member of this family of drugs. It is unstable in neutral solutions and is therefore only available for injections as artesunic acid. It is effective against *P. falciparum* resistant to all other operationally used anti-malarial drugs. It does not have hypnozoiticidal activity. It reduces gametocyte carriage rate (Barnes and White, 2005; WHO 2001).

2.5.5.4 Dihydroartemisinin

Dihydroartemisinin is the active metabolite of artemisinin and its derivatives. These derivatives have more potent blood schizonticidal activity than the parent compound. Dihydroartemisinin is the most potent anti-malarial of this group of compounds but it is also the least stable (WHO, 2001). Dihdroartemisinin-piperaquine combination is currently the second line of malaria treatment in Kenya.

2.5.5.5 Arteether

Arteether is the oil-soluble ethyl derivative of dihydroartemisinin. Clinical trials in India have indicated that it is an effective and rapidly-acting drug for the treatment of uncomplicated and severe *P. falciparum* malaria (WHO, 2001).

2.5.5.6 Artelinic Acid

Artelinic acid is a water-soluble derivative of artemisinin and is thought to be more stable than artesunate in solution thus offering the potential for oral administration. The compound is still under investigation (WHO, 2001).

2.5.6 KAF156

This is a new anti-malarial drug belonging to imidazolopiperazines group that has been tested in South Eastern Asia (White *et al*, 2016). However, in vitro studies have identified mutations in *Plasmodium falciparum* Cyclic Amine Resistance Locus (PfCARL) gene to aid parasite resistance to KAF156, a potential threat to its wide use in the population.

2.6 Drug resistance and insecticide resistance

The term "parasite resistance" has been defined as "the ability of a parasite strain to survive and/or multiply despite the administration and absorption of a drug in doses equal to or higher than those usually recommended but within the limits of tolerance of the subject". This definition was subsequently modified to specify that "the drug must gain access to the parasite or the infected red blood cell for the duration of the time necessary for the normal action of the drug" (D'Alesandro and Butteins,2001). Resistance has also been used when referring to therapeutic failure after administration of a standard dose of a drug; this definition is used in the WHO standard *in vivo* test protocol (Tran and Saier, 2004). However, in this test, serum drug levels are not normally measured and the observed therapeutic failure might be due to malabsorption, rapid or abnormal metabolism, or the presence of latent infections other than malaria. Therefore, it is important to define the terms being used in order to avoid confusion.

Resistance is also thought of as a shift to the right of the drug response curve, thus requiring higher drug concentrations to achieve the same parasite clearance (White, 2004). Resistance emerges de-novo through spontaneous gene mutations (single nucleotide polymorphisms [SNPs]) or gene duplications. These are thought to be independent of drug pressure, but the mutations are then selected for and spread as a result of drug pressure which provides a selective advantage to resistant parasites (Barnes and White, 2005).

Development of resistance therefore involves two distinct phases: de-novo emergence followed by subsequent spread. Resistance arises mainly during the asexual reproduction and may require a single genetic event as is the case in antifolate or atovaquone resistance, or multiple events (epistasis) as seen in chloroquine resistance. Meiosis however, does take place in the mosquito. Additionally many infections are multi-clonal and sometimes a single mosquito may bite two infected gametocyte carrying individuals. Thus there exists in the mosquito the possibility of recombination with the formation or breakdown of multigenic resistance.

2.7 Malaria vaccine

Concerted efforts have been employed over the last decade to develop a functional malaria vaccine. These have seen utilization of various forms of immunogens, ranging from whole parasite to specific molecules thought to evoke protective immunity against malaria infection (Coelho *et al,* 2017). Some of malaria vaccine candidates have been developed from attenuated sporozoites while others utlise various antigenic molecules of parasite origin. Vaccines targeting pre-erythrocytic phase of malaria aim to inhibit development of the parasite at early stages of infection and thereby limit symptoms and transmission (Coelho *et al,* 2017). The most recent effort in the development of malaria vaccine has seen large scale clinical trails of RTS,S vaccine. This is a fusion protein derived from circumsporozoite genetically fused to B virus surface antigen (HBsAg) that is indicated for children aged 6 to 17 months (Kaslow and Biernaux, 2017). Clinical trials RTS,S/AS01 vaccine candidate in Sub-Saharan showed and efficacy of about 38% offering protection for up to 3-4 years in infants and young children (RTSS,S Clinical Trial Partnership, 2015). Further anslysis have shown that variations in HLA molecules may affect the efficacy of this vaccine (Nielson *et al,* 2018). Experts have recommended the adoption of this vaccine in malaria control in Sub-Saharan Africa to lower malaria morbidity and avert infant mortally due to malaria.

2.8 Antigenic variation and immune-evasion in malaria

Antigenic variation has been defined as the process by which pathogens alter their surface proteins so as to evade the immune responses of a host (Ferreira *et al*, 2004) Plasmodium exhibit an extensive diversity of surface antigens. This is one of the main reasons why clinical immunity to malaria develops only after repeated exposure/infections to the same species for several years. Antigenic diversity could result from:

i) Complex parasite life cycle

ii) Allelic polymorphisms due to nucleotide replacement and recombination resulting in genetically stable alternative forms of genes coding for antigens

iii) Antigenic variation in which a parasite clone expresses alternative forms of an antigen without genetic variation (Ferreira *et al*, 2004)

Antigenic variation also results from surface proteins that *P. falciparum* parasite exports to the surface of the infected erythrocytes. Most malarial antigens have tandem arrays of short amino acid motifs. One such antigen is the circumsporozoite protein (CSP) that comprises of immune-dormant B-cell epitopes. In *P. falciparum*, CSP contains 37 to 50 copies of 4-mer motifs. Monoclonal antibodies to CSP repeats inhibit the infectivity of *P. falciparum* and *P. vivax* sporozoites in chimpanzees, but no clear association has been found between the levels of naturally acquired antibodies to CSP and human protection from malaria in Africa and Asia. Merozoite surface protein 1 (MSP-1) and MSP-2 are additional examples of blood-stage vaccine-candidate antigens with repetitive arrays. MSP-2 has been characterized so far only in *P. falciparum* and the closely related chimpanzee parasite, *P. reichenowi* (Zaval *et al*, 1985).

Surface antigens are characterized by possessing repetitive motifs that are highly variable. This variability contributes to immuno-evasive mechanisms employed by *P. falciparum* to survive in

the host. One such variable surface antigen is the *Plasmodium falciparum* erythrocyte membrane protein 1(PfEMP 1), a parasite protein exported to the surface of infected erythrocytes. PfEMP1 protein is coded by a family of hypervariable genes referred to as *var* genes. Each parasite haploid genome has about sixty *var* genes. The expression of *var* genes is mutually exclusive such that only one gene is expressed at a time. The expression of these genes varies is such a way that a parasite clone would express different *var* genes at different times. Since PfEMP1 protein is exposed on the surface of infected erythrocyte, they form epitopes that can evoke anti-plasmodial immune responses. However these responses are rendered ineffective due to frequent variation of these surface antigens. The implication is that immunity to *P. falciparum* infection develops over a long time, requiring several exposures to different parasite clones before an individual becomes semi-immune. Thus severe symptoms of malaria are exhibited among children below five years of age and immune-suppressed individuals like pregnant women. Furthermore, the acquired immunity is not sterile. One needs constant exposure to *P. falciparum* to maintain a minimum immunological memory that will protect against, limit or resolve subsequent infections. Thus a stay away from malaria endemic region results in loss of the immunity to malaria.

Var genes are largely grouped into three main groups based on their chromosomal location, direction of transcription and upstream promotor sequence(UPS) hence genes are also named as UPSA, UPSB and UPSC (Bengtsson *et al*, 2013; Rask *et al* 2010). UPSA and UPSB are located in the subtelomeric region in a tail to tail orientation (Howell *et al*, 2008) while UPSC genes are centrally located and are oriented head to tail in a tandem repeat manner. The hyper-variable gene repertoire is generated by frequent meiotic ectopic recombinations in the mosquito. Mitotic recombination may also occur to further diversify the gene repertoire. Group A genes are less

diverse compared to other genes and their transcription has been associated with severe symptoms of malaria (Warimwe *et al*, 2009; Bengtsson *et al*, 2013) in immune naïve patients (Figure 2.1).

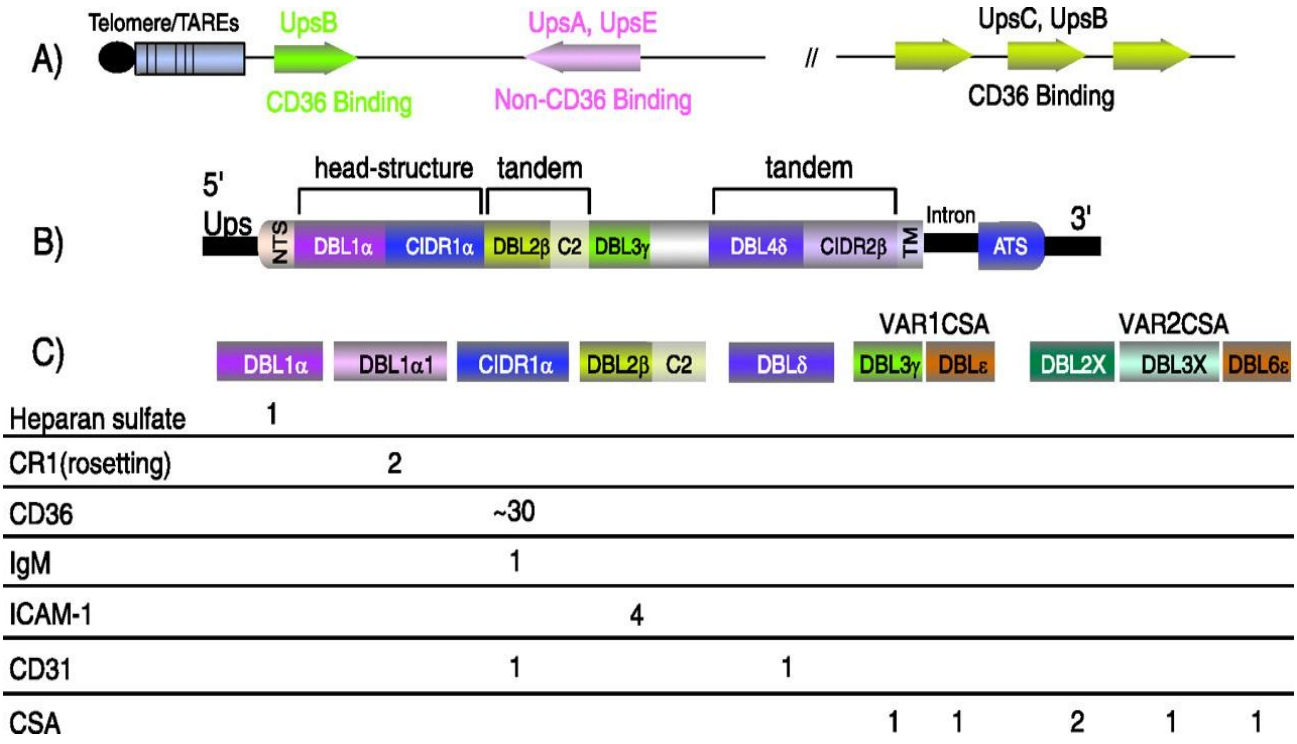


Figure 2.1Chromosomal organization of *var* gene- (Kyes *et al*, 2007).

(A) Genes are classified according to upstream promoter type and are found at the ends of most of the 14 *P. falciparum* chromosomes and clustered in internal regions on chromosomes 4, 7, 8, and 12. Arrows indicate the direction of transcription. TAREs, telomere associated repeat elements. (B) The extracellular binding region of the PfEMP1 proteins consists of four domain types: NTS, DBL, CIDR, and C2. DBL and CIDR domains are classified according to sequence similarity. (C) Relatively few PfEMP1 adhesion traits have been mapped to specific domains. The numbers of sequences with mapped adhesion domains are indicated for each host receptor-domain interaction

2.9 Duffy binding like alpha (DBLα) sequence

Duffy binding like alpha (DBLα) sequence tags from *var* genes although highly variable, possess some regions with limited variation that have been used to group them into various categories. These regions referred to as positions of limited variation (PoLVs) are used to study DBLα sequence tags and also the implications of their presence (Warimwe *et al*, 2009). In most studies, the PolVs are determined after an initial screening of 5' DIGDI and 3' PQFLR motifs on the sequences (Bull *et al*, 2007, 2008). The PoLV1 is therefore located relative to the 5' motif on the downstream side. PoLV2 and 3 are anchored on a conserved WW motif internal in the sequence tag while PoLV4 is located relative to 3' motif.

The DBLα sequence tags can be classified into several groups based on the number of cysteine residue and the motifs present at PoLV1 and PoLV2 (Bull *et al*, 2008). A majority of DBLα sequence tags contain four cysteine (cys4) residues. Others have two cysteine (cys2) residues while fewer sequences contain 0, 3,5 or 6 cysteine residues and are denoted as cysX (Bull *et al*, 2008). This classification system by Bull *et al* (2008) was based on the fact that the DBLα sequence tags have a 5' DIGDI and 3' PQFLR motifs. The objective of this project was to determine the types of *var* genes expressed in the field isolates and classify and compare PfEMP1 DBLα sequence tags in isolates from the two endemic sites. In order to analyse all the sequence tags isolated from the field isolates from Mbita and Tiwi, the classification system varied to include sequences with other 3' motifs. Thus selection of cysteine residues was not strictly based on the system adopted by Bull *et al* (2008). The sequence tags were further divided into several groups based on the number of cyteine residues and the motifs at PoLV1 and PoLV2 as described by Bull et al (2008). This system classifies DBLα sequence tags into six groups. Group1 consists of cys2 sequences with MFK motif at PoLV1. Group sequences are cys2 sequences with REY motif at PoLV2. Group3 sequences are cys2 DBLα sequences without MFK and REY motifs at PoLV1 and PoLV2 respectfully. Sequence tags with

four cystene residues (cys4 sequences) that do not have REY motif at PoLV2. Cys4 sequences with REY motif at PoLV2 belong to group5. Sequences with 0, 1, 3 5 of 6 cysteine. In this study all sequence tags were analysed to determine the number of cysteine residues present in the each sequence and the motif at PoLV1 and PoLV2. Sequence tags lacking 3' PQFLR motif were also included and the term group-like was used to imply that they had cys/PoLV features of the particular group they were classified into but lacked a 3' PQFLR. The results for this classification are as described below.

CHAPTER THREE
MATERIALS AND METHODS

3.1 Study sites

The study sites were Mbita Sub-District Hospital and Tiwi Health Centre. Mbita is located in western Kenya(0° 30′ 0′′ south, 34° 15′ 0′′ East) in Homa Bay County. It is next to Lake Victoria, and experiences moderate rainfall. It is a malaria hyperendemic zone with transmission of malaria taking place all year round. *P. falciparum* is the predominant malaria species in this region. Tiwi (4° 14′ 0′′ South, 39° 35′ 0′′ East) on the other hand is in Kwale County on the South Coast, along Indian Ocean shores. While year-long temperatures are ambient for mosquito breading, malaria transmission peaks between May and July. Inhabitants of these two sites show some level of "herd" immunity due to frequent exposure to the malaria parasite.

3.2 Sample collection and study approval

Samples were collected from patients visiting Mbita Sub-District Hospital in Homabay County, Western Kenya and Tiwi Health Centre in Kwale County, Coastal region during drug efficacy studies. All participants showed mild symptoms of malaria. Finger prick blood was collected on Whatmann filter paper and left to air dry to for dry blood spots on the filter paper. Dry filter papers were labeled for each patient and placed in an envelope then transported to Kenya

Medical Research Institute where they were stored at -20^{0}C. Patients showing severe malarial symptoms were referred for admission and were not included in the study. This study was approved by Scientific and Ethical Review Committees of Kenya Medical Institute (KEMRI). Blood spot samples on Whatman filter paper were collected from consenting patients after microscopic confirmation that they were positive for *P. falciparum* mono-infection. Filter papers were air dried and then transported to KEMRI, malaria laboratory for analysis.

3.3 Eligibility

3.3.1 Inclusion criteria

The inclusion criteria were:

i. Mono-infection with *P. falciparum* as detected by microscopy
ii. Signing of informed consent form
iii. Mild form of the disease

3.3.2 Exclusion criteria

i. Patients whose guardians were unwilling to sign the consent form
ii. Patients with severe forms of malaria were excluded from the study.

3.4 Laboratory procedures

3.4.1 DNA extraction

Parasite DNA was extracted using chelex method. Briefly, a piece of filter paper (approximately 2mm x 2mm) with the blood spot was incubated in 1000µl of 0.5% saponin in 1x PBS overnight at 4^{0}C. The resulting brown solution was then discarded and replaced with fresh 1xPBS, followed by an incubation of 15 to 30 minutes. The solution was removed and 100µl of DNAse free water was added, followed by 50µl of 20% chelex. The tubes were then incubated on a

heated block at 100^0C for ten minutes, being vortexed every two minutes. The solution was then centrifuged at 3000 rpm for 3minutes. The supernatant was removed, placed in fresh tubes and centrifuged again. The resultant supernatant was removed, this time ensuring that no chelex was picked, and stored at -20^0C awaiting PCR. 5µl of genomic DNA extracted by chelex was then used as DNA template in PCR.

3.4.2 Isolation and amplification of DBLα sequence tags by PCR

To isolate DBLα sequence tags, 5µl of genomic DNA was amplified using AF' GCACG (A/C) AGTTT(C/T) GC (forward primer) and DBLα BR, GCCCATTC (G/C) TCGAACCA (reverse primer) (Bull *et al,* 2005*)*. A PCR program of 35 cycles was carried out at a denaturation temperature of 94^0C, annealing temperature of 42^0C and 65^0C extension and a final extension of 65^0C. Each reaction tube had a total volume of 25µl consisting of 6.56µl ddH2O, 0.25µM of each outer primer, 1× standard PCR buffer (1.5mM MgCl2, 50mM KCl, 10mM TrisHCl(pH8.3), 0.5% DMSO), 200µm of each of the dNTPs,1 unit of Taq polymerase (KEMTAQ®) and 8µl of DNA template.

3.4.3 Sequencing of PCR products by 454-sequencing (ROCHE)

Amplified PCR products were sequenced by 454 sequencing, Roche^{TM} at the International Livestock Research Institute (ILRI), Nairobi campus. Briefly, cleaned PCR products were used to prepare sequence libraries which were then amplified to generate the sequence reads. The 454-sequence reeds were assembled using the Newbler 2.3.5 program from Roche. The SSF files were converted into Fasta format based on quality scores. The reads from each sample were then translated into amino acid sequence tags. One hundred (100) amino acids were used as the cut-

off for any single read to be translated from nucleotide to amino acid sequence. DBLα sequence tags were then isolated from the 454 sequence reads and grouped into contigs consisting of sequence tags in each sample that were similar or had overlapping reads implying they corresponded to the same *var* gene and/or region and singlets consisting of sequence reads that occurred only once in a sample using Newbler software.

3.4.4 Bioinformatics analysis of 454 sequence reads

The 454 sequencing data was released as SSF files (standard software data format files). A special bash script was used to convert these SSF files to fasta then to fastq and finally to fastaqual as shown appendix VI. This script thus converted the data into Fasta format and also gave the quality scores for each base in each 454 read.

3.4.4.1 Assembly of reads

The 454 reads from each were then assembled into contiqs and signlets using cap3 product. A bash script was first constructed then used to run all the sequences in Cap 3 product. Cap 3 uses quality scores to clip sequences so that base reads have the required quality score. Cap 3 grouped sequences from each isolate into contiqs and singlet sequences. Each contiq consisted of sequences with similar reads of slight over laps into the same contiq. Such sequences are likely to have been amplified from the same *var* gene or same chromosomal position. Singlets occur only once in the isolate and cannot be grouped with any other sequence. That is they are unique sequences that occur only once within the parasite isolate. Bioruby (https://github.com/georgeG) was used to concatenate all the reads in each isolate in *var* finder. These gave the number of DBLα tags in each isolate. The tags were then translated from nucleotide into amino acid sequences. This was on condition that a tag could only qualify for translation if it was at least

100 amino acids long. The 454 reads were deposited in GenBank, accession number: KP085750-KP087726.

3.4.5 Analysis of DBLα sequence tags

The amino acid sequence tags were aligned in Mega 5.2 (Tamura *et al*, 2011), and the aligned sequenced were the used to identify and group the DBLα tags based on the number of cysteine residues present in each tag. They were also used to assess polymorphisms within positions of limited variability (PoLV).

In order to assess the sequences, all sequence tags were screened for the motif present at 5' and 3' positions of the tags. The PoLVs were then identified based on either the 5' motif (PoLV1), 3' motif (PoLV4) or WW motif in the interior of the sequence (PoLV2 and PoLV3)

The mega such motif command was used to highlight cysteine residues in the sequence tags. The number of cysteine residues in each sequence tag was counted manually. This number of cysteines and the motifs at PoLV1 and PoLV2 were used as a basis for classifying the sequence tags in six groups.

3.5 Mapping of DBLα sequence tags of field isolates from the two malaria sites into a social network

3.5.1 Sequences used in this study

Sequence data GenBank KP085750-KP087726 was used in this analysis. Briefly, 1784 *var* sequences isolated from genomic DNA of 27 field isolates were sequenced. The sequencing was done by 454-Sequencing, Roche, at International Livestock Research Institute (ILRI), Nairobi campus. Of the 27 isolates, 23 were collected from Mbita while 4 were from Tiwi. A total of 1005 and 778 DBLα sequence tags were generated from isolates collected from Mbita and Tiwi

respectfully. Sequences lacking a 5' DIGDI and 3' PQFLR or PQYLR were excluded from analysis and hence were not included in construction of the networks (Fig 3.1).

Figure3.1: A screen shot of sequence tags used to construct networks

3.5.2 Construction of networks

Before construction of networks an Excel spreadsheet (Microsoft) was developed. It was then used to extract four blocks of amino acids from specific windows of DBLa sequences defined by three anchor points (Bull *et.*al, 2008). The following PSPBs were set as default: the 5′amino acid of PSPB1 was set 15 amino acids from the 5′ end of the tag region; the 3′end of PSPB2 was fixed 5 amino acids 5′ to the conserved central WW motif; the 5′ end of PSPB3 was fixed at 13 amino acids 3′to the central WW motif; the 3′end of PSPB4 was fixed 13 amino acids from the 5′end of the tag region. Determination of sequences with shared PSPBs and formatting of the information for import into network analysis package was done using Excel spreadsheet functions.

3.5.3 Visualization of networks

Pajek software was used to draw and visualize networks. Kamada Kawai algorithm (Kamada *et al,* 1989) and the Fruchterman Reingold algorithm (Fruchterman and Reingold, 1991) within Pajek were used to draw 2D and 3D networks respectively (Kamada and Kawai, 1989). Within the network each *var* sequence was represented by vertex with an edge being formed between two vertices that shared one or more PSPBs region. During this analysis no weighting was given to edges with respect to the number of PSPBs shared. Visualization of the divisions of the sequences into cys/PoLV groups and block sharing groups was achieved through formatting the data as Pajek partition files. The data was formatted as Pajek vector files so as to be visualized.

3.5.4 Cys/PoLV sequence grouping

Sequences were initially classified using positions of limited variability (PoLV) based on the Bull *et al* system (2008). Features used to group the sequences into one of six 'cys/PoLV groups' included PoLV1 motif, the PoLV2 motif and the number of cysteine residues within the tag sequence. Group 1 had MFK motif at PoLV1 and 2 cysteine residues, group 2 had a REY motif plus two cysteine residues, group3 had two 2 cysteine residues but lacked MEK or REY motifs at PoLV1 and 2 respectively, group 4 consisted of sequence tags with four cysteine residues but lacked REY motif at PoLV2, group 5 were sequence tags with four cysteine residues and REY motif at PoLV2 while group 6 consisted of sequences with 0, 3, 5 or 6 cysteine residues.

3.5.5 Searching for PSPBs within the sequences collected worldwide

The 14 aa PSPBs from block-sharing group 1 and 2 genes were used to search Fasta files of sequences for hits to any of the PSPBs associated with that block-sharing group. To test for overlap in genes containing 14 aa PSPBs from block-sharing groups 1 and 2, the number of

cys/PoLV group 2 genes were counted from the *var* network that matched PSPBs from blocks haring group 1 only, the number that matched PSPBs from block-sharing group 2 only, the number that matched PSPBs from both block-sharing group 1 and 2 and the number that did not match any. These numbers were expressed as a 2 x 2 table, and Fisher's two-sided exact test was used to determine whether there were less sequences that matched both block-sharing group 1 and 2 PSPBs than would be expected by chance.

3.5.6 Global sequence alignment and tree construction

Sequences were aligned using multiple sequence alignment with high accuracy and high throughput MUSCLE(Edgar, 2004) using default parameters. Neighbour-joining trees were constructed using MEGA3.1 (Kumar *et al*, 2004). Alignments were visualized using Genedoc (http://www.nrbsc.org/gfx/genedoc/index.html).

3.6 *PFEMP1* PCR product polymorphisms and clinical presentation of *P. falciparum* malaria

A reaction volume of 25µl was used to perform PCR.. Each PCR tube had a reaction mixture consisting of 8.86ul ddH_2O, 3.00µl of PCR buffer without Mg2+, (Roche), 2.80µl of MgCl2, 3µl of dNTPs, 1µl of both the DBLα-F (Forward: GACCGGCGACATTATAAGAGG) and DBLα-R (Reverse: TCGCAGGTATTGTGGCACGTAGTC) primers (Ozarkar *et al*, 2007), 0.24µl of Taq Polymerase, and 5ul and DNA template from each sample. The PCR process consisted of an initial denaturation temperature of 94^0c for 2minute, followed by 30 cycles consisting of a denaturation temperature of 94^0c for 1minute, annealing temperature of 51^0c for 2minutes and an extension temperature of 70^0c for 2minutes. This was followed with a final extension of 10minutes at 70^0c and the reaction halted at 4^0c.The PCR products were then separated by gel electrophoresis on 2% agrose gel at 80volts for 30 minutes and visualized on UV trans-illuminator.

CHAPTER FOUR
RESULTS

4.1 POLYMORPHISMS AT POSITIONS OF LIMITED VARIATION (POLVS) IN DBLα SEQUENCE TAGS

4.1.1 Analysis of polymorphisms in DBLα sequences from mbita samples based on positions of limited variation

4.1.1.1 Screening for 5' polymorphisms

All DBLα sequence tags were included in analysis by aligning them in Mega 5.2 and then screening for the motif present at the 5' DIGDI motif. The polymorphisms were identified through a quick search for 5' DIGDI. A total of 1198, sequence tags, both contiq sequence and singlets (those that appeared only once within the 454 reads from a given clinical isolate) were isolated from Mbita field isolates. After aligning in Clustal W, the5' DIGDI motif was found on 1134 (94.7%) sequence tags. These motifs were further separated on the basis of whether they were found on contiq or singlet DBLα tag. The identified polymorphic motifs in this 5' region were as shown in the table (4.1) below:

Table 4.1: 5' Motifs present in sequence tags from Mbita isolates

Polymorphic Motif	No. of contiq tags motif	No. of Singlet tags with motif
DLGDI	1	
DIGDV	7	
IFGDI	19	6
DIGNI		1
DIGDM		1
DIGGI	1	
DLGDI	7	
DIFGI		2
FLGDI	2	
YIGDI		1
NIGDI		1
DIGEY		1
DLVDI		1
XIGDI		1
ICGDI	1	
DIADI	3	

Three sequence tags were shorter than the rest of the sequence tags and could not align properly at the 5' end, hence they lacked corresponding sequences at this region. There were 43 (3.59%) sequence tags without the5' DIGDI consensus sequences. Only one of the unique sequence tags, **IFGDI**, was present on both contiq and singlet tags. The rest appeared on either a contiq or singlet but not both sequence types. DLGDI, DIGDV, DIGGI, ICGDI and DIADI were found on contiq sequence tags only while DIGNI, DIGDM, DIFDI, YIGDI, NIGDI, DIGEY, DLVDI and XIGDI were uniquely present on singlets.

4.1.1.2 PoLV1 polymorphism in non- 5'DIGDI sequences

The results of screening sequences lacking DIGDI motif showed that sequence tags with 5' DLGDI had LFL motif at PoLV1. Sequence tags bearing 5' DIGDV possessed LYL or MFK motifs at this position. Sequence tags with 5' IFGDV had either IYE, LKK, LST, ELK, KYG, LLT or a deletion at PoLV1. None had MFK motif. Majority of the deletions were on singlets. It was also noted that some sequence tags possessed YNDL, YKDV or YNEL insert immediately

after the 5'IFGDI sequence. All such sequence tags had a deletion in the PoLV1. The singlet beginning with 5' DIGNI had LYG at PoLV1, while that with 5' DIGGI had a MFK motif and the one with DIGDM had LYR. Contiq sequence tags with 5' DLGDI possessed either LYL or LFL at PoLV1. Singlets bearing 5' DIFGDI had LYD motif at PoLV1. Sequence tags with a 5' DLGDI had either LYL or LFL at their PoLV1. Two contiqs bearing a 5' DIGGI LYR while one had MFK within the PoLV1. Two singlets had a 5' FLGDI sequence. Both of them had a YNEV insertion next to FLGDI and a deletion of PoLV1. A singlet bearing 5' YIGDI had LYLwhile that with NIGDI had LYL motifs at PoLV1 instead of MFK. A singlet with 5'NIGDI carried LYL at PoLV1. The other singlets with non 5' DIGDI had various motif in their PoLV1 as follows: 5 DIGEY had ICM, 5' DLVDI had AFT and 5' XIGDI had IF followed with an insertion sequence, SLVX and then K. A contiq with 5'ICGDI had LFI at PoLV1. Three other contiqs had 5' DIADI. Two of them possessed LYR and one had LYI at PoLV1. 5' DLGDI, DIGDV, DIGGI, ICGDI and DIADI were found on contiq tags only while 5' DIGNI, DIGDM, DIFDI, YIGDI, NIGDI, DIGEY, DLVDI and XIGDI were found on singlets only.

4.1.1.3 PoLV1 polymorphism in sequence tags with 5'DIGDI

A total of 1134 sequence tags contained 5'DIGDI consensus sequence. These sequence tags were then analysed for variations within the PoLVs. The PoLV1 was searched for presence MFK motif and the number of sequence tags bearing this motif recorded. A total of 132 sequence tags had this motif at PoLV1. It was noted the All MK was separated from K on all sequence tags due to an insertion on some sequence tags at this region, causing MF……..K separation at this motif. Other motifs identified at this position were: LYI (on 32 tags), which was followed with a 3' R the either a T or N amino acid. The others included: LFV (2), LFR (35), LYS (257), LFR (31), LYR (257) (the 3' end of this motif had either a G or R so that LYRG=164 tags and LYRR=93 tags), LFY (5), LFL (152), LFH (2), LYL (229), IFS (2), LFI (50), MFL (12), LFC (2), LYE (2),

MLK (1), LYG (41), LSL (1), LYD (1), FYL (1), LYS (3), PFY (6), LFN (3), LYV (5), LYI (56), LFC (2), MFR (5), IYL (1 singlet), MLL (1 contiq), SIL (1contiq), and MYV (4).

A contig in sequence tags screened for 5' DIGDI and 3' PQYLR was found to contain MFK motif and a REY motif. While the REY motif was situated at PoLV2, the MFK motif was found several amino acids downstream the PoLV1 position. Thus this observation was not contradictory to findings from other studies which suggest that PoLV1 MFK and PoLV2 REY motifs were mutually exclusive, implying that both could not be found on the same DBLα sequence tag. It could thus be suggested that it is possible for the two motifs to co-exist on the same DBLα sequence tag for as long as the MFK motif was not at PoLV1. The occurrence was, however, found to be rare, implying the mutual exclusion principle could be the most preferred mode of selection in the parasite genome.

4.1.2 PoLV2 and PoLV3 Polymorphisms

The motifs found at PoLV2 and PoLV3 are shown in figure 4.1 below:

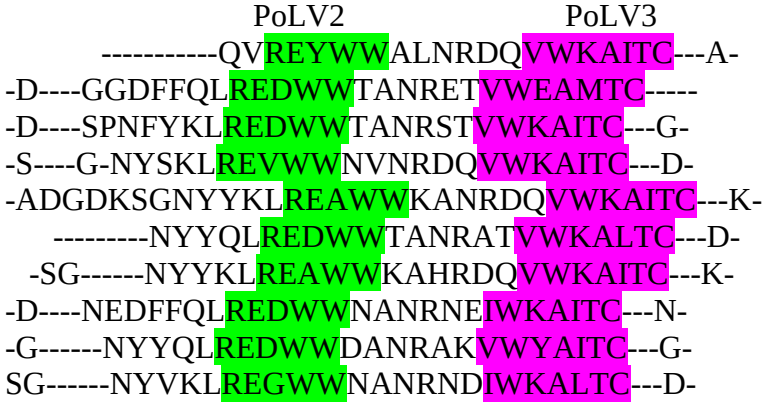


Figure 4.1: A section of DBL alpha sequence showing PoLV2 and PoLV3 motif used to define and classify the sequence tags

A total of 327 sequence tags from Mbita isolates possessed 5' DIGDI and 3' PQFLR that was initially used to define PoLVs and classify DBLα sequence tags into the four groups. At PoLV2, majority of the sequence tags included in the analysis had RED motif. 219 (66.9%) sequence

tags had this RED motif at their PoLV2. Other motifs present at this position were REY, 50 (15.3%); REA, 48 (14.6%). The other motifs were found to occur at a very low frequency within the Mbita isolates. These included RND found on 2 tags, REH on three (3) sequence tags, REG on two (2) tags, REV also on two (2) tags and RET on one (1) sequence tag. However, the numbers of the motifs could be have been different were all 454-reads to be used but there was a possibility that the frequency may not have changed.

On screening PoLV3 position, it was found that 175 (53.5%) sequence tags possessed KAIT motif while the rest, 152 (46.5%) had different motifs within the same position.

4.1.2.1 Variation in PoLV2 and PoLV3 on sequence tags with 3' PQYLR motif

The PoLV2 and PoLV3 motifs from Mbita sequence tags are shown in figure 4.2 below:

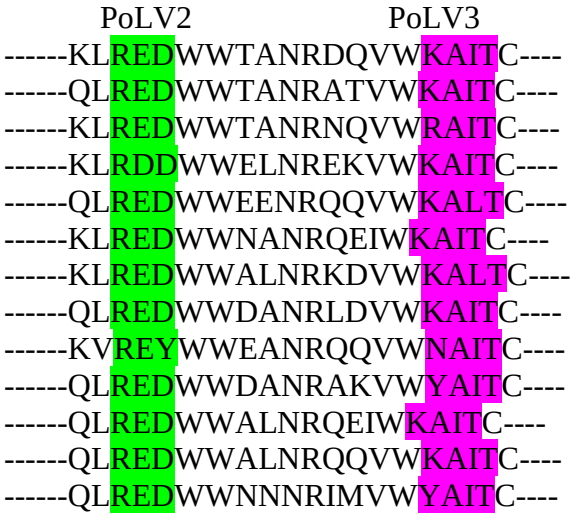


Figure 4.2: Polymorphisms in the PoLV2 and PoLV3 motifs sequence tag bearing PQYLR motif at their 3' ends

The screening of PoLV2 showed that 128 (16.6%) sequence tags out of the 770 tags had REY motif. However the majority of the sequence tags, 556 (72.2%), had RED motif at PoLV2. The rest of the sequence tags (84 or 10.9%) had a different motif other than RED and REY. These

included: RDD, REA, REH, RND, EKI, RNA, G--, GGM, RKD, DIK, GGK, RAD, EKQ, RKD, QED and RET.

4.1.3 Variations at PoLV4

In the 327 sequence tags screened for 3'PQFLR motif, 147 sequence tags had PTYF motif at the PoLV4. A total of 86 sequence tags had PTNL motif at this position. Of these 86 sequence tags, 16 were singlets while 70 were contiqs. The remaining sequence tags had other motifs at PoLV4 apart from PTYF and PTNL. They were P-TYL, P-TFF, L-TNF, P-TNM, --TYF, L-TNL, PLTNL, P-TSF and I-TNL. 18 sequence tags with 3' PQFLR had PLTNL motif at PoLV4, an indication that they had a leucine codon inserted between P (proline) and T (threonine). Sequence tags showing some of the motifs at PoLV4 with 3' PQFLR are as shown in figure 4.3 below:

```
                              PoLV4
----------------------VP-TYFDYVPQFLR
-----------NNG-DVL-TNFDYVPQFLR
-----------------E-GSPLTNLDYVPQFLR
-------------ETD-QVP-TYFDYVPQFLR
---------E-------G-IVP-TNLDYVPQFLR
-------------KEA-PVP-TYLDYVPQFLR
-----------AANI-DPP-TYFDYVPQFLR
-------------SSY-KVP-TYFDYVPQFLR
```

Figure 4.3: A section of DBL alpha a sequences showing polymorphism at PolV4 motifs in sequence tags bearing PQFLR at 3' end

A total of 548 (71.2%) sequence tags had PTFY motif at PoLV4. 56 (7.3%) sequence tags had PTNL motif and 146 sequence tags possessed one of the following motifs: PXYF, PTKL, PTNF, PTYL, PTYI, P-IPQPV-F, PTNM, PSYF, PRYF, LTNF, PTNI, PTNM, LTNF, FLHIF, PTFL, PLTNL, QTYL, STNF, TTYF, SINF, PAYF, PTYSI, PLT-N-L, ATYF, PXYF, QTYL, RTYF, PTSF, STYF, PTDL, PTKL, PHIF, PTIL, PHII-L, PAYF, PK-PI-S, LLQ-N-L, ATNF, TTYF, LTN-L, P-TRSPHIF, PTFF, TTNF, STNL

The PoLV4 of sequence tags with 3' PQYLR are shown in figure 4.4 below:

PoLV4

----------NKG-SVP-T---Y-LD-YVPQYLR
----------PIY-KVP-T---Y-FD-YVPQYLR
--------AT-------N-DVP-T---Y-ID-YVPQYLR
----------NVT-GVP-T---N-FD-YVPQYLR
----------INR-DPP-T---Y-FG-YVPQYLR
----------INR-DPP-T---Y-FD-YVPQYLR
----------LIG-DAP-T---Y-FD-YVPQYLR
-ETKITAGVATI-QIP-IPQPV-FD-YVPQYLR
----------FSG-DPP-T---N-MD-YVPQYLR
-----INET-------D-QVP-T---Y-FD-YVPQYLR

Figure 4.4: A section of DBL alpha sequence tags showing polymorphisms at PoLV4 in sequence tags with PQYLR motif at 3' end

4.1.4 Screening for 3' variations in sequence tags

A total of 327 DBLα sequence tags possessed 3' PQFLR while 3' PQYLR was found on 770 sequence tags. However 82 sequence tags (6.85%) did not possess either of the motifs.

4.1.5 Analysis of DBLα Sequences from Tiwi Isolates Based PoLVs

A total of 778 PfMEP1, DBLα sequence tags were isolated from the 454 reads from Tiwi field isolates.

4.1.5.1 Screening for 5' DIGDI motif in sequence tags from Tiwi isolates

This screening procedure showed that 477 (61.30%) sequence tags had 5' DIGDI motif at this position. Two sequence tags did not have any motif at their 5' end while 18 sequence tags had other motifs. These motifs were DIGDV, IFGDI, FIGLI and LIGDL.

4.1.5.2 Screening for 3' Motifs in sequence tags from Tiwi isolates

This showed that 177 sequence tags had 3' PQFLR, 278 contained 3' PQYLR, 12 had PQHLR while 23 sequence tags had different 3' motifs other those stated above. Out of the 23 sequence tags with "other" motifs in the 3' regions, 19 tags were singlets and only 4 tags were from contiqs.

4.1.5.3 PoLV1

A total of 164 DBLα sequences tags isolated from Tiwi isolates had both the 5'DIGDI and 3'PQFLR motifs. Their PoLV1 had either LYL, LYV, LYR, LFL, IFS, LFI, MFK, IFR, LHL, LYR or HYL motifs. Majority of the sequence tags containing PoLV1 MFK motif had either RED (25 tags) or REA (17 tags) motifs at PoLV2. Sequence tags with MFK motif at PoLV1 were further examined to establish the type of motif at their PoLV2. Two sequence tags with PoLV1 MFK motif had REA motif at PoLV2. However in this study, one sequence tag, a singlet, was found to possess both MFK and REY motifs at PoLV1 and PoLV2 respectfully. A total of 278 DBLα sequence tags were left after screening to selects tags with 5' DIGDI and 3' PQYLR. MFK motif at PoLV1 was found to be mutually exclusive with REY motif at PoLV2 and no single sequence was found with both motifs. A total of 29 sequence tags possessed MFK motif at PoLV1. Upstream this position was a conserved D region adjacent to the PoLV1. Downstream, there was a conserved L, F and I codons. L and F were separated by two amino acids same as F and I that were also separated by two amino acids. Majority of these sequence tags had RED motif at PoLV2. Others possessed REA motif while one sequence tag had RND motif.

In sequence tags with 3' PQHLR, the PoLV1 position had MFK, LFV, LFY, IFR, LYR, MFL, LFL and LFI motifs. Thus this position was less conserved compared to PoLV2 and PoLV3 as shown in figure 4.5 below.

```
                PoLV1
DIGDIVRGRDLYLG--NPQE---IKQRQQLENNLKTIFGKIHDD
DIGDIVRGKDLYLGY-DNKE----KNKEKLEQKLKEIFKEIH--
DIGDIVRGKDLFLG--NDKE----KKRKQLKEF-ENNFRKI---
DIGDIVRGKDPYLGYDDEEK----NRRKQLEKNLKEIFTQIY--
DIGDIVRGRDLY----RGGG----RGRKQLEENLQKIFGNIY--
DIGDIIRGKDHYLG--NNQE----K--EKLQENLRQIFQNI---

CNGDKPGDDNPNTDPP-T----YFDYVPQFLA
```

CSDKSKAGNGDVNIVP-X----YFDYVPQFLA
C------------------NGDEPKA---DKQNVDPQPIR
C----QKKT--------AKMTP-T----------GPTYLT
C------------------IGGDVP-T----YFDYVPQFLA
CNGDKPGDDNPNTDPP-T----YFDYVPQFLA
L-------ITTAGVTATSQRP-T----SKMSIPQLFR
C--DKDKGAKDGDQVP-T----YFDYVPQILR

Figure 4.5: DBL alpha sequence tags showing variations at PoLV1 motif

In the 23 DBLα sequence tags with 5' DIGDI and other 3' motif (non PQFLR, PQYLR, PQHLR), PoLV1 had the following motifs: LYL (15), LFL (4), PYL, LYR, HYL and LYG (all had a frequency of one). HYL, LYV and PYL motif were all found on singlet tags. The 5'DIGDI motif was separated from a conserved RG motif by a V/I codon, followed by an R/K codon then a conserved D region upstream but adjacent to PoLV1 region. Two other conserved regions were observed downstream the PoLV1 position before reaching the PoLV2. They had a conserved L and F regions separated by seven amino acids. These regions are shown below (figure 4.6) with the conserved regions highlighted.

5'Motif PoLV1
▲
DIGDIVRGRDLYLG--NPQE---IKQRQQLENNLKTIFGKIHD
DIGDIVRGKDLYLGY-DNKE--KNKEKLEQKLKEIFKEIH-
DIGDIVRGKDLFLG--NDKE----KKRKQLKEF-ENNFRKI--
DIGDIVRGKDPYLGYDDEEK-NRRKQLEKNLKEIFTQIY-
DIGDIVRGRDLY----RGGG------RGRKQLEENLQKIFGNIY-
DIGDIVRGRDLY----GGGG------RRKAELQDNLKEIFKQIH-

Figure 4.6 Conserved L and F region downstream PoLV1 of DBL alpha tags 2.9.4 PolV2

In the sequence tags screened for 5' DIGDI and 3' PQFLR, REA motif was the most prevalent motif at PoLV2, being present in 104 sequence tags out of the 164 tags screened. Other motifs present at this position included: REY (on 26 tags), REA (on 25 tags),REH (on 5 tags), RNA (on 2 tags), GGL (1 tag) and KRS (on 1 tag). All the 26 sequence tags bearing REY motif at PoLV2 were further examined to determine the motifs they carried at PoLV1. These sequence

tags were found to possess the following motifs at their PoLV1: LFL (14), LYL (4), LYR (2), LHL (2), MFK (1), HYI (1), LYI (1) and IFS (1).

The sequence tags screened for 5' DIGDI and 3' PQHLR had only one sequence tag with REA motif at PoLV2. The other eleven sequence tags had RED motif at this position. The upstream region of PoLV2 had a conserved Y codon located twelve amino acids before PoLV2. Six amino acids from this Y region was another conserved N codon located just three amino acids away from a conserved L region that was adjacent to PoLV2. Sequence tags showing PoLV2 and PoLV3 are given below. Conserved regions are highlighted in green, PoLV2 and 3 in red and anchoring points in purple in figure 4.7 below.

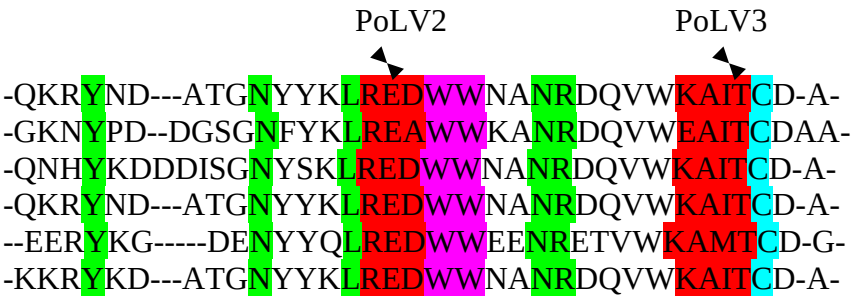


Figure 4.7 A section showing some conserved regions of upstream PoLV2 and between PoLV2 and PoLV3

In the 23 DBLα sequence tags with 5' DIGDI and other 3' motif (non PQFLR, PQYLR, PQHLR), PoLV2 had either RED (17 tags) or REY (6 tags). The PoLV2 was preceded by a Lysine/Valine region on its upstream position. The WW anchoring motif was separated by an NR conserved region by two amino acids. This region was located three amino acids upstream the conserved W region that is adjacent to PoLV3 position.

4.1.5.4 PoLV3 of sequence tags from Mbita isolates

The following motifs were found at PoLV3 of DBLα sequence tags screened for 5' DIGDI and 3' PQFLR were: KAIT, KALT, NSII, KAMI, EALT, EAMT, EAII, NAMI, YAIT, KSLT and

NAIT.In sequence tags screened for 5’ DIGDI and 3’ PQHLR had PoLV3 of nine tags contained KAIT while the remainder had KAMT, KATT or EAIT motif. This region was preceded by a conserved VW region and on its downstream side it has a conserved CD region.

In the 23 DBLα sequence tags with 5’ DIGDI and other 3’ motif (non PQFLR, PQYLR, PQHLR), PoLV3 had KAIT, KALT, EAIT, KALI, EAMT, NAMI, RALT and NAPMI where the proline was an insertion.

4.1.5.5 PoLV4 of sequence tags from Mbita isolates

The following motifs were present at this position: PTYF (70 sequence tags), PTNL (47 sequence tags), PTYL (11 sequence tags), LTNL (6 sequence tags), TTNL (3 sequence tags), LTNF (3 sequence tags), PLTNL (16 sequence tags) AND PLRVR (1 sequence tag). Two motifs, PLTNL and PLRVR had a lysine insertion while one sequence was not included in analysis because of deletions within the first two codons.

The following PoLV4 motifs were found on DBLα sequence tags with 5’ DIGDI and other 3’ motif (non PQFLR, PQYLR, PQHLR): PTYF, VPXY, PKADKQ, P-T-, P-TSK, --T------, P-GNI, P-TNL, T-MQKFYNF, P--------

All sequence tags screened for 5’ DIGDI and 3’ PQHLR had PTNL motif at PoLV4 as shown in figure 4.8 below:

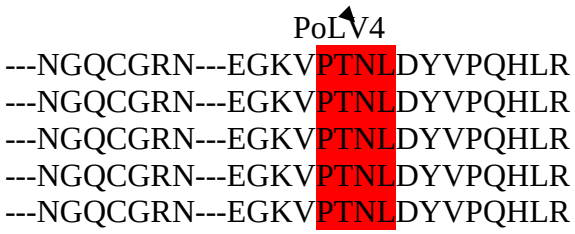


Figure 4.8 DBL alpha sequence tags with PTNL motif at PoLV4 in tags isolated from Tiwi field isolate

These sequence tags were found to be highly conserved in their 3' region with twenty amino acids being totally conserved.

4.2 Grouping of DBL*α* sequence tags from Mbita isolates based on the number of cysteine residues and positions of limited variations (cys/POLVS)

4.2.1 Group 1 DBL*α* sequence tags from Mbita

Sequence tags classified as group1/group1 like were 118 (9.84%). These were cys2 sequences with MFK motif at PoLV1. Majority of these sequences either had REA or RED motifs at PoLV2. A few other sequences had REG, RDD and RND motifs at PoLV2.The PoLV2 motif was preceded by a semi-conserved NYYKL region with YYKL being mostly conserved and N occasionally being substituted with E on some sequence tags. The PoLV3 motifs remained polymorphic in these sequences although KAIT was the most frequent motif at this position. Other motifs present were EAIT, RAIT, RALT and WALT. The PoLV4 position in these sequences had PTNL motif on mast sequence tags. Others had a PTYL motif. However, none of these sequence tags had PTYF motif at PoLV4. This motif was followed with a semi-conserved region consisting of DYV of it 3' side, adjacent to the 3' motif.Some group1 DBLα sequence tags from Mbita isolates are shown in figure 4.9 below:

DIGDIVRGKDMFKPNDKDAVWNGLRAVFKKIHEDLSPEVQKHYEDDGSGNYYKLREAWWNVN
RDKVWDALTCNAPDIGDIVRGKDMFKSNEDVEKGLRAVFGKINNSLTPKAKNHYKDDNGSGN
YYKLREDWWTVNRNQVWEAITCGALPKDIGDIVRGRDMFKSNNDVEKGLDVVFKKIQRKLNG
AAISYYNADEKGNFYKLREDWWMANRDQVWKAITCKAPQKADIGDIVRGRDMFKSNDNVEN
GLKAVFKKINNGLNDKGIRYYDDDGSGNYYKLREAWWTVNRDQVWKAITCGALPKSDIGDIVR
GRDMFKPNEEDAVQKGLREVFKKIKDDLNKNGINDYNDDGPEYYKLREAWWTANRDQVWKA
ITCKAPNGDIGDIVRGIDMFKPNVHDKVETGLREVFKKIHDGMEDEVKNDYNPDGSGNYYKLRE
AWWNVNRNKVWEAITCDASY

DDAEYFLHNSGGLLKFSNSKCHHNQGTVLTNLDYVPQYLR
SAYFMQSEDNKQLFSNPKCGHG--DKDVPTNLDYVPQYLR
NYFRKGSDGSDVFTSQ-GYCGRK---ELTVPTYLDYVPQFLR
AYFMQSEDNKQLFSNP--KCGHS-NGGAPLTNLDYVPQFLR
ANYFRKGPDGSDVFSNS-GPCGRK-EATVPTYLDYVPQFLR
KSGYFMQSESNTPLFSNP-KCGHK-QGKVPTNLDYVPQYLR

Figure 4.9: some group 1 DBL alpha sequence from Mbita isolates

Sequence tags categorized as group2/group2-like sequences containing two cysteine residues and REY motif at PoLV2were 65 (5.4%). Unlike group1 sequence, these group2 sequences did not have the semi-conserved NYYKL region just upstream PoLV2. Most of the sequence tags had LFL or LYL motif at PoLV1 although a few sequence tags had MFL, LHL and FFL motifs at this position. The 5' end of these sequence tags remained highly conserved with the 5'DIGDI motif being followed with either isoleucine (I) or valine (V) codon then a relatively conserved RGKD motif on the downstream side of the 5' DIGDI motif, but just upstream the PoLV1 position. This semi-conserved region was also observed in group1 sequences. Two of these sequence tags did not have this RGKD motif. Instead, one sequence tag had RGRM and another one had RGQK motif. The PoLV3 of these sequence tags remained polymorphic; show no bias towards any particular motif and the motifs present at this position varying with the field isolate from which they were isolated. PoLV4 had PTNL motif on most of the sequences. The variation at this position also depended on the field isolated from which the sequence originated. The 3' motif of the group2/group2-like sequences had either PQFLR or PQYLR motifs, although sequence tags from one isolate ha PQHLR motif at their 3' end. It was also noted that group2/group2-like sequences varied greatly in length depending on the field isolate samples from which they were isolated. Group2/group2-like sequences are as shown in figure 4.10:

5' Motif PoLV1 PoLV2 PoLV3
DIGDIVRGKDLFLGAPNKEKIKLEENLKKIFDNIKNENAELSKLSLEKVREYWWAIHRKELWEAL
TCSAPKGANYFVYKLDDIGDIVRGKDLYLDREKVKSRLEARLKTMFNNIKKSSKNLSHLSIHEVR
EYWWALNRDQVWKAITCKAEEDDIYRKTTDNDIGDIVRGKDLYLGTNEEKKPLEENLKNLFKR
LYEEIKDEQKSSYDDDGTGNYFKLREYWWNSNRNDVWKAITCGAEDDDIGDIVRGKDLYLVDQ
ETKQHLENRLESMFQNIQYNNTELKNIPLHELREYWWALNRVQVWNAITCKAKEGDIYSKTTDI
GDIVRGKDLYLVDQETKQHLENRLESMFQNIQYNNTELKNIPLHELREYWWALNRVQVWNAIT
CKAKEGDIYSKTT

PoLV4 3' Motif
RP-KFSSD-RCGHNYNGD---PLTNLDYVPQYLR
GKLLLWND-NCGHHENNN-VQTYLDYVPQYLR
DKYTKKKRNGETTQSSHIKCRN-VSDPPTNLDYVPQYLR
DNGKLLLWND-KCGHHVDKDVPTNLDYVPQFLR

DNGKLLLWND-KCGHHVDKDVPTNLDYVPQFLR

Figure 4.10: Group 2/group2-like sequences

Group3/group3-like sequences are cys2 sequence tags that do not have MFK and REY motifs at PoLV1 and PoLV2 respectfully. 133 sequence tags (11.10%) belonged to group3/group3-like category. Like the other cys2 sequences, the tags were highly conserved in the 5' region.

4.2.4 Polymorphisms in Cys4 Sequence tags from Mbita

There were 807 out of the 1198 (67.36%) sequence tags that had four cysteine residues and were therefore classified of cys4 sequences. Cys4 tags were classified into two groups based on presence or absence of REY motif at PoLV2. DBLα sequence tags with four cysteine residues but lacking REY motif at PoLV2 are referred to as group 4 sequences. A total of 695 cys4 DBLα tags (86.12% of the total cys4 sequences) from Mbita isolates were classified as group 4 sequences. On the other hand, 112, representing 13.87% of the cys4 tags from Mbita isolates had REY motif at PoLV2 hence they were classified as group 5 sequences.

4.2.5 Group4/group4-like DBLα sequence tags from Mbita field isolates

A total of 687 sequence tags belonged to this group. This constituted about 57.35% of all the DBLα sequences isolated from Mbita isolates after translation of the 454 sequence reads. The majority of group4 sequences from Mbita isolates, 645 sequence tags, had RED motif at PoLV2. This constituted 93.89% of the total sequence tags. All motifs found at this position and their frequencies are shown in the table 4.2 below:

Table 4.2: PoLV2 Motifs in Group4 Sequence Tags from Mbita Isolates

Motif	Sequence tags with motif	Frequency (%)
RED	645	93.89
RND	14	2.04
REA	5	0.73
RNA	11	1.60
REH	2	0.30
RET	2	0.30
REG	1	0.15
REV	1	0.15
RKD	2	0.30
RDD	2	0.30
RAD	3	0.42

Some of these motifs occurred only once or were found only in one field isolate. However, RND and RNA PoLV2 motifs were found in sequence tags isolated from various Mbita isolates. Twenty eight sequence tags did not have 5' DIGDI, instead they had either IFGDI or DIFGI motifs. Unlike group1/group1-like and group2/group2-like sequences that had a relatively conserved region of NYYKL just upstream PoLV2 motif, group4/group4-like sequences did not have this semi-conserved region of 5' to their PoLV2. The motifs present varied depending on the field isolate the sequence tags were isolated from. However, the lysine (L) codon that lies adjacent to PoLV2 on the 5' side remained conserved in all the sequence tags belonging to this group. These sequence tags remained very polymorphic at PoLV3, with no particular motif being dominant at this position. The PoLV4 motif in group4/group4-like sequences very conserved with most of them having PTYF motif at this position regardless of the 3' motif present in the tag. The 3' motif in these sequences was either PQFLR or PQYLR, although some tags had other motifs like PQXLR and PRFLA. Group4/group-like sequence tags are shown in figure 4.12 below:

5' PoLV1 PoLV2 PoLV3

DIGDIIRGKDLYRSYDKKEKEQRKELEQKLKGIFGKIYKDVTRGNNVALQKRYNDPNGDFYQLR
EDWWALNRKDVWKADIGDIIRGKDLFIGYDEKDRKEKAKLQENLKTIFGKIYEELLISTSGNNKE
VLKTRYKKDEDGNYYQLREDWWTANRETVWDIGDIIRGRDLYLGDRKEKVKLEKNLKEYFQKI
HGGLMNGAQNYYKGDADNNYYKLREDWWTANRETVWKAITCKADDIGDIIRGKDLFYGNTQ
ESAQRKKLDDKLKEIFKEIHDEVTKGRSASPLQARYQGDDNNNYSKLRKDWWTANRETVWKDI
GDIIRGRDLYRGDNREKTKLENNLKKIFGHIYEELKKDQTKKDGAQKHYKDAKGNYYQLRNDW
WEANRQEIWKAITDIGDIVRGRDLYRGNTKEKNRRDQLEQKLIEIFGKIHGGLTTTNAKKGQKSA
KDHYQDDNGGNFFQLREDWWNANRADLGDIIRGKDLYLGYDDEEKERRRKLEEKLKKIFGNIY
NDVTNGGKNMALQTRYGQDGQNFYQLREDWWTANRATVM

PoLV4 3’

TCDAP----HGAQYFRGTCGSNENTAT----QARDKCRCKGDQ--------VPTYFDYVPQFLR
KAITCGAPKESKYFRTTCSDTKGPSVANHYCRCNGDQPGQDKTNTDPPTYFDYVPQYLR
TG----NAYFRTTCSM-----NGIFSQANHYCRCNGDKPGEDKANID---PPTYFDYVPQFLR
ALTCDAPGD----ASYFRTTCS--DRQGG---AQARNKCTCN-NGD-----VPTYFDYVPQYLR
CGTHDG----DTYFRATCG----SS--TG-TGTQVRCRCPKTSDGKANDQ--VPTYFDYVPQYLR
KVWKAMTCSEDLK-NSSYFRVTCSDTH-GSSVAIHYCRCNGDKPDDDRPNTDPPTYFDYVPQFLR
WKAITCDEENK-IKDAQYFRGTCGG-EETATLAKDKCRCTTHV---------VPTYFDYVPQYLR

Figure 4.12: Group4/group4-like sequence tags

4.2.6 Group5/group5-like DBL*α* sequence tags from Mbita field isolates

Group5/group5-like sequence tags are those that posses four cysteine residues (cys4) with REY motif at PoLV2.120 DBLα sequence tags, constituting 10.02% of all sequence tags from Mbita isolates sequenced by 454 sequencing, Roche, were classified as group5/group5-like. Like the group4 sequences, these sequence tags lacked a semi-conserved NYYKL region 5’ but adjacent to PoLV2 motif. They however differed from the group4 sequences in that while the lysine (L) codon next PoLV2 on 5’ side was conserved in group4 sequences, it was not conserved in group5/group5-like sequences. The motifs at PoLV1 position of these sequences remained highly variable. Motifs at this region varied as per the field isolate they were extracted from. The PoLV4 was also polymorphic although majority of the sequence tags were observed to have KAIT motif at this position. Other sequences had other motifs like KALT, EALT, KAMT. The PoLV4 motif was predominantly PTYF that occurred on most of the sequences. PTNL motif was also found at PoLV4 of a few sequences. The level of variation differed with isolate.

4.2.7 Group6/group6-like DBLα sequence tags from Mbita field isolates

Sequence tags with zero, one, three, five or six cysteine residues were classified as group6/group6-like. There were 40 sequence tags that fell into this group. This constituted 3.34% of all the DBLα sequence tags isolated from patient samples collected from Mbita study site. These sequences were highly polymorphic and did not show conservation in any of the regions apart from the WW anchoring motif for PoLV2. Cys3 sequence tags constituted the majority of sequences in this group. Eighteen (18) sequences had three cysteine residues (cys3) and constituted 45% of all group6/group6-like sequences. Cys5 sequences were the next most abundant group6/group6-like sequences, with 13 sequence tags, constituting 32.50% belonging to this group. Cys1 and cys0 were the least frequent, with 5 (1.25%) and 4 (1%) sequence tags being cys1 and cys0 sequences respectfully.

Three group 6 DBLα sequences lacking 5' DIGDI motif were blasted at NCBI using blastp tool. The blast was done against all organisms in protein data bases to find out if these sequences were of human leukocyte or parasite origin. These sequences and their blast results are as shown appendix III.

The sequence shown below was very unique in that it lacked the 5' DIGDI motif.

"ICGDIMSDASALFISITTSIRTCLSKIFITFSCTTSDCIPYFFLLSVHQSSRNLKNLDRHLCS
DLLRRSSSLHYVFSRKFLSNVVVIVLFLVDPYHYNQEIDLFLLQYHLYLQKLR"

The sequence above was unique in that it did not only lack a 5' DIGDI motif, but it also had a cysteine residue within this region. This was not a common occurrence. On blasting using blastp tool against all proteins at NCBI, it mapped only onto three protein sequences. None of these sequences belonged to *P. falciparum.* They all had high e values with very low percent identity. This suggested that there was a possibility that none of the sequences in the data base was similar to this sequence. It could be a unique DBLα sequence found only in one sample from Mbita. Since the proteins it mapped onto had very low

identity to the query and the e value was very high, there were higher chances that the mapping was due to sheer coincidence in that they shared a few amino acids rather than having a phylogenetic relationship. The percent identity ranged from 24% to 39% among all sequences while the query cover ranged from 32% to 50% further suggesting lack of total identity/similarity and reducing chances of shared phylogeny with organisms from which the answer were got. The organisms whose sequences mapped to the query were *Enterococcus faecalis*, *Bdellovibrio bacteriovorus* str. Tiberius and *Schizosaccharomyces japonicas.* *Enterococcus faecalis* is a gram-positive commensal bacterium inhabiting the gastrointestinal tract but can become life threading under hospital conditions with high level antibiotic resistance. *Bdellovibrio bacteriovorus* is an intra-periplasmic predator of diverse Gram-negative bacteria while *S. japonicas* is a fission yeast similar to *Sz. pombe*. Based on both the low query coverage and percent identity together with high e values, it was deduced that this sequence belonged to neither of the organisms it mapped onto. However since the sequence lacked the PoLV motifs characteristic of DBLα sequence tags, it remained difficult to conclude that the sequence belonged to PfEMP1. The blastp results are as shown in appendix IV. The distribution of cys?PoLVs groups is summarized in figure 4.13 below:

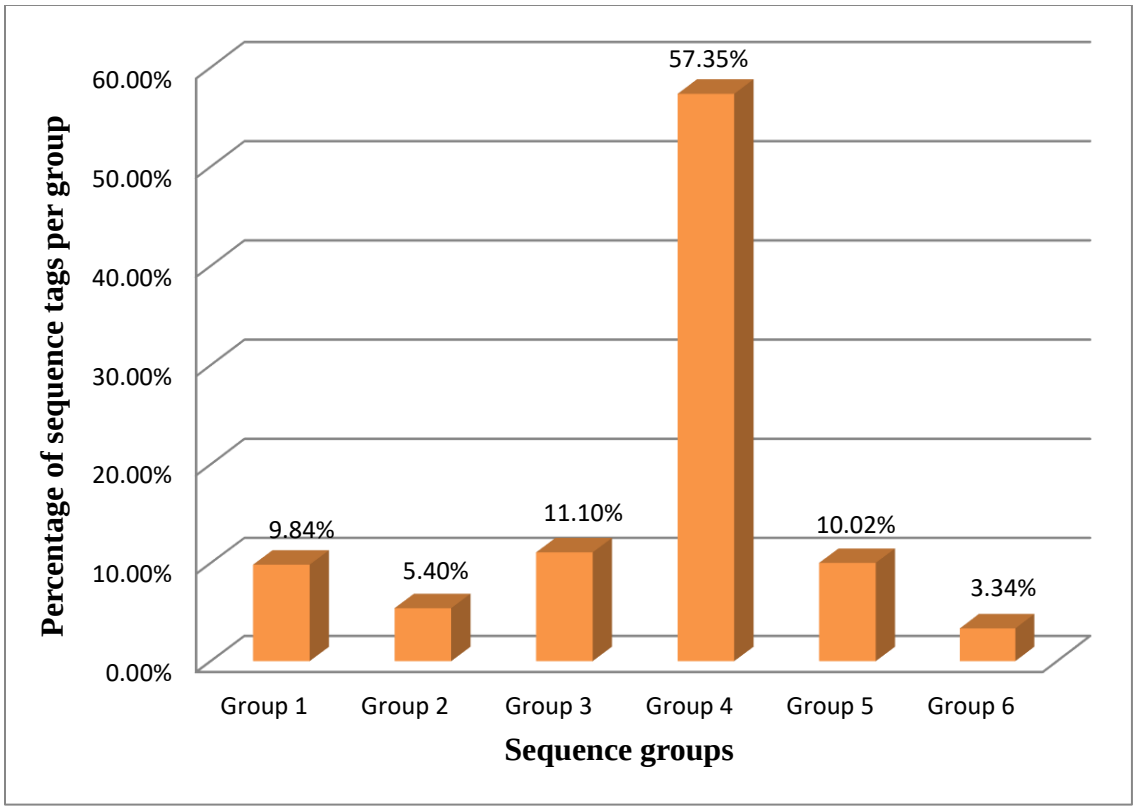


Figure 4.13: Frequency of cys/PoLV group of sequence tags from Mbita Isolates showing DBL*α* sequence groups on X-axis and frequency on Y-axis

4.3. Grouping of DBL*α*sequence tags from Tiwi isolates based on the number of cysteine residues and positions of limited variations (cys/PoLVS)

4.3.1 Group1/group1-like sequence tags from Tiwi samples

Sequence tags containing MFK motif at PoLV1 and lacking REY motif at PoLV2 were classified in as group1/group1-like sequences. A total of 94 sequence tags, constituting of 12.08% sequence tags from Tiwi were classified as group1/group1-like sequence. The 5' motif was majorly DIGDI in most of the sequences. However a few sequences lacked this motif. Some of the motifs substituting 5' DIGDI were DIGDV, IGDIV and NIGDI.

The RED motif at PoLV2 was found on 76% (71) of these sequence tags. The other tags had either REA or REH motifs. 21 sequence group/group1-like tags (22%) had REA motif at PoLV2. REH motif was the least frequent motif at PoLV2 of group1/group-like sequence tags, being

present on only 2 sequence tags. These sequences did not have a semi-conserved YYKL 5' just before the PoLV2 position. The KL motif however remained conserved in all group1/group1-like sequences.

The PoLV3 position was very polymorphic, with sequence tags showing no particular trend of selection or prominent motif at this position. PoLV4 however showed a trend of selection for certain motifs. PTNL was the most frequent motif at this position in group1/group1-like sequences. 64 (68%) sequence tags had this motif. Other motifs at this position included PTYL 15/16%), PLTNL (10/9%), PPTNL (1), ITNL (1), PTNF (2), LTNL (1) and TTNL (1).

The 3' side of Tiwi group1/group1-like sequence tags was predominated by PQFLR motif, being found on 71 sequence tags and constituting 76% of the total tags classified into this group. 18 (19%) sequence tags had PQYLR motif at the 3' side while 4 (4%) and 1 (1%) sequence tags had PQHLR and PPFLT motifs respectfully.

4.3.2 Group2/group2-like sequence tags from Tiwi samples

Group2 sequences are Cys2 sequence tags containing REY motif at PoLV2 but without MFK motif at PoLV2 were classified as group2/group2-like sequences. These sequences had 5' DIGDI motif apart from two sequence tags that contained DLGDI. This region was followed with a near conserved RGKD region with only two sequence tags containing RAKI and RGQK motifs on each sequence tags. The PoLV3 motifs were highly polymorphic. The PTNL motif was found on all but one sequence tags at their PoLV4. The sequence tag lacking PTNL motif had PTRSHL indicating an insertion at this region.

One cys2 sequence tag had MFK and REY motifs at PoLV1 and PoLV2 respectively. This sequence was not placed in any of the groups. This observation was also contrary to the expected

observation which presumes that MFK and REY motifs at PoLV1 and PoLV2, respectively, were mutually exclusive. The sequence mapped onto erythrocyte membrane protein of *Plasmodium falciparum*, when blasted at NCBI. However, none of the sequences had both MFK and REY motifs at PoLV1 and PoLV2 respectfully. The sequences had REY motif at PoLV2 but lacked MFK motif at PoLV1, implying that it was mapping to group2 sequences. It also implied that these motifs were still mutually exclusive and this could be a unique observation where a few sequences may be an exception given that the said sequence had 5' DIGDI and 3' PQFLR.

This sequence tag was also blasted against the human genome in order to confirm that it was from *P. falciparum* and not human (host) DNA. There was only a 40% identity with e-value of 8.2 to the PH and SEC 7domain containing protein 4 isoform X4 and also isoform X3, accession numbers XP 005265694.1 and XP 005265693.1 (Homo sapiens). These results proved that the sequence tag did not originate from human genomic DNA since it had very low similarity to a very small section of the Homo sapiens protein onto which it mapped.

4.3.3 Group3/group3-like sequence tags from Tiwi samples

Sequence tags containing two cysteine residues but lacking MFK and REY motifs at PoLV1 and PoLV2 respectfully were classified as group3/group3-like sequences. 85 sequence tags, an equivalent of 10.93% were classified as group3/group3-likes sequences.

All group3/group3-like sequences, apart from one had 5' DIGDI motif. Downstream this motif was a semi-conserved region consisting of RGRD motif, adjacent to PoLV1 motifs on the 5'side. The PoLV1 position was very polymorphic. This sequence was blasted at NCBI and it aligned to PfEMP1 sequence tags but the identity remained at a low percentage. The data base sequence with highest identity was 63% to three sequences from the database. The RED motif was the

most prevalent motif at PoLV2 of these group3/group3-like sequences. Other motifs at this position occurred in very low frequencies. They included RND, REA, REH and RNA. Two sequence tags included in this group did not have the WW motif that is used as the anchorage point for PoLV2 and PoLV3. These sequences did not have well defined PoLV2 motifs. These sequences however had 5' DIGDI and 3' PQFLR motifs that were used to define DBLα sequence tags. Since they had two cysteine residues and lacked MFK motif at PoLV1, they were classified as group3 sequences. These sequences and their blast results are shown in appendix V.

4.3.4 Group4/group4-like sequence tags from Tiwi samples

The sequence tags containing four cysteine residues (cys4 sequences) that lacked REY motif at PoLV2 position were classified as group2/group2-like sequences. 444 (57.07%) sequence DBLα tags from Tiwi isolates were grouped as group4/group4-like sequences. This group contained the majority of the sequence tags extracted from Tiwi field isolates.

4.3.5 Group5/group5-like sequence tags

The DBLα sequence tags containing four cysteine residues with REY motif at PoLV2 were classified as group5/group5-like sequences. A total 87 sequence tags from Tiwi field isolates, an equivalent of 11.18% of all sequence tags from this study site, were classified as group1/group1-like sequences.All sequences apart from the one except one had 5' DIGDI motif.When this sequence was blasted in the NCBI database, it aligned against PfEMP1 sequences in the database. It was however noted the sequences in the data bases had isoleucine amino acid after the second aspartic acid amino acid in their 5' DIGDI motif. The sequence contained 5' DIGDN, suggesting that isoleucine had been substituted by asparagine amino acid. While isoleucine is neutral and non-polar, asparagine is both neutral and polar. This substitution of hydrophobic with hydrophilic amino acid can impact protein function such as its binding capacity.

The PoLV1 of these sequence tags was highly polymorphic, with majority of the sequences having LYL and LFL motifs. The PoLV3 had either KALT, KAIT or NAMT motifs on most of the sequence tags in this group. Their PoLV4 contained mainly PTYF motif although some sequence had either LTNL, PKTL or PTNF motifs. The 3' motif on these sequence tags was either PQFLR or PQYLR.

4.3.6 Group6/group6-like sequence tags from Tiwi samples

This group consisted of DBLα sequence tags containing either one, three, five, six or no cysteine residues. A total of 38sequence tags, representing 4.88% of the total sequence tags isolated from Tiwi field isolates, were classified into this group.37 sequence tags had 5' DIGDI motif. This motif was followed by either valine (on most sequence tags) or isoleucine amino acid residue then RGRD motif that was semi-conserved on most sequence tags. It was absent on the two sequences that lacked 5'DIGDI motif. A third sequence with 5'DIGDI had RGQR in the place of RGRD.All these sequences contained five cysteine residues. One cys3 sequence tag contained two cysteine residues adjacent to each other as shown in the sequence below:

"DIGDIIRGKDLYRGDDKENIKLQNNLKQIFKEIYDKLDGKNGKAKDHYKDESGNFYQL
REDWWTANRETVWKALTCDDRLGVMHIFETHAMIVECCLKLITTAGVTATSQRPTSKM
SIPQLFR"

This sequence aligned to both cys3 and cys4 sequence tags whenblasted at NCBI database and it aligned to other PfEMP1 sequence tags in the data base.

One subject cys5 sequence, however, contained three cysteine residues adjacent to each other. This sequence had only 50% identity with the query sequence (sequence from Tiwi field isolates)at an e-value of 1e-30. Two cysteine residues on the subject sequence were in the same

position as the cysteine residues on the query sequence. The query however did not have CSDSDGRSE segment present on the subject sequence from the database

One cys5 subject sequence tag had two cysteine residues adjacent to each other although the two residues were located at different position from those present in the query. The subject and query sequences had only 48% identity with an E-value of 2e-28. This indicated that there was possibility that two cysteine residues occur adjacent to one another, even though this occurrence was at a very low frequency.

One sequence tag lacking cysteine residues also did not have defined PoLV2 and PoLV3 motifs. The sequence lacked the anchoring WW point that has been used to define the motifs at this position.

"DIGDIVRGRDLYLGNPQEIKQRQQLENNLKTIFGKIYEKLNGAEARYGNDPEFLNYEKI
GGLLIEKQYGKPSHVTLGVIHIFMQRAIEENELKVTAGVTTTSQVATRQMSIPXIFDYVP
QYLR"

When this sequence was blasted at NCBI, it mapped onto variant surface with 76% identity to one of the sequences in the database (accession number: AAC12725.1) and covering 71% of the query. Another subject (accession number: CAC 41296) covered 100% of the query had only 56% identity to the query. The implication is that the sequence is a DBLα tag with a well-defined 5' DIGDI motif. However, the sequences within the database did not have exact identity.

4.4. Cysteine residues, polymorphisms in PoLVS of DBL*α* and patient data

Three hundred and sixteen (316) of the total 1134 (27.86%) sequence tags collected from clinical isolates from Mbita, western Kenya, carried two cysteine residues and were classified as cys2 sequences. The cys2 screening did not strictly select for 5' DIGDI and 3' PQFLR motifs. Rather all DBLα tags were included in the analysis to widen the range of sequence tags considered.

However most of the cys2 sequences had 5' DIGDI with only a few tags containing other motifs at this region. When PoLVs were considered, PoLV1 was quite variable having MFK, LYL, LYF and other motifs. RED, REA or REY motifs were the most likely motifs to be found at PoLV2 within the cys2 sequence tags. Variation at PoLV3 depended on the isolate from which the sequence tags were isolated but KALT, KAIT and EALT were the most frequent motifs found at this position. At PoLV4 majority of the cys2 sequence tags had PTNL motif. Fewer sequence tags however, contained other motifs such as PTYF, PTYL and TTNF. This again depended on the isolate from which the sequence tag was isolated. The 3' region carried PQYFR motif in the majority of the sequence tags. A few sequences had 3' PQFLR or (much fewer) PQHLR. The length of the sequence tags was also polymorphic, with almost each isolate having its own length of cys2 sequences. This length polymorphism could not however be related to the presentation of malaria or patient age. This observation could be attributable to the fact that one of the inclusion criteria was uncomplicated malaria. Thus the real implication of length polymorphism or cys2 sequence tags in isolates could not be established as clinical diagnosis was largely based on presence of fibril disease and/or fever while inclusion was based on microscopy malaria positive. However the fact that the slide read malaria positive says very little about the extent of morbidity and disease severity.

The day zero (when the patients sought medical attention) clinical data available for these patients showed all patients from this site to have low haemoglobin (Hb) readings. Haemoglobin level ranged from 4.9 to 9.6.

The number of cys2 sequences in Mbita isolates was also not affected by the age of the patients, not did show variation with parasitemia in patients from Mbita study site.

The PoLV2 motifs of cys2 sequences bearing MFK at PoLV1were preceded by a region consisting of two threonine (Y) residues followed by lysine (K) and leucine (L) [YYKL followed by PoLV2 motif] that remained relatively conserved. The Y residue showed limited variation in with some (fewer) sequence tags having phenylalanine (F) in place of the first Y (FYKL) and others sequence tags bearing either valine (V), serine (S), phenylalanine (F) or isoleucine (I) (found on list number of tags) [YVKL, YSKL, YFKL or YIKL). However KL remained conserved. It was also observed that if there was a substitution at the first Y, there was no substitution at the second Y. That is none of cys2 sequence tags had both Y residues at this position substituted by the other amino acid residues (only one was substituted per sequence tag). K and L residues remained conserved in all cys2 sequence tags from Mbita isolates. This was not observed in cys2 sequences lacking MFK motif at PoLV1. However, presence KL just before PoLV2 did not seem to imply presence of MFK motif at PoLV1 in cys2 sequence tags since these residues were present on some cys2 sequence tags lacking MFK motif at PoLV1. It could be argued that all cys2 sequence tags bearing PolV1 MFK (therefore belonging to group A *var* genes), have KL just before PoLV2. However this could not be used to classify cys2 sequence with these residues (K and L) as belonging to group A*var* genes. Presence of these residues in non-group A *var* cys2 sequence tags varied within isolate, with some isolates having this (YYKL) region in their DBLα tags and others lacking it completed or having it on only few sequence tags. Some of these sequence tags are shown in figure 4.14 below.

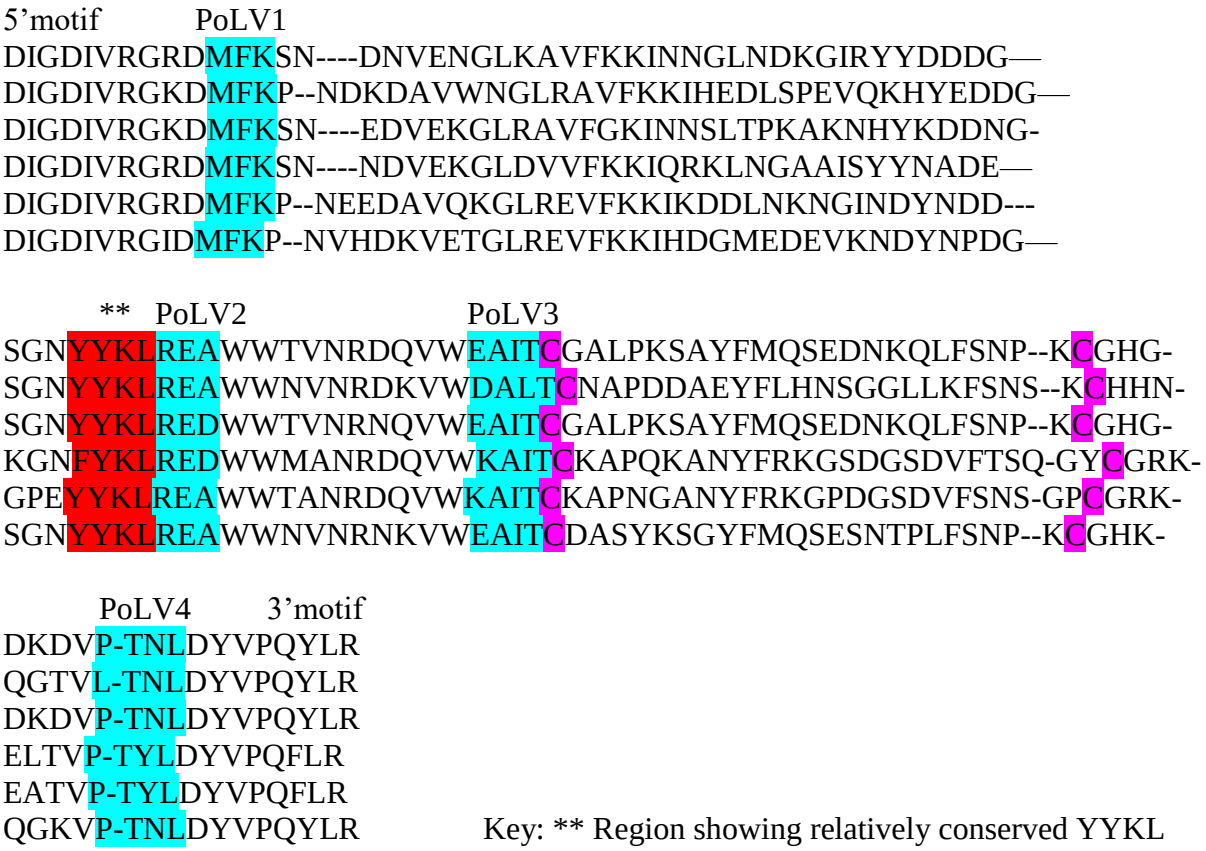
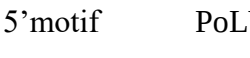


Figure 4.14: Sequence tags with relatively conserved YYKL region

A total of 208 cys2 sequence tags from Mbita isolates lacked MFK motif at PoLV1. Instead most tags contained either LYL, LFR or MFR motifs at this position. These sequence tags were considered to have originated from non group A *var* genes. AT PoLV3, majority of these cys2 sequences had either KAIT or KALT motifs but a few of these sequence tags had NAIT or KAMT motifs. PoLV4 of these tags remained polymorphic, with PTNL motif being the most frequent motif on most sequences. This remained true for cys2 sequence tags bearing MFK motif at PoLV1. Thus there was a high chance of cys2 tags carrying PTNL motif at PoLV4. Some cys2 sequences tags lacking MFK motif at PoLV1 are shown in figure 4.15 below:



DIGDIIRGRDLYGGNNKRRQQLDKKLKEIFGKIHSD-VTTN------GELKRRYKDDP---
DIGDIIRGKDLYFGNN-KKDKLQEQLKEYFKKIYDK-LKDK------EGAKERYKDTA---
DLGDIIRGKDLFLGIEQEKLYLESNLKRIFLKLKES-LHST--------IKSHYENEDE--
DIGDIIRGKDLFLGNNKEKKQLEENLKKIFAKIYNDVMKTSDKNGKKSALQKRYNDTT

PoLV2
EFFKLREDWWNANRQEIWKAITCNDDKKLASASYFRPT-CSDRNGN-
NYYELREDWWTANRETVWKAITCNA----HDSRYRKMG-A----DG-
NFYLLREDWWAVNRKEVWKAITCDAP---EEAKYTKKG-P----TG-EIIESKYHKCRNF--
DYFQLREDWWDANRLDVWKALTCHAP---ESAKYKVIGADG------

PoLV4 3'motif
SQYQNSKTNAGVGTTR---FPXYFDYVPQYLR
SITESTRGQCRDVSG----VPTNFDYVPQYLR
AD---VPTNFDYVPQFVR
STSESPMGKCGNVTG----VPTNFDYVPQYLR

Figure 4.15: Some cys2 sequences tags lacking MFK motif at PoLV1

A total of 210 DBLα sequence tags from Tiwi isolates were cys2 sequences. Most of these sequence tags had 5' DIGDI motif. However this occurrence was natural as there was no screening for this motif prior to select for this motif. MFK, LYL and LFL were the most common motifs at PoLV1 of cys2 sequence tags. A few other sequence tags had other motifs. MFK was the most prevalent PoLV1 motif in sequence tags from Tiwi isolates, with 101(48.1%) of cys2 tags bearing this motif at PoLV1. Cys2 tags from Tiwi isolates were longer that those from Mbita isolates. At PoLV2 majority of cys2 sequence tags from Mbita isolates had RED motif. REH, REY, REV and RND were also present at this position. When sequence tags with PoLV1 MFK motif were selected, none of the sequences had REY PoLV2 motif. However two sequence tags were found to have MFK motif located downstream PoLV1 and REY motif at PoLV2. The implication of this observation could not be determined but perhaps it was indicative that the motifs remain mutually exclusive only when located at PoLV1 and PolV2 respectively. However,one tag from Tiwi isolates had MFK and REY motifs at PoLV1 and PoLV2 respectfully. This rare occurrence could have been a chance occurrence rather than a departure from earlier findings. Alternatively, it could indicate a possibility of the two motifs

existing on same sequence tag under rare conditions at genomic level. These sequence tags are shown below in figure 4.16:

H6S3LLJ04JUXQM_10
DIGDIVRGKDMFKRNEEDAVQKGLRAVFKKIHEGLGTPEKHYKKDEDPNYYKLREYW
WALNRNDVWEALTCSAPGDAKYVKYFPSNTTTVSFDQCGHNDMHVPTNLDYVPQFLR

Contig142_10
DIGDIVRGKDLFLGHKQRKKYLEERLEQMFKNIQGNNSKLRTLSLNHVREYWWNLNRK
EVWKAITCRAEEKDTYFIKSSKGVYSFPNGRCGREDENVPTNLDYVPQYLR

Contig39_10
DIGDIVRGKDLFLGHKQRKKYLEERLEQMFKNIKENNNTELGDLSLAQFREYWWALNR
VQVWKAITCRAEEKDIYSKTTDNGKLLLWNYNCGHHVNKDVPTNLDYVPQFLR

Contig703_10
DIGDIIRGKDLYLDHEPGKQHLEERLEQMFKNIQENNNTELGELITAQVREYWWALNRQ
QVWNAMICKAKERDIYSKTTDNGKLLLWNDKCGHHVNKDVPTNLDYVPQFLR

DIGDIVRGKDLFLGHQQRKRKLEERLVEMFKNIQKKNSRLQDLPLDKVREYWWEANRQ
QVWNAITCDTGENDKYFKKSSTREFKFTGGRCGREDENVPTNLDYVPQYLR

Figure 4.16: DBL alpha sequence tags with MFK and REY motifs

The above sequence tags were isolated from Tiwi field isolates. Blasting results indicated that the MFK and REY motifs were only mutually exclusive when they were located at PoLV1 and PoLV2 respectfully. The MFK motif can also be located downstream the PoLV1 region. However, the MFK motif in all sequence tags was located upstream the PoLV2, regardless of whether it was at PoLV1 or another location. The REY motif was only found at PoLV2 and not at any other position within the sequence tag. The sequence that had MFK and REY motifs at PoLV1 and PoLV2respectfully was only found in a Tiwi isolate. Blast search indicated that there were no sequences from other studies that had these motifs at PoLV1 and PoLV2 respectfully, implying their mutual exclusivity. This was therefore a rare occurrence that could only have occurred by chance.

The distribution of Cys/PoLV groups in DBLα sequence tags isolated from samples from both endmic sites is shown in figure 4.17 below.

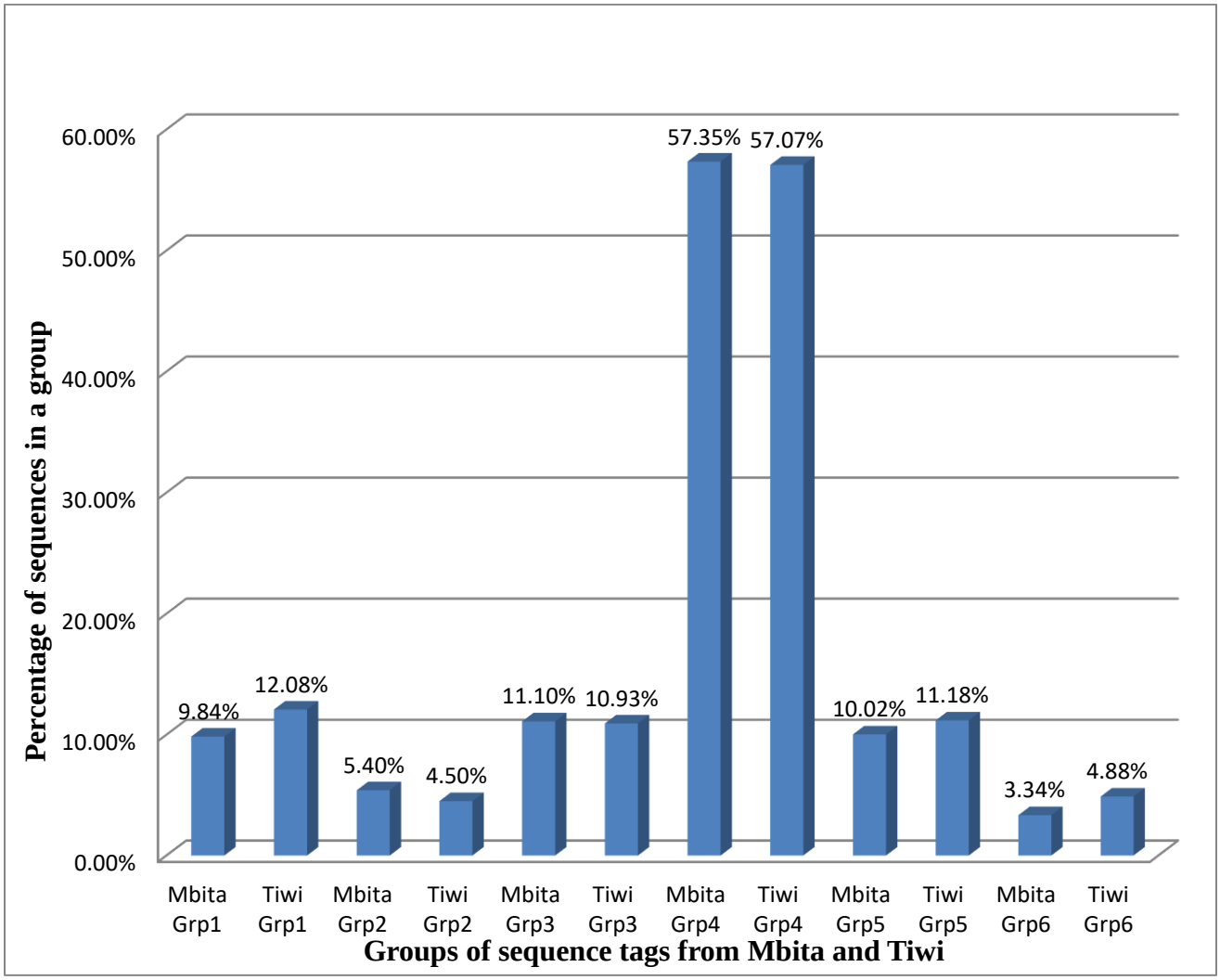


Figure 4.17: Comparison of CYS/PoLV groups from Mbita and Tiwi isolates

4.5 Mapping of DBLα sequence tags of field isolates from the two malaria sites into a social network

The analysis approach was based on a rationale which presupposes that blocks of sequences in highly polymorphic regions can be shared by two sequences that are totally different. Since DBLα sequence tags are highly variable, their alignment requires introduction of gaps. This was overcome by restricting the analysis to ungapped polymorphic sequence blocks locked within

var sequences fixed relative to one of the three conserved anchoring points at each end and in the middle to provide four independent windows. Single sequences were then used within these windows to find out if two sequences were identical within any of the sequence blocks. Thus sequences were analysed as multiple independent blocks, with each block, (that is the 'position specific polymorphic block' [PSPBs]) acting as a genetic marker of the sequence to which it was anchored. The PSBPs were used to construct networks. In the network, the nodes (vertices) representing sequences were joined by lines (edges) if they were identical at one or more of their constituent PSPBs. Giant components from the networks are shown in figure 4.18 below.

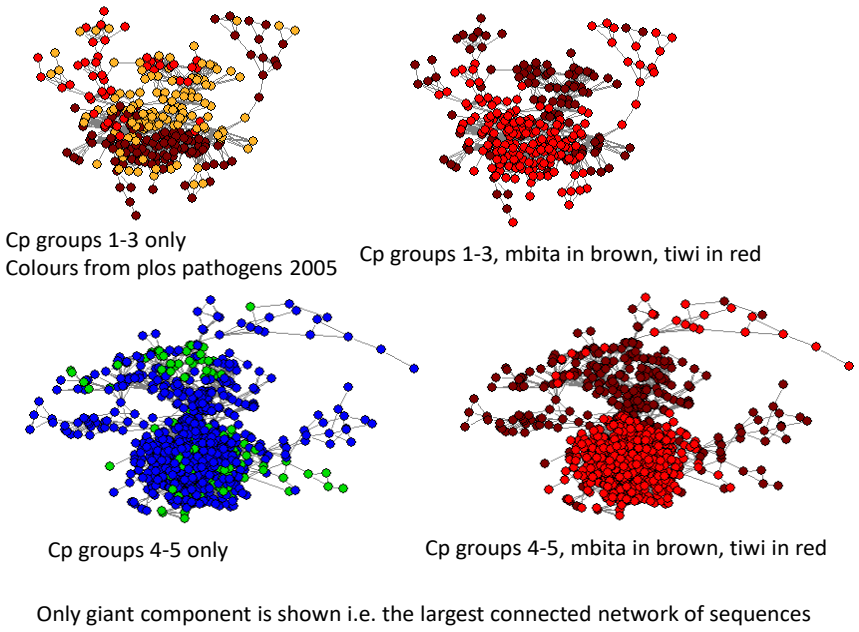


Figure 4.18: Giant Components of group1-5 sequence tags

The cys/PoLV groups have been used to classify DBLα sequence tags into six groups. Figure 4.19 shows the network of sequence groups 1-3 in which each group is represented by a different colour (Brown represents group1; red, group 2 and gold group 3)

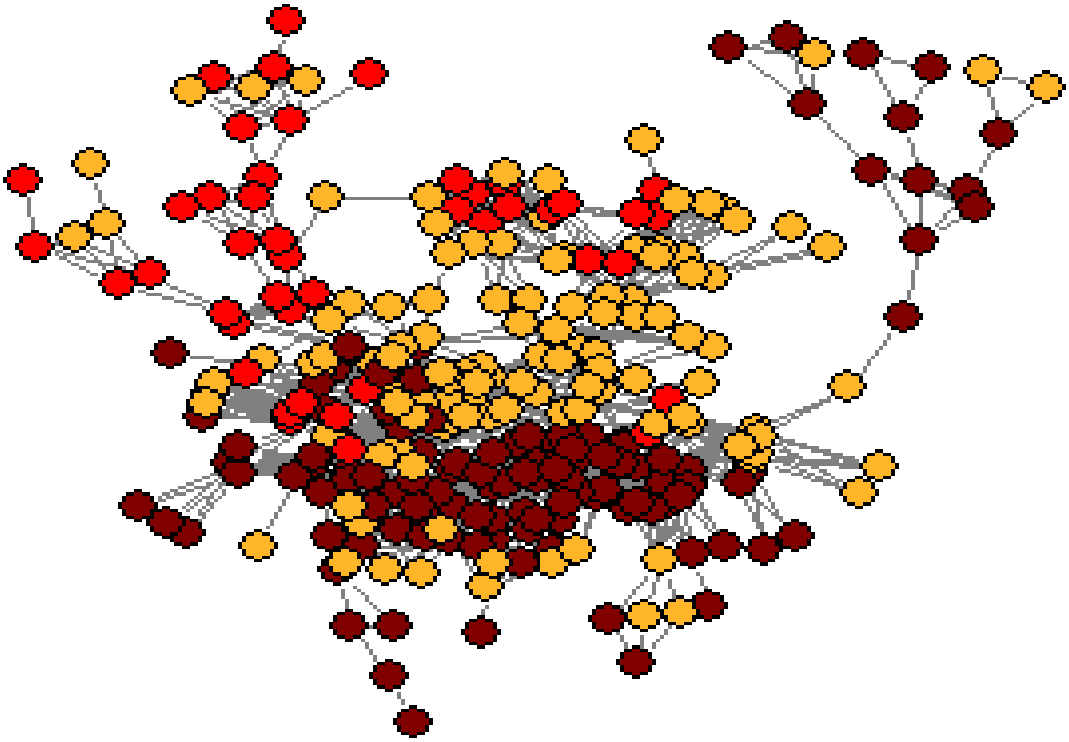


Figure 4.19: Group 1-3 sequence from Mbita and Tiwi isolates (brown=Mbita sequence tags, Red=Tiwi sequence tags)

Some DBLα sequence tags belonging to different cys/PoLV groups were joined by nodes. This suggested shared PSPBs on the sequence tags. Vertices of some group3 sequence tags had nodes linking them to groups 1 and 2 separately without any of these group 3 sequences being linked to groups 1 and 2 simultaneously. However, when groups 1-3 sequences from Mbita and Tiwi isolates were analysed together, they clustered separately into two distinct groups as shown in the figure 1B, in which brown colour represents group1-3 sequence tags from Mbita isolates while red represents group 1-3 sequence tags from Tiwi. These results suggested that parasites circulating in the two study sites have distinct DBLα sequence tags that do not share any PSPBs. Group 4 and group 5 tags from the two endemic sites mapped differently as shown in figure 4.20 below:

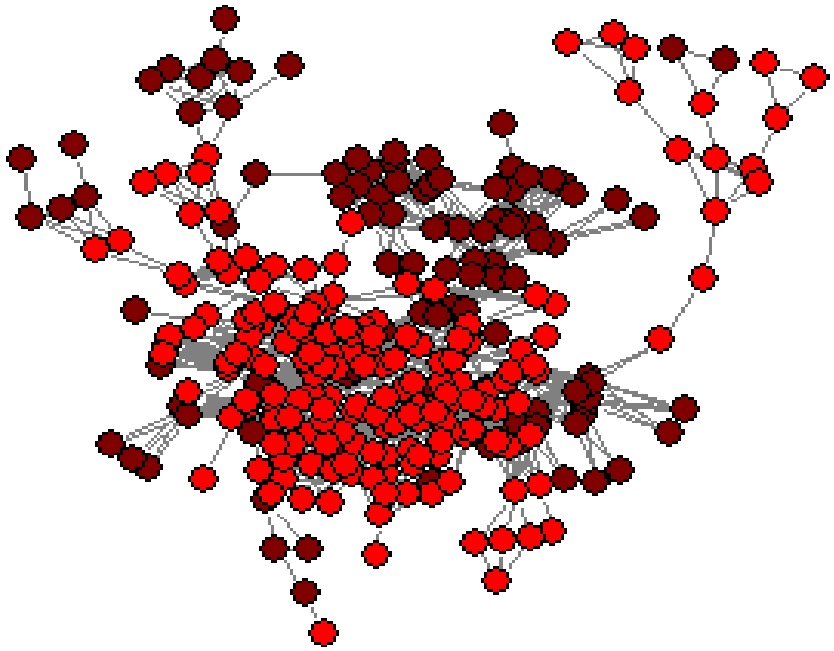


Figure 4.20: Group 4 and group5 from both study site (Mbita=Brown, Tiwi=Red)

In the analysis of group four (blue) and group five (green) sequence tags (figure 4.21) some group four sequence tags linked to group 5 sequences suggesting share PSBPs between these sequences. This suggested that some of group four sequences shared PSBPs.

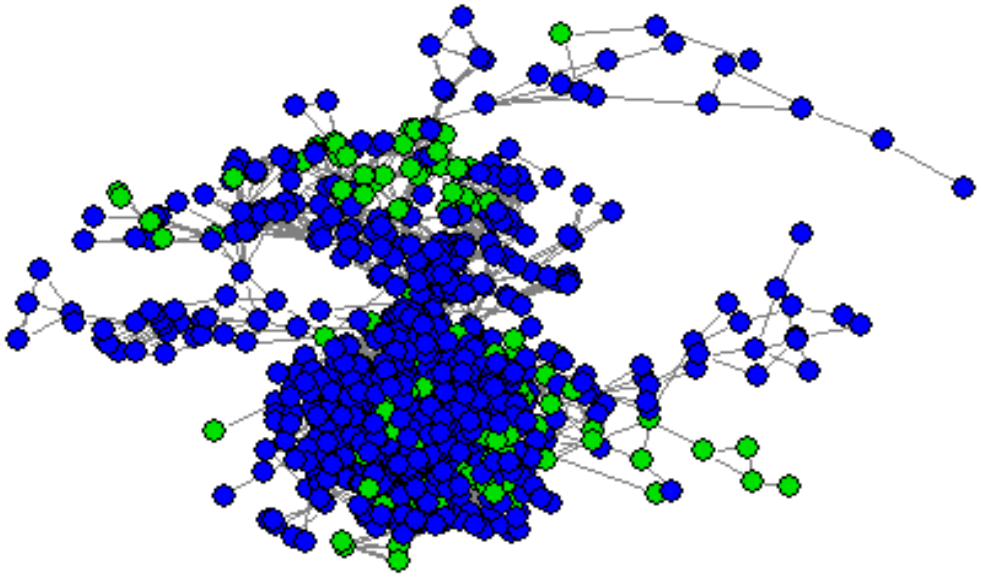


Figure 4.4: Group 4 (blue) and group 5 (green) sequence tags

When group 4 and group 5 sequence tags were analysed together (figure 1D), only a few sequence tags from both sites were linked together (Mbita= brown, Tiwi=red) as shown in figure 4.22 below:

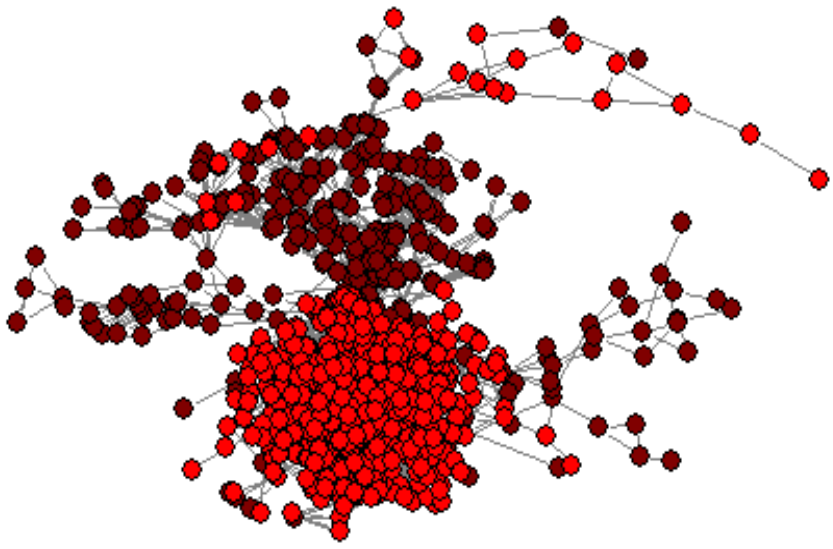


Figure 4.22: Group 4and group 5 from both study site (Mbita=Brown, Tiwi= Red)

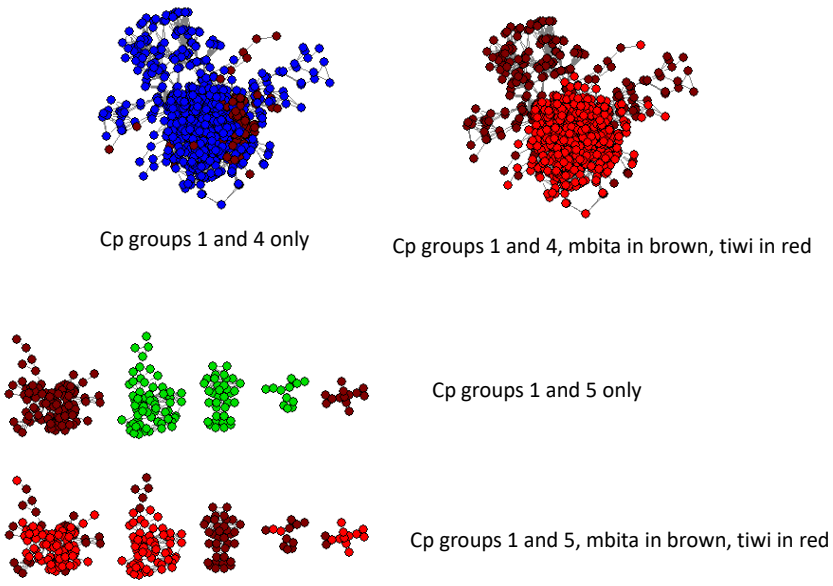


Figure 4.23: Small components showing occasional sharing of PSPB between 1 and 4 but never between 1and 5

Sequences tags belong to group 1 and were also analysed together. The figure (4.24) below shows the giant block of this analysis

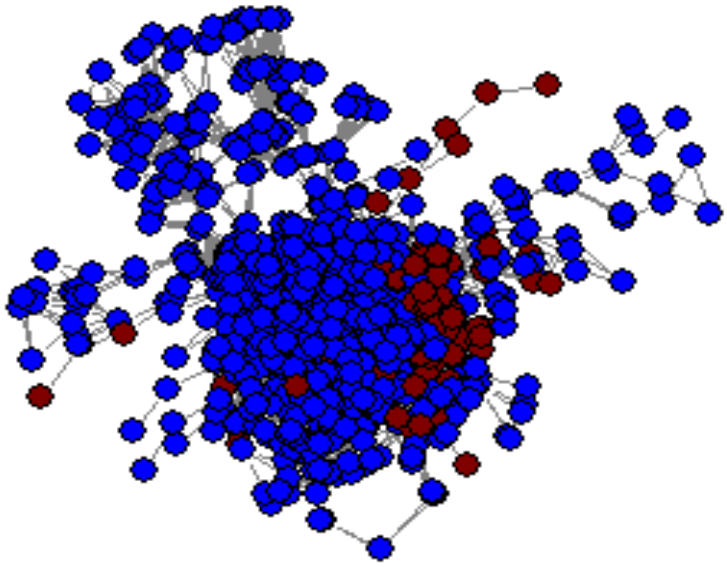


Figure 4.24: a giant component showing occasional sharing of PSPB between 1 and 4(Brown= group 2, Blue = group 4)

The analysis showed group 1sequences (brown) clustered within the group 4 (blue) sequences. It should however be noted that group one sequences were much fewer than group four sequence tags. When these sequences from the two endemic sites were analysed together, the results wereas shown in figure 4.25 below (Mbita=brown and Tiwi=Red):

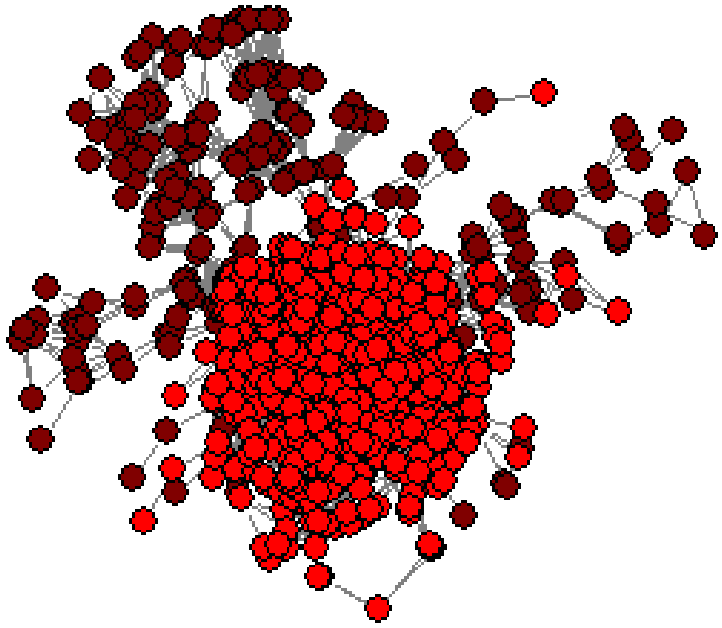


Figure 5: A giant component of group 1 and 4 sequence tags from Mbita (Brown) and Tiwi (Red)

These results indicated that group 1 and group 4 sequence tags did not share similar PSBPs, hence the clear separation between the two groups of the sequence tags. Similarly very few sequences from Tiwi shared similar PSPBs with those from Mbita at genomic level. This data could not clearly indicate if the expressed DBLα sequence tags would cluster in a similar manner.

Figure 4.26 below shows analysis of group 1 (green) and group 4 (Brown) from Mbita isolates. In this analysis, group 1 sequence tags clustered in three distinct sub-groups while group 5 sequence tags clustered in 2 distinct sub-groups. Now of these groups had their sequences clustering together

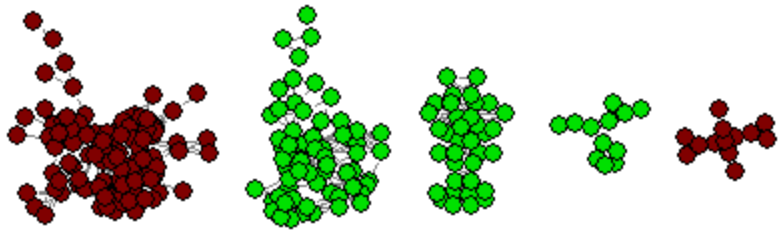


Figure 4.26: A visualization analysis of giant component of group 1 (green) and group 5 (brown)

Figure 4.27A below shows group 1 group 4 sequences only from Mbita (brown) and Tiwi (red), again these sequence tags segmented differently showing lack of similarity at genomic level.

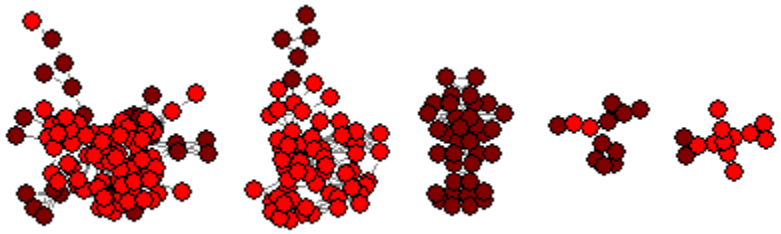


Figure 4.27A: Group 1 and group 4 sequence from Mbita (brown) and Tiwi (Red)

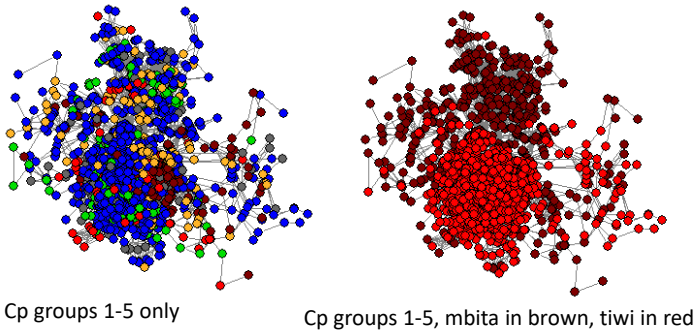


Figure 4.27B Comparison of group 1-5Figure

Another unique observation from analysis of these data is the fact that a few groups 1 and 4 sequence tags from both study sites linked to each other. This indicated that these sequence tags shared PSPBs, an observation that has not been made in previous studies. It could not be established however if these observation would be the same at the expression level. Figure 4.28 below shows these sequence tags linked to each other when the network containing both groups 1 and 4 sequences was blown up for closer observation.

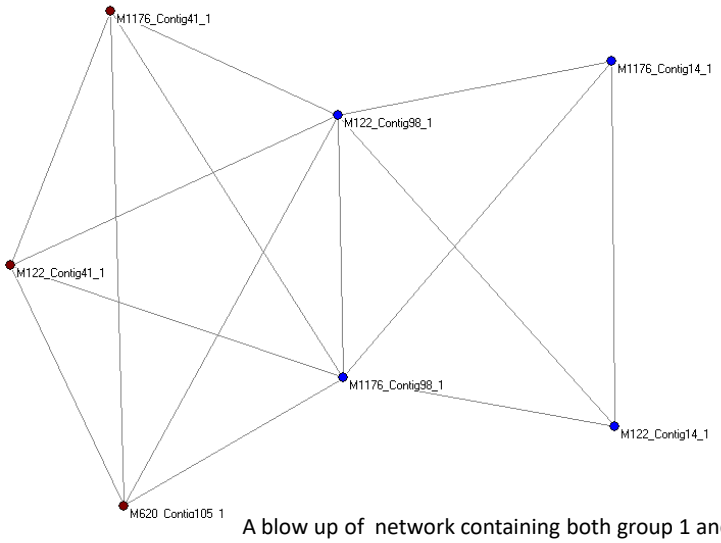


Figure 4.28: An experntion of network of both group 1 and 4 sequence to allow a closer visualization of sequences

4.6. *PFEMP1*PCR product polymorphisms and clinical presentation of P. *falciparum* malaria

A total of 96 samples were amplified by PCR. 60 samples (62.5%) gave a 455bp PCR product (referred to as PCR positive), while 37 (37.5%) did not amplify and were termed to be PCR negative. This product size is one of those found in African laboratory clone 3D7. Mbita samples gave bands only with the F-R primer combinations. None of the samples gave bands with the other primer combinations.

1 2 3 4 5 6 7 8 9 10 11 12

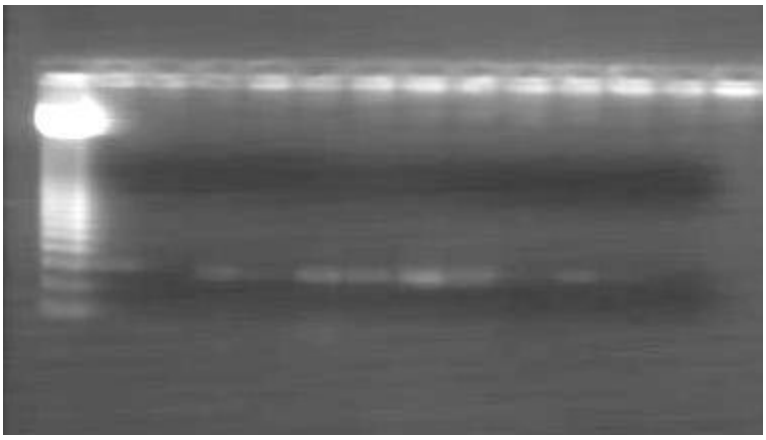


Figure 4.6: Gel photo of PCR amplification using F'R primers

SAMPLE ORDER: (1). 123 BP Ladder, (2). 3D7 (Positive control), (3). K398,(4). K196, (5). K782,(6).K399, (7).K1345, (8).K1176, (9).K1750, (10).K216, (11).K3069, (12). Negative Control

The amplification results were the compared to clinical data of patients from Mbita study site. When the patients' body temperature was considered, it was found that 19 subjects representing 32% of the samples that gave a 455bp band had a body temperature of above 38^0C suggesting high fever. 24 (40%) had body temperature of below 37^0C probably corresponding to the chills that are usually experienced by malaria patients. The majority of those with body temperature above 38^0C had parasitemias above 20,000/µl. Only one subject had parasitemia of 1160/µl. 10 (17%) subjects whose samples gave the amplicon had normal body temperature. Of these, four subjects were aged five years and above while six were below five years of age. Among the subjects that were PCR negative for the primer set, only 3 had a body temperature above 38^0C and 10 had body temperature below 37^0C. 12 had body temperature around 37^0C. Fishers' exact test gave a P-value of 0.074, suggesting a weak significance level.

From the clinical data, only 87 patients had their ages indicated. Among the subjects whose samples were PCR positive for the 455bp product, 47.1% (41 isolates) were aged six years and below while 16.1% (14) were above six years of age. In the PCR negative isolates, 23

representing 26.4% were collected from patients age six years and below while 5 (14%) were from patient aged above six years. These results are shown in figure 4.30 below.

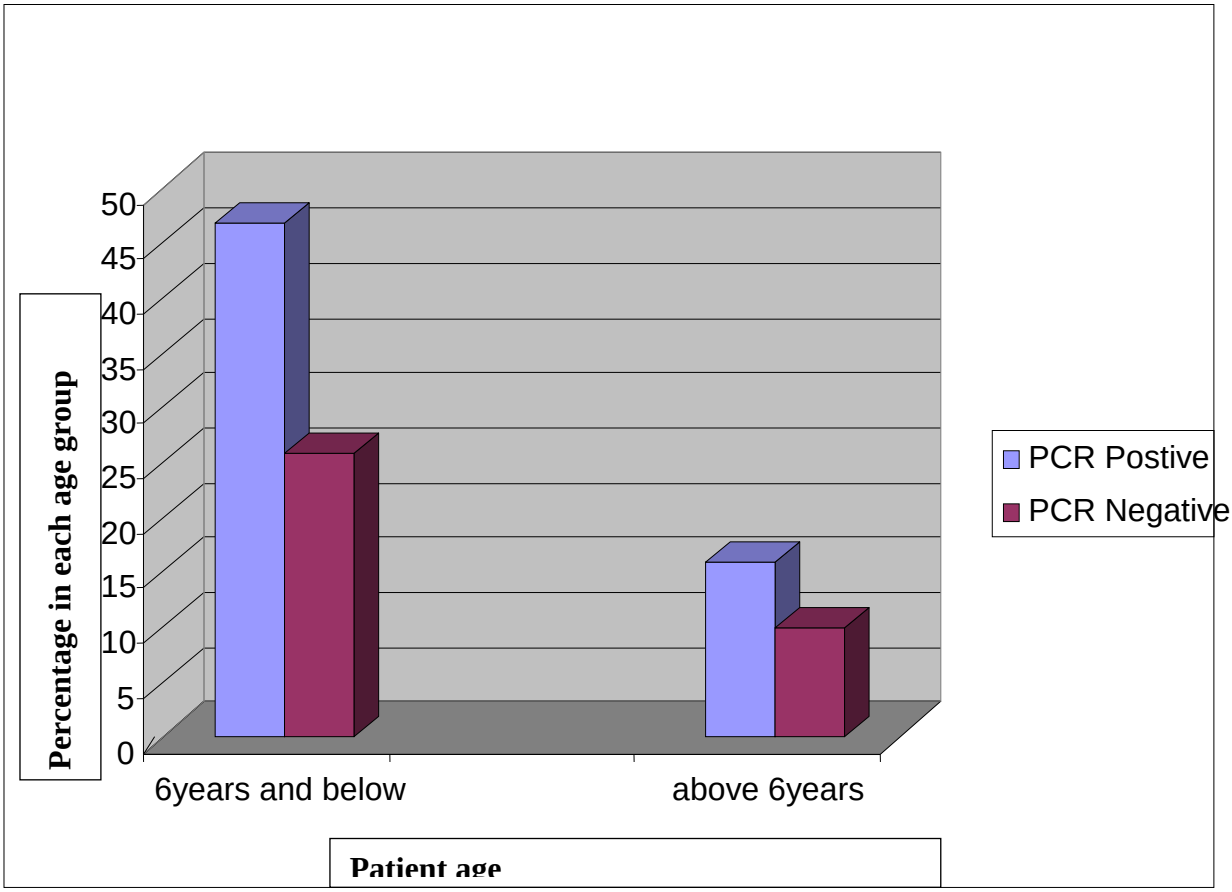


Figure 4.30: Comparison of patient ages among PCR positive and Negative samples

Haemoglobin (Hb) levels were recorded for 35 participants whose samples were analysed. These participants were divided into two groups: those whose Hb was 8 and below and those with Hb above 8. In the first category, 17 (49%) isolates were PCR positive while 10 (29%) were PCR negative. In the second category, only 3 (8.6%) isolates were PCR positive while 5 (14%) were PCR negative. This data is show in figure 4.31 below.

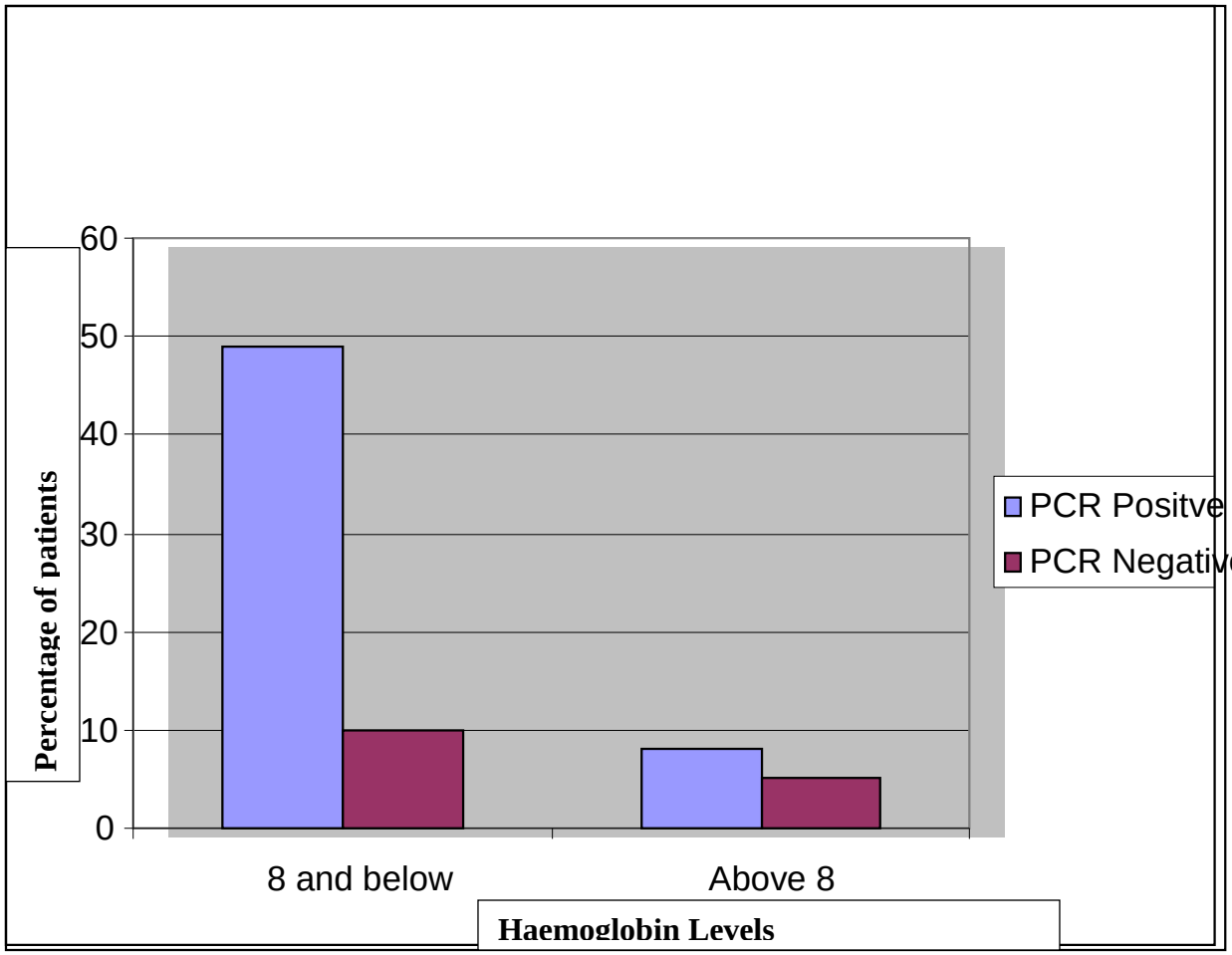


Figure 7: Comparison of patients' haemoglobin (HB) levels among PCR positive /negative samples

Parasite carriage (the approximate number of infected erythrocytes of the total number of erythrocytes in a microscope field) was recorded for 76 patients. 26 (34%) patients had a parasite carriage of 10,000ul and below while the other 50 (66%) patients had a parasitemia above 10,00ul. Among those with a parasite carriage of 10,00ul and below, 17 (22%) were PCR negative while 9 (12%) were PCR positive. Among the patients with a parasite carriage of above 10,000ul, only 11 (15%) were PCR negative. The rest 39 (51%) were PCR positive. This data is represented in the figure 4.33 below:

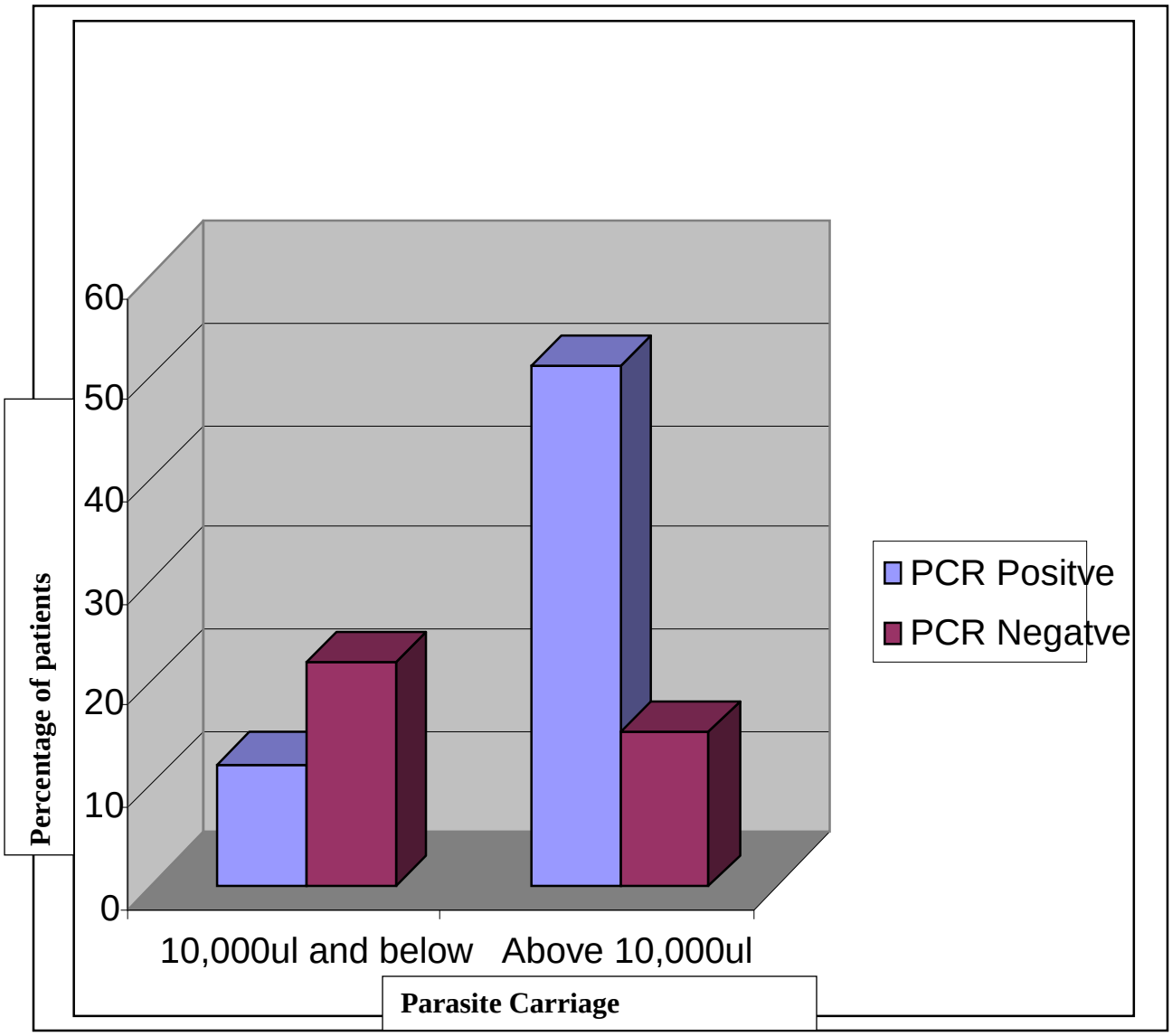


Figure 4.33: comparison of patients' parasite carriage among PCR positive and negative samples

This data indicates that majority of the patients with a parasite carriage above 10,000ul were also positive for the 455bp PCR product of the DBLα domain of *Pfemp*1 genes.

CHAPTER FIVE

DISCUSSION, CONCLUSION AND RECOMMENDATION

5.1 DISCUSSION

Although there has been a reported decline in malaria incidence in some countries (WHO, 2011), the disease still remains a major problem. In 2010, 216 million cases of malaria were recorded worldwide, with 81% of these cases being reported in sub-Saharan Africa. Mortality due to malaria was estimated to have been 655,000 cases in the same year. Reported reduction in malaria cases could be attributed to adoption of ACT as first line treatment in most malaria endemic countries. Increased distribution and use of ITNs has also greatly helped reduce malaria transmission in endemic areas. However, reports from Southern Asia indicating that *P. falciparum* has developed resistance against ACTs threatens the gains made towards reducing malaria burden and it eventual eradication.

Various attempts have been made to develop a potent and viable malaria vaccine. Some vaccine formulations have been subjected to clinical trials with only a few showing promising results. A malaria vaccine that would provide protection to non-immune individuals from non-malarious zones would go a long way in reducing the malaria morbidity burden. One factor that has hampered efforts of developing a viable malaria vaccine is antigenic variation in *P. falciparum* (Plebanski *et al*, 2002). Antigenic variation results from the multiple developmental stages of the parasite within the human host, with each stage expressing its unique antigens that evoke specific immune responses. Expression of polymorphic antigens on the erythrocyte surface also contributes to antigenic variation of the parasite at the erythrocytic stage.

Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) is a parasite protein coded for by *var* genes, that is exported to the erythrocyte surface. Each parasite haploid genome has

slightly over sixty *var* genes, of which only one *var* gene is expressed at a time. The figure below summarizes the impact of these parasite molecules in different organs.

5.2 Polymorphisms at positions of limited variation (PoLVs) in DBLα sequence tags

DBLα sequence tags extracted from genomic DNA of Mbita samples remained highly polymorphic. The functional significance of polymorphisms at the 5' and 3' motifs has not been established. However, these motifs are important in defining DBLα sequence tags. The polymorphic nature of these sequences at genomic level indicates that PfEMP1 proteins expressed by the parasite and exported to the surface of infected erythrocytes would also be very polymorphic thus rendering host immune responses ineffective. This observation also affirms the observation that malaria semi-immunity develops in a peace-meal manner, requiring several exposure episodes to a large repertoire of PfEMP1 surface antigens before developing protective immunity. It would also imply that developing a functional vaccine with universal coverage would remain a challenge. Thus this may, in part, explain why an effective malaria vaccine has not been found despite the fact that there have been decades of research with substantial amount of funding.

5.3 Grouping of DBLα sequence tags from Mbita isolates based on the number of cysteine residues and positions of limited variations (cys/PoLVS)

In the sequence tags isolated from Mbita field isolates, group4 sequences were the most prevalent, accounting for 57.35% of all the sequence tags. This was followed by group3, (11.10%), group5 (10.02%), group1 (9.84%), group2 (5.40%) and group6 had the least frequency (3.34%). Groups3, 4 and 5 DBLα sequence tags are all cys4 sequences that contain four cysteine residues. Cys4 sequences have been associated to uncomplicated and mild malaria symptoms of malaria (Warimwe *et al*, 2009). One of the inclusion criteria of subjects into the study was they had to be positive for malaria but show uncomplicated (non-severe) malaria

symptoms. Parasites causing uncomplicated malaria are more likely to express non-group A *var* genes which are characterized by possessing two cysteine residues (cys2 sequences). Expression of group A *var* genes, that cys/PoLV group1, is characteristic of non-immune patients who end up developing severe symptoms of malaria like cerebral malaria, convulsion and severe aneamia. The study was carried in a hyper-endemic area of malaria. There is therefore a high likely that most patients were semi-immune, having had several exposures to malaria and possessing anti-malarial memory T- and B-cells that limited the symptoms. However the existence of group1 sequence tags indicated that the patients could have developed severe malaria had they not received chemotherapeutic intervention. The fact that they sought treatment could also be an indicator that they felt ill and need medical intervention. It may thus be that this intervention was sought early during infection before the parasite over-came the immune resistance from the host or the host immune mechanisms themselves limited progression of malaria to its severe form.

5.4 Comparision of cys/PoLV groups from Mbita and Tiwi isolates

Figure 21 (Chapter 4) shows a comparison of distribution of the cys/PoLV groups between DBLα sequence tags from Mbita and Tiwi field isolates. The frequency of group1/group1-like sequences was slightly higher among sequence tags from Tiwi than those from Mbita isolates. These sequences were from four and twenty five Tiwi and Mbita field isolates respectively. The result imply that the occurrence of group1/gruop1-like DBLα sequence tags per isolate were much fewer in the Mbita isolates compared to isolates from Tiwi study site. The presence of these sequence tags indicated of expression of group A *var* genes (Warimwe et al, 2009). Group A *var* genes are mainly expressed in patients whose immunity is naïve, especially young patients. Expression of group A*var* has been associated to severe malaria symptoms. Studies in Brazil by Kirchgatter and del Portillo (2002) associated deletion of 1-2 cysteine residues with severe non-

cerebral malaria symptoms in malaria patients. In studies from Kenya (Bull *et al*, 2007, Warimwe *et al* 2009), expression of group1 DBLα has been associated with severe malaria symptoms including cerebral malaria and convulsions. In a study by Classens *et al*, (2012) it was established that a subset of group A-like *var* genes encodes ligands that bind to brain endothelial cells. The same study also demonstrated that infected cells selected for high binding affinity to brain endothelial also bound antibodies from African children with severe malaria with a higher affinity than the non-selected infected cells. The presence and expression of group A*var* genes implied that patients from both study sites could still have progressed to develop severe malaria symptoms. However, the study subjects were recruited on the basis of lacking symptoms of complicated malaria. The DBLα sequence tags were isolated from genomic DNA rather than RNA. The exact role of the sequences in disease pathology and presentation could therefore not be ascertained. It can however be postulated that their presence in genomic DNA would be indicative that the patients could have progressed to severe forms of the disease had they not sort early treatment before complicated symptoms arose. Further the role of immunity status could have also played a big role in limiting virulence of the parasite and therefore slowing down the development of severe symptoms. This fact is attributable to the fact that samples were collected from patients residing in malaria endemic sites where transmission continues year-long those it peaks at certain times of the year.

Group2/group2-like DBLα were slightly more in field isolates from Mbita than those from Tiwi. However, both sites yielded fewer gruop2-like DBLα sequences compared to group1 sequences. Cys2 sequences are expressed by but not limited to group A *var* genes (Warime et al, 2009). While group1 sequence tags are usually indicative of expression of group A *var* genes (located near chromosome telomeres), the other cys2 sequences can belong to either group B (located

either near the telomere or centromere) or group C (located near centromere) *var* genes or group A. Their presence therefore cannot be confidently associated to development of severe or mild clinical disease.

The Cys/PoLV group3 sequences tags had an almost similar frequency in field isolates. The same distribution pattern was observed in group4 sequence tags. However it should be noted that group4/group4-like sequence tags were the most frequent in isolates from both Mbita and Tiwi study cites. There was no significant difference between the frequency of group4 DBLα sequence tags from Mbita and Tiwi isolates. Cys4 sequences have been largely associated with mild malarial symptoms. They are also associated with expression of non-group A *var* genes. While group A*var* gene expression has been found in non-immune patients, especially young children, expression of the other *var* genes occurs is semi-immune patients. Sequences from the *var* genes are most likely to aid in immune-evasion of the host immune responses by the parasite.They may also be responsible for maintaining and equilibrium between host immune responses and persistence of *P. falciparum* infection in semi-immune hosts within malaria endemic zones. This may in turn maintain a constant population of parasites in circulation that sustain host responses, which keep the individual semi-immune to malaria infection. Such responses and immune memory maintained would limit the infection by *P. falciparum* or even resolve progressive infections and thus limit the development of severe malarial symptoms such as cerebral malaria, severe malarial aneamia and respiratory distress. The average age of the study participants was 3.40 years. These patients, having been born in malaria endemic zone are likely to have had several previous infections of *P. falciparum*, therefore had likely started developing some level of immunity although may have not become totally semi-immune as compared to mature individuals within the population. This is because malaria immunity is

developed in piece-meal, since it takes a long period for a person to be exposed to the various repertoires of PfEMP1, due to differential expression of *var* genes within the parasite and the high antigenic variation that accompanies expression of these parasite antigens. Thus the high frequency of cys4 sequences could be indicative that the patients were developing some level of semi-immunity to *P. falciparum* infections. However, the clinical significance of the high frequency of group4 (together with the other cys4) DBLα sequence tags from the field could not be clearly established.

Group6 sequences consist of DBLα sequence tags containing either zero, one, three, five or more cysteine residues. These sequence tags were least among the sequences isolated from Mbita field isolates and second last among sequence tags isolated from Tiwi isolates. While studies from Brazil have indicated that deletion of 1-2 cystein residues could be associated with severe non-cerebral malaria, the significance of the other group6 sequences could not be clearly established.

5.5 CYS/PoLV groups and patient data

While available data from severe malaria studies have linked the of some *var* genes, especially group A *var* genes which express cys/PoLV1 DBLα sequences, in this study, the symptoms could not be directly associated with *in vivo* (patient) data. This could be attributed to the fact that the study only recruited subjects who had uncomplicated malaria. Severe symptoms of malaria disqualified subjects from participating in the study. In vivo data available indicate that the study subjects had had Hb levels defined by WHO as severely aneamic. It was however difficult to conclusively attribute the observed aneamia to malaria. Since the data recorded did not include the nutritional status of the patient or malaria co-infections, it was difficult to ascertain if the observe aneamic condition was only to do malaria or other factors. The aneamic conditions could also be attributable to the fact that some patients have a residual infection of *P.*

falciparum but its virulence could be limited by host immune responses. In this case, there is a slow but continues heamolysis of the infected erythrocytes, especially those that sequester at privileged sites and escape splenic clearance. Thus this continuous haemolysis reduces the erythrocyte count leading to aneamic conditions. This could be the cost that the host has to bear inoder to develop protective immunity to *P. falciparum.*

There was also a mixture of temperatures at the time of diagnosis; with some patients appearing to have had high fever while a few other patients had body temperature less than 37^0C. This could have corresponded to the cold chills that accompany malaria infection. It is however presumed that since the subjects voluntarily sought treatment, they musthave had clinical symptoms corresponding to malaria. Clinical diagnosis was confirmed by microscopic examination of giemsa stained thin and thick smears of finger prick blood samples and only those patients whose slides were malaria positive were included in the study. The observed temperature variations were therefore attributed to malaria. The fact that some subjects who were positive for malaria still had normal body temperature could have indicated that they had had previous exposures and had developed some level of immunity and tolerance to *P. falciparum* infections. These studies have been conducted within malaria endemic areas, the subject from these areas are likely to have developed some level of immunity resulting from frequent exposures to *P. falciparum*. They however remained susceptible because of their age and were likely to have manifested other malaria symptoms such as aneamia, nausea and general malaise. The average age of the patients was 3.3 years, implying that most patients were still young and had not developed high immunity to *P. falciparum*. Development of semi-immunity to *P. falciparum* occurs in piece-meal due to high antigenic variation exhibited by the parasite and mediated by among others PfEMP1 coded for by the hyper-variable *var* genes. This requires

several years of exposure to the parasite before the individual develops some level of semi-immunity to resist and/or resolved parasite infection. The fact that group 4 and group 5 PfEMP1 DBLα sequence tags were the most frequent in field isolates from both study sites suggests that these sequences could be associated to mild malarial symptoms. Presence of these sequences has been largely attributed to expression of group B and C *var* genes. These genes are associated to mild malaria and it appears their main role is antigenic variation without contributing directly to disease severity. However, it should be noted that it is not the expression level that influences the binding capacity of expressed PfEMP1 proteins. Rather, the type of DBLα sequence tag that affects where the parasitized erythrocyte binds and the strength by which it binds.

Analysis of individual sequences revealed several unique sequences that aligned along PfEMP1 sequences in the data base but had very low identity. Some of these sequences were only found in isolates from one endemic site implying there were unique to parasites circulating at that site. The sequences shown below had 70% or less identity to sequences in the data base after blast search at NCBI, indicating they were more likely to be unique to the isolates used in this study.

Since PoLV1 MFK in cys 2 DBLα has been found to be expressed in group A*var* genes, all tags bearing this motif at PoLV1 were grouped as group A*var* genes. The fact that most cys2 sequences were long and also belonged to group A*var* genes could be used to postulate that patients from which these isolates were extracted had low immunity to *P. falciparum* (Warime *et al*, 2009). Some studies have shown that DBLα sequence tags from isolates collected from subjects with poorly developed immune responses to PfEMP1 antigens tend to be long. Expression of group A*var* genes has also been shown found in younger patients with a poorly developed repertoire of immune responses. It could therefore be argued that the shear presence these sequence tags in genomic DNA was indicative of fact that group A var genes were

expressed in this patients. Since group Avar genes have been associated to severe forms of malaria, the presence in DBLα sequence tags can be used to predict the likely hood of these patients developing severe forms. Since both Mbita and Tiwi studies were in vivo drug studies (clinical trials), the inclusion criteria of study patients required each participant to be both malaria positive but have uncomplicated disease. Patients with severe malaria symptoms were excluded from the study. Thus the data obtained could not be used to conclusively determine the impact of group A*var* genes on disease presentation.

One sequence had both MFK and REY motifs at PoLV1 and PoLV2 respectfully. This could be the first observation where the two motifs are present in the same DBLα sequence.Previous studies have indicated that these two motifs are mutually exclusive. The implication the two motifs being present the same sequence tag remains unknown.

5.6 Mapping of DBLαsequence tags of field isolates from the two malaria sites into a social network

Malaria caused by *P. falciparum* remains a public health problem in sub-Saharan Africa despite the control efforts put in place by governments in this region. One big challenge to malaria control and eventual eradication has been the lack of a viable vaccine to immunize naïve individuals in malaria endemic zones. The situation remains the same despite the fact that several decades of research have been dedicated to the search for a potent vaccine to control or even prevent malaria onset. The main challenge to the search for a viable vaccine has been antigenic variation exhibited by *P. falciparum*. This variation is mediated by parasite antigens exported to surface of infected erythrocytes. These antigens are usually coded for by genes located in subtelomeric regions of the chromosomes within the parasite genome.

One group of such surface antigens that is associated to antigenic variation in *P. falciparum* is the *P. falciparum* erythrocyte membrane protein 1(PFEMP1). Coded by about 60 *var* genes per haploid genome, these genes are expressed in a mutually exclusive manner such that only one *var* gene is expressed at a given time. Differential expression, recombination and differential splicing events can help maintain antigenic variations that enable the parasite escape host immune responses. This may explain the observation that children in malaria endemic regions only develop some level of immunity after several exposures to the parasite. This piecemeal development of malaria immunity could be the reason behind the observation that malaria episodes are more severe in children below five years, who have not been exposed to a large variety of the variable surface antigens and hence have not developed a pool of memory cells to respond to the wide repertoire of the surface antigens. The fact that children growing up in malaria endemic zones eventually develop protective immunity indicates that their immune system eventually learn to recognize a wide variety of the variable surface antigens and mount protective immune responses.

In this analysis, It was noted that sequences from Mbita in western Kenya and Tiwi, coastal region clustered separately within the same giant component of the network. This suggested that there were few shared polymorphic groups within these sequences. For instance when groups 1-3 sequences from Mbita and Tiwi isolates were analysed together, they clustered separately into two distinct groups, suggesting that parasites circulating in the two study sites had distinct DBLα sequence tags that did not share PSPBs. A previous study, (Bull *et al*, 2005) observed that block sharing groups generated independently in a world network showed geographical structuring. Although the sequences used in this analysis were isolated from parasites in the same country, they seemed to share very few position specific polymorphic blocks. This could be due to

sympatric distribution of the parasites isolates from Mbita and those from Tiwi (Mbita and Tiwi are located approximately 1100KM apart). The two sites are thus separated in such a way that there is reduced chance of genetic recombination between the parasite isolates within the mosquito vector. This could be attributed to the fact that the two study sites are separated central highlands, where malaria transmission is rare with only a few episodes being reported from travelers. Since these sequences were isolated from genomic DNA, it remains to be seen if the situation is similar at the expression level. It was also not possible to determine the effect of this observed difference on the ability of host immune responses from Mbita to effectively respond to parasites from Tiwi. That is, the ability of semi-immunity acquired from exposure to *P. falciparum* parasites circulating within Mbita to protect against infection of *P. falciparum* parasites circulating in Tiwi and vice versa.

This analysis also revealed that some group 1 and group 4 sequence tags from both study sites linked to each other suggesting they shared PSPBs. This was a rare observation since other studies have not reported similar results. The analysis could however not establish if this observation was only at genomic level or also in expressed sequences. However, when tags from these sequence groups from isolates of the two sites were analysed together, they clustered separately, indicating that they did not share PSPBs. The variations observed in sequences of surface antigens that would be immune response targets increase the challenge of developing viable vaccines. Further even within the same clone in infected erythrocytes, expression of PfEMP1 surface antigens is mutually exclusive manner. This together with antigenic switching makes these surface antigens to be difficult therapeutic and vaccine targets. These observations could explain the recent observation that the first potent anti-malarial vaccine to be approved by WHO, RTS, S has a low efficacy of 30% (The RTS,S Clinical Trials Partnership, 2014). This

vaccine targets the circumsporozoite protein present on the surface of sporozoites and liver stage schizonts. Since PfEMP1 plays a major role in immune evasion and since it is genetically diverse both at intra- and inter-population level, a single vaccine with universal efficacy may remain a challenge for some time. Thus these observations indicate that DBLα sequence tags from field isolates from Mbita and Tiwi study sites in Kenya are varied and only few share PSPBs. The majority, however, do not share PSPBS, suggesting that they genetically different.

5.7 *Pfemp1* PCR product polymorphisms and clinical presentation of *P. falciparum* malaria

The primers used in the PCR have previously been used to asses PCR product polymorphism as observed on agarose gel, among laboratory isolates from different regions (Ozarka *et al*, 2007). This work suggested that the polymorphisms observed reflected parasite clone origin and presence of the band could give vital information about virulence of the parasite. This could in turn provide some insights about malaria severity. Such information could be epidemiologically important that may inform intervention strategies and epidemic preparedness. Blood samples from malaria positive patients were collected on Whatmann filter paper during a clinical field study at Mbita, western Kenya. DNA was extracted from the samples followed by PCR analysis using primers that target the DBLa domain of *Pfemp*1 genes. The PCR product was then run on agarose gel and the results related to the patient data collected during the clinical study.

From the 96 samples amplified by PCR, 60 samples (62.5%) gave a 455bp PCR product, while 37 (37.5%) did not amplify. This product size is one of those found in African laboratory clone 3D7. On considering patients' body temperature; it was found that 19 subjects representing 32% of the samples that gave a 455bp band had a body temperature of above 38^0C suggesting high fever. 24 (40%) had body temperature of below 37^0C probably corresponding to the chills that are usually experienced by malaria patients. The majority of those with body temperature above

38^0C had parasitemias above 20,000/µl. Only one subject had parasitemia of 1160/µl. 10 (17%) subjects whose samples gave the amplicon had normal body temperature. Of these, four subjects were aged five years and above while six were below five years of age. Among the subjects that were PCR negative for the primer set, only 3 had a body temperature above 38^0C and 10 had body temperature below 37^0C. 12 had body temperature around 37^0C. Calculation of Fishers' exact test gave a P-value of 0.074, suggesting a weak significance level.

The PfEMP1 protein is important in malaria pathology. These results therefore suggest that presence of the 455bp PCR product in a sample from malaria patient, although not strongly significant, could help predict severity of the disease even if it is uncomplicated. This can form a good epidemiological tool that may influence treatment and intervention measures.

5.8 CONCLUSIONS

There following conclusions were arrived at:

1. The DBLα sequence tags of PfEMP1 protein are highly polymorphic with sequence from both study sites sharing very few similarities hence mapped into different clusters when studied using a network approach.
2. Group4/group4-like and group5/group5-like DBLα sequence tags were the most prevalent in field isolates from both Mbita and Tiwi study sites. These sequences have been associated to expression of *var* genes associated with non-severe/mild malarial symptoms. These sequences could explain why most patients who sort for treatment had uncomplicated malaria symptoms. They could also be responsible for developing and maintaining malaria immunity by retaining a residual infection of P. falciparum at immune privileged sites within the body to avoid splenic clearance and also maintain a

continual exposure of malaria antigens to the immune systems for continuous immunologic memory.

3. The DBLα sequence tags from both sites exhibit high inter-genomic and intra-genomic variations. The DBLα domain mediates binding of the PfEMP1 protein to its endothelial receptors resulting in cytoadherence or to receptors on uninfected erythrocytes resulting in rosseting. Variation occurs without compromising the binding capacity of the protein. This antigenic variation implies that the parasites can bind and sequester at certain privileged sites but at the same escape of host's immune responses. While this antigenic variation is a good survival strategy for the parasite, it impedes efforts to develop and viable and functional vaccine. Development of a potent vaccine would therefore require a combination of one or more fusion proteins containing fragments that correspond to conserved regions of PfEMP 1 and other surface proteins. However, even if some form of immunity is induced by the vaccine, it will not be effective against all parasite population. That is, its efficacy level would vary with population and is likely to be protective for only a short period of time.

5.9 RECOMMENDATIONS

1. Studies need be conducted using sequences from this study to construct proteins that can be used as antigens to stimulate immune responses. Such studies can form the basis of developing a fusion protein vaccine based on DBLα sequence tags either only or in combination with conserved sequences from other surface proteins.
2. There is a need to profile genomic and mRNA DBLα sequence tags in isolates from severe, mild and asymptomatic malaria cases. This will help identify the unique sequence tags expressed in each form of disease presentation and those found in all conditions.

Such information can inform the finding of the exact sequences that play a role in the presentation of different forms of malaria pathology and those whose role may only be purely evasion of host immune responses during antigenic variation.

3. While the role played by PfEMP1 proteins in malaria pathology is well known, their interactions with other Plasmodium surface antigens is little known. There is therefore need to establish if such interactions indeed do occur and the impact of such interactions on malaria pathology. Any synergetic interactions with other parasite surface antigens would greatly boast efforts of seeking for malaria vaccine candidates and perhaps culminated in the development of a potent vaccine that may effectively help control and/or prevent malaria infections.

REFERENCES

Baratin M., Roetynck S, Pouvelle B, Lemmers C, Viebig N.K, Johansson S, Bierling P, Scherf A, Gysin J, Vivier E and Ugolini S. (2007). Dissection of the Role of PfEMP1 and ICAM-1 in the Sensing of Plasmodium falciparum-Infected Erythrocytes by Natural Killer Cells. *Plos One* (Online). *Retrieved from* www.plosone.org

Barnes K.I and White N.J. (2005). "Population Biology and Antimalarial resistance: The Transmission of Antimalarial Drug resistance *in Plasmodium falciparum*". *Elseview/Acta Tropica*; *94*(3), 230-40

Barry AE, Schultz L, Buckee CO, and Reeder JC.(2009). Contrasting Population Structures of the Genes Encoding Ten Leading Vaccine-Candidate Antigens of the Human Malaria Parasite, Plasmodium falciparum. *PLoS ONE*; 4(12):e8497.

Baruch. D. I., Gormley. J. A.,. Howard C. and Pasloske B. L. (1996) Plasmodium falciparum erythrocyte membrane protein 1 is a parasitized erythrocyte receptor for adherence to CD36, thrombospondin, and intercellular adhesion molecule 1 (malaria/variant antigen). *Proc. Natl. Acad. Sci. USA. (93),* 3497-3502.

Bengtsson A,Joergensen L, Rask T.S, Olsen R.W, Andersen M.A, Turner L,Theander G.T, Hviid L, Higgins M.K, Craig A, Brown A and Jensen T. (2013). A Novel Domain Cassette Identifies Plasmodium falciparum PfEMP1 Proteins Binding ICAM-1 and Is a Target of Cross-Reactive, Adhesion-Inhibitory Antibodies. *The Journal of Immunology, 190*: 240–249.

Berger SS, Turner L, Wang CW, Petersen JEV, Kraft M, Lusingu JPA, Mmbando B., Marquard A.M., Dominique B. A. C. Bengtsson D. B.A.C., Hviid L., Morten N.A., Theander T. G.,

and Lavstsen T. (2013). Plasmodium falciparum Expressing Domain Cassette 5 Type PfEMP1 (DC5-PfEMP1) Bind PECAM1. *PLoS ONE, 8*(7):e69117

Beshir K.B., Sepúlveda N., Bharmal J., Robinson A., Mwanguzi J., Busula A.O., de Boer J Sutherland C. Cunningham J, and Hopkins H. (2017). *Plasmodium falciparum* parasites with histidine-rich protein 2 (*pfhrp2*) and *pfhrp3* gene deletions in two endemic regions of Kenya. *Nature: Scientific Reports* | 7: *14718* | DOI:10.1038/s41598-017-15031-2

Bull. P.C., Buckee C.O., Kyes S., Kortok M.M., Thathy V., Guyah B., Stoute J.A., Newbold C.A and Marsh K. (2008). Plasmodium falciparum antigenic variation: Mapping mosaic var gene sequences onto a network of shared, highly polymorphic sequence blocks. *Molecular Microbiology*: *68* (6), 1519–1534

Bull P.C., Kyes S., Buckee C.O., Montgomery J., Kortok M.M., Newbold C.I. and Marsh K.(2007). An approach to classifying sequence tags sampled from *Plasmodium falciparum var* genes. *Molecular & Biochemical Parasitology (154,)* 98–102

Bull PC, Berriman M, Kyes S, Quail MA, Hall N, Kortok MM, Marsh K and Newbold C.I. (2005). Plasmodium falciparum Variant Surface Antigen Expression Patterns during Malaria. *PLoS Pathog*: *1*(3). Retrieved from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1287908/

Bull P.C.,Lowe B.S., Kortok M., Molyneux C.S., Newbold C.I., and Marsh K. (1998). Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nat Med 4*:358–360

Claessens A., Adams Y., Ghumra A., Lindergard G., Buchan C. C., Andisi C., Bull P. C., Mok S., Gupta A. P.,Wang C., Turner L., Arman M., Raza A. Bozdech Z. and Rowe J. A.

(2012). A subset of group A-like var genes encodes the malaria parasite ligands for binding to human brainendothelial cells. *PNAS:E1772-E1781.Retrieved from* www.pnas.org/cgi/doi/10.1073/pnas.1120461109

Chelimo K, Sumba P.O, Kazura J.W, Ofula A. V and John C.C. (2003). Interferon-gamma responses to Plasmodium falciparum liver_stage antigen-1 and merozoite surface protein-1 increase with age in children in a malaria holoendemic area in western Kenya. *Malaria Journ. (2),* 37.

Chen Q, Fernandez V, Sundstrom A, Schlichtherle M, Datta S, Hagblom P, and Wahlgren M. (1998). Developmental selection of var gene expression in Plasmodium falciparum. *Nature 394,* 392–395.

CoelhoC.H., Doritchamou J.Y.A., Zaidi I. and Duffy P.E. (2017) Advances in malaria vaccine development: report from the 2017 malaria vaccine symposium. *npj Vaccines; 2:34,* doi:10.1038/s41541-017-0035-3

D'Alessandro U and Buttiens H. (2001).“History and Importance of Antimalarial Drug Resistance.”*Tropical Medicine and International Health*; *6 (11),* 845.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res; 32(5),* 1792–7.

Ferreira M. C., Nunes M. S. and Wunderlich G.(2004).Antigenic Diversity and Immune Evasion by Malaria Parasites. *Clinical and Diagnostic Laboratory Immunology*, 987–995

Frankland. S,Adisa A., Horrocks P., Taraschi T. F, Schneider T., Elliott S.R., Rogerson S. J., Knuepfer E., Cowman A. F., NewboldC. I., and Tilley L. (2006).Delivery of the Malaria Virulence Protein PfEMP1 to the Erythrocyte Surface Requires Cholesterol-Rich Domains. *Eukaryotic Cell; (5)* 849-860

Freitas-Junior L.H., Bottius E., Pirrit L.A., Deitsch K.W., Scheidig C, Guinet F, Nehrbass U, Wellems TE, and Scherf A. (2000). Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of P. falciparum. *Nature 407,* 1018–1022

Fruchterman TMJ, andReingold EM.(1991).Graph drawing by force-directed placement. Softw Pract Exp;21(11):1129–64

Furuya K. and Niki H. (2009).Isolation of heterothallic haploid and auxotrophic mutants of *Schizosaccharomyces japonicas*.*Yeast*. *Wiley InterScience*26: 221–233

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DM, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, and Barrell B. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature; 419(6906),* 498–511

Ho F., Baker J., Lee N., Luchavez L., Ariey F., Nhem S., Oyibo W., Bell D., González I., Chiodini P., Gatton M.L., Cheng Q. and McCarthy J.S. (2014). Circulating antibodies

against Plasmodium falciparum histidine-rich proteins 2 interfere with antigen detection by rapid diagnostic tests. *Malaria Journal:13*: 480. *Retrieved from* http://www.malariajournal.com/content/13/1/480

Hobley L., Lerner T. R., Williams L. E., Lambert C., Till R., Milner D. S., Basford S. M., Capeness M. J., Fenton A. K., Atterbury R. J., Harris M and Sockett E. R. (2012). Genome analysis of a simultaneously predatory and prey-independent, novel *Bdellovibrio bacteriovorus* from the River Tiber, supports *in silico* predictions of both ancient and recent lateral gene transfer from diverse bacteria. *BMC Genomics 13*: 670 doi:10.1186/1471-2164-13-670

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

Jacobs P, Radzioch D and Stevenson MM. (1996) A Th1-associated increase in tumor necrosis factor alpha expression in the spleen correlates with resistance to blood-stage malaria in mice. *Infect. Immun., 64(2),* 535-541.

Janes J. H., Wang C. P., Levin-Edens E., Vigan-Womas I., Guillotte M., Melcher M, Mercereau-Puijalon O, and Smith J.D. (2011). Investigating the Host Binding Signature on the *Plasmodium falciparum* PfEMP1 Protein Family. *PLoS Pathog 7(5*), e1002032. doi:10.1371/journal.ppat.1002032

Kamada T, Kawai S. (1989). An Algorithm for Drawing General Undirected Graphs. *Inf Process Lett. 31(1),*7–15.

Kaslow D.C and Biernaux S. (2015). RTS,S: Toward a first landmark on the Malaria Vaccine Technology Roadmap. *Vaccine; 33*, 7425–7432

Kraemer S.M., Kyes, S.A, Aggarwal, G1, Springer A.L, Nelson S.O, Christodoulou Z, Smith L.M, Wang W, Levin E, Newbold C.I, Myler P.J and Smith J.D. (2007). Patterns of gene recombination shape *var* gene repertoires in *Plasmodium falciparum:* comparisons of geographically diverseIsolates *BMC Genomics. Retrieved from* http://www.biomedcentral.com/1471-2164/8/45.

Kumar R.A., Sridevi K., Kumar N.V., Nanduri S. and Rajagopal S. (2004) Anticancer and immunostimulatory compoundsfrom Andrographis paniculata. *Journal Ethnopharmacollogy: 92(2-3),* 291–5.

Kyes S. A, Kraemer S.M, and Smith J.D.(2007).Antigenic Variation in *Plasmodium falciparum*: Gene Organization and Regulation of the *var* Multigene Family. *Eukaryotic Cell, 6(9),* 1511-1520

Nielsen C.M., Vekemans J., Lievens M., Kester K.E., Regules J.A., and Ockenhouse C.F. (2018). RTS,S malaria vaccine efficacy and immunogenicity during Plasmodium falciparum challenge is associated with HLA genotype. *Vaccine 36,* 1637–1642

Ozarkar A.D, Prakash D., Deobagkar D.N,. and Deobagkar D.D. (2007). Distinct polymorphism of DBLa domain of the *var* gene in laboratory-cultured and clinical field isolates of *Plasmodium falciparum* malaria parasites. *Current science, 93 (2)*

Plebanski M, Proudfoot O, Pouniotis D, Coppel V.A and Flannery G. (2002). Immunogenetics and the design of Plasmodium falciparum vaccines for use in malaria endemic populations.*J. Clin. Invest. 110:* 295-301.

Rask TS, Hansen DA, Theander TG, Gorm Pedersen A, Lavstsen T. (2010) Plasmodium falciparum Erythrocyte Membrane Protein 1 Diversity in Seven Genomes – Divide and Conquer. *PLoS Comput Biol 6(9):* e1000933. doi:10.1371/journal.pcbi.1000933

RTS,S Clinical Trials Partnership. (2015). Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. *Lancet,* 386: 31–45

Scherf A., Lopez-Rubio J.J, and Riviere L. (2008).Antigenic Variation in *Plasmodium falciparum. Annu. Rev. Microbiol; 62*, 445-470

Sinha A. K, Yadav S, Islam M. N., Yadav B.and Aryal B.(2012). Assessment of Anemia Based on Hb Levels in Children of 2 to 12 Years Age Group in Biratnagar, Nepal. International *Journal of Pharmaceutical & Biological Archives; 3(3)*, 552-554

Smith JD, Subramanian G, Gamain B, Baruch DI, Miller LH. (2000). Classification of adhesive domains in the *Plasmodium falciparum* erythrocyte membrane protein 1 family. *Mol Biochem Parasitol*; *110*, 293-310.

Smith JD,. Chitniq't C.E., Craig A.G., Roberts D.J., Hudson-Taylor D.E., Peterson D. S., Pinches R., and Newbold C.and Louis H. Miller. (1995). Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell 82*, 101–110.

Su XZ, Heatwole VM, Wertheimer SP, Guinet F, Herrfeldt JA, Peterson DS, Ravetch JA, Wellems TE. (1995). The large diverse gene familyvar encodes proteins involved in cytoadherence and antigenicvariation of *Plasmodium falciparum*-infected erythrocytes. *Cell, 82*:89-100.

Sutherland C.J,Lansdell P., Sanders M., Muwanguzi J., van Schalkwyk D.,Kaur H., Nolder D., Tucker J., Bennett H.M., Otto T.D., Berriman M., Patel T. A., Lynnn R., Gkrania-Klotsas E. and Chiodini P/L. (2017). *Pfk13*-independent treatment failure in four imported cases of *Plasmodium falciparum* malaria given artemether-lumefantrine in the UK. *Antimicrob. Agents Chemother*. doi:10.1128/AAC.02382-16

Tamura K., Peterson D., Peterson N.,Stecher G., Nei M., and Kumar s. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol. 28(10)*, 2731–2739.

Warimwe G.W, Keane T. M, Fegana G,Musyokia J.N, Newton C. R,Pain A, Berriman M, Marsh K, and Bull P.C. (2009). Plasmodium falciparum var gene expressionis modified by host immunity. *PNAS, 106 (51)*, 21801-21806

WHO.(2016). World Malaria Report, 2016. World Health Organization

WHO.(2011).Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity.Vitamin andMineral Nutrition Information System. Geneva,WorldHealth Organization.(WHO/NMH/NHD/MNM/11.1). *Retrieved from* http://www.who.int/vmnis/indicators/haemoglobin.pdf

WHO. (2011). World Malaria Report, 2011. World Health Organization

APPENDICES

Appendix I: *In vivo* data of samples from Mbita

SAMPLE CODE	455BP Band	PATIENT'S TEMPERATURE	HEAMOGLOBIN LEVEL	PARASITEMIA
K 025	VE+++	**38.8**	**12.9**	**50400**
K 026	**NEG**	**38.5**	**12.7**	**1700**
K 040	VE+	**39.3**	**7.7**	**44550**
K 041	**NEG**	**35.1**	**10.8**	**20250**
K 054	**NEG**	**37**	**12**	**3500**
K 303	**NEG**	**37.3**		**1000**
K 150	**NEG**	**36.9**	**4.2**	**16520**
K 291	**NEG**	**37.9**	**7.1**	**1120**
K 344	**NEG**			
K 385	**NEG**			
K 544	NEG	**37.5**		**1200**
K 598	NEG			
K 782	VE++			
K 701	VE++	**36.1**		**16960**
K 774	VE+	**37.3**		**4480**
K 815	VE+++	**38.6**		**40089**
K 840	VE+++			
K 875	VE++			
K 878	VE+	**35.8**		**3480**
K 881	NEG	**35.7**		**13720**
K 1010	NEG	**37**		**8880**
K 1149	NEG			
K 1152	NEG	**36.5**		
K 1295	VE+			
K 1302	VE++	**38.8**		**62460**
K 1307	VE++	**38.1**		**26320**
K 1348	VE+++	**37.1**		**76560**
K 1379	VE+++	**38.3**		**44800**
K 1438	NEG	**38.5**		**3520**
K 1476	NEG			
K 1478	VE+++	**37.1**		**38280**
K 1943	VE+	**36.3**		**9440**
K 2023	NEG	**37.7**		**40080**
K 2024	VE+	**37.1**		**18240**
K 3013	VE+	**38.6**		**32000**
K 3077	VE+	**37.8**	**7.2**	**12160**
K 3404	VE++	**37.0**	**7.8**	**2080**
K 3901	VE+++	**40.4**		**60080**
K 396	**VE+++**			
K122	**VE++**	**37.4**	**9.6**	**28600**

K225	NEG	**37.2**	**8.5**	**4640**
K259	NEG	**36.8**	**6**	**6720**
K296	**VE+++**	**37.2**	**5.2**	**3560**
K395	**VE++**	**36.1**		**24800**
K398	NEG	**36**	**6.9**	**1280**
K399	**VE+++**	**36**	**7.4**	**52000**
K404	**VE++**	**35.6**	**7.4**	**19200**
K1431	**VE+++**			
K656	VE+	**36.7**		**6600**
K802	**VE+++**	**38.6**		**34780**
K815	**VE++**	**38.6**		**40080**
K878	VE+	**35.8**		**3480**
K171	NEG	**36.4**	**5.2**	**1960**
K187	V+	**36.2**	**7.7**	**7320**
K237	VE+	**37**	**9**	**37760**
K1119	**VE+++**	**37.3**		**55920**
K1129	NEG	**36**		**6720**
K1167	NEG	**36.4**		**2080**
K1176	**VE+++**	**39.5**		**28320**
K1181	VE++	**38.9**	**7.6**	**20320**
K1275	NEG	**36.1**	**6.2**	**20560**
K1750	**VE++++**	**36.3**	**5.9**	**37600**
K3014	**VE++**	**36.6**	**5.4**	**17760**
K3018	**VE++**	**36.7**	**6.8**	**0**
K3060	**VE++++**	**39.4**	**5.8**	**20080**
K3067	NEG	**36.4**	**5.0**	**11520**
K3176	VE+	**36.7**		**9200**
K3418	NEG			
K1399	**VE++++**			
K 1574	**VE+++**	**36.1**		
K 3120	**VE+++**	**36.7**	**5.0**	**56360**
K3831	NEG	**39.1**		**45520**
K3648	VE+	**39.2**		**39520**
K3146	VE+	**36.8**		**0**
K3679	NEG	**36.2**		**pf+**
K3782	NEG	**36.6**		**12160**
K3302	NEG	**36.8**		**0**
K3286	**VE++**	**36.8**		**60920**
K3329	NEG	**40**		**60080**
K3793	NEG	**36.8**	**8.6**	**7760**
K3445	NEG	**37.4**		**2960**
K3403	VE++			
K3215	NEG			
K3309	VE+	**36.2**		**14240**
K3304	NEG	**36.9**		**41120**
K3263	VE+++	**36.4**		**20400**

K3248	NEG			**36360**
K3122	VE++	**37.3**	**6.9**	**14560**
K1499	VE++	**38.4**	**6.2**	**1160**
K1079	VE++++	**38.2**		**28080**
K1073	VE+++			
K1067	VE+++380	**37.3**		**14160**
K1296	NEG			
K3212	VE+	**38.9**		**34240**
K1142	VE++++	**34240**		**34240**
K1238	VE+++			**24880**
K1025	VE+++	**36.1**		**20720**
1238	V++	38.7	6.9	24880
1048	V++	36.4	7.1	16320
1076	V++			
09	V+++	38.6		21100
1226	V++	39.4	7.1	19200
027	V+	36.1		850
3176	V++			
3640	V++	38.8		

Appendix II: Amino acid code

Amino Acid Code:	Three letter Code:	Amino Acid:
A	Ala	Alanine
B	Asx	Aspartic acid or Asparagine
C	Cys	Cysteine
D	Asp	Aspartic Acid
E	Glu	Glutamic Acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
X	Xaa	Any amino acid
Y	Tyr	Tyrosine

Z	Glx	Glutamine or Glutamic acid

Appendix III: Blast results for group6 sequence tags from mbita isolates

IFGDIYNEVMRGTNGQTLQKRYKKDKQTGNYYKLREDWWTANRATIWEALTCHAPPD
AQYFRQTCCTGTGTQGRCRCGDGKSKVGKNETDQVPTYFDYVPQYLR

erythrocyte membrane protein 1 type k [Plasmodium falciparum]
Sequence ID: gb|AAC62726.1|Length: 136Number of Matches: 1
Related Information
Range 1: 29 to 136GenPeptGraphicsNext MatchPrevious Match

Score	Expect	Method	Identities	Positives	Gaps
164 bits(414)	2e-49	Compositional matrix adjust.	80/109(73%)	90/109(82%)	6/109(5

```
Query  1    IFGDIYNEVMRGTNGQTLQKRYKKDKQTGNYYKLREDWWTANRATIWEALTCHAPPDAQY  60
            IFGDIYNEVMRG NGQ ++ RYKKD + GN+++LREDWWTANR  IW+A+TC AP +A+Y
Sbjct  29   IFGDIYNEVMRGKNGQ-IEARYKKDDEDGNFFQLREDWWTANRQDIWKAITCGAPNNAKY  87

Query  61   FRQTCC-----TGTGTQGRCRCGDGKSKVGKNETDQVPTYFDYVPQYLR  104
            FRQTC     TGT T  RCRCGDGKSKVG NETDQVPTYFDYVPQYLR
Sbjct  88   FRQTCGGDNEKTGTLTPSRCRCGDGKSKVGTNETDQVPTYFDYVPQYLR  136
```

erythrocyte membrane protein [Plasmodium falciparum]
Sequence ID: emb|CAJ40209.1|Length: 137Number of Matches: 1
Range 1: 32 to 137GenPeptGraphicsNext MatchPrevious Match

Score	Expect	Method	Identities	Positives	Gaps
139 bits(349)	1e-39	Compositional matrix adjust.	71/108(66%)	82/108(75%)	6/108(5

```
Query  1    IFGDIYNEVMR-GTNGQTLQKRYKKDKQTGNYYKLREDWWTANRATIWEALTCHAPPDAQ  59
IF  I+  V + GTN + L+ RY+KD    NYY+LREDWWTANR T+WEALTC A  +A+
Sbjct  32   IFKQIHGNVTKVGTNAEELKARYQKDGD--NYYQLREDWWTANRHTVWEALTCDARDNAE  89

Query  60   YFRQTCC---TGTGTQGRCRCGDGKSKVGKNETDQVPTYFDYVPQYLR  104
            YFR+TC    TGT T  +CRCGDGKSK G NETDQVPTYFDYVPQYLR
Sbjct  90   YFRRTCGDEKTGTLTPNKCRCGDGKSKAGTNETDQVPTYFDYVPQYLR  137
```

EMP1 [Plasmodium falciparum]
Sequence ID: gb|ACT21596.1|Length: 132Number of Matches: 1
Related Information
Range 1: 33 to 132GenPeptGraphicsNext MatchPrevious Match

Score	Expect	Method	Identities	Positives	Gaps
135 bits(341)	2e-38	Compositional matrix adjust.	67/105(64%)	82/105(78%)	6/105(5

```
Query  1    IFGDIYNEVMRG-TNGQTLQKRYKKDKQTGNYYKLREDWWTANRATIWEALTCHAPPDAQ  59
            IFG+IYNE+  G TN +  ++RYK D  T NYY+LREDWWTANR T+W+A+TC AP D++
Sbjct  33   IFGNIYNELTSGKTNKEAAKERYKGD--TKNYYQLREDWWTANRETVWKAITCDAPDDSK  90

Query  60   YFRQTCCTGTGTQGRCRCGDGKSKVGKNETDQVPTYFDYVPQYLR  104
            YFR TC TGT T  +CRC   K + GK++TDQVPTYFDYVPQYLR
Sbjct  91   YFRHTCGTGTPTNDKCRC---KDENGKDDTDQVPTYFDYVPQYLR  132
```

IFGDIYNDLTSGRNGKKWAEAKNTTKMILQIIINYEKIGWTANRETVWKAITCDADGSYF
HATCSERNGGCSQANNKCRCKDENGKNTDQVPTYFDYVPQYLR
PfEMP1 DBLa, partial [Plasmodium falciparum]

Sequence ID: emb|CCF76422.1|Length: 128Number of Matches: 1
Related Information
Range 1: 32 to 128GenPeptGraphicsNext MatchPrevious Match

Score	Expect	Method	Identities	Positives	Gaps
142 bits(357)	7e-41	Compositional matrix adjust.	73/103(71%)	83/103(80%)	6/103(5

```
Query  1    IFGDIYNDLTSGRNGKKWAEAKNTTKMILQIIINYEKIGWTANRETVWKAITCDADGSYF  60
```

```
        IF +I++ L++ +NG K    +N      Q+  ++    WTANRETVWKAITCDADGSYF
Sbjct  32   IFKEIHSGLST-KNGVK-DRYQNDGDNYFQLREDW----WTANRETVWKAITCDADGSYF  85

Query  61   HATCSERNGGCSQANNKCRCKDENGKNTDQVPTYFDYVPQYLR  103
            HATCSERNGGCSQANNKCRCKDENGKNTDQVPTYFDYVPQYLR
Sbjct  86   HATCSERNGGCSQANNKCRCKDENGKNTDQVPTYFDYVPQYLR  128
```

erythrocyte membrane protein 1 [Plasmodium falciparum]

Sequence ID: gb|ACA30376.1|Length: 128Number of Matches: 1

Related Information

Range 1: 32 to 128GenPeptGraphicsNext MatchPrevious Match

Score	Expect	Method	Identities	Positives	Gaps
141 bits(356)	8e-41	Compositional matrix adjust.	73/103(71%)	83/103(80%)	6/103(5

```
Query  1    IFGDIYNDLTSGRNGKKWAEAKNTTKMILQIIINYEKIGWTANRETVWKAITCDADGSYF  60
            IF +I++ L++ +NG K    +N      Q+  ++    WTANRETVWKAITCDADGSYF
Sbjct  32   IFKEIHSGLST-KNGVK-DRYQNDGDNYFQLREDW----WTANRETVWKAITCDADGSYF  85

Query  61   HATCSERNGGCSQANNKCRCKDENGKNTDQVPTYFDYVPQYLR  103
            HATCSERNGGCSQANNKCRCKDENGKNTDQVPTYFDYVPQYLR
Sbjct  86   HATCSERNGGCSQANNKCRCKDENGKNTDQVPTYFDYVPQYLR  128
```

Appendix IV: Blast results of a rare DBL*α* sequence to confirm it was of parasite origin

hypothetical protein [Enterococcus faecalis]
Sequence ID: ref|WP_010712134.1|Length: 135Number of Matches: 1
See 3 more title(s)
Related Information
Identical Proteins-Proteins identical to the subject
Range 1: 48 to 96GenPeptGraphicsNext MatchPrevious Match

Score	Expect	Method	Identities	Positives	Gaps
35.8 bits(81)	1.3	Compositional matrix adjust.	12/49(24%)	29/49(59%)	2/49(4

```
Query  56   NLDRHLCSDLLRRSSSL--HYVFSRKFLSNVVVIVLFLVDPYHYNQEID  102
            N +R    D++ + +++  +Y   +K++ N+ ++ LF + PY  + E+D
Sbjct  48   NEERKKAIDIIEKENNILVYYAIEQKYMGNITMLYLFYISPYEEDWEMD  96
```

DownloadGenPeptGraphicsNextPreviousDescriptions
hypothetical protein Bdt_2815 [Bdellovibrio bacteriovorus str. Tiberius]
Sequence ID: ref|YP_007023763.1|Length: 217Number of Matches: 1
See 2 more title(s)
Related Information
Gene-associated gene details
Identical Proteins-Proteins identical to the subject
Range 1: 38 to 104GenPeptGraphicsNext MatchPrevious Match

Score	Expect	Method	Identities	Positives	Gaps
35.0 bits(79)	4.7	Compositional matrix adjust.	21/67(31%)	39/67(58%)	8/67(11%

```
Query  38   CIPYFFLLSVHQSSRNLKNLDR--HLCSDLLRRSSSLH--YVFSRKFLSNVVVI----VL  89
            +PYF LLS+ +  R LK++     +C+D+  + SSL    +F+ K ++ + V     +L
Sbjct  38   TLPYFNLLSIEEQERVLKDILTFFQVCTDVKSQGSSLKDTRLFTEKAIARIGVTADPAIL  97

Query  90   FLVDPYH  96
            L++P+H
Sbjct  98   DLIEPHH  104
```

DownloadGenPeptGraphicsNextPreviousDescriptions
DinB translesion DNA repair polymerase [Schizosaccharomyces japonicus yFS275]
Sequence ID: gb|EEB08917.2|Length: 524Number of Matches: 1
Related Information
Range 1: 303 to 340GenPeptGraphicsNext MatchPrevious Match

Score	Expect	Method	Identities	Positives	Gaps
34.3 bits(77)	8.6	Composition-based stats.	15/38(39%)	22/38(57%)	0/38(0%

```
Query  52   RNLKNLDRHLCSDLLRRSSSLHYVFSRKFLSNVVVIVL  89
            + L +L+ H C D+LRR   L+YVF  K   N++   L
Sbjct  303  QELSSLNVHNCGDILRRKGYLYYVFQEKSFQNLLSYAL  340
```

DownloadGenPeptGraphicsNextPreviousDescriptions

DNA polymerase kappa [Schizosaccharomyces japonicus yFS275]

Sequence ID: ref|XP_002175210.1|Length: 532Number of Matches: 1

Related Information

Gene-associated gene details

Range 1: 311 to 348GenPeptGraphicsNext MatchPrevious Match

Score	Expect	Method	Identities	Positives	Gaps
34.3 bits(77)	9.4	Composition-based stats.	15/38(39%)	22/38(57%)	0/38(0%

Query 52 RNLKNLDRHLCSDLLRRSSSLHYVFSRKFLSNVVVIVL 89
+ L +L+ H C D+LRR L+YVF K N++ L
Sbjct 311 QELSSLNVHNCGDILRRKGYLYYVFQEKSFQNLLSYAL 348

APPENDIX V: Sequences without well-defined PoLV2 motifs

DIGDIVRGRDLYLGKKEKKNQTETERDQLEKKLKEIFKEIHKDVMKTNGAQERYIDDAK
GGDFFSIKRSIGGRRIEKQYGKAITCHAPKEANYFIKTACNVGKGLVVNAIALVEMFPTY
FDYVPQFLR

DIGDIVRGRDLYRGNNGKDKLQDIKKIFKEIHDKLDNSIKSQYDDTAKTIQIYEKIGGLRI
GTPCGKLSRATLRVVIIIFDEHVVVHKIQLRLKITADVFLGRCPLRVFDYVPQFLR

The blast result is shown below.

Erythrocyte membrane protein [Plasmodium falciparum]
Sequence ID: gb|ACZ81978.1|Length: 124Number of Matches: 1
Related Information
Range 1: 1 to 124GenPeptGraphicsNext MatchPrevious Match

Score	Expect	Method	Identities	Positives	Gaps
162 bits(411)	1e-48	Compositional matrix adjust.	84/128(66%)	96/128(75%)	4/128(3

```
Query  1    DIGDIVRGRDLYLGKKEKKNQTETERDQLEKKLKEIFKEIHKDVMKTNGAQERYIDDAKG  60
            DIGDIVRG+DLYLG    N+ + +R +LE+KLKEIFKEIHKDVMKTNGAQERYIDDAKG
Sbjct  1    DIGDIVRGKDLYLGYD---NKEKEQRKKLEQKLKEIFKEIHKDVMKTNGAQERYIDDAKG  57

Query  61   GDFFSIKRSIGGRRIEKQYGKAITCHAPKEANYFIKTACNVGKGLVVNAIALVEMFPTYF  120
            GDFF ++       E  + KA+ CHAPKEANYFIKTACNVGKG      +    PTYF
Sbjct  58   GDFFQLREDWWTANRETVW-KALICHAPKEANYFIKTACNVGKGTNGQCHCIGGDVPTYF  116

Query  121  DYVPQFLR  128
            DYVPQ+LR
Sbjct  117  DYVPQYLR  124
```

Appendix VI: Bash script used to convert SSF files to Fasta then to Fastq and finally to Fasta Qual

```
#!/usr/bin/env python

#Takes a single FASTQ file and splits to .fasta + .qual files
import sys from Bio import SeqIO

if len(sys.argv) == 1:
        print "Please specify a  single .fastq file to convert."
        sys.exit()

filetoload = sys.argv[1]
basename = filetoload

#Chop the extension to get names for output files
if basename.find(".") != -1:
        basename = '.'.join(basename.split(".")[:-1])

SeqIO.convert(filetoload, "fastq", basename + ".fasta", "fasta")
SeqIO.convert(filetoload, "fastq", basename + ".qual", "qual")

#!/bin/sh

sff2fastq -o M1104_454Reads.RL16.fastq M1104_454Reads.RL16.sff

./fastq_to_fasta.py M1104_454Reads.RL16.fastq

sff2fastq -o M1119_454Reads.RL11.fastq M1119_454Reads.RL11.sff

./fastq_to_fasta.py M1119_454Reads.RL11.fastq
```

Appendix VII: Ethical Approvals

KENYA MEDICAL RESEARCH INSTITUTE

P.O. Box 54840 - 00200 NAIROBI, Kenya
Tel: (254) (020) 2722541, 2713349, 0722-205901, 0733-400003; Fax: (254) (020) 2720030
E-mail: director@kemri.org info@kemri.org Website:www.kemri.org

KEMRI/RES/7/3/1 **March 14, 2011**

TO: **EDWIN KIMELI TOO (PRINCIPAL INVESTIGATOR)**

THRO': **DR. KIMANI GACHUHI,**
THE DIRECTOR, CBRD,
NAIROBI

15/4/2011

RE: **SSC PROTOCOL NO. 1955 (2nd *RE- SUBMISSION*): COMPARING EFFICACY OF ARTEMETHER LUMEFANTRINE WITH DOU-COTECXIN IN THE CLEARANCE OF GAMETOCYTES IN UNCOMPLICATED PLASMODIUM FALCIPARUM MALARIA.**

Make reference to your letter dated March 14, 2011 received on the same day. Thank you for your response to the issues raised by the Committee. This is to inform you that the issues raised during the 186th meeting of the KEMRI/ERC meeting held on 15th February 2011, have been adequately addressed.

Due consideration has been given to ethical issues and the study is hereby granted approval for implementation effective this **14th day of March 2011,** for a period of twelve (12) months.

Please note that authorization to conduct this study will automatically expire on **13th March 2012.** If you plan to continue with data collection or analysis beyond this date, please submit an application for continuing approval to the ERC Secretariat by **15th December 2011.**

You are required to submit any amendments to this protocol and other information pertinent to human participation in this study to the ERC prior to initiation. You may embark on the study.

Yours sincerely,

RCKithinji

Caroline Kithinji,
FOR: SECRETARY,
KEMRI/NATIONAL ETHICS REVIEW COMMITTEE

Appendix VIII: Publication Abstracts

Journal of Natural Sciences Research
ISSN 2224-3186 (Paper) ISSN 2225-0921 (Online)
Vol.5, No.7, 2015

PfEMP1 DBLα Sequence Tags in Genomic DNA of P. falciparum Field Isolates from Two Malaria Endemic Sites in Kenya

Francis W. Makokha[1,2,3,4*] Sabah A. Omar[4] Francis T. Kimani[4] Gabriel Magoma[4] Rahma Udu[4,5]
Edwin Too[4] Nathan Shaviya[4] Charity Hungu[4]
1.Directorate of Research and Development, Mount Kenya, P.O Box 342-01000, Thika, Kenya
2.Institute of Tropical Medicine and Infectious Diseases, Jomo Kenyatta University of Agriculture and Technology, P.O Box 62000-00200, Nairobi Kenya
3.Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology, P.O Box 62000-00200, Nairobi Kenya
4.Centre for Biotechnology and Research Development, Kenya Medical Research Institute, P.O Box 54840-00200, Nairobi, Kenya
5.Masinde Muliro University of Science and Technology, Department of Biomedical Sciences, P.O 190-50100, Kakamega, Kenya
6.Department of Biochemistry, Pwani University, P.O Box 195-80108, Kilifi, Kenya
Email: makokhafw@gmail.com

Abstract

Background

Malaria caused by *Plasmodium falciparum* remains a major cause of childhood morbidity and mortality in sub-Saharan Africa. PfEMP1 protein, coded for by a family of about sixty variant *var* genes, is a parasite protein found on infected erythrocyte membrane. PfEMP1 protein mediates cytoadherence of infected erythrocytes on endothelial cells which may lead to severe symptoms of malaria. Although PCR amplification of the whole gene is difficult due to high variability, primers targeting the DBLα domain have been designed and used to study *pfemp1* genes. This objective of this study was to establish the distribution of DBLα sequence tags in isolates of *Plasmodium falciparum* from two malaria endemic sites in Kenya.

Method:

DNA extracted from field isolates collected from Mbita (Western Kenya) and Tiwi (Coastal region) was used to isolate and amplify DBLα domain sequence tags of *pfemp1* by PCR. PCR products were sequenced by 454 next generation sequencing. After assembly, the translated protein sequences (GenBank KP085750-KP087726) were then aligned in Mega 5.2 and classified into cys/PoLV groups based on the number of cysteine residues and the motifs at PoLV1 and PoLV2 within the sequence tag. Six sequence groups were found in sequences from both endemic sites. Group 4 sequences were the most prevalent (37.35% and 57.07% in isolates from Mbita and Tiwi respectively) in the isolates from both sites. Sequence tags from Tiwi had a higher proportion of cys2 (group 1 and 2) than sequences from Mbita although individual group 2 sequence tags were slightly higher in Mbita tags. Similarly the proportion of groups 5 and 6 sequence tags was higher in sequence tags from Tiwi than those from Mbita.

Conclusion

In conclusion, the frequency of the different cys/PoLV groups of DBLα sequence tags at both endemic sites follow almost similar pattern with group four sequence tags being the majority among the sequence tags isolated from patient isolates from both study sites. However, in the absence of expression data, the impact of this genomic distribution pattern on malaria pathology remains unknown.

Key Words: Malaria, PfEMP1, cys/PoLV, DBLα, var, Sequence tags

Journal of Natural Sciences Research
ISSN 2224-3186 (Paper) ISSN 2225-0921 (Online)
Vol.5, No.11, 2015

Mapping of DBL*α* Sequence Tags of Field Isolates from Two Malaria Endemic Sites in Kenya

Francis W. Makokha[1,2*] Peter C. Bull[4] Francis T. Kimani[5] Charity Hungu[5] Nathan Shaviya[6] Gabriel Magoma[3] Sabah O. Ahmed[4]

1.Directorate of Research and Development, Mount Kenya University, P.O Box 342-01000, Thika, Kenya
2.Institute of Tropical Medicine and Infectious Diseases, Jomo Kenyatta University of Agriculture and Technology, P.O Box 62000-00200, Nairobi Kenya
3.Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology, P.O Box 62000-00200, Nairobi Kenya
4. Kenya Medical Research Institute-Wellcome Trust, P.O Box 230-80108, Kilifi
5.Centre for Biotechnology and Research Development, Kenya Medical Research Institute, P.O Box 54840-00200, Nairobi, Kenya
6.Masinde Muliro University of Science and Technology, Department of Biomedical Sciences, P.O 190-50100, Kakamega, Kenya

Abstract
Plasmodium falciparum Erythrocyte Membrane Protein 1 (PfEMP1) found on the surface of infected erythrocytes (IEs) mediate antigenic variation during *P. falciparum* infection enabling the parasite evade host immune responses and prolong infection. These molecules mediate binding of IEs to host endothelial cells and uninfected erythrocytes. Cytoadhesion of IE to host cells leads to sequestration in tissues and PfEMP1 is thought to play an important role in parasite virulence. Here we analysed 1725 sequence tags sampled from the DBLa region of PfEMP1 encoding "*var*" genes from 27 patients in two different geographical regions in Kenya, Mbita in Western Kenya and Twiga on the Kenyan coast. The objective of this study was to construct a network to assess the extent of shared position specific polymorphic blocks (PSPBs) in sequences isolated from genomic DNA of field isolates from the two malaria endemic sites in Kenya.. Sequences from Mbita study site and those from Tiwi largely clustered into separate giant networks with only a limited number of sequences from the two sites linking to each other. This observation suggests that the parasite populations from the two endemic sites could be genetically varied and that PfEMP1 sequencing could be a useful tool of understanding the genetics of parasite populations. Thus the network approach of studying relationships between DBLα sequences is a useful tool of uncovering the genetic structure of parasite populations circulating in different malaria endemic regions.
Keywords: PfEMP1, Networks, Position Specific Polymorphic Groups, DBLα, Malaria, *P. falciparum*